

# Sentiment analysis of reviews related to Unesco Global Geopark (UGG) Ciletuh Palabuhanratu on google maps using the naive bayes algorithm

Adhitia Erfina<sup>1\*</sup>, Dian Siti Utami<sup>1</sup>, and Falentino Sembiring<sup>1</sup>

<sup>1</sup>Department of Information System, Nusa Putra University, Sukabumi, West Java, Indonesia

**Abstract.** Ciletuh Geopark is part of the UNESCO Global Geopark Network. This study will analyze a tourist review of the Ciletuh Pelabuhan Ratu Geopark based on reviews on Google Maps. The author considers that user reviews really need to be taken into account because it is to find out the reviews of tourists who share their experiences. Reviews from tourists who have visited geoparks are the most important thing because these reviews can be used as information to be used as data. This research uses the Naïve Bayes Algorithm, because this algorithm is considered to have a high enough level of accuracy so that it can determine the Unesco Global Geopark (UGG) Ciletuh Pelabuhanratu tourist attraction that is often visited based on tourist reviews on Google Maps. Then the highest accuracy value of this research is Palangpang with an accuracy value of 98.61%, the second position is the Cisolok Geyser tourist attraction with an accuracy value of 94.44%, then the third position is the Ujung Genteng tourist attraction with an accuracy value of 98.36%, the fourth namely the Cikaso tourist attraction with an accuracy value of 98.36%, the fifth is the Citepus tourist attraction with an accuracy value of 97.22%, the sixth is the Manic Peak tourist attraction with an accuracy value of 96.92%, the seventh is the Sodong tourist attraction with an accuracy value of 95.83%, the eighth is the Cipanarikan tourist attraction with an accuracy value of 95.01%, the ninth is the Teletubis Hill tourist attraction with an accuracy value of 94.48%, and the last is the Cimarinjung tourist attraction with an accuracy value of 94.44%.

## 1 Introduction

Indonesia is a country with very abundant natural resources. This diversity comes from the flora (plants) and fauna (animals) in Indonesia. Unesco Global Geopark (UGG) Ciletuh-Palabuhanratu was designated as Unesco Global Geopark (UGG) on April 17 2018. Ciletuh has an area of 126,100 hectares or 1,261 km<sup>2</sup> [1]. Because Ciletuh Pelabuhanratu has a very wide location, many tourist attractions in the Ciletuh Pelabuhanratu area are not well known. Therefore, based on the results of the data scrapping process, 10 tourist attractions were selected.

---

\* Corresponding author: [adhitia.erfina@nusaputra.ac.id](mailto:adhitia.erfina@nusaputra.ac.id)

To find out reviews from tourists on Google Maps regarding the Unesco Global Geopark (UGG) Ciletuh Pelabuhanratu, what must be done is to collect review data using the sentiment analysis method. Sentiment Analysis is a person's perspective on a particular topic.[2] . Sentiment analysis is used to analyze a review on Google Maps regarding tourist reviews about the Unesco Global Geopark (UGG) Ciletuh Pelabuhan Ratu. There are many tourist attractions included in the Unesco Global Geopark (UGG) Ciletuh Pelabuhanratu. However, 10 Ciletuh tourist attractions were selected which had the highest ratings according to tourists on Google Maps, including Teletubis Hill, Cikaso, Cimarunjung, Cipanarikan, Cisolok Geyser, Citepus, Palangpang, Puncak Manik, Sodong, Ujung Genteng. Then classify the reviews into positive, neutral or negative reviews. By implementing sentiment analysis, this research hopes that existing problems will be solved. The Naïve Bayes algorithm is the classification technique used in this research. The results of several studies that have been carried out conclude that this algorithm has quite good accuracy values.

Previous research analyzing digital wallet sentiment on Twitter found that in this study there were more negative reviews than positive because negative reviews were more than 50% [3]. Furthermore, research related to the new normal on Twitter has an accuracy value of 87%. Many Twitter users give positive responses to the new normal, because with the new normal their daily activities are more productive [4]. Then research on the use of antibiotics in Indonesia on Twitter. The results of the classification accuracy test were quite good with an accuracy value of 84 % [ 5].

