Sentiment Analysis Of Traveloka App Using Naïve Bayes Classifier Method

Ronal Watrianthos, Sudi Suryadi, Deci Irmayani, Marnis Nasution, Elida F. S. Simanjorang

Abstract: Traveloka is currently the most popular startup in Indonesia with share traffic reaching 78.49% using smartphone and monthly visits which reached 28.92 million based on a report in similarweb.com in May 2019. Traveloka, based on record, has been downloaded 10 million times since 2014 with rating reaches 4.4 out of 5 stars. As of May 2019, there were 386,646 reviews from users in the PlayStore, ranging from positive and negative reviews. However, it is necessary to analyze with certain methods to summarize the review. Every review given will get a conclusion after collected, and sentiment analysis will provide user experiences from the Traveloka application within certain period. This research was conducted using the Naïve Bayes Classifier method based on a review from the playstore to determine service quality. The purpose of this study is to find out the perceptions of users based on the measurement of service quality so that the results can be an evaluation for Traveloka in improving services. Studies show that during this period public opinion produced negative sentiments with Vmap value of 0.31020 greater than positive sentiment with a value of 0.16132.

Index Terms: App Store, Feedback, Machine Learning, Naïve Bayes Classifier, Sentiment Analysis,Text Mining, Vmap.

1 INTRODUCTION

The growth of internet users in Indonesia is currently happening so fast, data in January 2018 states that internet users have reached 132.7 million people, reaching 50% of Indonesia's population, which is up 23% from data in 2017[1][2]. This triggered the development of online commerce or e-commerce transactions in Indonesia to grow more rapidly. Google and Temasek's research states that Indonesia's e-commerce Gross Merchandise Value (GMV) in 2018 has reached US $ 12.2 billion and is predicted to increase US $ 53 billion in 2025[3]. Even Indonesia's e-commerce transactions were the highest in Southeast Asia, followed by Vietnam, Thailand, Malaysia, the Philippines and Singapore[3][4] The increase in e-commerce transactions has resulted in the birth of many startup companies such as Traveloka. Although ranked third in the first quarter of 2019 based on startupranking.com[5], it ranked first in Indonesia for startups specifically providing online tour and travel services with scores of 86.45 under Blibli 87.17 and Bukalapak 87.5. Traveloka is currently the most popular startup in Indonesia with share traffic reaching 78.49% using smartphone and monthly visits which reached 28.92 million based on the report similarweb.com, Mei 2019[6] Founded by Ferry Unardi in 2012, five years later Traveloka became a startup unicorn in Indonesia with a valuation of more than $ 1 billion[7]. Available on a variety of platforms such as websites on desktop computers, as well as in Playstore for Android and app stores for iOS smartphones, Traveloka listed that it has been downloaded 10 million times since 2014 with a rating of 4.4 out of 5 stars. As of May 2019, there were 386,646 reviews from users in the PlayStore, ranging from positive and negative reviews such as complaints, criticism, or suggestions. However, it is necessary to analyze with certain methods to summarize the review. Every review given will get a conclusion after collected, and through sentiment analysis it can be done so that user experience can be obtained from the Traveloka application for a certain period. The purpose of the sentiment analysis is other than to get sentiment information contained in a text[8], also to see the opinion of the text relating to a problem or object whether it tends to have a positive or negative outlook[9]. Sentiment analysis can be used as a tool to see the public response to a particular place, event, or phenomenon [10]. The way it works is by classifying whether there is polarization in the text in documents, sentences or opinions[11][12]so that future strategic steps can be taken. This classification of sentiment analysis can be done in three levels; the document level, sentence level, and aspect level[13][14] which in this study is carried out at the sentence level using the Naïve Bayes Classifier (NBC) method. This Naïve Bayes Classifier method is one of the machine learning methods that is a simple probability-based technique for sentiment analysis[11][15]. This method will be used in the sentiment analysis of the Traveloka application based on reviews from the playstore to determine service quality. The purpose of this study is to find out the perceptions of users based on the measurement of service quality so that the results can be an evaluation for Traveloka in improving services.

2 METHODOLOGY

2.1 Process of Sentiment Analysis

In sentiment analysis, the process will begin with identification of data. At this stage the data collection uses data crawling techniques and then labeling is carried to determine the sentiment of the comments obtained. The next stage is preprocessing data to select data and convert data into structured data which consists of 4 stages; cleaning, removing stopword, tokenization, and stemming. After preprocessing the data, then extracting the feature is done where features are made to facilitate the occurrence of machine learning processes through the Naïve Bayes Classifier method[8]. Then at the last stage the classification process is carried out based on the sentiments in the document. All of these stages will later produce a model that will be used to show the accuracy of the classification results[16]. This model will then be tested to measure the classification performance value that has been done. The results of this model then produce information sentiment.

