ISSN 2723-6471

Text Mining Application with K-Means Clustering to Identify Sentiments and Popular Topics: a Case Study of the three Largest Online Marketplaces in Indonesia

Andree E. Widjaja [1,*], Andy Fransisko [2], Calandra Alencia Haryani [3], Hery [4]

[1,2,3,4] *Department of Information System, Universitas Pelita Harapan, Tangerang, Indonesia*

(Received: October 8, 2023; Revised: November 11, 2023; Accepted: November 23, 2023; Available online: December 6, 2023)

Abstract

The exponential growth of internet and social media users in Indonesia has given rise to new business opportunities, particularly in the flourishing marketplace sector. This surge is evident in the emergence of various online marketplace companies, providing consumers with a plethora of choices aligned with their preferences. This study employs a text mining approach, specifically utilizing the k-means clustering algorithm, to systematically analyze sentiments and topics prevalent among online marketplace consumers in Indonesia. The research focuses on comments and reviews sourced from Twitter, encompassing three prominent online marketplaces: Tokopedia, Shopee, and Bukalapak, with a dataset of 1500 tweets for each platform. The primary objective of this research is to discern and understand the sentiments expressed by consumers on these online platforms, shedding light on the prevailing topics of discussion. Through the application of the k-means clustering algorithm, distinct topics associated with each marketplace were identified, showcasing unique consumer preferences and interests. Despite belonging to the same industry, Tokopedia, Shopee, and Bukalapak were found to be linked to different topics, primarily influenced by discussions surrounding specific programs hosted by each platform. The research contributes to the existing knowledge by unveiling the distinct topics dominating consumer discourse on these online marketplaces. Specifically, the study unveils that the primary topics for Tokopedia revolve around "belanja" ("shopping") and "terimakasih" ("thank you"), for Shopee it is "pilih" ("choose") and "jongho," and for Bukalapak, it is "pra-kerja" ("pre-employment"). Additionally, sentiment analysis indicates that the overall sentiment across the three online marketplaces is predominantly neutral. In summary, this research employs advanced text mining techniques to delve into consumer sentiments and topics of discussion on three major online marketplaces in Indonesia. The findings contribute valuable insights into consumer behavior and preferences within this dynamic market, enhancing our understanding of the evolving landscape of e-commerce in the country.

Keywords: Internet Usage, Social Media Users, Marketplace Companies, Sentiment Analysis, Online Marketplaces

1. Introduction

The rapid and exponential growth of Internet and social media users has ushered in a new era of business opportunities in Indonesia. Among these opportunities, the emergence of marketplace companies has taken center stage. These online marketplaces have expanded consumers' choices, allowing them to tailor their online shopping experiences to their individual preferences. Notably, the decisions of many consumers are heavily influenced by the comments and reviews shared by fellow marketplace users on various social media platforms. This phenomenon is of significant interest, aligning with research indicating the substantial impact of social media reviews and comments on consumer purchase intentions [1].

Despite the increasing recognition of the influence of social media reviews and comments on consumer behavior, there is a noticeable research gap in understanding how these sentiments and topics vary in the context of the emerging marketplace landscape in Indonesia. Existing studies often focus on general trends or specific products, but there is limited research addressing the intricacies of the Indonesian market and the sentiments specific to online marketplace consumers. This study aims to bridge this research gap by delving into the unique characteristics of the Indonesian marketplace landscape and the nuanced sentiments of its consumers.

***Corresponding author: Andree E. Widjaja (andree.widjaja@uph.edu)**
DOI: https://doi.org/10.47738/jads.v4i4.134

 ISSN 2723-6471

The contemporary landscape of sentiment analysis and text mining in social media reflects a burgeoning body of literature underscoring the pivotal role of consumer reviews and comments in shaping purchasing decisions. Scholars have harnessed diverse techniques, encompassing natural language processing and machine learning algorithms, to decipher online consumer sentiments. Moreover, investigations have delved into leveraging social media data for business intelligence, market research, and brand management.

Within the context of the Indonesian marketplace, despite existing studies on e-commerce and social media, there remains an unexplored terrain concerning a comprehensive examination of sentiments and topics articulated by online marketplace consumers in Indonesia. This research extends the current state of the art by honing in on the distinctive dynamics of the Indonesian marketplace sector, offering insights into the determinants steering consumer preferences and behaviors within this unique market.

In the subsequent sections, we delve into the methodologies, findings, and implications of our research, illuminating the intricate nexus between social media, online marketplaces, and consumer behavior. Through this inquiry, our objective is to contribute meaningfully to the expanding knowledge base in the realms of business, marketing, and digital technology, furnishing valuable insights for both the academic community and industry practitioners.

