Service Quality Analysis Based On Online Customer Review In Google Play Store (Study Case Of Telkomsel)

Wahyuning Hanifah Aprillia[1*], Maya Ariyanti[2], Sri Widiyanesti[3]

[1] Student, School of Economics and Business, Telkom University, Bandung, Indonesia
[2,3] Lecturer, School of Economics and Business, Telkom University, Bandung, Indonesia
[*]Corresponding Author:
Email: hanifah.aprillia@gmail.com

Abstract.

A corporation that offers internet services is known as an Internet Service Provider (ISP). Regional-scale networks and worldwide networks are offered at ISPs, allowing consumers to effortlessly connect with the outside world globally. Telkomsel is one of providers that is used the most widely in Indonesia. However, Telkomsel is also a provider with the most complaints than others. This study chooses Telkomsel as a case study to determine their quality of service based on customer review. This paper aims to analyze and identify the service quality of Telkomsel and topics that were discussed by Telkomsel users based on customer reviews in Google Play Store. We categorized the data according to the following service quality dimensions: network quality, customer service and technical support, information quality, security and privacy, and fulfillment. As a result, the Naive Bayes Classifier (NBC) was applied to assist in the sentiment analysis process. The accuracy for sentiment analysis using NBC was more than 75%. This study used Latent Dirichlet Allocation (LDA) models for topic modeling to identify themes that are often discussed by consumers. Hence, the result of this study can help a company to improve and develop their quality of product and service according to customer needs.

Keywords: *Service quality, big data, sentiment analysis and topic modeling, and text analytic.*

I. INTRODUCTION

Following the growth of economics in Indonesia, several industries also compete in improve the advancement of information technology to optimize the operational efficiency of the company. It requires internet support, that makes information will more easily and more quickly to access [1]. The internet is a computer network on a world scale that can interconnect to build communication between users. Currently, the internet is growing very rapidly and becoming the largest information network in the world. The rapid development of internet users makes new battles in the business world, especially in the field of internet service providers with a large population. This is due to the internet that can be easily accessed by everyone, and increasingly widespread internet package services offered by providers [1].According to a report by the Indonesian Internet Service Providers Association (APJII), in 2023 the internet penetration rate in Indonesia is increased to 78.19 percent, around 215.62 million people. This result is increased by 5 million people from the survey results in 2022, which only reached 210 million people [2]. Identified that needs and demands continue to increase every year, internet services are considered a potential business and able to generate profitable margins. Therefore, many telecommunications services companies try to get into the internet service business. An Internet Service Provider (ISP) is a corporate entity that offers internet connectivity services. ISPs provide consumers with access to a network that includes both regional-scale and multinational networks, enabling seamless global connectivity [1].

Each provider comes with a variety of characteristics and qualities, and each consumer determines their choices that suit their needs [3].APJII stated that the most widely used cellular operator by Indonesians is Telkomsel. It is known that as many 40.27 % of Indonesians use Telkomsel, which also includes Simpati, Halo, and by.U cards. In the survey, APJII explained that the reason of people to use this cellular operator is because the signal considered to be the strongest in their location, in addition it is also because the price of internet packages and attractive promo packages [4]. Thus, at the end of 2022, Telkomsel's internet service was reported to be down. According to the data on the Downdetector website, it shows that reports of disruptions to Telkomsel services reached 1.453 reports. And the highest report is on internet problems as much as 85 percent [5]. The Indonesian Consumer Foundation also recorded and explained that Telkomsel is

the provider company with the most complaints than the other provider [6]. In the previous study, many researchers mentioned that scale to measure quality service is based on dimensions, such as network quality, customer service and technical support, information quality, security and privacy, and fulfilment [7], [8]. Mostly, paper that related to service quality use conventional methods such as interview or questionnaire as data collection tool, but possible for the respondents to not be honest with their answer and restricted their opinions because it is restricted by the points given by the researcher [9], [10].

