A Technique for Tourism Visitor Feedback Collection Based on Google Maps Reviews Using Text-Mining

Achmad Maududie[1,a)], Aula Fajrun Khalilurahman[1,b)], Priza Pandunata[1,c)]

[1] *Faculty of Computer Science, Jember University Indonesia*

[a)] *maududie@unej.ac.id*
[b)] *192410101044@mail.unej.ac.id*
[c)] *priza@unej.ac.id*

Abstract. Jember Regency has enormous tourism potential which has the potential to be developed into a potential superior area. However, developing this potential is not easy considering that feedback on satisfaction with the services provided at tourist locations is also not easy to obtain. This research aims to develop a technique for collecting feedback, especially for services that are felt to be lacking, using a text-mining approach based on visitor reviews on Google Maps. This technique is based on Support Vector Machine as a sentiment classification algorithm, Latent Dirichlet Allocation (LDA) to detect topics, and Perceptron Tagger as a POS Tagger algorithm to recognize services that are considered lacking. The evaluation of this technique is promising because the classification and POS Tagger model in this research have a high level of accuracy, i.e., 92% and 96.6%, so they are good enough to be implemented. It is shown that based on review data from Google Maps from 2020 to 2023 in eight tourism areas in Jember Regency the entrance ticket price for Papuma Beach and Jember Mini Zoo is considered expensive, while Watu Ulo Beach, Payangan Beach, Pancer Beach, and Sukorambi Botanical Gardens are considered the tourist attractions are dirty, and Pancer Beach and the Gunung Gambir Tea Garden are difficult to access.

Keywords: Tourism feedback analysis, text mining in tourism, sentiment and topic detection

INTRODUCTION

Tourism has become one of the priority sectors and is expected to be one of the main drivers to accelerate economic growth in Indonesia [1]. The tourism potential in Jember is huge, but so far it has not been maximized due to a lack of creativity and development from the community and government [2]. Criticism, suggestions, and input from people who reside and have visited these tourist attractions are very important for the development of these tours [3]. To get criticism and suggestions is sometimes difficult because it requires interviews or giving feedback through a criticism and suggestion box which is often not filled in by visitors. Nevertheless, visitors to nature tourist sites often submit comments about the places they have visited through existing digital platforms, one of them is Google Maps. On Google Maps there are reviews and ratings feature that can be used as an indicator to support the quality and popularity of a tourist spot [4].

Several previous studies have been conducted related to this research. The first research was conducted by [9] with the title "Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning". There are 3 methods used, namely Decision Tree, Random Forest, and Support Vector Machine. This research uses manual labeling into positive, neutral, and negative as well as visited and not visited. It was found that Support Vector Machine is the most optimal in performing sentiment analysis. Further research was conducted by [10] entitled "Exploring customer concerns on service quality under the COVID-19 crisis: A social media analytics study from the retail industry". In this study, Latent Dirichlet Allocation (LDA) was used to find topics from the tweets so that researchers could evaluate customer concerns and attitudes when shopping for groceries in the UK during the pandemic crisis. Further research was conducted by [11] with the title "Part-of-speech (POS) tagging using conditional random field (CRF) model for Khasi corpora". Training and validation data are divided in a ratio of 80:20. This research compares several other models such as Bi-Gram, Tri-Gram and a combination of Bi-Gram and Tri-Gram. Evaluation was done using precision, recall, and F1-score. It was found that the CRF model gave better results with an accuracy rate of 92.12% than other models.

The main objective of this research is to develop a technique for collecting feedback, especially for services of tourist areas that are felt to be lacking (negative comments), using a text-mining approach based on visitor reviews on

I. H. Agustin (ed.), *Proceedings of the 2nd International Conference on Neural Networks and Machine Learning 2023 (ICNNML 2023)*, Advances in Intelligent Systems Research 183,
https://doi.org/10.2991/978-94-6463-445-7_9

reviews on Google Maps. Based on references to several previous studies, Support Vector Machine will be used as a classification method because it has high accuracy in the first study, Latent Dirichlet Allocation is used as a method to find topics contained in a sentence as in the second study, and the use of POS Tagging in the third study helps provide explanations related to words included in nouns and adjectives used as objects of improvement in this study.

