Exploring Sentiment Analysis on E-Commerce Business: Lazada and Shopee

Jenny Yow Bee Yin [1], Nor Hasliza Md Saad [1], Zulnaidi Yaacob [2]

[1] *School of Management, Universiti Sains Malaysia, Penang, Malaysia*
[2] *School of Distance Education, Universiti Sains Malaysia, Penang, Malaysia*

Abstract **– This study examined Twitter users' attitudes regarding e-commerce businesses, Lazada and Shopee. The primary goal of this research is to determine the sentiment of Twitter users toward the e-commerce company and the topic debate among Twitter users. The results demonstrate that both companies are receiving more positive than negative sentiment, indicating that Twitter users share positive opinions about Lazada and Shopee on social media. RapidMiner was chosen as a tool to extract data from Twitter through the Twitter API, data cleaning, and perform sentiment analysis. The sentiment is classified into three sentiment classes; Positive, Negative, or Neutral. These results will help Shopee and Lazada determine how the audience perceives the experience and recommendation. The results contribute to our understanding of e-commerce business services. It provides an overview of the public's viewpoint so that e-commerce businesses may comprehend their services' strengths and weaknesses and enhance marketing strategy.**

Keywords **– sentiment analysis, Twitter, Lazada, Shopee, e-commerce.**

DOI: 10.18421/TEM114-11
https://doi.org/10.18421/TEM114-11

Corresponding author: Nor Hasliza Md Saad,
School of Management, Universiti Sains Malaysia, Penang, Malaysia.
Email: norhasliza@usm.my

Received: 03 July 2022.
Revised: 05 September 2022.
Accepted: 20 September 2022.
Published: 25 November 2022.

1. Introduction

E-commerce is a digital marketplace and a mean of doing business that enables customers and sellers to transact business via the internet. Lazada and Shopee are examples of e-commerce platforms that lay a strong emphasis on providing users with an effective online marketplace for business-to-customer transactions (B2C). The B2C business model focuses the business activitivies and transactions with final consumers that involves the sale of both goods and services. The term "e-marketplace" refers to an online venue where various buyers and sellers may communicate with one another and conduct commercial transactions digitally. In general, there are two primary categories of e-marketplaces. Lazada and Shopee are two examples of e-marketplaces that fall into the horizontal category. This sort of e-marketplace sells items or services from a broad variety of categories and functions as a one-stop platform for several types of vendors and buyers. According to the e-commerce aggregator iPrice Group's 2021 report [1] on the e-commerce business performance for the Southeast Asian region, Shopee and Lazada are the leading competitors in the majority of the markets throughout the region. Shopee dominates Malaysia's e-commerce online traffic with 71%, followed by Lazada with 18% and PGMall with 9%.

As more new comers enter the e-commerce market, the rivalry between competitors heats up. As a result, every e-commerce company begins to build their brand on social media platforms such as Twitter, Facebook, and Instagram to communicate with customers and understand market needs. Those comments and postings are valuable to e-commerce enterprises since they allow them to understand what people think about their firm and enhance their service to keep a competitiveadvantage.

Lazada started operating in 2013, employing a marketplace place model that allowed third-party retailers to sell their products through the Lazadasite. In 2016, Alibaba Group brought a controlling stake in Lazada to support the Alibaba International expansion plans. Lazada serves most of Southeast

Asia, including Malaysia, Thailand, Indonesia, Singapore, Philippines, and Vietnam. Shopee started in 2015, and they have a hybrid business model including a consumer-to-consumer (C2C) marketplace and business-to-consumer (B2C). Besides serving the Southeast Asia country, Shopee has expanded its business to several countries in Latin America.

Both companies are active on the social media platform as one of the interaction channels to interact with customers, potential customers, and followers, promote their brand and establish customer service. Through social media, both companies can understand the most essential and most attractive marketing campaign the users would like to participate in, which can increase the company's market share. Lazada joined the Twitter platform in February 2012. As of May 2021, Lazada Malaysia has a total of 42.7K followers on Twitter. While Shopee Malaysia started in January 2015 and currently has 34.9K followers on Twitter [2].

1.1. Research Problem

Social media such as Facebook, Twitter, and Instagram have become effective communication channels across society. They feature a large amount of user-generated information, such as customer comments and feelings. Social media is quickly becoming crucial online information providers. People build relationships on social media by expressing their emotions, complaints, and opinions on topics ranging from current political concerns to the latest blockbusters. Building an online reputation through social media is valuable for most businesses, including e-commerce. As a result, many companies have used social media to connect with customers, and social media analytics are extensively utilised for research and business objectives [3,4].

The power of electronic word-of-mouth (e-WOM) can ruin a company's image and its profit in a short time. e-WOM communication refers to any positive or negative statement made by a customer about a product or company, which is made available to many people and institutions via the Internet [5,6,7]. The impact of negative eWOM is enormous, indicating that users are dissatisfied with the business's service and product [8, 9]. If Lazada and Shopee do not respond quickly to the bad statement, the company's image willsuffer, and other buyers will reconsider their purchases.

Business decision-makers are interested not only in what consumers think about their brand but also in what consumers think about the products and services offered by their competitors [10]. Businesses want to know how customers regard their product or service and whether they have a good market reputation. They also seek to understand their competitive advantage concerning competitors [11]. Thus, sentiment analysis is critical for Lazada and Shopee so that the companies can better understand their customers and create marketing plans and tactics for the future to win more customers and boost customer happiness. The main research objectives are: to figure out the sentiment of Twitter users towards Lazada and Shopee; and to explore the topic discussion of Twitter users about Lazada and Shopee. Few studies have addressed the elements that cause distinct forms of sentiment [12,13]. Clarifying the elements that determine certain forms of sentiment will assist manage the business [14].

2. Sentiment Analysis

Sentiment Analysis (SA) is NLP (NLP) that assesses and recognises polarity/sentiment in writing, such as social media postings or product descriptions [15]. Sentiment analysis helps businesses monitor online conversations about their brand, product, or service [16]. In addition, sentiment analysis examines people's views, attitudes, feelings, assessments, and evaluations of a service, product, organisation, individual, issue, contest, and attributes [17,18].

