



Sentiment Analysis and Service Aspects Prioritization of Indonesian Hotel Reviews Based on ABSA and Pareto Approach

Rais Sulaiman Rusid^{1,*} and Isti Surjandari²

^{1,2} Industrial Engineering Department, Faculty of Engineering, Universitas Indonesia, Kampus UI Depok, Indonesia

*rais.sulaiman@ui.ac.id

Abstract. In the hospitality industry, feedback reviews play a vital role in shaping perceptions and influencing decision-making. These reviews are not only valuable for stakeholders but also serve as important information resources, for instance, to improve customer satisfaction. However, the large volume of unstructured data poses challenges for the industry. This study addresses these challenges by implementing an aspect-based sentiment analysis (ABSA) approach to evaluate service quality, which is an advanced approach to sentiment analysis from Natural Language Processing (NLP). ABSA not only identifies the sentiment polarity of text but also important elements in feedback, including terms, categories, opinions, and sentiment polarity. Using IndoBERT, an Indonesian language pre-trained model of BERT, and feedback data from Indonesian hotels, the study classifies sentiments into "positive" and "negative" categories. Results from IndoBERT demonstrate impressive performance in this classification task, achieving an accuracy of 98.55% and an F1 score of 98.78%. This indicates its effectiveness in capturing the nuances of sentiment in Indonesian text. Additionally, the study employs a quality analysis approach utilizing a Pareto Diagram to prioritize issues. The Pareto 80/20 principle helps identify the 20% of aspects that contribute most significantly to 80% of customer dissatisfaction. The results suggest that management should pay attention to the aspects of services and beverages to improve customer satisfaction. By combining ABSA, IndoBERT, and the Pareto method, this study enhances sentiment analysis and supports informed decision-making based on data.

Keywords: Aspect-Based Sentiment Analysis, BERT, Hotel Reviews, Pareto Diagram, Quality Tools.

1 Introduction

In the digital era, feedback reviews in customer surveys play a crucial role in shaping perceptions and influencing customer decisions, including those in the tourism and accommodation industries. The feedback is input for management to improve quality and information resources for prospective customers. However, a method that accurately

analyzes the feedback will be necessary because of the unstructured and abundant data. One approach uses sentiment analysis (SA), a branch of text mining. SA is used to understand the entities' opinions, feelings, and attitudes, such as products, services, and events [1]. The advanced task of sentiment analysis is Aspect-Based Sentiment Analysis (ABSA). ABSA identifies sentiment polarity and specific aspects of the opinion target that encompass important elements, including terms, categories, opinions, and sentiment polarity [2]. The Term aspect depicts a word or phrase referring to features of the product or service discussed. The category is a term that groups aspects into a particular category, which is then grouped for simplified analysis. Opinion draws the opinion or judgment of the term. Sentiment polarity points to the opinion or judgment, whether positive, neutral, or negative [2]. ABSA is not like conventional sentiment analysis. It can provide more detailed information because it not only determines the sentiment polarity of the text but also identifies specific aspects of feedback.

There are numerous studies on ABSA for Indonesian reviews, including their algorithms for feedback analysis, utilizing traditional machine learning [3] and deep learning [4]. Then, transformers, as an advanced approach to deep learning, appear as a method for improving Natural Language Processing (NLP) tasks. Some researchers use transformers for an ABSA task, namely mT5 [5,6] and BERT [7]. The studies that use BERT compared with other algorithms show that BERT outperforms mT5 [8,9]. In the Indonesian language context, IndoBERT presents itself as an Indonesian language pre-trained model [10,11]. IndoBERT has demonstrated its capabilities in sentiment analysis, especially ABSA. Studies from [12–14] utilize IndoBERT for ABSA and yield better performance than other models, such as traditional machine learning and deep learning. This study will use IndoBERT to analyze hotel reviews with ABSA data from [5,6]. Additionally, using the IndoBERT model from Wilie, et al. [11] performs better than another IndoBERT model and traditional machine learning and deep learning [12]. Therefore, this study will use IndoBERT because of the advantages mentioned before, especially for Indonesian reviews.

Lastly, in this study, the quality approach is employed because NLP facilitates generating aspects and sentiment and how the results can improve the quality of the services. The studies about service quality evaluation with text mining are promising and have potential to develop, but they have not been satisfactory because they have not received wider attention [15,16]. Also, a study from Vencovský [16] shows that few studies still use quality tools for service quality evaluation. Therefore, for this study, implementing the Pareto Diagram represents a novelty in in-service quality evaluation. NLP and Pareto Diagrams offer an effective solution for identifying problem priorities through data visualization.

This study contributes to conducting aspect-based sentiment analysis with IndoBERT, which is rare in the studies of quality service evaluation in Indonesia. This study uses Indonesian hotel reviews in the Indonesian language, which makes it clear what aspects contribute to hotels and accommodation in Indonesia, and with Pareto analysis, knowing which mainly contribute to the problem faced by customers, so that quality improvement strategies can be implemented effectively and efficiently.

The remainder of this article is organized as follows: Section 2 covers related works; Section 3 presents the materials and methods that will be used in the study; Section 4

reports the results of the experimental findings; Section 5 discusses the results of the study; Section 6 ends with the conclusions and the future research.

