ALL THAT GLITTERS IS NOT GOLD: IDENTIFICATION, EXTRACTION, AND APPLICATIONS OF PRODUCT INFORMATION FROM ONLINE CONSUMER REVIEWS

Jorge E. Fresneda[1]
Martin Tuchman School of Management
New Jersey Institute of Technology
184-198 Central Ave, Newark, NJ 07103. USA
fresneda@njit.edu

David Gefen
LeBow College of Business
Drexel University.
3220 Market St, Philadelphia, PA 19104 USA
gefend@drexel.edu

Chelsey H. Hill
Feliciano School of Business
Montclair State University.
1 E Normal Ave, Upper Montclair, NJ 07043 USA
hillc@montclair.edu

ABSTRACT

Existing research showed that voting the helpfulness of online reviews may not be the most appropriate measure of the quality of the information provided by those reviews. This paper suggests a new approach to rank online consumer reviews based on three variables of the relative information content of the textual portion of each review and the readability of that text, as reflecting how well the information is conveyed. The study shows the value of this approach over and above previous measures based on human votes. The study then extends that approach to the evaluation of the contribution of individual reviewers. The methodology was tested on 49,143 online reviews from three dissimilar product categories available on Amazon.com: appliances, computers, and luxury beauty. The implications of the proposed methodology and its potential applicability are discussed.

Keywords: Online consumer reviews; e-WOM; e-Commerce; Review helpfulness; Reviewer expertise; Natural language processing; Information entropy

1. Introduction

Retailers acknowledge the importance of consumer reviews for online sales, recognizing that customers often rely on those reviews to assess products they cannot physically touch before making a purchase. Online reviews are a relevant source of purchase information for customers that helps them overcome this lack of physical interaction (You et al., 2015) and assists them in making purchase decisions (Zhu & Zhang, 2010). Previous research has indeed shown that online reviews affect consumers' behavior (e.g. Dellarocas et al., 2007; Floyd et al., 2014; Wang et al., 2015; You et al., 2015), finding that reviews that were assessed as 'helpful' had a greater impact on product sales (Chen et al., 2008). Despite the importance of online reviews as a source of pre-purchase information, some question the ability of these reviews to accurately and objectively reflect the experience of buyers (De Langhe et al., 2016; Kashyap et al., 2023; Schmalz et al., 2018; Wen & Ha, 2024; Zhuang et al., 2018).

This concern about the accuracy and the validity of the information contained in online reviews has also gained mass-media attention (Box & Croker, 2018; Dwoskin & Timberg, 2018; Engle, 2023) and even the Federal Trade Commission's attention (Viswanathan, 2022). Consumer skepticism toward online reviews has grown too (Wen & Ha, 2024), as online buyers question the credibility of the information provided and the reviewers' motivation to share

Cite: Fresneda, J. E., Gefen, D., & Hill, C. H. (2026). All that glitters is not gold: Identification, extraction, and applications of product information from online consumer reviews. *Journal of Electronic Commerce Research, 27*(4)**.**
[1] Corresponding author

information (Shan, 2016). To address those concerns, some online retailers started implementing actions to reassure the validity of the information contained in online reviews (e.g. Aral, 2014; ePlaybooks, 2023; Rubin, 2015). Presumably, by doing so, retailers attempt to recover customers' trust in online reviews as a reliable source of pre-purchase information.

One of the areas that raised most concerns was the way in which the information provided by reviews and reviewers is evaluated through helpfulness votes (Korfiatis et al., 2012; Mafael et al., 2016). An extensive literature reported that the current helpfulness vote system may not be a reliable measure of the review content quality (e.g., Wan & Nakayama, 2014) and multiple sources of bias (more in section 2.2 of the literature review) were reported as impacting the current helpfulness assessment methods through user votes (e.g., Liu et al., 2007). The suggested lack of reliability of these approaches may have important implications for the marketing discipline. To address this potential lack of reliability and other problems of people-driven helpfulness votes, this paper suggests a method to evaluate the helpfulness of reviews, and reviewers, more objectively, through measures of relative information content.

This study analyzes the textual portion of online reviews using Latent Semantic Analysis (LSA) to identify and quantify high entropy words. Information entropy is a commonly used measure of the amount of information contained in a system – in this case, an online review. The study classifies these high entropy words into three types of information that we label as corroboration, opinion, and recommendations, a framework previously used in existing literature (Fresneda & Gefen, 2020). Corroboration information is related to high entropy words used by sellers to describe their product and are also employed by reviewers in the review itself to confirm or disconfirm sellers' claims. Recommendation information are words, other than those in the corroboration group, that explicitly recommend the product, such as "buy" or "recommend". All the other high entropy terms, after accounting for both corroboration and these explicit recommendations made by the reviewer, are considered opinion information. Opinion, also referred as 'regular opinion' in the literature (Jindal & Liu, 2006; Qazi et al., 2016), is therefore a broad category that accounts for the remaining high entropy words in the textual portion of the product review.

After discussing the theoretical foundation in section 2 and developing the new methodology based on text-mining through LSA, information entropy, and principal component analysis (PCA) in section 3, section 4 elaborates on the application of our suggested method to classify reviews. Section 4 runs a PCA to show that the new suggested measures of information content, together with readability, load strongly only on the first principal component and very low on the other principal components. The other principal components are associated with variables suggested by previous research to be associated with known helpfulness biases. Section 5 then applies this objective methodological approach to rank the reviewers and, accordingly, identify the best contributors. Put into context, past literature acknowledged the influence and persuasive power of the most active online reviewers (Senecal & Nantel, 2004). Certainly, many e-commerce and online review sites developed rankings of their reviewers in an attempt to identify their more active and involved contributors. The study concludes with the discussion in section 6. The paper also includes an appendix, Appendix A, where the authors propose a validation method of the suggested semantic measure of online review helpfulness used for the review/reviewer ranking approach. The methodology was tested on 49,143 reviews of three dissimilar product categories – appliances, computers, and luxury beauty – available for purchase at Amazon.com. The major contributions of this research are in: (1) suggesting a new methodology to identify and quantify information in individual online reviews, (2) showing how this method can rank both reviews and reviewers, and (3) achieving these goals with a single, automatic approach, capable of assessing and ranking new reviews and reviewers in real time.

2. Theoretical Development

2.1. Online Reviews and the Current Measures of Review Helpfulness

Marketers recognize the important role of previous buyers as a source of purchasing information that influences new buyers' behavior in both offline and online settings (e.g. Brown & Reingen, 1987; You et al., 2015). The early literature on online reviews suggested that consumers consider the information provided by other peer consumers as valid and reliable (Bickart & Schindler, 2001; Godes & Mayzlin, 2004; Mayzlin, 2006) due to its non-commercial nature (Sen & Lerman, 2007). This literature also suggested that online reviews are effective and more influential than other sources of information, since they can be accessed anywhere and at any time through the Internet (Bakos & Dellarocas, 2011).

Common elements included in online consumer reviews are the number of stars, a textual field for reviewers to provide additional qualitative information, and a helpfulness evaluation of the review (Kashyap et al., 2023). This helpfulness evaluation is provided by potential new buyers assessing the quality of the information included in each review (Mudambi & Schuff, 2010). Review helpfulness was defined as the "the extent to which consumers perceive the product review as being capable of facilitating judgment or purchase decisions" (Li et al., 2013 p.103). Helpful

reviews are intended to provide consumers with information from previous buyers that can ease the purchase decision process and reduce the uncertainty associated with the lack of physical cues from the product (Sun et al., 2019; Tan, 1999). Highlighting the relevance of review helpfulness, the literature suggests that review helpfulness can impact sales (Floyd et al., 2014; Gopinath et al., 2014; Rosario et al., 2016; Wang et al., 2015). Although a few sites, such as Amazon.com, currently do not report the number of helpfulness votes that a review received, the mechanism for evaluating reviews has not changed significantly (Amazon, 2025; Harrison, 2025; Kung et al., 2024). Still, helpful votes are the dominant feature in selecting reviews shown. Users can now prioritize 'top reviews' or the 'most recent' reviews, but it is 'top reviews' the default option (as of 08/27/2025) – helpful votes dominates the 'top reviews' option, whereas recency dominates the 'most recent' option when presenting reviews to users (Amazon-Staff, 2025). Other popular review sites, such as TripAdvisor.com, are currently reporting helpfulness votes for each review.

Yet, several findings in the literature put into question the suitability of helpfulness ratings to assess the actual quality of the information provided. Previous research suggested that the helpfulness votes of product reviews may not constitute a reliable measure of review content quality (Fresneda & Gefen, 2019; Fresneda & Gefen, 2020; Kashyap et al., 2023; Lu et al., 2018; Salehan & Kim, 2016; Schmalz et al., 2018; Singh et al., 2017; Wang & Karimi, 2019). Wan and Nakayama (2014) and Li and Hitt (2008) found that, for those reviews labeled as the "most helpful," the helpfulness ratings were exaggerated and much higher than the evaluations of online reviews collected from a random population. These findings led some authors to address two different meanings of "helpfulness": helpfulness "in the narrow sense", as a sales assistant helping in making an informed purchase decision, or helpfulness "in the wild", defined by the way users of sites such as Amazon.com evaluate reviews in practice (Danescu-Niculescu-Mizil et al., 2009). These two different meanings represent a reflection of the gap between how review helpfulness has been conceptualized ("in the narrow sense") and how it has been operationalized based on helpfulness votes ("in the wild"). Studies by Fresneda and Gefen (2019; 2020) suggested a semantic measure of review helpfulness based on a text analysis of online reviews to reconcile these two different meanings.

Several papers suggested review helpfulness ranking methods based or partially based on text content. Ghose and Ipeirotis (2007) employed a stylistic approach that classifies information in the text as either objective or subjective as a tool to rank reviews. Saumya et al. (2018) developed a rank method based on features extracted from the body of the review and the product description and combined with customer questionnaire data. Qin et al. (2022) combined sentiment analysis and attributes extracted from the text of the review, together with extrinsic metrics such as the number of votes that the review received to rank reviews. Similarly, Dong et al. (2021) considered two intrinsic aspects of text (emotional factors and consumer's functional preference for products) together with text length to rank reviews. Finally, Yang and Li (2022) also employed information entropy from reviews, together with extrinsic factors, to rank only negative reviews in terms of their helpfulness. These two authors focused on product-related and service-related information of a review to quantify information entropy, using a domain dictionary to identify specific domain words related to these two types of information. In sum, previous research did not consider a holistic assessment of the information contained in online reviews by evaluating intrinsic elements, such as corroboration, opinion, and recommendation information.

Previous research identified several types of biases as a potential reason for the difference between these two meanings of review helpfulness. Since the purpose of this study is to develop a more objective measure of information content, the next section suggests variables that can potentially account for these sources of bias. The paper controls for these potential sources of bias by employing those suggested variables, as explained in the Methodology section.

2.2. Potential Sources of Bias in Helpfulness Evaluations of Online Reviews

Liu at al. (2007) suggested three different types of biases in the evaluation of review helpfulness: early bird bias, imbalanced vote bias, and winner cycle bias. The early bird bias suggests that the earlier a review is posted, the more votes it will receive. This potential source of bias has been reported in several other studies (e.g. Lu et al., 2018) and it is arguably the motivation for Ghose and Ipeirotis's (2007) work. The imbalanced vote bias suggests that Internet users tend to rate others' opinions positively rather than negatively. Finally, the winner cycle bias indicates that reviews with many helpfulness votes will continue to attract more votes, as reviews that are top ranked are more easily accessible to users. Literature also refers to the winner cycle bias as the "Matthew effect" (Singh et al., 2017; Wan, 2015).

