

# Sentiment Analysis of Data on Google Maps Reviews Regarding Tourism on Keraton Kasepuhan Cirebon Using the Lexicon Based Method

Faisal Akbar<sup>1,2</sup>, Hadiyanto<sup>1</sup> and Catur Edi Widodo<sup>1</sup>

<sup>1</sup>Doctoral Program of Information System, School of Postgraduate Studies, Diponegoro University, Semarang, Indonesia

<sup>2</sup>Department of Informatics Engineering, Sekolah Tinggi Ilmu Komputer Poltek Cirebon, Indonesia

**Keywords:** Sentiment Analysis, Lexicon Based Method, Keraton Kasepuhan Cirebon.

**Abstract:** Sentiment analysis is needed to find out a person's opinion of a particular object, by identifying the sentiments expressed by that person, then classifying the polarity value. One method for conducting sentiment analysis is Lexicon Based. In this study, it aims to carry out sentiment analysis by implementing the Lexicon Based method so that it can analyze the polarity of tourist perceptions of tourism at the Kasepuhan Palace in Cirebon. The dataset collected through Google Maps Reviews is sorted based on the most recent responses or comments. The dataset is 1117 scraped data using Python. Then the data is compressed to be processed so that it becomes 501 data that can be used. The library used is Sastrawi as the data dictionary. Based on the results of sentiment analysis, information was obtained that around 70% gave positive responses, then around 20% gave neutral responses, while the remaining around 10% gave negative responses to tourism at the Kasepuhan Palace, Cirebon.

## 1 INTRODUCTION

Developments in the field of Information and Communication Technology are very rapid every year and have impacts that can be felt directly in human life in various fields of activity, both individually and in groups (in a company or organization). Textual information found on the internet is generally divided into 2 (two) types, namely facts and opinions. Facts are objective statements about objects and events in the world, while opinions are statements that are subjective in nature by reflecting people's sentiments or perceptions about an object or event in the world. When an individual or group wants to obtain public opinion regarding a product, image and service, they no longer need to carry out conventional surveys and in a discussion group which costs quite a lot. With the existence of internet media, through a website service that has the feature of being able to provide online responses to a certain object subjectively based on the assessment of each of these people, so that it can generate large amounts of data that can be utilized directly and openly. Through online media, everyone can express anything, including their opinion that they think about a certain thing or object.

With easy access to various data needed to support related fields in human life, tourism is no exception.

The need for one's perspective on a tourist attraction is very important, because to be able to respond to various global challenges, a tourist attraction needs to adapt so that it is not easily abandoned by tourists. Especially in historical tourism objects, where these tourist objects are usually a form of relic from ancient times. It takes a heavier struggle to be able to continue to preserve this tourist attraction.

The importance of preserving tourism in the area, because with the development of the tourism sector, it will be proportional to the development of the economy in the area. This need is the basis for the importance of being able to know the sentiments of all tourists who come to the Cirebon Kasepuhan Palace.

Sentiment analysis is the process of understanding, extracting, and processing textual data automatically to obtain sentiment information contained in an opinion sentence. Sentiment analysis is carried out to be able to see how one's opinion or tendency towards an object, whether the opinion has a positive tendency or even vice versa towards negative, and also may contain a neutral tendency. One of them raised in this study is to be able to identify tourist tendencies and their opinions on the Keraton Kasepuhan Cirebon tourist attraction. The magnitude of the influence and benefits of this sentiment analysis has caused research and applications based on sentiment analysis

to develop rapidly. Even in America there are around 20 to 30 companies that focus on sentiment analysis services (Go et al., 2009).

## 2 PROPOSED METHOD: LEXICON BASED METHOD FOR SENTIMENT ANALYSIS

Sentiment Analysis (SA) or Opinion Mining (OM) is a computational study of people’s opinions, attitudes, and emotions towards an object, meaning that this object can represent individuals, organizations, or companies. When viewed from its duties, Opinion Mining has the task of extracting and analyzing one’s opinion about a particular object, while Sentiment Analysis is to identify the sentiments expressed in a text, then analyze it. Therefore, the main purpose of Sentiment Analysis is to find opinions, identify the sentiments expressed, and then classify the polarity (positive, negative, or neutral). In general, there are 3 (three) stages in sentiment analysis as shown in Figure 1.

