

Sentiment Analysis of Visitors in Borobudur Temple With Support Vector Machine Based on TripAdvisor Review

Edy Widodo^{1,*}, Gifa Delyani Nursyafitri¹, Ocktavia Nurima Putri¹, Yesih Nurmallasari¹

¹Department of Statistics, Universitas Islam Indonesia, Yogyakarta

*Correspondence: edywidodo@uii.ac.id

ABSTRACT

The number of online sites in Indonesia has recently experienced significant growth, both in tourism and other sectors. Borobudur Temple is one of the most well-preserved tourist attractions that is frequently visited by domestic as well as foreign tourist. TripAdvisor is a platform of an online forum that provides anyone to share information related to travel where enormous reviews in daily are posted. Tourists can read the accumulated opinions as a consideration for their travel plans. In this paper, we aim to analyze the English reviews with sentiment analysis by identifying and categorizing opinions as generally positive, negative, or neutral. Data used in this paper is a set of online reviews from TripAdvisor from December 29, 2017, to May 12, 2019. Data were analyzed using Support Vector Machine (SVM) classification methods. The result of each category shows that there are 305 positive reviews, 60 neutral reviews and the rest 36 are negative reviews. The accuracy performed by the SVM model is 100% that has been tested with 401 testing data and trained by 247 data.

Keywords: *TripAdvisor, Sentiment Analysis, Support Vector Machine.*

1. INTRODUCTION

According to ‘e-economy SEA 2018’ report by Google and Tematek, Indonesia has the largest and fastest-growing internet user in 2018. Various advantages for businesses are offered on the internet, such as for the tourism industry companies. It’s also caused by a lot of travel companies established online booking such as airplane tickets and hospitality through online platforms. Thus giving easy access to enormous information in terms of tourism, accommodation, and other sectors.

Borobudur Temple is one of the most well-preserved tourist attractions that is frequently visited by domestic as well as foreign tourist. This temple is located at Magelang, Central Java, Indonesia. This temple is listed as UNESCO World Heritage Site o. 592 and writer as the world’s largest Buddhist temple in the Guinness Book of World Records. **Figure 1.** below depict the growth of Borobudur Temple tourists from year to year, where the blue line defines foreign tourists, the orange line defines the local tourists and the purple line defines the total of both tourists. According to TRIBUNJOGJA.COM, the visitors amount of Borobudur Temple in 2018 increased after significantly dropped a year before. It is noted that there are 3,663,054 visits of domestic tourists and 192,231 foreign tourists.

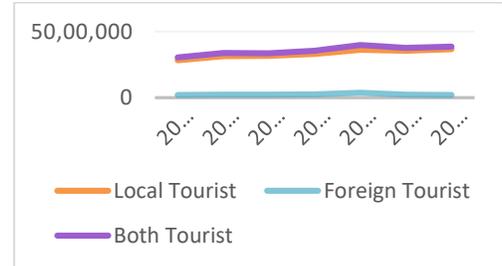


Figure 1. Borobudur Temple visitor diagram
Source: PT. TWC Borobudur, 2018

Customer satisfaction is the most crucial factor contributing to sustain customer growth and loyalty. TripAdvisor is a website with a concept where a tourist relies on the other tourist’s review to help them with their travel choice and planning.

TripAdvisor provides information for tourists to help them decide some travel-related content such as hospitality, flights, attractions, and restaurants with over 570 million reviews and opinion of 7,3 million accommodation, flights, attractions, and restaurants. TripAdvisor also let tourists to compare prices from more than 200 booking and official hotel websites. TripAdvisor is available in 49 markets with more than 455 million unique visitors every month that come to TripAdvisor. As it has become the largest travel community, the number of reviews on TripAdvisor keeps growing as the time goes

by that it will be hard for companies to keep track of the reviews.

Sentiment analysis is a process to determine whether an attitude from a text toward a particular topic is generally positive, negative or neutral. Information technology is currently providing a chance to develop a system in understanding and giving mood evaluation of an opinion. Mood analysis on an opinion called sentiment analysis that is referring to an automatic analysis of evaluative text with a focus to text classification based on its polarity. The impact and advantages of sentiment analysis make a significant growth to researches or application regarding sentiment analysis. There are around 20-30 companies using sentiment analysis to discover additional information about the public's sentiment toward the company.