We use sentiment analysis as an alternative method to assess the degree of service quality by using user-generated content sourced from online review platforms because Hasian et.al. [11] explained that this approach is more efficient in terms of time and cost.The objective of this paper are to examine user-generated content, specifically online customer reviews, by analyzing both quantitative features and textual content [12]. Sentiment analysis is a method that can be used to classify unlabeled reviews data into positive and negative sentiments. The sentiment analysis approach showed higher accuracy in taking and measuring service quality based on reviews compared to other text mining methods. For this reason, we choose sentiment analysis as a method to identify and classify sentiment based on customer review. Furthermore, we use topic modelling to identify topics discussed by users based on reviews [12].In this paper, we use Telkomsel as a study case to determine its quality of service based on customer review. As an internet service provider company, important for the company's management to design and improve for a better service system to see the expectations and achieve customer satisfaction by analyze the customer opinions and customer reviews [13]. This paper aims to analyze and identify the service quality of Telkomsel and topics that discussed by Telkomsel users based on customer reviews in Google Play Store. The result of this study can help a company to improve and develop their quality of product and service according to customer needs.

II. METHODS

In this study, the population is taken from Telkomsel users who create reviews on the Google Play Store. Meanwhile, the sample used in this study is taken by using non-probability sampling technique. And the sample is taken from data review of Telkomsel users that is uploaded in the period June 01, 2023 to August 31, 2023. As for the research design is shown in Fig 1 below.

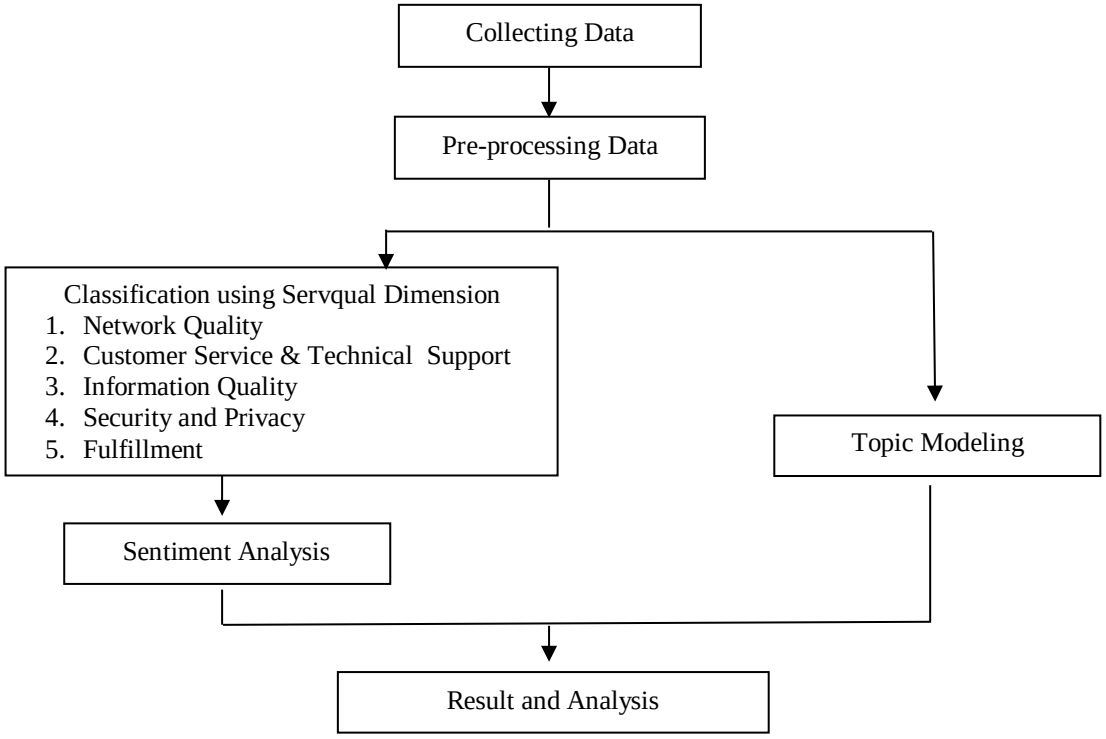

Fig 1. Research Method

A. Collecting Data

Data used in this research is secondary data. We managed to collect data from Google Play Store for Telkomsel's user reviews. Data is collected with timestamps in three months period from June 01, 2023 to August 31, 2023 with total data 50.000 user reviews. Data gained by crawling technique using Google Collaboratory with Python language and saved in CSV format.

