Original Research

Sentiment Analysis for Sumber Gempong Rice Field-Based Tourism Destination using Long Short-Term Memory

Njoto Benarkah [1]*, Vincentius Riandaru Prasetyo [1], Jehuda Rivaldo Soetiyono [1]

[1] Informatic Engineering, Faculty of Engineering, Universitas Surabaya, Surabaya-Indonesia
* corresponding author: benarkah@staff.ubaya.ac.id

Abstract*—Sumber Gempong is a rice field-based tourist destination located in Ketapanrame Village, Trawas District, Mojokerto Regency, East Java Province. It is managed by a village-owned company (BUMDesa Mutiara Welirang). BUMDesa evaluates tourist satisfaction manually by reviewing online comments and it consumes time and labor works. Data used in this research automatically collected from Google Maps Review. Long Short-Term Memory (LSTM) method analyze data of two sentiment labels, positive or negative, based on four categories: facilities, services, culinary, and attractions. The collected dataset has 674 comments consist of 420 positive sentiments and 254 negative sentiments with 320 facilities, 61 services, 125 culinary, and 192 attractions comments. Five LSTM models were trained on each of four categories and an overall category. The trained models of overall, facilities, services, culinary, and attractions categories achieved, respectively, 91.2%, 86.8%, 94.1%, 89.7%, and 95.6% of accuracies. The average result accuracy is 91.48%. The manager of BUMDesa Mutiara Welirang satisfied with the results of the system and the sentiment results can be used as evaluation material for Sumber Gempong.*

Keywords: *sentiment anaylsis, LSTM, deep learning, social media, tourism*

Abstrak—Wisata Sawah Sumber Gempong berada di Desa Ketapanrame, Kecamatan Trawas, Kabupaten Mojokerto dan merupakan tempat wisata alam yang dikelola oleh BUMDesa Mutiara Welirang. Evaluasi terhadap tempat wisata ini dilakukan dengan membaca secara manual ulasan-ulasan yang ditulis di media sosial dan pengamatan pribadi. Banyaknya jumlah ulasan yang ada menjadi kendala dalam melakukan evaluasi karena membutuhkan waktu yang cukup lama. Penelitian ini mengambil data ulasan secara otomatis dari media sosial yang diberi label positif atau negatif berdasarkan empat kategori, yaitu fasilitas, pelayanan, kuliner, dan wahana. Metode *Long Short-Term Memory* (LSTM) dipakai sebagai alat untuk melakukan analisis sentimen. Pengambilan data secara otomatis mendapatkan 674 ulasan yang dibagi menjadi 420 ulasan positif dan 254 ulasan negatif, dengan 320 ulasan fasilitas, 61 ulasan pelayanan, 125 ulasan kuliner , dan 192 ulasan wahana. Lima buah model dilatih berdasar tiap kategorinya dan kategori secara keseluruhan. Model yang telah dilatih mendapatkan nilai akurasi sebesar 91,2%, 86,8%, 94,1%, 89,7%, dan 95,6% berturut-turut untuk keseluruhan kategori, kategori fasilitas, layanan, kuliner, dan wahana. Rata-rata akurasi mencapai 91,48%. Hasil dari sistem telah diujicobakan kepada manajer BUMDesa Mutiara Welirang dan bisa dipakai sebagai bahan evalusi untuk peningkatan kualitas di Sumber Gempong.

Kata kunci: *analisis sentimen, LSTM, deep learning, media sosial, wisata*

INTRODUCTION

Data from 13 tourist destinations on December 2022 collected by a government department in Mojokerto Regency, called *Dinas Kebudayaan, Kepemudaan, Olahraga, dan Pariwisata* (Disbudporapar, 2023) recorded a significant amount of visitors as many as 99,090 tourist visitations on those locations, although it has not been covered all tourist destination in the region (Disbudporapar, 2023). Ketapanrame Village is located in Trawas District, Mojokerto Regency, East Java Province, Indonesia. It has beautiful natural scenery and many nature-based tourist destinations.

Sumber Gempong is a rice field-based tourist destination located in Ketapanrame Village, Trawas District, Mojokerto Regency managed by village-owned company, locally called *BUMDesa Mutiara Welirang*. It was inaugurated by the Mojokerto Regent (locally called Bupati Mojokerto) on December 18, 2021. It achieved the 2023 best tourist village award by the Ministry of Tourism of Republic of Indonesia on August 2023 (Hendriyani, 2023). This tourist destination gained its popularity since before its inauguration and crowded with visitors mostly on weekend. Visitor interest can be seen from the number of comments on social media, especially on Google Maps Review. The number of comments posted by August 2022 is 2,676 comments.

https://doi.org/10.24123/saintek.v5i2.6498

Social media has been used as a tool to find information related to contemporary culture or anything about social world (Budge, 2017). Research work using partial least square structural equation model (PLS-SEM) on online comments in a form of text data of visitor's experience in a museum visitation shows that the comments influence perspective and level of confidence of society (Fu, Liu & Li, 2024).

BUMDesa Mutiara Welirang uses comments posted on various social media owned by the company, including Google Maps Review as an evaluation material to gain insight of society opinions about Sumber Gempong for future. Figure 1 shows a negative sentiment of a comment on Google Map Review from a visitor regarding toilet facility and visitor's mismatched expectation of this tourist destination. The huge number of online comments posted by visitors gives analysis complication for the manager of BUMDesa to carry out evaluation process. Analysis on online comments is done manually by the manager as an evaluation material for future improvement of the tourist destination. Oral opinions from visitors and neighborhoods also considered as an evaluation material. This research proposed a sentiment analysis method to give overview of people perspective of this place as a positive or negative sentiments.

