Do Tourist Attraction Objects Implement Health Protocols? Analysis of Tourist Attraction Object in East Java Province Using Google Maps Review

D P Ratriatmaja[1], N W K Projo[1]

[1] Politeknik Statistika STIS, Jakarta, Indonesia

*Corresponding author's e-mail: 221709638@stis.ac.id

Abstract. The COVID-19 pandemic has impacted the tourism sector, particularly the Tourist Attraction Object (TAO) in Indonesia. This research aims: to analyse the implementation of health protocols and facility conditions at TAO, to analyse the change in visitor sentiment and rating towards TAO before and during the COVID-19 pandemic, to analyse the close relationship between ratings and reviews of visitor sentiment on TAO, to analyse the possibility of web scraping data to complement tourism data from BPS Statistics Indonesia. Using Google Maps review, this research uses the Multinomial Naïve Bayes (MNB), Term Frequency-Inverse Document Frequency (TF-IDF), pseudo-labelling, and word association methods. The results show that the health protocol has been implemented in TAO of East Java province, the available facilities are good, and there is no change in reviews during the TAO pandemic. The Stuart-Kendall Tau-c value shows a weak relationship in a positive direction between rating and review sentiment. According to Haversine, Jaro Winkler, and Levenshtein, the data calculation indicates that web scraping data can complement tourism data for BPS-Statistics Indonesia.

1. Introduction

The tourism sector has a significant influence on the Indonesian economy. Foreign exchange contributed by the tourism sector in 2018 rose to USD17.6 billion with a total of 15.8 million foreign tourist arrivals [1]. Every year tourism activities in Indonesia have an increasing trend, which can be seen from the growing number of tourists, local tourists, and foreign tourists. In 2018 the addition of local and foreign tourists was 12 percent and 5.4 percent from 2017 [2]. However, at the end of 2019, the COVID-19 pandemic was first detected in Wuhan, the People's Republic of China. Since then, Indonesian tourism and many other sectors have suffered from the COVID-19 pandemic impact.

In Indonesia, the first case of COVID-19 was found on March 2, 2020. Since the first case, Indonesia's number of COVID-19 cases has continued to increase [3]. The COVID-19 pandemic has impacted the tourism sector, particularly in the number of foreign tourist arrivals. The growth of foreign tourist visits and COVID-19 cases is shown in Figure 1. Positive cases of COVID-19 are increasing, while the number of foreign tourists is decreasing [4].
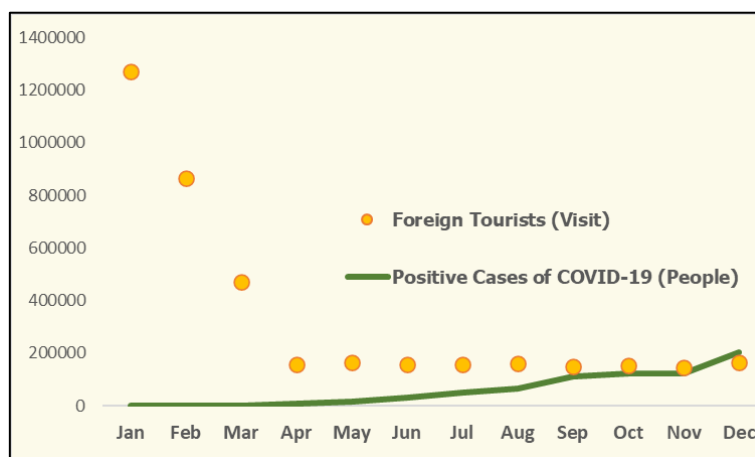
Figure 1. Number of Foreign Tourists and Positive Cases of COVID-19 per Month in 2020 [3] [4]

The Indonesian government has to find effective strategies so that the Indonesian economy can go along with COVID-19. The community must begin to adapt to the existence of this virus and continue to carry out productive activities. For this reason, the Decree of the Minister of Health of the Republic of Indonesia Number 382 about Public Health Protocols in Public Places and Facilities in the Context of Prevention and Control of Corona Virus Disease 2019 (COVID-19) was made and is effective from June 19, 2020 [5]. The regulation also manages health protocols implementation in TAO as a public place. TAO is everything that has uniqueness, beauty, and value in the form of a diversity of natural, cultural, and artificial wealth that is the target or purpose of tourist visits [6]. It is a commercial tourist destination with an entrance ticket system [7].

This research uses TAO data in East Java Province because it has the highest TAO businesses in Indonesia. In 2018 the number of TAO in East Java Province was 529 companies or enterprises [7]. On September 4, 2020, East Java Governor Regulation Number 53 of 2020 concerning the Implementation of Health Protocols in the Prevention and Control of COVID-19 has been launched [8]. On 16-17 September 2020, there were 7,003 health protocol violators [9]. The number of health protocol violations alerts the local government to periodically monitor the health protocol implementation to comply with the governor's regulation.

In addition, prospective TAO visitors must know the implementation of health protocols in TAO as their potential tourist destinations. Not only health protocol information but also information about the available facilities at the TAO. The condition of the available facilities can affect visitors' level of interest in a tourist spot [10]. This information can be obtained from experience shared by the previous visitors.

We collect Google Maps review using the web scraping method. Google Maps is a web-based mapping service, and it provides a review feature for users. Recommendations based on user experience usually are travel information choices and the most influential source of travel decisions [11]. The business owner takes advantage of online reviews to see the strengths and weaknesses of the service as material for customer relationship management [12].

