Aspect-Based Sentiment Analysis of Hotels in Bali on Tripadvisor Using BERT Algorithm

Dimas Samodra Bimaputra[1*], Edi Sutoyo[1,2]
[1]Department of Information Systems, Telkom University, Bandung, Indonesia
[2]Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen,
The Netherlands

Article Info	ABSTRACT
	The covid pandemic that began in 2020 has caused enormous losses worldwide, including in Indonesia. Human-to-human contact is the source of the transmission of the covid virus, so the government urges people to maintain cleanliness when interacting. Bali is a popular destination for foreign and domestic visitors in Indonesia. Hospitality businesses in Bali unquestionably face a high risk of covid transmission; consequently, changes in hotel business processes are unavoidable; the implementation of new business processes can have a negative effect on business performance. In order to maintain Bali's reputation as the most popular tourist destination in Indonesia, the government must evaluate the performance of several hotel services that have implemented new business processes. The Aspect-Based Sentiment Analysis (ABSA) methodology can be utilized for performance evaluation. One of the finest algorithms for analyzing text, Bidirectional Encoder Representations from Transformer (BERT), is required for sentiment analysis. The data consists of textual customer evaluations of hotels in Bali that have implemented a new protocol or Standard Operational Procedure (SOP), retrieved from the Tripadvisor website. In the form of a number of evaluations of various aspects of the hotel, the research results can assist the government in analyzing the performance of hotels in Bali based on predetermined criteria.

Corresponding Author:

Dimas Samodra Bimaputra,
Department of Information Systems, Telkom University,
Bandung, 40257, Indonesia.
Email: dimassamodra@student.telkomuniversity.ac.id

1. INTRODUCTION

The covid pandemic, which began in 2020, has brought many changes to human activities around the world, and Indonesia is no exception. People are required to change their lifestyle to survive the covid pandemic by always maintaining cleanliness, wearing masks, always washing hands using hand sanitizers and the like. Self-awareness about environmental hygiene is needed as a first step from the community. Crowded places have a high risk of spreading covid, one of the places in Indonesia that is known by foreign and local residents for its thick natural and cultural tourism, namely Bali [1]. Lodging is the main thing in tourism, of course, it has a high risk of covid transmission because all hospitality activities require direct contact starting from the check in to check out process. With the appeal of Indonesian President, Joko Widodo [2], there are many new

policies that require companies to change the way business works remotely, so hotels in Bali are required to design new business processes by following existing policies.

Changes to standard operating procedures (SOPs) will inevitably affect the performance of the business process. In order to maintain Bali's status as the most popular tourist destination in Indonesia, the government must conduct an evaluation of hotel performance in order to maintain control over these alterations. Several methods, such as interviews with relevant sources, can be used to assess hotel performance. However, in the current Covid pandemic, assessment can be conducted with a data mining approach using sentiment analysis based on data obtained from website or social media sources about customer impressions after using hotel services in Bali [3].

The use of sentiment analysis methods alone is insufficient to define hotel aspects; therefore, Aspect Based Sentiment Analysis (ABSA) is used to obtain more specific results. The selected aspects refer to the aspects that the hotel provides to its customers in creating its own competitiveness, broadly speaking these aspects such as hotel safety, service, comfort, attitude of hotel staff and cleanliness of the hotel environment. The results obtained from this research are in the form of information that can be used by other parties such as the government or companies to address the improvement of the quality of business that runs after the company is transformed. Research that has been done by [4] that measures customer satisfaction in terms of the hotel's service, comfort, cleanliness, and location, which is beneficial for maintaining company stability.
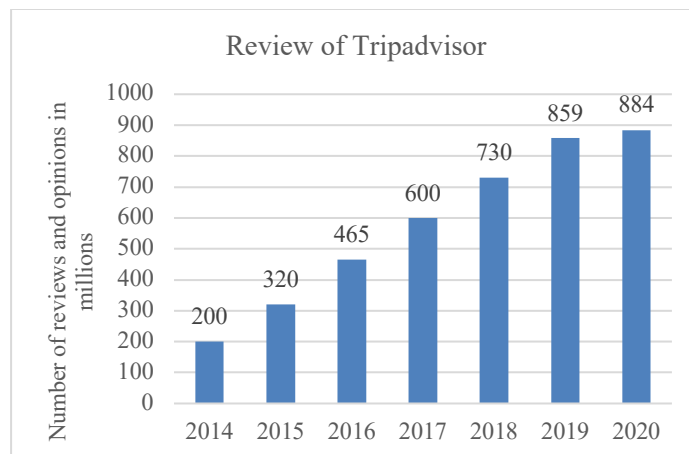

Fig. 1. Number of reviews on Tripadvisor website (Statista.com)

ABSA requires an optimal text processing and analysis algorithm. According to research [4], algorithms with Bidirectional Encoder Representations froms Transformers (BERT) architecture have the highest accuracy value compared to other architectural models such as Sent WordNet, Logistic Regression, and Long Short-Term Memory; therefore, BERT is ideally suited for use in this study's ABSA architecture. The necessary data for this research are customer reviews of Bali hotel services. Figure 1 from Statista.com demonstrates that Tripadvisor is a rich source of data due to the website's adoption of Web 2.0, which enables users to contribute to the creation of content, content, and other information, as evidenced by the number of reviews published each year, which rose between 2014 and 2020.