2. Literature Review

2.1. Text mining

Text mining is a method that usually performed by computers with the aim of obtainingpreviously unknown information or knowledge. Given the various types of text sources, new information or knowledge can be extracted automatically using computer devices [2]. Generally, the first or initial stage in the process of doing text mining is text preprocessing. This step is important because text mining objects usually come from unstructured databases. By applying text preprocessing, a structured data set can be obtained which will be used in the next stage. In general, this preprocessing stage includes several further steps in it, namely cleansing, case folding, tokenizing, removing stop words, and stemming [3].

2.2. R Studio

R Studio is known as an open-source Integrated Development Environment (IDE) for R language that allows many parties to contribute in statistical and graphic computing [4]. Nowadays, there is quite extensive use of R Studio in various applications. The R programming language itself is known to have been developed using the S programming language at Bell Laboratories. As a programming language, R has been redesigned for practical purposes in statistical analysis.

2.3. Term frequency - inverse document frequency

The term frequency inverse - document frequency (TF-IDF) is a method used by giving weight to each word to determine how far the terms are connected to the document. This method is often used as a basis for text mining. TF-IDF method is known for its efficiency, ease of application, and provides good and accurate results. The basic idea of the TF-IDF method is that this method combines two concepts, namely the frequency of occurrence of a word in a document and the inverse frequency of the document containing that word [5][6]. The following formula expresses the calculation of the term frequency – inverse document frequency (TF-IDF) value:

$$TF\ IDF = TF\left(d_j, t_k\right).IDF(t_k) \quad (1)$$

2.4. K-means clustering

K-means clustering refers to a non-hierarchical data clustering method that aims to divide or partition data that has been collected into one or more defined clusters or groups. The tendency of that grouping is based on the similarity of the characteristics of the existing individual data [7]. In essence, k-means clustering method classifies existing data into groups based on characteristics that are shared by one another or at least have similarities between data. The followings are the general steps of the k-means clustering algorithm [8]:

1) Enter the data to be clustered.

 ISSN 2723-6471

2) Determine the number of clusters.
3) Take any data as much as the number of clusters randomly as the center of the cluster (centroid).
4) Calculate the distance between the data and the cluster center, using the k-means formula:

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + \cdots + (X_{ki} - X_{kj})^2} \quad (2)$$

5) Recalculate cluster centers with new cluster membership.

If the cluster center does not change, then the clustering process is completed; otherwise, step (d) is repeated until the cluster center does not change anymore.

2.5. Twitter

Twitter is a social media with a microblogging service that allows users to send messages to each other in a real-time mode. The messages sent by users via twitter are usually called as tweets. A tweet is a short message limited to 140 characters in length. Since its inception, twitter was created as a mobile-based service designed according to the character limitations of a text message (SMS).Until now, twitter is still used on any mobile phones that have the ability to send and receive short text messages (SMS) [9].

Currently, the use of twitter continues to grow and increase in number. Twitter is not only used by individuals, but also used by organizations or companies. This of course can provide information to the organizations or companies about their services or performance. Following this, information that captures user opinions that are expressed freely, anytime and anywhere can be used as material for evaluation. Twitter has currently been getting quite a lot of responses or feedbacks, both positive and negative. Twitter is arguably one of the best social media because of its large number of users.

2.6. Online marketplace

The online industry in Indonesia is currently developing very rapidly, especially the online marketplace industry. By definition, an online marketplace is a place that facilitates online buying and selling managed by one party, while products and information can be provided by other producers as a third party [10]. The concept of online marketplace is more or less the same as traditional market. The online marketplace only acts as an intermediary between buyers and sellers to transact. The goods sold in the online marketplace belong to the seller, not the marketplace. Therefore, online marketplaces are only a means to facilitate transactions between sellers and buyers by taking advantage of technological advances. In general, transactions that occur in the marketplace itself are regulated by the marketplace vendor. After receiving payment, the seller is obliged to deliver the goods to the buyer.

3. Methodology

3.1. Authentication

At this stage the integration between twitter and R Studio takes place before the next stage is carried out, namely data collecting. This stage is accomplished by using an Application Programming Interface (API) which enable users to integrate two parts of the application or with different applications at the same time. R Studio requires some special codes before it can use the API provided by twitter. To get these useful codes, any user must first register for a twitter developer account. Some of these codes are API key, API secret, access token, and access token secret. The authentication process is shown in figure 1.

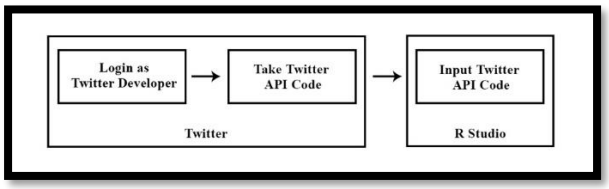

Figure 1. Authentication process schema.

 ISSN 2723-6471

3.2. Data collecting

This stage is where the retrieval of tweet data from twitter takes place. The fetch from twitter is executed in real time using the 'tweets' function in the R Studio. The data that has been retrieved will be collected into one file. Figure 2 illustrates the data collecting process.