METHOD

For this research, the author divided the method into several phases. The first phase is for generating models that are used to get negative data results. The first phase consists of 2 tasks, namely preparing data and creating classification models and POS tagger models.

Web Scraping

The data used in this research is review data on Google Maps. The technique used is web scraping. The data is taken in the form of ratings, review times, and reviews. The library used to do web scraping is Selenium. There are several stages when doing web scraping using Selenium, namely.

1. Determine the tourist attractions that will be used for scraping.
2. Studying the structure of the Google Maps review panel.
3. Getting the PATH of user reviews stored in the Google Maps review panel.
4. Extracting reviews for all tourist attractions using scraping technique.

Data Labeling

The data is labeled based on the ratings found on Google Maps. The rating on Google Maps is 1 star–5 stars. Based on the results of discussions with experts and by looking at several sample reviews obtained, 1 star and 2 stars are included in the negative class. While 3 to 5 stars go into non-negative.

Data Preprocessing

Data preprocessing is the initial stage in processing data. The data processed is review data from the Google Maps data source so that with preprocessing the resulting data will be better. Before being processed at a further stage, the data will be normalized first so that the data changes to become more standardized. Data is normalized using an alay dictionary [5]. The sub-processes carried out include the following.

Data Cleaning. The data cleaning stage is a stage to delete empty and duplicate data. At this stage, data with capital letters will also be converted to lowercase.

Stopword Removal. In the filtering stage, the set of words will be filtered using stopwords. If there are words included in the stopwords dictionary, then the word will be deleted.

Stemming. Stemming used in this stage is to convert words into their basic words. The stages for converting into basic words are removing prefixes (prefix words) and surfixes (suffix words) in the word.

Tokenizing. Tokenizing is a stage to separate document data in a sentence into a token. Before being broken down into tokens, characters other than the alphabet such as punctuation marks, spaces, tabs, and so on are removed.

Data Splitting

Based on previous research, the data is divided into 80% training data and 20% testing data.

Term Wighting

The term weighting stage aims to give value to each word so that it can be processed further. Term weighting in this research uses TF-RF weighting. TF-RF weighting consists of several processes, namely Term Frequency (TF) is the frequency of how often a word appears in a document. Furthermore, Relevance Frequency (RF) is a weighting that considers the distribution of relevant documents. TF-RF is a combined method of TF and RF so that the weighting considers the relevance between documents seen from the frequency of occurrence of terms in related categories [6]. The form of TF-RF calculation is as in equation (1).

$$TF_{td}RF = TF_{td} * RF$$
$$TF_{td}RF = TF_{td} * log(2 + \frac{b}{(1,c)}) \quad (1)$$

Description :
$TF_{td}RF$: TF-RF weighting
TF_{td} : Number of occurrences of word t in the document
b : Number of documents containing word t
c : Number of documents that do not contain word t

Normalization

This research uses Z-Score normalization or standard normalization. Normalization is to make all variables have a uniform scale range by changing the values to be centered around zero [7]. The calculation form of Z-Score normalization is as in equation (2).

$$v' = (v - \overline{A})/\sigma_A \quad (2)$$

Description :
v' : New Value
v : Old Value
$\overline{A}$: Average Attribute A
σ_A : Standard deviation of attribute A

Classification Model

Classification using Support Vector Machine will use a Linear kernel. After the classification is done, the accuracy will be calculated using the confusion matrix. If the accuracy result is good, the data will be re-divided into 8 tourist attractions to continue searching for topics.

POS Tagger Model

Tagger train aims to train the tagger that will be used in this research. There are 5 taggers that will be tested, namely Ngram Tagger, Classifier Tagger, TnT Tagger, Perceptron Tagger and CRFTagger. The dataset used comes from idn-tagged-corpus-CSUI [8] containing a corpus that uses tab-separated file (.tsv) format. Each line in this dataset contains a token along with the part-of-speech tag of the token separated by one tab character (\t). Sentences are separated by one blank line. The data will be supplemented with data from the author by manually tagging with a linguist named Nalendra Satyatama, S. Hum. A teacher at a private high school and a graduate of the Faculty of Letters, University of Indonesia. Evaluation will be done using confusion matrix by looking at the accuracy and F1-Score results and conducting experiments using several sentences so that the author can see firsthand the performance of the tagger.

