

Article

Trustworthiness of Review Opinions on the Internet for 3C Commodities

Ying-Chia Hsieh ^{1,*}, Long-Chuan Lu ¹ and Ruen-Jung Yang ²¹ Department of Business Administration, National Chung-Cheng University, Chia-Yi 621301, Taiwan; bmalcl@ccu.edu.tw² Department of Information Management, National Chung-Cheng University, Chia-Yi 621301, Taiwan; rjyang@mis.ccu.edu.tw

* Correspondence: carrie@mis.ccu.edu.tw

Abstract: The rapid development of the internet has resulted in rapid e-business growth, with online malls attracting many shoppers due to the privacy and convenience they offer. Like traditional malls, online malls can provide photos, specifications, prices, etc. However, consumers cannot touch the products in reality, which creates risks for the purchase. To date, there has been no research focusing on topic-specific search engines for 3C product reviews based on the trustworthiness of the reviews. This study is the first to sort the reviews of electronic products according to the degree of trust, by analyzing the characteristics of the reviews and the reviewers. This study proposes the criteria for features of the reviews and reviewers to consider to evaluate the trustworthiness of the reviews; builds a search engine to collect the product reviews scattered in opinion websites; and sorts the results by trustworthiness to provide a reliable e-commerce experience. To demonstrate the effectiveness of the proposed method, we conducted a set of experiments, and we adopted the Spearman's rank correlation coefficient to evaluate the similarity between our method and experts' opinions. The experimental results showed a high correlation coefficient with the opinions of experts, demonstrating that our method is effective at finding trustworthy reviews on the internet.

Keywords: trust; trustworthiness evaluation; review; topic-specific search engine; artificial intelligence



Citation: Hsieh, Y.-C.; Lu, L.-C.; Yang, R.-J. Trustworthiness of Review Opinions on the Internet for 3C Commodities. *Electronics* **2024**, *13*, 1346. <https://doi.org/10.3390/electronics13071346>

Academic Editor: Ying Tan

Received: 31 January 2024

Revised: 18 March 2024

Accepted: 29 March 2024

Published: 3 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Due to the rapid development of the ability to easily search e-businesses, many large e-malls have emerged, such as eBay, Yahoo, Amazon, and YouTube. E-malls attract many people to shop on the internet due to their convenience, low price, and privacy [1,2]. Compared with traditional malls, e-malls can only provide photos, specifications, prices, and other visualizations. Consumers cannot touch the products before ordering them, which forces consumers to bear risks in purchasing the items [3,4]. Thus, some consumers pursue professional opinions or the usage experience of other consumers before they make a decision about purchasing. With the development of the internet, more and more consumers are accustomed to sharing their experience and reviews of products, forming numerous online opinion-sharing communities. According to the literature [5,6], the users of e-commerce online opinion-sharing communities significantly affect product sales. The current well-known online communities, such as Amazon, eBay, and Taobao, provide users' comments on a product at the bottom of the product page. There are also communities that focus entirely on product reviews and opinion-sharing services [6,7], such as epinions.com, which categorizes discussion topics to facilitate the querying and sharing of comments according to the user's interests and needs.

To easily provide product reviews, some websites, or even the e-malls themselves, provide a platform or a board where the consumers can share their usage experience, including an evaluation of the product and/or comparisons with other products. For example, the Epinions and Amazon websites provide reviews and information about

their products to help potential consumers understand the function of the product and decide whether the product meets their requirements [8]. Since these reviews and those in opinion-sharing websites are written based on consumers' own experiences of using the product, they have a greater influence on potential consumers than advertisements. Since these platforms or boards are open to the public, some business managers may disguise themselves as consumers to advocate for their own products or criticize their opponent's products; even worse, some managers may hire, bribe, or give benefits to reviewers to promote their products or demote their opponents' products. This results in the problem of not all reviews being reliable [8]. To avoid spam or malicious reviews, some mechanisms, such as voting or comments attached to the reviews, have been adopted by Epinions and Amazon to provide methods through which consumers can judge the trustworthiness of the reviews [9]. However, the logic behind these mechanisms is too naive, and the reviews' votes and comments can also be manipulated by a group of collusive users [10,11]. Thus, consumers still need to decide the trustworthiness of reviews for themselves. In the case where consumers are not sensitive to potential deception, they may be misguided and make an inappropriate decision. In addition, there may be many reviews for a popular product on the same or different opinion websites. To obtain all the available information, consumers must take the time to collect and browse the reviews for the product (called target reviews) on all opinion websites. Clearly, a search engine that can collect all the opinions about a product and rank them according to their trustworthiness could help consumers make an efficient and appropriate decision. To this end, this study developed a search engine that collects the opinions on a product from the internet and ranks them according to proposed criteria for trustworthiness.

Previous studies proposed some mechanisms for detecting spam reviews and for predicting and ranking helpful reviews [12,13] on various opinion websites. Research [14] on spam detection has incorporated the technique of natural language processing (NLP) to process reviews. Since it is difficult to perform sentiment analysis, it is difficult to distinguish between spam and genuine reviews through natural language processing, and the accuracy of spam detection is not as high as expected. Some research on ranking helpful reviews proposed a machine learning-based approach for ranking comments [15,16]. Their approaches considered several factors that reflect the trustworthiness of reviews, which can be divided into three categories. The first set of factors is the content-based features of the reviews, such as their length and the complexity of the comments. The second category contains features related to the reviewer's reputation and characteristics, such as the reviewer's registration time and the number of reviews written by the reviewer. The factors in the third category are related to review visibility, such as the posting time of the review. Prior research evaluated the helpfulness of the reviews on an opinion website, but it did not consider some other important factors, such as the number of approvals of a review against that of other reviews and the reputation of the reviewers, each of which can significantly reflect the trustworthiness of the review. This study considers more factors than previous works to evaluate the trustworthiness of reviews and proposes a set of criteria (i.e., factors) for evaluating the trustworthiness of reviews, constructing a search engine to collect the target reviews scattered among different opinion websites. Finally, the collected target reviews are ranked using the proposed criteria. To demonstrate the effectiveness of the proposed search engine, it was used to collect more 80,000 reviews for mobile phones from the three main opinion websites (Epinions, CNET, and Amazon) and to rank them with the proposed "Trustworthy Review Ranking System" (TRRS). We invited several experts who are familiar with the functions of these mobile phones to evaluate whether the ranking of opinions by the proposed search engine was appropriate. We utilized the Spearman's correlation coefficient to compare the rankings between our method and the experts' opinions. We use three sets of weights to measure the effectiveness of our method, and the results showed a positive Spearman's rank correlation coefficient between our method and the experts' opinions, indicating that our method is effective in searching and ranking reviews based on trustworthiness. By analyzing the behavior of the

reviewers and the review characteristics and then sorting the reviews according to the level of trustworthiness, the users or potential customers can easily understand the reputation of the electronic product and reduce the screening time for the product. The results of this paper can be adopted by web-based commerce companies to calculate the impact of online reviews on their commercial online community and to help them to quickly detect fraudulent, biased, collusive, or exaggerated reviews.