The next research, namely zoom meetings in the Play Store, had an accuracy value of 74.37%. Responses from users of the application gave positive reviews because they considered Zoom cloud meetings to be an application that was easy to use in the learning or work process [6]. Then, research on the Female Daily application on the Play Store showed that the accuracy of the Naïve Bayes algorithm predictions was 82.5%. From the research results, users who gave positive reviews were generally because the Female Daily application contained complete information about beauty, be it skin care, hair care. , and make-up [7]. Furthermore, research regarding the Grab application on the Play Store had an accuracy value of 85.54% which was considered good because it produced an accuracy value above average [8].

Previous research on Google review sites using Naive Bayes obtained an accuracy value of 80.95% [9]. Research with West Kalimantan tourist attractions is in the very good group with an accuracy value of 84.85%. This research uses Naive Bayes with a Sentiment Analysis system in classifying tourist attraction reviews on Google Maps. To make it easier for tourists to access information, the classification produced by the system is displayed in web form [10]. Furthermore, research with beach tourist attractions in Karawang Regency using the RapidMiner application regarding Google Maps reviews obtained an accuracy value of 55% for Cibendo Beach, 65% for Pakis Beach, 85% for Samudra Baru Beach, 70% for Sedari Beach and 65% for Beach New Cape. Based on the reviews of tourists who visited the five beaches, the results showed that two of the five beaches received negative scores while the rest had positive scores, these results were calculated using Naive Bayes [11]. Finally, research related to Bali tourist attractions, obtained an accuracy value of 94.64% for Nusa Penida, 82.86% for Garuda Wisnu Kencana, 80% for The Edge, 90.71% for Pandawa and 85.54% for Pura Luhur Uluwatu. Based on this research, it can be seen that there are five Bali tourist attractions that are used as research objects with Busa Penida as the most recommended place [12].

Therefore, this research aims to analyze tourist reviews of Unesco Global Geopark (UGG) Ciletuh Pelabuhanratu. Based on 10 frequently visited tourist attractions in Sukabumi and the results of data collection regarding tourist reviews of the Ciletuh Pelabuhanratu Geopark.

2 Material and Methods

2.1 Web Scrapping

The data source is Google Maps. The reason why Google Maps is used as a data source is because the data that will be taken are reviews from tourists who have reviewed the Ciletuh Geopark. These reviews are negative reviews and positive reviews, therefore the data needs to be clarified. The following are examples of neutral, negative and positive reviews.