Topic Detection

Topics will be detected using Latent Dirichlet Allocation (LDA). LDA is done by calculating the probability of each word in each tourist spot.

Keyword Extraction

The results of topic detection will bring up words that are sorted according to their rank. The words will be searched for nouns and adjectives so that the results will appear in the form of the first noun and adjective after the

noun. The search in this research will use POS Tagging. This research will use perceptron tagger which has the highest model accuracy.

Feedback Detection and Expert Judgement

The tagged word will be interpreted according to the number of occurrences of the word in the keyword extraction result.

Results and Discussion

Web Scraping Result

Web scraping is done using selenium library. Data is taken from reviews contained in Google Maps of several tourist attractions in Jember Regency, namely Watu Ulo Beach, Papuma Beach, Pancer Beach, Payangan Beach, Sukorambi Botanical Garden, Gunung Gambir Tea Garden, Jember Mini Zoo and Gunung Pasang Jember Agrotourism. The total review data that is still raw and has empty data is 9120 data.

TABLE 1. Web scraping results

rating	review	duration_list
	Tadinya mau ke papuma, karna kurang ada info dari petugas terkait arah jalan ke papuma malah jadi bayar karcis buat ke watu ulo sangat sssangaaat mengecewakan	
1 bintang	ðŸ‘ŽðŸ‘ŽðŸ‘Ž â€¦	5 hari lalu
4 bintang		2 minggu lalu
5 bintang	bagus	2 minggu lalu

Preprocessing Data

Data Cleaning. The application of data cleaning results in a change in the amount of data due to the removal of empty columns. In addition, data cleaning will perform case folding.

TABLE 2. Data cleaning results

review	cleaning
Tadinya mau ke papuma, karna kurang ada info dari petugas terkait arah jalan ke papuma malah jadi bayar karcis buat ke watu ulo sangat sssangaaat mengecewakan ðŸ‘ŽðŸ‘ŽðŸ‘Ž â€¦	tadinya mau ke papuma karena kurang ada info dari petugas terkait arah jalan ke papuma malah jadi bayar karcis buat ke watu ulo sangat sangat mengecewakan

Stopword Removal. After cleaning, the data will be filtered based on Stopwords. Stop-words are words that are very common and have no or less meaning than other keywords.

TABLE 3. Stopword removal results

cleaning	Stopword removal
tadinya mau ke papuma karena kurang ada info dari petugas terkait arah jalan ke papuma malah jadi bayar karcis buat ke watu ulo sangat sangat mengecewakan	mau papuma kurang info petugas terkait arah jalan papuma jadi bayar karcis ke watu ulo sangat mengecewakan

Stemming. Stemming aims to reduce the number of different types of words in the text and may increase some of the frequency of occurrence of the word.

TABLE 4. Stemming results

Stopword removal	stemming
mau papuma kurang info petugas terkait arah jalan papuma jadi bayar karcis ke watu ulo sangat mengecewakan	mau papuma kurang info tugas kait arah jalan papuma jadi bayar karcis ke watu ulo sangat kecewa

Tokenizing. After successfully converting the sentence into basic words, perform tokenization so that the sentence changes into token form.

TABLE 5. Tokenizing results

stemming	tokenizing
mau papuma kurang info tugas kait arah jalan papuma jadi bayar karcis ke watu ulo sangat kecewa	['mau', 'papuma', 'kurang', 'info', 'tugas', 'kait', 'arah', 'jalan', 'papuma', 'jadi', 'bayar', 'karcis', 'ke', 'watu', 'ulo', 'sangat', 'kecewa']

Web Scraping Result

Term weighting or weighting aims to give value to each word so that it can be processed further. The term weighting used is TF-RF weighting.