Because it analyses data based on words or text, sentiment analysis is helpful for any type of brief communication from any medium. Sentiment analysis can be used as a brand monitoring technique by an organisation from time to time to gauge how people think about their brand and how satisfied consumers are with the brand. For example, in the presence of a marketing activities, sentiment analysis can assist the organisation in determining whether or not the activities was successful by understanding what people think and say about the activities. The sentiment result might be used to improve the marketing strategy.

In general, sentiment analysis is the text or word classified into three groups: positive, negative or neutral, according to their criteria. The lexicon-based approach is used in this research. It has a pre-made sentiment lexicon that adds up the sentiment scores of all the words in the document to give the document a score. A score can be a simple polarity value like +1, -1, or 0 for positive, negative, or neutral words, or it can be a value that shows how strong or intense the sentiment is. At the sentence and feature level, lexicon-based sentiment analysis is advantageous since it does not need training data and considered an unsupervised method.

Twitter is a microblogging system that allows users to send and receive short posts called tweets. In 2018, Twitter set a new limit for each post where it can up to 280 characters compared to the 140 characters. As of 2020, Twitter users reached 340M, and the total number of daily tweets was 500M. A large number of tweets per day can bring meaningful insights to a company and the government. This microblogging platform has been used as a form of public opinion. According to [19], Twitter users are 38% more likely to post opinions about brands and products than other social media users. Thus, the Twitter sentiment analysis focus on classifying the individual tweets based on polarity, which means tweets may be classified into positive, negative or neutral.

3. Social Media Involvement in Customer Buying Process

During this stage, consumers will assess all of their product and brand options based on a scale of qualities that may give the desired benefit. In the context of social media, the most important aspects of the decision-making process will be the comments and postings. since many individuals base their purchasing decisions on social media recommendations [20,21]. In this phase, consumers will read all comments and postings linked to the product or brand. They will browse publicly accessible information on the Internet or online reviews to get insight into the attitudes and experiences of other customers about the items or services [22]. The positive and negative sentiments expressed by users will impact the following phase of a consumer's decision-making process [23]. Classical consumer purchasing behaviour models are pillars of consumer behaviour research, which has led to consumer decision path analysis**,** there are five phases purchase decision procedures [24,25,26].

a) Awareness

Buyer decision-making starts with need awareness. A desire may be triggered by both internal and external stimuli. Social media postings and adverts represent external stimuli [20,21]. Those who use social media and see advertisements on social media that easily arouse their desire. The customer will identify a certain brand or product that fulfils their requirements.

b) Search Information

During this stage, a customer who has identified a certain need may seek information from internet sources, such as social media platforms. Consumers increasingly rely on internet sources and e-WOM when researching items or services they want to purchase or utilize [27]. The customers will browse several websites and social media to study the user-generated information and comments about a certain product or brand on social media. Many studies argue that when potential customers attempt to obtain information before making a purchase decision [28].

c) Evaluting Alternative

During this stage, consumer will evaluate all of their product and brand option on a scale of attributes which have the ability to deliver benefit that the customers seeking. In the context of social media, the comments and posts are the attributes that will be most influential to the decision-making process [29]. Consumers will read all the comment and posts related to the product orbrand in this stage. The positive and negative sentiment that express by the users will influence a consumer on the next step of decision process [30].

d) Purchase Decision

During this phase, the customer may establish an intention to acquire the product since he or she has analysed and determined its worth. The negative feedback on the social media about the brand or product will influence the personal preference and they might or might not purchase the item [9,11].

e) Post Purchase Behavior

Post-purchase behaviour involves customer satisfaction evaluation. Customer experiences affect whether they buy a product or brand again. When consumers post on social media, the experience affects the company or brand. If the total experience and after-sales service are great, consumers will share their pleasant experience on social media, and the excellent reputation will attract new customers. People post negative remarks on social media after a bad buying experience. Negative comments or complaints about a product, services or brand have a bigger impact than positive ones, and when something "goes viral," it can spread a lot faster [8,31]

4. Methodology

RapidMiner is the research tool used in this study. RapidMiner is a tool for conducting data mining workflows for various tasks, ranging from different areas of data mining applications to different parameter optimisation schemes [32, 33]. One of the main traits of RapidMiner is its advanced ability to program execution of complex workflows, all done within a visual user interface, without the need for traditional programming skills [34].

Opinions are subjective expressions of human thoughts, emotions and feelings. To conduct a sentiment analysis by using lexicon-based approach, a set of dictionaries of opinionated terms is used [35, 36]. There are several approaches for detecting the sentiment of phrases. The sentiment analysis output consists of 3 different polarities: *Positive*, *Negative* and *Neutral* .

In this research, the VADER (Valence Aware Dictionary Entiment Reasoner) model is chosen and used for the text analysis as this tool is specifically attuned to a sentiment expressed in social media [37, 38]. VADER model is a rule-based and lexicon sentiment analysis tool. It can be directly used to unlabeled text data. VADER sentimental analysis relies on a dictionary that maps lexical features to emotion intensities known as sentiment scores. The sentiment score of a text can be obtained by summing up the intensity ofeach word in the text. The support for sentence sentiment also considers booster words, for example, *very* in very satisfy and negation words (e.g. *not* in not satisfy). Every word in thetext will be valued and given a scoring string. In a tweet, users will include many elements in their sentences, such as emojis, capital words and conjunctions. VADER is excellent for social media text because it can understand well all the sentiments of a text that contains emoticons, slang, capital words, multiple punctuation marks, acronyms and much more text that can indicate the writer is happy or unhappy.

4.1. Research Design

Data collection, data preparation, sentiment classification, text preprocessing, and data visualisation are the five key procedures involved in sentiment analysis. The entire procedure will be discussed in the following section.

4.1.1. Data Collection

The data to be used in this research is secondary data obtained directly from Twitter. In RapidMiner, a Search Twitter operator can allow users to search and extract the updated tweets that are related to a specific topic or keyword from Twitter API. The operator requires a connection to a Twitter account. Two keywords related to E-commerce company, Lazada and Shopee, are selected as the query and the English language is selected.