After implementing the new health protocol and collecting data from the Tripadvisor website, the government must evaluate hotel performance, which is segmented based on service, comfort, safety, and sanitation, using the sentiment analysis method. This investigation has yielded an evaluation of every aspect of the hotel.

2. RESEARCH METHOD
2.1. Aspect-based Sentiment Analysis (ABSA)

Aspect Based Sentiment Analysis (ABSA) is an aspect-based sentiment analysis method. The purpose of aspect-based sentiment analysis is to detect terms in the text that may appear explicitly so that each word has a different meaning when combined with other words. For

example, the following text "phone is out of date but the camera is still good" from these words there are two aspects, namely "camera" which means good and "hp" which means old. The method used by machine learning in analyzing sentiment needs to repeat terms and language structures (grammar) to recognize these two meanings. The purpose of grouping aspects as above is to produce recommendations from the customer's point of view [5]. Neural network architecture is very popular in making ABSA like the example above, because of its capability in encoding and decoding in the process of making architecture and can use vector dimensions in data processing [6].

2.2. Neural Networks

Neural Networks (NN) is one of the Machine Learning (ML) techniques that employs the deep learning architecture. The representation of NN is a human-like neural network composed of layered, interconnected nodes. The operation of NN is identical to that of the human brain. In NN, there are three primary layers: the input layer, the hidden layer, and the output layer, and each neuron or node is connected to those in the next layer, as shown in Figure 2.
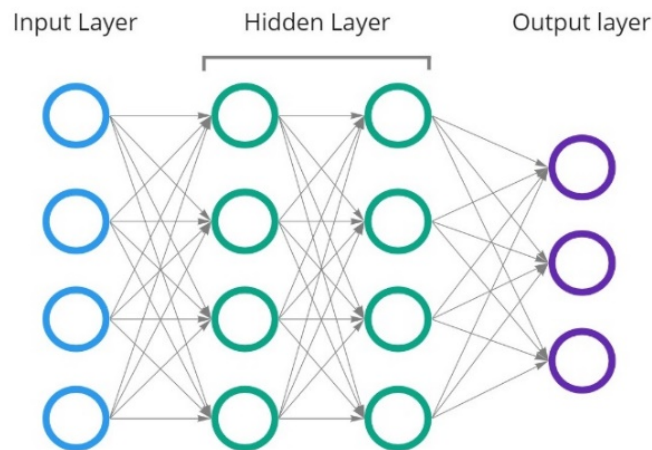

Fig. 2. Neural Network Representation

Data enters through the input layer and is forwarded to the hidden layer, where the hidden layer is the defined number, which then enters the next hidden layer and ultimately the output layer as a result of the NN process [7]. Each NN process can be repeated or referred to as a Recurrent Neural Network (RNN) to generate an optimal value, which gives NN an advantage over other ML techniques. RNN is a simple NN that loops each output back into the concealed layer. The cycle is repeated for the number of epochs determined [8].

2.3. Bidirectional Encoder Representation from Transformer (BERT)

BERT or stands for Bidirectional Encoder Representation from Transformer is a deep learning architecture for natural language processing (NLP) that was published in 2018. BERT uses the transformer technique to analyze the contextual relationship between one word and another in a sentence [9], [10]. Transformer uses data obtained from the self-attention mechanism; this mechanism changes the meaning of each word associated with the word that will be processed next. Transformer has two mechanisms called Encoder and Decoder as shown in Figure 3. The following is an explanation of each mechanism:
a. Encoder

Encoder serves to read the entire text input. Encoder is a stack of six layers. Each layer has a self-attention layer and a feed-forward neural network layer. Both layers help each node to focus on the overall semantic context of the word.
b. Decoder

Functioning to generate a sequence of prediction outputs, the Decoder is a stack containing the exact same six layers as the Encoder, but in each Decoder layer an attention-layer is added that helps nodes retrieve key content from other nodes. This process is done in parallel so that there will

be more than one attention layer that is processed and produces one output in the encoder process, this is referred to as multi-head attention [11].
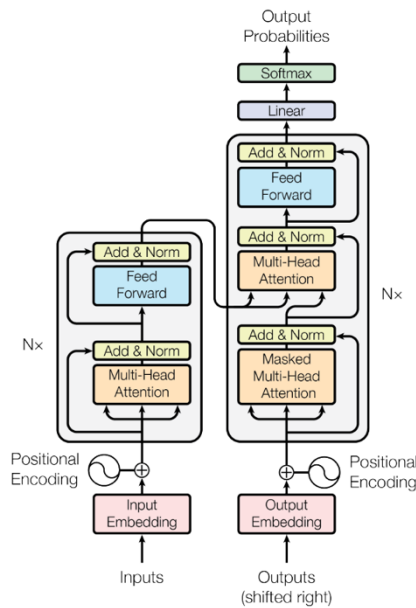