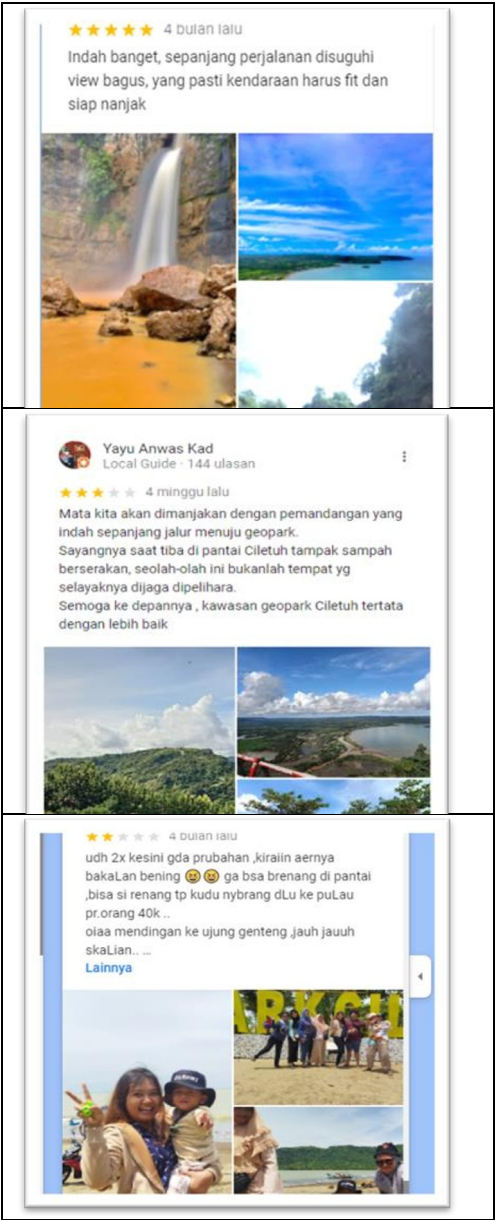


Fig. 1. Example of Positive – Neutral – Negative Review

From Figure 1 it can be concluded that there are 3 types of public comments on tourist objects in the Unesco Global Geopark (UGG). There were those who gave positive comments because Geopark tourism was undoubtedly beautiful, but there were also those who gave negative comments because they were dissatisfied with the cleanliness of the tourist spot.

2.2 Pre-Processing

The preprocessing stages are shown in Figure 2 Preprocessing can also be considered as a classification process. Done through a feature selection process with little impact on the classification process. The data entered at this stage is original data that has not been processed at all and then processed for sentiment classification.

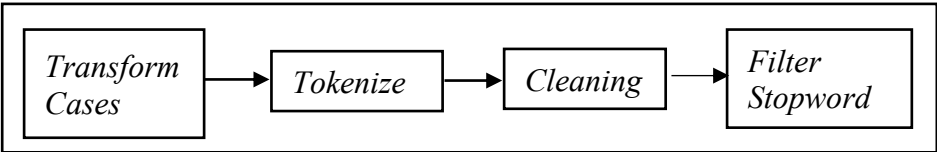


Fig. 2. Pre-processing Step

The data that has been obtained in the Web Scraping process will then go through the next stage, namely the preprocessing stage. Where at this stage it is a process to remove words that have no effect in the data for the classification process later. The following is a process with the R programming language using R Studio tools.

```
23 #cleaning the text
24 # Convert the text to lower case
25 docs <- tm_map(docs, content_transformer(tolower))
26
27 #Remove punctuation
28 docs <- tm_map(docs, tospace, "[[:punct:]]")
29
30 #Remove numbers
31 docs <- tm_map(docs, tospace, "[[:digit:]]")
32
33 # add two extra stop words: "available" and "via"
34 mystopwords = readLines("D:/SRIPSI BISMILLAH/r studio/stopword.csv")
35
36 # remove stopwords from corpus
37 docs <- tm_map(docs, removewords, mystopwords)
38
39 # Remove your own stop word
40 # specify your stopwords as a character vector
41 docs <- tm_map(docs, removewords, c(mystopwords))
42
43 # Eliminate extra white spaces
44 docs <- tm_map(docs, stripwhitespace)
45
46 # Remove URL
47 removeURL <- function(x) gsub("http[[:alnum:]]*", " ", x)
48 docs <- tm_map(docs, removeURL)
49
50 #Replace words
51 docs <- tm_map(docs, gsub, pattern="However", replacement="However")
52
53 #Build a term-document matrix
54 dtm <- TermDocumentMatrix(docs)
55 m <- as.matrix(dtm)
56 v <- sort(rowSums(m),decreasing=TRUE)
57 d <- data.frame(word = names(v),freq=v)
58 head(d, 15)
```

Fig. 3 Pre-processing Step

After going through the pre-processing stage, the data will be saved in (.csv) form in the folder mentioned. In Figure 3, the pre-processing stages include.

### 2.2.1 Transform cases

In the preprocessing step, a lowercase conversion step (to lowercase) is used, changing the text in the data set. The goal is to eliminate the occurrence of inconsistent data, in the sense that the attributes are the same and can have different values just by letters.

### 2.2.2 Tokenize

The process continues with the Tokenizing stage, namely breaking down the words in the text sentences which are separated by spaces according to each word that makes them into one part. So this process produces single words for weighting purposes.