Fig. 1. An example of comment posted on Google Maps Review by a visitor that has negative sentiment, based on phrases "*minim toilet*" and "*apa yang anda lihat di Instagram dan google tidak seperti kenyataan*".

Methods based on Natural Language Processing (NLP) or deep learning can be used to analyze online comments posted by visitors. The development of Internet applications and huge amount of online data, especially online comments posted on social media, can be automatically analyze using NLP or deep learning methods to get useful information. Sentiment analysis has been applied to a broad range of fields, such as in politics during political change period (Arsi & Waluyo, 2021), in business to get insight of people opinions of a product (Negara, Muhardi, & Putri, 2020), (Ahn & Park, 2023), (Jassim, Abd, & Omri, 2023), (Kaur & Sharma, 2023), in tourism to provide an overview of tourism products and future developments (Azzahra & Wibowo, 2020), (Ginantra et al., 2022), (Mahardika et al., 2022), in medical field to improve the quality of health services (Samah et al., 2023) or to give overview of public opinion about a certain disease (Benrouba & Boudour, 2023), (Caldo et al., 2023), in education to find out student comments over administration process in a university, and university policy during and after pandemic, (Yuliska et al., 2021), (Oktaviana, Sari, & Indriati, 2022), (Qaqish, Aranki, & Etaiwi, 2023), and to give sentiment overview in metaverse world (Tunca, Sezen, & Wilk, 2023). Sentiment analysis also has been used to detect emotions given by society through social media (Mohammad et al., 2018), (Liu et al., 2023). Research work by Talaat gives comparison of several sentiment analysis methods on various data on social media and achieved varying accuracy results (Talaat, 2023).

Long Short-Term Memory (LSTM), as one of the methods used for sentiment analysis, also has been used intensively to solve everyday problems, such as disease classification, and as a feature to gain device access based on face recognition. Sentiment analysis using LSTM also successfully help to gain public perception (Liliana, Hikmah, & Harjono, 2021; Maulana, Wijoyo, & Mursityo, 2023). LSTM also has been used to predicts things in electronical field (Ningrum et al., 2021), (Selle, Yudistira, & Dewi, 2022). Research works to analyze positive and negative sentiments using LSTM also incorporate convolutional neural network (CNN) and give varying accuracy results, 90.75% (Ombabi, Ouarda, & Alimi, 2020), and 81.31% (Mohammed & Kora, 2019). Another research using CNN and LSTM to analyze positive and negative sentiments on eight educational tourist destinations gives a promising 91.24% of accuracy result using 60,365 comments (Wang, Chu, & Lan, 2022). LSTM also useful in finding correlation between online comments and taste ratings with 90% accuracy result using 48,152 positive sentiments and 55,603 negative sentiments (Fu & Pan, 2022). Google Maps Review data can be used for sentiment analysis. The data is used by Valence Aware Dictionary for Sentiment Reasoning (VADER) model for sentiment analysis to evaluate customer opinions related to food at several restaurants (Mathayomchan & Taecharungroj, 2020). The research evaluates the relation among four factors, which are, food, services, atmosphere, and rating values. Google Maps Review data also have been used to analyze people opinions about public library (Borrego & Comalat Navarra, 2020).

Automatic sentiment analysis using LSTM regarding Sumber Gempong using online data has never been done. Previous research works shows LSTM method gives good results for sentiment analysis. Sumber Gempong is a nature-based tourist destination and it is managed by village-owned business (BUMDesa). BUMDesa is a legal form for village ownership to support community welfare.

The purpose of this research is to develop web-based application using sentiment analysis on online data from Google Maps Review or Instagram for Sumber Gempong. Features of the application are automatic data retrieval using crawling method to get online comments, set data label into two kinds of positive and negative labels, and group data into four categories of facilities, services, culinary, and attractions. Labelled data in each category trained by the corresponding LSTM model and classified as positive or negative sentiments. *BUMDesa Mutiara Welirang* can use the results as evaluation material for quality improvement of the tourist destination.

METHODS

This research implements sentiment analysis into a web-based application for Sumber Gempong using online comments automatically retrieved from Google Maps Review or Instagram posts of Sumber Gempong account. For this study, data from Google Maps Review were used and automatically retrieved by system and limited to a predetermined time period. Preprocessing methods on retrieved data are carried out by giving label for each comment, to do stemming process, and convert slang words into standard words. Preprocessing methods give result in a form of data vectors, and feed into LSTM models for training process. The methodology of this research work uses the following steps: data crawling, dataset construction, preprocess, convert text data into numerical vector, LSTM model building and training, and evaluation of trained LSTM model.

Data Crawling

Crawling is an automatic process to traverse each page of websites and automatically retrieve data to get the information (Hanifah & Nurhasanah, 2018). Data crawling collects data from a certain location and automatically download the data. The research design crawls data from Instagram using Instaloader or Google Maps Review using Simple HTML DOM, limited to a predetermined time period. Successfully downloaded data stored in database as the input for subsequent process.

https://doi.org/10.24123/saintek.v5i2.6498

Dataset Construction

The amount of data collected from Google Map Review are 674 comments. Each data labelled as positive or negative sentiments and group into facilities, services, culinary, and attractions categories. An expert on Indonesian language labels each data manually and the results checked by the tourist manager of *BUMDesa Mutiara Welirang*. Data consists of 420 positive sentiments, 254 negative sentiments. The distribution of data in each category are 320 comments in facilities category, 61 comments in services category, 125 comments in culinary category, and 192 comments in attractions category. Table 1 shows examples of data from the dataset.