TABLE 6. Term weighting result

kata	nilai TF-RF
Mau	0.3026332174752666
Papuma	0.6116955698088149
Kurang	0.3049351038757156
Info	0.3012959716742768
Tugas	0.30214146292852173
Kait	0.3012959716742768

First, the TF must be calculated first. Then RF will be calculated. The results of TF and RF will be multiplied so that the results are obtained as in the table. For example, the word "mau" has a TF-RF value of 0.3026332174752666 which is obtained from the TF value = 1 and RF value = 0.3026332174752666.

Normalization

The normalization used is Z-Score normalization or standard normalization. Normalization is used to equalize the value of the data so that it is not too far apart and will not bias the results of the classification.

TABLE 7. Normalization Result

kata	nilai TF-RF	z-score
Kali	0.3026780050083807	11.467934
Tarik	0.3052994951275524	6.425113
Saran	0.3032615003264074	9.851903

After the TF-RF results are known, the results of the weight per word will be normalized so as to get the results as in the table. For example, the word "saran" has a TF-RF weight value = 0.3032615003264074. After normalization, the value changes to 9.851903.

Data Classification Result

Data classification is done using SVM (Support Vector Machine). Data is divided into training data and testing data. The data division uses 80% training data and 20% testing data. Linear kernels are used because the data used has so many features that the use of linear kernels is still one of the best options than other kernels.

TABLE 8. Classifying Google Maps Data Reviews

review	tokens	class
Paling gak jelas 2 kali di tarik,saran kalo orang mau ke papuma jangan ditarikin,jadi double bayar 2 kalo 20k di papuma 55k non tunai,gak apa :) kan kalo ke watu ulo belok kiri,nah nariknya pas belokan,jangan di tengah jalan mau ke papuma -_	['kali', 'tarik', 'saran', 'kalo', 'orang', 'papuma', 'ditarikin', 'double', 'bayar', 'kalo', 'di', 'papuma', 'non', 'tunai', 'enggak', 'apa', 'kalo', 'ke', 'watu', 'ulo', 'belok', 'kiri', 'nariknya', 'pas', 'belok', 'di', 'jalan', 'ke', 'papuma']	Negative
Tadinya mau ke papuma, karna kurang ada info dari petugas terkait arah jalan ke papuma malah jadi bayar karcis buat ke watu ulo sangat sssangaaat mengecewakan	['mau', 'papuma', 'kurang', 'info', 'tugas', 'kait', 'arah', 'jalan', 'papuma', 'jadi', 'bayar', 'karcis', 'ke', 'watu', 'ulo', 'sangat', 'kecewa']	Negative
bagus	['bagus']	Non-negative

After classification, take the negative data or rating that is worth 0 and will be processed further. The evaluation results are carried out using the confusion matrix and get a model accuracy of 92%.

TABLE 9. Extracting Negative Label Only

review	tokens	class
Paling gak jelas 2 kali di tarik,saran kalo orang mau ke papuma jangan ditarikin,jadi double bayar 2 kalo 20k di papuma 55k non tunai,gak apa :) kan kalo ke watu ulo belok kiri,nah nariknya pas belokan,jangan di tengah jalan mau ke papuma -_	['kali', 'tarik', 'saran', 'kalo', 'orang', 'papuma', 'ditarikin', 'double', 'bayar', 'kalo', 'di', 'papuma', 'non', 'tunai', 'enggak', 'apa', 'kalo', 'ke', 'watu', 'ulo', 'belok', 'kiri', 'nariknya', 'pas', 'belok', 'di', 'jalan', 'ke', 'papuma']	Negative
Tadinya mau ke papuma, karna kurang ada info dari petugas terkait arah jalan ke papuma malah jadi bayar karcis buat ke watu ulo sangat sssangaaat mengecewakan	['mau', 'papuma', 'kurang', 'info', 'tugas', 'kait', 'arah', 'jalan', 'papuma', 'jadi', 'bayar', 'karcis', 'ke', 'watu', 'ulo', 'sangat', 'kecewa']	Negaitve

POS Tagger Model Result

There are several taggers used in this research namely Ngram Tagger, Classifier Tagger, TnT Tagger, Perceptron Tagger and CRFTagger. The data is divided into 80% training data and 20% testing data. Evaluation is done using confusion matrix.

