Aspect-Based Sentiment Analysis of Borobudur Temple Reviews Use Support Vector Machine Algorithm

Muhammad Resa Arif*, Pristi Sukmasetya, Rofi Abul Hasani and Maimunah

1 Department of Informatics Engineering, Universitas Muhammadiyah Magelang, Magelang, Indonesia

Abstract. As one of the most popular tourist attractions in Indonesia, Borobudur Temple is currently included in the top ten list of tourism priorities by the Ministry of Tourism. To increase the number of tourists, it is very important to maintain the quality of tourist attractions. Tourist growth is directly related to the number of online reviews of tourist attractions. Tourism managers need more than just reviewing good and negative sentiments to maintain and improve the quality of tourist attractions. Many aspects serve as benchmarks for visitors to come to a tourist spot, such as aspects of ticket prices, location, attractiveness, facilities, accessibility, visual image, and human resources. Therefore, sentiment analysis is needed for each of these aspects to find out aspects that need to be improved in order to increase the number of visitors. Support Vector Machine (SVM) is an algorithm used to categorize aspect-based sentiments. Analyzed using SVM, the dataset must first be cleaned and normalized through preprocessing. The results of the analysis show that the aspects of accessibility and visual image need to be improved to maintain and increase the number of visitors. This is because these two aspects have the most negative reviews compared to other aspects. The results of model testing only get an average accuracy value of 0.8148 because the distribution of data for all aspects and reviews is not balanced.

1 Introduction

The Indonesian government’s new policy regarding the repeal of the Community Activity Restrictions (PPKM) has allowed various sectors to grow and operate freely, one of which is the tourism sector. Various tourism managers are trying to improve the quality of tourist attractions because this is one of the important factors contributing to the high level of visitors [1]. Borobudur Temple is one of the tourist attractions that was included in the World Heritage List in 1991 and has attracted many visitors [2]. The sustainability of the number of tourist visitors is a challenge in the current era. One important factor on which tourism managers must focus is tourist loyalty. Tourist loyalty can be maintained through service
quality and tourist attractiveness as well as by increasing independent variables or certain factors that make visitors feel satisfied and want to come back [1].

The various determining factors or independent variables include safety factors, accessibility, tour duration, accommodation costs/entrance ticket prices, tour packages offered, attractiveness, facilities, season or climate at tourist attractions, tourist infrastructure, service quality, visual image, location, and human resources [2,3]. The number of determining factors to attract tourist visitors makes it difficult for managers to identify the advantages and disadvantages of each factor. Information technology is increasingly developing, and has a positive impact in the form of access and convenience for humans in various ways [4]. One of them is that now everyone can provide an assessment or review of tourist attractions that have been visited online.

All review data provided by each visitor does not have a certain meaning if no processing of the data is carried out. One method of text data processing that can be done is sentiment analysis. Previous research that discusses review data processing involves conducting a sentiment analysis of hotel reviews to find out the comparison of positive and negative sentiments from visitors who have stayed before [5,6]. Other research [6] analyzes reviews on culinary or restaurant businesses by looking for visitor satisfaction levels. The two studies [5,6] could only distinguish the number of positive and negative reviews. Other research [7] have developed sentiment analysis with results in the form of product information that is highly recommended and products that are not recommended.

Several previous studies on sentiment analysis were only able to find general results. Meanwhile, the assessment of positive and negative reviews is not limited to one aspect. The sentiment analysis results obtained do not provide specific information for stakeholders. Therefore, a more in-depth analysis of each aspect is needed as in the research [8]. This study analyzed hotel reviews on aspects of room cleanliness, amenities, bathroom conditions, prices, and food and beverages. The results obtained were that the most negative sentiment values were obtained from the food and beverage aspect. Stakeholders obtain information through the results of sentiment analysis, that the food and beverage aspects are the most urgent to improve so that there are always a lot of visitors.

Sentiment analysis can be completed through machine learning and deep learning-based approaches. Some very varied methods or algorithms can be used to solve sentiment analysis problems. Comparison of machine learning methods between logistic regression algorithms, random forest, and KNN on review text analysis on the BBC news portal [9]. The research focuses on accuracy parameters, of the three algorithms compared to logistic regression obtaining the highest accuracy value of 97%. Another research is comparing machine learning algorithms using Support Vector Machine (SVM) and Naïve Bayes [10]. The experiment was carried out on an unbalanced dataset, in which negative sentiment was more dominant than positive and neutral sentiment. The experimental results show that the SVM algorithm is superior to Naïve Bayes in terms of accuracy, precision, and recall.