In our study, the variable Posting Order of the online review accounts for the early bird bias and potentially for the imbalanced vote bias and for the winner cycle bias. Since early reviews get more helpfulness votes than late ones (Lu et al., 2018), they are more likely to receive more future helpfulness votes through the winner cycle bias. Early reviews gain default authority among users (Liu et al., 2007), eventually influencing the objectivity of the helpfulness votes due to the imbalance vote bias – as earlier reviews' helpfulness may be evaluated more positively rather than negatively. These three sources of bias may lead to the earliest reviews being more accessible than late reviews. Hence, review posting order may suitably account for the accessibility of the online review.

Another source of potential bias is the result of existing ranking algorithms. Amazon's algorithm for ranking helpful reviews is based on the aggregated number of evaluations that the reviews received. This feature has an "anchoring effect", leading to the more voted helpful reviews receiving even more votes, since they are made more accessible to review users (Cao et al., 2011; Wan & Nakayama, 2014). This research employs the variable *Number of Evaluations of Review Helpfulness* to account for this anchoring effect and to account for the possible divergences with review posting order. Additionally, the use of the total number of helpfulness evaluations allows to control for product heterogeneity and popularity because the total number of helpfulness evaluations, practically, uniquely identifies each product ID.

The number of stars is another potential source of bias. Several studies addressed the impact of negative or positive valence (number of stars) on the evaluation of the information provided by reviews, leading to a negative or positive bias (e.g., Hu et al., 2009; Pan & Zhang, 2011; Purnawirawan et al., 2015; Sen & Lerman, 2007; Wu, 2013). These studies found that negative or positive star ratings are more influential over variables, such as sales or review helpfulness. Corroborating this, De Langhe et al. (2016) found a relevant lack of consistency between star ratings and 'more objective' measures of product quality obtained from ConsumerReports.org scores. Furthermore, previous studies suggested that the number of stars may not reflect product performance accurately, since its distribution sometimes follows a J-shaped pattern (Hu et al., 2017; Hu et al., 2009) – with very few reviews rating products with 2 or 3 stars. Ullah et al. (2016) suggested that review ratings, such as the number of stars, can be better understood as a sort of catharsis, i.e. emotional release. These authors also found that on the extreme positive side of the review ratings, reviews encompass a greater balance of positive emotional content in contrast to the extreme negative side of the review ratings, where the negative emotional content is smaller. Hu et al. (2017) suggested an acquisition bias. The acquisition bias is a self-selection bias that implies that product buyers – and therefore potential product reviewers – are more likely to have a positive predisposition towards the product, rather than negative. These findings may help to explain the aforementioned J-shaped distribution of the number of stars. Villarroel Ordenes et al. (2017) suggested that the role of the number of stars in a review may be closer to a summary of the overall sentiment of a review than to an evaluative tool. By incorporating the variable *Number of Stars*, this study attempts to correct for the negativity and positivity bias – including the acquisition bias – suggested in previous research. Based on the aforementioned literature, this study contends that the number of stars is more closely related to a potential source of bias, influencing the actual evaluation of the information content.

Another source of potential bias is sentiment. Sentiments included in the text were consistently found to impact the assessed helpfulness of online consumer reviews, regardless of the actual information content of the review (e.g. Chong et al., 2016; Hu et al., 2012; Maidar et al., 2024; Salehan & Kim, 2016; Villarroel Ordenes et al., 2017). To account for this potential source of bias a sentiment analysis of each review included in the dataset was implemented. This sentiment analysis provides a *Sentiment Score* that may be a good representation of the emotional balance of the content included in each review.

Table 1. Summary of potential sources of bias in helpfulness vote evaluations.

Source of Bias	**Brief Description**	**Example of Relevant Literature**	**Suggested Variable to Control for Source of Bias**
Early bird bias	The earlier a review is posted, the more votes it will receive	Lu, Wu and Tseng (2018)	*Posting Order*
Imbalanced vote bias	Internet users rate others' opinions positively rather than negatively	Liu et al. (2007)	*Posting Order**
Winner cycle bias/ Matthew effect	Reviews with many helpfulness votes will attract more votes	Wan (2015); Singh et al. (2017)	*Posting Order**
Anchoring effect	Reviews with more review votes are more accessible	Cao, Duan and Gan (2011)	*Number of Evaluations of Review Helpfulness*
Number of stars	Negative or positive star ratings are more influential	Purnawirawan (2015)	*Number of Stars*
Sentiments	Sentiments included impact the assessed helpfulness	Villarroel Ordenes et al. (2017)	*Sentiment Score*

Table 1 summarizes the potential sources of bias that the suggested methodology in this paper controls for by incorporating four variables: *Number of Stars*, review *Posting Order*, *Number of Helpfulness Evaluations*, and *Sentiment Score*. The methodology proposed shows that these potential sources of bias can be separated, largely, from the information included in online reviews. If that is the case, the methodology suggested may be a suitable alternative to review helpfulness in evaluating the information content of each individual review. To estimate the amount of information of each type included in each review we employed information entropy.

2.3. The Proposed Alternative. Accounting for Information Content with Information Entropy

The length of the text was a variable widely employed in previous literature as a proxy of information content in online reviews, even in the early attempts of processing the textual data included in each review (Godes & Mayzlin, 2004). Some authors found a positive relationship between the length of the text and review helpfulness (Lu et al., 2018; Schindler & Bickart, 2012). Nevertheless, a limiting issue of using textual length, as a measure of information content, is that specific words that can be informative for one product category may not be informative for another. Even more important, more words may not imply more information or better information. Previous literature found that new buyers are more influenced by the quality of information – defined as information breadth and depth, credibility, relevance, and factuality – than by information quantity (Filieri, 2015). This study employs information entropy as an approach to overcome those limitations.

Information entropy was suggested by Shannon (1948) in his seminal work "A Mathematical Theory of Communication." Entropy of information is defined as a "measure of the amount of information the system contains" (Belzer, 1973 p.301). In the context of online consumer reviews, 'system' refers to the textual portion included in each review. Two of the most important contributions of the seminal work by Shannon are related to the properties of information: information can be actually measured and has an additive property. If a message contains elements that are statistically independent of each other, the total information contained is the sum of the information content of the individual elements (Machta, 1999; Shannon, 1948). Shannon's entropy of a categorical random variable with size p and with associated probabilities $\theta_1,...,\theta_p$ with $\theta_k > 0$ and $\sum_k \theta_k = 1$ is given by (Hausser & Strimmer, 2009):

$$H = -\sum_{k=1}^{p} \Theta_k \, log(\Theta_k) \qquad (1)$$

The methodology suggested in this study accounts for three types of information included in online reviews that we named "corroboration", "recommendation" (explicit recommendations from previous reviewers), and "opinion". The selection of these specific types of information is based on (1) their prevalence in the literature (e.g. Qazi et al., 2016); and (2) their prevalence in a content analysis conducted by Lu et al. (2014) over a large data set of consumer-generated reviews.

Mudambi and Schuff (2010) and Li and Huang (2013) found that reviews discussed product features and provided the experiences of previous buyers. Furthermore, Moe and Trusov (2011) and Min and Park (2012) pinpointed that the text in online reviews describes not only the personal experience of former buyers with the product but also with the whole purchase process (i.e. including broad elements such as shipping, packaging, or customer service interactions). Other authors considered product features and characteristics cited in online reviews in their analyses (e.g. Ghose et al., 2012; Lee & Bradlow, 2011; Sánchez-Franco et al., 2016). Previous literature referred to recommendations to potential new buyers as another informational element included in reviews (Duan et al., 2008; Vermeulen & Seegers, 2009). Packard and Berger (2017) suggested that explicit endorsement (of the type "I recommend it") included in online reviews is very persuasive and increases purchase intent by itself. This highlights the importance of this type of information and arguably, its impact on helpfulness. In an extensive content analysis of text reviews, Lu et al. (2014) found that 86.4% of reviews are focused on information about what the reviewer learned from the whole transaction process, including elements such as service attitude, congruity between seller's claims and real products, shipping, and order fulfillment time. 7.1% of the reviews on their database included content related to tips or recommendations to other buyers (Lu et al., 2014). Finally, 5.2% of the reviews analyzed focused on cognitive processing of product features and characteristics.

Given this literature, this study suggests three types of information that together with readability constitutes the foundation of the ranking methodology: corroboration, recommendation, and opinion. This framework was already employed in existing literature (Fresneda & Gefen, 2020). The classification of information into these three types is, to some extent, akin to Qazi et al. (2016), except that our study applies their theoretical framework at a more granular level, as explained in a subsequent paragraph.

The information type that we labeled as *corroboration* is information provided by manufacturers and sellers that is validated or refuted by reviewers in the textual portion. Corroboration information is important because, to reduce purchase uncertainty, individuals rationally seek to verify from other sources, such as peer consumers, the claims that

manufacturers make (Fresneda & Gefen, 2020; Racherla et al., 2012; Ramirez et al., 2002; Sánchez-Franco et al., 2016). This information type is conceptually close to *comparative opinions* in Qazi et al. (2016).

The second category of information is *recommendation*. This information is derived from text that was not selected into the corroboration category but included words that clearly recommend the product such as "buy" or "recommend". (The Methodology section of this study shows that this information category accounts for a reasonably comprehensive set of words related to recommendations.) Packard and Berger (2017) suggested that endorsement is important to consumers. Explicit recommendations are also prevalent in online reviews. Previous literature identified explicit recommendations as an important component of online reviews (e.g. Vermeulen & Seegers, 2009). The content analysis conducted by Lu et al. (2014) found that 7.1% of online reviews contain recommendations or explicit tips from previous buyers. Our conceptualization of recommendations is conceptually close to *suggestive opinions* in Qazi et al. (2016).

The third information category is *opinion*. This category includes the high entropy words that were not included in the previous two categories. Conceptually, the opinion type of information correspond to *regular opinions* in Qazi et al. (2016) or merely *opinion* in prior literature (Jindal & Liu, 2006). This information category addresses positive or negative facets of a product, excluding both corroboration and recommendation words (i.e. comparative opinions and suggestive opinions in Qazi et al. (2016), respectively). Again, our study does more than merely replicating previous research.

Although conceptually similar to the theoretical framework from Qazi et al. (2016), our approach differs significantly from these authors in the unit of analysis. Qazi et al. applied their theoretical framework to label entire reviews as either regular, suggestive, or comparative opinions, whereas this study applies an approach with a finer level of granularity by identifying individual terms included in each review. A finer level of granularity allows an arguably more realistic approach, as reviews can contain these three types of information simultaneously (i.e. a single review can contain recommendations and at the same time opinions and/or corroborations). Reviews that provides different types of information simultaneously might be perceived as more relevant and more helpful, as they address more than one product information facet at once.

In addition to measures of information content, this study also employs a readability index to account for the ability of the reviewer to convey the information in the review. Due to the open and available nature of online reviews, the way the information is conveyed may impact the ability of new buyers to understand and use the information. In this sense, previous research suggested that higher readability has a positive effect on variables such as review helpfulness (Ghose & Ipeirotis, 2011; Korfiatis et al., 2012; Liu & Park, 2015). As one of the most standard measures of readability, this study employs the Automated Readability Index (ARI) developed by Senter and Smith (1967). Applying that measure, this study estimates information content through information entropy while controlling for text readability through ARI and controlling for the aforementioned sources of bias. In this manner the study extends previous literature in terms of the information that is provided.