2. Literature Review

2.1. Concept of Trust

Trust is a subjective concept in which an individual believes that another person can reliably bring benefits and/or give satisfaction [17,18]. Although trust is an abstract concept, some scholars have devoted themselves to quantifying trust [3]. The trustworthiness of someone is defined as the probability that their behavior can meet another person's expectation. A higher trustworthiness of an individual indicates a higher possibility that this individual can satisfy the expectations of others. In traditional society, people establish trust between each other based on face-to-face interactions. However, due to the convenience of the internet, people have fewer face-to-face interactions. Relational trust is not applicable for calculating the trustworthiness of an individual in a network society, since most people interact with each other through email, telephone, or blogs, without face-to-face interactions. For example, in an online shopping scenario, consumers purchase merchandise through an online platform. They do not have any physical interaction with the sellers during the whole purchasing process. The trustworthiness of the seller is calculated based on the reputation of the seller and/or feedback on transactions from other consumers before trading with them. The calculation of the trustworthiness of the participating persons in an auction website is another example. An auction website can calculate the trustworthiness of a participant from feedback on the participant's past transactions [19]. The public can use this information to help them judge whether to trust the person they are considering dealing with. There are also no face-to-face interactions between users on opinion websites. We adopted calculative trust in this study to evaluate and rank the trustworthiness of reviewers and the collected reviews through the reviewer's reputation and the content of the reviews. With this ranking, consumers can efficiently and effectively access trustworthy reviews.

2.2. Spam Detection

Previous studies have shown that consumers' purchasing intentions are significantly influenced by online reviews [20]; clearly, whether the reviews are genuine is important for the readers of the reviews. Due to the low cost of spreading information through the internet, unsolicited messages (often called spam) permeate the internet, including in emails, blogs, forums, online communities, and video streaming websites. Spam is often spread without permission from the receivers, and the content of the spam is often unhelpful to the receivers. There are several reasons for the spread of spam, such as disseminating pornography, phishing, and spreading advertisements and viruses. Spam not only occupies the network but also wastes the user's time filtering it out. Automatically detecting or filtering spam is an important task to prevent users from being exposed to high volumes of information that is essentially useless. The spam detection methods used in previous studies included machine learning methods used to build a model such as a neural network that is trained using the appropriate data. The model can learn the features of the training data and is then used to classify new input data. Machine learning methods can be further divided into two subtypes: supervised learning and unsupervised learning. The supervised learning method feeds spam and non-spam data into the model to train an inferred model, with classifier and regression functions, for spam prediction.

Unlike the supervised learning methods, unsupervised learning methods do not require pre-labeled data to establish an inferred model. Since the training data given to the method are unlabeled, there are no feedback rewards from the training data to the method

when establishing the inferred model. The unsupervised learning method is more suitable for problems where the characteristics of the problem continuously evolve or change. For example, owing to the different types of email spam, spam cannot be easily labeled as a specific class. The classifiers of supervised learning for email spam often suffer from low discriminant rates. Some researchers [14] proposed an incremental cluster-based method to address this problem. This method extracts characteristics from each group as criteria to classify the email. The cluster-based method then uses these criteria to classify the emails and spam. Compared to the supervised learning method, the execution time of unsupervised learning methods is relatively large. Since the training data are unlabeled, the unsupervised learning method has to classify the training data by itself.

In general, machine learning methods can quickly detect spam. However, nowadays, many spam reviews can disguise themselves as genuine reviews with different appearances to avoid being detected [21]. A review is considered spam if there are many duplicates across different opinion websites. This type of spam review often disguises itself by replacing the title or contents with synonymous or similar words to avoid detection. Other prejudiced spam reviews may have overly enthusiastic or negative attitudes towards the product. These two types of spam reviews cannot be identified by means of machine learning methods, since these methods do not consider semantics. Natural language processing, like sentiment analysis, has been proposed as a solution to this problem [15,22]. Natural language processing can explore the structure and meaning of a text, transforming the semantics or sentiments, which can be analyzed by the system to better understand the sentiment and attitude of the reviewer and to detect extraordinary attitudes for the purpose of spam detection. The natural language processing method can find more hidden spam on the internet than machine learning methods, but the execution time of natural language processing methods is very long since it is difficult to retrieve the semantics of a review. This study adopted the NLP technique to determine the attitudes of both the reviews and the comments on the reviews to consider whether the reviews met with approval in comments from others and whether any reviews were outliers compared to other reviews for the same product.

2.3. Topic-Specific Search Engines

Search engines are the most commonly used tool on the internet. Traditional search engines are general-purpose tools that can find a variety of results for a query. However, for a user who is only interested in a specific category of results, the wide array of results from general-purpose search engines will slow the user in browsing through the results. Therefore, topic-specific search engines that primarily focus on specific content have been proposed to meet the demands for information on a specific topic [23]. The concept of topic-specific search engines is widely applied in different domains to provide a convenient way to search for information on the web. For example, Pricewatch is a price comparison search engine [24]. Consumers can use it to find the cheaper and cheapest products on the internet. Compared to general-purpose search engines that equip a program (called a crawler or spider) to retrieve all the information on the internet, topic-specific search engines equip a similar spider (called a focus crawler) to only retrieve information related to the specific topic. Although the goal and collected information of the general-purpose and topic-specific search engines are different, the topic-specific search engines, like general-purpose search engines, also rank the results with similar popularity metrics. However, these topic-specific search engines rarely take into consideration the possibility that the content of a webpage may be fake. The concept of trust is not considered as a factor in ranking the search results. Therefore, users cannot judge whether the ranked search results can be trusted or not. The amount of information on the internet is huge, and the demand for searching for specific topics has increased. Some studies in the literature have used new computer architectures, such as Spark architecture, to efficiently analyze the huge amounts of data generated by mobile devices [25]. However, to date, there has been no research focusing on topic-specific search engines for product reviews. This paper proposes

a topic-specific search engine for product reviews with its ranking criterion being the trustworthiness of the review. Since the product reviews may be fake, the ranking of these reviews by the proposed topic-specific search engine plays an important role in reducing the risk of consumers buying inappropriate products.

3. Proposed Calculation of Trustworthiness

The proposed topic-specific search engine, focusing on reviews, collects most of the product reviews published on the internet. The search engine first retrieves all the product reviews on the internet through its focus crawler and stores them in a database. Then, the TRRS analyzes the characteristics of the reviews and calculates the trustworthiness of these reviews. When a user enters a query about a specific product, the TRRS returns the pre-ranked reviews to the user. The structure of the TRRS can be divided into three modules: the crawler module, the processing module, and the user interface module (Figure 1).