2 Related Works

ABSA, as the core of this research, has evolved over many years, and ABSA, along with the Indonesian language, has developed recently. There are three steps for conducting ABSA in this study: data collection, data preprocessing, and sentiment analysis. Before conducting sentiment analysis for the aspects, some studies will extract the aspects or opinions using an algorithm or a manual approach. The first study on the ABSA task in Indonesian languages was conducted by Gojali & Khodra [17] used Conditional Random Field (CRF) algorithm for aspect and sentiment extraction at once, which could extract the aspect and its opinions. The researchers will determine the sentiment polarity based on the results, where the sentiment will be labelled as negative if the results contain the negation words. Also, the extracted aspect can be categorized by the group that researchers can determine. A study from Ekawati & Khodra [18] conducted aspect categorization with a modified MaxEnt algorithm after the aspect extraction step with CRF algorithm. For sentiment classification, the researchers use the same algorithm as aspect categorization. The studies mentioned earlier used a supervised approach for the ABSA task. Sasmita, et al. [19] started using an unsupervised approach because many NLP tools are mainly for English, and under-resourced Indonesian language NLP tools. A study from Surjandari, et al. [3] used Latent Dirichlet Allocation (LDA) as a topic modeling method, identified topics within documents and a corpus [20] for specifying the aspects, and also categorizing at once. Each review will be classified based on the aspect. With Support Vector Machine (SVM) as a supervised machine learning algorithm for sentiment analysis for each aspect.

The emergence of deep learning for NLP adds knowledge for ABSA tasks for the Indonesian Language. Several studies have utilized deep learning for ABSA tasks and generated higher performances than traditional NLP and machine learning. Besides aspect and sentiment, the opinion that determined the sentiment of the aspect can be extracted. A study by Cahyadi & Khodra [21] employed a feedforward neural network with a one-vs-all strategy for aspect category classification, CRF for opinion target expression extraction, and Convolutional Neural Network (CNN) for sentiment polarity. Achieved the highest score, 87%, 76.4%, and 78.7%, respectively. Lastly, Chamid et al. [22] conducted research ABSA task with a graph-based Graph Convolutional Network (GCN) and Graph Recurrent Network (GRN) for aspect and opinion extraction, and semi-supervised with CNN and Recurrent Neural Network (RNN) for sentiment classification. The study showed that with GRN and CNN results were good performance with 97.14% and 94.02%.

As mentioned, transformers, developed from deep learning, are the basis of a state-of-the-art model for NLP. Transformers use attention mechanisms instead of recurrence and convolutional, where these two approaches were removed entirely [23]. There are a few methods using transformers for Indonesian ABSA tasks. A study by William & Khodra [5] applied a pre-trained model, T5, for the Indonesian language IndoT5 to

extract an opinion triplet (aspect term, opinion term, and sentiment polarity) from Indonesian hotel reviews. This study's advantage lies in extracting explicit opinion triplets and implicit opinion triplets through the Generative Aspect-based Sentiment Analysis (GAS) framework, which utilizes a sequence-to-sequence approach to generate sentiment elements directly from the input. With IndoT5 and GAS framework, researchers achieved an F1 score of 79.5%. Improvement was conducted by William & Khodra from Suchrady & Purwarianti study [6], where another framework was LEGO-ABSA. LEGO-ABSA was a methodology that could build an assembly task prompt for the specific ABSA [24]. This framework, being improved by Indo LEGO-ABSA, combines multitask learning and prompting to increase the generative ABSA model for Indonesian language and achieves better performance than the GAS framework for extraction. Another model is Bidirectional Encoder Representations from Transformers (BERT), proposed by Devlin et al. [25]. Some studies compare BERT and T5, demonstrating that BERT outperforms T5 [8,9]. In the Indonesian review context, studies have compared BERT, using the Indonesian pre-trained model IndoBERT, and other models, showing that BERT is better than other models like deep learning or machine learning [12–14]. For this study, data have been labelled with aspect, opinion, and sentiment. The experiment will conduct the ABSA with IndoBERT based on aspects and sentiment formed to predict the sentiment of the reviews.

After conducting ABSA task, aspects from the review will sometimes have significant amounts and need to be categorized to continue the analysis. For example, to understand aspects that are frequently written by customers that probably have many aspects, clustering can be helpful when the goal of the study continues to the exploration of the aspects, where in this study, quality analysis from the voice of the customers, collecting aspects of reviews, clustering aspects, and topic modeling will be relevant for this study. Topic modeling has been used to find topics in documents. In this context, it is used to find topics from aspect terms. LDA has been used frequently, as seen in a study by [3], which focused on finding topics from public opinion regarding ride-hailing service providers, and in Hou et al. [26], for identifying regional flood issues and needs identification in China. Furthermore, the grouping aspect that integrates with LDA results in a more accurate grouping of textual data [27]. The study conducted by Chang et al. [28] employed various topic modeling methods, including Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA), and Non-negative Matrix Factorization (NMF), and then found that LDA got a higher coherence score than other methods. K-Means and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) were utilized for clustering. Another study conducted by Parfenova [29] investigated various topic modeling methods, such as BERT and LDA, as well as clustering methods like K-Means and HDBSCAN for interview text analysis and found that combining BERT and HDBSCAN results in the best coherence score. Using BERT for topic modeling, Grootendorst [30] built a BERTopic that generates document embeddings with pre-trained models, clusters these embeddings, and generates topic representations with class-based Term Frequency-Inverse Document Frequency (c-TF-IDF). Many studies have applied BERTopic for topic modeling and clustering approaches. A study by Araújo et al. [8] utilized BERTopic for the topic modeling evaluation of three datasets of product reviews. BERTopic starts used for Indonesian text.

