Extracting Consumer Opinion on Indonesian E-Commerce: A Rating Evaluation and Lexicon-Based Sentiment Analysis

A Setiyawan¹, A W Wijayanto², H Youshi¹

¹ Jiangsu University, No. 301 Xuefu Road, Zhenjiang, People’s Republic of China
² Politeknik Statistika STIS, Jl. Otto Iskandardinata No. 64C, Jakarta, Indonesia

*Corresponding author’s e-mail: 5103200250@stmail.ujs.edu.cn

Abstract. E-commerce as a business platform offers abundant advantages in modern life all over the world. Sellers and buyers at online marketplaces may get benefits and advantages from e-commerce. One of the advantages is that e-commerce can be accessed anywhere and anytime. Despite providing advantages, e-commerce also has disadvantages including product quality fraud and data theft. Online marketplaces provide facilities for consumer evaluation, through star rating and consumer reviews. In this paper, we focus on the Business-to-Consumer (B2C) e-commerce type and extract consumer opinion data from a leading online marketplace in Indonesia and use text mining approaches to compare the rating evaluation and sentiment analysis on consumer reviews. With 2,937 records, we investigate the relationship between star rating and lexicon-based sentiment analysis. From the results, we found that most consumers do not hesitantly provide a good evaluation indicated by a 5-star rating and positive sentiment of reviews. A quite polarized rating distribution is found and indicates a straightforward consumer opinion. However, a further examination of the relation between rating and review, we discover inconsistencies in consumer opinion where the good rating may also contain negative reviews. Our result findings provide an insight to build a more integrated consumer opinion indicator in e-commerce and that online marketplace sellers need to look deeper at the detailed reviews rating.

1. Introduction
The advancement of current technology in the digital era has been improving modern human daily life. One of the prominent conveniences in the digital era is the emergence of electronic commerce (e-commerce). E-commerce is a business platform that connects sellers/providers and buyers/consumers to perform market activities over the internet [1]. E-commerce provides a very broad market space, in contrast to conventional markets which are limited to a location in a place, a city, or a country [2]. E-commerce even reaches sellers and buyers across continents [3]–[7].

E-commerce provides numerous advantages for sellers and buyers. E-commerce can be accessed anywhere and anytime [1]. Not only unlimited by time, but e-commerce also does not require buyers to have warehouses and stores. On the other hand, e-commerce also provides product variety. For buyers, product variety is an advantage because it provides product choices to be purchased. Fast delivery is also an e-commerce advantage. Another advantage is competitive prices, ease of payment, and ease of communication [3]–[5], [8].

Besides providing benefits and advantages, e-commerce also has disadvantages and potential risks. The disadvantages include products not in accordance with the order, product damage or poor quality, longtime delivery, fraud in e-commerce, and privacy issues such as potential data theft [2], [8]–[10].
For evaluation and to determine consumer satisfaction, e-commerce provides facilities for consumers to give feedback or assessment. Consumers can give a rating and write a review. Rating is an assessment of e-commerce which is usually on a scale of 1 to 5 or sometimes 1 to 10, or it can also use certain symbols such as stars or hearts or other symbols. Review is a summary, a consumer evaluation review of e-commerce which is usually delivered in text. Through a review, it will be known in more detail about advantages and disadvantages, quality of e-commerce, stores or products, shipping quality, and another aspect. Consumer evaluation and feedback help and improve e-commerce services.

To quickly find out e-commerce evaluation, it is usually enough to consider only the rating. In 1-star to 5-stars scale rating, 1-star indicates bad rating, 5-stars indicates good rating. To reveal a more detailed evaluation, it is necessary to read the review. Reviews provide more assessment of what aspects are good and what aspects are bad. Reviews also contain sentiment, which can be a positive, negative, or normal sentiment. Rating and review influence consumers for purchasing e-commerce products.