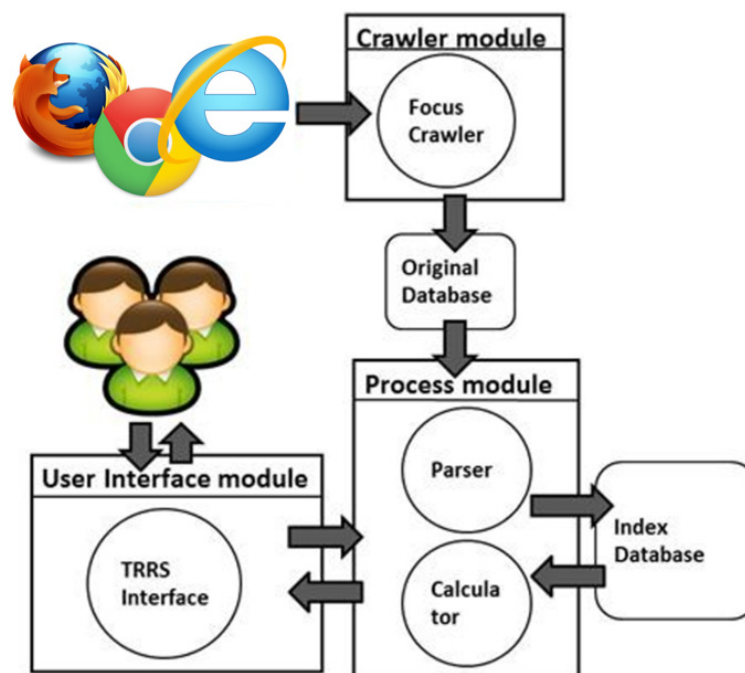


Figure 1. Operation process of TRRS.

In the crawler module, a focus crawler is used to retrieve the product reviews from the internet. The focus crawler periodically searches for product reviews on major opinion websites such as Amazon, CNET, etc. The specified products whose reviews can be retrieved are given in advance; the crawler will only retrieve the reviews whose context contains the keywords related to the specified product. Then, the crawler divides the related reviews into several fields, such as title, content, date, author name, etc., and stores the divided data in a database, which is processed by the next module. In the processing module, there are two functional operators, namely, a parser and a calculator. The parser analyzes the content of the reviews in the database and analyzes the attitude of the reviews through natural language processing (discussed later), then stores the categorization of the attitude in the index database. The attitude category is used to filter out outliers whose attitude is opposite to the attitude of most of the other reviews for the same products. The method for identifying outliers is discussed in the next subsection. The calculator filters out certain factors, such as the length of the review and the number of visitors to the review, which are used to calculate the trustworthiness of the review (also discussed later). Then, the calculator sorts the reviews by trustworthiness; the ranked reviews are used by the user interface module. In the user interface module, users only need to type their keywords into the interface of the TRRS, and then the interface returns results ranked

based on trustworthiness. Compare to traditional search engines, this topic-specific ranking allows users to browse reviews with ease and confidence. Thus, users can access more trustworthy reviews through the TRRS.

The processing module containing the parser and calculator ranks the reviews according to the trust factors that affect the trustworthiness of the reviews. In this study, these trust factors were grouped into two categories, namely, reviewer authority and review authority, as depicted in Figure 2. Since the influence of these trust factors varies, the TRRS gives individual factors different weights to balance the inherent bias in developing the criteria for the trustworthiness of reviews.

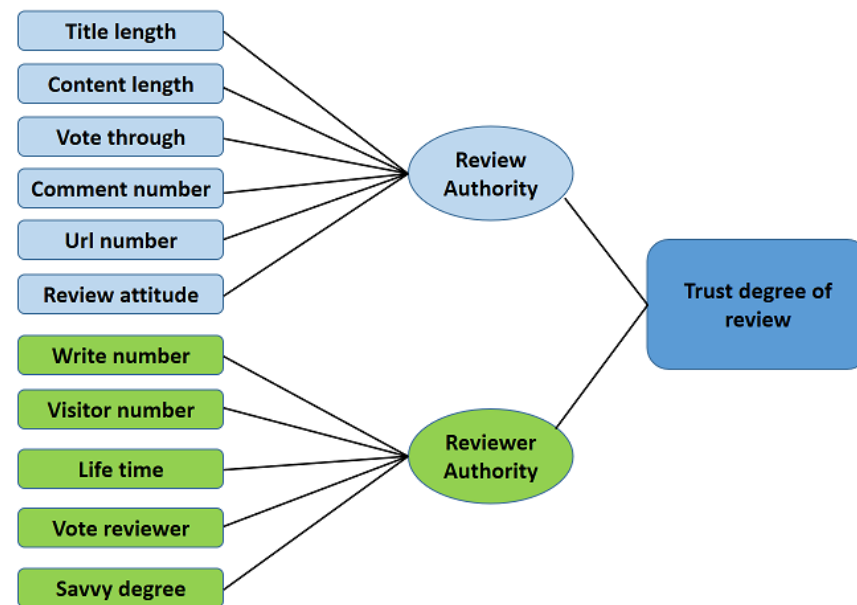


Figure 2. Classification of the trust factors.

The review authority contains the following factors: the title length and content length of the review, the number of comments and the vote through for the review, and the URL number and the attitude of the review, which are related to the characteristics of a review.

The factor of the title length of a review is denoted by C_{TL} . Jindal and Liu [26] indicated that the longer the review's title, the more comments and feedback the review will receive, and that if a review attracts users to give feedback, the review has a lower possibility of being a spam review. Thus, we can reason that if a review has a longer title, the review has a lower possibility of being spam; in other words, the review has a higher possibility of being trustworthy. Let $|C_{TL}(c)|$ denote the length of the title of review c . The lengths of the titles of reviews may vary significantly. Thus, this study used the natural logarithm to reduce the variation in the lengths. Therefore, the measurement of the factor of the title length of a review is defined as the value of $\log(|C_{TL}(c)|)$. Note that for the other factors discussed later, if their values varied significantly, we also used the natural logarithm to reduce their variation to a smaller range.

Similarly, the factor of a review's content length is denoted by C_{CL} . Hsu, Khabiri, and Caverlee [27] indicated that the length of a review's content can be a criterion to measure the usefulness of the review. The longer the content of a review, the more useful the review is. In general, a useful review has a lower possibility of being spam. Thus, we can also reason that if a review has a longer content, the review's content has a lower possibility of being spam; in other words, the review has a higher possibility of being trustworthy. Let $|C_{CL}(c)|$ denote the length of the content of review c . Then, the measurement of the factor of the content length of a review is defined as the value of $\log(|C_{CL}(c)|)$.