Each search is set to the 500-1000 limit search results due to the limited capacity of the computer used for the analysis. The whole process of data collection was done from 25 to May 28 2021. From the collection, we collected 4000 tweets for Lazada and a total of 2404 tweets for Shopee. The data were accumulated for several days because of the Twitter API rule of not providing past data that is more than seven days.

4.1.2. Data Preprocessing

Before conducting the sentiment analysis, the data collected directly from Twitter will go through processing steps to clean data. Data Preparation is the process of collecting,cleaning, and consolidating data into one file or data table, primarily for use in the analysis. Data preparation can improve the accuracy, quality and processing speed of data will increase. The replacement of above symbols and characters are replaced by white space. Through data preprocessing steps, the number of tweets will be reduced. A summary of the tweets is shown in Table 1.

Table 1. Summary of Number of Tweets

Keywords	No. of Collected Tweets	Clean Data
Lazada	4000	927
Shopee	2404	1322

4.1.3. Sentiment Classification

A lexicon-based method is used to conduct the sentiment analysis. To analyse the people's opinion, polarity-based is chosen where the pieces of text are classified as either positive or negative. It is range between -1 to +1, where -1 is extremely negative sentiment and +1 is extremely positive sentiment.

In this research, the users' tweets are analysed using VADER, and the words aregiven a score. The scoring string is then summed up and a final score is produced. If the score is more than 0 then the specific tweet is considered positive sentiment; if the final score is less than 0 then the tweet is considered a negative sentiment. If the score is 0, the tweet will be labelled as neutral.

4.1.4. Text Preprocessing

In this step, the tweets will go through another process to extract the most frequent words thatoccur in the text for different polarities. TF-IDF (Term Frequency-Inverse Document Frequency)technique is used as the schema for creating the word vector. TF-IDF is a weight-based metric to identify reviews or documents relevant to some query terms. The underlying idea of TF-IDF is that rare terms are weighted higher than common terms. In other words, rare terms are regarded to be more important than common terms due to their discriminating nature. For example, the words that are very common on every document, such as this, that and if will rank low even they may appear many times in the tweets, since they do not mean much to the document in particular. In contrast, if the word *Buy* appears many times in a document while not appearingmany times in others, it probably means that it is very relevant.

TF-IDF consists of two parts, Term Frequency and Inverse Document Frequency. The term frequency is the number of times a term is found in the document. IDF is the logarithm of document frequency which indicates the rarity of a term. A statistical weight is used to measurethe importance of the term in the text document collection. The most common word is, the value is 0; otherwise, it will approach as 1.

The tweets that have to undergo a data cleaning process are in a phrase. Thus, to generate the word list of each polarity, the sentences need to be split into words through tokenisation. The text processing has

a) Tokenize:

Tokenisation is breaking a stream of textual content into words, terms, symbols, or some other meaningful elements called tokens [39]. The tokenisation aims to explore the words in a sentence and identify the meaningful keywords. Furthermore, the tokeniser can cater for consistency in the documents. "Non-letter" mode was selected as the parameter for splitting the text into word tokens whenever it is non-letter like a full stop, space and numerical value.

b) Transform Cases

In writing a tweet, there must be a different form of letters. Transform cases are the process of transforming all the words into a particular case, either in capital letters or lower cases. The purpose of transforming all characters into one case is for simplification in reading andanalysing the text. In this research, all the characters are transformed to lower case.

c) Filter Stopwords (English)

Stop words are a division of natural language. Articles, prepositions, and pronouns are the most common words in text documents, yet they do not convey the meaning of the documents. Stop-words are removed from a text because they make it appear heavier and less essential for the analysis. The dimensionality of term space is reduced when stop words are removed, improving the system performance. Example of the stop words is *the, in,a, an*, *with*, and so forth.

d) Stem (Porter)

Stemming is the process of removing the suffix of the words. This method is used to identify the root/stem of a word. The main purpose of stemming is to reduce the number of wordsand can have accurately matching stems. For example, the words *enjoy, enjoyed* can be stemmed to the word *"enjoi"*. In RapidMiner, a Porter stemming algorithm is used and thisalgorithm is based on the idea that the suffixes in the English language are mostly made up of a grouping of smaller and simple suffixes [39].

e) Filter Token (by Length)

Filter token is a process of removing the word shorter or longer than the configured number of characters. The filtering process can provide flexibility to the researchers in designing the data structure and mining structure so that a single mining structure can be created based on the comprehensive data source view [40]. In this particular research, the range is set to 5-25 characters in a token because the words less than five characters do not bring significant insight to this research

4.1.5. Data Visualization

Data visualisation is the practice of translating information into a visual context, such as a graph or map, to make people understand and pull insights. The main goal of data visualisation is to make it easier to identify the patterns, trends and outliers in a large dataset. Data visualisation is one of the critical steps in the analysis process, which states that after the data has been collected, processed and modelled, it must be visualised to draw a conclusion. Tableau data visualisation software is used in this research to create the graph and word cloud.

4.2. Content Analysis

Content analysis is a method for methodically analysing written, verbal, or visual evidence thatcan be used qualitatively or quantitatively [41]. After the materials have been sorted and coded, specific themes emerge. Books, manuscripts, paintings, photographs, recorded conversations, videotaped events, communications on electronic mailing lists and online forums, blog postings, and other sources can all be used to create content. The content is broken down into conceptual parts, which are subsequently coded or labelled. The categories for qualitative analysis develop as the analysis progresses. The findings are utilised to form assumptions about the text's messages.. In this study, a total of 50 tweets for each keyword and each polarity are classified into different themes based on the content of the tweets. The theme's findings can help the company determine what issues are being discussed by Twitter users.

5. Findings

5.1. Sentiment Analysis

From the data collected through Twitter, 4000 tweets related to Lazada underwent a data cleaning process, and a total of 927 clean data was used to categorise as positive, negativeand neutral sentiment. Out of 927 clean data, there are 483 tweets or 52.10% labelled as positive, 289 tweets (31.18%) labelled as neutral and only 155 tweets (16.72%) labelled as negative.

