Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)
Vol. 5, No. 1, June 2019, pp. 44~49
ISSN: 2338-3070, DOI: 10.26555/jiteki.v5i1.13087

Sentiment Analysis Using Maximum Entropy on Application Reviews (Study Case: Shopee on Google Play)

Ulinnuha Rhohmawati[1], Isnandar Slamet[2], Hasih Pratiwi[3]
[1,2,3]Department of Statistics Universitas Sebelas Maret, Surakarta, Indonesia

ARTICLE INFO

Article history:
Received May 10, 2019
Revised June 16, 2019
Accepted June 26, 2019

Keywords:
Shopee
Google Play
Sentiment Analysis
Maximum Entropy
Word Association

ABSTRACT

Shopee was one of the e-commerce application that could found on Google Play. The amount of Shopee application reviews on Google Play continues to grow over time. These make the company trying to get the overall information from all reviews because it would take a long time to read each of the reviews on Google Play. Therefore analysis was used using text mining. One part of text mining was sentiment analysis that applied the maximum entropy method to classification. Based on the results of the analysis found an accuracy of 97.32%. By using the maximum entropy method it could be concluded that word association obtained related to "application", "promo", "satisfy", and "discount" for positive sentiment. Meanwhile for negative sentiment, the reviewers of Shopee application on Google Play were related to "problematic", "login", "old", "verification", and "expensive". The results of this research in Indonesian.

Corresponding Author:

Ulinnuha Rhohmawati,
Department of Statistics,
Universitas Sebelas Maret,
Sutami Street 36A, Surakarta 57126, Indonesia
Email: liabagaskara2018@gmail.com

1. INTRODUCTION

Shopee is top five shopping application and most popular in Indonesia. Most Shopee application users who provide reviews in the form of praise or criticism. A review is critical to know the feedback from the product [1]. Currently there are a lot of Shopee application reviews in Indonesia on *Google Play*. It makes difficult for the company to obtain information as a whole because it takes a long time to read the reviews that enter *Google Play* one by one. Sentiment analysis can be used to resolve this problem [2].

This research aims to analyze the sentiment of *Shopee* application reviews on *Google Play*. The analysis can classify whether a review is positive or negative [3]. By using sentiment analysis the company can find out positive and negative reviews from users discuss the application. Then the result of both sentiments will be visualized in the word cloud. Next look for word associations based on the workload. From the word associations the company will get information to what things discussed by user based on positive and negative sentiment. When the company knows what things discussed in positive sentiment, the company can also improve the quality so that users satisfaction fulfilled. Meanwhile, when the company knows the information related to negative sentiment then the company can quickly take action to mitigate it. Sentiment analysis is the process of understanding, extracting and processing textual data automatically to get information [4].

Research of sentiment analysis using Chinese reviews took a long time in pre-processing data [5]. While in this study using reviews of Shopee applications that use Indonesian, so that pre-processing does not take a long time because there is already a package to help to pre-process the Indonesian reviews. Maximum entropy is very useful in several natural language processing applications [6]. So, our method to be used for

sentiment analysis is maxent. Distribution search which results in maximum entropy value aims to get the best probability distribution that is closest to reality [7],[8].

2. RESEARCH METHOD

Shopee application reviews *on* Google Play from 16th February to 8th March 2019 used in this research. Data retrieval using scrapping techniques with extensions from Google Chrome, namely Data Miner [9]. After the data is collected then pre-processing or cleaning data [10]. After that, do manual labelling and divide the data into training and testing [11]. Then using maximum entropy method to sentiment analysis the followings are [12] :

a. Identify specific words in the document (sentence).
b. Form a matrix that contains the value of the occurrence of these specific words with the following index:

$$f_j(a,b) = \begin{cases} 1; \text{if } f_j \text{ appears in document } b \text{ on class } a \\ 0; \text{if } f_j \text{ not appears in document b on class a} \end{cases} \quad (1)$$

c. Make a maximum entropy model with training data that a_j is counting for each class with a generalized iterative scaling procedure

$$a_j^{(0)} = 1 \quad (2)$$

$$a_j^{(n+1)} = a_j^{(n)} \left[\frac{E_p f_j}{E^{(n)} f_j} \right]^{\frac{1}{C}}, \quad (3)$$