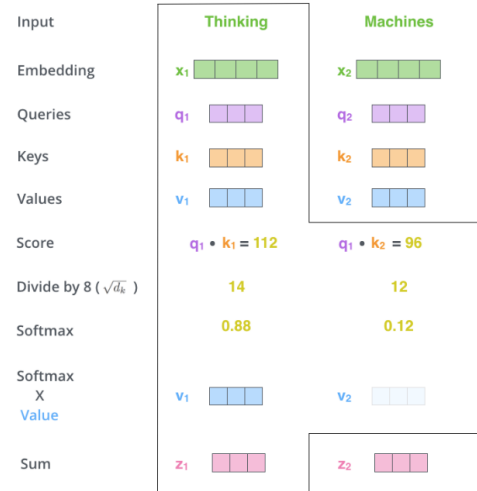Fig. 3. Encoder and Decoder Process [11]   Fig. 4. Calculation of Softmax [12]

Figure 3 shows the encoder and decoder process performed by the transformer in parsing text data. Each text input per word entered into the encoder is converted into a list vector using embeddings. Positional encoding is added to each word to indicate the position of each word. The input vector then passes through two layers, namely the self-attention layer and the feed-forward neural network. In the self-attention layer, three vectors will be created namely Query, Key, and Value. The self-attention value will be divided by 8 because the square root of the vector dimension is 64. The value is calculated from softmax so that the vector value is multiplied by the value of softmax as in Figure 4. Finally, the vector value is summed up to be the output of the self-attention layer, which is then continued into the feed forward neural network layer to input the next word. Every Encoder process is completed the key and value values enter the Decoder, all processes in the self-attention and feed-forward neural network are processed in the add and normalize layer. The process that occurs in Decoder, namely in the add and normalize layer, helps analyze relevant words. The output of each Encoder step will enter the Decoder and produce a vector value [12]. During the training process BERT will convert the word into a special token namely CLS, SEP and id token before proceeding to the Transformer process to give the position of each number [13].

3.   RESEARCH METHODOLOGY

The research begins by identifying problems regarding the need for hotel performance assessment which then aligns the objectives and solutions to be made. The next stage is data processing which contains a small stage in the form of data collection from the Tripadvisor website source using the webscrapping technique which is then labeled according to the selected aspects. The data is then cleaned at the data preprocessing stage, then the data is sampled using random oversampling and separated into two parts, namely 90% for training and 10% for testing. The next stage is the creation of the BERT architecture using the previous data and continued with the evaluation of the architecture using the calculation of accuracy, precision, recall, f1-score, confusion matrix, and ROC curve. The next stage implements the architecture that has been made on all data and analyzes the results of the architecture prediction. All of the above stages are illustrated in Figure 5.
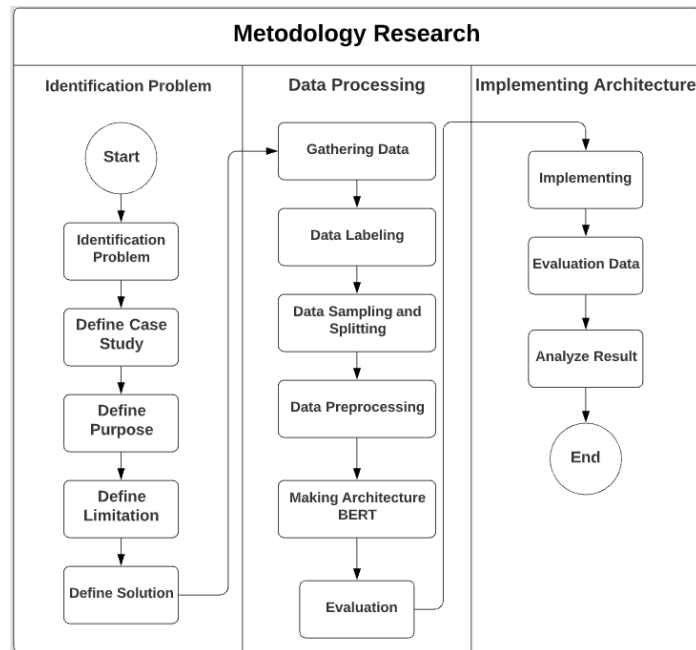
Figure 5. Flow of Research Methodology

The data used in the research was obtained from the Tripadvisor website. Data were selected from five hotels that have implemented health protocols; the hotels are sorted based on the number of customer reviews. Data retrieval is carried out with the help of the Prowebscrapper tool [14]. The flow of data retrieval is described in Figure 6.
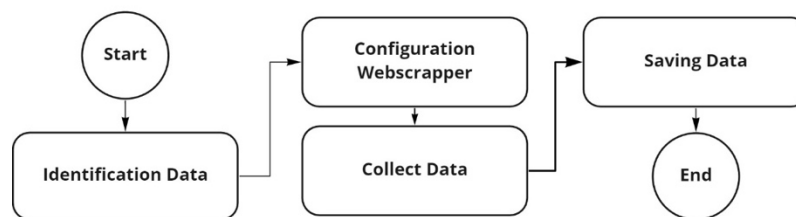

Figure 6. Text Preprocessing Flow

The first stage is identifying data needs and determining the attributes to be taken, then configuring Prowebscrapper according to the data needed, then running the previously created configuration to start data collection, and finally downloading the data that has been stored on Prowebscrapper. The results of the data obtained are shown in Table 1.