Therefore, in this study, aspect-based sentiment analysis was carried out on the Borobudur Temple review using the SVM algorithm. This algorithm was chosen because, in previous studies, the SVM algorithm was superior in processing text data with high accuracy values. In addition, a smaller number of datasets is also more appropriate if the data analysis process is carried out using machine learning algorithms as opposed to deep learning, which requires a large number of datasets [11]. Some of the aspects analyzed in this study include aspects of ticket prices, location, attractiveness, facilities, accessibility, visual image, and human resources. It is hoped that the results of the sentiment analysis of the review of Borobudur Temple on six aspects can assist Borobudur Temple managers in improving the quality of each aspect to maintain and increase the number of visitors.
2 Method

The dataset used in this study was obtained from internet sources that have been used in previous studies [14]. In this study, two destination locations were used, namely Borobudur Temple and Prambanan Temple, with a total of 2,122 data points. This study did not use all of this data but only took review data at the Borobudur Temple location, totaling 988 review data. The entire data is distributed into 6 aspects, namely attractiveness, amenities, accessibility, visual image, entrance ticket prices, and managers or human resources. The data obtained is unbalanced and tends to be dominant in certain aspects.

![Fig. 1. Sample datasets](image1)

One review data can contain more than one aspect as shown in Fig. 1. There are some data that are not normal such as abbreviated words, inappropriate spelling and the presence of several illegal characters. If the data is not normal, if the preprocessing stage is not carried out first, it will have an impact on reducing the level of accuracy in the machine learning model [12]. Therefore, in this study, several stages of data preprocessing were carried out before the data analysis process was carried out.

![Fig. 2. Research flow](image2)

The overall research stages are shown in Fig. 2. There are 7 main stages carried out in this study up to the testing stage. Explanation of the research flow in more detail as follows.

2.1 Data preprocessing

The first stage is data preprocessing, which aims to improve the quality of the dataset. Some of the stages in the data preprocessing are as follows:

a. Text cleansing

This stage aims to remove some illegal characters, decode ASCII characters if some text contains ASCII characters, and remove emoticons, punctuation marks, numbers, and incomplete links. According to previous research [13] The data collected came from Twitter; therefore, in this study, a process of removing mentions was carried out. Apart from that, this stage also normalizes paragraphs by deleting tabs, new lines, and back slices. The final treatment in the text cleansing process is to change all the words
b. Tokenizing
The next stage is to separate words in a sentence to facilitate the next process. The tokenizing process uses the NLTK library.

c. Stopwords
After the dataset is divided into text/words, the process of removing connecting words is carried out. This stage is useful for highlighting more specific features in a text by removing all connecting words that have no meaning. This stopword process uses a literary library because the dataset used in this study is in Indonesian.

d. Normalization
After all the connecting words are removed, the next step is to normalize. This stage is carried out because in the dataset used, some of the sentences contain non-standard words (slang words). The process of changing non-standard words into proper and standard words uses the help of a dictionary of Indonesian non-standard words obtained in the study [14].

e. Stemming
The last stage in the preprocessing process is to change each word that has been carried out in the following stages into its basic form. This process is carried out with the aim of facilitating the process of data analysis because, because the root words are given different affixes that were previously recognized as different, after the stemming process is carried out, these words will be considered the same because they have changed to basic word forms.

2.2 Label conversion

Each sentence in the dataset may have a positive (1), negative (-1), and neutral (0) sentiment value and not have any sentiment value for an aspect ("-"). So if a sentence only relates to 4 aspects then the other 2 aspects will be given the notation ".". Therefore, before being processed at the model training stage, it is necessary to carry out a conversion process where aspects that have a value of "," will be replaced with NaN, which is easily understood by the Python programming language. This stage is a continuation of the preprocessing process, namely changing the value indicators for each sentence in the dataset.

2.3 Count vectorizer model

After the sentence is changed into its basic word form at the stemming stage, the next step is to count the number of occurrences of a word in the sentence. This process will output an array consisting of word documents and the number of occurrences.

2.4 TF-IDF

This process is a continuation of the previous stage. Documents/words that have counted the number of occurrences in each sentence are then processed to calculate the weight of each word. TF-IDF is one of the methods used to determine the weight of words that affect the sentiment features and aspects of each sentence to increase the value of model accuracy [15].
2.5 Dataset split

The dataset in this study does not only consist of positive/negative/neutral sentiment information but also related aspects of each sentence. Therefore, to facilitate the process of data analysis, it is necessary to divide the dataset into two main parts, namely, datasets about sentiments and aspects. After the division process, each data is divided into a ratio of 80:20. This division is based on previous research which is the ideal ratio to obtain a better accuracy value [16].

