IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.17, No.1, January 2023, pp. 79~90
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: https://doi.org/10.22146/ijccs.77354

Aspect-Based Sentiment Analysis in Bromo Tengger Semeru National Park Indonesia Based on Google Maps User Reviews

Cynthia As Bahri*[1], Lya Hulliyyatus Suadaa [2]
[1,2] Program Studi Komputasi Statistik, Politeknik Statistika STIS, Jakarta, Indonesia
e-mail: ***[1]221810227@stis.ac.id**, [2]lya@stis.ac.id

Abstrak

Keberadaan teknologi mampu mempengaruhi dan membentuk pola perilaku seseorang saat merencanakan wisata, sedang berwisata dan setelah berwisata. Review yang diberikan oleh pengunjung terhadap destinasi wisata dapat dimanfaatkan sebagai bahan evaluasi untuk peningkatan kualitas destinasi wisata serta menjadi faktor penentu bagi wisatawan lain untuk berkunjung ataupun bagi wisatawan lama untuk berkunjung kembali. Proses pemanfaatan review tersebut dapat dilakukan dengan menilai aspek-aspek yang terdapat pada destinasi wisata berdasarkan ulasan dari pengunjung. Penelitian ini bertujuan untuk melakukan analisis sentimen berbasis aspek pada salah satu destinasi wisata di Indonesia yaitu Taman Nasional Bromo Tengger Semeru berdasarkan review dari pengguna Google Maps. Aspek yang digunakan diantaranya atraksi, fasilitas, akses, dan harga. Model klasifikasi sentimen yang digunakan adalah model machine learning yang terdiri dari SVM, Complement Naïve Bayes, Logistic Regression, dan transfer learning dari pre-trained BERT, IndoBERT dan mBERT. Berdasarkan hasil eksperimen, transfer learning dari model IndoBERT mencapai performa terbaik dengan akurasi dan F1-Score masing-masing sebesar 91.48% dan 71.56%. Selain itu, diantara model machine learning yang digunakan, model SVM memberikan hasil terbaik dengan nilai akurasi sebesar 89.16% dan F1-Score sebesar 62.23%.

Kata kunci*—Analisis Sentimen Berbasis Aspek, Google Maps Review, Machine Learning, Transfer Learning*

Abstract

Technology can influence and shape a person's behavior patterns when planning tours, traveling, and after traveling. Visitors' reviews can be used as evaluation material to improve the quality of tourist destinations and become a determining factor for other tourists to visit or revisit the destinations. The process of utilizing these reviews can be done by assessing the aspects of tourist destinations based on reviews from visitors. This study aims to conduct an aspect-based sentiment analysis on one of the tourist destinations in Indonesia, namely Bromo Tengger Semeru National Park, based on reviews of Google Maps users. The aspects consist of attractions, facilities, access, and price. The sentiment classification model used is a machine learning model consisting of SVM, Complement Naïve Bayes, Logistic Regression, and transfer learning from pre-trained BERT, IndoBERT, and mBERT. Based on the experimental results, transfer learning from the IndoBERT model achieved the best performance with accuracy and F1-Score of 91.48% and 71.56%, respectively. In addition, among the machine learning models

Received August 28th,2022; Revised January 31th, 2023; Accepted January 31th, 2023

used, the SVM model gives the best results with an accuracy of 89.16% and an F1-Score of 62.23%.

Keywords— *Aspect-Based Sentiment Analysis, Google Maps Review, Machine Learning, Transfer Learning*

1. INTRODUCTION

Indonesia is a country with various natural beauties that can attract the attention of tourists to visit, both domestic and foreign tourists. The increasing number of tourist visits makes tourism a priority sector to encourage economic growth by creating new jobs, opportunities to build businesses, foreign exchange earnings, and infrastructure development [1].

Efforts to improve the tourism sector cannot be separated from the existence of technology. According to the Ministry of Communication and Information of the Republic of Indonesia, technology influences and shapes a person's behavior in carrying out tourism activities, starting from planning a trip until after the trip [2]. For example, when someone wants to do tourism activities, they will make plans in advance by looking for related information about the tourist destinations they want to visit. The information sought can be in the form of activities or interesting things that can be found in tourist destinations, supporting transportation, accommodation that can be used on trips, or the necessary funds. When you are on tour, Google Maps can use to find out how to get to the place. Even when they have finished their tour, tourists can share their experiences or reviews of these tourist attractions through social media, which can be a determining factor for other tourists to visit or revisit the destinations.