Table 1

Examples of Comments as the Result of Crawling Process and Have Been Given Labels in Each Category

Data	Label	Category
pernah mau masuk ke situ, tapi gagal masuk karena penuh, tdk ada parkir, datang sekitar jam 11 siang	Negative	Facilities
Harga tiket masuk terjangkau, jenis makanan bermacam-macam, harga manakan standar, harga tiket wahana juga standar. Meskipun weekend tapi tidak terlalu ramai, jadi bisa menikmati pemandangannya.	Positive	Facilities, Culinary, Attractions
Tempat nya sangat ramai, minim toilet, apa yang anda lihat di Instagram dan google tidak seperti kenyataan	Negative	Facilites
Tempatnya bersih sejuk jg semua petugasnya ramah	Positive	Services
Kesini sudah berulang kali, mulai dari yang masih ada pedagang yang motoran, sampai sekarang banyak pembangunan yang sudah berkembang banyk.. Dari kereta sawah, perahu bebek, becak langit	Positive	Attractions
harga tiket masuk 5k.parkir mobil 6k.naik kereta 10k.harga makanan standar.Lumayan lah	Positive	Facilities, Culinary, Attractions

Preprocess

Preprocess step converts text data into standard format needed by the subsequent process (Khairunnisa, Adiwijaya, & Faraby, 2021). This step eliminates unnecessary words that can interfere the subsequent process (Normawati & Prayogi, 2021). The processes used in this step are case folding, data cleaning, tokenizing, slang conversion, stopword removal, and stemming.

Case Folding

The objective of case folding is to reduce to complexity for the process of finding comments. All letters converted into standard format (Alita and Isnain, 2020). All letters converted into small case letters. Table 2 shows examples of the results of conversion process in case folding.

Table 2

Examples of Case Folding Process

Input data	Conversion result
Tempatnya sejuk tapi tempat untuk teduh kurang ..	*tempatnya sejuk tapi tempat untuk teduh kurang ..*
Tempat nya sangat bersih, krg toilet	*tempat nya sangat bersih, krg toilet*
Tempatnya bersih sejuk, jg teduh	*tempatnya bersih sejuk, jg teduh*

Data Cleaning

Data cleaning remove unnecessary characters, such as punctuation and emoticon, in a comment posted by visitor in Google Maps Review. The removal of punctuation of data in order to reduce data noise (Yulita, 2021). Table 3 shows three examples of data and the results from data cleaning.

Table 3
Examples of Cleaning Process

Input data	Cleaning result
tempatnya sejuk tapi tempat untuk teduh kurang ..	*tempatnya sejuk tapi tempat untuk teduh kurang*
tempat nya sangat bersih, krg toilet	*tempat nya sangat bersih krg toilet*
tempatnya bersih sejuk, jg teduh	*tempatnya bersih sejuk jg teduh*

Tokenizing

Tokenizing divides a comment into words (Saptari, 2021). Each word from tokenizing result is given its weight. Table 4 shows the results of tokenizing process.

Table 4
Examples of Tokenizing Process

Input data	Tokenizing result
tempatnya sejuk tapi tempat untuk teduh kurang	*tempatnya | sejuk | tapi | tempat | untuk | teduh | kurang*
tempat nya sangat bersih krg toilet	*tempat | nya | sangat | bersih | krg | toilet*
tempatnya bersih sejuk jg teduh	*tempatnya | bersih | sejuk | jg | teduh*

Slang Conversion

Slang word conversion converts slang words or abbreviations contained in data into standard word based on Indonesian Dictionary. *Kamus Besar Bahasa Indonesia* (KBBI) is used as the reference dictionary. Table 5 shows the conversion of slang words and abbreviations.

Table 5
Examples of slang conversion process

Input data	Conversion result
tempatnya | sejuk | tapi | tempat | untuk | teduh | kurang	*tempatnya | sejuk | tapi | tempat | untuk | teduh | kurang*
tempat | nya | sangat | bersih | krg | toilet	*tempat | nya | sangat | bersih | kurang | toilet*
tempatnya | bersih | sejuk | jg | teduh	*tempatnya | bersih | sejuk | juga | teduh*

Stop word removal

Stopword removal step removes unnecessary words or meaningless words and considered unimportant. Some examples of words to be removed are ‘yang’, ‘di’, ‘dari’, and its kind. The removal process refers to a list of stopwords that have been determined and created previously (Putra, Pratomo and Perwira, 2022). Sastrawi or Natural Language Toolkit (NLT) libraries can be used as a reference for this process. Tabel 6 shows examples of stopword removal.

Table 6
Examples of Stopword Removal Process

Input data	Stopword removal
tempatnya | sejuk | tapi | tempat | untuk | teduh | kurang	*tempatnya | sejuk | tapi | tempat | teduh | kurang*
tempat | nya | sangat | bersih | kurang | toilet	*tempat | nya | bersih | kurang | toilet*
tempatnya | bersih | sejuk | juga | teduh	*tempatnya | bersih | sejuk | teduh*

Stemming

Stemming step converts words in root form of the word by eliminating prefixes and suffixes in a word. This step normalizes the data. Sastrawi or Natural Language Toolkit (NLT) libraries can be used as a reference for this process. Table 7 shows the result of stemming process.

Table 7
Examples of Stemming Process

Input data	Result of stemming
tempatnya | sejuk | tapi | tempat | teduh | kurang	*tempat | sejuk | tapi | tempat | teduh | kurang*
tempat | nya | bersih | kurang | toilet	*tempat | bersih | kurang | toilet*
tempatnya | bersih | sejuk | teduh	*tempat| bersih | sejuk | teduh*

Conversion of Text Data into Numerical Vector

Input for LSTM model must be in a form of vector. Vector data helps the method to find combination of words from sentiment data (Nurvania, Jondri, & Lhaksamana, 2021). The conversion process constructs a dictionary of words from the words collection from the previous result. Each word is translated into round number. Figure 2 shows conversion process illustration from text data into numerical vectors.