Review on Google Maps is included in Big Data because it can store large data with insight that can be analysed. A review of an TAO contains information on health protocols, available facilities, and other information related to TAO. Exploring the visitor sentiment can be done by conducting sentiment analysis using the Multinomial Naïve Bayes (MNB) method as a classifier, Term Frequency-Inverse Document Frequency (TF-IDF) as a weighting factor, and pseudo-labelling algorithms in labelling data.

BPS-Statistics Indonesia (BPS), as the national statistical agency in charge of collecting data on Indonesia, also collects data related to TAO. Data collection is carried out every year in the Survey of Companies/Businesses of Tourist Attraction Objects. The data collected includes data on basic information, expenses, and operating income of TAO. One of the publications of this survey is the

Directory of Companies/Businesses of Tourist Attraction Object. Just as BPS collects general information related to TAO, Google Maps also includes some of this information. So, there is an opportunity that the data on Google Maps can complement BPS data. This objective will be analysed by matching the location and name of TAO from the results of data collection with BPS data.

2. Methods

2.1. Literature Review
A literature study is a technique that collects information from various sources and scientific references relevant to the research topic. Scientific references that can be used as sources are in print media and digital media. The information that has been collected is used as a theoretical basis and guide in designing a data collection system and data analysis.

2.2. Data Collection
TAO review data is collected from Google Maps using the web scraping method of the Selenium package in Python. Web scraping is carried out in two stages. The first is collecting TAO data published on google maps until January 26, 2021. The second step is scrapping review data published until February 1, 2021. TAO scraping is done by searching for keywords, namely natural tourism, artificial tourism, and cultural tourism, along with mentioning the name of each district/city in East Java Province, for example, "Wisata Alam di Kabupaten Nganjuk." These keywords are based on the definition of TAO according to Law Number 10 of 2009. In East Java Province, there are 38 districts/cities, so that the keywords created are 114 keywords. Scraping reviews are done by re-accessing tourist attraction links that have been saved during the previous scraping for the TAO.

2.3. Pre-Processing of Review Text
Before making a review analysis, the pre-processing stage of visitor review text is carried out as follows,
- Cleaning: removes numbers, punctuation marks, line spaces, spaces at the beginning and end of sentences, non-ASCII characters, emojis and converts all text letters to lowercase,
- Tokenizing: the stage of word-solving in the review sentence, generally separated by spaces,
- Eliminating stop words and slang words: stop words are common words that often appear and have no meaningful meaning. Slang words are slang words that are written not following standard words,
- Stemming: change the word into a root word by removing the suffix on the word.

The pre-processing stage of review text is carried out using the Python package nltk and sastrawi.

2.4. Review Sentiment Classification
This stage is carried out to classify the review text into positive, neutral, and negative classes. Term Frequency-Inverse Document Frequency (TF-IDF) is weighted before classifying the reviews using the Multinomial Naïve Bayes (MNB) method. The following is the MNB classification formula with TF-IDF [13],

$$P(c|term\ dokumen\ d) = P(c)x\ P(t_1|c)x\ P(t_2|c)\ x\ ...\ x\ P(t_n|c) \qquad (1)$$

Where $P(c|term\ dokumen\ d)$ is the probability that a document belongs to class c, $P(c)$ is the prior probability of class c, t_n is the word in the n-th document d, and $P(t_n|c)$ is the probability of the n-th word given class c.

With the prior probability of class c ($P(c)$) and the probability of the n-th word given class c ($P(t_n|c)$) as follows [13].

$$P(c) = \frac{N_c}{N} \qquad (2)$$

$$P(t_n|c) = \frac{W_{ct}+1}{(\sum W'\,\epsilon VW'_{ct})+B'} \qquad (3)$$

Where $P(c)$ is the prior probability of class c, N_c is number of class c in all documents, N is total number of documents, $P(t_n|c)$ is the probability of the n-th word given class c with word weighting TF-IDF, W_{ct} is TF-IDF or W weighting value of term t in class c, $\sum W' \in V W'_{ct}$ is total number of W of all terms that are in class c, and B' is W number of unique words in all documents.

Implementation of the Multinomial Naïve Bayes classifier formula is used in the Python package sklearn in the naïve_bayes library, module MultinomialNB. Word weighting TF-IDF also uses the Python package sklearn, that is TfidfVectorizer with unigrams and bigrams range.