For Shopee, there is a total of 1322 clean data used to classify the tweets into different polarities. There is approximately 57.87% of the tweets labelled as positive. At the same time, 448 tweets (33.88%) are labelled as neutral. The tweets labelled as negative only have a total of109 or 8.25%.

Both companies have more positive tweets than negative tweets. The number of positive tweets is over 50% of the total clean data collected. This means that social mediausers share the positive opinion about Lazada and Shopee. The results are shown in Figure1 and a summary in Table 2.

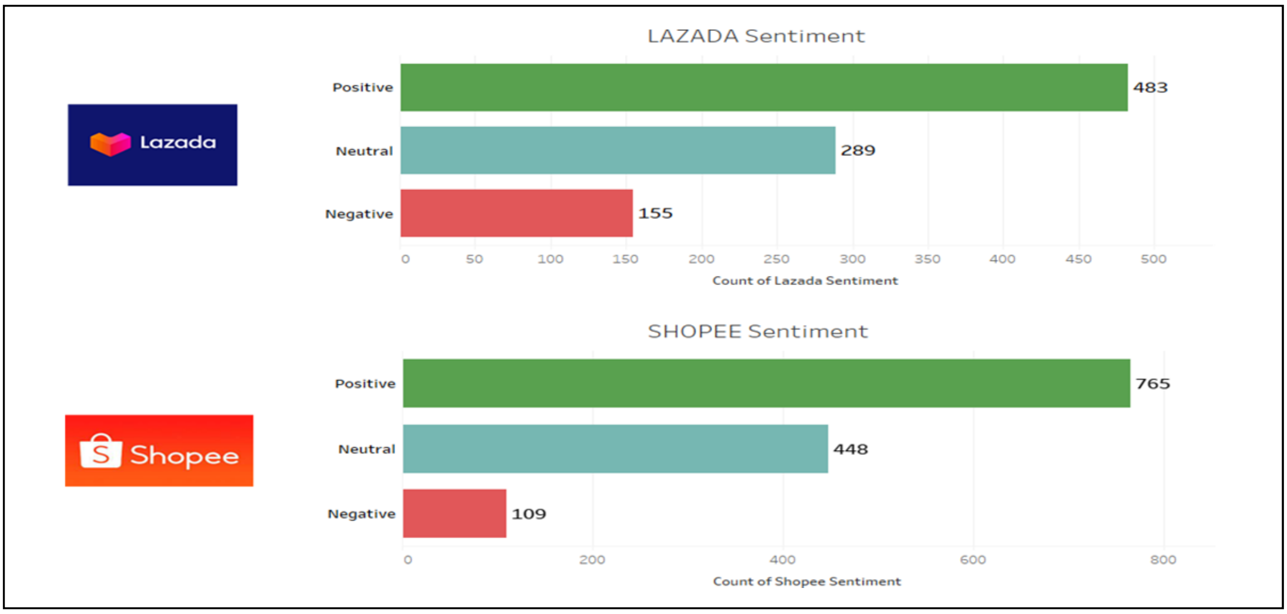

Figure 1. Sentiment Analysis Result for Lazada and Shopee

Table 2. Sentiment Analysis Result for Lazada and Shopee

Lazada		
Sentiment	Total	Percentage (%)
Positive	483	52.10
Negative	155	16.72
Neutral	289	31.18
Total	927	100

Shopee		
Sentiment	Total	Percentage (%)
Positive	765	57.87
Negative	109	8.25
Neutral	448	33.88
Total	1322	100

5.2. Content Analysis

Content analysis is a research technique that identifies the presence of specific words, topics, or concepts in qualitative data (i.e. text). In this section, the tweets are classified into different themes accordingly:

5.2.1. Lazada

For the positive sentiment, a total of 6 themes are identified. Out of the 50 tweets that have been selected, there are 23 tweets (46%) discussing about promotion. The promotion is referring to the seller promoting their store and product on social media. The next theme is the contest with atotal of 18 %. For the contest, it refers to the Give Away contest that the seller conducts. Throughpromoting and conducting the contest, users are gaining more attention from social media users and creating brand awareness among the users. There are five tweets that discussed rewards. Users are sharing the getting rewards chance on social media. Finally, thetweets about persuade and inquiry are three, respectively. From the tweets, we know that users are using the social media platform to ask questions about products and places to buy. It shows that users are searching and evaluating before making any decision. Users also use the social media platform to persuade others to join Lazada for online shopping. Seven tweets are categorised as others because the tweets do not show a specific theme. Below is the summary of the results of Lazada positive themes and sample tweets for each theme.

Table 3. Lazada: Themes in Positive Sentiment Analysis

Theme	**Total**	**Percentage (%)**
Promotion	23	46
Contest	9	18
Others	7	14
Reward	5	10
Inquiry	3	6
Persuade	3	6
Total	**50**	**100**

For negative sentiment, 84% of the tweets are emotion tweets. The users express his or her dissatisfaction toward Lazada by sharing a hashtag and boycott action. The content of the tweets is almost the same. There are also four tweets about the user's experience. The tweets are discussing their bad experience with Lazada previously. The next theme is a product with a total of 3 tweets (6%) followed by others (2%). For products, users are complaining about the defective product that they receive.

Table 4. Lazada: Themes in Negative Sentiment Analysis

Theme	Total	Percentage (%)
Emotion	42	84
User's Experience	4	8
Product	3	6
Others	1	2
Total	**50**	**100**

5.2.2. Shopee

There is a total of 765 positive sentiments for Shopee. The 50 top highest scores were selected, and each tweet is classified into different themes. There are five themes. The promotion has the highest percentage with 44%. Most of the tweets are about seller promoting their store and product. The next theme is reward with a total of 11 tweets, followed by ten tweets. Both themesrefer to the users sharing the reward available on the Shopee and Give Away contests, respectively. Users also share their experience (8%) with Shopee and feel happy about the experience they have with Shopee. The last theme is inquiry (3 out of 50 tweets), where users also use the social media platform to ask questions and collect opinions before making purchases.