2.6 Training data

The main stage in this research is the process of forming a machine learning model through the training process. The dataset that has been divided by the ratio of 80:20, which is 80% of the total dataset, will be used in this training process. This process is carried out six times, according to the number of aspects that have been previously divided at the dataset split stage. Each aspect will be trained so that 12 models are created, consisting of 6 aspect sentiment models and 6 positive, negative, or neutral sentiments for each aspect.

2.7 Model evaluation

The final step is to test the machine learning model that has been created through the data training process. The testing process is carried out using the confusion matrix method. This method is used to determine the performance of the model as a whole through several test parameters [19]. The way this method works is class-based testing by comparing the classification results with the original label, with 4 possible predictive results: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Several parameters that can be used as good or bad criteria for a model are as follows:

a. Accuracy
   Describes the level of accuracy of a model in classifying new data. The formula can be written as follows:
   \[ \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \]  (1)

b. Precision
   Describes the level of accuracy between the requested data and the results of the model's predictions or can be referred to as the ratio of true positive predictions to all data that has been successfully predicted positively.
   \[ \text{Precision} = \frac{TP}{TP + FP} \]  (2)

c. Recall
   The success rate of a model in retrieving information. The formula can be written as follows.
   \[ \text{Recall} = \frac{TP}{TP + FN} \]  (3)

d. F1-score
   This criterion is the midpoint when a trade-off occurs, where if the recall value is high, then the precision value tends to be low, and vice versa. As a result, the F1-score can be used to provide a more precise assessment of a model's recall and precision values. The greater the F1-score (closer to 1), the better the recall and precision values of the model, and vice versa. The formula can be written as follows:
\[
F1-score = 2 \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}
\]  

3 Result and discussion

The data used in this study amounted to 988 data which were divided into positive, negative, and neutral classes. The dataset obtained does not have a balanced composition. To make it easier to distinguish the distribution of data, the distribution of the dataset is pre-processed into 6 aspects along with their sentiment classes.

Fig. 3 shows that the dataset tends to be dominant in the attractiveness and visual image aspects. Fig. 3 depicts the dominance of positive sentiment in both aspects, in addition to the number of unbalanced aspect data. Even so, there is also a negative sentiment which is quite dominated by aspects of accessibility, visual imagery, and amenities. Of all the aspects analyzed, the aspect of attractiveness shows the best results, namely the high number of positive sentiments and the minimum number of negative sentiments.

![Distribution of datasets on aspects and sentiments](image)

Fig. 4 shows a visualization of the frequency of occurrence of words contained in the attractiveness aspect. These words become an important feature in terms of attractiveness. The larger the size of the letters in these words indicates that the frequency of occurrence of the word is high in the entire dataset.

![The dominance of the appearance of the word on the aspect of attractiveness](image)

Table 1 shows the results of the evaluation of the classification model for the six aspects. The accuracy value is not too high on the attractiveness aspect because the amount of data related to that aspect is the largest. However, the F1-score is a more appropriate assessment.
because it gets the highest value, namely 0.8653, which means that the model has good performance in recognizing new data through the information obtained in the training process. Meanwhile, the human resources aspect shows the highest accuracy value compared to other aspects. This is because the amount of data on these aspects is the least, resulting in the worst model performance when tested with new data only obtaining a value of 0 on the precision, recall and F1-score criteria, which means that the model is unable to find information back from new data based on existing knowledge. The average value of all aspect models formed is good enough with an accuracy value of 0.8148. The lack of data in several aspects resulted in the average F1-score being quite low. As a result, the solution to improving model performance must increase the dataset, particularly for aspects with a small number. The balance of the dataset in each aspect also affects the performance of the model.

Table 1. Evaluation results of the aspect classification model

<table>
<thead>
<tr>
<th>No.</th>
<th>Aspect</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Attractiveness</td>
<td>0.7626</td>
<td>0.7626</td>
<td>1.0000</td>
<td>0.8653</td>
</tr>
<tr>
<td>2.</td>
<td>Amenities</td>
<td>0.8535</td>
<td>1.0000</td>
<td>0.2564</td>
<td>0.4082</td>
</tr>
<tr>
<td>3.</td>
<td>Accessibility</td>
<td>0.8737</td>
<td>0.9090</td>
<td>0.2941</td>
<td>0.4444</td>
</tr>
<tr>
<td>4.</td>
<td>Visual Image</td>
<td>0.5909</td>
<td>0.5882</td>
<td>0.9009</td>
<td>0.7117</td>
</tr>
<tr>
<td>5.</td>
<td>Price</td>
<td>0.8989</td>
<td>1.0000</td>
<td>0.2000</td>
<td>0.3333</td>
</tr>
<tr>
<td>6.</td>
<td>Human Resources</td>
<td>0.9090</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td><strong>Average value</strong></td>
<td><strong>0.8148</strong></td>
<td><strong>0.7099</strong></td>
<td><strong>0.4419</strong></td>
<td><strong>0.4605</strong></td>
</tr>
</tbody>
</table>