B. Pre-processing Data

Pre-processing data involves the organization and filtration of unnecessary data from the original dataset. This technique aims to prevent the occurrence of redundant, contradictory, and flawed data. The sequential steps involved in pre-processing data are as follows:

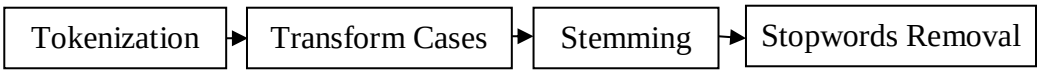

Fig 2. Pre-processing Data

a) Tokenization: the process of separating a sentence into a separate word.
b) Transform cases: the process of turning the sentences or words into small letters.
c) Stopwords removal: the process of removing words that don't have meaning.
d) Stemming: the process of converting words into native languages in Indonesian.

Based on the explanation of the pre-processing data steps, Table 1 below shows the examples of pre-processing data.

Table 1. Pre-processing Data

Step	Review	Result
Tokenization	Duh ampun telkomsel ANJRIT gaada angin gada ujan sinyalnya ilang terus	'Duh' 'ampun' 'telkomsel' 'ANJRIT' 'gaada' 'angin' 'gada' 'ujan' 'sinyalnya' 'ilang' 'terus'
Transform Cases	Duh ampun telkomsel ANJRIT gaada angin gada ujan sinyalnya ilang terus	'duh' 'ampun' 'telkomsel' 'anjrit' 'gaada' 'angin' 'gada' 'ujan' 'sinyalnya' 'ilang' 'terus'
Stopwords Removal	duh ampun telkomsel anjrit gaada angin gada ujan sinyalnya ilang terus	ampun telkomsel gaada angin gada ujan sinyalnya ilang terus
Stemming	ampun telkomsel gaada angin gada ujan sinyalnya ilang terus	ampun telkomsel tidak ada angin tidak ada hujan sinyal hilang terus

C. Classification using Servqual Dimension

The data that has completed the pre-processing step will be categorized based on the dimension of service quality. Machine learning is necessary to categorize this type of data. Machine learning involves the stages of data training and data testing. Prior to training and testing the data, it is necessary to label the data, which involves assigning categories to the data depending on its properties. Table 2 below shows a method of labelling based on servqual dimensions.

Table 2. Labelling Data using Servqual Dimension

Customer Review	Dimension
Signal selalu buruk	Network Quality
Memberikan layanan yang cepat dan memuaskan pelanggan	Customer Service & Technical Support
Paket 4G mingguan tidak ada lagi, tidak konsen buat kuota murahnya.	Information Quality
Bagaimana si nomor pribadi bisa bocor kemana mana	Security and Privacy
Kok semakin lama semakin mahal harga kuotanya	Fulfillment

D. Classification using Sentiment Anaysis

In this research, the model used for classification process is Naïve Bayes Classifier (NBC). There are three classes of sentiment in a dataset, including positive, neutral, and negative. After pre-processing, the data is labeled and subsequently split into training data and testing data based on customer sentiment towards the service quality of Telkomsel. Table 3 shows instances of labeling for training data with sentiment analysis.

Table 3. Labelling Data using Sentiment Analysis

Customer Review	Sentiment
Pilihan yang cocok karena jaringan lancar dan banyak promo	Positive
Ok pelayanan sudah meningkat. saya harap dapat diditingkatkan lagi	Neutral
Layanannya jelek banget	Negative

E. Topic Modeling

The company needs a comprehensive understanding of the frequently discussed issues by consumers in order to identify opportunities for improvement. For example, entities in a topic group might be important evidence to consider. In this research, we operate Latent Dirichlet Allocation (LDA) for topic modeling based on sentiment. To explore the relationships between terms and topics in a flexible manner, we utilize LDAvis, a web-based interactive visualization. The cumulative term frequency is denoted by shade light blue, whereas the term frequency predicted for the selected subject is depicted in red. The topic was determined with the most extensive-term distribution, the larger the distribution means, and the more often the word appears.