Where

$$E_p f_j = \sum_{x \in \varepsilon} p(x) f_j(x) \quad (4)$$

$$E^{(n)} f_j = \sum_{x \in \varepsilon} p(x) f_j(x) \quad (5)$$

$$p^{(n)}(x) = \pi \prod_{j=1}^{k} (a_j^{(n)})^{f_j^{((x)}} \quad (6)$$

$$\forall x \in \sum_{j=1}^{k} f_1(x) = C \quad (7)$$

d. Find the joint probability *p (a, b)* for testing data

$$a = \{positif,\ negatif\} \quad (8)$$

$$p^*(a,b) = \pi \prod_{j=1}^{k} a_j^{f_j(a,b)} \quad (9)$$

e. Determine the topic of testing data document by looking at the largest a^* value in a class

$$a^* = \arg\max\ p*(a,b) \quad (10)$$

The next steps calculate the accuracy from the classification using the confussion matrix. In the classification evaluation process, that are four possibilities to occur from the process of classifying a row of data [13]. If the positive data and positive predictions will be counting as true positive, but if the data is predicted to be negative then it will be counted as false negative. If the data is negative and negative predictions will be counted as true negative, but if the data is positive predicted it would be counted as a false positive [14],[15]. The confussion matrix shown in Table 1.

Table 1. Confussion Matrix

Class	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

$$accuracy = \frac{\text{true positive} + \text{true negative}}{\text{true positive} + \text{false positive} + \text{true negative} + \text{false negative}} \text{ x } 100\% \quad (11)$$

The final steps are visualization and word association the positive and negative reviews. The research step can be simplified using a flowchart in Fig. 1.

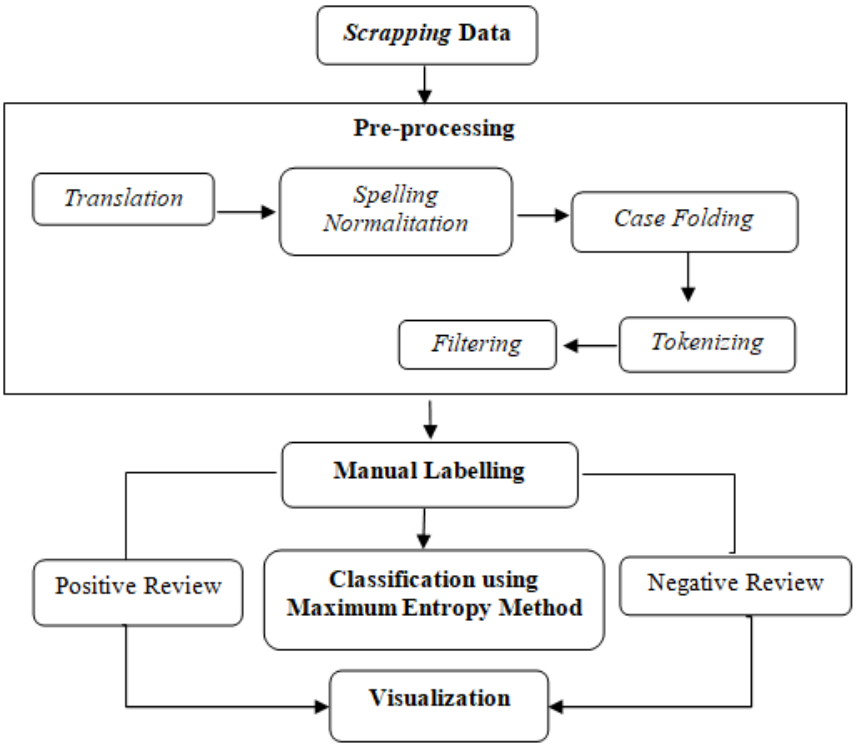

Fig. 1. Flowchart of research design

3. RESULTS AND DISCUSSION

Review data on Shopee application users on *Google Play* from 16th February to 8th March 2019 using scrapping technique with Data Miner obtained 1305 reviews. Then, do pre-process that starts with translation, spelling normalization, case-folding, tokenizing, and filtering. The example of pre-processing shows in Fig. 2.

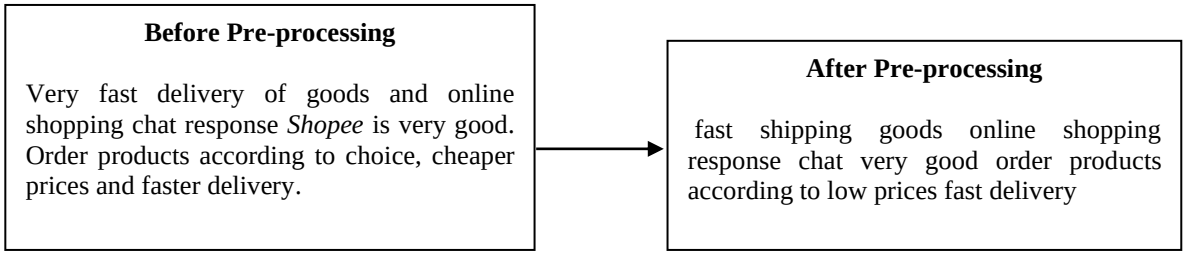

Fig. 2. Pre-processing result of Indonesian reviews

Fig. 2. it is known that input data Shopee application review data from Google Play as input data, then after pre processing so that results such as output data. Then perform manual labelling into positive or

negative. Review data of Shopee application users who entered the positive sentiment class amounted to 891, while those the negative sentiment class amounted to 414.