Labelling data uses the pseudo-labelling algorithm. We make two sentiment classification models as explained in Figure 2. Model 1 trains data that has been labelled manually. While Model 2 trains the combined data from labelled data and pseudo data. Data is divided into labelled data and unlabelled data with a ratio of 1: 99. The text labelling reviews are performed manually on labelled data. As appears in Figure 2, labelled data will be used to build Model 1. After Model 1 is evaluated, then Model 1 will be used to predict the pseudo data. Pseudo data is a portion of unlabelled data, which is 20% of unlabelled data. These prediction results are referred to as pseudo-label data, and the value is considered as true value. The next step is creating Model 2. It starts with combining labelled data and data pseudo. Like Model 1, Model 2 is also evaluated and used to predict the rest of the unlabelled data.
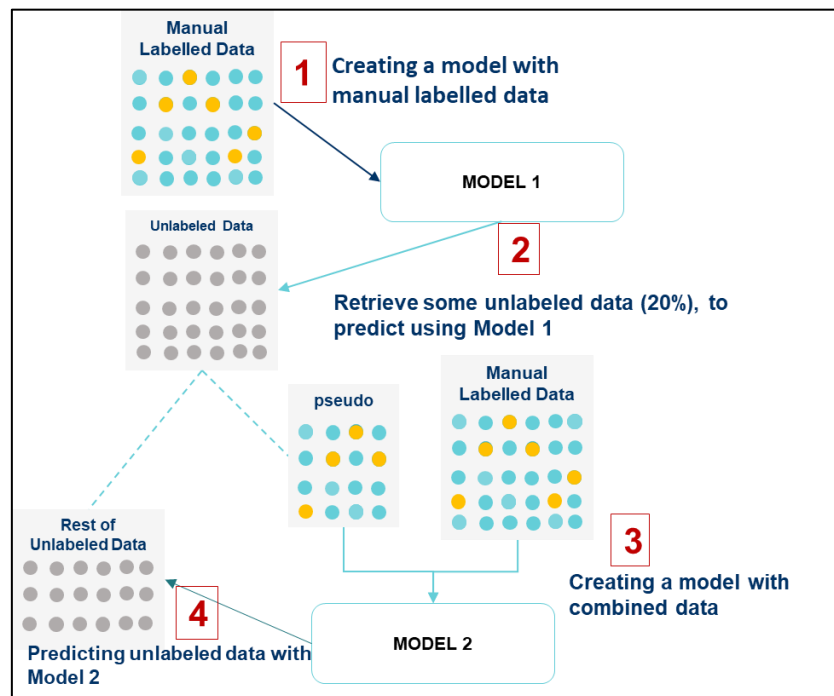

Figure 2. Sentiment Classification Program Flow [14]

2.5. Model Evaluation

Model evaluation is carried out to measure how accurately the model predicts the testing data on the sentiment prediction model. The model evaluation method used is as follows,

- Accuracy Value: the number of correct predictions divided by the number of predicted data,
- Cohen Kappa value: a coefficient that evaluates the agreement between two rates, namely the predicted and actual values. Calculation of the value of Cohen Kappa with the formula [15]

$$\kappa = (p_0 - p_e)/(1 - p_e) \qquad (4)$$

Where κ is the Cohen Kappa value, p_0 is the observed agreement ratio, and p_e is the expected agreement when the two rates assign labels at random. Table 1 is the classification of the interpretation of the Cohen Kappa accuracy value [16],

Table 1. Cohen kappa value classification

Cohen Kappa Nilai Value	Strength of Agreement
< 0.0	Poor
0.01 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 0.99	Perfect

Model evaluation is carried out for model validation with *10-fold Cross-Validation*. It is a model evaluation method by dividing the data 10-*fold* to validate the model's accuracy. All stages in model evaluation uses Python package sklearn.

2.6. Word Association

Word association shows the close relationship between two or more qualitative variables [17]. Word association is used to find related that indicated the application of health protocols and facilities available at TAO. So that it can be seen what topics often arise and are associated with the application of health protocols and facilities available at TAO. The word association formula used is as follows [18]

$$r = \frac{N \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{\{N X_i^2 - (\sum X_i)^2\}\{N \sum Y_i^2 - (\sum Y_i)^2\}}} \tag{5}$$

Word association calculations are performed on Rstudio with the R function, namely findAssocs() in the "tm" package. In this study, the minimum correlation value limit is 0.26. According to Table 2, the correlation value of 0.26 indicates enough correlation between the two variables [19].

Table 2. Classification of the correlation value

Correlation Value	Interpretation
0	No correlation
>0-0.25	Very weak correlation
>0.25-0.5	enough correlation
>0.5-0.75	Strong correlation
>0.75-0.99	Powerful correlation
1	Perfect correlation

2.7. Stuart-Kendall Tau-c Correlation (τ_c)

Stuart-Kendall Tau-c correlation is a method of calculating the correlation of two variables with ordinal data types. This correlation is used when the number of categories in the first variable is not equal to the number of categories in the second variable. The cross-table is not square (the number of rows is not equal to the number of columns). The value of the tau-c correlation is at $-1 \leq \tau_c \leq 1$. A negative correlation value indicates a non-unidirectional correlation, while a positive correlation indicates a unidirectional correlation [19]. The calculation of the tau-c correlation is suitable for data that are not normally distributed [20]. Here's the formula for calculating the value of τ_c [21],

$$\tau_c = 2m(P - Q)/n^2(m - 1) \tag{6}$$

P is the number of concordances, Q is the number of discrepancies, n is the total number of observations, and m equals min(R, C). R is the number of rows, while C is the number of columns. The calculation of the Stuart-Kendall Tau-c correlation is performed on the Rstudio application with the StuartTauC() function from the DecsTools library.

2.8. Data Matching

Data matching is done to see if the scraped TAO matches the BPS directory. The matching method used is string similarity with Jaro Winkler and Levenshtein and the distance of the coordinates with Haversine.

Before performing data matching, it is necessary to convert the data to obtain the coordinates of the tourist attraction. The latitude and longitude of the scraping data are obtained from the tourist attraction link on Google Maps, which contains the coordinates of the tourist attraction. While in BPS data, the coordinate points are obtained from converting the tourist object's address variable into coordinate points. Converting address text into coordinates is also known as geocoding. Geocoding is done with the geopy package from Python. The ArcGIS library is used to find the coordinates of an address in the world from this package so the TAO coordinates can be obtained and used to calculate the location distance.