Research by Mahfudiyah & Alamsyah [31] utilized BERTopic with IndoBERT as a pre-trained language model for BERTopic. Additionally, Asnawi et al. [32] used this model to generate coherent and interpretable topics, especially for short and unstructured text from Indonesian news platforms. By integrating ABSA with IndoBERT and topic modeling and clustering with BERTopic, this kind of research can provide a deeper insight into service aspects that receive positive and negative attention.

Lastly, in this research, the quality approach is employed because NLP is applied to generate aspects and sentiment, and how the results can improve the quality of the services. Text mining has been a promising method for evaluating service quality [16]. One of the quality approaches that can be utilized is quality tools. Some studies have utilized control charts for monitoring process stability through statistical variation analysis [33,34]. However, a systematic literature review by Vencovský [16], aside from [33,34], revealed no quality tools for service quality evaluation using text mining. Therefore, for this study, implementing the Pareto Diagram is a novelty for service quality evaluation with NLP. From all related works discussed, this study will specifically focus on developing frameworks integrating ABSA, clustering aspects generated with topic modeling, and finally, the Pareto Diagram will help evaluate the service quality. The scope of this study is a comprehensive evaluation of hospitality service quality based on customer feedback.

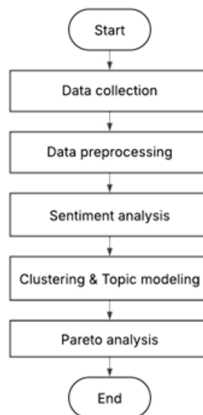


Fig. 1. Research methodology

3 Materials and Methods

Figure 1 depicts the research methodology. The first step focuses on collecting data. After that, the raw data is entered into data preprocessing, sentiment analysis, clustering, and topic modeling. The final step is evaluation, where the Pareto Diagram is used to analyze and prioritize findings.

3.1 Dataset

A dataset used in this study is a collection of Indonesian hotel reviews taken from [5,6]. The data contains 5000 review texts broken down into 3000 train data, 1000 validation data, and 1000 test data. The change removes neutral sentiment data because only training data and test data contain this, not as much as other sentiments (positive and negative). Moreover, the amount of data will change because the data must be melted for use with BERT, which will be discussed in the next section. The composition of data is feedback (the column named "text") and triplet opinion (aspect-opinion-sentiment), where a triplet opinion will be transformed into the index number of the "text" column (the column named "num_targets"). For example, the feedback "cukup bersih dan wifi okelah." ("quite clean and wifi is okay") with num_targets ([[3], [4], 'POS'], ([-1], [0, 1], 'POS'])). The "wifi" is an aspect term with index number [3], "okelah" ("okay") is an opinion term with a number index [4], and the sentiment polarity is positive or "POS". Determining the index for num_targets needs attention to punctuation. Additionally, there is an aspect term with an index of [-1], which [4] refers to as an implicit aspect and will transform into "NULL".

3.2 Preprocessing Data

BERT is a model that inputs one sentence or a pair of sentences as a single token sequence. The classification task will be represented by one of those tokens [25]. Therefore, because the raw data has more than one label/class, the first step of preprocessing data is to melt the text and num_targets per row with the exact text. Because the data is melted, the amount of data changes, unlike raw data. This approach corresponds to the BERT used for the ABSA task in previous studies, known as Aspect-Based Sentiment Analysis (ABSA) via constructing auxiliary or sentence-pair classification [12,35]. This approach involved merging the review text or text column with auxiliary sentences that contain num_targets, which include aspect terms, opinion terms, and sentiment polarities. These terms must be transformed into words related to the index. However, sentiment polarities will not be included in the input for the BERT model. Lastly, data must be preprocessed using other methods, such as lowercase conversion, padding, and tokenization. An example of input data for the model is presented in Table 1.

Table 1. The example of input data for the model

text	aspect and opinion	label
<i>cukup bersih dan wifi okelah.</i> (quite clean and wifi is okay)	NULL cukup bersih	positive
<i>cukup bersih dan wifi okelah.</i> (quite clean and wifi is okay)	wifi okelah	positive

3.3 Sentiment Analysis with IndoBERT Fine-Tuned Model for ABSA via Constructing Auxiliary (Sentence-Pair Classification)

Previous studies using this data employed the T5 model for the ABSA task [5,6]. For this study, IndoBERT, a pre-trained BERT model using the Indonesian corpus, will be employed [11]. The IndoBERT model that will be used is from Wilie et al. [11], as it has been pre-trained on 4 billion Indonesian words and outperforms another IndoBERT model [12]. Therefore, this model will be applied to the ABSA task with a modification that incorporates an auxiliary construct as an input [12,35], as shown in Table 1. As mentioned earlier, there is a preprocessing phase that involves tokenization. The tokenization will utilize BERT, specifically WordPiece [12]. Each input will be tokenized, and there will be a classification token ([CLS]) placed in the first position, then a separator token ([SEP]) placed between sentences in one input [12]. Because this model uses a sentence-pair classification approach, the separator tokens will be placed between feedback text and aspect and opinion sentences, as well as at the end of the input. The formula of the BERT token for sentence-pair classification is written as:

$$S = ([CLS], C, [SEP], A, [SEP]) \quad (1)$$

Where C is the feedback text, A is the aspect and opinion text, and [CLS] and [SEP] are unique tokens given by BERT. The IndoBERT pre-trained model will be fine-tuned with input sequences for adaptation to the ABSA task, and an output layer will be added, including a softmax layer and a linear layer, which will be used as the last layer in the IndoBERT architecture. The final vector will result from the model, and only [CLS] will be used as a fixed-dimensional composite representation of the entire input sequence, which has been processed using the IndoBERT output layer to determine the probability of sentiment in each aspect. Fine-tuned will follow a recommendation from [12] with a change in the number of epochs to 5. For other parameters: the learning rate = $2e-5$, batch size = 32, maximum sequence length = 128, loss function = Cross Entropy Loss, optimizer = Adam, and dropout rate = 0.1, while for evaluation strategy and save strategy on each last epoch and evaluation with F1 score based on [6]. This configuration ensures good results and evaluation for this task.

3.4 Evaluation

To evaluate the effectiveness of sentiment analysis in classification, performance metrics enable the quantification of the task's performance. For a classification task, the results can be depicted in a matrix called the confusion matrix [36]. The confusion matrix categorizes the results into four categories, depending on the actual label for this study, which is the label derived from raw data and the predicted label: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). To summarize the evaluation, four basic metrics will be used for the classification task: Accuracy, Precision, Recall, and F1 Score [37]. For each definition, accuracy is the percentage of data correctly predicted, that is, $(TP + TN)/(TP + TN + FP + FN)$. Precision is the percentage of data predicted to be "positive" that is actually "positive," which is

$(TP)/(TP + FP)$. Recall is a percentage of "positive" samples that were correctly predicted, that is, $(TP)/(TP + FN)$. F1 Score is the harmonic mean of precision and recall.

3.5 Clustering and Topic Modeling

This step is taken because of several aspects that necessitated the clustering for the Pareto Diagram. The advantages of BERTopic for clustering and topic modeling include its ability to process natural language, conduct topic modeling dynamically, and its compatibility with the diverse and complex nature of reviews, allowing for more nuanced and accurate feedback analysis [8]. The phases of BERTopic involve yielding document embeddings with a pre-trained transformer model, clustering embeddings, and, finally, creating topic representations using Class Term Frequency-Inverse Document Frequency (c-TF-IDF) [30]. For topic clustering, HDBSCAN is the default method of BERTopic clustering, which is an improvement of DBSCAN that creates clusters based on density. HDBSCAN will then convert these clusters into a clustering hierarchy [30]. Before clustering, BERTopic will reduce the vector from embedding with Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP). The dimensionality-reducing technique UMAP has been proven to maintain global information while maintaining a reasonable processing time [8]. Moreover, a study by Allaoui et al. [38] showed that this technique improves accuracy and reduces the time required for K-Means and HDBSCAN. In the final step of BERTopic, after clustering the topics, the representative topics will be modeled based on the documents in each cluster, where each cluster will be assigned one topic. BERTopic modifies TF-IDF to c-TF-IDF, where the importance of words in a document is measured. However, for c-TF-IDF, the term "document" is replaced by "class", which is the merged representation of all texts in one cluster [30]. The formula for c-TF-IDF from Grootendorst is:

$$W_{t,c} = tf_{t,c} \cdot \log\left(1 + \frac{A}{tf_t}\right) \quad (2)$$

Where $tf_{t,c}$ is frequency term, as the frequency of term t in class c , that is how t frequently appears in class, and for document inverse frequency that replaced by class inverse frequency is measuring how much the information given by t term to c class as the logarithmic of average the numbers of a word each class divided by the frequency of term t tf_t and add one for resulting positive value.

4 Results

This section explains the application of IndoBERT for ABSA task. The results will be used for clustering and topic modeling with BERTopic, where the aspects are grouped to determine the aspect that will be solved first based on frequency through visualization with a Pareto Diagram. This process involves sorting the aspects and selecting the problem that will be solved.

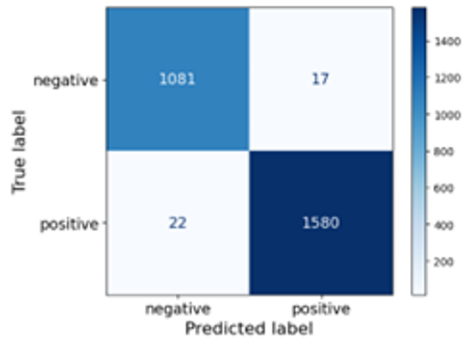


Fig. 2. Confusion matrix

4.1 Sentiment Analysis

In this section, the results of sentiment analysis with IndoBERT will be shown. The model achieves good performance in each evaluation, with a fast training time and a small number of epochs. The training time was 16.5 minutes, and the evaluation scores were accuracy of 98.86%, precision of 98.81%, recall of 98.77%, and F1 score of 98.79%. To understand whether the model classifies data well, it is then used again to predict test data. The confusion matrix in Figure 2 shows that the model can predict test data effectively, with an accuracy of 98.55%, precision of 98.62%, recall of 98.93%, and F1 score of 98.78%. From 1097 negative sentiment data points, 17 did not predict correctly (False Positive), while from 1602 positive sentiment data points, 22 did not predict correctly (False Negative). The results show that the model struggles to recognize the implicit aspects and the feedback that combines Indonesian and English