After a review is posted on the internet, some readers may browse the review for reference. These readers may agree or disagree with the viewpoint of the review, and they can vote for or against the review or provide feedback through comments. Most opinion

websites provide a mechanism that allows readers to vote for the review. Thus, other readers can judge the reliability of the review from the voting ratio. If the review is a spam review, the voting ratio will be low. Jindal and Liu [26] also indicated that the voting ratio can be a criterion to measure whether a review is a spam review. Let the factor of the voting ratio of a review be denoted as C_{VT} ; then, the measurement of the factor of the voting ratio for review c is defined as the value of $\log(C_{VT}(c))$.

In addition to voting for a review, most opinion websites also provide other mechanism like commenting to compliment or criticize the review. Generally, the more comments a review receives, the more attractiveness it has. In most cases, if a review is a spam review, it will not attract people's attention and receive feedback. Thus, we can infer that if a review obtains many comments, the review has a lower possibility of being a spam review; in other words, the review has a higher possibility of being trustworthy. Let the factor of the number of comments for review c be denoted as $C_{CN}(c)$; then, the measurement of the factor of the number of comments for review c is defined as the value of $\log(C_{CN}(c))$.

The number of the hyperlinks contained in a review can be a clue for identifying advertisement reviews. Benevenuto et al. [28] indicated that spam reviews usually contain hyperlinks in their content. Thus, we can reason that if a review has more hyperlinks in its content, the review has a higher possibility of being spam; in other words, the review has a lower probability of being trustworthy. Let the factor of the number of URLs for review c be denoted as $C_{UN}(c)$; then, the measurement of the factor of the number of URLs in the review is defined as the value of $\log(C_{UN}(c))$.

The above factors implicitly influence the trustworthiness of a review. The content of a review contains more clues to identify spam reviews. In general, a review for a product always reveals a positive or negative attitude towards the product. Yoo and Gretzel [29] indicated that some spam reviews have extraordinary attitudes among all the reviews for the same product. That is, if an attitude of a review is against that of most of the other reviews, the review has a higher probability of being a spam review. In this study, we adopted the technique of natural language processing [30] to preprocess the reviews and determine the attitudes of the reviews. In detail, the content of a review is first parsed into tokens, and then the neutral tokens are filtered, such as nouns, pronouns, "be" or similar verbs, articles, etc. The remaining tokens are used to search the WordNet to obtain the attitude of these tokens, and the numbers of positive and negative words are summed. The attitude score is defined as the number of positive words minus the number of negative words. The same calculation for attitude is applied to all the other reviews for the same product. The mean and the standard deviation of the attitude score for the same product can then be derived. A review is judged as an extraordinary one if its attitude score falls three standard deviations away from the mean of the attitude scores for the product.

Another obvious clue in the content of a review is the attitude of the comments from other users. Let P_k denote the k th product in the opinion website, $C_i^{P_k}$ denote the i th review for the k th product, and $AP(C_i^{P_k})$ denote the attitude score of review i for the k th product. The mean, μ_k , of a review's attitude for product k is defined as the expression $\sum_{i=1}^{N_k} (AP(C_i^{P_k}) / N_k)$, where N_k denotes the number of reviews on the website for product P_k . The standard deviation, σ_k , of a review's attitude score for product k is defined as the expression $\sqrt{1 / N_k \sum_{i=1}^{N_k} (AP(C_i^{P_k}) - \mu_k)^2}$. In summary, for a review C that discusses the k th product, the standard score (i.e., its Z-value) of the review's attitude is denoted as $Z(C^{P_k})$. Using the definition of Z-values, $Z(C^{P_k})$ equals $(AP(C^{P_k}) - \mu_k) / \sigma_k$. Note that if the attitude of a review is significantly different from the average attitudes of all the reviews, that review may be biased, faulty, and unreliable. Thus, the trustworthiness of a review is inversely proportional to the absolute Z-value, denoted as $|Z(c)|$, of the attitudes of the review.

In addition to the characteristics of a review, this study also considered the characteristics of the reviewer to measure the trustworthiness of a review. In general, a trustworthy reviewer often writes trustworthy reviews; thus, we can use reviewer authority to measure

the trustworthiness of reviews. Relational trust is not applicable in a network society, so calculative trust is the main way to measure the trustworthiness of an individual in an online community. In this study, we considered the reviewer's reputation and past behavior to measure the trustworthiness of a review. By analyzing reviewers' behaviors, we can obtain the number of reviews, number of visitors, and registration time (or life time). The factor of the number of reviews written by a reviewer is denoted by R_{WN} . Hsu, Khabiri, and Caverlee [27] indicated that the number of reviews can be a criterion to measure a reviewer's usefulness. The more reviews a reviewer writes, the more experience the reviewer has. In general, an experienced reviewer writes more useful reviews than novice reviewers. Thus, we can reason that if a reviewer writes more reviews, the reviewer has a lower possibility of being a spam reviewer; in other words, the reviewer has a higher possibility of being trustworthy. Let $R_{WN}(r)$ denote the number of reviews written by reviewer r . Then, the measurement of the factor of number of reviews of the reviewer is defined as the value of $\log(R_{WN}(r))$.

Similarly, the factor of the number of visitors of a reviewer, denoted as R_{VN} , is the number of people to ever visit the reviewer's profile in this social community. Generally, the more visitors the reviewers have, the more attractiveness they have. If a reviewer is a spam reviewer, they cannot attract people to visit them, and the visitors are fewer than those for genuine reviewers. Thus, we can reason that if more people visit the reviewer, the reviewer will have a lower possibility of being a spam reviewer, in other words, the reviewer has a higher possibility of being trustworthy. Let $R_{VN}(r)$ denote the number of people who have ever visited the reviewer r . Then, the measurement of the factor of the number of visitors of the reviewer is defined as the value of $\log(R_{VN}(r))$.

The factor of the registration time (denoted as life time) of a reviewer, denoted as R_{LT} , is the period of time for which the reviewer has been registered in this social community. Benevenuto et al. [28] indicated that the accounts of spam reviewers are usually live for shorter periods compared to genuine reviewers, and the life time of spam reviewers usually does not exceed one year. Thus, we can reason that if the reviewer has a longer life time, the reviewer has a lower probability of being a spam reviewer; in other words, the reviewer has a higher probability of being trustworthy. Let $R_{LT}(r)$ denote the period of time that the reviewer r has been registered in this social community. Then, the measurement of the factor of the reviewer's registration time is defined as the value of $\log(R_{LT}(r))$. In addition to the reviewer's behavior, the reputation of the reviewer is another important factor for measuring the trustworthiness of a reviewer. In this paper, the reviewer's reputation was based on the voting ratio and the degree of savviness of the reviewer.