TABLE 10. POS Tagger Model Result

tagger	accuracy score	f1-score
Ngram Tagger	0.9375145676326626	0.9490659066840355
Classifier Tagger	0.9421567865744698	0.9429991302143553
TnT Tagger	0.9380001553880817	0.9496166245950404
Perceptron Tagger	0.966494444876078	0.9660803019450952
CRFTagger	0.9538108927045296	0.9530943345181548

Based on the training data evaluation results, Perceptron Tagger has the highest accuracy with 96.6%. The author experiments by using several sentences in Google Maps review data and sentences made by the author to see the accuracy of tagging directly on a sentence. Ngram Tagger, TnT Tagger, and Perceptron Tagger produce better results such as the word "kotor" which means adjective, is in accordance with its tag, namely "JJ". While in CRFTagger, the word "kotor" is classified into "NN" which means noun. Based on the accuracy results and experiments, the tagger that will be used for this research is Perceptron Tagger.

Topic Detection

Topic detection is done to find out what are the main topics in 1 tourist spot. If there is no data on a tourist spot, the topic will not be displayed. This research takes 1 topic for each tourist spot with each containing 10 words. After the topic is found, then take the nouns that are in the topic of each tourist spot.

TABLE 11. Topic Detection Result

group	topic	noun in topic
0	pantai, sampah, bagus, kotor, ada, ulo, masuk, mahal, pasir, watu	watu, ulo, sampah, pasir, pantai
1	pantai, ada, mahal, tiket, bayar, masuk, sampah, indah, banget, bagus	tiket, sampah, pantai
2	ada, pantai, sampah, unjung, payang, kotor, sana, masuk, indah, istimewa	sampah, payang, indah, unjung, istimewa, pantai
3	sampah, kotor, pantai, tempat, ada, bagus, sayang, masih, jalan, parkir	sampah, tempat, parkir, jalan, pantai
4	ada, kolam, anak, mahal, tempat, renang, masuk, unjung, jalan, lokasi	lokasi, tempat, anak, jalan, unjung, renang, kolam
5	jalan, ada, bagus, tempat, sejuk, cantik, istimewa, asli, oleh, tolong	tolong, tempat, sejuk, istimewa, jalan
6	ada, istimewa, sejuk, masuk, sungai, wisata, tempat, cantik, parkir, alami	wisata, cantik, sungai, tempat, alami, sejuk, istimewa, parkir

7	ada, anak, tempat, di, yang, bagus, hewan, makan, mahal, pandu	hewan, tempat, pandu, anak

In **TABLE 11**, there is a group column that means the name of the tourist spot. The topic column is used to accommodate any topics found at the tourist attractions. The noun in topic column is used to accommodate nouns that are in the topic of each tourist spot. In group 0, which means Watu Ulo Beach, the topic is found in the form of "pantai, sampah, bagus, kotor, ada, ulo, masuk, mahal, pasir, watu". Of the 10 words that are nouns are "watu, ulo, sampah, pasir, pantai".

Keywords Extraction

After finding the topic, keyword extraction is performed using the results of the nouns in the topic. If there is a noun in the review that is the same as the noun in the topic, then the noun will be retrieved and become the object. Once the noun is found, take the first adjective after the noun. This adjective will be the information about the object/noun.

TABLE 12. Keywords Extraction

sentence	noun	adjective
pantai watu ulo ombak besar umum pantai selatan hampar pasir main tepi pantai tidak saran mandi renang deret batu karang bentuk punggung ular cukup	pantai	Adjective not found
pantai watu ulo ombak besar umum pantai selatan hampar pasir main tepi pantai tidak saran mandi renang deret batu karang bentuk punggung ular cukup	watu	Adjective not found
pantai watu ulo ombak besar umum pantai selatan hampar pasir main tepi pantai tidak saran mandi renang deret batu karang bentuk punggung ular cukup	ulo	umum
pantai watu ulo ombak besar umum pantai selatan hampar pasir main tepi pantai tidak saran mandi renang deret batu karang bentuk punggung ular cukup	pantai	Adjective not found
pantai watu ulo ombak besar umum pantai selatan hampar pasir main tepi pantai tidak saran mandi renang deret batu karang bentuk punggung ular cukup	pasir	Adjective not found
pantai watu ulo ombak besar umum pantai selatan hampar pasir main tepi pantai tidak saran mandi renang deret batu karang bentuk punggung ular cukup	pantai	Adjective not found
ombak gede banding papuma ikon batu bentuk ular karena ombak gede ada tulis boleh foto atas batu overall bagus banget pantai although memang kotor ya	pantai	kotor