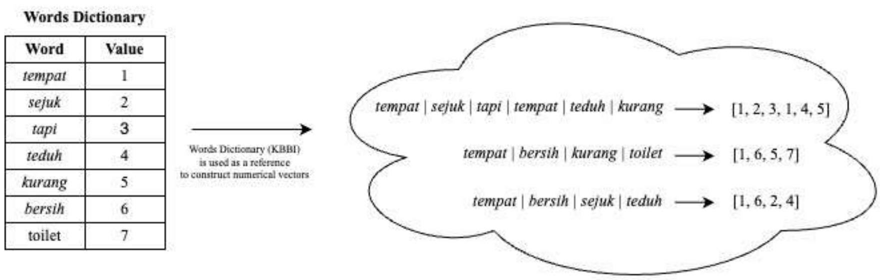


Fig. 2. Word dictionary construction and converting text data into numerical vectors.

Next, padding process to ensure all numerical vectors has the same size based on the longest data size. Zeros are added to all data shorter than the longest data size, either with pre-padding or post-padding. Figure 3 shows the pre-padding method by adding zeros in front of number.

https://doi.org/10.24123/saintek.v5i2.6498

Number Order		*Pre-padding*
[1, 2, 3, 1, 4, 5]	→	[1, 2, 3, 1, 4, 5]
[1, 6, 5, 7]	→	[0, 0, 3, 1, 4, 5]
[1, 6, 2, 4]	→	[0, 0, 3, 1, 4, 5]

Fig. 3. Pre-padding method to add zeros in numerical vector.

Training Process of LSTM Models

LSTM model constructed with several layers as shown in Table 8. The model uses loss function, binary cross-entropy, and for optimization uses Adam optimizer. Evaluation metric uses accuracy. The training process uses epoch=100, validation split=0.1, and batch size=64. Training process has early stop condition when validation loss value remains stable in 5 folds. Five categories of data use the same model and generate five models for each category.

Table 8
LSTM Model

Layer (type)	Output Shape	Number of param
embedding (Embedding)	(None, None, 50)	72,900
bidirectional (Bidirectional)	(None, None, 128)	58,880
dense (Dense)	(None, None, 64)	8,256
dropout (Dropout)	(None, None, 64)	0
dense_1 (Dense)	(None, None, 32)	2080
bidirectional_1 (Bidirectional)	(None, 64)	16640
dropout_1 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 1)	65
Total params: 158,821		
Trainable params: 158,821		
Non-trainable params: 0		

Model evaluation

Accuracy, precision, recall, and f1-score are used as evaluation metrics of the classification results of the trained LSTM model. Accuracy measure how well the model classifies correct sentiment classifications against all data. Equation (1) is the formula to compute an accuracy value.

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

Equation (2) is the formula to compute precision. Precision refers the capacity of a model to classify correct positive sentiments. Precision computes the number of correct positive sentiment classification divided by all positive sentiment classifications.

$$precision = \frac{TP}{TP + FP} \quad (2)$$

Equation (3) is the formula to compute recall. Recall refers the capacity of a model to classify positive sentiment data. Recall computes the number of correct positive sentiment classification divided by all positive sentiment data.

$$recall = \frac{TP}{TP + FN} \tag{3}$$

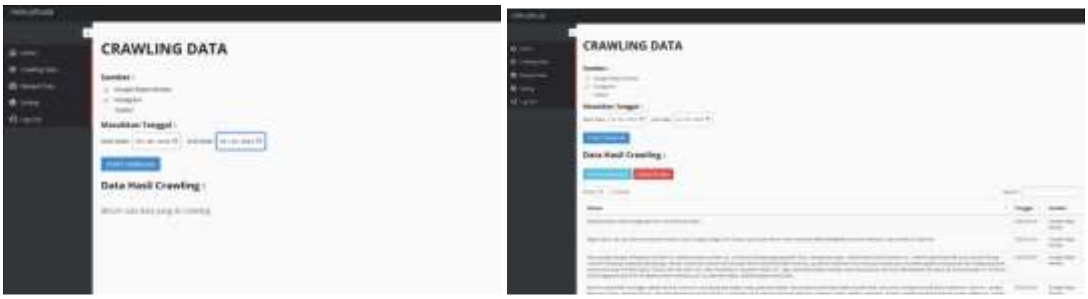
Equation (4) is the formula to compute f1-score. F1-score is based on recall and precision values. F1-score computes the multiplication of recall and precision divided by the addition of recall and precision.

$$f1score = \frac{recall \times precision}{recall + precision} \tag{4}$$

Where TP (*true positive*) = positive sentiment that classified as positive sentiment, TN (*true negative*) = negative sentiment that classified as negative sentiment, FP (*false positive*) = negative sentiment that classified as positive sentiment, and FN (*false negative*) = positive sentiment that classified as negative sentiment.