Table 5. Shopee: Themes in Positive Sentiment Analysis

Theme	Total	Percentage (%)
Promotion	22	44
Reward	11	22
Contest	10	20
User's Experience	4	8
Inquiry	3	6
Total	**50**	**100**

For Shopee's negative sentiment, it is categorised as five different themes. The first theme is service.It has a total of 8 tweets sharing about the service the seller provides on the platform. Followed by user experience, with 3 tweets (17.64%). Users are sharing the bad service they encounter, especially customer service where the Shopee does not follow up their case. The next three themes are inquiry, product and price, with two tweets each.

Table 6. Shopee: Themes in Negative Sentiment Analysis

Theme	Total	Percentage (%)
Service	8	47.05
User's Experience	3	17.64
Inquiry	2	11.77
Product	2	11.77
Price	2	11.77
Total	**17**	**100**

5.3. Frequent Words

Determining word frequencies in any manuscript provides a clear picture of the patterns of words used. It is also necessary to gain knowledge about the most often appearing words in tweets; by doing so, one can learn about the structure and type of tweets; a word cloud can beused to show this. The greater the word within the word cloud, the more frequently that term appears in the corpus [42]. The results are discussed in comparison Lazada and Shopee.

5.3.1. Lazada

The text processing steps produced a total of 20 most frequent words for both companies. In addition, each company has a list of positive and negative words. The results were shown in a word cloud where the bigger the size, the more frequent the word, as illustrate in Figure 2.

Among the 20 positive words, "lazada" have a total of 351, which is considered a default word, so it will not be discussed in this section. The second word is "shope", with a total of 89 occurrences in the document. The word "shope" is the root word of Shopee. The next word is "avail", which means available, with a total occurrence of 42 times. "Thank", "order," and "voucher" is the top five most frequent words. Finally, for the negative sentiment, "boycott", "aldub", "project", "tbadnboycottmzetxapt", and "goldilocks" are the most frequent words that appear more than 40 times in the text. These five frequent words are from similar tweet content where users advocate boycott action toward Lazada. The results and sample tweets are summarised in Table 6.

5.3.2. Shopee

From the analysis in Figure 3, "shope" is the most frequent word for both positive and negative polarity. Since "shope" is the keyword, this particular word will not be included in the result. For frequent positive words, "album" has a total number of 149 occurrences. It is followed by "follow", "repli", "belanjasmsdishopee", "mantul", "miminshopeekasihhadiah", and "shopeepaymantulsal" which occur more than 90 times in the document.

These six frequent words refer to similar content where users participate in a give away contest and retweet the content that the organiser requires, as illustrated in Table 7 . On the other hand, the overall number of negative sentiments for Shopee is lower, so the total number of occurrences for frequent words is also lower when compared to positive sentiments. The words "album", "crazi", and "limit" have the occurrence of more than 10 times compare to the other words.

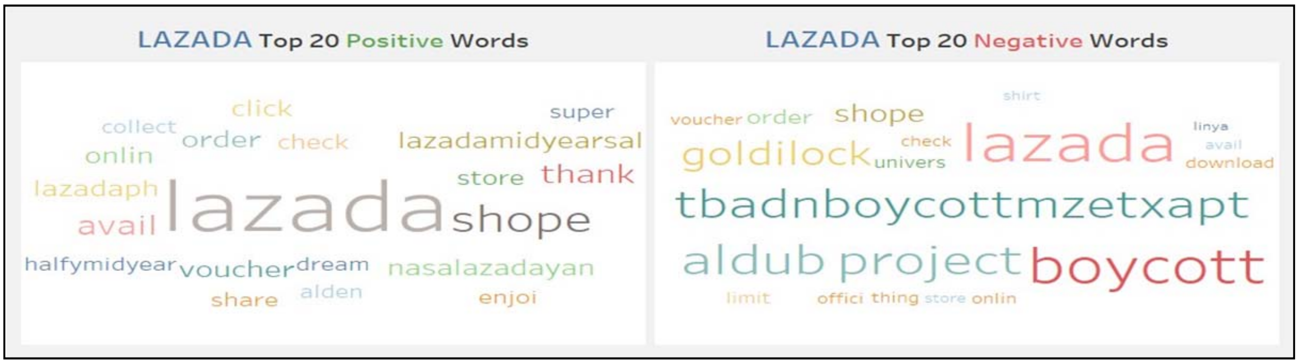

Figure 2. Lazada Word Cloud

Table 7. Positive and Negative Frequent Words for Lazada

Frequent Words in Positive Tweets	Total
lazada	351
shope	89
avail	42
thank	39
order	32
voucher	30

Frequent words in Negative Tweets	Total
lazada	109
boycott	91
aldub	62
project	62
tbadnboycottmzetxapt	61
goldilock	40

Figure 3. Shopee Word Cloud

Table 8. Positive and Negative Frequent Words of Shopee

Frequent Words in Positive Tweets	Total
shope	522
album	149
follow	110
repli	107
belanjasmsdishope	96
mantul	96

Frequent Words in Negative Tweets	Total
shope	90
album	12
crazi	11
limit	11
chill	9
price	8

6. Discussion

The primary objective of this study was to investigate the public perception of Lazada and Shopee on social media. Based on the sentiment analysis results, both companies have more positive than negative sentiments. This can indicate that Twitter users are sharing a reasonable opinion on social media about Lazada and Shopee.

From the content analysis and frequent word analysis, we can identify the topic discussion of Twitter users and the trending case. Based on the positive sentiment finding of the first keyword, Lazada, the users are sharing more about promotion on social media. Users and the seller from Lazada are utilising the Twitter platform to do advertising on their online store and products. Social media allow marketers to target audience and consumers easily and effectively reach out to the people most interested in what they offer.