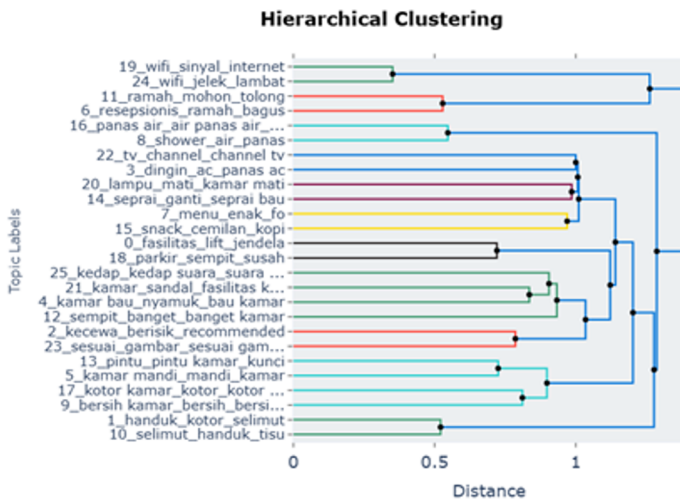


Fig. 3. Negative sentiment topic dendrogram

4.2 Clustering and Topic Modeling

Aspects and opinion sentences from data used for this step are from the previous step, where the data was split into positive and negative data. With BERTopic, the model generates positive and negative sentiment data clustering. HDBSCAN results cluster -1 if the data does not belong to any clusters. Cluster -1 will not be used in the analysis. HDBSCAN is a hierarchical clustering algorithm, so the visualization is presented as a dendrogram, which can be seen in Figure 3 for negative sentiment and Figure 4 for positive sentiment. The left side of each graph indicates the number of topics and their names, represented by the words representing the clusters. There are 26 clusters for negative sentiment and 27 clusters for positive sentiment.

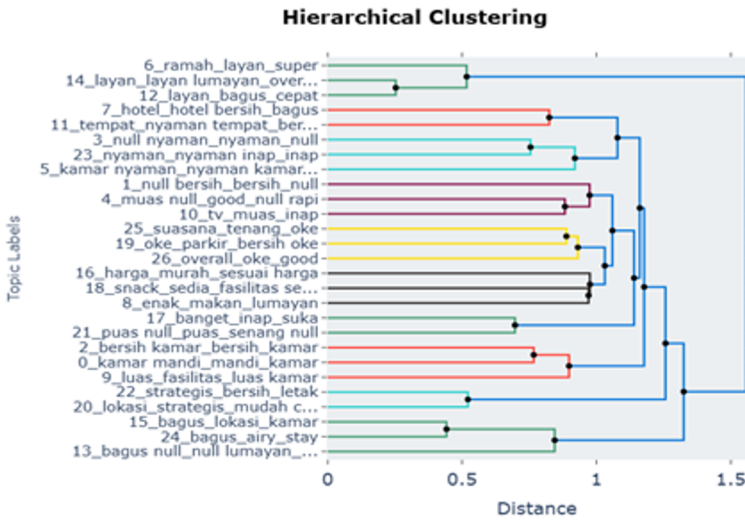


Fig. 4. Positive sentiment topic dendrogram

4.3 Pareto Diagram

The final step of the study is to create the Pareto Diagram. BERTopic can show the frequency of each cluster. The frequency can be used to calculate the cumulative frequency. Figures 5 and 6 are Pareto diagrams for negative sentiment and positive sentiment, respectively. According to the Pareto Principle, prioritizing problems based on negative sentiment ranges from topic 0 to topic 17, and for positive sentiment, it ranges from topic 0 to topic 18. For each topic generated from aspect and opinion sentences by the BERT model, these topics will be visualized using a Pareto Diagram. Each topic will be named based on what is contained in each topic.