Generally, if the voting ratio for a reviewer is high, it means that the reviewer is a reliable reviewer. In turn, a reliable reviewer can write more trustworthy reviews than a novice reviewer. Thus, we can reason that if a reviewer has a higher voting ratio, the reviewer has a higher reputation, and their reviews have a higher probability of being trustworthy. Let the voting ratio for reviewer r be denoted by $R_{VR}(r)$. Then, the measurement of the factor of the voting ratio is defined as the value of $\log(R_{VR}(r))$.

Jøsang et al. [3] indicated that reputation systems, such as auction websites, forums, and opinion websites, are trustworthy. The ranking result of opinion websites can also be trusted. Thus, we can reason that if a reviewer has a higher rank, the reviewer has a lower probability of being a spam reviewer; in other words, the reviewer will have a higher probability of being trustworthy. The factor of the degree of savviness of reviewer r is denoted by $R_{SD}(r)$ for this opinion website. Then, the factor of the degree of savviness of the reviewer is defined as the inverse value of $R_{SD}(r)$, i.e., $1/R_{SD}(r)$.

Since the above two authorities in Figure 2 affect the trustworthiness of reviews, we can take each factor into consideration to represent their influence in the trustworthiness evaluation. For simplicity, we used the linear combination of each factor with a weight. Then, the above factors can be summed up into a measurement. The trustworthiness of a review can be formulated as below. In detail, let $TDR(c)$ denote the trustworthiness of

review c written by reviewer r . The weights in the formula associated with the factors are aligned in the same order as in Figure 2. Then,

$$\begin{aligned} \text{TDR}(c) = & W_{TL} \times \log(|C_{TL}(c)|) + W_{CL} \times \log(|C_{CL}(c)|) + W_{VT} \times \log(R_{VT}(r)) + \\ & W_{CN} \times \log(C_{CN}(c)) + W_{UN} \times \log(C_{UN}(c)) - W_{RA} \times |Z(c)| + \\ & W_{WN} \times \log(R_{WN}(r)) + W_{VN} \times \log(R_{VN}(r)) + W_{LT} \times \log(R_{LT}(r)) + \\ & W_{VR} \times \log(R_{VR}(r)) + W_{SD} \times \left(\frac{1}{R_{SD}(r)} \right), \end{aligned}$$

where the total weight is 1, namely, $W_{all} = 1$

$$= W_{TL} + W_{CL} + W_{VT} + W_{CN} + W_{UN} + W_{RA} + W_{WN} + W_{VN} + W_{LT} + W_{VR} + W_{SD}.$$

Since the content, comments on the review, and (limited) information on the reviewer can be obtained from online reviews, this information was considered the easiest to obtain from websites. We surveyed nearly all the recent papers that investigated the characteristics of reviews and reviewers, and we proposed potential factors that influence the trustworthiness of reviews, which are summarized in Figure 2. Referring to the literature, this study considered the trustworthiness of reviews from their intrinsic characteristics, like their content length, title length, number of URLs contained, comments on the review (voting ratio), and the number of comments. In addition, the reviewer's characteristics were also considered, since they can also reflect the trustworthiness of the review; these included the number of reviews, number of visitors, life time, and degree of savviness.

Note that each factor in the review authority (i.e., the title length, content length, number of comments, voting ratio, review attitude, and number of URLs) can be considered to be independent from the others. For example, the title length has no relationship with the content length, number of URLs, voting ratio, comment attributes, and number of comments in most cases, and the title length has no relationship with all the other considered factors (shown in Figure 2) of the reviewer authority. Regarding the other considered factors of the reviewer authority (i.e., the number of reviews, number of visitors, life time, voting ratio for reviewers, and degree of savviness), the first three factors (namely, the number of reviews, number of visitors, and life time) in some situations may have positive relationships with each other. For example, if the reviewer has a longer life time, they may write more reviews and have a higher number of visitors, and vice versa. However, in most cases, their relationships are too vague to evaluate. The influences of the first three factors are summed in the formula, thus influencing the trustworthiness of the review. Similarly, in some websites, the last two factors of the reviewer authority, i.e., the voting ratio and the degree of savviness of the reviewer, are also summed up to show their influences.

4. Results and Discussion

4.1. Experiments and Results

To demonstrate the effectiveness of the proposed search engine for trustworthy reviews for 3C products, we constructed a search engine in the PHP language. Since there are too many product reviews on the internet to be completely crawled in a limited amount of time, we selected three well-known opinion websites, i.e., Epinions, Amazon, and CNET, to retrieve the reviews for mobile phones. We used a focus crawler to retrieve product reviews from November 2021 to February 2022 from these three websites, resulting in more than 80,000 product reviews being added to our database. Note that the Epinions website is a paid review website. We obtained reviews from the website <https://paperswithcode.com/dataset/epinions> (accessed on 29 March 2023). In this study, three experiments were performed, since the influence of the selected trust factors may vary. The TRRS gives each factor a weight to balance the inherent bias in developing criteria for the trustworthiness of reviews. To ensure generalizability, in our experiments, we used three different sets of weights for each experiment. The first set of weights assumes that all trust factors have the same influence on the measurement of trustworthiness. The second set of weights

assumes that the reviewer authority has more influence than the review authority, since many online websites use calculative trust to evaluate the trustworthiness of consumers and suppliers; thus, the reviewer's reputation and past behavior are important in measuring trustworthiness on the internet. The third set of weights was decided by general rules of thumb.

In our experiments, we compared our results with the opinions of experts who have used the specific mobile phones. Seven experts participated, five of them male and two of them female. All of them have a great deal of experience in e-shopping for 3C commodities; two of them come from the faculties of business departments of universities, another two are colleagues in our laboratory, and the remaining three were invited experts from the internet. We applied the Spearman's rank correlation coefficient to evaluate the similarity between our results and the experts' opinions. A high correlation coefficient indicates that our method is effective in finding trustworthy reviews on the internet. We analyzed the reviews for three mobile phone models: Apple iPhone, Samsung Galaxy, and Google Pixel.

In our experiment, we selected the top 10 reviews using the TRRS for the three products, and we then compared the result with the experts' ranking (Figure 3). The blue lines in each sub-figure of Figure 3 denote the review results from the experts for the three products, which were the baseline for comparing the review results obtained using the TRRS. The red lines in Figure 3 denote the ranking of the reviews for the three products. The Spearman's rank coefficient, ρ , which was used to measure the correlation between the experts' opinions and the TRRS results, is shown in each sub-figure. The smaller the value of ρ , the higher the correlation.