Travel experiences or reviews of tourist destinations can be used for the advancement of tourist destinations by looking at the shortcomings and advantages of these tourist destinations as evaluation material or consideration in improving the quality of tourist destinations. The process of using a review to find out opinions about an object is called sentiment analysis.

Sentiment analysis is generally used only to identify whether the review given to an object has positive, negative, or neutral sentiments. This makes sentiment analysis insufficient in identifying sentiments for certain aspects of an object. Suppose the review given to an object has a positive sentiment. In that case, it does not mean that the author of the review has a positive sentiment for all aspects of the object being reviewed, as well as for negative sentiment. For this reason, it is necessary to carry out a more complete analysis by identifying aspects of the review and determining each sentiment. This analysis is referred to as aspect-based sentiment analysis.

Aspect-based sentiment analysis can be used to analyze reviews of tourist destinations since the reviews have many aspects with different sentiments. Research by M. Pontiki, et al. [3] discusses the subtasks that can be performed on aspect-based sentiment analysis, namely aspect term extraction, aspect term polarity, aspect category detection, and aspect category polarity. This study uses the 4th subtask, namely aspect category polarity, where the aspect category has been determined previously. This subtask aims to determine the polarity for each aspect that has been determined.

D. Arianto and I. Budi [4] have carried out research on aspect-based sentiment analysis in tourist destinations by taking advantage of reviews from Google Maps users on tourist destinations of Borobudur and Prambanan Temples with the aspect used were Attractions, Amenities, Accessibility (ability to access), Pictures, Price, and Human Resources. By utilizing five machine learning models, this study concludes that the Logistics Regression, Decision Tree, and Extra Tree methods provide a higher score than other methods.

In this study, researchers conducted an aspect-based sentiment analysis on one of the tourist destinations in Indonesia, namely Bromo Tengger Semeru National Park, based on reviews from Google Maps users. Bromo Tengger Semeru National Park (TNBTS) was chosen as a case study in this research because TNBTS is a tourist destination that is included in the UNESCO list as one of the world networks of biosphere reserves (The World Network of Biosphere Reserves) since 2015 [5] and also includes into the list of national tourism strategic areas [1]. Aspects used in this study are Attractions, Amenities, Access, and Price. The classification model used in this study is a machine learning model, namely SVM, Naïve Bayes, and Logistic Regression. In addition, researchers also use a transfer learning model from a pre-trained model that has been previously trained on a larger corpus that can classify text better than machine learning models. For the model evaluation stage, comparisons of accuracy, precision, recall, and F1-Score values are used to find out which model is better in classifying aspect-based sentiment in the Bromo Tengger Semeru National Park review data based on Google Maps users.

2. METHODS

This study uses an aspect-based sentiment analysis method using Machine Learning and Transfer Learning as the classification models. The stages to be carried out in this research are illustrated in Figure 1.

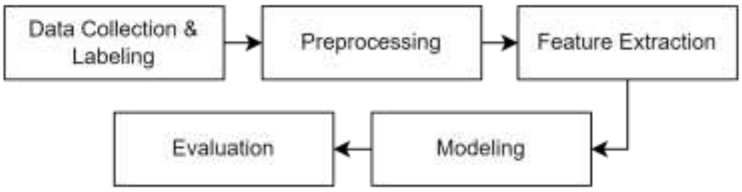

Figure 1 Research Method Flowchart

2.1 Data Collection and Labeling

The dataset used in this study is data obtained from scraping on Google Maps user reviews of one of the tourist destinations in Indonesia, namely Bromo Tengger Semeru National Park. An example of a Google Maps user review in Bromo Tengger Semeru National Park is illustrated in Figure 2. The reviews used in this study are reviews in Indonesian, English, and a mixture of Indonesian and English languages.