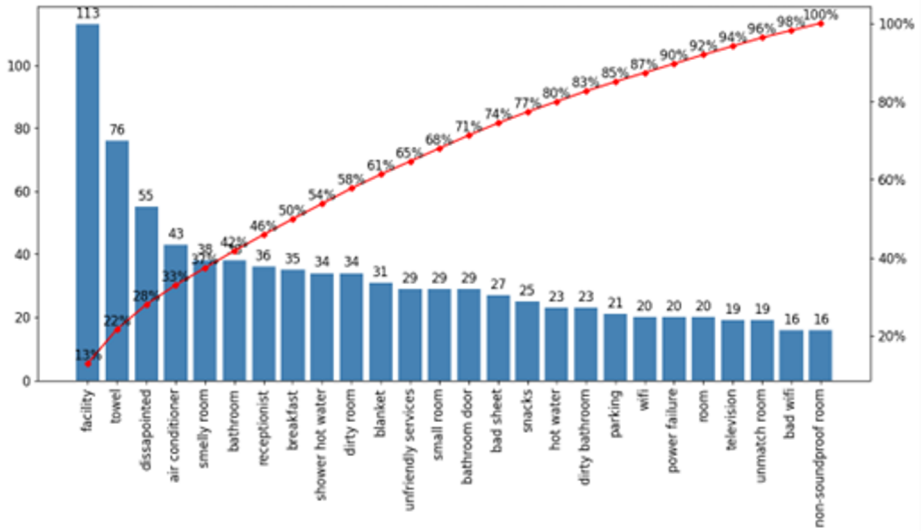


Fig. 5. Pareto diagram for negative sentiment

5 Discussion

The application of IndoBERT for the ABSA task yields better performance than previous studies using the same data and the T5 model [5,6]. Furthermore, according to their studies, IndoBERT's training is faster than that of T5. However, this study did not re-conduct the T5 training. However, this is related to other studies mentioned earlier, which indicate that BERT outperforms T5 [8,9]. IndoBERT outperforms other algorithms, such as deep learning and machine learning, because the IndoBERT model is a pre-trained model based on an Indonesian corpus, allowing it to understand the meaning of Indonesian feedback. However, in previous findings, IndoBERT cannot be classified effectively if the feedback is in English. Then, for the following discussion about clustering and topic modeling with BERTopic, this study can obtain more nuanced and representative topics with feedback. Embedding from IndoBERT pre-trained results on a topic more clearly aligned with studies from [31,32]. Lastly, after generating negative and positive sentiment aspects and conducting clustering and topic modeling, resulting topics, each topic will be visualized using a Pareto Diagram. This approach represents a novel approach to service quality evaluation using NLP and text mining. Applying the Pareto Diagram involves selecting the aspects that will be refined first, focusing on those that account for 80% of the effect. There are 18 topics or aspects related to negative sentiment and 19 topics related to positive sentiment.

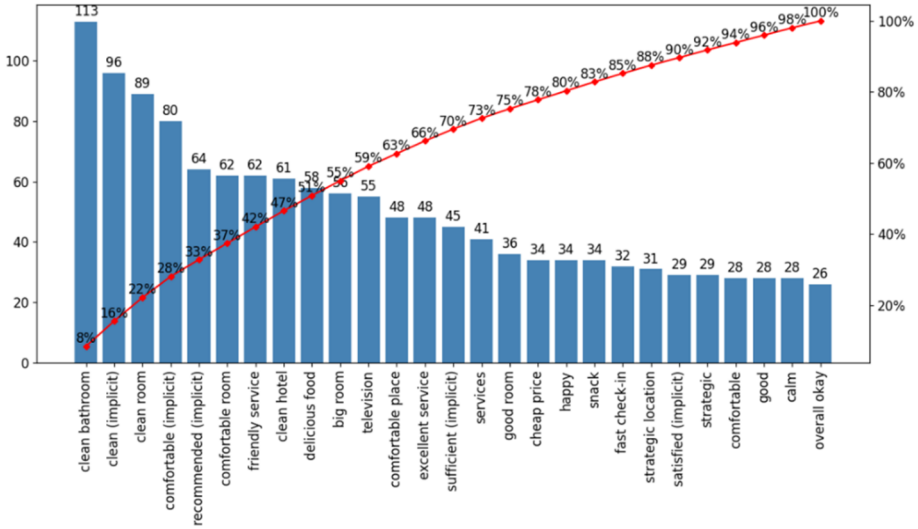


Fig. 6. Pareto diagram for positive sentiment

Each topic will be named based on what is contained in the topic. For the topics discussed by customers in negative feedback, the most categories mentioned in cluster 0 are: *facility, towel, disappointed (implicit), air conditioner, smelly room, bathroom, receptionist, breakfast, shower hot water, dirty room, blanket, unfriendly services, small room, bathroom door, bad sheet, snacks, hot water, dirty bathroom, and parking.* For positive sentiment data are *clean bathroom, clean (implicit), clean room, comfortable (implicit), recommended (implicit), comfortable room, friendly services, clean hotel, delicious food, big room, television, comfortable places, excellent service, sufficient (implicit), services, good room, low price, happy, snack, and fast check-in.*

Based on the Pareto analysis, some points can be discussed. First, the implicit words including "disappointed", "clean", "comfortable", and "recommended" indicate that the aspect from the data is "NULL", assuming that the implicit category is overall hotel services. Second, there are more implicit aspects mentioned in positive sentiments than in negative sentiments, indicating that customers tend to pay more attention to unsatisfactory aspects. Lastly, the majority outcome of the category from negative and positive sentiment is about services and some words for beverages, which is similar to the result from Kim & Kim [39], who used text mining to analyze customer satisfaction in hotel reviews, which showed that services and dining influence customers' satisfaction. The researchers also suggest increasing satisfaction with those aspects shaping positive customer reviews.

This study provides academic and managerial implications. For academic purposes, this study gives a framework for opinion mining customer reviews with advanced methods to understand customer opinions better. Combining text mining and a quality approach enhances the effectiveness of opinion mining by extracting text and analyzing it directly to understand the problem. For managerial implications, industries aware of the aspects identified through feedback can effectively address and resolve problems

likely to be addressed more efficiently. Additionally, with a quality approach like the Pareto Diagram, stakeholders can identify the problems that frequently occur to customers. Then, the problems can be resolved first, and the main issues can be addressed. This can happen if the company has limited resources and cannot resolve it simultaneously. This study evaluates customer hotel selection aspects necessary for improving service quality and customer satisfaction to obtain a competitive advantage [39]. Additionally, this study can provide a nuanced analysis of feedback using text mining, not only for hospitality services but also for other services. However, different services probably have different problems. Therefore, this study will be reproduced for other services for future research.