Data matching is done with the attributes of the district's name, the name of the TAO, and the coordinates. Two places are said to be the same if they have the Jaro Winkler value of the district name = 1 or the Levenshtein value of the district name = 100; the Jaro Winkler value of the TAO name 0.85 or the Levenshtein value of the name 85; and the Haversine value 2 km [22]. The Jaro distance calculation formula is [23],

$$d_j = \frac{1}{3} \, x \, \left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m} \right) \tag{7}$$

with d_j is the Jaro distance, m is the same number of characters, $|s_1|$ is the length of string 1, $|s_2|$ is the length of string 2, and t is the number of transpositions. So, the Jaro Winkler distance calculation formula is as follows: [23],

$$d_w = d_j + (l.p(1 - d_j)) \tag{8}$$

d_w is the Jaro Winkler distance, d_j is the Jaro distance, l is the length of the common prefix at the beginning of the string with a maximum value of 4 characters. (The length of the same character before finding the maximum inequality 4), and p is a constant of the scaling factor. According to Winkler, the default value of p is 0.1. The Jaro Winkler distance is calculated with the jaro_winkler function in the Python jellyfish package.

The following Levenshtein formula is used for string matching [24],

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j) & if \min(i,j) = 0 \\ \min \begin{cases} lev_{a,b}(i-1,j) + 1 \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & others. \end{cases} \tag{9}$$

With $lev_{a,b}(i-1,j)+1$ is a delete operation, $lev_{a,b}(i,j-1)+1$ is an insert operation, $lev_{a,b}(i-1,j-1)$ is a delete operation, and $1_{(a_i \neq b_j)}$ value is 1 if element $a_i \neq b_j$. And the calculation uses the Fuzzy Wuzzy library in Python

Meanwhile, to calculate the distance between two locations by calculating the length of the coordinates, the Haversine distance formula is used as follows [25]

$$a = sin^2 \left(\frac{\emptyset_2 - \emptyset_1}{2} \right) + \cos(\emptyset_1).\cos(\emptyset_2).sin^2(\frac{\lambda_2 - \lambda_1}{2}) \tag{10}$$

$$d = 2r.\arcsin(\sqrt{a}) \tag{11}$$

Where d is the distance of Haversine, $\emptyset_1$ and $\emptyset_2$ is the latitude point, and λ_2 and λ_1 is the longitude points. This method is calculated with Python's math package.

The results of matching data that have met the threshold above are evaluated by string evaluation in Character Error Rate (CER). Calculation of CER with the formula [26]

$$CER = (S + I + D)/N \tag{12}$$

S is the number of characters that are replaced/exchanged (Substitution), I is the number of characters that are inserted (Insertion), D is the number of characters that are deleted (Deletion), and N is the character in the string or reference word. The calculation is done with the fastwer Python package. The CER accuracy value is obtained, which is 100-CER.

3. Results

TAO data that has been collected is 13,708 and after pre-processing stage, the total data is 2,694. Meanwhile, data from the BPS TAO directory shows that East Java has 526 TAO businesses. The review data collected are 645,875 reviews. The pre-processing and equalization of periods are carried out so that the number becomes 238,274 review data. The equalization is a process of cutting data, which is only retrieving data published from 2019-2021. The average number of reviews per TAO is 88.45.

Sentiment analysis uses 2,383 data for Model 1, 47,179 data for pseudo label data, and Model 2 predicts 188,712 data. The following tables are the results of the model evaluation.

Table 3. Results of the evaluation of the sentiment classification model

Model	Accuracy	Cohen Kappa
Model 1	76.37%	39.14%
Model 2	92.29%	59.12%

Table 3 shows the accuracy and Cohen Kappa values of the two models. It shows fair results for Model 1 and moderate for Model 2 [16]. After predicting unlabelled data with Model 2, the data are recombined into one. The final result of sentiment classification is as follows,

Table 4. The final result of sentiment classification

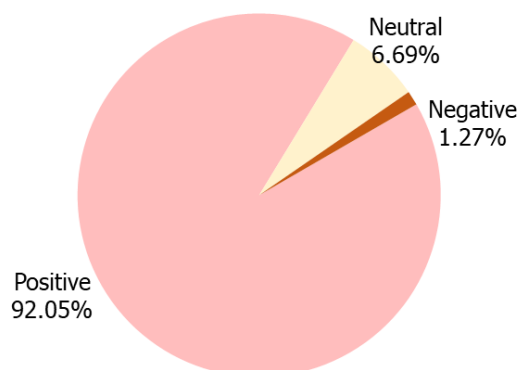| Data	Positive	Neutral	Negative	Total Data
The final result	219,322	15,929	3,023	238,274

Figure 3. Classification of Reviews Based on Sentiment Class

The number of reviews based on sentiment class is shown in Table 4. In figure 3, 92.05% of the data is labelled positive, 1.27% data is labelled negative, and 6.69% is labelled neutral.

The keywords available for public facilities include "fasilitas", "parkir", "kamar mandi", and "musala" [10]. While the public place health protocol keywords are taken from the East Java Governor Regulation Number 53 of 2020, include "masker", "cuci tangan", "handsanitizer", "jaga jarak", "cek suhu", "prokes", "pandemik", dan "protokol". The reviews with positive and negative labels are used to search for available word associations. Meanwhile, to find information related to health protocols, data are written after the East Java Governor Regulation Number 53 of 2020 to see TAO's compliance with these regulations. Before calculating word associations, the data is filtered again. Data filtering is carried out to select data with pre-processing attributes that have the keywords mentioned before. We use the correlation limit of 0.26, as it shows enough correlation [19].