Figure 2 Examples of Google Maps User Reviews On Bromo Tengger Semeru National Park

After the data is collected, the next step is to label the data manually. To avoid subjectivity, the labeling is carried out by three annotators, with the final results of the labels being selected based on majority voting. Sentiment labeling is carried out for each review on aspects of Attractions, Amenities, Access, and Prices. The selection of attractions, amenities, and access aspects is based on the book Basic Knowledge of Tourism Science [6] which says that the main components that must be owned by a Tourist Destination Area (DTW) are attractions, amenities, access, and additional services. However, this study does not use additional service aspects. Based on the preliminary results of the sentiment label quality for additional service aspects in the annotated dataset, achieving Krippendorf's alpha value of 0.178 means that the data for this aspect is not consistent enough to be used as a dataset. In addition, the price aspect is chosen by considering its positive influence on customer satisfaction [7]. Sentiment labels used are "positive", "negative", and "neutral". In addition, for review data that does not contain pre-defined aspects, the sentiment will be classified as "none."

An example of data labeling is illustrated in Figure 3.

Review			
Bromo Tengger Semeru dapat dinikmati siapa saja, mulai dari anak muda hingga orang-orang yang sudah berusia lanjut. **Aksesnya mudah dijangkau**, kita tidak perlu mengeluarkan banyak energi untuk **menikmati berbagai keindahannya**. Sebagai bukti, dalam waktu setengah hari saja, kita sudah bisa menjangkau hampir semua objek wisata disana.			
Aspect			
attractions	amenities	access	price
positive	none	positive	none

Figure 3 Example of Data Labeling

After data labeling by several annotators, the quality of the labeling results was measured using Krippendorff's alpha as an inter-rater reliability test [8].

2. 2 Preprocessing

The data that has been collected and labeled will proceed to the preprocessing stage. Preprocessing is a technique used to clean the data. The preprocessing used in this study is as follows.

a. Lower Casing: converts all letters in the text into the same form, namely lowercase.
b. Cleaning: eliminate unnecessary characters such as emoji, username, hashtag, and URL.
c. Remove punctuation and extra spaces.

The preprocessing stage in this study does not use stopword removal because this process can eliminate some words that are important for classifying sentiments [9].

2.3 Feature Extraction

In this study, the feature extraction used to classify the model is using one of the word embedding methods, namely TF-IDF. TF-IDF is a word weighting method to calculate how important a word is in a collection of documents [10]. TF-IDF consists of Term Frequency (TF) and Inverse Document Frequency (IDF). TF means that the more often a term appears in the document, the greater its weight will be. While the IDF gives the meaning that the more often the term appears in several documents, the smaller the weight will be. Here's the formula for calculating TF-IDF [11]:

$$TF - IDF = tf_{ij} idf_i = tf_{ij} \times \log_2 \left(\frac{D}{df_i}\right) \tag{1}$$

where,

tf_{ij} : the number of i-words in j-document

D : total documents

df_i : the number of documents containing i-words

Below is an example of calculating TF-IDF in three simple documents or sentences, as follows:

D1: "Saya suka hujan."

D2: "Hari ini hujan."

D3: "Minggu adalah hari."

Examples of calculations are shown in Table 1.

Table 1 Example Of TF-IDF Calculation

Document Term	TF			DF	IDF	TF-IDF		
	D1	D2	D3			D1	D2	D3
saya	1	0	0	1	$\log(3/1) = 0.48$	0.48	0	0
suka	1	0	0	1	$\log(3/1) = 0.48$	0.48	0	0
hujan	1	1	0	2	$\log(3/2) = 0.18$	0.18	0.18	0
hari	0	1	1	2	$\log(3/2) = 0.18$	0	0.18	0.18
ini	0	1	0	1	$\log(3/1) = 0.48$	0	0.48	0
minggu	0	0	1	1	$\log(3/1) = 0.48$	0	0	0.48
adalah	0	0	1	1	$\log(3/1) = 0.48$	0	0	0.48

Feature extraction will only be used for machine learning models. Meanwhile, for transfer learning, the feature extraction stage is not carried out from the beginning but uses the token vector weight from the pre-trained model that has been trained previously using a larger dataset.