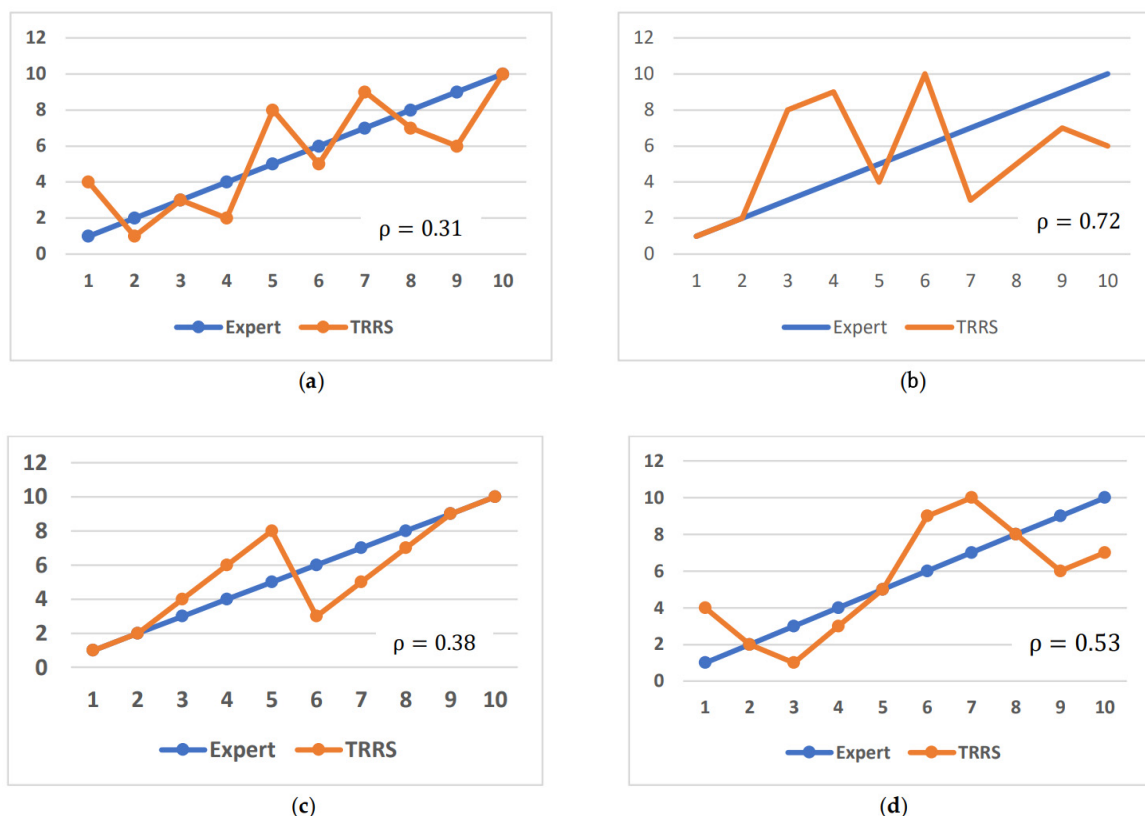


Figure 3. Cont.

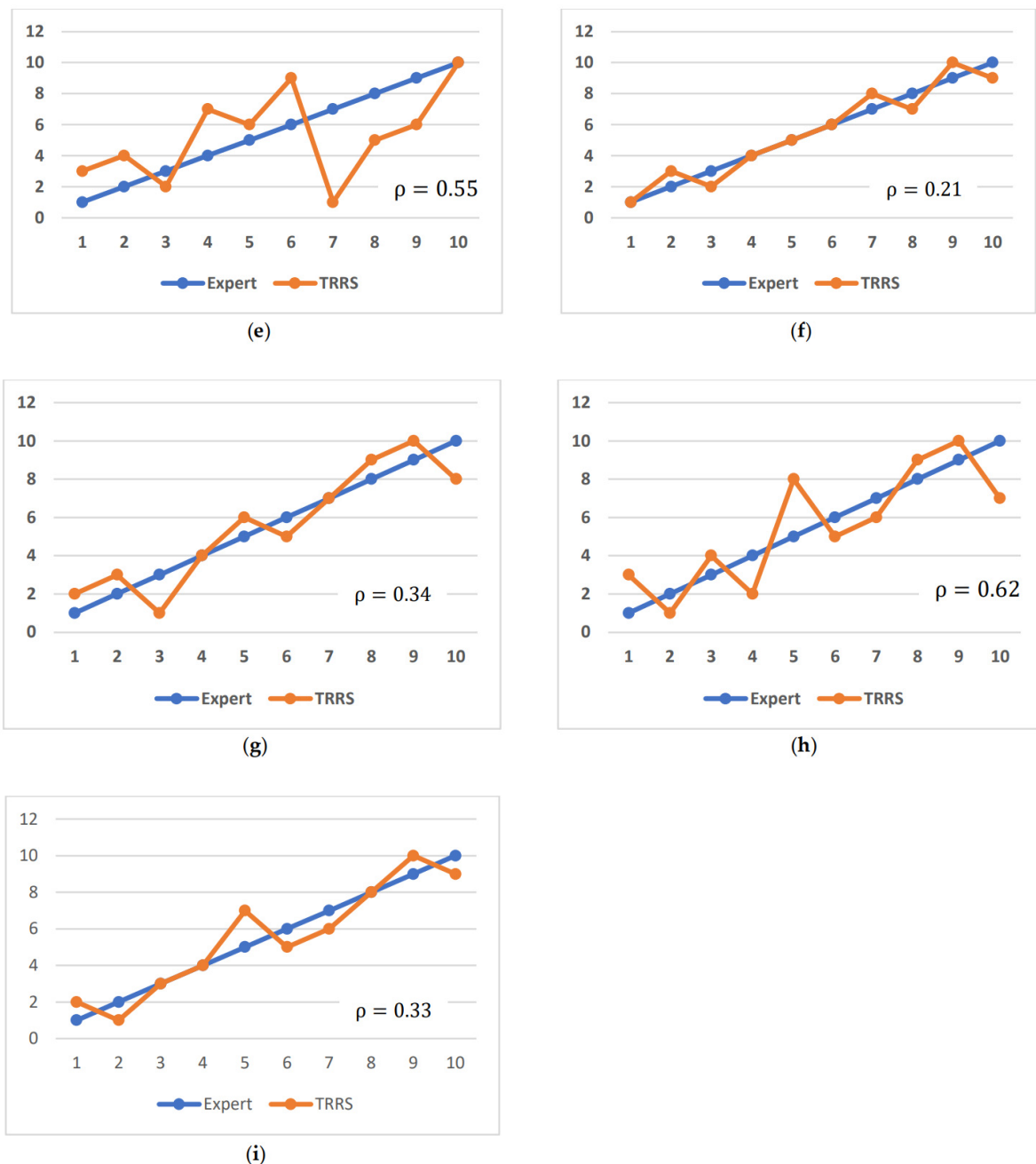


Figure 3. Ranking of reviews by experts and the TRRS using three different weight sets: (a) First product with the first weight set; (b) second product with the first weight set; (c) third product with the first weight set; (d) first product with the second weight set; (e) second product with the second weight set; (f) third product with the second weight set; (g) first product with the third weight set; (h) second product with the third weight set; and (i) third product with the third weight set.