Table 1. Data Collection Results

Text	Title	Date Post	Date of Stay
An excellent hospitality, clean and cozy place. Very recommended for family. The animal zone really good and so fun for kids. The Spa treatment was excellent and so relaxing. The pool is amazing and so beautiful	Beautiful place	Galalima Glamping wrote a review Nov 2020	Date of stay: November 2020

The total amount of data retrieved was 3,419, and the selected hotels were Double-Six Luxury Hotel Seminyak, Le Meriden Bali Jimbaran, Mason Elephant Lodge, Padma Resort Ubud and Viceroy Bali. The five hotels were selected based on user ratings and the number of reviews given by hotel customers on the Tripadvisor.

The data that has been collected is then selected only the required attributes, namely text, and given additional labels in the form of sentiment labels (reviews) which contain positive and negative values and the second label is the safety aspect, cleanliness aspect, comfort aspect, and service aspect where the value 1 indicates that the text is indicated by a certain category, while the

value 0 is the opposite does not indicate a certain category. After adding the labels, the data will be in the form of Table 2.

Table 2. Labeling results

Text	Review	Aspect Safety	Aspect Cleanliness	Aspect Comfort	Aspect Service
Had an amazing time at the chillhouse - food was amazing, met some lovely people, learning to surf was awesome and the yoga was excellent. Recommended if you're looking to socialise but still have a really relaxing break.	Positive	0	0	1	1
Had a great time. Rooms were clean and air conditioned, food was amazing and the atmosphere was great for solo travelers. Yoga, surfing and organic food. I was there for 6 days and would definitely go back.	Positive	0	0	1	1

The data that has been labeled is then resampled due to the unbalanced data shown in Table 3. This table shows the number of labels for each aspect used, in this table there are differences in numbers which result in imbalanced data. The imbalanced data can produce an architecture that is biased towards the most classes because the few classes are considered as noise data, it can be likened to the architecture will be more able to predict the value that has more classes, but it will be difficult to categorize the fewer classes due to the mislead of the prediction architecture [15]. Therefore, sampling is needed to overcome the imbalance of data in this study, one of the sampling methods suitable for use is random oversampling, this technique equalizes the data by increasing the number of the lowest class by duplicating data from the lowest class until the number is equal to the highest class. The result of oversampling will change the data into more. This technique is used in all aspects of hotels and reviews so that the total data used in this study is as shown in Table 3.

Table 3. Number of classes for each aspect

Label	Number of class 1	Number of class 0
Review	2,847	453
Aspect Cleanliness	564	2,736
Aspect Safety	113	3,187
Aspect Service	2,575	725
Aspect Comfort	3,026	274

Before the data is processed, it must be set aside first through the data preprocessing stage, in this research the preprocessing stage is described in Figure 7. Text cleaning on stopwords using the NLTK[1] library that is already available for English.
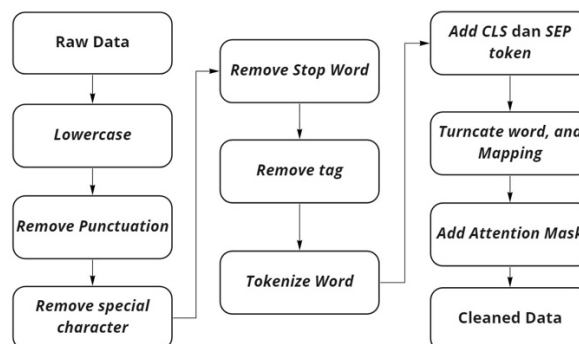

Figure 7. Flow of Text Preprocessing

[1] https://www.nltk.org/

Data that has been sorted and sampled is then cleaned at the text preprocessing stage before entering the training process. The data that is ready to be processed beforehand is then divided into two parts, namely training data which is used to create an architecture and testing data which serves as a validation of predictions by the architecture that has been made. Training data is taken 90% and testing data is taken 10% of the total data. This is followed by converting words into lowercase letters and changing English abbreviations such as "don't" to "do not" and other words that use the abbreviation "not" followed by the removal of tags such as "@" and the like, as well as punctuation or special characters. Followed by stopwords removal process that often appear and white spaces that come from the previous stage. The results of text cleaning can be seen in Table 4. The text that has been cleaned is then given a token in the form of SEP, CLS, and token id in each text.