The research question in this paper is to find out whether a good rating always has a positive review and a bad rating always has a negative review. To do this, in this paper, we use text mining approaches to study consumer opinion behavior on evaluating e-commerce through ratings and detailed reviews measured as positive or negative sentiments. In the age of data science, the researcher can explore the data both quantitatively and qualitatively. Here, text mining can be used as one of the statistical tools to address qualitative research and knowledge management. To begin with, text mining of the zakat administration in times of COVID-19. However, used text mining as the perfect preference in job finding. Furthermore, we examine the relationship between consumer ratings and reviews, whether a good rating also has a positive sentiment review, vice versa. We will perform analysis on consumer ratings and perform sentiment analysis on consumer reviews.

2. Theoretical Background

2.1. E-Commerce
Many researchers have discussed and explored e-commerce. About the definition of e-commerce, e-commerce is a business platform that allows seller or service provider and buyer or consumer to meet for marketing/business activities over the internet. The development of computer technology and network infrastructure has triggered the emergence of e-commerce today.

Activity in e-commerce includes various types of marketing/business activities, such as shopping, banking, transportation, services, and investment. Amazon.com and eBay became the first and most popular e-commerce in the world. Then followed by the emergence of e-commerce companies in other countries such as Alibaba.com in China. In Indonesia, e-commerce has also emerged, such as Tokopedia, Bukalapak, Lazada, Go-Jek, Shopee and the others.

There are at least 5 types of e-commerce include Business-to-Business (B2B), Business to Consumer (B2C), Business-to-Government (B2G), Consumer-to-Consumer (C2C), and Mobile commerce (M-Commerce). E-commerce is easily accessible anytime and anywhere, not limited to time and region. E-commerce can also be accessed through various devices, such as desktop computers, laptops, tablet PCs, and mobile phones.

2.2. Reviews on E-Commerce
E-commerce evaluation is necessary for consumer satisfaction and improving e-commerce services. Evaluation and feedback from consumers are usually through ratings and reviews. Rating and reviews are needed to provide a consumer experience with a product. The consumer when purchasing a product in e-commerce could not physically experience the product. In this exploring and mitigating fraud rating, this research about product rating and reviews (sentiment). Many scholars also explore e-commerce reviews, the sentiment of reviews, and customer purchasing behavior.
2.3. Lexicon-Based Sentiment Analysis
Lexicon-based sentiment analysis (LSA) is one of the major approaches in sentiment analysis. LSA works by using the word semantic orientation in a given text to calculate the sentiment [22]. LSA requires a dictionary consisting of positive or negative sentiment labels assigned to each of the words. To compose the dictionary, different existing ways have been introduced, not only manually but also automatically. Given an input text, in LSA, the text will be converted as a bag of words (BoW) and all of its positive and negative words are labeled with sentiment values based on the pre-defined dictionary [24]. To make an overall sentiment value of the input text, we apply sum or average to the input text as a combining function. In this study, we utilize the lexicon-based approach which helps us to generate a labeled training set.

3. Methodology

3.1. Dataset
The data used in this article is collected reviews data from e-commerce in Indonesia. We use data from Shopee (2,937 records/instances), one of the top e-commerce in Indonesia. The original extracted review is in the Indonesian language (Bahasa Indonesia).

3.2. Pre-processing
There is an initial process that needs to be done when performing text analysis in machine learning. This process is called text pre-processing. Data about review from web crawling needs to be cleaned up first. Because in the text review there is a noise that will interfere with the analysis process. There are common methods that can be used in text pre-processing, including text transformation, tokenization, and filtering [1], [14], [25].

a. Text Transformation
Transformation is the process to transforms text from original text to modified text. The most common and most frequently performed process is to apply the text to lowercase transformation. We need to transform text to lowercase because the computer will identify a word as a different word if written in different cases. For example, the word "Car" and "car" will be interpreted differently by computer. This will make duplication of words in the dictionary we created later [14], [25].

Another transformation method is removing accents, parsing HTML, and removing web URLs. Remove accents will remove all diacritics or accents in the original text. Parse HTML will detect HTML tags on the original text and parse out text only. Remove URLs will remove URLs from the original text. In this article we perform 3 transformations, lowercase transformation, parse HTML and remove URL. Remove accents are not performed because in Bahasa Indonesia there is no accent word [14], [25].