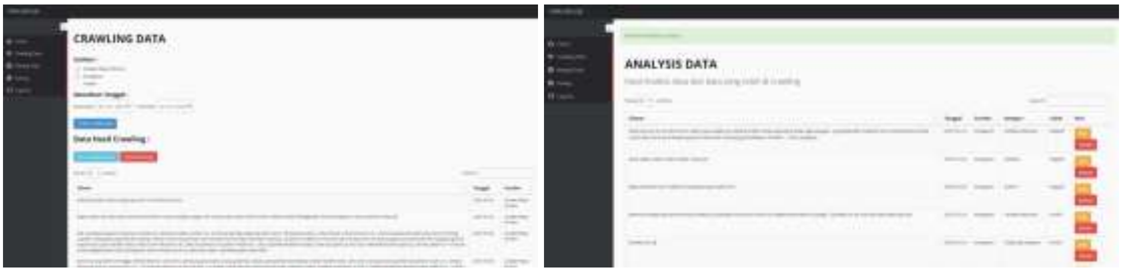
RESULTS

The sentiment analysis research with the development of web-based application is developed based on interview with the tourist unit manager of *BUMDesa Mutiara Welirang* to solve problem faced by *BUMDesa Mutiara Welirang* to analyze online comments as an evaluation material for future improvement of Sumber Gempong. The web-based application is developed using Laravel v9.52.4, Python v3.8.8, and MySQL as the database of the system. The initial development features are login, crawling, sentiment analysis, sentiment analysis history management, and account setting. Whole system has been verified to ensure the system works correctly and it has been tested to the tourist unit manager of *BUMDesa Mutiara Welirang*. Figure 4 – Figure 6 shows three features of the system. Figure 4 shows the crawling feature. The feature automatically collect data from Google Maps Review or Instagram. User needs to provide starting date and ending date of data retrieval and choose the social media source. Figure 5 shows the sentiment analysis feature as the main feature. User can give label to data that has been collected from crawling process. A button provided to start the analysis process and the result of sentiment analysis is given in a list. Figure 6 shows the history of the analysis and user can see the graphical proportion of both sentiments.



(a) (b)

Fig. 4. Crawling data feature (a) initial parameters setting and (b) result of the crawling process.



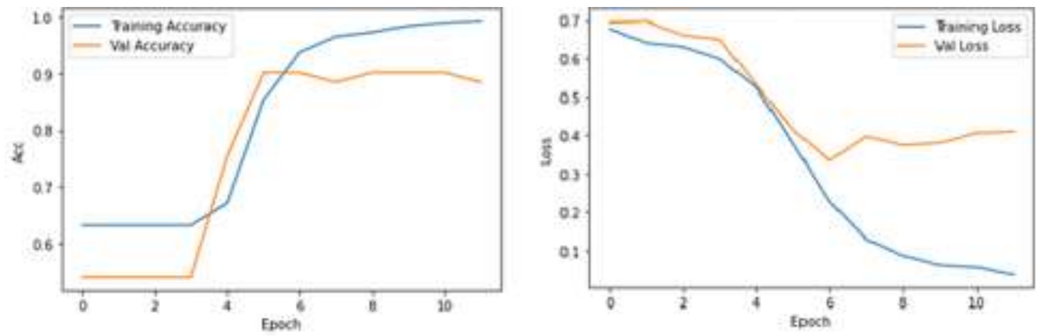
(a) (b)

Fig. 5. Sentiment analysis feature (a) initial parameters setting and (b) result of the analysis.

Fig. 6. History feature with graphical diagram of the analysis result.

The system has been validated to measure the success rates of the system based on the objective of the research. Dataset has been validated by the tourist manager of *BUMDesa Mutiara Welirang*. Data were labelled by an expert in Indonesian language. The manajer also validated the classification results of the system, and the system showed good results. The results of the sentiment analysis used as evaluation material by *BUMDesa Mutiara Welirang*.

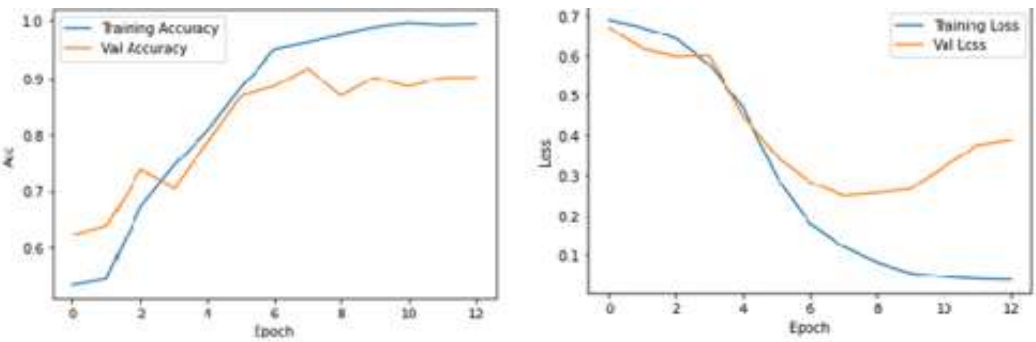
Performance graphic of training process accuracy to classify positive and negative sentiments can be seen in Figure 7. Figure 7 – Figure 11 shows the performance graphics for training process accuracy to classify facilities (Figure 8), services (Figure 9), culinary (Figure 10), and attractions (Figure 11) categories.



(a) (b)

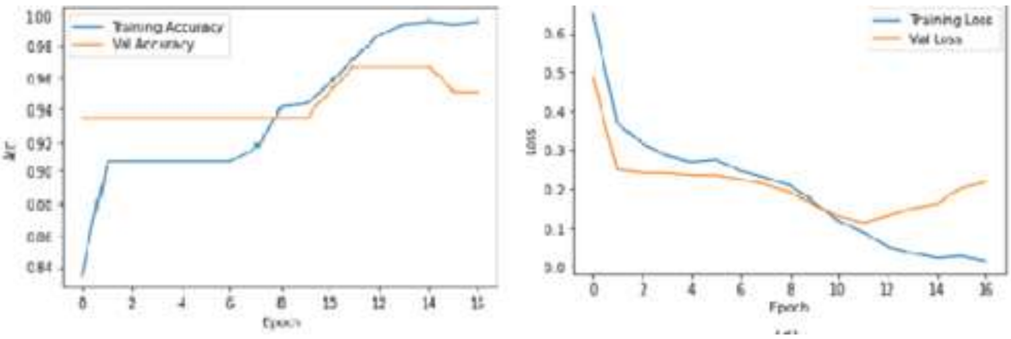
Fig. 7. Performance of training process graphics (a) accuracy) and (b) loss of LSTM model for overall category.