Since this study focused on the trustworthiness of reviews, metrics such as precision, the recall rates, and others, which are used for evaluating the performance of classifiers, were not suitable for this problem since the evaluation results of the trustworthiness of a review fall on a spectrum ranging from untrustworthy to trustworthy, rather than binary results. Thus, we did not use these metrics. Instead, we adopted the Spearman's correlation coefficient to compare our ranking results with those from the experts.

In the above three experiments with different sets of weights, it was evident that the influences of the selected trust factors were different, and we could achieve a better performance by adjusting the weights. According to the results of the experiment, we can

see that the results with the second set of weights were worse than those with the other sets. Thus, we can reason that consumers may not significantly consider the reviewer's reputation and past behavior, and they only focus on information about the characteristics of the review in judging the trustworthiness of reviews. Despite the sets of weights being different, the proposed method still produced a positive result for all experiments, thus demonstrating the effectiveness of the proposed ranking mechanism.

4.2. Discussion

The weight of each factor is crucial to the final ranking. Generally, it is easy to precisely gauge the weights for factors for tangible materials/phenomena, such as the weather, earthquakes, nutrition, etc. However, the weights of the factors influencing the trustworthiness of a review are complex, since trustworthiness is an intangible concept. In this study, we attempted an Analytic Hierarchical Process (AHP) [31], where we invited experts to try to weight the factors through pairwise comparisons of the factors. However, the experts could not finish the comparisons of the factors; for example, they could not give a weight for 'number of comments' against the 'title length', 'content length', etc., since these factors are related to cognition and are hard to gauge numerically.

For such intangible concepts, trial-and-error or rule-of-thumb may be a possible way to weight the factors. A Principal Component Analysis (PCA) is a linear dimensionality reduction technique that is used to identify the principal components. Some redundant components can be removed if they consist of a set of other orthogonal components. We weighted each factor but did not use a solid analytic technique to determine the correlation between the considered factors. The reason for this is that the review data were written and responded to by humans; thus, some of the above considered factors, such as the voting ratio and number of comments for the review and the number of visitors and number of votes for the reviewer, may be null. In addition, the correlations among the factors are vague as there have been no studies in the literature quantifying these correlations. We finally used three different sets of weights for each factor (Tables 1–3) to determine which set of weights was more suitable by comparing the results with the experts' rankings. This is a feasible way to obtain suitable weights for factors for such an intangible concept after several trials.

Table 1. The first set of weights.

Weight	W_{TL}	W_{CL}	W_{VT}	W_{CN}	W_{UN}	W_{RA}	W_{WN}	W_{VN}	W_{LT}	$W_{VR'}$	W_{SD}
Value	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09

Table 2. The second set of weights.

Weight	W_{TL}	W_{CL}	W_{VT}	W_{CN}	W_{UN}	W_{RA}	W_{WN}	W_{VN}	W_{LT}	$W_{VR'}$	W_{SD}
Value	0.033	0.033	0.033	0.033	0.033	0.033	0.16	0.16	0.16	0.16	0.16

Table 3. The third set of weights.

Weight	W_{TL}	W_{CL}	W_{VT}	W_{CN}	W_{UN}	W_{RA}	W_{WN}	W_{VN}	W_{LT}	$W_{VR'}$	W_{SD}
Value	0.1	0.3	0.1	0.1	0.05	0.05	0.1	0.05	0.05	0.05	0.05

Clearly, trial-and-error can also be applied to product types or categories other than 3C products since the characteristics of reviews and reviewers for the different categories of products are similar. A possible difference between 3C and products of other categories is the terms used in the review and the evaluation method for the degree of savviness of the reviewers. The weights of the considered factors in this paper for 3C commodities may be different to those for different categories of products, such as travel and restaurant evaluations [13], since reviews for different products will have different focuses and attitudes and

a different reviewer group. Thus, our method needs to be modified accordingly to adapt to different categories of products.

This paper investigated the factors of reviews posted on 3C commercial websites. The proposed trustworthiness equation is based on the reviews and information of the reviewers obtained via crawling from the websites. Clearly, the disclosed information about the characteristics of the reviews, the characteristics of the reviewers, and linguistic considerations, such as the compatibility of the language, the judgment of sentiments, etc., significantly influenced the effectiveness and completeness of the data when evaluating the trustworthiness of reviews. A confidentiality and protection policy for the reviewers is respected by each e-commercial website. In cases where the unveiled data are de-identified or even when the reviews are converted into statistical data, the proposed method cannot work well. Clearly, the degree of data disclosed to the public is a limitation to this method.

5. Conclusions

In this paper, a search engine to find trustworthy reviews was designed using an innovative method of ranking reviews for users. The method included many trust factors used to evaluate the trustworthiness of reviews. Since product reviews are scattered across many opinion websites and the number of product reviews is huge, users have to take the time to collect product reviews. However, there has been no prior research focusing on topic-specific search engines for product reviews. Our proposed search engine collects reviews and ranks them based on trustworthiness, reducing a user's risk of buying an inappropriate product. Biased, fraudulent, or exaggerated opinions may influence or mislead customers to make the wrong decision. If their shopping experience is negative, consumers may not refer to the review mechanism provided by the website in the future. Thus, a trustworthiness evaluation and ranking of reviews is important and needed in e-commerce. Our experiments showed the effectiveness of the proposed ranking mechanism. The Spearman's rank coefficient showed a positive correlation between the experts' opinions and our results. Thus, our method can allow users to browse reviews with ease and confidence. In addition, the results of this paper can be used in industry and academia to calculate the impact of opinions on the online community and to help to quickly detect fraudulent, biased, collusive, or exaggerated opinions.

Author Contributions: Conceptualization, R.-J.Y. and Y.-C.H.; methodology, R.-J.Y. and Y.-C.H.; software, R.-J.Y. and Y.-C.H.; validation, L.-C.L. and Y.-C.H.; formal analysis, R.-J.Y. and Y.-C.H.; investigation, Y.-C.H.; writing, Y.-C.H. and R.-J.Y.; supervision, L.-C.L.; project administration, L.-C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Derived data supporting the findings of this study are available from the corresponding author on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Saxena, I. Comprehensive Consumer perspective of E-Commerce models. *Res. Inspir.* **2019**, *5*, 8–17.
2. Megdadi, Y.; Alghizzawi, M.; Hammouri, M.; Megdadi, Z.; Haddad, R.; Ibrahim, E. The Impact of Electronic Sales Channels on Customers Response of Convenience Products Outlets Stores. *Int. J. Prof. Bus. Rev.* **2023**, *8*, e01379. [\[CrossRef\]](#)
3. Jøsang, A.; Ismail, R.; Boyd, C. A survey of trust and reputation systems for online service provision. *Decis. Support Syst.* **2007**, *43*, 618–644. [\[CrossRef\]](#)
4. Bobadilla, J.; Ortega, F.; Hernando, A.; Gutiérrez, A. Recommender systems survey. *Knowl. Based Syst.* **2013**, *46*, 109–132. [\[CrossRef\]](#)
5. Wan, Y.; Ma, B.; Pan, Y. Opinion evolution of online consumer reviews in the e-commerce environment. *Electron. Commer. Res.* **2017**, *18*, 291–311. [\[CrossRef\]](#)
6. Mabrouk, A.; Redondo, R.P.D.; Kayed, M. SEOpinion: Summarization and Exploration of Opinion from E-Commerce Websites. *Sensors* **2021**, *21*, 636. [\[CrossRef\]](#)