### 2.2.3 Cleaning

Cleaning is the process of removing all non-alphabetic characters, such as symbols and numbers, in an evaluation sentence.

### 2.2.4 Stopwords

The next step is the Stopword Filter (Stopword Removal) step, namely the stage of removing words that are not very important or meaningless, then these words are deleted. Examples of these words are that, or, in, with, but, and to.

## 2.3 Sentiment Classification

Data that has gone through pre-processing and is ready to go through the next process, namely the classification process, where the review text will be labeled into 3 (three) sentiments, namely positive, neutral and negative sentiments based on the sentiment score.

```

77 kata.positif = c(positif, "is near to")
78 kata.negatif = c(negatif, "cant")
79 score.sentiment = function(kalimat2, kata.positif, kata.negatif, .progress='none')
80 {
81   require(plyr)
82   require(stringr)
83   scores = laply(kalimat2, function(kalimat, kata.positif, kata.negatif) {
84     kalimat = gsub('[:punct:]', '', kalimat)
85     kalimat = gsub('[:cntrl:]', '', kalimat)
86     kalimat = gsub('\\d+', '', kalimat)
87     kalimat = tolower(kalimat)
88
89     list.kata = str_split(kalimat, '\\s+')
90     kata2 = unlist(list.kata)
91     positif.matches = match(kata2, kata.positif)
92     negatif.matches = match(kata2, kata.negatif)
93     positif.matches = !is.na(positif.matches)
94     negatif.matches = !is.na(negatif.matches)
95     score = sum(positif.matches) - (sum(negatif.matches))
96     return(score)
97   }, kata.positif, kata.negatif, .progress=.progress )
98   scores.df = data.frame(score=scores, text=kalimat2)
99   return(scores.df)
100 }
101
102 hasil = score.sentiment(kalimat2$text, kata.positif, kata.negatif)
103 view(hasil)

```

**Fig. 4.** Sentiment Classification

2.4 Naive Bayes

After going through all the stages of data collection, text pre-processing, and sentiment classification (data labeling), the final stage is the implementation of the Naïve Bayes algorithm to determine the accuracy value.

```
116
117 #tahap 3 perhitungan akurasi dengan naive bayes
118 library(e1071) #untuk naive bayes
119 library(caret)
120 library(globals)
121
122 #cara membagi data menjadi dua
123 dataset <- read.csv("D:/SKRIPSI BISMILLAH/r studio/ugg_ulasansentimen.csv",stringsAsFactors = F)
124 set.seed(12345)
125 sampel=sample(1:nrow(dataset), 0.75*nrow(dataset), replace = TRUE)
126 training=data.frame(dataset[sampel,])
127 testing=data.frame(dataset)[-sampel,]
128
129 #membuat model naive bayes
130 modelNB= naiveBayes(klasifikasi~,data = training)
131
132 #melakukan prediksi
133 prediksi=predict(modelNB,testing)
134 hasil=confusionMatrix(table(prediksi,testing$klasifikasi))
135 hasil
```

Fig. 5. Naive Bayes

Before data processing is carried out, the data is first divided into training data and testing data. Where from Teletubis Hill, Cikaso, Cimarinjung, Cipanarikan, Cisolok Geyser, Citepus, Palangpang, Puncak Manik, Sodong, Ujung Genteng, each tourist spot has 120 data available, 50% is used as training data and 50% as test data. Then, after dividing it into two parts, namely training data and test data, a naïve Bayes model was created using the training data. Next, make predictions between the model that was created with the training data and compare it with the test data.