2.4. Another Important Aspect. Reviewers and Source Credibility

Reviewers are assessed on their credibility as reliable/unreliable sources of information as potential new buyers search for cues and signs that those reviewers have credibility (Agnihotri & Bhattacharya, 2016; De Langhe et al., 2016; Karimi & Wang, 2017). Bennett (1999) asserted that online messages are evaluated through the question "Who is telling me this?" (p. 4). Credible sources are critical in e-commerce as the anonymity of reviewers can make it difficult for potential buyers to determine the reliability of the information provided (Folse et al., 2016). In the context of product reviews, source credibility is the extent to which the reviewer is perceived as a reliable source of product information and can be trusted as giving objective opinions about products (Goldsmith et al., 2000; Lu et al., 2018; Weathers et al., 2015). Credible sources of information induce more positive attitudes toward products than other less credible sources. Sources evaluated as credible can also increase the purchase intention of consumers through the adoption of the messages that they conveyed (Shan, 2016).

Indeed, different review sites developed systems to identify and rank reviewers on their contribution as information providers. Some of those sites allow users to evaluate reviewers on the perceived helpfulness of the information provided – e.g. Tripadvisor.com as shown in Figure 1. Accounting for the importance of recognizing their best reviewers, other sites even provided financial incentives to those of them who have proven themselves as credible reviewers – as it was the case of Epinions.com.

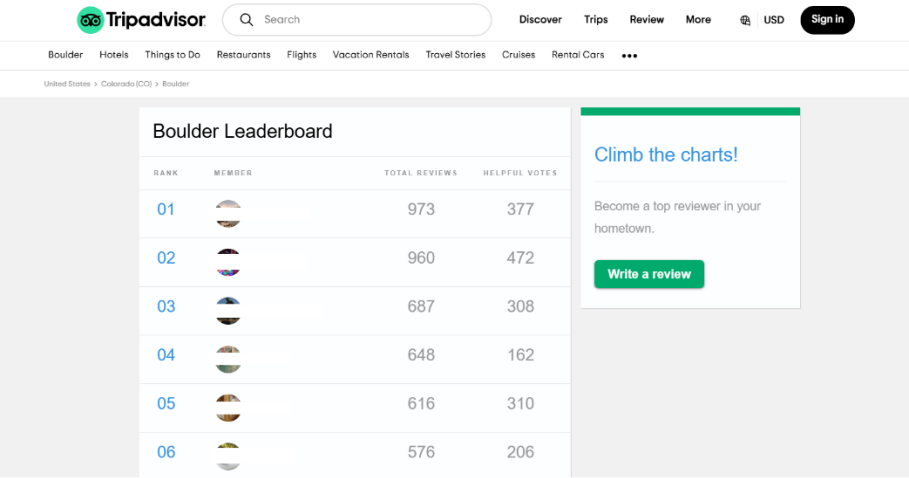

Figure 1. Tripadvisor.com Reviewer Ranking for Boulder (CO) (as of 12/16/2024).

The common practice of recognizing the more relevant contributors as "top reviewers" or of assigning badges to identify top reviewers (e.g. Yelp Elite Squad badge) is also well supported in the literature. Previous research suggests that potential buyers are more receptive to messages conveyed by reviews that exhibit these signals of quality reviewers (Baek et al., 2012; Ghose & Ipeirotis, 2011). Publicly recognizing the best contributors provides "good reasons" for consumers to trust the review of a complete stranger (Willemsen et al., 2012). The next section introduces the methodology employed to rank both reviews and reviewers.

3. Methodology

The dataset employed in this research consists of 49,143 reviews from Amazon.com corresponding to three product categories: appliances, computers, and luxury beauty (16,381 reviews per category). The dataset was collected in 2014 and updated in 2023 by Julian McAuley and was previously used in scholarly research (e.g. Fresneda & Gefen, 2019; McAuley, Pandey, et al., 2015; McAuley, Targett, et al., 2015).[2] The available variables are: reviewer ID, product ID (Amazon's ASIN number), the text of the review, number of stars, helpfulness votes, number of helpfulness evaluations of each review, and review posting date. The availability of variables such as *helpfulness votes* and *number of helpfulness evaluations of each review*, variables that are no longer reported by Amazon (Figure 2 shows the limited information currently reported by Amazon.com) makes this dataset very suitable for our study. The date of the review combined with the ID of the product allows for the calculation of the posting order of each review. As the number of helpfulness evaluations uniquely identifies each product ID, the inclusion of number of helpfulness evaluations practically adds a control of product ID, and, with it, product type and its heterogeneity, into the analysis. The sentiment score was calculated for each of the reviews in the dataset following a lexicon-based sentiment analysis (Hu & Liu, 2004; Taboada et al., 2011). Specifically, Liu's opinion lexicon was employed (Liu & Hu, 2019), consisting of a list of around 6,800 English positive and negative opinion words or sentiment words.

[2] The 2023 version of the dataset was expanded with new reviews and several improvements as compared to the 2014 version, such as: Newer Interactions; Richer Metadata; Fine-grained Timestamp; Cleaner Processing; and Standard Splitting. Further information can be found in the following link: https://amazon-reviews-2023.github.io/.

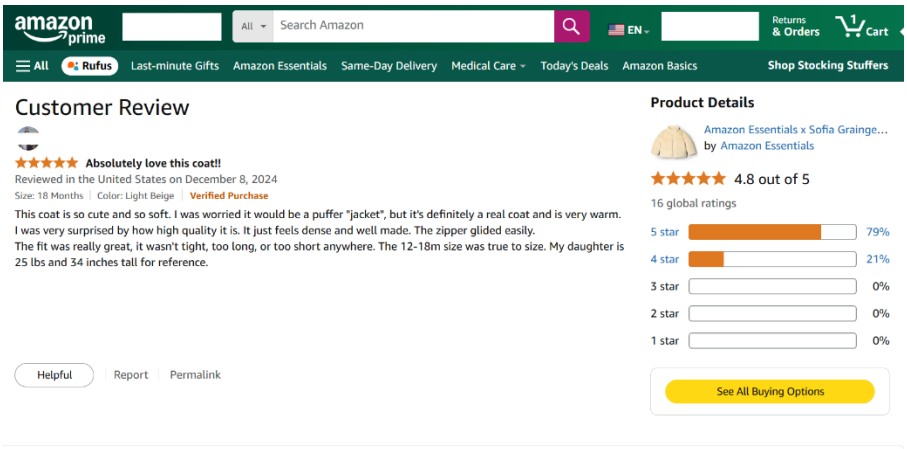

Figure 2. Amazon,com Review Example.

LSA was employed to analyze the textual part of each of the 49,143 online reviews of the dataset. Through LSA, this study identifies and quantifies the number of terms ascribed to recommendation, corroboration, and opinion types of information in the text of each review. The terms belonging to the three types of information are applied to estimate the recommendation, corroboration, and opinion entropy embedded in the text of each review. LSA is an automatic statistical method that defines the similarity of meanings across words (named "terms" in LSA) and documents (e.g. online consumer reviews) by analyzing large bodies of text. LSA identifies through singular value decomposition (SVD) what terms and what documents factor together in a Document-to-Terms frequency Matrix (DTM). Terms factoring together are assumed to carry some latent shared meaning, similar to what factors are assumed to do in a principal component analysis (PCA). As in a PCA, only those factors that explain a predetermined high degree of variance are kept. The retained factors thus identify those terms that carry the most weight in explaining variance in the DTM. As LSA applies SVD, the terms in the text can be associated with each other even if they do not co-occur together. It is enough for two terms to both co-occur with a third term to make the former two terms close in meaning (Landauer et al., 1998). Research suggests that LSA approximates some aspects of human learning and discrimination of meaning similarity between words (Landauer et al., 2013). LSA allows for the analysis of many documents – online product reviews – with minimal human intervention. All of these characteristics make LSA especially suitable for the analyses performed in this study.

LSA is a bottom-up approach. In text analysis, bottom-up approaches study patterns in the text first, to suggest more elaborated theoretical explanations subsequently (Humphreys & Jen-Hui Wang, 2018). Differing from common topic modeling approaches such as Latent Dirichlet Allocation (LDA) that identify topics in the text (e.g. Buschken & Allenby, 2016; Mankad et al., 2016), LSA identifies how semantically close terms are to each other. Discovering latent topics would not serve the purpose of the study. Differing from top-down approaches such as Sentiment Analysis (e.g. Li & Wu, 2010) or Conditional Random Field (CRF) (Netzer et al., 2012), LSA is not a dictionary-based or predefined rule-based technique. This allows LSA more flexibility than other text analysis tools. Our methodology, based on LSA, partly fills the gap suggested by Humphreys and Wang (2018). LSA runs as an unsupervised classification method, and is more suitable for insight interpretation.

Explicit recommendations in the text of each review were identified by using the lexical database WordNet (University, 2010). WordNet classifies words into groups of cognitive synonyms, which are presented to the user by their semantic relations – by words that share meaning. We chose WordNet synonyms for the words "recommend" and "recommendation" (such as "suggest") and "buy" (such as "acquire"), since many of the recommendations are associated with suggestions about buying or not buying a product. Potentially conflicting words, such as "get" – as one of the suggested words for "buy" – were discarded. Words in past tense were discarded as well, as potentially narrating the experience with the product, such as "I bought this washing machine and it worked well for many loads." The resulting list contained the words "acquire," "advise," "buy," "procure," "propose," "recommend," "suggest," "tip," and "warn" and the associated words to those – such as "recommendations," "warning," and "suggestions." This list of words allowed us to identify recommendations in the text of reviews and estimate the recommendation entropy of each of those reviews. Looking at many of the recommendations included suggests that this list seems comprehensive. This methodology automates a task implemented by human coders in previous research (Negi & Buitelaar, 2015).

Product descriptions provided by the retailer/manufacturer were collected by using the ASIN number of each of the products reviewed as a reference (as shown in Figure 3). These product descriptions were employed to identify corroboration terms, as reviewers confirm or deny these claims from the sellers in the text of the review.

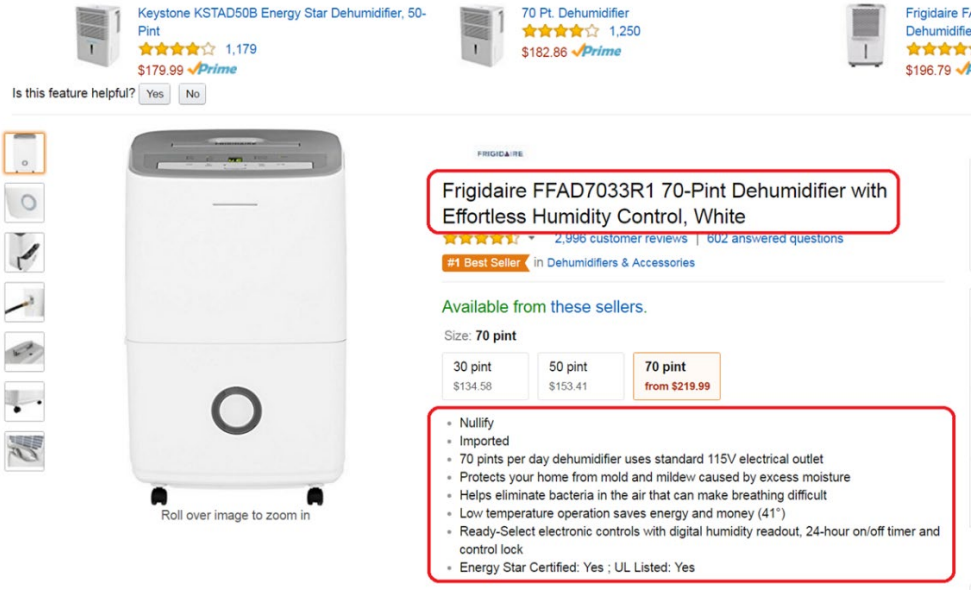

Figure 3. Operationalization of Corroboration Terms.