Tabel 4. Text Cleaning

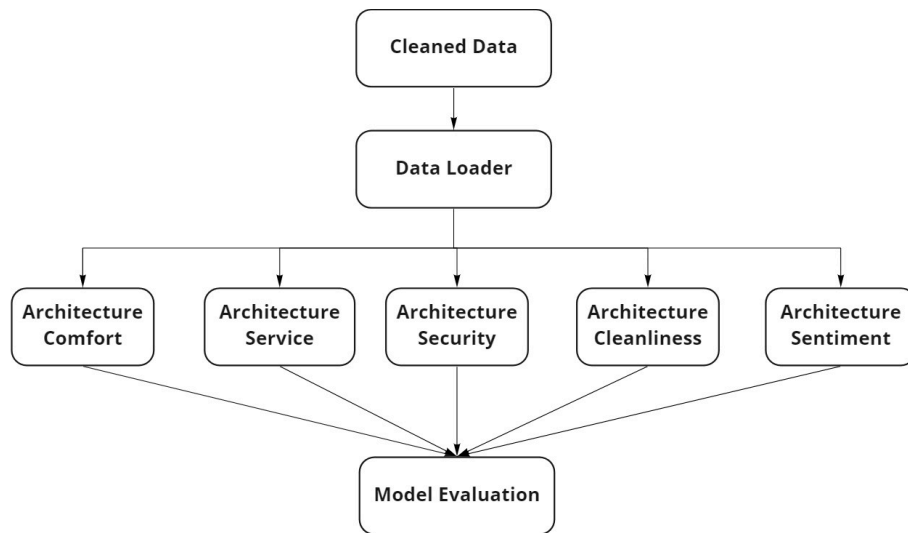| Before	After
I couldn't imagine a more beautiful place to stay. The staff were amazingly friendly. Could not fault the facilities and service. We didn't want to leave! We will be back for sure and won't stay anywhere else!	imagine beautiful place stay staff amazingly friendly could fault facilities service want leave back sure stay anywhere else

Fig. 8. BERT Architecture Generation

The BERT architecture is made by putting the cleaned data into the data loader, which splits the data into several groups. The training process comes next, and the architecture is then evaluated (as shown in Figure 8). The bert-base-uncease algorithm will be used to build the design, which will have 12 layers, 768 hidden layers, 12 heads, and 110M parameters. BERT utilizes the idea of "fine-tuning," which involves a little bit of extra setup in the form of an "optimizer" to handle processing of parameters during the training process. The batch size is 16, and the process is repeated twice (Epoch). Table 5 shows the BERT configuration used in this research.

Table 5. BERT Configuration

Architecture	Bert Base Uncased
Hidden size BERT	768 layers
Hidden size classifier	50 layers
Learning rate (Optimizer)	Adam
Label	2
Epoch	2

In this study, there are five kinds of labels used, so the results of the architecture created are also five, namely the review architecture (sentiment), cleanliness architecture, service architecture, comfort architecture, and safety architecture whose entire architecture uses the configuration and the process is repeated as much as the number of architectures created.

The training process produces training values lost (train loss), validation values lost in the process (val loss), the resulting accuracy (val accuracy), and the time required (elapsed) on each amount of data entered into the process (batch) and repetition done recursively (epoch) as in Table 6.

Table 6. Architecture Training Results

Architecture	Epoch	Train Loss	Val Loss	Val Acc	Elapsed
Review	1	0.19	0.14	96.1	73.02
	2	0.04	0.06	98.6	71.63
Service	1	0.24	0.10	97.5	66.02
	2	0.08	0.09	97.9	64.63
Comfort	1	0.12	0.05	98.8	6.46
	2	0.02	0.02	99.5	75.46
Cleanliness	1	0.22	0.05	98.9	69.88
	2	0.03	0.02	99.4	68.61
Safety	1	0.07	0.001	100	80.89
	2	0.001	0.0006	100	79.93

Each architecture that has been developed previously is evaluated using three methods: confusion matrix, classification report, and ROC curve [16]. The first method used is the confusion matrix where the confusion matrix is shaped like Table 7.

Table 7. Confusion Matrix

Classification	False	True
False	True Negative	False Negative
True	False Positive	True Positive

It is assumed that if the data is correct and predicted to be correct, it will be classified as True Positive (TP), if the actual amount of data is incorrect but the predicted value is correct, it will be classified as False Positive (FP), if the actual data is correct but the prediction is incorrect, it will be classified as False Negative (FN), and if the actual value is incorrect and predicted to be incorrect, it will be classified as True Negative (TN) [17], [18].

The next evaluation uses the classification report which includes the calculation of precision, recall, f1-score, and accuracy to measure the accuracy of the architecture [13]. Each of these calculations is obtained through the following formula:

$$Accuracy = \frac{TP + TN}{Total\ data} \qquad\qquad 1$$

$$Recall = \frac{TP}{(TP\ +\ FN)} \qquad\qquad 2$$

$$Precission = \frac{TP}{(TP\ +\ FP)} \qquad\qquad 3$$

$$F1\ Score = 2\ \text{x}\ \frac{(Recall\ x\ Precision)}{(Recall\ +\ Precision)} \qquad\qquad 4$$

4.   RESULTS AND DISCUSSION

For example, in the comfort aspect, the TP value is 335, and the TN value is 303, while both FP and FN values are 0. Based on the results of the confusion matrix on the five architectures created, the architecture can predict easily so that the TP and TN values are higher than FP and FN so that the accuracy value obtained is high around 90% and above on all architectural aspects. Table 8 summarizes the results of the confusion matrix on all aspects.