https://doi.org/10.24123/saintek.v5i2.6498



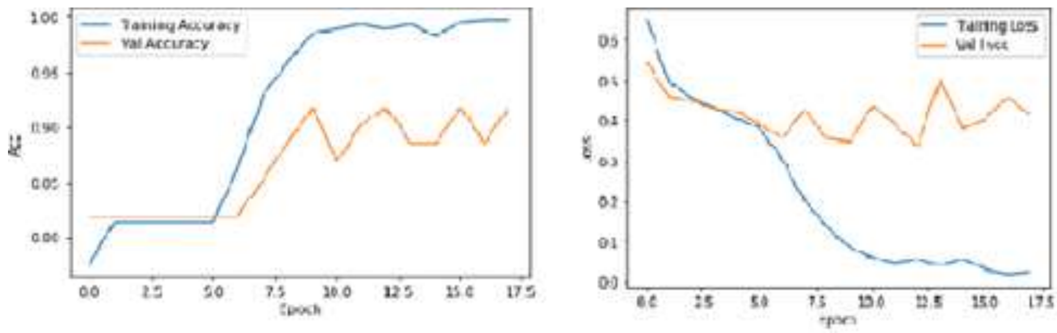
(a) (b)

Fig. 8. Performance of training process graphics (a) accuracy) and (b) loss of LSTM model for facilities category.



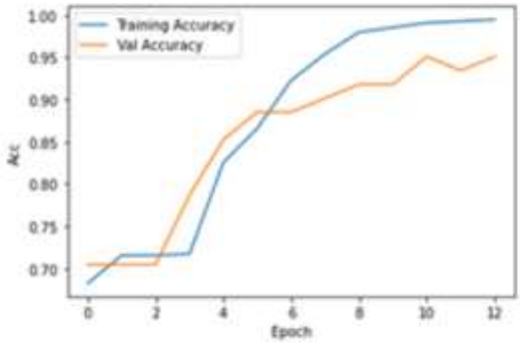
(a) (b)

Fig. 9. Performance of training process graphics (a) accuracy) and (b) loss of LSTM model for services category.

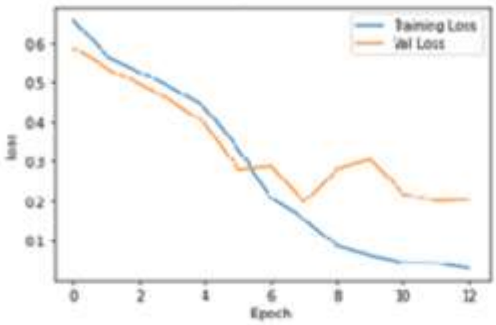


(a) (b)

Fig. 10. Performance of training process graphics (a) accuracy) and (b) loss of LSTM model for culinary category.



(a)



(b)

Fig. 11. Performance of training process graphics (a) accuracy) and (b) loss of LSTM model for attractions category.

All five models achieve high accuracy results using proportion 90% of training data and 10% of test data, and using validation split=0.1. Evaluation measurement results based on precision, recall, f1-score, and accuracy are shown in Table 9.

S=-1 means negative sentiment, and S=+ means positive sentiment. M means model number, 1 for overall model, 2 for facilities model, 3 for services model, 4 for culinary model, and 5 for attractions model. Three models achieve accuracy results above 90%, model 1 with 91.2% accuracy result, model 3 with 94.1% accuracy result, and model 5 with 95.6% accuracy result.

Table 9
Performance Evaluation of The Models

M	S	Precision	Recall	F-1 Score	Accuracy
1	-	92%	85%	88%	91.2%
	+	91%	95%	93%	
2	-	83%	94%	88%	86.8%
	+	93%	78%	85%	
3	-	95%	98%	97%	94.1%
	+	75%	50%	60%	
4	-	96%	91%	93%	89.7%
	+	69%	85%	76%	
5	-	98%	96%	97%	95.6%
	+	90%	95%	92%	

DISCUSSION

Sentiment analysis for Sumber Gempong built using five LSTM models to classify negative and positive sentiments achieves average accuracy result of 91.48%. The highest accuracy, 95.6%, achieved by the fifth model, the attractions category. Although attractions data contribute 27.51% of all category data, the model gives consistent results showing it has the highest accuracy and with high precision and recall values of this category, which achieved above 90%. LSTM has good performance on various amount of data. With data distribution 45.84% of facilities, 8.74% of services, 17.91% of culinary, and 27.51% of attractions, LTSM shows its capability as sentiment analysis classifier by having high accuracy results in of each category. LSTM still gives good result when applied to all data and classifies positive or negative sentiments in general with 91.2% accuracy result.

The system implemented as web-based application using the trained LSTM models and successfully crawls Google Maps Reviews data using Simple HTML DOM. Limitation happened at the end of the development when the system encountered difficulty attempting data Instagram post of an account because of the policy changes of the company.

CONCLUSION

The tourist unit manager of *BUMDesa Mutiara Welirang* have tested the system and can automatically get the sentiment analysis results crawled from Google Maps Review. The system allows the manager to choose a period of date for which the data must be analysis. Further development mobile-based application of the system may give more advance and mobility access for the company. Update on slang words dictionary may also improve the result of sentiment analysis. An alternative method to crawl data on Twitter and Instagram may need to be developed due to limitation to access the content in those platforms.