Product descriptions provided a large body of text, so additional text-mining operations were required. As a first step, stop-words (such as "the", "a", and "an"), numbers, and punctuation marks were removed from the text. The remaining words were included in another DTM where each term had a frequency for each of the documents included. A weighting transformation was used. This transformation reduces the influence of frequent words, which presumably carry less differentiating meaning, and gives more weight to those words that, being infrequent, convey more information (Landauer et al., 2013). We used the default *tfidf* transformation (Dumais, 2004). This approach requires multiplying the log of the local term frequencies (localWeight) by the inverse of the entropy of the word in the text body (globalWeight):

$$P_{ij} = \frac{wordLocalFrequency}{wordGlobalFrequency}\ ; \qquad localWeight = log_2(wordLocalFrequency + 1)\ ;$$

$$globalWeight = 1 + \frac{\sum_j^{ncontexts} P_{ij} * log_2 P_{ij}}{\log ncontexts}\ ; weight = localWeight * globalWeight \quad (2)$$

Where:

- **i** = term number
- **j** = document number
- **wordLocalFrequency** = number of appearances of term *i* in document *j*
- **wordGlobalFrequency** = total number of appearances of term *i* in all the documents
- **ncontexts** = total number of documents

This entire process was repeated independently for the three product categories included in the dataset. As a consequence, three different lists of product description words were generated and were employed later to identify corroboration terms in the reviews. Working with three different product description lists allowed us to analyze each product context separately.

The analysis of the text included in each online review started by generating a weighted DTM. The weighted DTM was calculated over the textual portion of all the reviews. As standard, a reduced-rank SVD was calculated from the weighted DTM. This process resulted in a SVD reduced-dimension representation, which was the best k-dimensional approximation to the original weighted DTM by using least-squares. After this process, each term and each document were represented in a semantic space as a k-dimensional vector (the number of dimensions of this vector is automatically selected by the *lsa* function in R). This reduced-dimension space of meaning allowed us to compute similarities and associations between terms, between documents, and between documents and terms, since

documents and terms were represented as vectors in the semantic space (Dumais, 2004). To summarize the information available, the closest 20 terms in meaning to each of the 49,143 reviews in the dataset were computed using this semantic space.[3]

Recommendation terms were labeled in the three pools of review words using the original list of recommendation words. The number of times that any of the recommendation terms was mentioned in the text of the review was counted and summed. Likewise, corroboration terms were identified in the three pools of review words from the three lists of product description words. The number of times that any of the corroboration terms appeared in the text of each review was counted and added. From the three pools of review words, all the terms that were previously ascribed to recommendation or corroboration were removed. The remaining words were labeled as opinion terms. The number of times that any of those opinion terms were mentioned in the text of the review was counted and summed.

After this identification process, the entropy of the recommendation, corroboration, and opinion types of information was estimated as the number of terms for each review added beyond the number of words included in the product description by the seller. This also allows studying the contribution of individual reviews beyond the information that is provided by sellers. Information entropy is estimated using the Maximum Likelihood (ML) method (Hausser & Strimmer, 2009) of the number of terms of the recommendation type added beyond the number of terms in the product description provided by the seller. ML can estimate entropy also from discrete counts. Because the underlying probability mass function in Equation (1) is of a discrete random variable that in practice is unknown, H and θ_k in Equation (1) need to be estimated from observed counts $y_k \geq 0$ – i.e. the number of times that recommendation, corroboration, or opinion terms were used in each review. Equation (3) is modified from equation (1) as follows:

$$\hat{H}^{ML} = -\sum_{k=1}^{p} \hat{\theta}_k^{ML} \log\left(\hat{\theta}_k^{ML}\right) \qquad (3)$$

Equation (3) is constructed by plugging the ML frequency estimates in Equation (4):

$$\hat{\Theta}_k^{ML} = \frac{y_k}{n} \qquad (4)$$

into Equation (1), with $n = \sum_{k=1}^{p} y_k$ as the total number of counts. (As an example, a review in the dataset that contains 3 terms ascribed to recommendations and 53 words in the product description would have a value of recommendation entropy of

$recommendation\ entropy(53, 3) = 0.2088998$).

As a measure of the readability of the text, the ARI score is calculated through the following standard formula (Senter & Smith, 1967) for each of the reviews in the dataset (as using the number of characters, the number of words, and the number of sentences in each individual review):

$$ARI\ Score\ =\ 4.71\left(\frac{number\ of\ characters}{number\ of\ words}\right) + 0.5\left(\frac{number\ of\ words}{number\ of\ sentences}\right) - 21.43 \qquad (5)$$

After describing our methodology, the following sections of the study introduce its potential applications. First, section 4 presents its potential application to develop review rankings.

4. A Suggested New Method of Review Classification

The current helpfulness vote rankings of reviews raise several issues, addressed in the Theoretical Development section of this study. This section suggests and empirically tests a new approach that combines the three types of high entropy words that we categorized as recommendation, corroboration, and opinion entropy, together with the readability index of the review (ARI), to rank reviews. As a starting point, a PCA with Varimax rotation was run including these four information-related variables and the previous four variables suggested to account for the different potential sources of bias (see last column in Table 1): *Posting Order*, *Number of Helpfulness Evaluations*, *Number of Stars*, and *Sentiment Score*. Tables 2-4 show the results of the PCA analyses of the three product categories with *Eigenvalue* ≥ 1 (loadings above .40 are emphasized in red). In all three product categories the PCAs show that the four information-related variables are in the first principal component (i.e. the component that accounts for most of the variance), that they load very low on the other principal components, and that those other principal components combine the variables known from previous research to be associated with potential sources of bias. Also note that the four information-related variables all have positive loadings and that the values of those loadings exceed .5 in all cases. This Factor 1 will be used in the remainder of the paper.

[3] Different dimensions sizes were tested. We found that the best balance of term saturation vs. computation burden was at 20 terms.

Table 2. PCA Analysis for the Appliances Category (Varimax Rotation).

Rotated Component Matrix [a]

	Component		
	1	2	3
Number of Stars	-0.185	**0.812**	0.153
ARI	**0.534**	0.005	-0.261
Posting Order	0.105	0.112	**0.792**
Sentiment Score	0.069	**0.868**	-0.088
Corrob. Entropy	**0.765**	-0.016	-0.057
Opinion Entropy	**0.701**	-0.156	-0.145
Recom. Entropy	**0.594**	0.015	0.219
Number Help. Eval.	0.219	0.052	**-0.587**

Rotation Method: Varimax with Kaiser Normalization

[a.] Rotation converged in 4 iterations

Initial Eigenvalues

Component	Total	% of Variance	Cumulative %
1	**1.971**	24.634	24.634
2	**1.384**	17.305	41.939
3	**1.054**	13.180	55.119
4	0.912	--	--

Table 3. PCA Analysis for the Computers Category (Varimax Rotation).

Rotated Component Matrix [a]

	Component		
	1	2	3
Number of Stars	-0.049	-0.029	**0.965**
ARI	**0.587**	-0.016	0.095
Posting Order	-0.293	**0.744**	-0.087
Sentiment Score	0.317	**0.710**	0.085
Corrob. Entropy	**0.767**	0.094	-0.008
Opinion Entropy	**0.751**	-0.012	-0.017
Recom. Entropy	**0.519**	-0.067	-0.114
Number Help. Eval.	0.310	0.153	0.186

Rotation Method: Varimax with Kaiser Normalization

[a.] Rotation converged in 4 iterations

Initial Eigenvalues

Component	Total	% of Variance	Cumulative %
1	**2.058**	25.726	25.726
2	**1.090**	13.621	39.348
3	**1.002**	12.520	51.868
4	0.982	--	--

Table 4. PCA Analysis for the Luxury Beauty Category (Varimax Rotation).

Rotated Component Matrix [a]

	Component		
	1	2	3
Number of Stars	-0.167	**0.845**	0.093
ARI	**0.563**	-0.023	-0.280
Posting Order	0.051	-0.109	**0.819**
Sentiment Score	0.348	**0.742**	-0.061
Corrob. Entropy	**0.779**	0.168	-0.133
Opinion Entropy	**0.795**	-0.070	-0.078
Recom. Entropy	**0.539**	0.075	0.291

Number Help. Eval.	0.177	-0.121	**-0.505**

Rotation Method: Varimax with Kaiser Normalization

[a.] Rotation converged in 5 iterations

Initial Eigenvalues

Component	Total	% of Variance	Cumulative %
1	**2.137**	26.707	26.707
2	**1.305**	16.314	43.022
3	**1.044**	13.048	56.069
4	0.924	--	--

The study next compared the ranking results of our approach, i.e. Factor 1, with the common approach of using positive helpfulness votes.[4] At first glance, both approaches exhibit important differences. The helpfulness ranking contains old reviews in the very first positions, while the information ranking contains more recent reviews in its top positions, possibly overcoming some of the noted biases of helpfulness ranking. As an illustration, Figure 4 shows the 15 first reviews ordered through both approaches, namely (1) helpfulness votes and (2) information score, for the appliances product category. The difference between the dates in the fifth column is noticeable, as is the number of helpfulness evaluations as shown in the last column. Spearman's rank correlation and Kendall rank correlation, shown in Table 5, suggests that there is a very small positive correlation between the two ranking approaches. Finally, Table 6 illustrates these discrepancies between the information ranking and the helpfulness ranking with specific examples from each of the three product categories, showing the textual portion of each of these examples.

Helpfulness Rank	ASIN	Helpful Votes	Product Name	Posting Date	Number Eval
1	B000O3I2CG	546	SPT SU-4010 Ultrasonic Dual-Mist Warm/Cool Humidifier with Ion Exchange Filter - Blue	1/17/2009	552
2	B0014X7B54	416	Waste King Legend Series 1/2 HP Continuous Feed Operation Garbage Disposal - (L-2600)	2/20/2009	422
3	B0006MQCA4	367	KAZ INC. PUR 2-stage Dispenser - DS-1800Z	12/31/2006	427
4	B002QUAPSO	298	Breathing Mobile Washer Classic - Portable Clothes Washing Machine - Handheld - Manual - Mobile Hand Powered Laund	4/30/2011	306
5	B000O3I2CG	256	SPT SU-4010 Ultrasonic Dual-Mist Warm/Cool Humidifier with Ion Exchange Filter - Blue	1/16/2009	294
6	B001FDV410	207	Haier HLP23E Electronic 1-1/2-Cubic-Foot Touch Pulsator Top-Loading Portable Washing Machine	9/8/2010	210
7	B000FFS0V6	206	Danby DDW496W Countertop Dishwasher	11/26/2007	206
8	B003JN13S8	206	Samsung WF210ANW 3.5 cu. Ft. High Efficiency Front-Load Washer - White	3/13/2011	209
9	B002C8HR9A	195	The Laundry Alternative Wonderwash Non-electric Portable Compact Mini Washing Machine	8/3/2009	199
10	B003O4QDWS	166	Speed Queen AWN412 3.3 Cu. Ft. White Top Load Washer	12/12/2009	175
11	B0002KXMT4	160	Haier HLP21N 6.6-Pound Pulsator Wash with Stainless Steel Tub	3/20/2008	163
12	B002GEDBIG	129	Centrifugal Clothes Portable Spin Dryer	9/26/2009	134
13	B000G837TW	128	InSinkErator Evolution Compact 3/4 HP Household Garbage Disposer	2/29/2008	129
14	B00005O64S	127	Haier HDT18PA Space Saver Compact Countertop Dishwasher	1/11/2010	136
15	B0039V7JFG	127	LG 3.6 CF FRONT LOAD WASHER DRYER COMBO	2/15/2011	152