F. Result and Analysis

The result of the sentiment analysis model performance is then evaluated using confusion matrix. Confusion matrix is a method usually used to calculate the accuracy of data mining concepts. The confusion matrix classification is shown in Table 4 below.

Table 4. Confusion Matrix

Predicted Class	**Actual Class**			
		Positive	**Neutral**	**Negative**
	Positive	T-Pos	F-PosNet	F-PosNeg
	Neutral	F-NetPos	T-Net	F-NetNeg
	Negative	F-NegPos	F-NegNet	T-Neg

Based on the values of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), the results of sentiment analysis performance evaluation in text classification can be seen from the accuracy, precision, and recall values [14]. Accuracy measures correctness of classification models in predicting labels and accurately the machine can correctly classify the data. The accuracy value of a model can be considered good if the value > 70% and vice versa. The formula for measuring accuracy is as follows:

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

Precision shows the ratio of correct predictions and total positive and negative predictions. The precision value is used to measure the accuracy of the recommendations produced by the system. The formula for calculating precision is as follows:

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

However, recall indicates the model's ability to correctly identify members of positive or negative classes within a collection of unlabeled data. The formula for measuring recall is as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

Precision refers to the degree of accuracy in determining the extent to which the classifier's predictions align with the actual data. Recall measures the classifier's ability to accurately predict data, considering both its completeness and sensitivity [7].

III. RESULT AND DISCUSSION

From 21.588 data reviews, we classify the data based on service quality dimensions, including network quality, customer service and technical support, information quality, security and privacy, and also fulfillment. The classification data based on service quality dimensions is shown in Table 5 below.

Table 5. Classification based on Servqual Dimension

Dimension	**Percentage**
Network Quality	64.5%
Customer Service & Technical Support	9.4%
Information Quality	7.5%
Security and Privacy	0.7%
Fulfillment	17.9%

According to Table above, the result of classification based on service quality dimensions shows that the most problem that talks by user is about network quality. Because according to the result, 64.5% users talked about network quality of Telkomsel provider, 0.7% user reviewed the customer service and technical support. 7.5% user commented about information quality, 7.5% user talked about security and privacy, and 17.9% user talked about fulfillment of Telkomsel.

Meanwhile, the classification data using sentiment analysis is shown in Table 6 below.

Table 6. Classification based on Sentiment Analysis

Sentiment	Percentage
Positive	20.5%
Neutral	20.8%
Negative	58.7%

Table 6 shows the results of the sentiment analysis on Telkomsel, out of 58.7 percent of customers give a negative sentiment, 20.8 percent give a neutral sentiment, and for the remaining 20.5 percent of customers respond. And based on the confusion matrix values, the results of machine learning performance evaluation in classification can be seen from accuracy, precision, and recall. The values describe how accurate the machine can classify data correctly. The result of the text classification performance model shown in Table 7 below.

Table 7. Result of Classification Performance Model Accuracy = 75.6%

	True Positive	True Neutral	True Negative	Class Precision
Pred. Positive	12033	65	579	73.8%
Pred. Neutral	615	3392	412	84.4%
Pred. Negative	3023	558	911	47.8%
Class Recall	94.4%	20.2%	76.7%	

To comprehend the developing concepts, it is crucial to observe that the terms that have been offered are interconnected. The topics must be customized to the desired customer experience dimension to ease the analysis. For the topic modeling result, it shown on Fig 3.

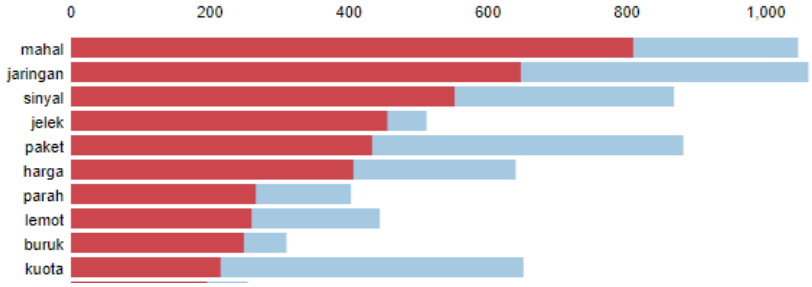

Fig 3. Topic 1 of Telkomsel

In Fig 3, the topic frequently discussed by the Telkomsel user, concerns the price and quality of provider. These are shown in words mahal, jaringan, sinyal, jelek, paket, and harga. This topic relates to network quality and fulfillment dimension.