ACKNOWLEDGMENT

Authors would like to express our sincere gratitude to H. Zainul Arifin, S.E., NI.P. as Head of Ketapanrame Village and Saifudin as Tourist Unit Manager of *BUMDesa Mutiara Welirang*, Ketapanrame Village, Trawas District, Mojokerto Regency, East Java. Time and opportunity has been given to us to conduct this research for the village.

REFERENCES

Ahn, H. and Park, E., 2023. Motivations for user satisfaction of mobile fitness applications: An analysis of user experience based on online review comments. *Humanities and Social Sciences Communications*, 10(1). https://doi.org/10.1057/s41599-022-01452-6.

Alita, D. and Isnain, A.R., 2020. Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *jurnal komputasi*, 8(2). https://doi.org/10.23960/komputasi.v8i2.2615.

Arsi, P. and Waluyo, R., 2021. Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(1), p.147. https://doi.org/10.25126/jtiik.0813944.

Azzahra, S.A. and Wibowo, A., 2020. Analisis Sentimen Multi-Aspek Berbasis Konversi Ikon Emosi dengan Algoritme Naïve Bayes untuk Ulasan Wisata Kuliner Pada Web Tripadvisor. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(4), p.737. https://doi.org/10.25126/jtiik.2020731907.

Benrouba, F. and Boudour, R., 2023. Emotional sentiment analysis of social media content for mental health safety. *Social Network Analysis and Mining*, 13(1). https://doi.org/10.1007/s13278-022-01000-9.

Borrego, Á. and Comalat Navarra, M., 2020. What users say about public libraries: an analysis of Google Maps reviews. *Online Information Review*, 45(1), pp.84–98. https://doi.org/10.1108/OIR-09-2019-0291.

Budge, K., 2017. Objects in Focus: Museum Visitors and Instagram. *Curator: The Museum Journal*, 60(1), pp.67–85. https://doi.org/10.1111/cura.12183.

Caldo, D., Bologna, S., Conte, L., Amin, M.S., Anselma, L., Basile, V., Hossain, M.M., Mazzei, A., Heritier, P., Ferracini, R., Kon, E. and De Nunzio, G., 2023. Machine learning algorithms distinguish discrete digital emotional fingerprints for web pages related to back pain. *Scientific Reports*, 13(1). https://doi.org/10.1038/s41598-023-31741-2.

Disbudporapar, 2023. *Data Wisatawan Desember 2022*. Dinas Kebudayaan, Kepemudaan, Olahraga dan Pariwisata, Pemerintah Kabupaten Mojokerto.

Fu, M. and Pan, L., 2022. Sentiment Analysis of Tourist Scenic Spots Internet Comments Based on LSTM. *Mathematical Problems in Engineering*, 2022, pp.1–9. https://doi.org/10.1155/2022/5944954.

Fu, X., Liu, X. and Li, Z., 2024. Catching eyes of social media wanderers: How pictorial and textual cues in visitor-generated content shape users' cognitive-affective psychology. *Tourism Management*, 100, p.104815. https://doi.org/10.1016/j.tourman.2023.104815.

Ginantra, N.L.W.S.R., Yanti, C.P., Prasetya, G.D., Sarasvananda, I.B.G. and Wiguna, I.K.A.G., 2022. Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan K-NN. *Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI*, [online] 11(3), pp.205–215. https://doi.org/10.23887/janapati.v11i3.49450.

Hanifah, R. and Nurhasanah, I.S., 2018. Implementasi Web Crawling untuk Mengumpulkan Informasi Wisata Kuliner di Bandar Lampung. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 5(5), p.531. https://doi.org/10.25126/jtiik.201855842.

Hendriyani, I.G.A.D., 2023. *Siaran Pers: Menparekraf: ADWI 2023 Perkuat Konsistensi Masyarakat Bangun Desa Wisata*. [online] Kementerian Pariwisata dan Ekonomi Kreatif Republik Indonesia. Available at: <https://kemenparekraf.go.id/hasil-pencarian/siaran-pers-menparekraf-adwi-2023-perkuat-konsistensi-masyarakat-bangun-desa-wisata>.

Jassim, M.A., Abd, D.H. and Omri, M.N., 2023. Machine learning-based new approach to films review. *Social Network Analysis and Mining*, 13(1). https://doi.org/10.1007/s13278-023-01042-7.

Kaur, G. and Sharma, A., 2023. A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. *Journal of Big Data*, 10(1). https://doi.org/10.1186/s40537-022-00680-6.

Khairunnisa, S., Adiwijaya, A. and Faraby, S. Al, 2021. Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2), p.406. https://doi.org/10.30865/mib.v5i2.2835.

Liliana, D.Y., Hikmah, N.N. and Harjono, M., 2021. Pengembangan Sistem Pemantauan Sentimen Berita Berbahasa Indonesia Berdasarkan Konten dengan Long-Short-Term Memory. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(5), p.995. https://doi.org/10.25126/jtiik.2021854624.

Liu, X., Shi, T., Zhou, G., Liu, M., Yin, Z., Yin, L. and Zheng, W., 2023. Emotion classification for short texts: an improved multi-label method. *Humanities and Social Sciences Communications*, 10(1). https://doi.org/10.1057/s41599-023-01816-6.

Mahardika, F.R., Supianto, A.A., Setiawan, N.Y., Yuwana, R.S. and Suryawati, E., 2022. Rekomendasi Pengembangan Fasilitas Wisata Tugu Pahlawan Surabaya Melalui Visualisasi Dashboard Hasil Klasifikasi Analisis Sentimen Ulasan Pengunjung. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(2), p.363. https://doi.org/10.25126/jtiik.2022925655.