6 Conclusion

This study demonstrates that the aspect-based sentiment analysis (ABSA) approach, utilizing the IndoBERT model, is practical in categorizing Indonesian-language hotel reviews. This model achieved 98.55% accuracy, 98.62% precision, 98.93% recall, and 98.78% F1 score in a relatively short time, around 16.55 minutes. These results indicate that IndoBERT effectively and efficiently classifies sentiment in hotel customer reviews. Although there are obstacles in recognizing reviews that use a mixture of Indonesian and English, the confusion matrix results show that this mode has a low error rate. Additionally, clustering offers valuable insights into the key factors influencing positive and negative customer sentiment. Thus, IndoBERT is a reliable model for Aspect-Based Sentiment Analysis in Indonesian-language hotel reviews. Lastly, the quality approach can help resolve some problems with the Pareto diagram, where the tool visualizes aspects that frequently occur for customer problems. This study only covers hotel reviews. The following research can be expanded to other domains, such as restaurants, transportation, or public services, to test the generalizability of the ABSA model in various contexts and compare the performance of IndoBERT with other models.

Acknowledgments. This work was supported by the Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan - LPDP), in the form of support for research and publication.

Disclosure of Interests. The authors declare no conflicts of interest.

References

1. Liu, B.: Sentiment Analysis and Opinion Mining. Springer International Publishing, Cham (2012)
2. Zhang, W., Li, X., Deng, Y., Bing, L., Lam, W.: A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges. *IEEE Trans. Knowl. Data Eng.* 35, 11019–11038 (2023). <https://doi.org/10.1109/TKDE.2022.3230975>

3. Surjandari, I., Wayasti, R.A., Zulkarnain, Z., Laoh, E., Masbar Rus, A.M., Prawiradinata, I.: Mining Public Opinion on Ride-Hailing Service Providers using Aspect-Based Sentiment Analysis. *IJTech.* 10, 818 (2019). <https://doi.org/10.14716/ijtech.v10i4.2860>
4. Fernando, J., Khodra, M.L., Septiandri, A.A.: Aspect and Opinion Terms Extraction Using Double Embeddings and Attention Mechanism for Indonesian Hotel Reviews. In: 2019 International Conference of Advanced Informatics: Concepts, Theory and Applications (ICAICTA). pp. 1–6. IEEE, Yogyakarta, Indonesia (2019)
5. William, Khodra, M.L.: Generative Opinion Triplet Extraction Using Pretrained Language Model. In: 2022 9th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA). pp. 1–6. IEEE, Tokoname, Japan (2022)
6. Suchrady, R.Z., Purwarianti, A.: Indo LEGO-ABSA: A Multitask Generative Aspect Based Sentiment Analysis for Indonesian Language. In: 2023 International Conference on Electrical Engineering and Informatics (ICEEI). pp. 1–6. IEEE, Bandung, Indonesia (2023)
7. Azhar, A.N., Khodra, M.L.: Fine-tuning Pretrained Multilingual BERT Model for Indonesian Aspect-based Sentiment Analysis. In: 2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA). pp. 1–6. IEEE, Tokoname, Japan (2020)
8. De Araújo, P.A.M., Haddadi, S.J., Silva, F.D.S., Reis, M.D.S., Reis, J.C.D.: Topic Modeling Influence in Sentiment Analysis from User-generated Product Reviews. In: 2024 32nd International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE). pp. 110–115. IEEE, Reggio Emilia, Italy (2024)
9. Shaik Vadla, M.K., Suresh, M.A., Viswanathan, V.K.: Enhancing Product Design through AI-Driven Sentiment Analysis of Amazon Reviews Using BERT. *Algorithms.* 17, 59 (2024). <https://doi.org/10.3390/a17020059>
10. Koto, F., Rahimi, A., Lau, J.H., Baldwin, T.: IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP, <https://arxiv.org/abs/2011.00677>, (2020)
11. Wilie, B., Vincentio, K., Winata, G.I., Cahyawijaya, S., Li, X., Lim, Z.Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., Purwarianti, A.: IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. In: Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing. pp. 843–857. Association for Computational Linguistics, Suzhou, China (2020)
12. Yulianti, E., Nissa, N.K.: ABSA of Indonesian customer reviews using IndoBERT: single- sentence and sentence-pair classification approaches. *Bulletin EEI.* 13, 3579–3589 (2024). <https://doi.org/10.11591/eei.v13i5.8032>
13. Jazuli, A., Widowati, Kusumaningrum, R.: Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback. *Applied Sciences.* 15, 172 (2024). <https://doi.org/10.3390/app15010172>