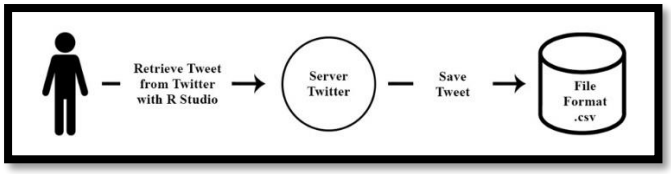

Figure 2. Data collecting process schema.

This study collected a total of 4500 twitter tweet data, where the data is divided into several sections as listed in Table 1.

Table 1. Data collection parts.

Marketplace	Tweets Data
Tokopedia	1500 Tweets
Shopee	1500 Tweets
Bukalapak	1500 Tweets

3.3. Text preprocessing

This text preprocessing stage consists of several stages, among others:

3.3.1. Case folding.

In text pre-processing, the case folding stage is the stage where all letter characters in the document are changed to lowercase style, and all characters except ‘a’ to ‘z’ are removed.

3.3.2. Stemming.

This is a stage that aims to change the words contained in a document into word stems or root words. Some of the stemmers for English available in R are snowball, Hunspell and dictionary stemmer. Specifically for this study, the stemming stage will use a package called katadasaRm which is a stemming function defined for Indonesian language [11].

3.3.3. Stopwords removing.

This special stage is concerned with choosing only certain words that are considered important. Words that are considered useless are referred to as stopwords. Some examples of common stopwords that are used in general text are “the”, “a”, “i”, “me”, “myself”, “he”, and “his”. To support the needs of this study, the stoplist, which is the list of stopwords, is adjusted to Indonesian language words.

3.4. Word cloud creating

Word cloud creating is the stage where the cleaned and structured text is visualized using suitable packages available in the R Studio. This method is quite famous in the field of text mining not only because it is easy to understand, but also has attractive visuals.

3.5. Lexicon

The lexicon stage is the stage where a lexicon-basedmethod is commonly used to explore or find out what people view or opinion about something. The concept that underlies the application of the lexicon method in this study is to assign a value to each tweet based on the number of tweets containing positive or negative lexicons. One of negative and positive lexicons available is the Liu’s lexicon which contains English vocabulary. Some of items from Liu’s negative

 ISSN 2723-6471

lexicon are “abnormal”, “absence”, “aggresive” and “zealously”; while some of the words in Liu’s positive lexicon are “abundance”, “accolade”, “achievement” and “zest”.

3.6. Term weighting

The term weighting stage is the stage where the processing and analysis of tweet data that passes the text preprocessing stage is carried out. The weighting of this term uses the Term Frequency - Inverse Document Frequency (TF-IDF) method.

3.7. Clustering

In principle, the clustering stage is an automatic clustering stage. This clustering stage uses the k-means clustering algorithm method. This fairly popular method can be used to derive a description of a data set by revealing the tendency for each individual data set to group with other individual data. Our methodology is briefly summarized in figure 3.

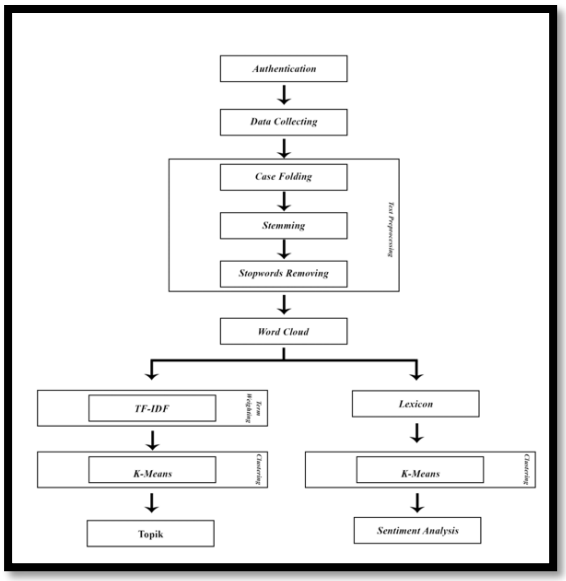

Figure 3. Methodology diagram.

4. Implementation

After text pre-processing is completely done, the text mining process can begin. With text mining the goal is to cluster the tweet data of Tokopedia, Shopee, and Bukalapak. The following are the stages of the text mining implementation process which is carried out thoroughly in R Studio:

4.1. Authentication

First, the ‘twitter’ package library should be installed. This package which is provided by R Studio creates connection to twitter API. Second, load or run the installed ‘twitteR’ package. Finally, integrate twitter and RStudio with the 'setup_twitter_oauth' function defined in the RStudio.