For example, in the sentence "ombak gede banding papuma ikon batu bentuk ular karena ombak gede ada tulis boleh foto atas batu overall bagus banget pantai although memang kotor ya" there is a noun "pantai" and after the noun there is an adjective "kotor", so the resulting phrase will be "pantai kotor".

Feedback Detection and Expert Judgement

Based on the keywords extraction results, the number of objects that appear in the keywords extraction results is counted. The results are sorted based on the highest number.

a) Pantai Watu Ulo

TABLE 13. Feedback Detection Pantai Watu Ulo

group	phrase	count
0	pantai kotor	2
0	pasir hitam	2
0	pantai asli	1
0	pantai bagus	1
0	pantai bersih	1
0	pantai hitam	1
0	pantai indah	1
0	pantai negatif	1
0	pantai tua	1
0	sampah kotor	1
0	ulo lepas	1
0	ulo murah	1

Based on the results of keywords extraction, on Watu Ulo Beach several words are formed in the form of dirty beaches, black beaches, negative beaches so that objects that need attention are dirty beaches.

b) Pantai Papuma

TABLE 14. Feedback Detection Pantai Papuma

group	phrase	count
1	pantai indah	4
1	tiket mahal	4
1	pantai bagus	3
1	pantai populer	2
1	pantai bersih	1
1	pantai mulus	1
1	pantai nyata	1
1	pantai salah	1
1	pantai terjal	1
1	sampah minimal	1
1	tiket bagus	1
1	tiket indah	1
1	tiket luas	1

Based on the results of keywords extraction, Papuma Beach brings up results such as beautiful beaches, expensive tickets, good beaches, and so on. Based on the results obtained, the object that needs attention is expensive tickets.

c) Pantai Payangan

TABLE 15. Feedback Detection Pantai Payangan

group	phrase	count
2	istimewa indah	1
2	pantai bersih	1
2	pantai kotor	1
2	pantai penting	1
2	payang keras	1
2	payang tepat	1
2	unjung biasa	1

Based on the results of keywords extraction, Payangan Beach brings up the results of dirty beaches and several other words. Based on the results obtained, the object that needs attention is the cleanliness of the beach.

d) Pantai Pancer

TABLE 16. Feedback Detection Pantai Pancer

group	phrase	count
3	pantai kotor	5
3	jalan hancur	1
3	jalan rosak	1
3	pantai bagus	1
3	pantai bersih	1
3	pantai lepas	1
3	pantai penuh	1
3	sampah bagus	1
3	sampah kotor	1
3	sampah lokal	1
3	sampah strategis	1
3	tempat bagus	1
3	tempat utama	1

Based on the results of keywords extraction, Pancer Beach brings up results such as dirty beaches, destroyed roads, damaged roads, good beaches, good trash, dirty trash, and so on. So the objects that need attention in Pancer Beach are dirty beaches and garbage. In addition, there are also visitors who complain about the damaged road.

e) Taman Botani Sukorambi

TABLE 17. Feedback Detection Taman Botani Sukorambi

group	phrase	count
4	anak selamat	1
4	jalan bagus	1
4	kolam curam	1
4	lokasi jauh	1
4	lokasi luas	1
4	tempat bagus	1
4	tempat kecil	1
4	tempat kotor	1
4	tempat misterius	1
4	tempat tua	1

Based on the results of keywords extraction, Sukorambi Botanical Garden brings up results such as safe children, good roads, steep ponds, far locations, wide locations, good places, small places, dirty places, mysterious places, and old places. Based on the results obtained, the objects that need attention in the Sukorambi Botanical Garden are mysterious places and dirty places.

f) Kebun Teh Gunung Gambir

TABLE 18. Kebun Teh Gunung Gambir

group	phrase	count
5	sejuk indah	2
5	jalan khawatir	1
5	jalan pas	1
5	jalan sempit	1
5	sejuk luas	1
5	sejuk lumayan	1
5	tempat lumayan	1