Furthermore, word of mouth can be used to market items in ways that are not possible through traditional advertising. Besides promotion, the users also tweet more about the contest organised by the seller and

Lazada on social media and the reward available. The giveaway contest serves as an engaging method for connecting sellers and consumers. The retweet action can enhance the number of positive interactions with customers and their likelihood of recommending the company to others they know. In addition, users sharing the available reward will drive increased traffic to the e-commerce websites or apps, which will benefit both company and the seller.

Twitter users also share their pleasant experiences on Lazadaand encourage people to purchase through Lazada. The persuasive tweet will convince others, readers, to perform a transaction on Lazada. From the frequent word analysis, we can indicate that users who own an online store on Lazada also own another store on different platforms such as Shopee. This can be proved by the word "shope", "avail", and "order". When users promote products on social media, they will include the other online store they own in the post to enhance the odds of obtaining more sales and reaching out to other consumers, as some social media users may only have one e-commerce account. Mainly, the term "thanks" refers to an appreciation note from Lazada to users where users contact them through Twitter. Using Twitter as a customer service tool, Lazada can act quickly and respond to users by direct messaging or responding to users. This action shows Lazada's concern about the problem and their awareness of the users' issues. Because of the excellent customer service, customers are eager to convert from being a one-time consumer to a loyal customers. In addition, "thanks" also refers to Lazada's appreciation from customers for the service they provided and also refers to Lazada for appointing Alden Richards, one of the famous celebrities in the Philippinesas the brand ambassador during the 6.6 Mid-Year Sale Campaign [43].

On the other hand, the most prominent negative topic of the Twitter users about Lazada was the boycott issues. Users are expressing their dissatisfaction with Lazada regarding the new marketing move. This is in line with the most frequent negative words: "boycott", "aldub", "project", and "tbadnboycottmzetxapt". These four words refer to similar tweets where an ALDub Nation fan raises a boycott campaign for solo projects of the celebrity. ALDub Nation is a fans club on a show for Alden Richards and Maine Mendoza. [44]. In March 2021, a variety show announced that actress Mendoza would be paired in the show with her real-life partner, and this announcement irritated ALDub followers. The followersat first boycotted the production company and created a hashtag #TBADNBoycottMZETxAPT to protest real-life pairing [45].

After that, the followers' boycott action upgraded to boycott all ALDub's solo projects and the company that collaborated with the single artist, including Lazada, as Lazada only appointed Alden as a brand ambassador without Mendoza. The boycott movement will harm Lazada, as consumers will stop buying from the site, resulting in their reputation suffering. Users also shared their opinion toward Lazada and Shopee by comparing these two main rivals in the SEA e-commerce sector. Users share that they have had unpleasant experiences with Lazada, such as delayed delivery and receiving a fake product from the seller. The negative frequent words "shope" and "order" can be proved this. A defective or fraudulent product sent to customers would tarnish the company's reputation and cause customers to lose trust in the products sold on the platforms . Negative information is more compelling, informative, and beneficial, say consumers. Negativity bias is when buyers pay more attention to negative online reviews than positive ones [46,47].

Moving on to the next keyword, Shopee**.** Based on the findings, the positive sentimenttopic discussion is similar to Lazada, where users share about promotions, contests, rewards and personal experiences on Shopee. There are also some questions on Twitter about where to buy a particular product on Shopee. This shows that Twitter users are very confidence with the Shopee platform when performing the online transaction. From the frequent wordanalysis, "album" appears as the most common word as users promote the album they sell on the platform. Some sellers also use the album as a prize for giving away contests. Through such an contest, the seller creates brand awareness among social media users and targets more audiences interested in the product. Shopee also extensively uses the social media platform as a consumer engagement and marketing tool, hosting a sale contest and a giveaway contest. Shopee is using the giveaway contest to raise awareness among social mediausers about the current marketing program they are working on and to encourage more people to visit their applications. This can be proved by frequent words "follow", "repli", "belanjasmsdishopee", "mantul", "miminshopeekasihhadiah", "shopeepaymantulsal", which appear more than 90 times in the document, as depicted in Figure 3 and Table 8 Shopee's positive sentiment, Twitter users are more likely to complain about the service provided by the seller, such as do not reply messages. On the Shopee platform, the seller and consumer can contact directly through the chat service. Once customers do not receive any reply from the seller, users will feel frustrated with the store and find another alternative. In addition, some users are complaining about the customer service of Shopee, where their issues do not follow up by the teams. Customer service is a crucial area for an e-commerce business. [48] discovered that perceived customer service quality significantly impacts satisfaction, which in turn influences a customer's future purchase intention. As a result, Shopee must consider this issue in order to avoid an increase in customerchurn

7. Conclusion

Since people develop relationships through social media and share sentiments, complaints, and opinions of all kinds. Social media networks such as Twitter have become essential marketing sensing channels. Those online reviews significantly impact customers' actions and attitudes toward the company or brand. This research aims to detect the polarity and the topic discussed by the users about Lazada and Shopee on Twitter. It also includes how social sentiment will affect a consumer's buying decision [22,25,27]. This study analysed 927 and 1322 clean data for Lazada and Shopee, respectively, and measured their sentiment using the Lexicon-based approach. The result shows that both keywords receive more favourable opinions than negative opinions. The major theme addressed in Lazada is the promotion of positive sentiment and emotion for negative sentiment. For Shopee, users are discussing more promotions, contest and complaining the services provided. Boycott difficulties that Lazada faced was identified, and their goal on this movement through frequent word analysis. Lazada must resolve the issue as soon as possible because the impact of Negative Word of Mouth will tarnish the firm'simage, influence a customer's purchasing decision, and the company may lose market share.

Sentiment analysis is the most recent trend in understanding the needs of the general public; itis a more straightforward and cost-effective technique to understand how people feel about a specific subject of matter and the brand impact of micro-blogging. In this scenario, we studied people's attitudes regarding the e-commerce industry and identified recent concerns with Lazada and Shopee via Twitter. The sentiment result can be used to improve the company's operations.