We use 1,249 reviews that has been filtered according to keywords to get information on the implementation of health protocols. The following is the result of calculating word associations,

Table 5. Health protocol words association

Governor's review (corlimit= 0.26)					
masker		cuci tangan		protokol kesehatan	
pakai	0.61	sedia	0.36	terap	0.32
		sabun	0.34	patuh	0.29

Table 5 shows the word "masker" often appears with the word "pakai", the word "cuci tangan" often occurs with the word "sedia" & "sabun" and the word "protokol kesehatan" often appears with the word "terap" and "patuh". The word association calculation above shows that the implementation of the health protocol has been implemented in the TAO of East Java Province. The health protocols, including masks, handwashing facilities with soap, and the implementation of health protocols, have complied.

The data with positive and negative labels are used to find information related to available facilities, with 13,494 and 422 reviews, respectively. The result is in Table 6.

Table 6. Association of facility words available on positively labelled data

Positive Reviews (corlimit= 0.26)			
fasilitas		parkir	
lengkap	0.3	masuk	0.26

The positive reviews in Table 6 show that the word "fasilitas" often appears together with the word "lengkap", with a correlation value of 0.3. While the word "parkir" often occurs together with the word "masuk", with a correlation value of 0.26.

Table 7. Association of facility words available on negative labelled data

Negative Reviews (corlimit= 0.26)			
parkir		kamar_mandi	
masuk	0.32	kotor	0.33
		renovasi	0.33
		kunci	0.27

There are still visitor perceptions regarding the available facilities with negative sentiments, as shown in Table 7. The word "parkir" often appears together with the word "masuk", with a correlation

value of 0.32. The word "kamar mandi" often appears along with word "kotor", "renovasi", and "kunci", with correlation values of 0.33, 0.33, and 0.27. It is necessary to evaluate and improve facilities with negative sentiments.

The word association available facilities calculation above shows that TAO businesses in East Java Province have complete facilities, such as parking lots. But there are still facilities that need special attention because they still get negative reviews, namely parking lots and bathrooms.
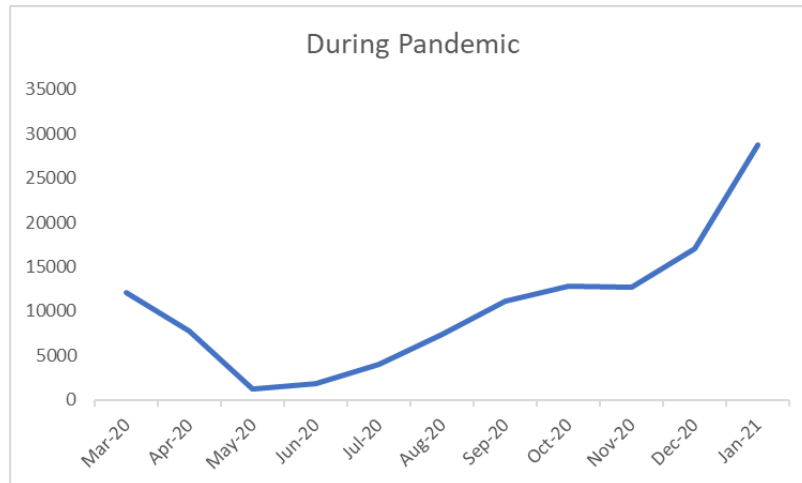
Figure 4. Timeseries Reviews During the Pandemic

The reviews data is divided into two parts to see reviews and ratings changing before and during the pandemic. The before pandemic data are reviews data with a publish date attribute one year before the pandemic begin in Indonesia, or from March 2020 until March 2019. The during pandemic data are reviews data with a publish date attribute from March 2020 until February 2021. The time series review of TAO in East Java on Google Maps during the pandemic is shown in Figure 4.

The number of reviews data is calculated based on time and sentiment labels. The changes in sentiment labels and changes in the average rating value given by visitors to each TAO are presented in the figure below.
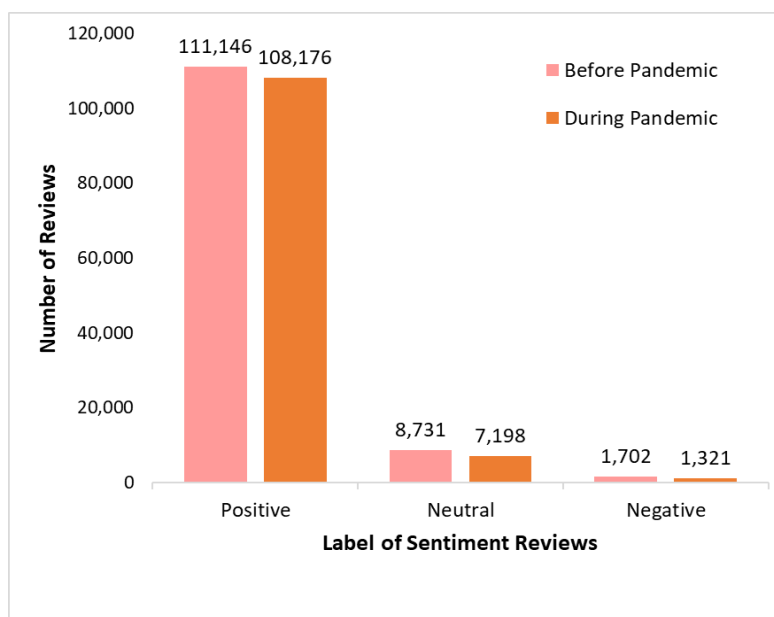