2. 4 Modeling

At this stage, the model is built using machine learning models, namely SVM, Complement Naive Bayes, and Logistic Regression, and transfer learning models from pre-trained BERT, IndoBERT, and mBERT.

2.4.1Machine Learning

a. SVM

Support Vector Machine (SVM) is a machine learning classification method based on statistical theory built from a limited amount of training data to get the best classification results [12].

b. Complement Naïve Bayes

Naive Bayes is a machine learning classification method based on probability theory [13] and Bayes theorem. Complement Nave Bayes (CNB) is a heuristic solution of Multinomial Nave Bayes (MNB). MNB estimates by calculating the probability of words appearing in the training data for one class, for example, c. In contrast, CNB calculates the probability of words appearing in all classes except c. Each class other than c is calculated for its probability, and the value with the smallest probability is taken, thus giving the meaning that the probability of class c has the highest value compared to classes other than c.

c. Logistic Regression

Logistic Regression is a machine learning classification method that is used to get the odds ratio on more than one explanatory variable. This method is similar to multiple linear regression, with the exception that the response variable is binomial [14]. This model generally has a good evaluation score and is close to the SVM score.

2.4.2 Transfer Learning

a. BERT

BERT is a two-way transformer model that was previously trained on a large corpus, namely the Toronto Book Corpus and Wikipedia [15].

b. IndoBERT

IndoBERT is one of the BERT models trained on a large Indonesian language corpus (Indo4B) which includes formal and non-formal languages such as Indonesian Wikipedia, news articles, social media, blogs, websites, and video recording subtitles [16].

c. mBERT

mBERT is a BERT model that is trained in 104 languages, including Indonesian. Indonesian language training data were taken from all Wikipedias in Indonesian, consisting of 532,806 articles [17].

2. 5 Evaluation

In this study, the evaluation of the model was carried out by comparing the values of precision, recall, accuracy, and F1-Score of machine learning and transfer learning models.

a. Precision

Precision is a comparison between true positives and the amount of data that is predicted to be positive.

$$precision = \frac{TP}{TP + FP} \tag{2}$$

b. Recall

Recall is a comparison between a true positive value with the number of data that is positive.

$$recall = \frac{TP}{TP + FN} \tag{2}$$

c. Accuracy

Accuracy is the ratio of true predictions or correctly predicted data on the whole data.

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{2}$$

d. F1-Score

F1-Score is the harmonic average of the precision and recall values.

$$F1 - Score = \frac{2 \times (recall \times precision)}{recall + precision} \tag{2}$$

The evaluation phase of the model in this study was carried out using the Cross Validation method. Cross Validation is a method used to select the appropriate model to make predictions. The data will be divided into two parts, one used to fit the model, and the other used to assess the model's ability to predict [18]. This study uses 5-fold cross-validation.

3. RESULTS AND DISCUSSION

The collected dataset of 1890 reviews of Bromo Tengger Semeru National Park was obtained by scraping Google Maps user reviews using the Selenium library. This review data consists of reviews in Indonesian, English, and a mixture of Indonesian and English.

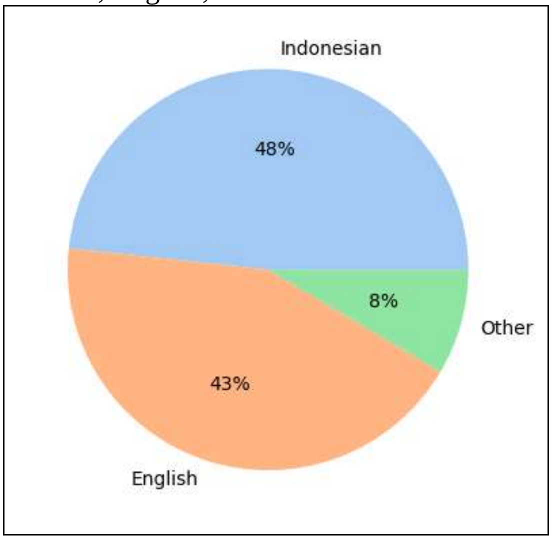

Figure 4 Language Percentage Based on Langdetect Library

Figure 4 shows the percentage of languages identified in the review data using the langdetect library. The results obtained from the language identification process stated that Indonesian had the largest percentage at 48% and followed by English at 43%. The rest were identified as several other languages, apart from Indonesian or English. There are several errors in the identification results due to the limitations of the model used by Langdetect. For example, there is a review in Indonesian but not formal, so the library identifies the review as not in Indonesian.