Information Rank	ASIN	Information Score	Product Name	Posting Date	Number Eval
1	B0029CZQIS	5.14355	LG 5231JA2006A Refrigerator Water Filter	9/12/2013	0
2	B000AST3AK	5.07341	General Electric MWF Refrigerator Water Filter	3/13/2014	1
3	B001FDV410	5.05065	Haier HLP23E Electronic 1-1/2-Cubic-Foot Touch Pulsator Top-Loading Portable Washing Machine	9/8/2010	210
4	B002QUAPSO	4.87006	Breathing Mobile Washer Classic - Portable Clothes Washing Machine - Handheld - Manual - Mobile Hand Powered Laund	4/30/2011	306
5	B003LB86W4	4.74469	LG LFX28978ST27.6 Cu. Ft. Stainless Steel French Door Refrigerator - Energy Star	4/3/2013	6
6	B000FFS0V6	4.71536	Danby DDW496W Countertop Dishwasher	11/26/2007	206
7	B000AST3AK	4.54844	General Electric MWF Refrigerator Water Filter	5/2/2013	0
8	B002GQQWYY	4.5216	Samsung DMR78AHS 24 Integrated Console Tall Tub Dishwasher In Stainless Steel	12/3/2010	3
9	B003RCBX4K	4.48903	Samsung DMT400RHS 400 24" Stainless Steel Fully Integrated Dishwasher - Energy Star	5/21/2011	36
10	B003XRA3CM	4.43321	Whirlpool 4391960 Element	4/27/2014	0
11	B001AAEGB8	4.40235	Whirlpool 8212490RC 7-Foot Industrial Braided Ice Maker Hose	10/28/2012	5
12	B0009793KC	4.3692	EveryDrop by Whirlpool Refrigerator Water Filter 5 (Pack of 1)	6/4/2012	0
13	B001DTB2GI	4.32838	Countertop Portable Dishwasher with Digital Controls - White	5/25/2011	0
14	B0021JYYMI	4.32679	LG : WM2301HR 27 Front-Load Washer with 4.2 cu. ft. Capacity Wild Cherry Red	12/6/2010	7
15	B0053Y2XJ4	4.25554	Whirlpool 61004441 Actuator Pad for Refrigerator	8/8/2012	1

Figure 4. Fifteen First Reviews of the Appliances Category Ordered by Helpfulness Votes and Information Scores.

[4] The reviews included in the dataset were ordered according to the number of positive helpfulness votes, replicating approaches used by e-retailers and online review sites such as Yelp.com.

Table 5. Spearman's ρ and Kendall's τ Correlation Coefficients for the Three Product Categories.

Product Category	Spearman's rank correlation ρ	*p*-value	Kendall's rank correlation τ	*p*-value
Appliances	.203	< 2.2e-16	.136	< 2.2e-16
Computers	.326	< 2.2e-16	.219	< 2.2e-16
Luxury Beauty	.174	< 2.2e-16	.116	< 2.2e-16

Table 6. Ilustration of the Differences Between the Helpfulness Ranking and the Information Ranking.

Product Category: Appliances		**Review Text**
Information Rank	8	A Fantastic Dishwasher And Another Amazing Samsung Experience. Like Jim Logan, I just don't understand why people are so dissatisfied with this dishwasher. I LOVE mine. It works amazingly well. I use the Smart Autowash setting most often, but when I need them the other settings are fantastic. I read all the reviews on this site prior to purchasing, as well as other sites, but as with the other Samsung appliances I own (specifically the WF448AAP/XAA washer and the DV448AEP/XAA dryer, as well as our RS275ACRS/XAA refrigerator), I have learned from experience reading the instruction manual cover to cover helps immensely. In every case, all the common complaints I read about on this dishwasher (as well as the other items I own) were alleviated by simply following the instructions to the letter. I make certain to only use the detergent type and rinse agent recommended, I tweaked the setting to the recommended number (see previous reviews from other satisfied owners), I make sure I don't expect to grab my clean dishes straight out of the dishwasher to be used / put away immediately unless I have run the Sanitize setting (I find running the cycle at night gives the dishes plenty of time to dry by the next morning). My dishes come out perfectly clean, spotless, shiny, and like new 99% of the time, and that's without rinsing. There's the odd occasional bit of stuff every fifteenth load or so, and I'm talking a spec or two here, that I have to remove but beyond that the dishwasher is really quite perfect in my opinion. Looks great, runs SUPER quiet . everything Samsung promises, I have seen delivered and I absolutely LOVE this dishwasher. Don't let the negative reviews scare you - I'm confident you'll love it too.
Helpfulness Rank (# Helpfulness Votes)	4,445 (1)	
Posting Order	4	
Product Category: Computers		**Review Text**
Information Rank	8	Solid Home Video Editing Package. The \Pinnacle Studio Ultimate 12\" package is a solid video editing and publication package. Having used previous versions of the Pinnacle Studio software, there really wasn't much of a learning curve for me to come up speed on the great old features or pick up the newer additions. I should point out that I selected this software for my children to use since my higher-end software (Vegas Pro 8) is a little more complicated. Here are my observations: Pros:* Easy to use timelines allow for intuitive \"drag and drop\" insertion of video footage, transitions, and sounds* Easy to use footage editing tools make short work of trimming footage* SmartMovie and Montage templates are easy to use and really liven up home videos (Montage is great for opening scenes)* ScoreFitter is a very powerful and easy to use tool for generating and adding mood setting background music to the video footage - VERY useful tool* Powerful Effect Plug-ins (including \"Dream Glow\", which is activated by registering the software)* Lots of effects and transitions (be sure to read the printed manual and package inserts for using these)* One of the best DVD Menu creation tools of any non-pro software available* Easy to use Picture-in-Picture (PIP) and Chroma Key (Green Screen) tool - The included Green Screen material is good quality and worth about 1/3 of the package price Cons:* Only two video tracks available on the Timeline (Why I deducted one \"STAR\" from my rating) - This limits the ability to do PIP and titles, or multiple PIPs, at the same time - The work around to this is to do the first PIP, render the video, then import the new video and add the next PIP/Title overlay (repeat as needed)* The advanced plug-ins are not as intuitive as the rest of the program (read the manual and included package inserts) Comments:* Other reviewers have reported stability issues, I have not had any issues on \"Vista Ultimate\" with 4 Gigs RAM* Another good package in this price range is Sony Vegas Movie Studio Platinum Pro Packand should be considered as a viable alternative
Helpfulness Rank (# Helpfulness Votes)	898 (6)	
Posting Order	11	

		(less themes, an additional video track)Overall, I am very pleased with \"Pinnacle Studio Ultimate 12\" and it's ease of use. The huge number of effects, transitions, and themes will keep this software fresh for the foreseeable future. Recommended.".
Product Category: Luxury Beauty		**Review Text**
Information Rank	3	Noticeable Results after just 5-7 days. After being skeptical for some time, I decided to try the StriVectin-SD Eye Concentrate for Wrinkles since some of my friends had recommended it to me, so after 5 days of buying a small sample size, and following the instructions for using it I was really surprised that a product could give you results in just a few days. I started noticing the area around my eyes way more smooth, very firm and it was reducing bags under my eyes.as a result you feel more confident and people start noticing something “different on your look” of course my secret was that I've been using StriVectin-SD Eye Concentrate for Wrinkles. Highly recommended for a day to day use for a clear and bright skin. I just would say a bit pricey, but excellent results. hopefully they come up with specials to buy more of this product. Try it and enjoy the results.
Helpfulness Rank (# Helpfulness Votes)	5,749 (1)	
Posting Order	67	

5. A Suggested New Method of Reviewer Classification

Similar to the classification of reviews, ranking reviewers and identifying relevant contributors raise comparable issues. They too are based on helpfulness votes or on the number of contributions – number of reviews posted – with no reference or evaluation of the actual content. This section suggests an analogous approach to the one shown for individual reviews and tests it empirically as well. Differing from the previous section, we consolidated the 49,143 reviews from the three different product categories together, at the reviewer level. In other words, we placed all the reviews in a new unique dataset and ordered the dataset using the ID of the reviewer. The 49,143 reviews in the dataset were developed by 46,701 reviewers. Putting the reviews of different categories together allows the study to reproduce the ability of reviewers to contribute to different product categories at the same time. A PCA analysis with Varimax rotation was implemented. The results of this analysis are shown in Table 7. A fixed number of 4 factors was employed as extraction criterion, since the fourth factor has an Eigenvalue of 0.983 while allowing for a better interpretation of the loadings of variables among the 4 components (loadings higher than .40 are highlighted in red).

Table 7. PCA Analysis for the Consolidated Variables (Varimax Rotation).

Rotated Component Matrix [a]

	Component			
	1	**2**	**3**	**4**
Number of Stars	-0.004	-0.005	-0.005	**0.999**
ARI	**0.707**	0.061	0.159	0.016
Posting Order	-0.116	**0.893**	-0.099	-0.013
Sentiment Score	**0.486**	**0.475**	0.158	0.021
Corrob. Entropy	**0.825**	0.057	0.086	-0.001
Opinion Entropy	**0.765**	-0.043	0.089	-0.008
Recom. Entropy	**0.557**	-0.146	-0.277	-0.024
Number Help. Eval.	0.125	-0.055	**0.928**	-0.008

Rotation Method: Varimax with Kaiser Normalization

[a.] Rotation converged in 5 iterations

Initial Eigenvalues

Component	**Total**	**% of Variance**	**Cumulative %**
1	**2.413**	30.159	30.159
2	**1.060**	13.244	43.403
3	**1.000**	12.502	55.905
4	0.983	11.723	67.628
5	0.874	--	--

The pattern of association among the information variables is observed again. However, the sentiment score cross-loaded also on Factor 1, but at a much lower coefficient than the four information variables. To verify the potential impact of Sentiment Score on our measure, we conducted a correlation analysis, included in Appendix A, which suggests that sentiment score has a weak correlation with Factor 1 (Cohen, 1988), of the same magnitude as Number

of Helpfulness Evaluations, which increases our confidence in the new measure to rank reviews and reviewers in a more objective way. None of the remaining variables that potentially account for the sources of bias interferes with the variables with relevant loadings in Factor 1. The scores of Factor 1 are employed again as a measure of information content as well as a measure of how well that information is conveyed. (The four information-related variable loadings are positive and with values over .5.) The reviewer ranking was developed by ordering reviewers from the largest to the smallest value of the score of Factor 1.

To compare reviewer ranking based on information with the helpfulness votes rankings, we aggregated the total number of reviews developed by each reviewer. This approach is common for other review sites that prioritize the total number of reviews posted, such as TripAdvisor (hereafter TripAdvisor Rank), as shown in Figure 1. Figure 5 shows the ten first reviewers as ranked by the number of reviews method (TripAdvisor Rank) and our suggested information method (Info Rank). Table 8, shows that there is practically a zero correlation between the information ranking and the ranking of the raw number of reviews posted by individual users.