The impact of the negative sentiment was significant where the microblogging platform involved many users, and the news could spread in a short time, and it became Negative WOM. Thus, Lazada and Shopee have to consider the negative opinion and tackle the problem as quickly as possible. To reduce the negative WOM impact, the companies need to know what and why people share the negative thoughts on the social media platform. Finding out the reason behind the topic will give the companies a better understanding and better strategise their future marketing plan. Furthermore, the company need to consider the nature of the complaint. If the company is in a fault situation, it is important to issue an apology statement to the customers with the further action that the company will take.

The boycott action advocated by ALDub fans was identified as one of the most severe challenges confronting Lazada in this study. The number of boycott messages was numerous, and persuading others to join this boycott movement was easy. Lazada was in a dilemma in which they should or should not respond on this issue. Lazada can actively respond to consumer boycotts by offering further discounts or a free shipping program to address this issue. In addition, it could assist Lazada in removing or distancing itself from the boycott's original source or concern.

Next, the company should not ignore unfavourable feedback but prioritise it. Ignoring a nasty comment is one of the worst things to happen to a brand. When confronted with a significant complaint, engage those customers as soon as possible [49]. Listening to their concerns and responding with a solution can increase their satisfaction and experience with the organisation. By giving the immediate consumer attention, they may show their gratitude with a positive review, which can improve the perception of the company to hundreds of other customers [28,42].

Although the research has found useful information and results, this study still encounters several limitations that need to be identified and known. First and foremost, this study was limited to collecting the data due to time constraints. The data collected in this research are from 25-28 May 2021, only four days. Therefore, the data might not be sufficient to make an assumption and conclusion on the Twitter users' opinion sharing.

In addition, there is some constraint on the RapidMiner tool used in this research. In this research, the tool only manages to search the keyword in general and cannot extract certain location data, for example, a tweet about Lazada and Shopee in Malaysia. This is due to the privacy issue that Twitter users do not disclose their location when they tweet. In RapidMiner, the tool uses Twitter standard search API to scrap the Twitter data, and there is a limit on search data where the data that extract is only passed seven days before the data collection date.

For future research on this topic, it is recommended to acquire a larger data set, focus on a specific location, and collect data over an appropriate period of time in order to obtain more accurate estimates of public opinion.

Finally, the themes that emerged from the sentiment analysis could be used to develop variables into quantitative survey questions for additional validation or exploratory research.

References

[1]. TechWire Asia (2021). Mapping E-commerce Companies in Southeast Asia. Retrieved from: https://techwireasia.com/2021/11/mapping-e-commerce-companies-in-southeast-asia/ [accessed: 01 May 2022].

[2]. Kim, Y., Dwivedi, R., Zhang, J., & Jeong, S. R. (2016). Competitive intelligence in social media Twitter:: iPhone 6 vs. Galaxy S5. *Online Information Review*, *40*(1), 42-61. https://doi.org/10.1108/OIR-03-2015-0068

[3]. Ali, Y. S., Hussin, A. R. C., & Dahlan, H. M. (2020). Electronic Word of Mouth engagement in social commerce platforms: An empirical study. *Information Development*, *36*(3), 438-456. https://doi.org/10.1177/0266666919867488

[4]. Goraya, M., Jing, Z., Shareef, M. A., Imran, M., Malik, A., & Akram, M. S. (2021). An investigation of the drivers of social commerce and e-word-of-mouth intentions: Elucidating the role of social commerce in E-business. *Electronic Markets*, *31*(1), 181-195. https://doi.org/10.1007/ s12525-019-00347-w

[5]. Sheikh, Z., Yezheng, L., Islam, T., Hameed, Z., & Khan, I. U. (2019). Impact of social commerce constructs and social support on social commerce intentions. *Information Technology & People*, *32*(1), 68-93. https://doi.org/10.1108/ITP-04-2018-0195

[6]. Yang, Y., Park, S., & Hu, X. (2018). Electronic word of mouth and hotel performance: A meta-analysis. *Tourism management*, *67*, 248-260. https://doi.org/10.1016/j.tourman.2018.01.015

[7]. Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of general psychology*, *5*(4), 323-370.

[8]. Esmark Jones, C. L., Stevens, J. L., Breazeale, M., & Spaid, B. I. (2018). Tell it like it is: The effects of differing responses to negative online reviews. *Psychology & Marketing*, *35*(12), 891-901.

[9]. Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism management*, *47*, 140-151.

[10]. Chevalier, J. A., Dover, Y., & Mayzlin, D. (2018). Channels of impact: User reviews when quality is dynamic and managers respond. *Marketing Science*, *37*(5), 688-709.

[11]. Sohaib, M., Akram, U., Hui, P., Rasool, H., Razzaq, Z., & Khan, M. K. (2019). Electronic word-of-mouth generation and regulatory focus. *Asia Pacific Journal of Marketing and Logistics*, *32*, 23-45

[12]. Young Kim, J., & Hwang, J. (2022). Who is an evangelist? Food tourists' positive and negative eWOM behavior. *International Journal of Contemporary Hospitality Management*, *34*(2), 555-577.

[13]. Lis, B., & Fischer, M. (2020). Analyzing different types of negative online consumer reviews. *Journal of Product & Brand Management*, *29*(5), 637-653. doi: 10.1108/JPBM-05-2018-1876

[14]. Park, S., & Woo, J. (2019). Gender Classification Using Sentiment Analysis and Deep Learning in a Health Web Forum. *Applied Sciences (2076-3417)*, *9*(6). https://doi.org/doi:10.3390/app9061249

[15]. Hassan, M. K., Hudaefi, F. A., & Caraka, R. E. (2022). Mining netizen's opinion on cryptocurrency: sentiment analysis of Twitter data. *Studies in Economics and Finance*, *39*(3), 365-385.

[16]. Loureiro, M. L., & Alló, M. (2020). Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the UK and Spain. *Energy Policy*, *143*, 111490.

[17]. Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006, August). Yale: Rapid prototyping for complex data mining tasks. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 935-940).

[18]. Aslam, S. (2021). Twitter by the numbers: Stats, demographics & fun facts. *Omnicoreagency. com*. Retrieved from: https://www.omnicoreagency.com/twitter-statistics/ [accessed: 06 January 2022].