Furthermore, the labeling was done manually by three annotators by applying the majority voting system to produce 1660 labeled data which is illustrated in Figure 5. The reduced data occurred after the labeling process using the majority voting method because the labels that at least two annotators did not agree upon were not used in the analysis.

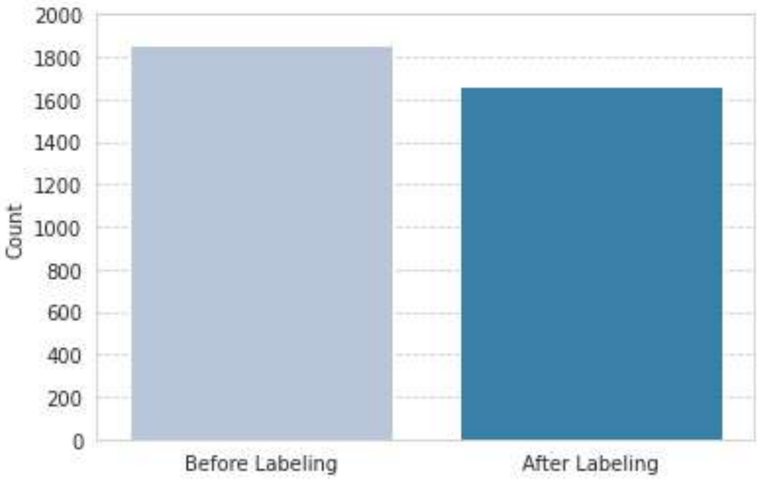

Figure 5 Comparison of the Amount of Data Before and After Labeling

In the sentiment distribution graph for each aspect shown in Figure 6, it can be seen that the aspect of tourist destinations that are often discussed in Google Maps user review data for tourist destinations of Bromo Tengger Semeru National Park are attractions, with the majority label being "positive". Meanwhile, other labels such as amenities, access, and prices are not discussed much in the review data, so the majority label is "none".

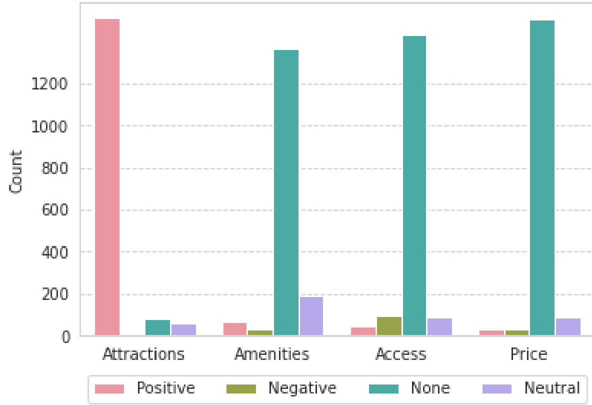

Figure 6 Sentiment Distribution for Each Aspect

To see the inter-rater reliability of the data that has been labeled, a comparison is made on the alpha value. The higher the alpha value, the higher the inter-rater reliability of the data.

Table 2 Inter-Rater Reliability Test

Aspect	Alpha
Attractions	0.509
Amenities	0.568
Access	0.518
Price	0.779

In Table 2, it can be seen that the labels on the attractions, amenities, and prices have an average alpha of 0.5 which means that the labels on these aspects are quite consistent. Likewise, the label on the price aspect has a fairly high alpha value of 0.779, which means that the label on that aspect is consistent.

In the preprocessing stage, by lowering the casing, cleaning, and removing punctuation and excess spaces, an example of the results from the preprocessing stage can be seen in Figure 7 below.