Figure 5. Total Sentiments by Published Time and Sentiment Class of Reviews

Figure 5 shows the change in the number of sentiments by time and class of sentiment. It shows a decrease in the number of reviews on TAO in East Java during the pandemic in the three sentiment classes. During the pandemic, the number of TAO visitors decreased. It is due to restrictions on activities in public facilities/places to prevent the spread of COVID-19.
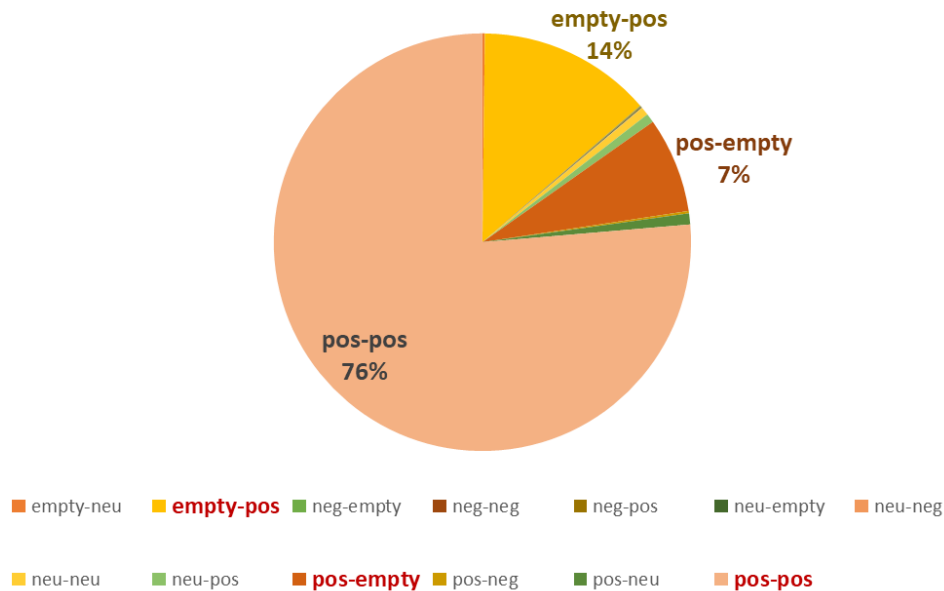

Figure 6. TAO Percentage Based on Changes in Sentiment Label

From Figure 6, about 76% of TAO with positive labels before the pandemic still get positive labels from visitors during a pandemic. It means that there is no change in sentiment. In addition, 14% of TAO has changed from no reviews or is empty to positive sentiment during a pandemic. While only 7% of TAO has changed from positive sentiment becomes no reviews during a pandemic. Other changes can be categorized as minor percentages.
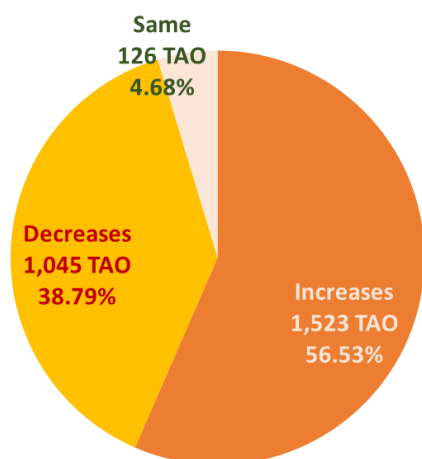

Figure 7. TAO Percentage Based on Changes in Average Rating Value

About 56.53% of TAO experience an average increase during the pandemic, as shown in Figure 7. During the pandemic, the ratings are better than before the pandemic. However, some TAO experiences decreased in the average rating. In addition, there are TAO with the same assessment before the pandemic and during the pandemic.

The Stuart-Kendall Tau-c correlation on the Rstudio application show the relationship between rating and visitor sentiment labels. We develop a contingency table in Table 8 before calculating the correlation.

Table 8. Contingency table

		Sentiment Labels		
		Positive	Neutral	Negative
Rating	1	4,594	385	312
	2	4,034	277	411
	3	19,869	1,315	1,089
	4	51,098	3,222	701
	5	139,727	10,730	510

Table 8 is a contingency table of rating values with sentiment labels. Calculation of the Stuart-Kendall Tau-c correlation at Rstudio obtains a correlation value of 0.016. This value is categorized as a very weak correlation with a positive direction [19]. The positive correlation shows that when visitors give an excellent rating value, the sentiment label is also good or positive. But the correlation between rating and review sentiment is very weak, indicating that the two are less interconnected. Other factors might be more closely related to ratings and sentiment.