Table 2. Sentiment model evaluation results for aspects

<table>
<thead>
<tr>
<th>No.</th>
<th>Aspect</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Attractiveness</td>
<td>0.8129</td>
<td>0.6608</td>
<td>0.8129</td>
<td>0.7290</td>
</tr>
<tr>
<td>2.</td>
<td>Amenities</td>
<td>0.8139</td>
<td>0.8269</td>
<td>0.8139</td>
<td>0.7605</td>
</tr>
<tr>
<td>3.</td>
<td>Accessibility</td>
<td>0.4848</td>
<td>0.2350</td>
<td>0.4848</td>
<td>0.3166</td>
</tr>
<tr>
<td>4.</td>
<td>Visual Image</td>
<td>0.7731</td>
<td>0.7830</td>
<td>0.7731</td>
<td>0.6820</td>
</tr>
<tr>
<td>5.</td>
<td>Price</td>
<td>0.5555</td>
<td>0.7874</td>
<td>0.5555</td>
<td>0.5007</td>
</tr>
<tr>
<td>6.</td>
<td>Human Resources</td>
<td>0.4000</td>
<td>0.1600</td>
<td>0.4000</td>
<td>0.2285</td>
</tr>
<tr>
<td></td>
<td><strong>Average value</strong></td>
<td><strong>0.6400</strong></td>
<td><strong>0.5755</strong></td>
<td><strong>0.6400</strong></td>
<td><strong>0.5362</strong></td>
</tr>
</tbody>
</table>

Table 2 shows the results of the evaluation of the sentiment model on different aspects. The distribution of data that has good positive/negative/neutral sentiments but is not balanced on all aspects results in an unfavorable accuracy value, namely only getting 0.6400. However, it is clear that the amenities aspect sentiment model receives the highest score across all test parameters. This is because the ratio of the number of positive, negative, and neutral sentiment data is almost balanced, as shown in Fig. 3. Therefore, an unbalanced dataset greatly affects the performance of the machine learning model.

The development of text analysis on aspects of an object obtains better results than in previous studies [6–8]. Through the application of aspect-based sentiment analysis, it is able to provide more specific and clear information compared to previous research, which only provided information on positive or negative tendencies in an object. This study produced more detailed information on several aspects that were considered to have a tendency of negative evaluations, so that it would be easier for stakeholders to map the aspects that had to be prioritized for improvement, so that there would still be a lot of visitors and it was hoped that this would increase. However, if viewed from the performance side of the model, it is not good with an accuracy value of no more than 90%. This is because the dataset used is unbalanced, so the model is less able to recognize several aspects where the amount of data is not large enough. Therefore, to improve this research in the future, pre-processing can be
done first if the dataset used is not balanced so that the accuracy of the model performance can be maintained as in the research [17].

4 Conclusion

From the aspect sentiment analysis at Borobudur Temple tourist attractions through visitor review data, it can be concluded that the attractiveness and visual image aspects are the ones most reviewed by visitors. Borobudur Temple has a high attractiveness for prospective visitors because the positive sentiments given by visitors are very high and inversely proportional to the number of negative assessments on this aspect, which is very small. Meanwhile, to maintain and increase the number of visitors, Borobudur Temple managers must pay attention to aspects of easy access to the location and its visual image. This is because these two aspects have the highest number of negative ratings compared to the other four aspects. The number of datasets that are not balanced between Aspect 1 and other aspects, or the review value, has a significant effect on the performance of the model. The average accuracy value of the aspect classification model is better, namely getting a value of 0.8148 compared to the aspect sentiment model, which only gets an average accuracy value of 0.6400. This is due to the extreme differences in the distribution of sentiment data. As for the distribution of aspect classes, even though it is also not balanced, the difference is not as significant as in the sentiment data. Therefore, to improve model performance, it is necessary to increase the amount of data, especially in small classes. Another alternative that can be done is if the dataset conditions are very large, the dataset can be treated at the preprocessing stage, namely by balancing the datasets between classes.

References

1. Hermawan, H.; Wijayanti, A. Loyalty on Ecotourism Analysed Using the Factors of Tourist Attraction, Safety, and Amenities, with Satisfaction as an Intervening Variable; 2019; Vol. 8;.