3 Results and Discussion

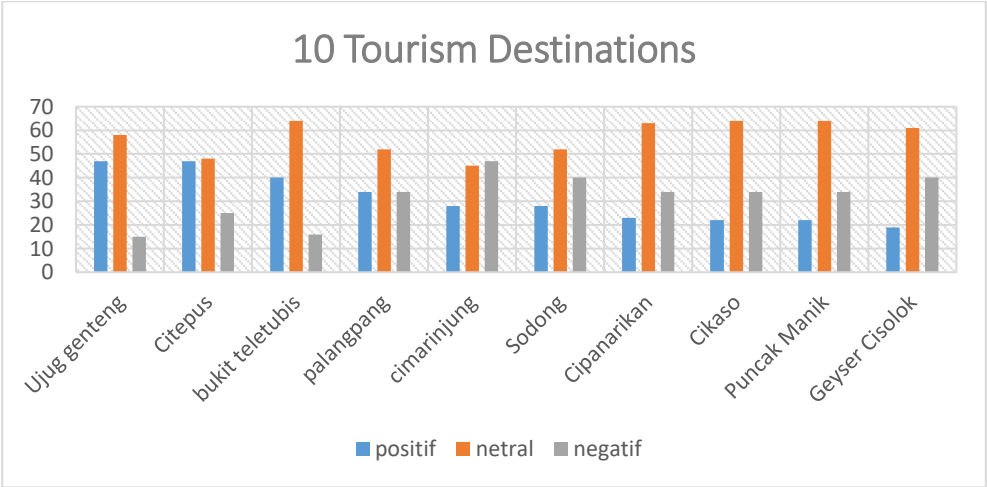
Based on document processing using the Naïve Bayes algorithm for reviews of tourist objects in the Unesco Global Geopark Ciletuh, then accuracy values will be calculated after previously carrying out the classification process, the following are the results of the calculations.

Table 1. Classification Results

No	Tour name	Sentiment			Accuracy Value
		Positive	Neutral	Negative	
1	Ujung Genteng	47	58	15	98.36
2	Citepus	47	48	25	97.22
3	Teletubbies Hill	40	64	16	94.48
4	Palangpang	34	52	34	98.61
5	Cimarinjung	28	45	47	94.44
6	Sodong	28	52	40	95.83
7	Cipanarikan	23	63	34	95.01
6	Cikaso	22	64	34	98.36
9	Puncak Manik	22	64	34	96.92
10	Cisolok geyser	19	61	40	98.44

Based on Table 1, it is found that the sentiment analysis that has been carried out, the most positive sentiment is the Ujung Genteng tourist attraction with 47 positive reviews. In second place is Citepus with 47 positive reviews, then in third place is Teletubis Hill with 40 positive reviews, fourth place namely Palangpang with 34 positive reviews, and fifth place is Cimarunjung with 28 positive reviews, sixth place is Sodong with 28 positive reviews, seventh place is Cipanarikan with 23 positive reviews, eighth place is Cikaso with 22 positive reviews, ninth place namely the peak with 22 positive reviews and finally the Cisolok Geyser with 19 positive reviews. So it was concluded that based on the sentiment of the reviews, the recommended application was the Dana application with the most positive sentiment, namely 182 positive reviews.

It can be seen from the results of calculating the accuracy value using the Naïve Bayes algorithm for the Palangpang tourist attraction with the highest accuracy value, namely with an accuracy value of 98.61%, the second position is the Cisolok Geyser tourist attraction with an accuracy value of 94.44%, then the third position. namely the Ujung Genteng tourist attraction with an accuracy value of 98.36%, fourth namely the Cikaso tourist attraction with an accuracy value of 98.36%, fifth namely the Citepus tourist attraction with an accuracy value of 97.22%, sixth namely the Puncak Manik tourist attraction with a value accuracy of 96.92%, seventh is the Sodong tourist attraction with an accuracy value of 95.83%, eighth is the Cipanarikan tourist attraction with an accuracy value of 95.01%, ninth is the Teletubis Hill tourist attraction with an accuracy value of 94.48%, and lastly is the Cimarunjung tourist attraction with an accuracy value of 94.44%.



**Fig. 6.** Classification Results

So it can be concluded that the accuracy value using the Naïve Bayes algorithm produces a fairly high accuracy value, namely above 90%. However, this accuracy value is not a benchmark for awarding the title of highest rating tourist attraction.