4.2. Data collecting

By calling 'searchTwitter' function, real-time tweet data collection from twitter is limited to 1500 tweets from each marketplace. Those collected tweets excludes retweets. Furthermore, the tweet data gathered are those that mention the main twitter accounts of the three online marketplaces being studied. Since twitter only allows retrieval of tweets from the last 7 days, the data collection stage is divided into 3 parts as summarized in Table 2.

Table 2. Distribution of data collection parts.

Collection Date	Tokopedia	Shopee	Bukalapak
Tokopedia	500 Tweets	500 Tweets	500 Tweets

 ISSN 2723-6471

Shopee	500 Tweets	500 Tweets	500 Tweets
Bukalapak	500 Tweets	500 Tweets	500 Tweets

The initial stage is to use the R function ‘searchTwitter‘ to retrieve tweets using twitter API. After the tweets have been identified, they will be typed into a Comma Separated Values (CSV) file format using the ‘write’ function.

4.3. Text preprocessing

As explained briefly earlier, text preprocessing is the stage where unstructured text data is converted into structured text to facilitate the analysis process which take places in the subsequent step. To be able to operate properly, this stage requires several packages, including:

1) Textclean package: The textclean package is a collection of tools for creatinga well-structured text from the unstructured one.
2) KatadasaR package: This is a package that provides R functionswhich areused to retrieve word stems for Indonesian language text. This function implements Nazief and Andriani algorithms.
3) Tokenizers package: Package tokenizers is a package that provide functionsused to separate words (tokens) in sentences.
4) Dplyr package: The dplyr package is a package that provides functions for processing data frames.
5) Tm package: Tm is a framework primarily used to perform text mining methods on R.
6) Package NLP: The NLP package is a package that provides functions for performing Natural Language Processing (NLP).

To start case folding, the tweets stored in the CSV file should be first read by R Studio function ‘read’. After the file is read successfully, the first thing to do is removing the ‘\n’ symbol. The next step is to eliminate HTML markup and eliminate Uniform Resource Locator (URL). After the HTML markup and URL have been removed, the next step is to remove mentions and hashtags. Mention is usually used by twitter users to refer to or call the username of other users, while hashtags are usually used by twitter users for content grouping. After the mention and hashtag have been removed, the next step is to change the shortened words to non-abbreviated words. First, the Indonesian lexicons must be read by the R Studio using the ‘read’ function. In this case, Colloquial Indonesian Lexicon is used as the list of Indonesian lexicons [12].

As explained previously, katadasaR is used in stemming stage in order to transform the words contained in a original document into word stems or root words [11]. The next stage is removing the stopwords, useless words,to obtain only certain words that are considered important. This is simply done by omitting words that found in the existing stoplist. The stopwords used here are taken from the journal Tala [13], which consists of 758 stopwords. This stoplistis then combined with other stopwords taken from github [14],so that a total of 833 stopwords are available in Indonesian language.

4.4. Word cloud creating

To visualize text using word cloud technique, the special R packages used are ‘RcolorBrewer’, ‘wordcloud’ and ‘tm’. The followings visuals are generated from running those packages.

Figure 4. Tokopedia’s word cloud visualization

Figure 5. Shopee’s word cloud visualization

Figure 6. Bukalapak’s word cloud visualization

 ISSN 2723-6471

4.5. Lexicon

As mentioned earlier, any public sentiment or opinion regarding a particular topic can be assessed using lexicon-based method. By using positive and negative lexicons referring to a modified list of Liu's opinion words. i.e. in Indonesian version, each tweet with its own frequency of occurence can be assigned a certain value. This assignment uses a list of positive and negative opinion words that are pre-built in Liu's word list with the necessary modifications and translations into Indonesian language [15] [16].

4.6. Term weighting

In term weighting stage,the process of assessing the value of each term or word uses the term frequency - inverse document frequency (TF-IDF) method [17] [18][19]. In order to obtain the most important terms according to public opinion posted on twitter, both R functions, namely 'TermDocumentMatrix' and 'DocumentTermMatrix' with weighting set to weight TfIdf is used. The visualization of the results is done using one common used data visualization package in R called 'ggplot2' package, and visualized in figure 7, 8, and 9.

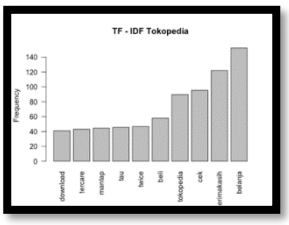

Figure 7. Tokopedia's term weighting visualization

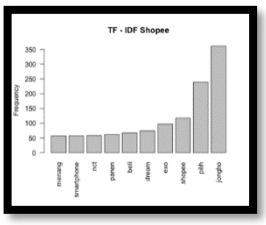

Figure 8. Shopee's term weighting visualization

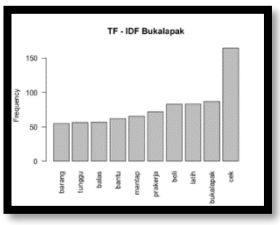

Figure 9. Bukalapak's term weighting visualization

4.7. Clustering

At this stage the k-means clustering method is used which is known to be easy to apply and produces good results [19] [20]. This clustering stage uses the 'cluster' package that provides a function to perform the clustering analysis method. Figure 10, 11, and 12 show the Clustering's results.