There are a lot of research works related to sentiment analysis. Review data classification based on the sentiment category is necessary for the company to extract the insight of the large amount of data from TripAdvisor. The classification process of review data is performed with Support Vector Machine (SVM) methods by machine learning. SVM is a machine learning method that implements the Structural Risk Minimization (SRM) to find the ideal hyperlane that divides two classes in the input space. Prior researches have shown that SVM methods provides a good performance in classifying text document. SVM is one of a classifier with supervised learning methods that predict classes based on the pattern from the training process result. This method was invented by Vladimir Vapnik.

2. MATERIALS AND METHODS

This paper examined secondary sources of data from the TripAdvisor website. The object of this research is the review from Borobudur Temple visitors from December 29, 2017, to May 12, 2019. The flow of research is shown in **Figure 2.** below. Data is collected by scraping the visitors of Borobudur Temple review from TripAdvisor website in the form of text by using R. Then, word weighting is performed by measuring the frequency of a term in a document. The weighted value will define the sentiment label as positive, negative or neutral. The

classification process is then executed with the Support Vector Machine (SVM) approach by learning the data pattern with training data. Based on the specific learning result of the SVM algorithm, the accuracy of the model is adjusted with testing data. Then, the classification result is visualized with the word cloud.

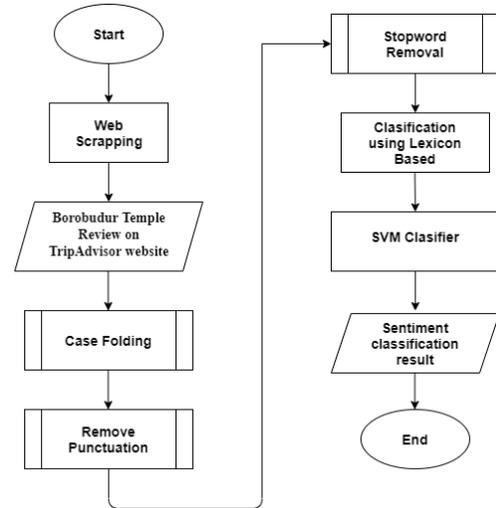


Figure 2. Research flowchart

3. RESULTS AND DISCUSSION

648 reviews are collected from scraped Borobudur Temple Review on TripAdvisor website from December 29, 2017, to May 12, 2019. Data labeling process is generated automatically through scoring the sentiment using the lexicon dictionary. Sentiment analysis is commonly performed to classify the text document into three sentiment classes, such as positive, negative and neutral. Sentiment class is defined by subtracting the positive word score with the negative word score. Based on the result of data labeling of the first five reviews in **Table 1.**, it shows that the sentences with score > 0 will be classified into positive, sentences with score < 0 will be classified into negative and sentences with score equal to 0 will be classified as neutral.

Table 1. Sample of data labeling

No	Classification	Score	Text
1	Positive	1	If you plan to visit this temple I advise to stay around the area...
2	Positive	1	This is a must see place in yogyakarta and if you're already in the island of java...
3	Positive	1	You've seen it on postcards. You've watched it on screen. Your friends posted it on social media...
4	Positive	5	We went for the sunrise tour and engaged the services of a tour guide which turned out to be a smart...
5	Neutral	0	This is a must visit for anyone in Indonesia. It is a very large...

Meanwhile, in **Table 2.** the total of each positive, negative and neutral label from all the observation are shown. There are 486 observations defined as positive, 65 negative observations and the rest 97 are neutral.

Table 2. Total of each label

Class	Observation
Positive	486
Negative	65
Neutral	97
Total	648

Training data is used by the algorithm classification to form a model classifier. This model is a representation of knowledge that will be used to predict a new class of data. The machine will understand the data pattern and predict a classification better as the data training is bigger. Meanwhile, testing data is used to measure how far the classifier managed to make classifications correctly. Data that is used for training and testing is the labeled data performed by using slovin formula with 5% level of significance.

$$n = \frac{648}{1 + 648 \cdot 0.05^2} = 247.32 \approx 247$$

Table 3. shows the comparison of testing data and training data. There are 181 training data and 305 testing data belong to the positive class, 29 training data and 36 testing data belong to the negative class and 37 training data, also 60 testing data are neutral.

Table 3. Train and test data of each classes

Class	Train Data	Test Data
Positive	181	305
Negative	29	36
Neutral	37	60

After the data getting analyzed with the SVM method approach, the classification result is shown in **Table 4.**

Table 4. SVM classification result

Support Vector Machine model			
	Positive	Negative	Neutral
Positive	305	0	0
Negative	0	36	0
Neutral	0	0	60
Accuracy	100%		

From the table above it can be seen that the SVM model classified positive sentiment of Borobudur Temple visitors as many as 305 review data, 36 data is correctly classified as negative and 60 others are neutral. The results of classifications stated that the SVM model obtains 100% accuracy. The most common words from the review data can be visualized by word cloud. Based on **Figure 3.** it is known that the “sunrise” word has the biggest shape that means the “sunrise” word has the highest frequency in positive class.