The purpose of Figure 7 is that words that often appear in a review will be used as data. From this picture it can be concluded that water attractions are favorite tourist attractions in the Ciletuh Geopark.

word	freq
air	36
panas	32
pemandian	12
bagus	12
wisata	10
tempatnya	10
geyser	9
cisolok	9
nya	9
keluarga	8
cocok	8
lokasi	7
nyaman	7
pengunjung	7
rekreasi	6

Fig. 7. Word Frequency

3.1 Confusion Matrix

Tabel 2. Confusion Matrix

No	Tour name	Sentiment		
		Positive	Neutral	Negative
1	Cimarinjung	30	29	11
2	Edge of roof tiles	28	39	5
3	Palangpang	24	33	14
4	Cipanmarikan	20	45	7
5	Citepus	14	29	21
6	Teletubbies Hill	13	38	21
7	Sodong	13	38	21
8	Cikaso	13	38	21
9	Bead peak	10	35	24
10	Cisolok geyser	9	39	24

The conclusion from the picture above is that after doing the confusion matrix in R studio the final result of the 10 tourist attractions in Ciletuh Pelabuhanratu turned out to be 14 positive reviews, then 3 tourist attractions which had 13 positive reviews, namely Teletubis Hill, Sodong Waterfall and Cikaso Waterfall, then Puncak Manik with 10 positive reviews, the last one is Geyser Cisolok with 9 positive reviews. The purpose of the image above is predictive analysis which displays and compares actual values with model predicted values which can be used to produce accuracy values.

4 Conclusion

Many tourists give positive reviews because they are satisfied with the Unesco Global Geopark (UGG) Ciletuh Pelabuhanratu tourist attraction. In terms of the beauty of the place and the trip it is quite a feast for the eyes. However, some tourists are less than satisfied with the cleanliness of this tourist attraction and there are several extortion charges that must be eradicated at this place. Therefore, the final conclusion of this research is that the Unesco Global Geopark (UGG) Ciletuh Pelabuhanratu tourist attraction which has the highest rating based on tourist reviews is Ujung Genteng, Citepus, Teletubis Hill, Palangpang Beach, Cimarinjung Waterfall, Sodong Waterfall, Cinarikan Beach, Cikkaso Waterfall, Puncak Manik, Cisolok Geyser. In this research, the author realizes that there are still shortcomings and limitations in this research. Therefore, for further development and research, the researcher has several suggestions, including: 1. In future research, it is hoped that the data



collection process can use web crawling techniques. 2. In this study, many comment reviews used non-standard language and there were also comment reviews that used emoticons. It is hoped that in future research, non-standard language and emoticons can be included in the calculations to get more accurate results. 3. In this research the data comes from Google Maps, for further research the data can be sourced from Twitter or YouTube.