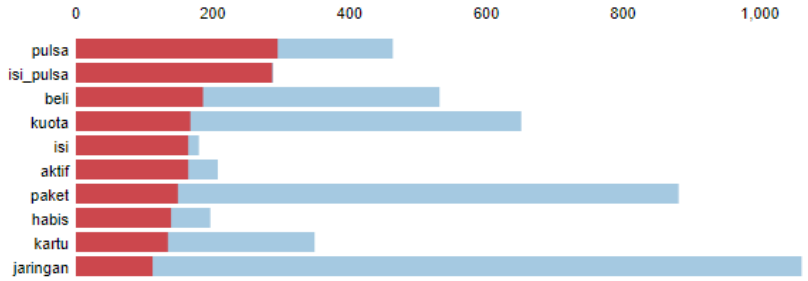

Fig 4. Topic 2 of Telkomsel

In Fig 4, the topic frequently discussed by the Telkomsel user, concerns the service and product of provider. These are shown in the words pulsa, isi pulsa, beli, kuota, isi, hilang, and aktif. This topic relates to information quality and also security and privacy dimension.

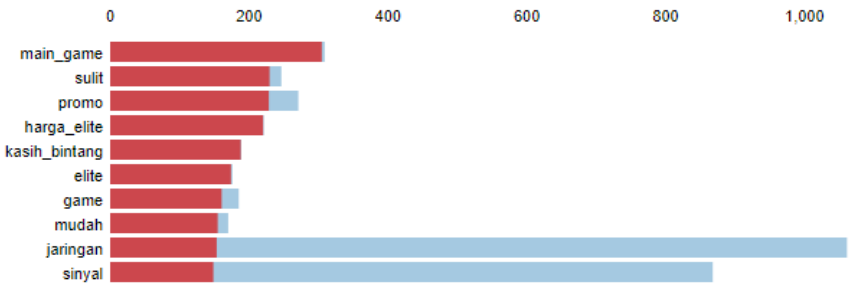

Fig 5. Topic 3 of Telkomsel

In Fig 5, the topic frequently discussed by the Telkomsel user concerns the network, price, and product of provider. These are shown in the words main game, sulit, promo, harga elit, kasih bintang, tolong tingkatkan, and elite. This topic relates to network quality, fulfillment, customer service and technical support, and information quality dimension.

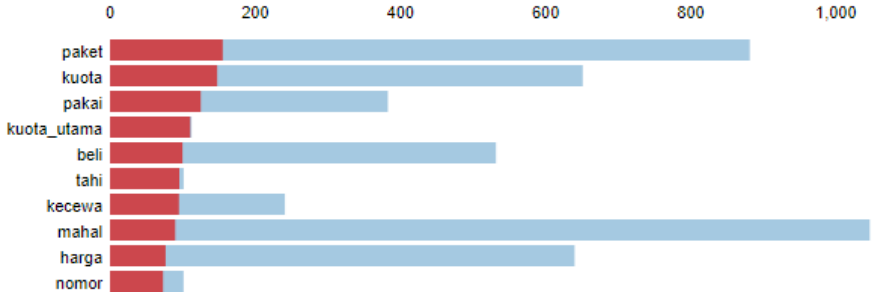

Fig 6. Topic 4 of Telkomsel

In Fig 6, the topic frequently discussed by the Telkomsel user, concerns the product and expression of customers. These are shown in the words paket, kuota, pakai, kecewa, tahi, and masuk akal. This topic relates to fulfillment and information quality dimension.