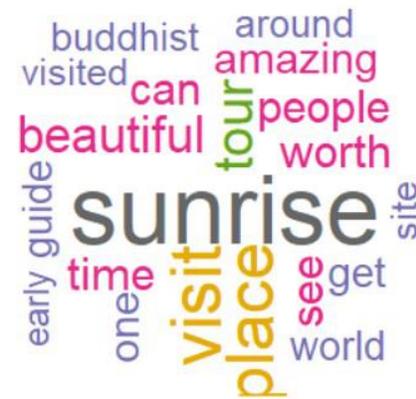


Figure 3. Positive class word cloud

The amount of 20 most common words based on Borobudur Temple visitor review on TripAdvisor is shown in **Table 5.** and it is proven that the “sunrise” word has the most frequency, followed by visit, place, tour, beautiful, worth and others.

Table 5. Positive words frequency

No	Word	Frequency	No	Word	Frequency
1.	Sunrise	193	11.	Amazing	74
2.	Visit	135	12.	One	71
3.	Place	127	13.	Get	70
4.	Tour	98	14.	World	66
5.	Beautiful	90	15.	Guide	65
6.	Worth	80	16.	Around	63
7.	Time	80	17.	Buddhist	63
8.	See	76	18.	Site	57
9.	People	76	19.	Visited	57
10.	Can	75	20.	Early	56

Meanwhile, in **Figure 4.** below, the negative class word cloud shows that the “expensive” and “ticket” word has the most frequency among the other. It may represent the ticket price of the Borobudur Temple that can be considered as expensive.



Figure 4. Negative class word cloud

Table 6. is the result of the top 20 words in negative class with the most amount of frequency. It can be seen that “expensive” and “ticket” word has the most frequency than the other with each frequency is 10.

Table 6. Negative words frequency

No	Word	Frequency	No	Word	Frequency
1.	Expensive	10	11.	Chained	6
2.	Ticket	10	12.	Hours	6
3.	Buddhist	7	13.	Site	6
4.	Cloudy	7	14.	Really	6
5.	Complex	7	15.	Crowded	6
6.	Sun	7	16.	Around	6
7.	Tourist	6	17.	Lot	6
8.	Hotel	6	18.	Early	6
9.	Tickets	6	19.	Feel	6
10.	Yogyakarta	6	20.	Went	6

4. CONCLUSION

Based on the research performed in this paper, it is concluded that the result of data scraping of Borobudur Temple visitor review from TripAdvisor website given 648 reviews with 305 positive sentiments, 60 neutral sentiments and the rest 36 are negative. The sentiment classification result shows that based on 247 training data and 401 testing data, SVM achieves good performance with 100% level of accuracy.

REFERENCES

- [1] Afifah, A. (2015). Respon Pelanggan Pada Situs Tripadvisor.Com Sebagai Bentuk Cyber Public Relations The Premiere Hotel Pekanbaru. Jom FISIP. 1-15.
- [2] Supratman, L. P. (2018). Penggunaan Media Sosial oleh Digital Native. Jurnal ILMU KOMUNIKASI . 47-60.
- [3] Hidayat, I. K., Sunarto, P., Guntur, T. (2014). Mengenal Relief, Mudra dan Stupa Candi Borobudur untuk Anak-Anak Usia 9-12 Tahun melalui Edugame. Journal of Visual Art and Design. 6(1): 58-68.
- [4] Angkiriwang, D. I., Susanto, K. C. (2018). Pengaruh Ulasan Online Di Tripadvisor Terhadap Minat Menginap Generasi Milenial Di Surabaya. Jurnal Hospitality dan Manajemen Jasa. 56-65.
- [5] Pang, B., Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval. 1–2.
- [6] Rozi, I. F., Pramono, S. H., Dahlan, E. A. (2013). Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. Jurnal EECCIS. 6(1): 37-43.
- [7] Feldman, J. S. R. (2007). The Text Mining Handbook. New York: Cambridge University Press.
- [8] Indrayuni, E. (2016). Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization. EVOLUSI: Jurnal Sains dan Manajemen. 4(2).
- [9] Saraswati, N. W. (2011). Text Mining dengan Metode Naïve Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis. Universitas Udayana Denpasar.