## References

1. "UNESCO Global Geopark (UGG) – Indonesian National Commission for UNESCO." [https://knui.kemdikbud.go.id/?page\\_id=492](https://knui.kemdikbud.go.id/?page_id=492) (accessed Feb. 09, 2022).
2. "Sentiment Analysis: Definition, Steps and Application to Twitter Data - Exsight." <https://exsight.id/blog/2021/10/10/analysis-sentimen-pada-data-twitter/> (accessed Mar. 21, 2022).
3. MWA Putra, Susanti, Erlin, and Herwin, "Sentiment Analysis of Electronic Wallets on Twitter Using the Naïve Bayes Classifier Method," *IT J. Res. Dev.*, vol. 5, no. 1, pp. 72–86, 2020, doi: 10.25299/itjrd.2020.vol5(1).5159.
4. TWU Imelya Susianti, Sri Soerya Ningsih, M. Al Haris, "Sentiment Analysis on Twitter Regarding New Normal Using the Naïve Bayes Classifier Method," *Pros. Semin. Edusainstech FMIPA UNIMUS*, pp. 354–363, 2020, [Online]. Available: <https://prosiding.unimus.ac.id/index.php/edusaintek/article/view/576/578>.
5. O. Dwirawati and KN Siregar, "Sentiment Analysis on Twitter Regarding the Use of Antibiotics in Indonesia Using the Naive Bayes Classifier," *Media Inf.*, vol. 15, no. 1, pp. 1–9, 2019, doi: 10.37160/bmi.v15i1.232.
6. N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, "Sentiment Analysis of Zoom Cloud Meetings on the Play Store Using Naïve Bayes and Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.
7. AK REGGALI, "Sentiment Analysis of Female Daily Application Review Data on the Google Play Website Using the Naïve Bayes Algorithm," July, 2018, [Online]. Available: <https://repositori.usu.ac.id/bitstream/handle/123456789/44520/170403105.pdf?sequence=1&isAllowed=y>.
8. R. Wahyudi and G. Kusumawardana, "Sentiment Analysis on the Grab Application on Google Play Store Using Support Vector Machine," *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
9. H. Herlawati, RT Handayanto, PD Atika, FN Khasanah, AYP Yusuf, and DY Septia, "Sentiment Analysis on Google Review Sites with Naïve Bayes and Support Vector Machine," *J. Komtika (Computing and Inform.)*, vol. 5, no 2, pp. 153–163, 2021, doi: 10.31603/komtika.v5i2.6280.
10. A. Rifa'i, H. Sujaini, and D. Prawira, "Sentiment Analysis of West Kalimantan Tourist Attractions on Google Maps Using the Naive Bayes Method," *J. Edukasi dan Perelit. Inform.*, vol. 7, no. 3, p. 400, 2021, doi: 10.26418/jp.v7i3.48132.
11. W. Khofifah, DN Rahayu, and AM Yusuf, "Sentiment Analysis Using Naive Bayes to See Public Reviews of Beach Tourist Attractions in Karawang Regency on Google Maps Reviews," *J. Interkom J. Publ. Ilm. Bid. Technol. Inf. and Commun.*, vol. 16, no. 4, pp. 28–38, 2022, doi: 10.35969/interkom.v16i4.192.

12. DS Utami and A. Erfina, "Sentiment Analysis of Bali Tourist Attractions on Google Maps Using the Basically Indonesian Naive Bayes Algorithm," vol. 6, pp. 418–427, 2022.
13. Edelweis Lararenjana, "Google Maps is a regional mapping service via the web, know its benefits | merdeka.com," Nov. 28, 2020. <https://www.merdeka.com/jatim/google-maps-sebuah-jasa-pemetaan-region-via-web-ketahui-secepatnya-kln.html> (accessed Jan. 29, 2022).
14. [14] "Ciletuh Sukabumi Geopark, One of the UNESCO Geological Parks in Indonesia | kumparan.com." <https://kumparan.com/viral-food-travel/geopark-ciletuh-sukabumi-lahan-satu-taman-geologi-unesco-di-tanah-air-1vxRnKjciWj/full> (accessed Apr. 14, 2022).
15. "Sentiment Analysis (Sentiment Analysis) : Definition, Types and How It Works. " <https://lp2m.uma.ac.id/2022/02/21/analysis-sentimen-sentiment-analysis-defined-tipe-dan-cara-kerjanya/> (accessed Apr. 14, 2022).
16. ATJ H, "Text Preprocessing to Minimize Meaningless Words in the Text Mining Process," Inform. UPGRIS, vol. 1, pp. 1–9, 2015.
17. "Naive Bayes algorithm | BINUS UNIVERSITY BANDUNG - Creative Technology Campus." <https://binus.ac.id/bandung/2019/12/algorithm-naive-bayes/> (accessed Feb. 09, 2022).
18. "Chapter 1 Introduction to R and RStudio | Introduction to R Programming and RStudio." <https://bookdown.org/aepstk/intror/intro.html> (accessed Apr. 14, 2022).
19. "What is R and How is it Different from RStudio – GEOSPATIALIST." <https://geospacialis.com/bedanya-r-dengan-rstudio/> (accessed Apr. 17, 2022).