Table 8. Confusion Matrix Results

Aspect Name	TP	TN	FP	FN
Review	278	284	2	6
Cleanliness	283	262	0	3
Comfort	295	308	0	3
Service	247	257	1	10
Safety	335	303	0	0

The results of the ROC curve are depicted in Figure 10 where if the blue line of AUC score is close to the value of 1 on the TPR axis then the algorithm has a ratio of correct predictions more than the FPR or wrong predictions. The average of the five architectures created has an AUC value close to the value of 1 which forms a blue line on the TPR as shown in Figure 9 which is close to the value of 1 on the TPR axis. From all evaluation methods used, all architectures that have been created can predict correctly with an average accuracy of 97% and above, so that they can continue to be implemented on all data.
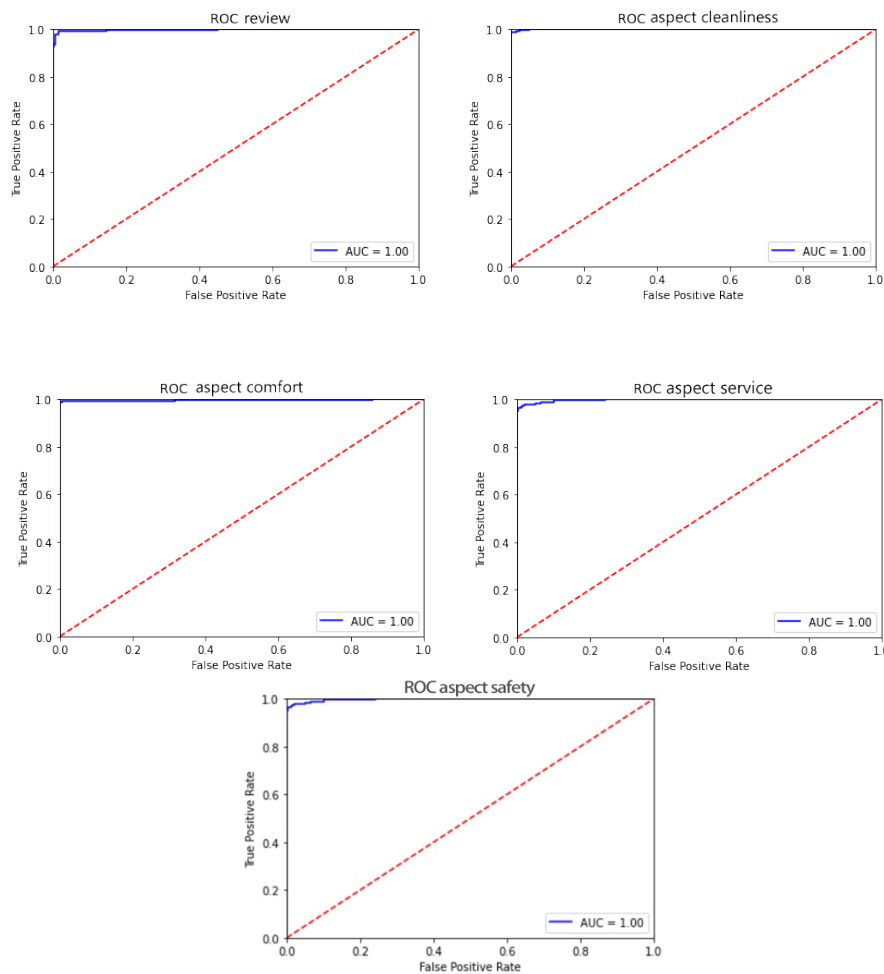

Fig. 10. ROC Curve results

The architectures that have already been made is then used on all of the data to predict each owned aspect. This is done in the same way that the architecture is cleaned by preprocessing text before processing, which is then put into the dataloader, and then the algorithm is used on the text. The outputs of all the data are shown in the following figure:

The results of this study which started from defining the problem, designing the architecture, and implementing the architecture made on data totaling 3,419 produced graphs Figure 10. The four figures show the number of positive and negative review relationships on hotel aspects. At a glance at all the results of the numbers and graphs, the average hotel in Bali has more positive impressions on each aspect than the negative reviews.

In Figure 10.a, the total number of predicted review labels on all data is displayed. Based on this figure there are five hotels with the number of positive reviews in orange and negative in maroon, from these results all hotels have more positive impressions than negative impressions, and the hotel that has the most positive impression is Le Meriden Bali Jimbaran.