[19]. Yan, Q., Wu, S., Zhou, Y., & Zhang, L. (2018). How differences in eWOM platforms impact consumers' perceptions and decision-making. *Journal of Organizational Computing and Electronic Commerce*, *28*(4), 315-333.

[20]. Yen, C. L. A., & Tang, C. H. H. (2019). The effects of hotel attribute performance on electronic word-of-mouth (eWOM) behaviors. *International Journal of Hospitality Management*, *76*, 9-18.

[21]. Lee, J., Park, D. H., & Han, I. (2011). The different effects of online consumer reviews on consumers purchase intentions depending on trust in online shopping malls An advertising perspective. *Internet Research*, *21*(2), 187-206.

[22]. Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of consumer psychology*, *24*(4), 586-607.

[23]. Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of marketing*, *80*(6), 69-96.

[24]. Liang, T. P., & Lai, H. J. (2002). Effect of store design on consumer purchases: an empirical study of on-line bookstores. *Information & management*, *39*(6), 431-444.

[25]. Puccinelli, N. M., Goodstein, R. C., Grewal, D., Price, R., Raghubir, P., & Stewart, D. (2009). Customer experience management in retailing: understanding the buying process. *Journal of retailing*, *85*(1), 15-30.

[26]. Cui, G., Lui, H. K., & Guo, X. (2012). The effect of online consumer reviews on new product sales. *International Journal of Electronic Commerce*, *17*(1), 39-58.

[27]. Le, L. H., & Ha, Q. A. (2021). Effects of negative reviews and managerial responses on consumer attitude and subsequent purchase behavior: An experimental design. *Computers in Human Behavior*, *124*, 106912.

[28]. Cheung, C. M., & Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision support systems*, *54*(1), 461-470.

[29]. Nam, K., Baker, J., Ahmad, N., & Goo, J. (2020). Determinants of writing positive and negative electronic word-of-mouth: Empirical evidence for two types of expectation confirmation. *Decision Support Systems*, *129*, 113168.

[30]. Smith, N. K., Larsen, J. T., Chartrand, T. L., Cacioppo, J. T., Katafiasz, H. A., & Moran, K. E. (2006). Being Bad Isn't Always Good: Affective Context Moderates the Attention Bias Toward Negative Information. *Journal of Personality and Social Psychology*, *90*(2), 210-220.

[31]. Singh, S., & Pareek, A. (2022). Sentiment Analysis on Public Transportation Using Different Tools and Techniques: A Literature Review. In *International Conference on Emerging Technologies in Computer Engineering* (pp. 99-110). Springer, Cham.

[32]. Miloš, J., Milan, V., Boris, D., & Milija, S. (2014). Using RapidMiner for Research: Experimental Evaluation of Learners.

[33]. Proellochs, N., & Feuerriegel, S. (2021). Sentiment analysis: dictionary-based sentiment analysis. Retrieved from: https://cran.rproject.org/web/packages/SentimentAnalysis/index.html [accessed: 18 January 2022].

[34]. Hota, H. S., Sharma, D. K., & Verma, N. (2021). Lexicon-based sentiment analysis using Twitter data: a case of COVID-19 outbreak in India and abroad. In *Data Science for COVID-19* (pp. 275-295). Academic Press.

[35]. Chinnasamy, P., Suresh, V., Ramprathap, K., Jebamani, B. J. A., Rao, K. S., & Kranthi, M. S. (2022). COVID-19 vaccine sentiment analysis using public opinions on Twitter. *Materials Today: Proceedings*.

[36]. Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).

[37]. Vijayarani, S., & Janani, R. (2016). Text mining: open source tokenization tools-an analysis. *Advanced Computational Intelligence: An International Journal (ACII)*, *3*(1), 37-47.

[38]. Verma, T., Renu, R., & Gaur, D. (2014). Tokenization and filtering process in RapidMiner. *International Journal of Applied Information Systems*, *7*(2), 16-18.

[39]. White, M. D., & Marsh, E. E. (2006). Content analysis: A flexible methodology. *Library trends*, *55*(1), 22-45. https://doi.org/10.1353/lib.2006.0053.

[40]. Sharma, A., & Ghose, U. (2020). Sentimental analysis of twitter data with respect to general elections in india. *Procedia Computer Science*, *173*, 325-334. https://doi.org/10.1016/j.procs.2020.06.038

[41]. Fang, J., Hu, L., Hossin, M. A., Yang, J., & Shao, Y. (2019). Polluted online reviews: The effect of air pollution on reviewer behavior. *International Journal of Electronic Commerce*, *23*(4), 557-594.

[42]. Bmplus. (2021, June 4). *Business Mirror*. Retrieved from: https://businessmirror.com.ph/category/bmplus/ [accessed: 05 May 2022]

[43]. Cabbuag, S. (2021). Striving for authenticity: The foundation and dynamics of the AlDub nation fandom. *Plaridel*. Advance online publication. Retrieved from: https://www.plarideljournal.org/article/striving-for-authenticity-the-foundation-and-dynamics-of-the-aldub-nation-fandom/ [accessed: 15 May 2022].

[44]. Biong, I. (2021, March 26). *ALDub fans urge boycott of 'Daddy's Gurl' over Maine-Arjo pairing*. Retrieved from: https://entertainment.inquirer.net/406604/aldub- fans-launch-boycott-movement-of-daddys-gurl-over-maine-arjo-pairing [accessed: 15 May 2022].

[45]. 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.

[46]. Osman, H., D'Acunto, D., & Johns, N. (2019). Home and away: Why do consumers shy away from reporting negative experiences in the peer-to-peer realms?. *Psychology & Marketing*, *36*(12), 1162-1175.

[47]. Kassim, N., & Abdullah, N. A. (2010). The effect of perceived service quality dimensions on customer satisfaction, trust, and loyalty in e-commerce settings. *Asia Pacific Journal of Marketing and Logistics*, *22*(3), 351-371.

[48]. Grégoire, Y., Salle, A., & Tripp, T. M. (2015). Managing social media crises with your customers: The good, the bad, and the ugly. *Business Horizons*, *58*(2), 173-182.