Data conversion to obtain latitude and longitude points is carried out on both TAO data from scraping results and the BPS directory. Then data matching is done with the following results,
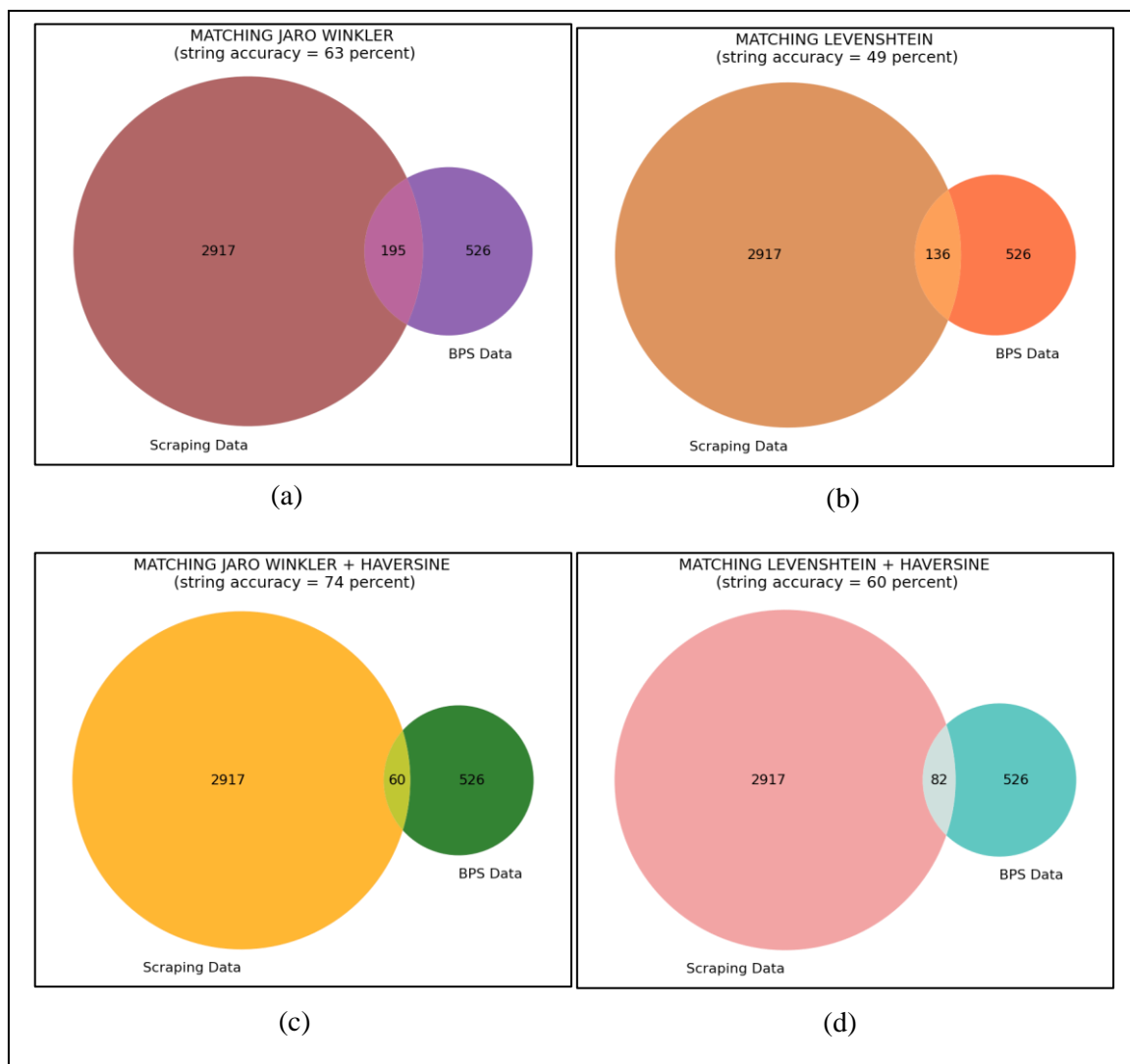

Figure 8. TAO Matching Data Results

Figure 8 shows the results of the combination method used. The matching string method with Jaro Winkler is better than the Levenshtein method from the calculation of string accuracy. The best string accuracy is obtained from the combination of Jaro Winkler and Haversine methods, with a string accuracy of 74% (Figure 8(c)). However, the most suitable amount of data is obtained from implementing the Jaro Winkler method, 195 data or 37% from BPS directory data with a string accuracy of 63% (Figure 8(a)). This accuracy is categorized as a high value compared to the accuracy value of the other combinations of methods. It can be said that the TAO data from Google Maps potentially complement BPS data because there is a match between the two.

4. Conclusions

We conclude that,

1. Sentiment analysis using the MNB + TF-IDF method and applying a pseudo-labelling algorithm has been successfully carried out to label TAO reviews on Google Maps,
2. The health protocol following East Java Governor Regulation Number 53 of 2020 has been implemented in East Java Province TAO, namely wearing masks and providing handwashing places with soap,
3. Complete facilities are available at TAO, such as an entrance parking lot. However, it is still necessary to improve the bathroom facilities in East Java TAO,
4. Visitors' opinions on East Java TAO did not change during the pandemic, and this can be seen from the sentiment change. Visitor sentiment towards TAO remains positive during the pandemic,
5. Visitors' assessment of East Java TAO has increased, it can be seen from the number of TAO with a better rating during the pandemic,
6. Calculation of correlation shows a weak relationship with a positive direction between rating and visitor sentiment,
7. Google Maps web scraping data can potentially be complementary data for BPS as indicated by a matching value of 37%.

The tourism sector, especially the TAO business, is a potential business during a pandemic as long as they implement health protocols, follow the existing regulations, provide good facilities, and improve existing facilities.

We suggest business owners to pay attention to the facilities available in their TAO business. It is necessary to repair or renovate the available facilities, especially the bathroom. In addition, both visitors and business owners must always maintain the application of health protocols in TAO. For further research, we can explore more information in reviews on Google Maps, not only tourist attractions. It considers the ability of Google Maps reviews to store information based on visitor experiences.