a. Review

b. Aspect Service

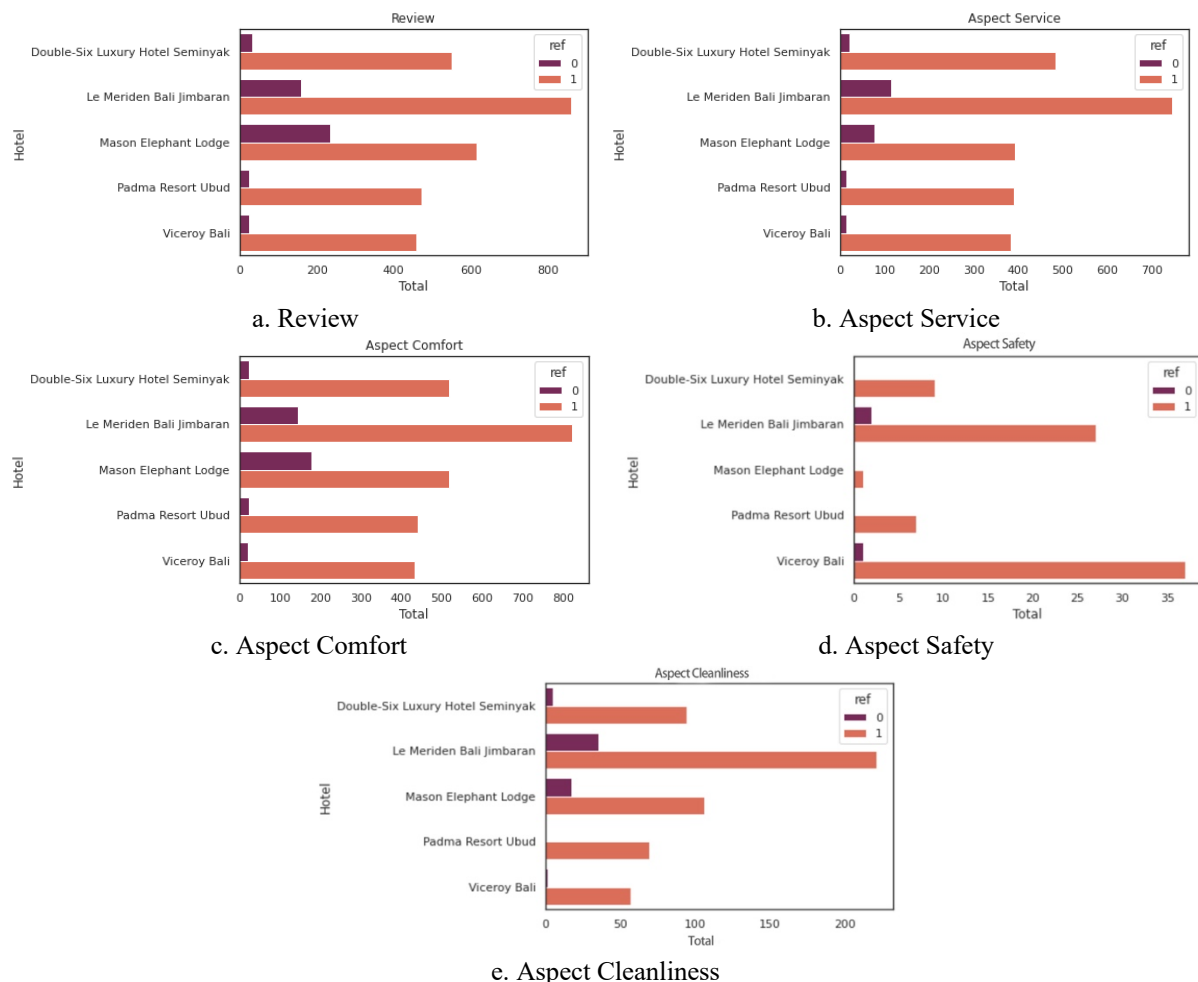c. Aspect Comfort

d. Aspect Safety

e. Aspect Cleanliness

Fig. 10. Final Label Results on Label Review and Each Aspect

The total amount of data labelled as service aspects is shown in Figure 10.b. Based on this figure shows hotel services that have positive and negative impressions, from these results all hotels have more positive impressions than negative impressions on the services provided covering everything from customer interactions with hotel employees, room booking services, and so on.

Figure 10.c illustrates the total amount of data labeled as comfort aspects. Based on this figure shows the number of positive and negative impressions of hotel comfort, the average of the five hotels tends to have a positive impression of both hotel rooms, cafes, swimming pools and others that include customer comfort in staying at the hotel.

The amount of data labeled with safety aspects is shown in Figure 10.d. Based on the figure shows a graph of the number of positive and negative impressions of the safety provided by the hotel including privacy, confidentiality, safety officers, and the like related to hotel safety. Of the five hotels, there are not too many texts that indicate safety, it can be seen that the number is around 0 to 35. Of the five hotels, Viceroy Bali hotel has safety that seems good among other hotels.

The number of positive and negative reviews on the aspect of hotel cleanliness is shown in Figure 10.e. Hotel cleanliness is assessed by how clean the hotel facilities are such as swimming

pools, rooms, restaurants and others. From these results, the average of the five hotels has a more positive impression than its negative impression.

In terms of comfort aspects, the amount of data available is around 0-800 and the positive impressions have more than negative in the five hotels. In the service perspective, the amount of data available is 0-700 which shows that the five hotels in Bali have satisfactory service. In the aspect of safety, the amount of data available is 0-35, in terms of data for safety aspects less than other aspects, but in this data positive reviews remain more than negative. In the aspect of cleanliness, the amount of data available is around 0-200, where the five hotels are kept clean as evidenced in the graph which shows more positive reviews than negative ones.

Based on the results of the implementation of this research, it can be concluded that the use of BERT architecture produces good accuracy values, the average value produced is more than 90% in each aspect used by using a batch size configuration of 16, Adam optimizer, and epoch 2. From the prediction data of 570, it can correctly predict 562 data. The results of the precision, recall, f1-score, and accuracy values of the five models created have an average value of 98%. The results of the implementation on all data, namely the Double-Six Luxury Hotel Seminyak, Le Meriden Bali Jimbaran, Mason Elephant Lodge, Padma Resort Ubud, and Viceroy Bali hotels, on average, have more positive impressions than negative reviews after implementing the new SOP calculated from various aspects ranging from cleanliness, comfort, service, and safety provided by hotel companies in Bali. It can be concluded that the average hotel in Bali has good performance in terms of aspects.