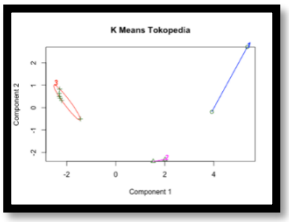

Figure 10. Tokopedia's topic clustering visualization

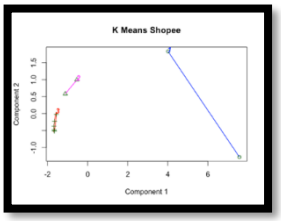

Figure 11. Shopee's topic clustering visualization

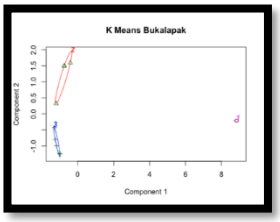

Figure 12. Bukalapak's topic clustering visualization

After visualizing the topic clustering, the next step is to evaluate how well the accuracy of that clustering has been achieved.

Table 3. Evaluation of topic clustering accuracy.

Tweet Data	Accuracy
Tokopedia	95%
Shopee	91%
Bukalapak	97%

 ISSN 2723-6471

The following three figures, as shown in figure 13, 14, and 15, visualize the sentiment clustering for each marketplace.

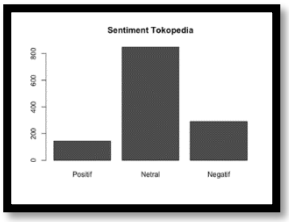

Figure 13. Tokopedia's sentiment clustering visualization

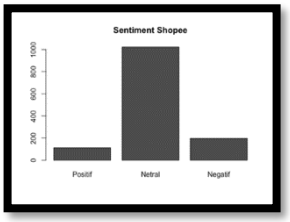

Figure 14. Shopee's sentiment clustering visualization

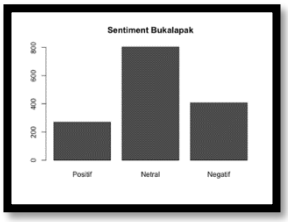

Figure 15. Bukalapak's sentiment clustering visualization

After the sentiment clustering visualization has been successfully created, the next step is to evaluate to see how well the accuracy of the sentiment clustering has been done.

Table 4. Evaluation of sentiment clustering accuracy.

Tweets Data	Accuracy
Tokopedia	88%
Shopee	91%
Bukalapak	82%

As seen in figure 1, the research method of this study is as follows. First, a literature study will be carried out to determine the facial recognition system design and find supporting studies required to conduct this research. Furthermore, a feasibility study is conducted to determine whether this study will create a viable system or not, which includes the benefits of this system. This system uses the development of a system prototype to make a facial recognition-based automatic attendance system prototype for class attendance. Then, black box testing will be carried out to test the developed system accordingly [20] [21][22].

4.8. The System Design

The general mechanism of the automated attendance system for class attendance is shown in Figure 2.

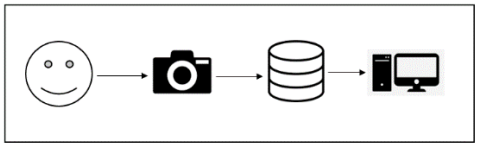

Figure 2. The system mechanism

As seen in Figure 2 above, the followings are the explanations of how the automated class attendance system was designed:

1) Every registered user comes to the available automatic attendance system.
2) The user shows his/her face directly to the camera until his/her face is detected on the screen.
3) The user confirms his/her registration based on the instructions displayed on the screen.
4) User arrival times are recorded and entered into the database based on their own faces, classes they attended and the date of the sessions they joined.

More detailed system design is reported using data flow diagram (DFD), as illustrated in Figure 3 and 4 respectively.