7. Dawn, S.; Das, M.; Bandyopadhyay, S. Singer: A recommendation system based on social-influence-aware graph embedding approach. In Proceedings of the 2021 IEEE 18th India Council International Conference (INDICON), Guwahati, India, 19–21 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.
8. Miao, Q.; Li, Q.; Zeng, D. Fine-grained opinion mining by integrating multiple review sources. *J. Am. Soc. Inf. Sci. Technol.* **2010**, *61*, 2288–2299. [[CrossRef](#)]
9. Mayzlin, D.; Dover, Y.; Chevalier, J. Promotional reviews: An empirical investigation of online review manipulation. *Am. Econ. Rev.* **2014**, *104*, 2421–2455. [[CrossRef](#)]
10. Gupta, V.; Aggarwal, A.; Chakraborty, T. Detecting and Characterizing Extremist Reviewer Groups in Online Product Reviews. *IEEE Trans. Comput. Soc. Syst.* **2020**, *7*, 741–750. [[CrossRef](#)]
11. Pereira, R.H.; Gonçalves, M.J.; Magalhães, M.A.G. Reputation Systems: A framework for attacks and frauds classification. *J. Inf. Syst. Eng. Manag.* **2023**, *8*, 19218. [[CrossRef](#)]
12. Moniz, N.; Torgo, L. A review on web content popularity prediction: Issues and open challenges. *Online Soc. Netw. Media* **2019**, *12*, 1–20. [[CrossRef](#)]
13. Zelenka, J.; Azubuike, T.; Pásková, M. Trust model for online reviews of tourism services and evaluation of destinations. *Adm. Sci.* **2021**, *11*, 34. [[CrossRef](#)]
14. Ghiassi, M.; Lee, S.; Gaikwad, S.R. Sentiment analysis and spam filtering using the YAC2 clustering algorithm with transferability. *Comput. Ind. Eng.* **2022**, *165*, 107959. [[CrossRef](#)]
15. Zhao, H.; Liu, Z.; Yao, X.; Yang, Q. A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. *Inf. Process. Manag.* **2021**, *58*, 102656. [[CrossRef](#)]
16. Sivaramakrishnan, N.; Subramaniaswamy, V.; Vilorio, A.; Vijayakumar, V.; Senthilselvan, N. A deep learning-based hybrid model for recommendation generation and ranking. *Neural Comput. Appl.* **2021**, *33*, 10719–10736. [[CrossRef](#)]
17. Robbins, B.G. What is Trust? A Multidisciplinary Review, Critique, and Synthesis. *Sociol. Compass* **2016**, *10*, 972–986. [[CrossRef](#)]
18. Strömbäck, J.; Tsfati, Y.; Boomgaarden, H.; Damstra, A.; Lindgren, E.; Vliegenthart, R.; Lindholm, T. News media trust and its impact on media use: Toward a framework for future research. *Ann. Int. Commun. Assoc.* **2020**, *44*, 139–156. [[CrossRef](#)]
19. Wu, F.; Li, H.-H.; Kuo, Y.-H. Reputation evaluation for choosing a trustworthy counterparty in C2C e-commerce. *Electron. Commer. Res. Appl.* **2011**, *10*, 428–436. [[CrossRef](#)]
20. Zhang, N.; Yu, P.; Li, Y.; Gao, W. Research on the Evolution of Consumers' Purchase Intention Based on Online Reviews and Opinion Dynamics. *Sustainability* **2022**, *14*, 16510. [[CrossRef](#)]
21. Wang, G.; Xie, S.; Liu, B.; Yu, P.S. Identify Online Store Review Spammers via Social Review Graph. *ACM Trans. Intell. Syst. Technol.* **2012**, *3*, 1–21. [[CrossRef](#)]
22. Ali, A.H.; Kumar, H.; Soh, P.J. Big Data Sentiment Analysis of Twitter Data. *Mesopotamian J. Big Data* **2021**, *2021*, 1–5. [[CrossRef](#)]
23. Hsu, C.-C.; Wu, F. Topic-specific crawling on the Web with the measurements of the relevancy context graph. *Inf. Syst.* **2006**, *31*, 232–246. [[CrossRef](#)]
24. Passyn, K.A.; Diriker, M.; Settle, R.B. Price comparison, price competition, and the effects of shopbots. *J. Bus. Econ. Res. (JBER)* **2013**, *11*, 401–416. [[CrossRef](#)]
25. Azeem, M.; Abualsoud, B.M.; Priyadarshana, D. Mobile Big Data Analytics Using Deep Learning and Apache Spark. *Mesopotamian J. Big Data* **2023**, *2023*, 16–28. [[CrossRef](#)]
26. Jindal, N.; Liu, B. Opinion spam and analysis. In Proceedings of the 2008 International Conference on Web Search and Data Mining, Palo Alto, CA, USA, 11–12 February 2008; pp. 219–230.
27. Hsu, C.F.; Khabiri, E.; Caverlee, J. Ranking comments on the social web. In Proceedings of the 2009 International Conference on Computational Science and Engineering, Vancouver, BC, Canada, 29–31 August 2009; IEEE: Piscataway, NJ, USA, 2009; Volume 4, pp. 90–97.
28. Benevenuto, F.; Magno, G.; Rodrigues, T.; Almeida, V. Detecting spammers on twitter. In Proceedings of the Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS), Redmond, WA, USA, 13–14 July 2010; Volume 6, p. 12.
29. Yoo, K.H.; Gretzel, U. Comparison of deceptive and truthful travel reviews. In *Information and Communication Technologies in Tourism 2009*; Springer: Vienna, Austria, 2009; pp. 37–47.
30. Eler, D.M.; Grosa, D.; Pola, I.; Garcia, R.; Correia, J. Analysis of Document Pre-Processing Effects in Text and Opinion Mining. *Information* **2018**, *9*, 100. [[CrossRef](#)]
31. Canco, I.; Kruja, D.; Iancu, T. AHP, a reliable method for quality decision making: A case study in business. *Sustainability* **2021**, *13*, 13932. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.