Process	Result
Data	Taman Nasional Bromo Tengger Semeru 😍 mantap 👍 👍 #khentravelindo | paket wisata Bromo ID
Lower casing	taman nasional bromo tengger semeru 😍 mantap 👍 👍 #khentravelindo | paket wisata bromo ID
Cleaning, remove punctuation & white space	taman nasional bromo tengger semeru mantap paket wisata bromo

Figure 7 Example of Preprocessing Results

The next step is feature extraction using the TF-IDF method using the TfidfVectorizer library from Sklearn. The TF-IDF hyperparameters for each machine learning model were obtained from the grid search.

At the model development stage, machine learning classification models are used, namely SVM, Complement Naive Bayes, and Logistic Regression which are available in the Scikit-learn or Sklearn library. All machine learning models use hyperparameters based on grid search results. Then for the development of transfer learning models from pre-trained models BERT Cased, BERT Uncased, IndoBERT Uncased, mBERT Cased, and mBERT Uncased, the Python library and transformers were used using Adam as an optimization algorithm and equipped with a batch size of 10 and a learning rate of 1×10^{-5}.

Table 3 Experimental Results Of Sentiment Classification Model Based On Aspects Using 5 Fold Cross-Validation

Model	Precision	Recall	Accuracy	F1-Score
Machine Learning				
SVM	**0.7006**	**0.5951**	0.8916	**0.6223**
Complement Naive Bayes	0.5561	0.5860	0.8142	0.5515
Logistic Regression	0.6911	0.5869	**0.8949**	0.6127
Transfer Learning				
BERT Cased	0.6852	0.5540	0.8904	0.5683
BERT uncased	0.6353	0.5605	0.8952	0.5718
IndoBERT	**0.7478**	**0.7016**	**0.9148**	**0.7156**
mBERT Cased	0.6417	0.5612	0.8889	0.5699
mBERT uncased	0.6508	0.5961	0.9000	0.6087

The experimental results of the built model are shown in Table 3. Based on Table 3, it can be seen that all fine-tuned models which are transfer learning from pre-trained models have higher F1-Scores than machine learning models. This is because the machine learning model only learns from the training data used, while the transfer learning model has been previously trained on a large language corpus.

Of all the fine-tuned models, the IndoBERT classification model has the higher evaluation score. The accuracy values and F1-Score are 91.48% and 71.56%, respectively. Based on the corpus used by each model, Multilingual BERT (mBERT) is more suitable for multilingual data than the IndoBERT model because the mBERT model has been trained on 104 language corpora, including Indonesian. However, in this study, the IndoBERT model achieved the best performance. This can happen because the IndoBERT model was trained on several Indonesian-speaking corpus, both in the form of formal and non-formal language, which is more suitable to be applied to the data used in this study. Google Maps user review data is generally in the form of non-formal language. In addition, the data used has a higher percentage of data in Indonesian than in English. Therefore, the IndoBERT model is more suitable for use in this case than the Cased and Uncased mBERT models. This also applies to the Cased and Uncased BERT models whose data are trained on an English-language corpus. These results are in line with the results given in the study of B. Wilie et al. [16], namely IndoBERT provides better performance than mBERT.

The distribution of scores from the classification results of each sentiment for each aspect of the IndoBERT fine-tuned model can be seen in Table 4.

Table 4 Results Of Aspect Classification For Each Label On Indobert Using 5-Fold Cross-Validation

Aspect		Precision	Recall	F1-Score	Support
Attractions	**Positive**	**0.9459**	**0.9899**	**0.9670**	**153**
	Neutral	0.6133	0.4150	0.3968	5
	Negative	0	0	0	1

	None	0.4000	0.1581	0.2204	7
Amenities	Positive	0.2571	0.0618	0.0756	7
	Neutral	0.5271	0.6502	0.5662	21
	Negative	0.4000	0.1467	0.2143	3
	None	**0.9336**	**0.9364**	**0.9480**	**135**
Access	Positive	0.6000	0.5022	0.4839	5
	Neutral	0.5809	0.3168	0.3968	9
	Negative	0.6230	0.4909	0.5023	10
	None	**0.9300**	**0.9851**	**0.9565**	**142**
Price	Positive	0.4000	0.1733	0.2267	3
	Neutral	0.7681	0.7411	0.6965	10
	Negative	0.3667	0.2400	0.2467	3
	None	**0.9684**	**0.9975**	**0.9825**	**151**

In Table 4, it can be seen that the aspect-based sentiment classification using the IndoBERT fine-tuned model produces various F1-Scores. Some achieved very high scores, such as the attraction aspect with a “positive” label, as well as the amenities, access, and price aspects for the “none” label, each of which had a score of more than 90%. And there is also a label on the aspect that has the lowest score reaching 0%. This is caused by the imbalance of available datasets for each aspect and sentiment label, as shown in the support column in Table IV.