2.2 Naïve Bayes Classifier

Naïve Bayes Classifier is a classification technique based on the Bayes theorem discovered by British scientist Thomas Bayes. This theorem works based on the independent assumptions of predictors so that future opportunities can be identified based on previous experience. In the Naïve Bayes algorithm, a number of clues called attributes are needed to assist in forming the appropriate class for the sample being analyzed[17], where the presence of certain features in the class is not related to other features. There are several steps that must be taken in the Naïve Bayes algorithm, namely feature formation, probability calculation for each class

sentiment, and calculation of the maximum value of Vmap. In Naïve Bayes, Vmap is a calculation to determine the probability of testing data from each class based on learning processes[18]. In determining the value of the Naïve Bayes equation, formula (1) can be used as follows:

$$P(C|F1,\dots,Fn) = \frac{p(C)p(F1,\dots,Fn|C)}{P(F1,\dots Fn)} \quad (1)$$

After feature formation, the process is continued by calculating the probability of each class, which is done in equation (2):

$$p(ci) = \frac{fd(Ci)}{|D|} \quad (2)$$

Description of equation (2):
fd (ci) = the number of documents in the class
ci | D = the amount of training data

When the probability of each class has been obtained, then the probability calculation of each feature in the sentiment class is done with equation (3) below:

$$p(wk|ci) = \frac{f(wki,ci)+i}{f(ci)+|W|} \quad (3)$$

Description of equation (3):
f (wk, ci) = value of occurrence of the word wk in ci class
f (ci) = total number of words that appear in ci class
|W|= total number of wk

3 RESULT AND DISCUSSION

3.1 Data Input and Preprocessing

The data input used in this study was taken from comments that appeared in the Traveloka application in the PlayStore from April to May 2019. Before the data classification process using Naïve Bayes Classifier is carried out, the data will be processed first at the pre-processing stage such as case folding, tokenizing, stemming, and filtering. This pre-process is needed so that during the classification stage the results obtained can be more optimal.

3.2 Formation of Data Train

After the pre-process is done, the training data can then be formed by providing keywords for each opinion. Each keyword will be given a value of zero (0) for neutral, one (1) for positive, and minus 1 (-1) for negative. Table 1 shows the results of the training data that have been obtained at this stage.

Table 1. Train Data Formation

Data	Emergence	Sentiment Class
T1	Tested [1]	Positif
T2	Helping [1]	Positif
T3	Complicated [-1]	Negatif
T4	Easy [1]	Positif
T5	Dissapointed [-1]	Negatif

After obtaining training data, the next step is to calculate the probability of each class whose results can be seen in table 2.

Table 2. Probability Class Training

Sentiment Class (c)	Data (j)					fd(cj)	p(cj)
	T1	T2	T3	T4	T5		
Positive	1	1	0	1	0	3	3/5
Negative	0	0	1	0	1	2	2/5

3.3 Determining the Probability of Data Train

After getting the probability of each class, then the next step determines the probability of each training data from each class so that the probability of excercise data is obtained, as shown in table 3 below.

Table 3. Probability Exercise Data

Data f (Wki, Ci)	Sentiment Class	
	Positive	Negative
Tested	$\frac{1+1}{3+5}=\frac{2}{8}$	$\frac{0+1}{2+5}=\frac{1}{7}$
Helping	$\frac{1+1}{3+5}=\frac{2}{8}$	$\frac{0+1}{2+5}=\frac{1}{7}$
Complicated	$\frac{0+1}{3+5}=\frac{1}{8}$	$\frac{1+1}{2+5}=\frac{2}{7}$
Easy	$\frac{1+1}{3+5}=\frac{2}{8}$	$\frac{0+1}{2+5}=\frac{1}{7}$
Dissapointed	$\frac{0+1}{3+5}=\frac{1}{8}$	$\frac{1+1}{2+5}=\frac{2}{7}$

3.4 Implementation of Naïve Bayes Classifier

After obtaining the train data, the next step continues to the classification process using the Naïve Bayes Classifier algorithm by testing the Vmap value. For this classification test data is needed as shown in table 4 below.

Table 4. Data Test

No	Data Test
1	Helps and makes it easier if I need airplane tickets and hotels, but a little complicated in use….

Table 5. Result of Preprocessing Data Test

Help	Easy	Complicated

The Vmap calculation will be used in the Naïve Bayes Classifier classification in determining the largest probability value based on learning processes in the test data for each class. The Vmap calculation can be done with the following formula:

$$Vmap = p(wk|c)xp(c) \quad (4)$$

a. Vmap for positive sentiment:
Vmap (positive) = P(positive) P(help|positive) P(easy|positive) P(complicated|positive) Vmap(positive)
$= \frac{3}{5}+\frac{2}{8}+\frac{2}{8}+\frac{1}{8} = 0{,}16132$

b. Vmap for negative sentiment:
Vmap (negative) = P(negative) P(help|negative) P(easy| negative) P(complicated|negative) Vmap(negative)
$= \frac{2}{5}+\frac{1}{7}+\frac{1}{7}+\frac{2}{7} = 0{,}31020$

Vmap calculation found that based on the category, the test data has a negative sentiment because it has a greater value than positive sentiment which is equal to 0.31020. It can be concluded, public opinion on the Traveloka application in the

playstore is classified as having negative sentiments.