Based on the results of keywords extraction, Gunung Gambir Tea Garden brings up results such as beautiful cool, worried roads, fitting roads, narrow roads, and others. Based on the results obtained, the object that needs attention at Gunung Gambir Tea Garden is about the road.

g) Agrowisata Gunung Pasang Jember

TABLE 19. Feedback Detection Agrowisata Gunung Pasang Jember

group	phrase	count
6	cantik asli	2
6	alami asli	1
6	istimewa murah	1
6	parkir luas	1
6	sejuk segar	1
6	sungai asli	1
6	sungai segar	1
6	tempat baru	1

Based on the results of keywords extraction, Gunung Pasang Jember Agrotourism brings up results such as beautiful original, natural original, special cheap, spacious parking, fresh cool, and others. From the reviews filtered into negative reviews, there are no negative things that need further attention to this tourist spot.

h) Jember Mini Zoo

TABLE 20. Feedback Detection Jember Mini Zoo

group	phrase	count
7	tempat bagus	2
7	anak aneka	1
7	anak train	1
7	tempat bersih	1
7	tempat mahal	1

Based on the results of keywords extraction, Jember Mini Zoo brings up results that are good places, various children, children train, clean places, and expensive places. The object that needs attention is an expensive place.

Conclusion

Based on the evaluation, this technique is promising because the classification and POS Tagger model in this research have a high level of accuracy, i.e., **92%** and **96.6%**. The results given also look quite good where each tourist spot can be seen what objects need to be considered by the manager to develop tourist attractions.

Acknowledgment

This research was supported by University of Jember. I thank all those who have helped my research.

References

1. BPS, Statistik Kunjungan Wisatawan Mancanegara 2021, Badan Pusat Statistik, 2021.

2. Aries, "Radarjember.id," 21 Juli 2022. [Online]. Available: https://radarjember.jawapos.com/opini/21/07/2022/perda-kepariwisataan-urgensi-menuju-jember-konkret/.
3. mg1, "Radarjember," 10 Maret 2021. [Online]. Available: https://radarjember.jawapos.com/pemerintahan/10/03/2021/potensi-di-jember-besar-tapi-pengembangannya-belum-optimal/.
4. A. Dinakaramani, F. Rashel, A. Luthfi and R. Manurung, "Designing an Indonesian part of speech tagset and manually tagged Indonesian corpus," in *Proceedings of the International Conference on Asian Language Processing 2014, IALP 2014*, 2014.
5. N. A. Kartini, F. A. Bachtiar and I. , "Klasifikasi Rating Berdasarkan Komentar Tempat Wisata Di Media Sosial Dengan Menggunakan Metode Fuzzy K-Nearest Neighbor," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* pp. 4164-4169, 2019.
6. N. A. Salsabila, Y. A. Winatmoko, A. A. Septiandri and A. Jamal, "Colloquial Indonesian Lexicon," in *2018 International Conference on Asian Language Processing (IALP)*, Bandung, 2018.
7. Abdi, H. , "Normalizing Data," *Experiments of the Mind,* 2022.
8. D. Kurniawan, Pengenalan Machine Learning Dengan Pyhton, Jakarta: PT Elex Media Komputindo, 2020.
9. N. Leelawat, S. Jariyapongpaiboon, A. Promjun, S. Boonyarak, K. Saengtabtim, A. Laosunthara, A. K. Yudha, and J. Tang. (2022). "Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning". *Heliyon*. Vol. 8, page 1-11.
10. X. Li, M. Xu, W. Zeng, Y. K. Tse, and H. K. Chan. (2023). "Exploring customer concerns on service quality under the COVID-19 crisis: A social media analytics study from the retail industry". *Journal of Retailing and Consumer Services*. Vol. 70, page 1-15.
11. S. Warjri, P. Pakray, S. A. Lyngdoh, and A. K. Maji. (2021). "Part-of-speech (POS) tagging using conditional random field (CRF) model for Khasi corpora". *Internationa Journal of Speech Technology*. Vol. 24, pages 853-864.