In the machine learning classification model, SVM achieves the best F1-Score value compared to other machine learning models. The experimental results for the SVM classification model for each aspect and sentiment label can be seen in Table 5. In general, the sentiment label scores for each aspect have a similar pattern to the IndoBERT fine-tuned results due to data imbalances.

Table 5 Results Of Aspect Classification For Each Label On SVM
Using 5-Fold Cross-Validation

Aspect		Precision	Recall	F1-Score	Support
Attractions	Positive	**0.9154**	**0.9863**	**0.9491**	**302**
	Neutral	0.3667	0.1005	0.1416	13
	Negative	0.000	0.000	0.000	1
	None	0.2800	0.0829	0.1119	16
Amenities	Positive	0.2349	0.0948	0.1151	13
	Neutral	0.5576	0.5285	0.5273	39
	Negative	0.2000	0.0286	0.0500	6
	None	**0.9045**	**0.9651**	**0.9333**	**274**
Access	Positive	0.7000	0.3425	0.4137	9
	Neutral	0.1952	0.1248	0.1483	17
	Negative	0.4291	0.2695	0.3176	19
	None	**0.7947**	**0.8368**	**0.9394**	**287**
Price	Positive	0.2000	0.0400	0.0667	7
	Neutral	0.5589	0.3358	0.3896	18
	Negative	0.5000	0.1119	0.1800	6
	None	**0.9368**	**0.9895**	**0.9623**	**301**

In Figures 8, you can see an example of a review with a gold label or the actual label given by the annotator manually, along with the prediction results using the SVM and IndoBERT models. Review example 1 in Figure 8 shows that the IndoBERT model produces the same prediction as the gold label, while the SVM model makes an error where the model does not succeed in identifying the label on access.

Review			
Tempat yang keindahannya tak kan terlupakan, walau di perlukan perjuangan untuk mencapai puncak penanjakan tp semua tergantikan saat melihat matahari baru. Bagi pengunjung yg melalui jalur tumpang tidak disarankan untuk membawa motor matic apalagi saat musim kemarau,karena akan menyulitkan diri sendiri di lautan pasir.			
SVM Prediction Results			
attractions	amenities	access	price
positive	none	none	none
IndoBERT Prediction Results			
attractions	amenities	access	price
positive	none	negative	none
Gold Label			
attractions	amenities	access	price
positive	none	negative	none

Review			
Pemandangan alam yang menajubkan.. Saya baru pertama kali kesini.. Ini pengalaman luar biasa bagi saya pribadi.. Padang pasir luas.. Padang rumput juga ada nan hijuau yang luas.. Dan warga masyarakat khususnya disekitar bromo yang sangat ramah kepada wisatawan.. Tapi ada hal yang kurang untuk kamar mandinya yang dekat parkiran jeep sebelum ke kawah.. Kurang bersih..			
SVM Prediction Results			
attractions	amenities	access	price
positive	none	none	none
IndoBERT Prediction Results			
attractions	amenities	access	price
positive	neutral	none	none
Gold Label			
attractions	amenities	access	price
positive	negative	none	none

Figure 8 Review Example 1 & 2, Gold Label, and Prediction Label Using SVM and IndoBERT

From this example, it can be concluded that the IndoBERT model can capture sentence context better than SVM. However, there are still some errors in IndoBERT as shown in example 2, where the IndoBERT model gives a “neutral” prediction on the facility aspect while the gold label is “negative”.