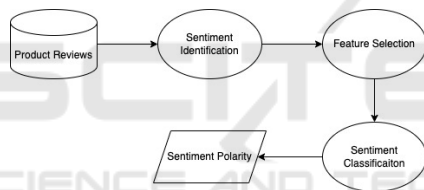


Figure 1: Stages of Sentiment Analysis (Qiu et al., 2009).

Product reviews are a dataset that contains a collection of responses or comments from many people on a particular object. For example, in this case the responses or comments from tourists are to find out their opinions on the Cirebon Kasepuhan Palace. Then at the sentiment identification stage is the stage where identification of all incoming comments or comments is carried out. In this stage, in general, it can be seen how the average tourist opinion of the Cirebon Kasepuhan Palace tourist attraction is. In the feature selection process it is used to select the features that will be used in the next process, sentiment classification. Tools or features that can be used at the feature selection stage include term frequency, part of speech tagging, dictionaries of words or phrases that contain opinions, and negated words (for example, not good that shows bad meaning) (Qiu et al., 2009).

The method used in this sentiment analysis is Lexicon Based, where there are 3 (three) approaches for the sentiment classification stage, namely the first approach by utilizing the use of Machine Learning, then carrying out Lexiconbased sentiment analysis, and the

last is the Hybrid Approach to incorporate Machine Learning. and Lexicon-based sentiment analysis.

The Lexicon-based approach is one of the methods when conducting sentiment analysis that utilizes a data dictionary which contains a list of words containing opinions, where each word in the dictionary has been given a polarity score by giving a value between -1 (for a negative class) to +1 (for a negative class). for the positive class). By using the Literary Library, developers can use the sentiment.polarity property to be able to find out the sentiment score for a word or sentence in Figure 2

	content	clean_text	word_length	sentiment
242	Good...	good	1	5
36	Heritage	heritage	1	0
187	Good	good	1	5
299	Top	top	1	5
178	Beautiful	beautiful	1	0

Figure 2: Sentiment Score Example.

## 3 DATA AND EXPERIMENTAL SETUP

The data used comes from Google Maps Reviews provided by users of the Cirebon Kasepuhan Palace tourism. The data collection method uses data scraping tools via Python, so that 1117 reviews are obtained based on the latest data provided by users. However, all of this data cannot be processed immediately because there are still other columns apart from the responses or comments provided by the user, so the next step is to eliminate unnecessary columns into just one column, namely the Caption column. From the results of this elimination, the remaining 501 data can be processed for sentiment analysis.

	content
0	Tempatnya nyaman dan juga luas, banyak ilmu se...
1	Nice keraton
2	Terdapat batu pasujudan di bagian selatan bagi...
3	Mengenal sejarah kesultanan keraton kasepuhan ...
4	Sejarah dan situsnya perlu dipelajari dan dile...
5	(Translated by Google) Good place (Original) ...
6	Kangen dengan muludan, utk tahun ini tepatnya ...
7	(Translated by Google) OK... (Original) Sip...
8	Seru lihat sejarah jaman dulu
9	(Translated by Google) Lots of cool photo spot...

Figure 3: Results of Scraping Data.

Based on Figure 3 is the result of the entire dataset

collection process, so that it becomes a ready-to-use dataset. Whereas in Figure 4 is a graphical map of word distribution based on the highest frequency after the process of removing punctuation marks, numbers, and cleaning sentences.

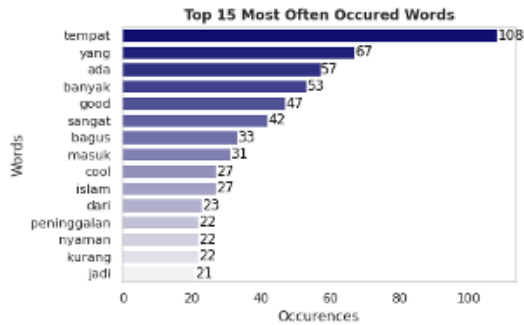


Figure 4: Most Word Frequency Graph.

Several stages were carried out in carrying out research to analyze sentiment as shown in Figure 5. Based on Figure 5, the initial step is to collect the datasets that have been described previously, namely obtaining a dataset of 501 data to be processed. Next, processing is carried out at the Preprocessing stage, where at this stage the entire text is cleaned so that the text is clean from noise. The Preprocessing stage is carried out in 4 (four) steps, as follows.