TripAdvisor Rank	Unique Reviewer ID	Reviews Written	Percent Helpful	Days Last Review
1	ATDE9JYCPI0L1	22	79.570	1311
2	A3KEZLJ59C1JVH	19	66.667	977
3	A2KSBMNVDCCTYH	15	62.712	2979
4	A1M04H40ZVGWVG	11	32.000	1182
5	A34BZM6S9L7QI4	11	100.000	99
6	AZWXG6KBXXC2N	10	100.000	127
7	A3ATKSR23880M1	9	33.333	324
8	A3R9H6OKZHHRJD	9	100.000	25
9	A3T6HX48F66512	9	60.000	157
10	A1IBKN1GXYJ3K4	8	97.333	1813
Info Rank	**Unique Reviewer ID**	**Cumulative Factor 1**		**Days Last Review**
1	ATDE9JYCPI0L1	38.023		1311
2	A3KEZLJ59C1JVH	25.023		977
3	A2V5R832QCSOMX	21.700		809
4	A34BZM6S9L7QI4	21.586		99
5	ATLA5CJH0ZZHP	16.449		196
6	A231WM2Z2JL0U3	16.316		4244
7	A1IBKN1GXYJ3K4	16.124		1813
8	A2KSBMNVDCCTYH	15.936		2979
9	A3T6HX48F66512	15.769		157
10	A1S1SSE7HQDZEW	15.578		288

Figure 5. Ten First Reviewers Ordered by Tripadvisor Rank and by Info Rank.

Table 8. Spearman's ρ and Kendall's τ Correlation Coefficients for the Two Reviewer Ranking Methods.

Ranking 1	Ranking 2	Spearman's rank correlation ρ	*p*-value	Kendall's rank correlation τ	*p*-value
TripAdvisor Rank	Info Rank	.083	< 2.2e-16	.056	< 2.2e-16

6. Discussion

Through different measures of information content and a measure of readability, this methodological paper suggests a new, single approach to evaluate online reviews and reviewers in real-time. This is in support of the need to improve the process of review helpfulness evaluation in addition to building trust in online reviews as conveying valid information to make informed purchase decisions. The low correlation between the current measures of review helpfulness – based on user votes – and more objective measures of information content, as the one developed in this study – shown in sections 4 and 5 – suggest the existence of the theorized gap between the two meanings of helpfulness (Danescu-Niculescu-Mizil et al., 2009). Indeed, several sources of bias were reported in the literature, impacting helpfulness measures based on votes and shaping the so-called helpfulness "in the wild" – i.e. the way helpfulness is currently evaluated by users. This study controlled for these in its methodology development, trying to disclose the underlying "narrow sense" – i.e. product information assisting consumers making decisions. Ultimately, these results

contribute to the emerging literature that put into question the reliability of the current helpfulness assessment approaches and urge academicians and practitioners to rethink the way they tackle consumer review helpfulness.

In addition to this methodological contribution of our study, the suggested methodology can equally benefit consumers and marketing professionals. More informative reviews can allow consumers to overcome the common problem of information overload (Filieri, 2015; Hong et al., 2017; Wang et al., 2024). A study conducted by Brightlocal.com (BrightLocal, 2019) found that consumers read, on average, ten reviews before trusting a company. Therefore, if more informative reviews are presented in the top ten list, this might allow consumers to make better purchase decisions, leading to higher levels of satisfaction. High quality information presented to customers through online consumer reviews may offer increased potential value to customers and provide a source of differentiation for e-retailers (Mudambi & Schuff, 2010). Nevertheless, these might not be the only benefits for marketers.

Marketing professionals can focus their analysis and base their decisions on more informative reviews, as a major source of insights for marketing intelligence (Schweidel et al., 2011; Yang et al., 2015) and product improvement and development (Decker & Trusov, 2010). Equally relevant, practitioners can grasp a better knowledge of how information posted online might impact sales. Research also suggests that, because of their information content, every review posted might have an intrinsic economic value (Wu et al., 2015), higher for "truly" helpful reviews. This new methodology can better lead marketers in their search of those high-informational reviews, and those reviewers who provide abundant and quality information.

Understanding the role of information entropy may allow for better modeling of what constitutes a better online review, and thereby possibly identifying reviews that are most beneficial to customers. As a final potential contribution of the study, these benefits can be achieved through one single methodology, reducing complexity and even computation burden.

6.1. Limitations and Avenues for Future Research

This study has several limitations that open new avenues for future research. The suggested methodology selected one of the many Sentiment Analysis approaches available. Future research may apply other methods to account for sentiments in the text of reviews, such as Machine Learning methods or methods that include a dependency parser or even extend the analysis to identify and asses emotions included in online reviews. Additionally, this research only accounts for explicit recommendations as they were highlighted as prevalent in the literature (Packard & Berger, 2017) and in the content analysis implemented by Lu et al. (2014). Although Packard and Berger (2017) found that implicit recommendations' role might be less relevant than explicit ones, future research may also include these types of recommendations in the methodological approach. Future research can also assess the tradeoff between more granularity in the opinion category – a more detailed sub-categorization of opinions – versus less generalizability of this more granular opinion category. Finally, our suggested methodology is based on information content, as previous literature suggested that more information can be more helpful (on average) in reducing uncertainty (Sun et al., 2019; Tan, 1999). Future research can empirically valuate the suitability of our proposed information measure in reducing consumer's uncertainty.

7. Conclusion

Online consumer reviews are a relevant source of pre-purchase information that help consumers reduce uncertainty. As a common practice, consumers rely on product reviews to make more informed purchase decisions. However, the ability of these reviews to fulfill this task is hampered by several issues that affect the validity and accuracy of the information provided in these reviews. One of the major problems is how the quality of this information is assessed through helpfulness votes. The over-abundance of online reviews led to another key problem: information overload (Fang et al., 2016; Wang et al., 2024). Arguably more concerning, consumer skepticism toward online reviews has also grown in the last few years (Wen & Ha, 2024). This study suggests an approach to identify and present to consumers reviews that contain valuable information, as well as a way to recognize reviewers for their contribution. A less subjective approach than the current helpfulness votes systems might be increasingly relevant for consumers, practitioners, and scholars.

Data Availability Statement

Data Availability Statement: Data available upon reasonable request.

REFERENCES

Agnihotri, A., & Bhattacharya, S. (2016). Online Review Helpfulness: Role of Qualitative Factors. *Psychology & Marketing*, *33*(11), 1006-1017. https://doi.org/10.1002/mar.20934

Amazon-Staff. (2025). Everything you need to know about Amazon reviews. *News*. Retrieved 09/02/2025, from https://www.aboutamazon.com/news/retail/amazon-customer-reviews-star-ratings

Amazon. (2025). Understanding Customer Reviews and Ratings. *Help & Customer Service*. https://www.amazon.com/gp/help/customer/display.html?ref_=hp_gcs_csd_d2_other_5_G8UYX7LALQC8V9KA_zXU06txjse&nodeId=G8UYX7LALQC8V9KA&qid=1755625760312&sr=5

Aral, S. (2014). The Problem With Online Ratings. Retrieved 01/03/2025, from http://sloanreview.mit.edu/article/the-problem-with-online-ratings-2/

Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of Online Consumer Reviews: Readers' Objectives and Review Cues. *International Journal of Electronic Commerce*, *17*(2), 99–126. https://doi.org/10.2753/JEC1086-4415170204

Bakos, Y., & Dellarocas, C. (2011). Cooperation Without Enforcement? A Comparative Analysis of Litigation and Online Reputation as Quality Assurance Mechanisms. *Management Science*, *57*(11), 1944-1962. https://doi.org/10.1287/mnsc.1110.1390

Belzer, J. (1973). Information Theory as a Measure of Information Content. *Journal of the American Society for Information Science*, *24*(4), 300-304. https://doi.org/10.1002/asi.4630240411

Bennett, P. (1999). Understanding responses to risk: Some basic findings. In P. Bennett & K. Calman (Eds.), *Risk communication and public health* (pp. 3-32). Oxford University Press.

Bickart, B., & Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing*, *15*(3), 31-40. https://doi.org/10.1002/dir.1014

Box, D., & Croker, S. (2018). Fake five-star reviews being bought and sold online. Retrieved 01/03/2025, from https://www.bbc.com/news/technology-43907695

BrightLocal. (2019). Local Consumer Review Survey. Retrieved 01/03/2025, from https://www.brightlocal.com/learn/local-consumer-review-survey/

Brown, J. J., & Reingen, P. H. (1987). Social Ties and Word-of-Mouth Referral Behavior. *Journal of Consumer Research*, *14*(3), 350-362. https://www.jstor.org/stable/2489496

Buschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science*, 35(6), 953. https://doi.org/10.1287/mksc.2016.0993

Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems*, *50*(2), 511-521. https://doi.org/10.1016/j.dss.2010.11.009

Chen, P.-Y., Dhanasobhon, S., & Smith, M. D. (2008). All reviews are not created equal: The disaggregate impact of reviews and reviewers at amazon. com. *Available at SSRN: http://ssrn.com/abstract=918083*.

Chong, A. Y. L., Li, B. Y., Ngai, E. W. T., Ch'ng, E., & Lee, F. (2016). Predicting online product sales via online reviews, sentiments, and promotion strategies A big data architecture and neural network approach. *International Journal of Operations & Production Management*, *36*(4), 358-383.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Lawrence Erlbaum Associates.

Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., & Lee, L. (2009). How opinions are received by online communities: a case study on amazon. com helpfulness votes. Proceedings of the 18th international conference on World wide web, 141-150, https://doi.org/10.1145/1526709.1526729

De Langhe, B., Fernbach, P. M., & Lichtenstein, D. R. (2016). Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Ratings. *Journal of Consumer Research*, *42*(6), 817-833. https://doi.org/10.1093/jcr/ucv047

Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, *27*(4), 293-307.

Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, *21*(4), 23-45.

Dong, H., Hou, Y., Hao, M., Wang, J., & Li, S. (2021). Method for Ranking the Helpfulness of Online Reviews Based on SO-ILES TODIM. *IEEE Access*, *9*, 1723-1736. https://doi.org/10.1109/ACCESS.2020.3040151

Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter? — An empirical investigation of panel data. *Decision Support Systems*, *45*(4), 1007-1016.

Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology*, *38*(1), 188-230.

Dwoskin, E., & Timberg, C. (2018). How merchants use Facebook to flood Amazon with fake reviews. Retrieved 01/03/2025, from https://www.washingtonpost.com/business/economy/how-merchants-secretly-use-facebook-to-flood-amazon-with-fake-reviews/2018/04/23/5dad1e30-4392-11e8-8569-26fda6b404c7_story.html?noredirect=on&utm_term=.1278d5cb59a6

Engle, J. (2023). Do You Trust Online Reviews? *The New York Times*. https://www.nytimes.com/2023/11/16/learning/do-you-trust-online-reviews.html

ePlaybooks. (2023). Amazon Vine Program: Everything Sellers Should Know. *ePlaybooks*. Retrieved 12/18/2024, from https://www.eplaybooks.com/post/amazon-vine-program

Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, *52*, 498-506. https://doi.org/10.1016/j.tourman.2015.07.018

Filieri, R. (2015). What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. *Journal of Business Research*, *68*(6), 1261-1270.

Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., & Freling, T. (2014). How Online Product Reviews Affect Retail Sales: A Meta-analysis. *Journal of Retailing*, *90*(2), 217-232. https://doi.org/10.1016/j.jretai.2014.04.004

Folse, J. A. G., Porter Iii, M., Godbole, M., & Reynolds, K. E. (2016). The Effects of Negatively Valenced Emotional Expressions in Online Reviews on the Reviewer, the Review, and the Product. *Psychology & Marketing*, *33*(9), 747-760. https://doi.org/10.1002/mar.20914

Fresneda, J. E., & Gefen, D. (2019). A semantic measure of online review helpfulness and the importance of message entropy. *Decision Support Systems*, *125*, 113117. https://doi.org/10.1016/j.dss.2019.113117

Fresneda, J. E., & Gefen, D. (2020). Gazing at the stars is not enough, look at the specific word entropy, too! *Information & Management*, *57*(8), 103388. https://doi.org/10.1016/j.im.2020.103388

Ghose, A., & Ipeirotis, P. G. (2007). Designing novel review ranking systems: predicting the usefulness and impact of reviews. *Proceedings of the ninth international conference on Electronic commerce*, 303 - 310, https://doi.org/10.1145/1282100.1282158

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics. *IEEE Transactions on Knowledge and Data Engineering*, *23*(10), 1498-1512.

Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science*, *31*(3), 493-520. https://doi.org/10.1287/mksc.1110.0700

Godes, D., & Mayzlin, D. (2004). Using Online Conversations to Study Word-of-Mouth Communication. *Marketing Science*, *23*(4), 545-560. https://doi.org/10.1287/mksc.1040.0071

Goldsmith, R. E., Lafferty, B. A., & Newell, S. J. (2000). The Impact of Corporate Credibility and Celebrity Credibility on Consumer Reaction to Advertisements and Brands. *Journal of Advertising*, *29*(3), 43–54. https://doi.org/10.1080/00913367.2000.10673616

Gopinath, S., Thomas, J. S., & Krishnamurthi, L. (2014). Investigating the Relationship Between the Content of Online Word of Mouth, Advertising, and Brand Performance. *Marketing Science, 33*(2), 241-258. https://doi.org/10.1287/mksc.2013.0820

Harrison, A. (2025). How to Increase Upvotes on Amazon? Retrieved 8/18/2025, from https://sagemailer.com/blog/how-to-boost-upvotes-on-amazon-and-improve-your-products-rating-honestly/

Hausser, J., & Strimmer, K. (2009). Entropy Inference and the James-Stein Estimator, with Application to Nonlinear Gene Association Networks. *Journal of Machine Learning Research*, *10*, 1469-1484.

Hong, H., Xu, D., Wang, G. A., & Fan, W. (2017). Understanding the determinants of online review helpfulness: A meta-analytic investigation. *Decision Support Systems*, *102*, 1-11. https://doi.org/https://doi.org/10.1016/j.dss.2017.06.007

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 168-177.* https://doi.org/10.1145/1014052.1014073

Hu, N., Bose, I., Koh, N. S., & Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems*, *52*(3), 674-684. https://doi.org/10.1016/j.dss.2011.11.002

Hu, N., Pavlou, P. A., & Zhang, J. (2017). On Self-Selection Biases in Online Product Reviews. *MIS Quarterly*, *41*(2), 449-A417.

Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. *Communications of the ACM*, *52*(10), 144-147.

Humphreys, A., & Jen-Hui Wang, R. (2018). Automated Text Analysis for Consumer Research. *Journal of Consumer Research*, *44*(6), 1274-1306.

Jindal, N., & Liu, B. (2006). Mining comparative sentences and relations. *Proceedings of the 21st National Conference on Artificial Intelligence*, 2, 1331-1336

Karimi, S., & Wang, F. (2017). Online review helpfulness: Impact of reviewer profile image. *Decision Support Systems*, *96*(Supplement C), 39-48. https://doi.org/https://doi.org/10.1016/j.dss.2017.02.001

Kashyap, R., Kesharwani, A., & Ponnam, A. (2023). Measurement of online review helpfulness: a formative measure development and validation. *Electronic Commerce Research*, *23*(4), 2183-2216.

Korfiatis, N., García-Bariocanal, E., & Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, *11*(3), 205-217.

Kung, H., Orlando, D., & Kirimlioglu, E. (2024). Were You Helpful - Predicting Helpful Votes from Amazon Reviews. *arxiv* Retrieved 08/27/2025, from https://arxiv.org/html/2412.02884v1

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2013). *Handbook of latent semantic analysis*. Psychology Press.

Lee, T. Y., & Bradlow, E. T. (2011). Automated Marketing Research Using Online Customer Reviews. *Journal of Marketing Research*, *48*(5), 881-894. https://www.jstor.org/stable/23033526

Li, M., Huang, L., Tan, C.H., & Wei, K.K. (2013). Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features. *International Journal of Electronic Commerce*, *17*(4), 101-136.

Li, N., & Wu, D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, *48*(2), 354-368. https://doi.org/10.1016/j.dss.2009.09.003

Li, X., & Hitt, L. M. (2008). Self-Selection and Information Role of Online Product Reviews. *Information Systems Research 19*(4):456-474. https://doi.org/10.1287/isre.1070.0154

Liu, B., & Hu, M. (2019). *Opinion Lexicon*

Liu, J., Cao, Y., Lin, C.-Y., Huang, Y., & Zhou, M. (2007). Low-Quality Product Review Detection in Opinion Summarization. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 334–342, Prague, Czech Republic. Association for Computational Linguistics

Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, *47*, 140-151. https://doi.org/https://doi.org/10.1016/j.tourman.2014.09.020

Lu, S., Wu, J., & Tseng, S.-L. (2018). How Online Reviews Become Helpful: A Dynamic Perspective. *Journal of Interactive Marketing*, *44*, 17-28. https://doi.org/https://doi.org/10.1016/j.intmar.2018.05.005

Lu, X., Li, Y., Zhang, Z., & Rai, B. (2014). Consumer Learning Embedded in Electronic Word of Mouth. *Journal of Electronic Commerce Research*, *15*(4), 300-316.

Machta, J. (1999). Entropy, information, and computation. *American Journal of Physics*, *67*(12), 1074-1077.

Mafael, A., Gottschalk, S. A., & Kreis, H. (2016). Examining Biased Assimilation of Brand-related Online Reviews. *Journal of Interactive Marketing*, *36*, 91-106.

Maidar, U., Ra, M., & Yoo, D. (2024). A Cross-Product Analysis of Earphone Reviews Using Contextual Topic Modeling and Association Rule Mining. *Journal of Theoretical and Applied Electronic Commerce Research*, *19*(4), 3498-3519.

Mankad, S., Han, H., Goh, J., & Gavirneni, S. (2016). Understanding Online Hotel Reviews through Automated Text Analysis. *Service Science*, *8*(2), 124-138.

Mayzlin, D. (2006). Promotional Chat on the Internet. *Marketing Science*, *25*(2), 155-163.

McAuley, J., Pandey, R., & Leskovec, J. (2015). Inferring networks of substitutable and complementary products. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 1-10. https://doi.org/10.1145/2783258.2783381

McAuley, J., Targett, C., Shi, J., & van den Hengel, A. (2015). Image-based recommendations on styles and substitutes. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 10, 43-52. https://doi.org/10.1145/2766462.2767755

Min, H.-J., & Park, J. C. (2012). Identifying helpful reviews based on customer's mentions about experiences. *Expert Systems with Applications*, *39*(15), 11830-11838.

Moe, W. W., & Trusov, M. (2011). The Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing Research*, *48*(3), 444-456. https://doi.org/10.1509/jmkr.48.3.444

Mudambi, S. M., & Schuff, D. (2010). What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly*, *34*(1), 185-200.

Negi, S., & Buitelaar, P. (2015). Towards the Extraction of Customer-to-Customer Suggestions from Reviews. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal.

Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine Your Own Business: Market-Structure Surveillance Through Text Mining. *Marketing Science*, *31*(3), 521-543.

Packard, G., & Berger, J. (2017). How Language Shapes Word of Mouth's Impact. *Journal of Marketing Research*, *54*(4), 572-588.

Pan, Y., & Zhang, J. Q. (2011). Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews. *Journal of Retailing*, *87*(4), 598-612.

Purnawirawan, N., Eisend, M., De Pelsmacker, P., & Dens, N. (2015). A Meta-analytic Investigation of the Role of Valence in Online Reviews. *Journal of Interactive Marketing*, *31*, 17-27.

Qazi, A., Shah Syed, K. B., Raj, R. G., Cambria, E., Tahir, M., & Alghazzawi, D. (2016). A concept-level approach to the analysis of online review helpfulness. *Computers in Human Behavior*, *58*, 75-81.

Qin, J., Zheng, P., & Wang, X. (2022). Comprehensive helpfulness of online reviews: A dynamic strategy for ranking reviews by intrinsic and extrinsic helpfulness. *Decision Support Systems*, *163*, 113859.

Racherla, P., Mandviwalla, M., & Connolly, D. J. (2012). Factors affecting consumers' trust in online product reviews. *Journal of Consumer Behaviour*, *11*(2), 94-104. https://doi.org/10.1002/cb.385

Ramirez, A., Walther, J. B., Burgoon, J. K., & Sunnafrank, M. (2002). Information Seeking Strategies, Uncertainty, and Computer-Mediated Communication. Toward a Conceptual Model. *Human Communication Research*, *28*(2), 213-228.

Rosario, A. B., Sotgiu, F., De Valck, K., & Bijmolt, T. H. A. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, *53*(3), 297-318. https://doi.org/10.1509/jmr.14.0380

Rubin, B. F. (2015). *Amazon looks to improve customer-reviews system with machine learning*. CNET. Retrieved 01/03/2025 from http://www.cnet.com/news/amazon-updates-customer-reviews-with-new-machine-learning-platform/

Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, *81*, 30-40.

Sánchez-Franco, M. J., Navarro-García, A., & Rondán-Cataluña, F. J. (2016). Online Customer Service Reviews in Urban Hotels: A Data Mining Approach. *Psychology & Marketing*, *33*(12), 1174-1186.

Saumya, S., Singh, J. P., Baabdullah, A. M., Rana, N. P., & Dwivedi, Y. K. (2018). Ranking online consumer reviews. *Electronic Commerce Research and Applications*, *29*, 78-89.

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. https://doi.org/10.1002/cb.1372

Schmalz, M., Carter, M., & Lee, J. H. (2018). It's Not You, It's Me: Identity, Self-Verification, and Amazon Reviews. *SIGMIS Data Base for Advances in Information Systems, 49*(2), 79-92.

Schweidel, D. A., Moe, W. W., & Boudreaux, C. (2011). Social media intelligence: measuring brand sentiment from online conversations. https://doi.org/10.2139/ssrn.1874892

Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the Web. *Journal of Interactive Marketing*, *21*(4), 76-94.

Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, *80*(2), 159-169.

Senter, R. J., & Smith, E. A. (1967). *Automated readability index*.

Shan, Y. (2016). How credible are online product reviews? The effects of self-generated and system-generated cues on source credibility evaluation. *Computers in Human Behavior*, *55, Part B*, 633-641. https://doi.org/10.1016/j.chb.2015.10.013

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Tech J*, *27*, 379-423, 623-656.

Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Kumar Roy, P. (2017). Predicting the "helpfulness" of online consumer reviews. *Journal of Business Research*, *70*, 346-355. https://doi.org/10.1016/j.jbusres.2016.08.008.

Sun, X., Han, M., & Feng, J. (2019). Helpfulness of online reviews: Examining review informativeness and classification thresholds by search products and experience products. *Decision Support Systems*, *124*, 113099. https://doi.org/10.1016/j.dss.2019.113099

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, *37*(2), 267-307.

Tan, S. J. (1999). Strategies for reducing consumers' risk aversion in Internet shopping. *Journal of Consumer Marketing*, *16*(2), 163-180.

Ullah, R., Amblee, N., Kim, W., & Lee, H. (2016). From valence to emotions: Exploring the distribution of emotions in online product reviews. *Decision Support Systems*, *81*, 41-53.