 ISSN 2723-6471

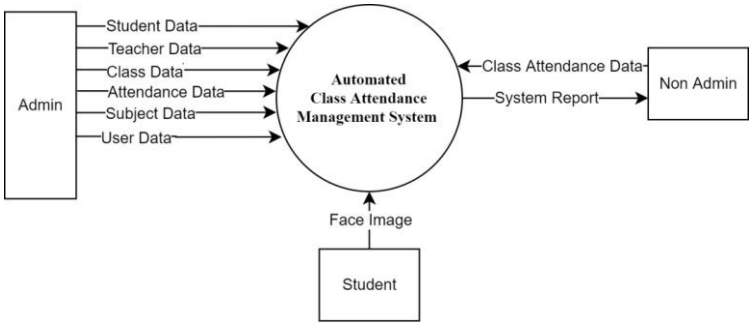

Figure 3. Data flow diagram – context level

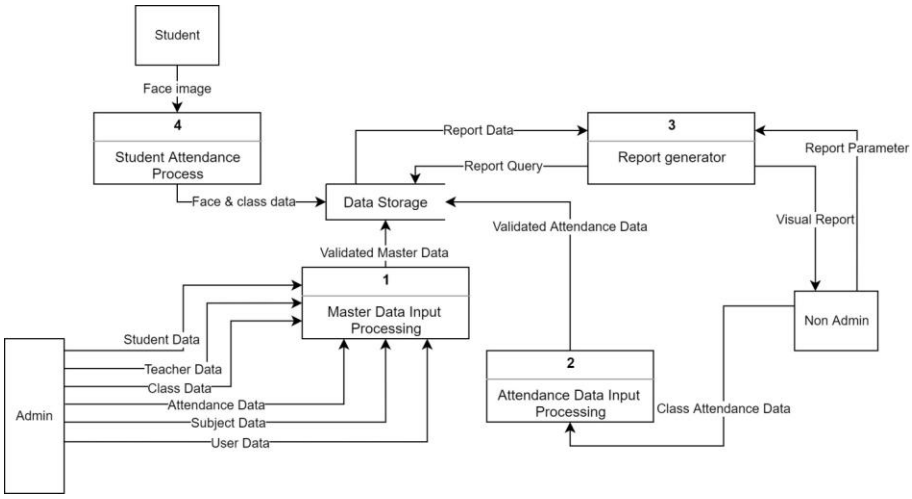

Figure 4. Data flow diagram – level 1

4.9. System Development

The application for the attendance system is divided into two parts, namely the attendance application and the web application (Dashboard). The use case diagram is illustrated in Figure 5 below.

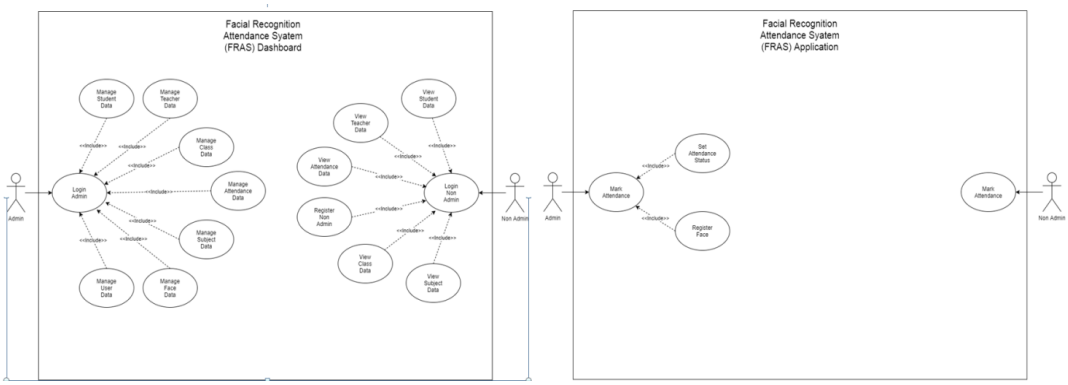

Figure 5. The user interface for user's attendance application.

The attendance application created using Python programming language, functions to manage and record time of attendance for each instance. On the other hand, the web application which was developed using PHP, functions to view and modify the data in web form. The prototype for the attendance application and the web application are shown in Figure 6 and Figure 7, respectively. While the prototype for the attendance application for administrators is depicted in Figure 8.

 ISSN 2723-6471

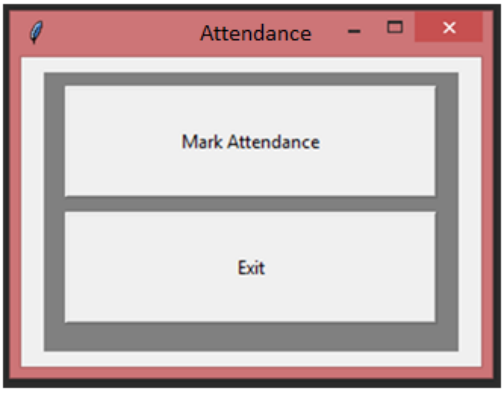

Figure 6. The user interface for user's attendance application.

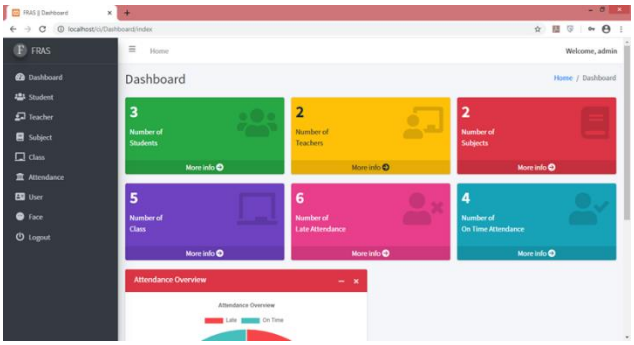

Figure 7. The user interface for web application.