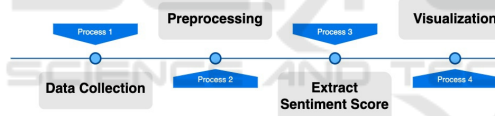


Figure 5: Sentiment Analysis Process Stages.

Comment selection, in this step comments are selected based on the latest posts, with the Google Maps Reviews function for sorting based on the most recent comments. Cleansing, The sentences obtained usually still contain noise, namely random errors or variances in the measured variable. Therefore, it is necessary to remove the noise. The omitted words are characters, icons, URLs, and so on (Azhar et al., 2013). Parsing, The parsing process is the process of breaking a document into words by analyzing a collection of words by separating the words and determining the syntactic structure of each of these words (Liu et al., 2005). Sentence Normalization, The purpose of this process is to normalize sentences so that non-standard sentences or typos become normal again according to KBBI rules, so that these sentences can be recognized as the correct language (Buntoro, 2017). What needs to be done in the process of normalizing sentences is as follows

1. Stretch punctuation and symbols other than the al-

phabet. The intention is to provide a distance for punctuation from the following or previous words, with the aim that the punctuation marks and symbols other than the alphabet do not become one with the words during the tokenization process.

2. Change to all lowercase.
3. Normalization of words with the normalization process rules, among others, can be seen in Figure 6.

Tidak Normal / gaul	Normal
Akhiran -ny	Akhiran -nya
Akhiran -nk	Akhiran -ng
Akhiran -x	Akhiran -nya
Akhiran -z	Akhiran -s
Akhiran -dh	Akhiran -t
Kata berulang: sama2	Kata berulang: sama-sama
Ejaan: oe	Huruf: u
Ejaan: dj	Huruf: j

Figure 6: Word Normalization Rules (Putranti and Winarko, 2014).

4. Eliminate repeated letters in a sentence. Usually a person can write a combination of letters to express his feelings in a sentence, but it is also possible that there are mistakes. For example, the word “good” is used to express people really like a certain thing. However, this word is not justified in KBBI, so it is necessary to remove repeated letters to become “good”.
5. Removing emoticons is the removal of facial expression icons that are embedded in a sentence but have no meaning in KBBI, usually this is done for people who want to give their facial expressions when conveying something. Some examples of feelings and sentiment emoticons can be seen in Figure 7.

Emoticon	Feeling	Sentiment
:) :-)	Happy	Positive
:( :-)	Sad	Negative
:D :-D	Very Happy!	Positive
D: D=	Very Sad	Negative
* * * * *	Fascinated	Positive
D:< D: D8	Horror, disgust, sadness	Negative
xD XD	Laughing, big grin	Positive
:  =  :-	Straight face no expression	Neutral

Figure 7: The Meaning of Emoticons.

Tokenization, after doing preprocessing until the process of normalizing the sentence, then the sentence is broken into tokens using a space delimiter. The token used in this study is the unigram, a token consisting of only one word. Part of Speech (POS) Tagger, POS Tagger is a process for giving a class to a word. In the POS tagger process it is done by parsing, then the class of each word is determined using

the help of a self-made dictionary based on KBBI using the Maximum Entropy method. The POS tagging process is divided into three processes, namely separating each token in the document by checking each word in the document, then identifying each word in the document by providing the type of word, checking the words that have not been identified for the form of affixes and suffixes so that basic words are obtained. Based on the linguistic rules on the word temporary sentiment is obtained (Saputra et al., 2021; Buntoro et al., 2014; Nafan and Amalia, 2019). Sentiment determination is done by looking at the presence of words that contain sentiments that have positive or negative polarity from comments that have been labeled as word class. The word classes chosen are adjectives, adverbs, nouns, and verbs, in accordance with previous research references that these four types of words are the types of words that contain the most sentiments. In this system, if a comment contains a noun (NN) before or after the adjective (JJ) or adverb (RB) and the noun (has opposite polarity to the adjective or adverb), the polarity obtained is based on the adjective or adverb. adverb, because adjectives or adverbs give affirmation to nouns (Putro, 2011). Load Dictionary, after preprocessing and tokenization, the next step is to carry out a Load Dictionary, the purpose of which is to determine the type of data dictionary used in this study. For example, a dictionary with positive, negative, negative sentiments, as well as a dictionary of normalized language abbreviations like the following.