3.1. Accuracy testing of maximum entropy method

The accuracy test of the maximum entropy method using a comparison of training and testing data at 80%: 20% [15]. The training data in this research are 1044 reviews, while for the testing data amount to 261 reviews. The test parameters used is accuracy. The test parameters can be calculated using the confussion matrix. The result of the confussion matrix in this research shown in Table 2.

Table 2. Confussion Matrix

Class	Positive	Negative
Positive	182	4
Negative	3	72

Table 2, it is known that the total data is 261, where there are 254 review data classified into the appropriate sentiment class, and there are 7 review data that have not classified into the appropriate sentiment class. The accuracy has achived of 97.32% using (11).

3.2. Visualization and associations of positive reviews

Extraction of information on positive reviews repeated. So, that information about the positive reviews of Shopee application users is most often reviewed. A positive review can identify based on word frequency of words in the review. The visualization of information extraction result obtained from positive reviews of users as follows:

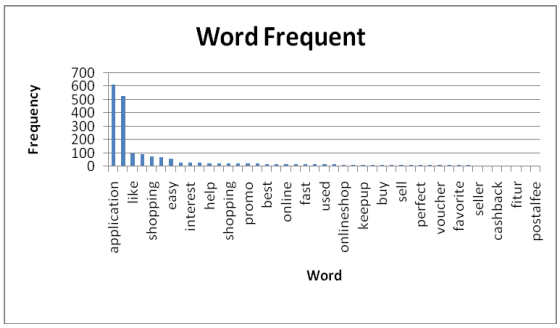

Fig. 3. Most frequent word in positive

Fig. 3. shows some of the most frequently occurring words, namely the word *"*application*"* 627 times, the word" like" 523 times. The words that appear in Fig. 3. have positive sentiments and also are the topic of positive reviews that are most widely reviewed by Shopee application users. Then, the association words are shown in Table 3.

Table 3. Association Positive Word

Application		Discount		Satisfy		Promo	
Word	**Value *r***	**Word**	**Value *r***	**Word**	**Value *r***	**Word**	**Value *r***
Good	0,49	Big	0,33	Humble	0,38	Interest	0,73
Nice	0,17	Super	0,33	Security	0,27		
				Customer service	0,23		

Based on Table. 3, we obtained several word associations in the classification of positive sentiment. Table 3, the word associations related to the word "application", there is information obtained that many users who access the Shopee application are overall good and great. Then, the word "discount" also give information on how the Shopee application provides large discounts to users. After that, the word "satisfy" informs that the friendliness and response of the seller and the customer service in the Shopee application are very satisfying for the users and the security in shopping also satisfying for users. After that, the word "promo" give information that the promo held by Shopee is very attractive for users, so Shopee application users are interested in shopping at the Shopee application.

3.3. Visualization and associations of negative reviews

The visualization of information extraction obtained from negative reviews of Shopee application users as follows in Fig. 4.

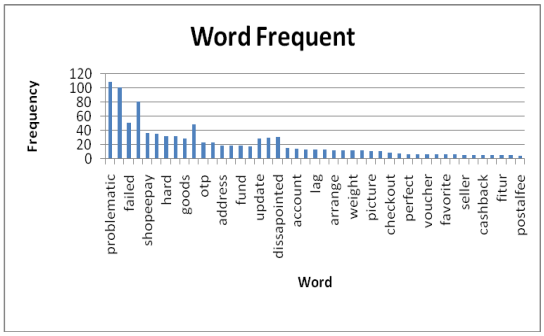

Fig. 4. Most frequent word in negative reviews

Fig. 4. shows some of the most frequently occurring words, namely the word *"*problematic*"* 127 times, the word application*"* 103 times, the word *"*failed*"* 5 times and so on. Same as positive reviews, word frequent from negative reviews also can be used to find the word association obtained in Table 4.

Table 4. Association Negative Word

Problematic		Old		Login		Verification		Expensive	
Word	**Value *r***	**Word**	**Value *r***	**Word**	**Value *r***	**Word**	**Value *r***	**Word**	**Value *r***
Code	0,33	Loading	0, 64	Android	0,43	Photo	0, 21	Postal fee	0,76
Post	0,26	Picture	0, 32	Web	0, 43	Shopee pay	0,29		
OTP	0,22	Response	0,25	OTP	0, 42	Fund	0,27		
		Application	0,24	Hard	0.41	Failed	0,26		
		Fixed	0,18			Withdrawal	0,24		

Based on Table 4, word associations related to the word "problematic", information obtained that there are many complaints from Shopee application users regarding the delivery of problematic OTP codes and problematic and difficult postal code settings. If seen in Table 3, word associations with the word "old"give information that users complain about things related to the process take a long time of loading the image. Users also complain about the old application loading and response processes. Regarding the complaint, users also asked Shopee to improve the system. Then, the word "login" provide information that users have difficulty logging in to the Shopee application via android or website. Users also have difficulty logging in because of the problematic OTP code. After that, the word "expensive" give information that user complains about expensive shipping or postage costs.