4 Conclusions

This research shows how public opinion on the Traveloka application in the playstore in the May 2019 period. The Naïve Bayes Classifier method is used in the data train to produce probabilities from each of the different class criteria which are then optimized to produce sentiment analysis. Studies show that during this period public opinion produced negative sentiments with Vmap values of 0.31020 greater than positive sentiments with a value of 0.16132. These negative results were obtained regarding aspects of application usage and consumers who were disappointed with the expensive ticket prices. For further research, it is expected to apply different algorithms to get accurate results in assessing public sentiment. should be referenced in the body of the paper.

References

[1] We Are Social, "Indonesia Digital Landscape," 2018.

[2] R. Watrianthos, I. R. Munthe, and R. Muti, "Motives Influencing Facebook Usage by Students and Its Relationship with Facebook Addiction Disorder," Int. J. Sci. Technol. Res., vol. 8, no. 3, pp. 2–3, 2019.

[3] Google and TEMASEK, "e-Conomy SEA 2018 Southeast Asia's internet economy hits an inflection point," pp. 1–32, 2018.

[4] katadata.co.id, "Transaksi e-Commerce Asia Tenggara Diproyeksi Mencapai Rp 1.469 Triliun pada 2025," katadata.co.id, 2018. [Online]. Available: https://databoks.katadata.co.id/datapublish/2018/11/30/trasaksi-e-commerce-asia-tenggara-diproyeksi-mencapai-rp-1469-triliun-pada-2025.

[5] Startupranking.com, "Startup Ranking," Startupranking.com, 2018. [Online]. Available: https://www.startupranking.com/countries.

[6] Similarweb, "Website Analysis Overview Report May 2019 traveloka.com," Similarweb, 2019. [Online]. Available: https://www.similarweb.com/website/traveloka.com.

[7] Liputan6, "4 Starup Unicorn Indonesia," Liputan6, 2019. [Online]. Available: https://www.liputan6.com/news/read/3898189/4-startup-unicorn-di-indonesia.

[8] M. O. Pratama et al., "The sentiment analysis of Indonesia commuter line using machine learning based on twitter data," J. Phys. Conf. Ser., vol. 1193, no. 1, 2019.

[9] A. Razia Sulthana, A. K. Jaithunbi, and L. Sai Ramesh, "Sentiment analysis in twitter data using data analytic techniques for predictive modelling," J. Phys. Conf. Ser., vol. 1000, no. 1, 2018.

[10] Y. Nurdiansyah, S. Bukhori, and R. Hidayat, "Sentiment analysis system for movie review in Bahasa Indonesia using naive bayes classifier method," J. Phys. Conf. Ser., vol. 1008, no. 1, 2018.

[11] S. Wahyu Handani, D. Intan Surya Saputra, Hasirun, R. Mega Arino, and G. Fiza Asyrofi Ramadhan, "Sentiment analysis for go-jek on google play store," J. Phys. Conf. Ser., vol. 1196, no. 1, 2019.

[12] J. Ling, T. B. Oka, and I. P. E. N. Kencana, "Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square," E-Jurnal Mat., vol. 3, no. 3, pp. 92–99, 2014.

[13] S. Aji et al., "Review Sentiment Analysis of World Class Hotel Using Naive Bayes Classifier And Particle Swarm Optimization Method," 2019.

[14] S. K and F. F, "Survey on aspect-level sentiment analysis," IEEE Trans. Knowl. Data Eng., vol. 28, no. 3, pp. 813–830, 2016.

[15] T. INUI and M. OKUMURA, "A Survey of Sentiment Analysis," J. Nat. Lang. Process., vol. 13, no. 3, pp. 201–241, 2006.

[16] F. C. Permana, Y. Rosmansyah, and A. S. Abdullah, "Naive Bayes as opinion classifier to evaluate students satisfaction based on student sentiment in Twitter Social Media," J. Phys. Conf. Ser., vol. 893, no. 1, 2017.

[17] E. U. Artha et al., "Klasifikasi Model Percakapan Twitter Mengenai Ujian Nasional," vol. 03, no. 01, pp. 121–125, 2018.

[18] R. Zhen, Y. Jin, Q. Hu, Z. Shao, and N. Nikitakos, "Maritime Anomaly Detection within Coastal Waters Based on Vessel Trajectory Clustering and Naïve Bayes Classifier," J. Navig., vol. 70, no. 3, pp. 648–670, 2017.