Text transformation example:
Lowercase transformation: Original TEXT will transform to the original text.
Remove accents: “naïve” will be transformed to “naive”.
Parse HTML: <strong>Original text</strong> will transform to original text.
Remove URL: This is a https://www.example.com/ URL will transform to This is a URL.

b. Tokenization
Tokenization is a method for separating a text into smaller text units. The smaller text units are called tokens. The token can be a word, a character, or a sub-word. Usually, the tokenization process is classified into 3 types: words, characters, and sub-words (n-gram characters) tokenization. The goal of tokenization is to create a vocabulary of the text [14]. In this article, we perform tokenization by word only.

c. Filtering
Filtering is a process to remove or keep a selection of words. In this article, we remove stop words in Bahasa Indonesia. Stop words are common words that usually appear in large numbers and are
considered meaningless [14], [25]. Examples of stop words for English include "of", "the", and "on". Meanwhile, Bahasa Indonesia includes "yang", "di", and "ke". For this goal, we need a stop words list in Bahasa Indonesia. In this article, we use a list of stop words in Bahasa Indonesia from the list by [7].

3.3. Lexicon-Based Sentiment Analysis

Sentiment analysis in this article uses lexicon-based sentiment analysis with a positive and negative dictionary in Bahasa Indonesia from the word list created by [6]. We calculate the sentiment score for reviews based on the dictionary and add a tag or label to the text with (-1) for negative sentiment, add a tag to the text with (1) for positive sentiment and add tag (0) for the normal sentiment.

![Figure 1. Reviews Analysis Step](image1)

4. Results and Discussion

4.1. Pre-processing Results

After performing text pre-processing, we will have a dictionary/word list in lowercase that is clean of stop words, the meaningless common words. In this dictionary we also calculate the frequency of each word. Figure 2 shows us the different list of the words before pre-processing and after pre-processing. Before pre-processing is performed, there are many meaningless words and punctuation in the data/corpus.

<table>
<thead>
<tr>
<th>Weight</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>3225</td>
<td>,</td>
</tr>
<tr>
<td>2443</td>
<td>.</td>
</tr>
<tr>
<td>2226</td>
<td>shopee</td>
</tr>
<tr>
<td>1795</td>
<td>di</td>
</tr>
<tr>
<td>1505</td>
<td>saya</td>
</tr>
<tr>
<td>1113</td>
<td>dan</td>
</tr>
<tr>
<td>999</td>
<td>bisa</td>
</tr>
<tr>
<td>942</td>
<td>nya</td>
</tr>
<tr>
<td>930</td>
<td>belanja</td>
</tr>
<tr>
<td>884</td>
<td>..</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Weight</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>2226</td>
<td>shopee</td>
</tr>
<tr>
<td>930</td>
<td>belanja</td>
</tr>
<tr>
<td>644</td>
<td>barang</td>
</tr>
<tr>
<td>629</td>
<td>aplikasi</td>
</tr>
<tr>
<td>613</td>
<td>ongkir</td>
</tr>
<tr>
<td>472</td>
<td>gratis</td>
</tr>
<tr>
<td>439</td>
<td>bagus</td>
</tr>
<tr>
<td>436</td>
<td>banget</td>
</tr>
<tr>
<td>352</td>
<td>gratis ongkir</td>
</tr>
<tr>
<td>304</td>
<td>suka</td>
</tr>
</tbody>
</table>

![Figure 2. Top 10 words before and after pre-processing](image2)

One of the interesting outputs in the text analysis is the word cloud. Word cloud is a method for displaying text data visually. Word cloud is popular in text analysis/text mining because it is easy to understand. Word cloud has a function to bring up a visual image of text. The use of word clouds will
help us to get a complete understanding of an idea from a collection of texts. The word cloud will visualize the words based on the number of times they appear in the dictionary. The more frequency of a word, the larger size of the word is visualized. Word cloud in Figure 3 shows us words with large size are “belanja” (shopping), “barang” (things/product), “shopee” (name of e-commerce), “bagus” (good), “ongkir” (shipping charges), “suka”, “aplikasi”, “gratis”, some of these words represent positive reviews. We also saw negative words like "kecewa" (disappointed) and “tolong” (help) printed in a smaller size.