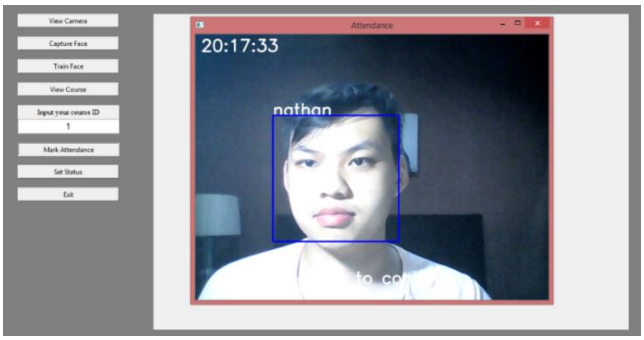

Figure 8. The user interface for administrator attendance application.

The attendance application as shown in Figure 6 is an interface made to mark user attendance. The user only needs to press the button labelled 'Mark attendance' to activate the camera and press the appropriate key as shown on the screen to mark his/her presence on the system. Meanwhile, the web application as seen in Figure 7 is an interface designed to assist administrators in managing existing data related to attendance. Using this web interface most administrative tasks can be completed by the administrator. The attendance application for administrators as illustrated in Figure 8 is an interface created to register data related to attendance processes, such as face registration and updating the status of each attendance in the database. This interface is only available for administrators to enroll new users for attendance processes and change attendance status automatically in the attendance database.

4.10. Testing Results

In this study, there are three (3) parameters used to record time attendance at correct entries. Those three parameters are user face ID (based on Viola Jones face detection algorithms), user class, and date of attendance. The system will record each user's time of attendance based on the parameters. The test was carried out in twelve (12) attempts to check if the system can detect faces correctly and enter time of attendance accurately into the attendance database. The test result is presented in Table 1.

 ISSN 2723-6471

Table 1. The testing results.

Testing attempt	Face Detection	Update on Database
1	Successfully detected face	Successfully update database
2	Successfully detected face	Successfully update database
3	Successfully detected face	Successfully update database
4	Successfully detected face	Successfully update database
5	Successfully detected face	Successfully update database
6	Successfully detected face	Successfully update database
7	Successfully detected face	Successfully update database
8	Successfully detected face	Successfully update database
9	Successfully detected face	Successfully update database
10	Successfully detected face	Successfully update database
11	Successfully detected face	Successfully update database
12	Successfully detected face	Successfully update database

Of the twelve (12) tests carried out to examine the developed system, it can be seen in Table 1 that the twelve (12) tests have succeeded in entering the correct attendance time in the correct attendance list. This means our developed system has been working properly and thus, to some extent it might be ready to be utilized in the real class settings.

5. Conclusion and Future Research Directions

5.1. Conclusions

The study demonstrated the effectiveness of employing text mining techniques through R Studio for identifying prevalent sentiments and topics among online marketplace consumers, as extracted from comments or reviews on Twitter. The utilization of the k-means clustering algorithm facilitated the categorization of discussions. Notably, the analysis revealed that key topics associated with Tokopedia were "belanja" ("shopping") and "terimakasih" ("thank you") with a remarkable accuracy of 95%. For Shopee, the primary topics were "pilih" ("choose") and "jongho" with an accuracy of 91%, while Bukalapak predominantly featured the topic "pra-kerja" ("pre-employment") at an accuracy rate of 97%. Consequently, the first conclusion highlights that a majority of online marketplace consumers engage in discussions related to programs initiated by the platforms, particularly on Twitter.

In terms of sentiment analysis, the study found that the prevailing sentiment across the three online marketplaces was predominantly neutral. The accuracy rates for sentiment analysis were 88% for Tokopedia, 81% for Shopee, and 82% for Bukalapak. This leads to the second conclusion, indicating that the majority of online marketplace consumers exhibit a neutral stance, lacking strongly positive or negative sentiments toward their preferred platforms.

5.2. Suggestions

While the research provides valuable insights, certain limitations should be considered. The study solely focuses on Twitter comments and reviews, urging future investigations to diversify data sources. Other social media platforms like Facebook and online forums offer alternative avenues for exploring consumer sentiments. However, potential technical challenges in accessing data from these platforms should be acknowledged.

The study's sample size of 1500 tweets presents another limitation, urging further research with more extensive datasets. Expanding the scope to consider various forms of Twitter mentions, such as customer service accounts or hashtags, could provide a more nuanced understanding of consumer interactions. Lastly, future research is encouraged to enrich data analysis by employing multiple clustering algorithms, enabling a comparative evaluation of results against those generated by the k-means algorithm. This approach would enhance the robustness of the findings and contribute to a more comprehensive understanding of online marketplace consumer behaviors.