4. CONCLUSION

Shopee application users discuss matters related to "application", "promo", "satisfy", and "discount" for positive sentiments. Meanwhile for negative sentiments, users talk about "problematic", "login", "old", "verification", and "expensive" and the company can evaluate the application's performance regarding this matter. The accuracy from maximum entropy method is 97.32%. The results of the accuracy show that the maximum entropy method is excellent to use sentiment analysis on Shopee application reviews. For further research can be used other methods such as naive Bayes classifiers or super vector machine.

REFERENCES

[1] S. Anwar Hridoy, M. Ekram, M. Islam et al., "Localized Twitter Opinion Mining using Sentiment Analysis," *Decision Analytic*, 2015 available at: Google Scholar.
[2] N. Mehra, S. Khandelwal, P. Patel, "Sentiment Identification using Maximum Entropy Analysis of Movie Reviews," *International Journal of Computer Applications*, 2016 available at: Google Scholar.
[3] B. Wagh, J. Shinde, P. Kale, "A Twitter Sentiment Analysis using NLTK and Machine Learning Techniques," *International Journal of Emerging Research in Management and Technology.*, vol. 6, pp. 37, 2018 available at: Google Scholar.
[4] W. Medhat, A. Hassan, and H.Korashy, "Sentiment Analysis and Applications: A Survey," *AIN Shams Engineering Journal.*, vol. 5, pp. 1093-1113, 2014, doi: 10.1016/j.asej.2014.04.011.
[5] L. Zheng, H.Wang, and S.Gao, "Sentimental Feature Selection for Sentiment Analysis pf Chinese Online Reviews," *International Journal of Machine Learning and Cybernetics.*, vol.9, pp. 75-84, 2018 available at: Google Scholar.

[6] H. Htet, and Y. Myint, "Social Media (Twitter) Data Analysis using Maximum Entropy Classifier on Big Data Processing Framework (Case Study : Analysis of Health Condition, Education Status, State of Business)," *Journal of Pharmacognosy and Phytochemistry.*, vol. 7, pp. 695-700, 2018 available at: Google.
[7] A. Gupte, S. Joshi, P. Gadhul, *and* A. Kadam, "Comparative Study of Classification Algorithm used in Sentiment Analysis," *International Journal of Computer Science and Information Technologies.*, vol. 5, pp. 6261-6264, 2014 available at: Google Scholar
[8] D. Patel, S. Saxena, and T.Verma, "Sentiment Analysis using Maximum Entropy Algorithm in Big Data," *International Journal of Innovative Research in Science, Engineering and Technology*., vol. 5, pp. 8356-8361, 2016 available at: Google Scholar.
[9] P. Milev, "Conceptual Approach for Development of Web Scraping Application for Tracking Information," *Economic Articles.*, issue 3, pp. 475-485, 2017 available at: Google Scholar.
[10] T. Agrawal, and A.Singhal, "An Efficient Knowledge-Based Text Pre-processing Approach for Twitter and Google+", *Springer.*, pp. 379-389, ISBN 978-981-13-9942, 2019 available at: Google Scholar.
[11] A. B. Panwar, M. A. Jawali, and D. N. Kyatanavar, "Fundamentals of Sentiment Analysis: Concepts and Methodology," *Sentiment Analysis and Ontology Engineering.*, pp. 25-48, ISBN 978-3-319-30317-8, 2016 available at: Google Scholar.
[12] X. Xie, S.Ge, F.Hu, M.Xie, A and N. Jiang, "An Improved Algorithm for Sentiment Analysis Based on Maximum Entropy," *Soft Computing*, vol.23, pp. 599-611, 2017 available at: Google Scholar.
[13] Z. Zhao, T. Liu, B. Li, and X. Bu, "Correction to: Guiding the Training of Distributed text Representation With Supervised Weighting Scheme for Sentiment Analysis." *Data Science and Engineering*, vol. 2, Issue. 4, 2017, doi: 10.1007/s41019-017-0040-6.
[14] C. C. Aggrawal, Data Mining*:* The Textbook, Swiss: Springer International Publishing Switzerland, 2015, available at: Google.
[15] S. Suthaharan, "Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning," *Springer.*, pp.10, 2015 available at: Google Scholar.