Princeton University.(2010). *WordNet*. Princeton University. Retrieved 01/03/2025 from http://wordnet.princeton.edu

Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management*, *30*(1), 123-127. https://doi.org/ 10.1016/j.tourman.2008.04.008

Villarroel Ordenes, F., Ludwig, S., De Ruyter, K., Grewal, D., & Wetzels, M. (2017). Unveiling what is written in the stars: Analyzing explicit, implicit, and discourse patterns of sentiment in social media. *Journal of Consumer Research*, *43*(6), 875-894.

Viswanathan, S. (2022). Should we trust online reviews? *Consumer Alerts*. Retrieved 12/16/2024, from

Wan, Y. (2015). The Matthew Effect in social commerce. *Electronic Markets*, *25*(4), 313-324. https://doi.org/10.1007/s12525-015-0186-x

Wan, Y., & Nakayama, M. (2014). The reliability of online review helpfulness. *Journal of Electronic Commerce Research*, *15*(3), 179-189.

Wang, F., & Karimi, S. (2019). This product works well (for me): The impact of first-person singular pronouns on online review helpfulness. *Journal of Business Research*, *104*, 283-294. https://doi.org/10.1016/j.jbusres.2019.07.028.

Wang, F., Liu, X., & Fang, E. (2015). User Reviews Variance, Critic Reviews Variance, and Product Sales: An Exploration of Customer Breadth and Depth Effects. *Journal of Retailing*, *91*(3), 372-389.

Wang, L., Che, G., Hu, J., & Chen, L. (2024). Online Review Helpfulness and Information Overload: The Roles of Text, Image, and Video Elements. *Journal of Theoretical and Applied Electronic Commerce Research*, *19*(2), 1243-1266.

Weathers, D., Swain, S. D., & Grover, V. (2015). Can online product reviews be more helpful? Examining characteristics of information content by product type. *Decision Support Systems*, *79*, 12-23. https://doi.org/10.1016/j.dss.2015.07.009

Wen, T., & Ha, H.-Y. (2024). An Empirical Test of the Impact of the Online Review–Review Skepticism Mechanism on Behavioral Intentions: A Time-Lag Interval Approach between Pre- and Post-Visits in the Hospitality Industry. *Journal of Theoretical and Applied Electronic Commerce Research*, *19*(3), 2070-2087.

Wild, F., Stahl, C. (2007) Investigating Unstructured Texts with Latent Semantic Analysis, in: Lenz, D. (Ed.), Advances in Data Analysis. Springer.

Willemsen, L. M., Neijens, P. C., & Bronner, F. (2012). The ironic effect of source identification on the perceived credibility of online product reviewers. *Journal of Computer-Mediated Communication*, *18*(1), 16-31. https://doi.org/10.1111/j.1083-6101.2012.01598.x

Wu, C., Che, H., Chan, T. Y., & Lu, X. (2015). The Economic Value of Online Reviews. *Marketing Science*, *34*(5), 739-754. https://doi.org/10.1287/mksc.2015.0926

Wu, P. F. (2013). In Search of Negativity Bias: An Empirical Study of Perceived Helpfulness of Online Reviews. *Psychology & Marketing*, *30*(11), 971-984.

Yang, C.-S., Chen, C.-H., & Chang, P.-C. (2015). Harnessing consumer reviews for marketing intelligence: a domain-adapted sentiment classification approach. *Information Systems and e-Business Management*, *13*(3), 403-419. https://doi.org/10.1007/s10257-014-0266-z

Yang, L., & Li, Y. (2022). A new method for ranking the usefulness of negative online reviews based on combined weighting method and improved TOPSIS. *Journal of Intelligent & Fuzzy Systems*, *42*(4), 3719-3736. https://doi.org/10.3233/JIFS-211928

You, Y., Vadakkepatt, G. G., & Joshi, A. M. (2015). A Meta-Analysis of Electronic Word-of-Mouth Elasticity . *Journal of Marketing*, *79*(2), 19-39.

Zhu, F., & Zhang, X. (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*, *74*(2), 133-148.

Zhuang, M., Cui, G., & Peng, L. (2018). Manufactured opinions: The effect of manipulating online product reviews. *Journal of Business Research,* 87, 24-35. https://doi.org/10.1016/j.jbusres.2018.02.016

Appendix

Part 1

The purpose of this first part of the appendix is to suggest potential measures to validate the semantic measure of online review helpfulness. To do so, 500 reviews were selected to be employed as a benchmark due to their unanimously positive evaluation through helpfulness votes. Table A.1 shows descriptive statistics for this subset of reviews from our data set (this subset hereinafter referred as BEST500).

Table A1. Descriptive statistics for the BEST500 subset.

Variable	N	Mean	Standard Deviation	Median	Min	Max	Skewness	Kurtosis
Semantic Review Helpfulness	500	0.14	0.02	0.14	0.06	0.21	-0.80	1.78
Helpfulness Vote Ratio	500	0.91	0.13	0.94	0.00	1.00	-3.95	22.68
Helpfulness Thumb-up Votes	500	38.88	45.70	26	15.00	546	5.87	47.19
Information Entropy Increment	500	0.36	0.23	0.35	-0.95	0.69	-0.67	1.54
Sentiment Score Deviation	500	5.67	5.95	4.00	0.00	48.00	2.63	10.98
ARI	500	6.91	2.56	6.76	-1.44	16.09	0.38	0.85
Number of Helpfulness Evaluations	500	42.26	48.02	30.00	15.00	552.00	5.70	43.81
Number of Words in Review	500	280.18	276.78	194.50	26.00	2137.00	2.61	9.54
Number of Words in Product Description	500	107.76	140.44	63.00	3.00	1157.00	4.08	23.13
Posting Order	500	3.72	10.03	1.00	1.00	110	7.35	59.89
Number of Stars	500	3.38	1.71	4.00	1.00	5.00	-0.44	-1.55

The remaining 20,222 reviews constitute a second subset of the dataset (this subset hereinafter referred as MINUS500). As a first remark, we observed that the semantic helpfulness measure is on average 21% higher for the BEST500 subset than for the MINUS500 one (0.14 versus 0.11, respectively). A new semantic space was built from the subset MINUS500. The cosine similarity of each review with all the remaining reviews in the semantic space was calculated within this new space. In other words, we calculated the cosine similarity of review 1 with review 2, review 1 with review 3, review 1 with review 4 and so forth. Then, we calculated the average of this cosine similarity of each review with the remaining reviews in the data set. Following the previous example, we calculated the average similarity of review 1 with all the remaining reviews, from review 2 to review 20,222. This procedure created a new variable: the average semantic similarity of each review with all the remaining reviews in MINUS500.

The subset BEST500 was then folded-in the semantic space created with MINUS500. Folding-in new documents into an existing semantic space allows to project those documents into the space without changing its structure (Wild and Stahl, 2007). We calculated the average cosine similarity of each of the reviews within BEST500 using the space created from MINUS500. This average was used as an actual benchmark to be compared with the new variable calculated previously (the average semantic similarity of each review with all the remaining reviews in MINUS500).

We calculated the deviation of each of the reviews in MINUS500 from the benchmark calculated previously as a simple subtraction. Correlation coefficients of this deviation (subtraction) and the semantic measure of review helpfulness, the helpfulness vote ratio, and the helpfulness thumb-up votes were finally calculated and shown in Table A.2 (The helpfulness vote ratio and the helpfulness thumb-up votes are popular operationalizations of review helpfulness in previous literature (Fresneda and Gefen, 2019).)

Table A2. Correlation coefficients of the different helpfulness measures and the deviation.

Correlated variables	Correlation coefficient
Deviation / Semantic Review Helpfulness	0.31
Deviation / Helpfulness Vote Ratio	0.09
Deviation / Helpfulness Thumb-up Votes	0.12

Then, we created a new subset (named MINUS500.2) to contain only those reviews with thumb-up votes. The new subset contains 6,115 reviews. Removing the many reviews that were not evaluated in their helpfulness might eliminate noise in the analysis. No thumb-up votes over a review can mean either that the review was not evaluated

or that the information conveyed is not helpful. Therefore, within MINUS500.2, those reviews with positive helpfulness votes should be clearly more similar to those included in BEST500 and vice versa.

The entire analysis was replicated using the subset MINUS500.2 instead of the previous MINUS500. Table A.3 shows the results of the correlation coefficients of this replication, showing a much stronger correlation of the deviation and the semantic measure of review helpfulness than in Table A.2. Figure A.1 summarizes our validation process.

Table A3. Correlation coefficients of the different helpfulness measures and the deviation (replication with MINUS.500.2).

Correlated variables	Correlation coefficient
Deviation / Semantic Review Helpfulness	0.53
Deviation / Helpfulness Vote Ratio	0.02
Deviation / Helpfulness Thumb-up Votes	0.20

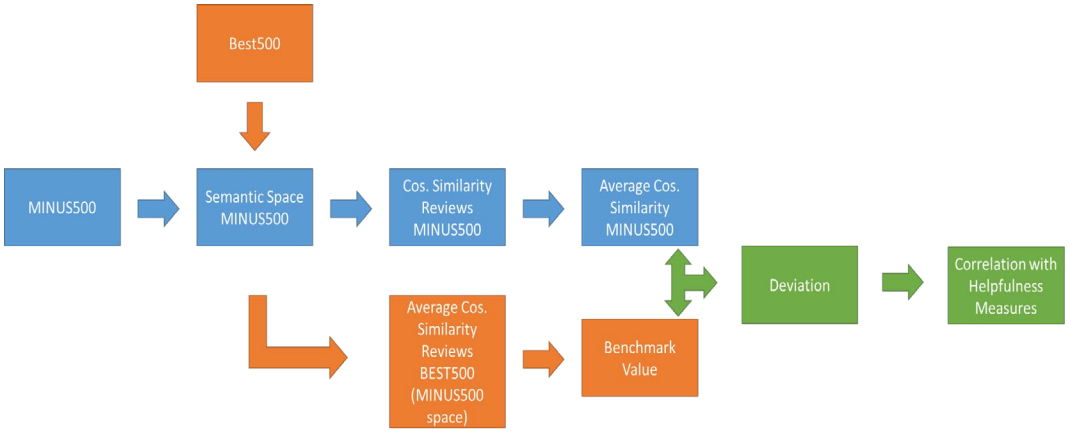

Figure A1. Validation method process

Part 2

This second part of the appendix presents a correlation analysis of all the variables included in the dataset used to develop the reviewer rank and Factor 1, in other words, we analyzed the correlation among all the variables included in Table 6 in the manuscript plus the newly developed Factor 1. We marked in red those variables with a strong correlation using Cohen's (1988) rule of thumb. As shown, sentiment score has a weak correlation with Factor 1 (Cohen, 1988), of the same magnitude as Number of Helpfulness Evaluations, which increases our confidence in the new measure to rank reviews and reviewers in a more objective way.

	N. Stars	ARI	Post. Order	Sent Sc.	Corr. Ent.	Op. Ent.	Recom. Ent.	N. Help. Eval.	Factor 1
N. Stars	**1.000**								
ARI	0.002	**1.000**							
Post. Order	-0.003	-0.079	**1.000**						
Sent Sc.	0.001	0.091	0.073	**1.000**					
Corr. Ent.	-0.005	0.329	-0.058	0.172	**1.000**				
Op. Ent.	-0.005	0.260	-0.071	0.107	0.457	**1.000**			
Recom. Ent.	-0.003	0.111	-0.027	0.077	0.246	0.250	**1.000**		
N. Help. Eval.	-0.002	0.103	-0.055	0.054	0.115	0.136	0.058	**1.000**	
Factor 1	-0.028	**0.557**	-0.133	0.229	**0.765**	**0.742**	**0.551**	0.233	**1.000**