1. Positive: good, great, cool, excellent, etc.
2. Negative: ugly, bad, evil, etc.
3. Negation: no, not, away, etc.
4. Abbreviation language conversion: brp=how much, sp=who, spt=like, etc

Extract Sentiment Score, The results of all previous processes that have been carried out are in the form of a collection of adjectives, adverbs, nouns, and verbs. For each of these words, the sentiment value is then extracted using the Lexicon Based method. In this case, the extraction utilizes the sentiment score in the Literary Library. Determination of thresholds for positive, negative, and neutral labels is shown in the following algorithm.

```
def sentimentlabels(senscores):
    labels = []
    for score in senscores:
        if(score > 0.1):
            labels.append('positive')
        elif(score < -0.05):
            labels.append('negative')
        else:
            labels.append('netral')
    return labels
```

Figure 8: Threshold Determination.

Kasepuhan Cirebon can be seen in Figure 9 some of the results.

	content	clean_text	word length	sentiment	label
0	Tempatnya tempatnya		31	-7	negatif
1	Nice kerat nice kerato		2	4	positif
2	Terdapat b' terdapat be		32	-9	negatif
3	Mengenal s' mengenal s		5	1	positif
4	Sejarah da' sejarah da		7	7	positif
5	(Translated'good place		4	4	positif
6	Kangen de' kangen de		12	7	positif
7	(Translated'sip		1	0	netral
8	Seru lihat s' seru lihat si		5	6	positif
9	(Translated'lots cool ph		8	8	positif
10	(Translated'historical tc		4	0	netral
11	(Translated'extraordina		3	2	positif
12	Ramai n ny' ramai sejar		5	4	positif
13	Selamat da' selamat da		5	-2	negatif
14	Sebuah ke' sebuah ke		13	15	positif
15	Tetap tegu' tetap teguh		5	4	positif
16	(Translated'full history j		4	0	netral
17	Suka bang' suka bange		11	8	positif
18	Sejarah bu' sejarah buc		6	2	positif
19	(Translated'the place o		9	6	positif
20	(Translated'great mant		2	8	positif
21	(Translated'like home s		5	-1	negatif
22	(Translated'wong wors		6	1	positif

Figure 9: Example of Sentiment Analysis Results.

In Figure 10 you can see the results of the word cloud based on the dataset used, where you can see a collection of words that are most often used based on the size of the word, the larger the size means the word is used more and more, so vice versa if it is smaller, the word is less used.



Figure 10: Example of Sentiment Analysis Results.

#### 4 RESULTS AND DISCUSSION

The results of the sentiment analysis using the Lexicon Based method for the dataset that has been collected to see how tourists respond or comment regarding their opinions on the tourism of the Keraton

The results of sentiment analysis using the Lexicon Based method are shown in Figure 11 and Figure 12. Where the graph states that the trend of tourist sentiment is around 70% positive, around 20% neutral, while negative is only around 10% towards the

Keraton Kasepuhan Cirebon tourism

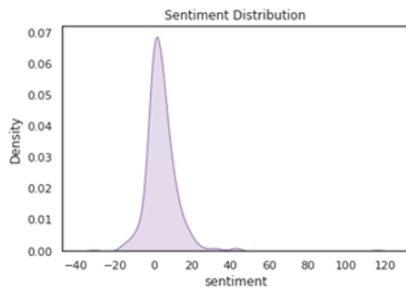


Figure 11: Sentiment Distribution.

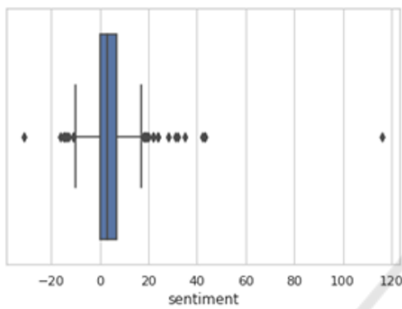


Figure 12: Sentiment Trend Chart.

Whereas in Figure 13 it can be seen that the relationship between numerical data or the number of words in a sentence has a positive sentiment tendency, which means that the number of words has a positive influence on tourists' opinions about tourism at the Kasepuhan Palace in Cirebon.