Figure 3. Word cloud before and after text pre-processing

4.2. Rating Distribution

The distribution of ratings could be seen in Figure 4. From Figure 4, consumer rating with 5-star reaches 56.79% (1,668 instances), 4-star reach 7.73% (227 instances), there are no instances with 3-star, rating with 2-star reach 7.9% (232 instances) and the rating with 1-star is 27.58% (810 instances).

Figure 4. Rating Distribution

Figure 4 shows us that consumers tend to give a 1-star rating and 5-star rating. Most of the consumers give 5-stars or a good rating (56.79%). The second is a 1-star or bad rating (27.58%). Statistics for the rating are mean (µ) = 3.58 and standard deviation (σ) = 1.78. Based on this rating evaluation, we can conclude that e-commerce is good.
4.3. Sentiment Analysis
From total data of 2,937 instances, there are 30.61% (899 instances) negative reviews, 23.12% (679 instances) were normal reviews and 46.27% (1,359 instances) were positive reviews. This can be seen in Figure 6. If we look at the distribution of sentiment scores in Figure 5, we find mean ($\mu$) = 2.11 and standard deviation 8.09. From this, we know that data tends to be positive reviews.

Figure 5. Sentiment Score Distribution

Figure 6. Sentiment Label Distribution

Therefore, from sentiment analysis, we can conclude that consumers tend to give positive reviews for e-commerce. Figure 7 and Figure 8 show the top 5 negative sentiment scores (maximum score = -50) and top 5 positive sentiment scores (maximum score = 40). Some of the negative keywords from Figure 7 are “lemot” (slowly), “gagal” (failed), “buruk” (bad), “jek” (bad), “kecewa” (disappointed), and “lambat” (slowly). Then, some of the positive keywords from Figure 8 are “bagus banget” (very

<table>
<thead>
<tr>
<th>Ulasan</th>
<th>Rating</th>
<th>Sent_Score</th>
<th>Sentimen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Makin lemot ...</td>
<td>1.0</td>
<td>-60</td>
<td>-1.0</td>
</tr>
<tr>
<td>Sudah beberapa kali cekout,gagal malah kembali ke beranda shopee sekarang makin ...</td>
<td>4.0</td>
<td>-31.25</td>
<td>-1.0</td>
</tr>
<tr>
<td>Aplikasi buruk banget...</td>
<td>5.0</td>
<td>-28.5714</td>
<td>-1.0</td>
</tr>
<tr>
<td>Aplikasinya jelek...</td>
<td>5.0</td>
<td>-28.5714</td>
<td>-1.0</td>
</tr>
<tr>
<td>Diupdate ko malah makin jelek,load konten/image lama dan tidak muncul,sangat kecewa.</td>
<td>5.0</td>
<td>-27.7778</td>
<td>-1.0</td>
</tr>
<tr>
<td>Mau belanja jadi terhambat karna lambat respon klik barang nya,(loding apps)parah...</td>
<td>1.0</td>
<td>-25</td>
<td>-1.0</td>
</tr>
</tbody>
</table>

**Figure 7.** Top 5 Negative Sentiment Score

<table>
<thead>
<tr>
<th>Ulasan</th>
<th>Rating</th>
<th>Sent_Score</th>
<th>Sentimen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wah aplikasinya bagus banget!!!!!!!!!!!!!!!</td>
<td>5.0</td>
<td>40</td>
<td>1.0</td>
</tr>
<tr>
<td>Mudah diakses cepat dan baik pelayanannya banyak kemudahannya ,saya suka banget.</td>
<td>5.0</td>
<td>38.4615</td>
<td>1.0</td>
</tr>
<tr>
<td>pokonya shope hebat,orang2nya pada amanah,balıktenutama sangat mengutamakan ...</td>
<td>5.0</td>
<td>33.3333</td>
<td>1.0</td>
</tr>
<tr>
<td>Kualitas Produk bagus,pengiriman cepat,responsif...dan banyak sekali keuntungan lainnya.</td>
<td>5.0</td>
<td>33.3333</td>
<td>1.0</td>
</tr>
<tr>
<td>Sangat Membantu dan Bermanfaat,Terima kasih banyak Shopee,Terbaik Sepanjang masa Go...</td>
<td>5.0</td>
<td>31.25</td>
<td>1.0</td>
</tr>
<tr>
<td>Barang sesuai deskripsi,respon penjual cepat dan ramah,package rapi aman,recomend ...</td>
<td>5.0</td>
<td>31.25</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**Figure 8.** Top 5 Positive Sentiment Score