The data used is recommended to be balanced in terms of both aspects and the amount of data from the selected hotels, the data used can be taken from other sources so that the data is more varied not only sourced from the Tripadvisor website. The aspects studied in the next research can be more reproduced so that they can assess the hotel from a wider variety of perspectives, both in terms of customers, employees, and so on.

5.   CONCLUSION

Based on the results of the implementation of sentiment analysis based on aspects using the BERT algorithm for assessing hotels, the following conclusions are obtained.

1. Sentiment analysis using BERT produces an accuracy value on the review architecture of 95%, cleanliness aspect architecture of 99%, comfort aspect architecture of 99%, service aspect architecture of 95%, and safety aspect architecture of 100% by using a batch size configuration of 16, learning rate Adam, and epoch 2.

2. The sampling method used can affect the accuracy value of the architecture when implementing the BERT analysis sentiment. Random undersampling makes the data become less so that it makes the architecture use less training data so that the accuracy value will be lower at 74% on the service aspect than random oversampling, which duplicates the data to be more but the resulting accuracy is higher at 95% on the service aspect. Other labels also experience the same thing, so that random oversampling is suitable for this research.

3. From the implementation of all architectures in several hotels in Bali, namely Double-Six Luxury Hotel Seminyak, Le Meriden Bali Jimbaran, Mason Elephant Lodge, Padma Resort Ubud, and Viceroy Bali used in this study, on average, they have more positive impressions than negative impressions calculated from various aspects ranging from cleanliness, comfort, service, and hotel safety.

REFERENCES

[1]   O. : Zein, M. Muktaf, and E. R. Zulfiana, "Persepsi Wisatawan Asing Terhadap Wisata Indonesia," *J. Cakrawala ISSN*, vol. 1693, no. 1, pp. 83–106, 2018.

[2]   P. S. Pratiwi, "'Jokowi Imbau Masyarakat Bekerja dan Beribadah di Rumah' selengkapnya di sini:," *CNN Indonesia*, 2022. https://www.cnnindonesia.com/nasional/20200315141316-32-483586/jokowi-imbau-masyarakat-bekerja-dan-beribadah-di-rumah.%0A%0A

[3]   S. Nahavandi, "Industry 5.0-a human-centric solution," *Sustain.*, vol. 11, no. 16, 2019, doi: 10.3390/su11164371.

[4]   D. A. K. Khotimah and R. Sarno, "Sentiment analysis of hotel aspect using probabilistic latent semantic analysis, word embedding and LSTM," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 4, pp. 275–290, 2019, doi:

10.22266/ijies2019.0831.26.

[5]   M. E. Mowlaei, M. Saniee Abadeh, and H. Keshavarz, "Aspect-based sentiment analysis using adaptive aspect-based lexicons," *Expert Syst. Appl.*, vol. 148, p. 113234, 2020, doi: 10.1016/j.eswa.2020.113234.

[6]   X. Li *et al.*, "Enhancing BERT Representation with Context-Aware Embedding for Aspect-Based Sentiment Analysis," *IEEE Access*, vol. 8, pp. 46868–46876, 2020, doi: 10.1109/ACCESS.2020.2978511.

[7]   H. H. Do, P. W. C. Prasad, A. Maag, and A. Alsadoon, "Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review," *Expert Systems with Applications*, vol. 118. pp. 272–299, Mar. 2019. doi: 10.1016/j.eswa.2018.10.003.

[8]   D. Jurafsky and J. H. Martin, "Deep Learning Architectures for Sequence Processing," *Speech Lang. Process. (3rd ed. Draft.*, 2022.

[9]   X. Li, L. Bing, W. Zhang, and W. Lam, "Exploiting bert for end-to-end aspect-based sentiment analysis_," in *W-NUT@EMNLP 2019 - 5th Workshop on Noisy User-Generated Text, Proceedings*, 2019, pp. 34–41. doi: 10.18653/v1/d19-5505.

[10]   S. Alaparthi and M. Mishra, "Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey," *arXiv Prepr. arXiv2007.01127*, 2020.

[11]   A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[12]   J. Alammar, "The Illustrated Transformer 'https://jalammar.github.io/illustrated-transformer/,'" *jalammar.github.io*, 2018, [Online]. Available: https://jalammar.github.io/illustrated-transformer/

[13]   J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 2019, vol. 1, pp. 4171–4186.

[14]   A. V Saurkar, K. G. Pathare, and S. A. Gode, "An overview on web scraping techniques and tools," *Int. J. Futur. Revolut. Comput. Sci. \& Commun. Eng.*, vol. 4, no. 4, pp. 363–367, 2018.

[15]   J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, pp. 1–54, 2019.

[16]   S. Yang and G. Berdine, "The receiver operating characteristic (ROC) curve," *Southwest Respir. Crit. Care Chronicles*, vol. 5, no. 19, pp. 34–36, 2017.

[17]   R. Kumar, *Machine Learning Quick Reference : Quick and Essential Machine Learning Hacks for Training Smart Data Models*. Packt Publishing, 2019.

[18]   J. Han, M. Kamber, and J. Pei, *Data mining: Data mining Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2012. doi: 10.1109/ICMIRA.2013.45.