4. CONCLUSIONS

Based on the discussion that has been described above, several conclusions are obtained from a review of data obtained from Google Maps. The "positive" label is mostly found in the Attraction aspect. As for the other aspect, it has the "none" label as the majority of the labels. The results obtained from the analysis of model development in this study are the transfer learning model of the pre-trained model is better than the machine learning model. Then, the best model produced in this study is the IndoBERT model, which is refined with an accuracy value of 91.48% and an F1-Score of 71.56%.

REFERENCES

[1] Badan Pusat Statistik (BPS), “Statistik Wisatawan Nusantara 2020,” BPS RI, Jakarta, 2021.

[2] Kominfo. (2019, April) Pentingnya Teknologi dalam Sektor Pariwisata [Online]. Available: https://aptika.kominfo.go.id/2019/04/pentingnya-teknologi-dalam-sektor-pariwisata/

[3] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, “SemEval-2014 Task 4: Aspect Based Sentiment Analysis,” in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, pp. 27-35, August 2014.

[4] D. Arianto and I. Budi, “Aspect-based Sentiment Analysis on Indonesia’s Tourism Destinations Based on Google Maps User Code-Mixed Reviews (Study Case: Borobudur and Prambanan Temples),” in Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation, Hanoi, Vietnam: Association for Computational Linguistics, 24-26 October 2020, pp. 359-367.

[5] UNESCO. (2015) Bromo Tengger Semeru-Arjuno Biosphere Reserve, Indonesia [Online]. Available: https://en.unesco.org/biosphere/aspac/bromo-tengger-semeru-arjuno
[6] I. K. Suwena and I. G. N. Widyatmaja, Pengetahuan Dasar Ilmu Pariwisata. Denpasar: Pustaka Larasan, 2017.
[7] V. R. Khansa, and N. Farida, "Pengaruh Harga Dan Citra Destinasi Terhadap Niat Berkunjung Kembali Melalui Kepuasan (Studi pada Wisatawan Domestik Kebun Raya Bogor)," Jurnal Ilmu Administrasi Bisnis, vol. 5, no. 4, pp. 104-114, Jun. 2016.
[8] K. Krippendorff and R. Craggs, “The Reliability of Multi-Valued Coding of Data,” Communication Methods and Measures, vol. 10, no. 4, pp. 181-198, 2016.
[9] C. H. Yutika, Adiwijaya and S. A. Faraby, “Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes,” Jurnal Media Informatika Budidarma, vol. 5, no. 2, pp. 422-430, April 2021.
[10] I. M. Yulietha, S. A. Faraby and Adiwijaya, “Klasifikasi Sentimen Review Film Menggunakan Algoritma Support Vector Machine Sentiment Classification of Movie Reviews Using Algorithm Support Vector Machine,” e-Proceeding of Enginering, vol. 4, no. 3, pp. 4740-4750, Desember 2017.
[11] B. Trstenjak, S. Mikac, and D. Donko, “KNN with TF-IDF Based Framework for Text Categorization,” 24th DAAAM International Symposium on Intelligent Manufacturing and Automation 2013, pp. 1356-1364, 2014.
[12] V. N. Vapnik, “Statistical learning theory,” Springer, 1995.
[13] W. Zhang and F. Gao, “An Improvement to Naive Bayes for Text Classification,” Procedia Engineering, vol. 15, pp. 2160-2164, 2011.
[14] S. Sperandei, “Understanding logistic regression analysis,” Biochemia Medica, pp. 12-18, 2014.
[15] J. Devlin, M. Chang, L. Kenton, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” 2019.
[16] B. Wilie, K. Vincentio, G. I. Winata, S. Cahayawijaya, X. Li, Z. Y. Lim, S. Soleman, R. Mahendra, P. Fung, S. Bahar, and A. Purwarianti, “IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding,” CoRR, 2020.
[17] A. N. Azhar and M. L. Khodra, "Fine-tuning Pretrained Multilingual BERT Model for Indonesian Aspect-based Sentiment Analysis," 2020 7th International Conference on Advance Informatics: Concepts, Theory, and Applications (ICAICTA), pp. 1-6, 2020.
[18] J. Shao, “Linear model selection by cross-validation,” Journal of the American Statistical Association, vol. 88, no. 422, pp. 486-494, 1993.