4.4. Ratings and Review Relation

In the previous section, results of the e-commerce evaluation have been discussed for rating and review separately. Both rating and review show a good e-commerce evaluation from consumers. In this section, we will evaluate the rating and review simultaneously. When we examine more deeply, by looking at the reviews provided by consumers, it is found that at the 4-star and 5-stars ratings (good ratings) there are many negative reviews from consumers, see Region A in Figure 10. This case is shown in Figure 9 and Figure 10. Figure 9 is a scatter plot between sentiment score and rating from consumers. Figure 10 is a scatter plot between sentiment (label) and rating from consumers. From these results, we know that consumers still give a good rating on e-commerce even though the consumer experiences with e-commerce were disappointing (which can be seen from the review). We can also see that in a bad rating, there is a positive review. This part is the main article contribution, about rating evaluation paired with review sentiment.

**Figure 9.** Rating and Sentiment Score Scatter Plot
Region A in Figure 10 is instances with ratings more than 4 (4-star and 5-star rating) and have negative sentiment (-1). Number instance for region A is 392 instances from 2,937 instances (13.34%). Region B in Figure 10 is instances with bad ratings (1-star and 2-star) and have positive sentiment (1). Figure 11 shows the top 5 negative sentiment scores but has a good rating (4-star and 5-star).

Figure 11. Top negative sentiment score and has a good rating (rating 4 and rating 5)

4.5. Discussion
Consumer behavior on a business platform may be different from one consumer group to another, in certain regions, or other aspects. In this article we work with e-commerce data from Indonesia, we can see the behavior or trends of consumers in giving ratings. Based on the data and our investigation results, consumers in Indonesia prefer to choose 1-star and 5-star rather than other ratings. In theory, a rating with 1-star indicates a bad rating, while 5-stars indicates a good rating. Our finding shows that consumer behavior still gives a good rating, even though the consumer is experiencing disappointment. This could be due to the culture of the Indonesians who are full of politeness even though there are inconveniences in the reviews but the rating can still be good. This behavior needs to be studied further.

Another topic that can be learned more is about what aspect has the most negative review and what aspect has the most positive review. We will know which part of e-commerce is the best and which part has received criticism from consumers to make improvements. For this purpose, we can perform “topic modeling”.

5. Conclusion
E-commerce opinion evaluation may refer to both of the ratings given by the consumer as well as the consumer reviews. In this study, we extract consumer opinion data from a leading online marketplace in Indonesia and use text mining approaches to compare the rating evaluation and sentiment analysis on consumer reviews. Based on the rating evaluation results, there is a polarized rating distribution where the majority of consumers tend to give an extreme 1-star or 5-star rating rather than the more moderate ones such as 2-star, 3-star, and 4-star ratings. We also found that most consumers are not hesitantly expressing a good evaluation indicated by a 5-star rating and positive sentiment of reviews.
However, in our further assessment of the relationship between rating and reviews, we find irregularities in the opinion distribution where the good rating may likewise contain negative comments. It can be seen that even though the consumer is not satisfied with e-commerce, the consumer still gives a good rating. In other words, from the data, a 4-star and 5-star rating are not always reflecting good opinions, there are many negative reviews from consumers. Finally, to get a rich e-commerce evaluation result we need to look deeper at the reviews provided in e-commerce, it is not enough to just look at the rating.

References