Mathayomchan, B. and Taecharungroj, V., 2020. "How was your meal?" Examining customer experience using Google maps reviews. *International Journal of Hospitality Management*, 90, p.102641. https://doi.org/10.1016/j.ijhm.2020.102641.

Maulana, A.R., Wijoyo, S.H. and Mursityo, Y.T., 2023. Analisis Sentimen Kebijakan Penerapan Kurikulum Merdeka Sekolah Dasar dan Sekolah Menengah pada Media Sosial Twitter dengan Menggunakan Metode Word Embedding dan Long Short Term Memory Networks (LSTM). *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(3), p.523. https://doi.org/10.25126/jtiik.20231036977.

Mohammad, S., Bravo-Marquez, F., Salameh, M. and Kiritchenko, S., 2018. SemEval-2018 Task 1: Affect in Tweets. In: *Proceedings of the 12th International Workshop on Semantic Evaluation*. [online] New Orleans, Louisiana: Association for Computational Linguistics. pp.1–17. https://doi.org/10.18653/v1/S18-1001.

Mohammed, A. and Kora, R., 2019. Deep learning approaches for Arabic sentiment analysis. *Social Network Analysis and Mining*, 9(1), p.52. https://doi.org/10.1007/s13278-019-0596-4.

Negara, A.B.P., Muhardi, H. and Putri, I.M., 2020. Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes dan Seleksi Fitur Information Gain. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(3), p.599. https://doi.org/10.25126/jtiik.2020711947.

Ningrum, A.A., Syarif, I., Gunawan, A.I., Satriyanto, E. and Muchtar, R., 2021. Algoritma Deep Learning-LSTM untuk Memprediksi Umur Transformator. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(3), p.539. https://doi.org/10.25126/jtiik.2021834587.

Normawati, D. and Prayogi, S.A., 2021. Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer dan Informatika (J-SAKTI)*, 5(2).

Nurvania, J., Jondri, J. and Lhaksamana, K.M., 2021. Analisis Sentimen Pada Ulasan di TripAdvisor Menggunakan Metode Long Short-Term Memory (LSTM). *eProceedings of Engineering*, 8(4).

Oktaviana, N.E., Sari, Y.A. and Indriati, I., 2022. Analisis Sentimen terhadap Kebijakan Kuliah Daring Selama Pandemi Menggunakan Pendekatan Lexicon Based Features dan Support Vector Machine. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(2), p.357. https://doi.org/10.25126/jtiik.2022925625.

Ombabi, A.H., Ouarda, W. and Alimi, A.M., 2020. Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks. *Social Network Analysis and Mining*, 10(1), p.53. https://doi.org/10.1007/s13278-020-00668-1.

Putra, R.P., Pratomo, A.H. and Perwira, R.I., 2022. Text Message Classification using Multiclass Support Vector Machine on Information Service Chatbot in the Informatics Department UPN "Veteran" Yogyakarta. *Telematika*, 19(3), p.295. https://doi.org/10.31315/telematika.v19i3.7418.

Qaqish, E., Aranki, A. and Etaiwi, W., 2023. Sentiment analysis and emotion detection of post-COVID educational Tweets: Jordan case. *Social Network Analysis and Mining*, 13(1). https://doi.org/10.1007/s13278-023-01041-8.

Samah, K.A.F.A., Azharludin, N.M.N., Riza, L.S., Jono, M.N.H.H. and Moketar, N.A., 2023. Classification and visualization: Twitter sentiment analysis of Malaysia's private hospitals. *IAES International Journal of Artificial Intelligence*, 12(4), pp.1793–1802. https://doi.org/10.11591/ijai.v12.i4.pp1793-1802.

Saptari, R., 2021. Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Unit Gawat Darurat Rumah Sakit Umum di Indonesia Menggunakan Seleksi Fitur Information Gain dan

Support Vector Machine. *Joined Journal (Journal of Informatics Education)*, 4(2), p.104. https://doi.org/10.31331/joined.v4i2.1925.

Selle, N., Yudistira, N. and Dewi, C., 2022. Perbandingan Prediksi Penggunaan Listrik dengan Menggunakan Metode Long Short Term Memory (LSTM) dan Recurrent Neural Network (RNN). *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(1), p.155. https://doi.org/10.25126/jtiik.2022915585.

Talaat, A.S., 2023. Sentiment analysis classification system using hybrid BERT models. *Journal of Big Data*, 10(1). https://doi.org/10.1186/s40537-023-00781-w.

Tunca, S., Sezen, B. and Wilk, V., 2023. An exploratory content and sentiment analysis of the guardian metaverse articles using leximancer and natural language processing. *Journal of Big Data*, 10(1). https://doi.org/10.1186/s40537-023-00773-w.

Wang, Y., Chu, C. and Lan, T., 2022. Sentiment Classification of Educational Tourism Reviews Based on Parallel CNN and LSTM with Attention Mechanism. *Mobile Information Systems*, 2022, pp.1–13. https://doi.org/10.1155/2022/6177427.

Yuliska, Y., Qudsi, D.H., Lubis, J.H., Syaliman, K.U. and Najwa, N.F., 2021. Analisis Sentimen pada Data Saran Mahasiswa Terhadap Kinerja Departemen di Perguruan Tinggi Menggunakan Convolutional Neural Network. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(5), p.1067. https://doi.org/10.25126/jtiik.2021854842.

Yulita, W., 2021. Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Data Mining dan Sistem Informasi*, 2(2), p.1. https://doi.org/10.33365/jdmsi.v2i2.1344.