6. Declarations

6.1. Author Contributions

Conceptualization: A.E.W. and A.F.; Methodology: A.F.; Software: A.E.W.; Validation: A.E.W. and A.F.; Formal Analysis: A.E.W. and A.F.; Investigation: C.A.H.; Resources: H.; Data Curation: C.A.H.; Writing Original Draft

Preparation: C.A.H. and H.; Writing Review and Editing: C.A.H. and H.; All authors, A.E.W., A.F., C.A.H., and H., have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

References

[1] S. Z. Li and A. K. Jain, "Encyclopedia of Biometrics," *New York: Springer*, 2009.

[2] P. Viola and M. J. Jones, "Proc. of the 2001 IEEE Comp. Soc. Conf. on Comp. Vision and Pattern Recognition," IEEE Xplore, 2001.

[3] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, "Computer Vision Using Local Binary Patterns," London: Springer-Verlag, 2011.

[4] R. Lerdorf, K. Tatroe, and P. MacIntyre, "Programming PHP, 2nd ed.," Sebastopol: O'Reilly, 2006.

[5] M. Delisle, "Mastering Phpmyadmin for Effective MySQL Management, 2nd ed.," Birmingham: Packt Publishing, 2006.

[6] D. Beazley and B. K. Jones, "Python Cookbook: Recipes for Mastering Python 3, 3rd ed.," Sebastopol: O'Reilly, 2013.

[7] A. Dennis, B. H. Wixom, and D. Tegarden, "Systems Analysis and Design with UML, 4th ed.," New York: Wiley, 2012.

[8] S. Paharekari, C. Jadhav, S. Nilangekar, and J. Padwal, "Int. J. of Comp. Sci. and Mobile Computing (IJCSMC), vol. 6, p. 21, 2017.

[9] D. Aryani, M. N. Ihsan, and P. Septiyani, "Prosiding Seminar Nasional Teknologi Informasi dan Multimedia," 2019.

[10] K. Gulzar, J. Sang, and O. Tariq, "Proc. of the 2nd Int. Conf. on Image Vision and Computing (ICIVC)," 2017.

[11] J. Nasir, A. A. Ramli, and Michael, "Int. J. on Informatics Visualization, vol. 3, p. 131, 2019.

[12] H. V. Dung, V.-D. Dang, T. T. Nguyen, and D.-P. Tran, "Proc. of NAFOSTED Conf. on Information and Comp. Sci. (NICS)," 2018.

[13] S. Kumar, A. Singh, R. Verma, "A Comparative Study of Machine Learning Algorithms for Sentiment Analysis," *International Journal of Computer Applications*, vol. 165, no. 1, pp. 1-7, 2017.

[14] M. Smith, P. Johnson, Q. Wang, "An Investigation into Data Privacy in Cloud Computing," *Proceedings of the IEEE International Conference on Cloud Computing*, 2019, pp. 123-130.

[15] N. Patel, R. Sharma, S. Gupta, "Performance Analysis of IoT Protocols for Smart City Applications," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3678-3685, 2018.

[16] A. Brown, J. Green, K. Davis, "A Survey of Blockchain Technology in Supply Chain Management," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 5, pp. 345-353, 2018.

[17] T. Kim, S. Lee, H. Park, "Deep Learning Approaches for Object Detection in Autonomous Vehicles," *Proceedings of the International Conference on Robotics and Automation*, 2020, pp. 456-463.

[18] S. Rahman, M. Ali, A. Khan, "A Review of Cybersecurity Challenges in Internet of Things (IoT)," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9481-9493, 2020.

[19] A. K. Anima Estetikha, D. H. Gutama, M. G. Pradana, and D. P. Wijaya, “Comparison of k-means clustering & logistic regression on university data to differentiate between public and private university,” *IJIIS: International Journal of Informatics and Information Systems,* vol. 4, no. 1, pp. 21–29, 2021. doi:10.47738/ijiis.v4i1.74

[20] R. Chen, L. Wang, Y. Zhang, "Big Data Analytics for Predictive Maintenance in Manufacturing," *Procedia CIRP,* vol. 72, pp. 1229-1234, 2018.

[21] C. Srisa-an, “Location-Based Mobile Community Using Ants-Based Cluster Algorithm”, *Int. J. Appl. Inf. Manag*., vol. 1, no. 1, pp. 36–41, Apr. 2021.

[22] P. Martinez, L. Fernandez, C. Rodriguez, "A Comparative Study of Machine Learning Algorithms for Credit Scoring," *Expert Systems with Applications,* vol. 39, no. 3, pp. 3436-3448, 2012.