References

[1]   Menpar RI 2019 *Buku Saku Manajemen Krisis Kepariwisataan Kemenpar RI* (Jakarta: Kemenpar RI)
[2]   Menparekraf RI 2020 *Rencana Strategis 2020-2024* (Jakarta: Kemenparekraf RI)
[3]   KPC-PEN. *Perkembangan Kasus Per-hari* [Internet]. Jakarta. 2020 [cited January 9, 2021]. Available from: https://covid19.go.id/peta-sebaran
[4]   BPS-Statistics Indonesia (BPS). *Berita Resmi Statistik BPS RI* [Internet]. Jakarta. 2020 [cited March 29, 2020]. Available from: https://www.bps.go.id/.
[5]   Decree of Health Minister of Republic Indonesia Number 382 of 2020 concerning Public Health Protocols in Public Places and Facilities in the Context of Prevention and Control of Corona Virus Disease 2019 (COVID-19). Jakarta. 2020.
[6]   Law of the Republic of Indonesia Number 10 of 2009 concerning Tourism. Jakarta. 2009.
[7]   BPS-Statistics Indonesia (BPS) 2020 *Statistik Objek Daya Tarik Wisata 2018* (Jakarta: BPS-Statistics Indonesia (BPS))
[8]   East Java Governor Regulation Number 53 of 2020 concerning the Implementation of Health Protocols in the Prevention and Control of Corona Virus Disease 2019. Surabaya. 2020.

[9] Kurniawan D. *Polda Jatim Jaring 7.000 Pelanggar Protokol Kesehatan, Denda Rp 63 Juta* [Internet]. Surabaya. 2020 [cited January 10, 2021]. Available from: https://surabaya.liputan6.com/

[10] Suchaina 2014 *Pengaruh Kualitas Fasilitas Sarana dan Prasarana Terhadap Peningkatan Jumlah Pengunjung Wisata Danau Ranu Gratis. Jurnal Psikologi* **2**(2) 89-109.

[11] Pan B, Crotts J C and Maclaurin T 2007 *Travel Blogs and the Implications for Destination Marketing. Journal of Travel Research* **46**(1) 35-45

[12] Cho Y, Im I and Hitlz R 2003 *The Impact of E-Services Failures and Customer Complaints on Electronic Commerce Customer Relationship Management. Journal of Consumer Satisfaction, Dissatisfaction & Complaining Behavior* **16** 106

[13] Rahman A, Wiranto and Doewes A 2017 *Online News Classification Using Multinomial Naive Bayes. Jurnal Ilmiah Teknologi dan Informasi* **6**(1)

[14] Fudholi D H and Juwairi K P 2019 *Pemanfaatan Teknik Semi-supervised Learning untuk Klasifikasi Dokumen Medis* [Thesis]. Yogyakarta: Universitas Islam Indonesia

[15] Cohen J 1960 *A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement* **20**(1) 37-46

[16] Poesio, Artstein R and Massimo 2008 *Inter-Coder Agreement for Computational Linguistics. Association for Computational Linguistics* **34**(4)

[17] Ulwan N M 2016 *Pattern Recognition pada Unstructured Data Teks Menggunakan Support Vector Machine dan Association* [Thesis]. Yogyakarta: Universitas Gajah Mada

[18] Kurniasari S R 2018 *Implementasi SVM dan Asosiasi untuk Sentiment Analysis Data Ulasan The Phoenix Hotel Yogyakarta pada Situs TripAdvisor* [Thesis]. Yogyakarta: Universitas Islam Indonesia

[19] Sarwono J. *Korelasi* [Internet]. 2011. [cited March 29, 2020]. Available from: https://www.jonathansarwono.info/korelasi/korelasi.htm.

[20] Vusvitasari R, Nugroho S and Akbar S 2008 *Kajian Hubungan Koefisien Korelasi Pearson, Spearman-rho, Kendall-Tau, Gamma, dan Somers. Gradien* **4**(2)

[21] SAS Institute 2009 *SAS/STAT 9.2 User's Guide, Second Edition: The FREQ Procedure (Book Excerpt) 2nd Edition* (United States of America: SAS Publishing)

[22] Annisa C F and Pramana S 2020 *Kajian Pemanfaatan Data Google Maps dalam Official Statistics (Studi Kasus: Usaha Sektor Penyedia Makan Minum di Pulau Jawa dan Bali). Seminar Nasional Official Statistics 2020* **2020**

[23] Kurniawati A, Puspitodjati S and Rahman S 2010 *Implementasi Algoritma Jaro-Winkler Distance untuk Membandingkan Kesamaan Dokumen Berbahasa Indonesia.* Universitas Gunadarma, Depok

[24] Putra D P 2020 *Pemanfaatan Algoritma Pencocokan String dalam Pencarian Kata Tidak Baku pada KBBI.* Sekolah Teknik Elektro dan Informatika. Bandung

[25] Miftahuddin Y, Umaroh S and Karim F R 2020 *Perbandingan Metode Perhitungan Jarak Euclidean, Haversine, dan Manhattan dalam Penentuan Posisi Karyawan (Studi Kasus: Institut Teknologi Nasional Bandung). Jurnal Tekno Insentif* **14**(2) 69-77

[26] Read J, Mazzone E and Horton M 2005 *Recognition Errors and Recognizing Errors – Children Writing on the Tablet PC* (Preston: University of Central Lancashire)