14. Bahri, C.A., Suadaa, L.H.: Aspect-Based Sentiment Analysis in Bromo Tengger Semeru National Park Indonesia Based on Google Maps User Reviews. *Indonesian J. Comput. Cybern. Syst.* 17, 79 (2023). <https://doi.org/10.22146/ijccs.77354>
15. Liu, X.-X., Chen, Z.-Y.: Service quality evaluation and service improvement using online reviews: A framework combining deep learning with a hierarchical service quality model. *Electronic Commerce Research and Applications.* 54, 101174 (2022). <https://doi.org/10.1016/j.elerap.2022.101174>
16. Vencovský, F.: Service Quality Evaluation Using Text Mining: A Systematic Literature Review. In: Buchmann, R.A., Polini, A., Johansson, B., and Karagiannis, D. (eds) *Perspectives in Business Informatics Research.* pp. 159–173. Springer International Publishing, Cham (2020)
17. Gojali, S., Khodra, M.L.: Aspect based sentiment analysis for review rating prediction. In: 2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA). pp. 1–6. IEEE, Penang, Malaysia (2016)
18. Ekawati, D., Khodra, M.L.: Aspect-based sentiment analysis for Indonesian restaurant reviews. In: 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA). pp. 1–6. IEEE, Denpasar (2017)
19. Sasmita, D.H., Wicaksono, A.F., Louvan, S., Adriani, M.: Unsupervised aspect-based sentiment analysis on Indonesian restaurant reviews. In: 2017 International Conference on Asian Language Processing (IALP). pp. 383–386. IEEE, Singapore (2017)
20. Blei, D.M.: Probabilistic topic models. *Commun. ACM.* 55, 77–84 (2012). <https://doi.org/10.1145/2133806.2133826>
21. Cahyadi, A., Khodra, M.L.: Aspect-Based Sentiment Analysis Using Convolutional Neural Network and Bidirectional Long Short-Term Memory. In: 2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA). pp. 124–129. IEEE, Krabi (2018)
22. Chamid, A.A., Widowati, Kusumaningrum, R.: Graph-Based Semi-Supervised Deep Learning for Indonesian Aspect-Based Sentiment Analysis. *BDCC.* 7, 5 (2022). <https://doi.org/10.3390/bdcc7010005>
23. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention Is All You Need, <https://arxiv.org/abs/1706.03762>, (2017)
24. Gao, T., Fang, J., Liu, H., Liu, Z., Liu, C., Liu, P.: LEGO-ABSA: A Prompt-based Task Assemblable Unified Generative Framework for Multi-task Aspect-based Sentiment Analysis.
25. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, <http://arxiv.org/abs/1810.04805>, (2019)
26. Hou, H., Shen, L., Jia, J., Xu, Z.: An integrated framework for flood disaster information extraction and analysis leveraging social media data: A case study of the Shouguang flood in China. *Science of The Total Environment.* 949, 174948 (2024). <https://doi.org/10.1016/j.scitotenv.2024.174948>
27. Yang, P., Yao, Y., Zhou, H.: Leveraging Global and Local Topic Popularities for LDA-Based Document Clustering. *IEEE Access.* 8, 24734–24745 (2020). <https://doi.org/10.1109/ACCESS.2020.2969525>

28. Chang, P., Yu, Y.-T., Sanders, A., Munasinghe, T.: Perceiving the Ukraine-Russia Conflict: Topic Modeling and Clustering on Twitter Data. In: 2023 IEEE Ninth International Conference on Big Data Computing Service and Applications (Big-DataService). pp. 147–148. IEEE, Athens, Greece (2023)
29. Parfenova, A.: Automating the Information Extraction from Semi-Structured Interview Transcripts. In: Companion Proceedings of the ACM Web Conference 2024. pp. 983–986. ACM, Singapore Singapore (2024)
30. Grootendorst, M.: BERTopic: Neural topic modeling with a class-based TF-IDF procedure, <http://arxiv.org/abs/2203.05794>, (2022)
31. Mahfudiyah, N., Alamsyah, A.: Understanding User Perception of Ride-Hailing Services Sentiment Analysis and Topic Modelling using IndoBERT and BERTopic. In: 2023 International Conference on Digital Business and Technology Management (ICONDBTM). pp. 1–6. IEEE, Lombok, Nusa Tenggara Barat, Indonesia, Indonesia (2023)
32. Asnawi, M.F., Hanafi, M., Kurniawan, N.F., Suwondo, A., Nasrullah, A., Setyawan, C.: Topic Modelling Analysis on Indonesian News Using BERT Topic Model. In: 2024 6th International Conference on Cybernetics and Intelligent System (ICORIS). pp. 1–6. IEEE, Surakarta, Indonesia (2024)
33. Ashton, T., Evangelopoulos, N., Prybutok, V.R.: Exponentially weighted moving average control charts for monitoring customer service quality comments. *IJSS*. 8, 230 (2013). <https://doi.org/10.1504/IJSS.2013.057237>
34. Ashton, T., Evangelopoulos, N., Prybutok, V.R.: Quantitative quality control from qualitative data: control charts with latent semantic analysis. *Qual Quant*. 49, 1081–1099 (2015). <https://doi.org/10.1007/s11135-014-0036-5>
35. Sun, C., Huang, L., Qiu, X.: Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence, <https://arxiv.org/abs/1903.09588>, (2019)
36. Varoquaux, G., Colliot, O.: Evaluating Machine Learning Models and Their Diagnostic Value. In: Colliot, O. (ed.) *Machine Learning for Brain Disorders*. pp. 601–630. Springer US, New York, NY (2023)
37. Miller, C., Portlock, T., Nyaga, D.M., O’Sullivan, J.M.: A review of model evaluation metrics for machine learning in genetics and genomics. *Front. Bioinform*. 4, 1457619 (2024). <https://doi.org/10.3389/fbinf.2024.1457619>
38. Allaoui, M., Kherfi, M.L., Cheriet, A.: Considerably Improving Clustering Algorithms Using UMAP Dimensionality Reduction Technique: A Comparative Study. In: El Moataz, A., Mammass, D., Mansouri, A., and Nouboud, F. (eds) *Image and Signal Processing*. pp. 317–325. Springer International Publishing, Cham (2020)
39. Kim, Y.-J., Kim, H.-S.: The Impact of Hotel Customer Experience on Customer Satisfaction through Online Reviews. *Sustainability*. 14, 848 (2022). <https://doi.org/10.3390/su14020848>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