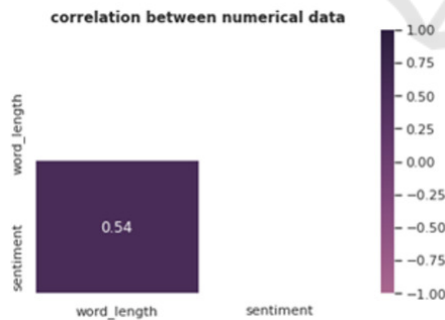


Figure 13: The Meaning of Emoticons.

## 5 CONCLUSION

Based on the research that has been done, it is found that the Lexicon Based method for sentiment analysis can be used on datasets originating from Google Maps Reviews in Indonesian, with the support of the Google Translate Library to translate first if there are sentences or words in English. While the percent-

age of tourist tendencies resulting from the sentiment analysis carried out found that most tourists had a positive response to tourism at the Kasepuhan Palace in Cirebon, around 70% gave a positive response, then around 20% gave a neutral response, while the remaining around 10% gave a negative response negative. However, for future research, the level of accuracy of the results can be calculated whether they are good, then it is also necessary to pay attention to the data dictionary used along with the keywords in order to increase the accuracy even better according to the object of research.

## ACKNOWLEDGEMENTS

This research was supported by the Doctoral Program of Information System at Diponegoro University and also Department of Informatics Engineering at Sekolah Tinggi Ilmu Komputer Poltek Cirebon indicates that both of these organizations have provided support and resources to the research in question. This type of acknowledgement is often included in research papers or other academic documents to thank the organizations and individuals who have contributed to the research and to give credit to them for their contributions. The inclusion of both the Doctoral Program of Information System at Diponegoro University and also Department of Informatics Engineering at Sekolah Tinggi Ilmu Komputer Poltek Cirebon in the acknowledgement suggests that the research has received support from a variety of sources, which can be beneficial in helping to ensure the success and thoroughness of the research.

## REFERENCES

Azhar, Y., Arifin, A., and Purwitasari, D. (2013). Otomatisasi perbandingan produk berdasarkan bobot fitur pada teks opini. *Jurnal Ilmu Komputer*, 6:31–34.

Buntoro, G. (2017). Analisis sentimen calon gubernur DKI Jakarta 2017 di twitter. *INTEGER: Journal of Information Technology*, 2:32–41.

Buntoro, G., Adji, T., and Purnamasari, A. (2014). Sentiment analysis twitter dengan kombinasi lexicon based dan double propagation. In Nugoho, H., editor, *Leveraging Research and Technology through University-Industry Collaboration*, volume VI, page 39–43. Department of Electrical Engineering and Information Technology Universitas Gadjah Mada, Yogyakarta, Indonesia.

Go, A., Huang, L., and Bhayani, R. (2009). Twitter sentiment analysis. *Final Project Report*, 1:1–12.

Liu, B., Hu, M., and Cheng, J. (2005). Opinion observer: Analyzing and comparing opinions on the web. In EL-

- lis, A. and Hagino, T., editors, *World Wide Web 2005*, volume XIV, page 342–351. Association for Computing Machinery, Chiba, Japan.
- Nafan, M. and Amalia, A. (2019). Kecenderungan tanggapan masyarakat terhadap ekonomi indonesia berbasis lexicon based sentiment analysis. *Jurnal Media Informatika Budidarma*, 3:268–273.
- Putranti, N. and Winarko, E. (2014). Analisis sentimen twitter untuk teks berbahasa indonesia dengan maximum entropy dan support vector machine. *Indonesian Journal of Computing and Cybernetics Systems*, 8:91–100.
- Putro, M. (2011). *Analisis Sentimen Pada Dokumen Berbahasa Indonesia dengan Pendekatan Support Vector Machine, Master's project*. Bina Nusantara University, Department of Informatics Engineering.
- Qiu, G., Liu, B., Bu, J., and Chen, C. (2009). Expanding domain sentiment lexicon through double propagation. In Cohn, A. and D., editors, *The Interdisciplinary Reach of Artificial Intelligence*, volume XXI, page 1199–1204. United States of America, Pasadena, California.
- Saputra, F., Nurhadryani, Y., Wijaya, S., and Defina, D. (2021). Analisis sentimen bahasa indonesia pada twitter menggunakan struktur tree berbasis leksikon. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8:135–146.