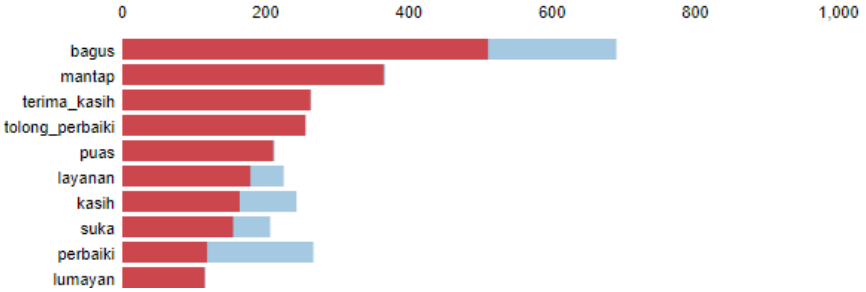

Fig 7. Topic 5 of Telkomsel

In Fig 7, the topic frequently discussed by the Telkomsel user concerns consumer gratitude for support in solving. These are shown in the words bagus, mantap, terima kasih, tolong perbaiki, and puas. This topic relates to customer service and technical support dimension.

IV. CONCLUSION

By analyzing the customer review, it can help the company or business organization to get a chance to develop their quality of products and services, create new innovations, and increase sales and demand of products. This research presents the combination of sentiment analysis and topic modeling for analyze the customer review of Telkomsel and classify the data based on service quality dimensions.Based on service quality dimensions, the classification result shows that the customer review mostly response about network quality of Telkomsel provider. Moreover, based on the sentiment analysis, the result shows that Telkomsel needs to improve their quality of service, customer service, and their products.

Meanwhile, topics that often discussed by user or customer is about the variable of network quality, fulfillment, and customer service and technical support.Another variable that is relevant and influences customer reviews should be included for future research. It is recommended to increase the quantity of data in order to achieve accurate categorization results. To gain a comprehensive perspective, it is recommended

to include additional subjects of investigation from different companies and industries. Integrate additional means of data extraction, such as from websites and several other social media platforms, in addition to Twitter. The subject modeling is contingent of each facet of the client experience, therefore yielding more comprehensive findings and analysis.

REFERENCES

[1] S. Nurajizah, N. A. Ambarwati, and S. Muryani, "Sistem Pendukung Keputusan Pemilihan Internet Service Provider Terbaik Dengan Metode Analytical Hierarchy Process**" *JURTEKSI (Jurnal Teknologi dan Sistem Informasi),*** vol. 6, no. 3, pp. 231–238, Aug. 2020, doi: 10.33330/jurteksi.v6i3.632.

[2] Lavinda, "APJII: Pengguna Internet Indonesia 215 Juta Jiwa pada 2023, Naik 1,17%," katadata.co.id. Accessed: May 21, 2023. [Online]. Available: https://katadata.co.id/lavinda/digital/646342df38af1/apjii-pengguna-internet-indonesia-215-juta-jiwa-pada-2023-naik-1-17

[3] D. Kurniawan, W. Wardhana, and N. A. Ito, "Penggabungan Dua ISP Guna Menstabilkan Koneksi Internet Dengan Metode Failover," ***Jurnal Komputasi***, vol. 4, no. 2, 2016.

[4] A. Putri, "Operator Seluler Paling Banyak Digunakan Masyarakat Indonesia," data.goodstats.id. Accessed: May 22, 2023. [Online]. Available: https://data.goodstats.id/statistic/embed/operator-seluler-paling-banyak-digunakan-masyarakat-indonesia-Ahoxy

[5] N. P. Bestari, "Telkomsel dan Indihome Gangguan, Netizen Ramai di Twitter," www.cnbcindonesia.com. Accessed: May 16, 2023. [Online]. Available: https://www.cnbcindonesia.com/tech/20221221171821-37-398958/telkomsel-dan-indihome-gangguan-netizen-ramai-di-twitter

[6] R. Pahlevi, "YLKI: Indihome dan Telkomsel Paling Banyak Dikeluhkan Konsumen," databoks.katadata.co.id. Accessed: Jul. 07, 2023. [Online]. Available: https://databoks.katadata.co.id/datapublish/2022/01/10/ylki-indihome-dan-telkomsel-paling-banyak-dikeluhkan-konsumen

[7] Alamsyah and E. A. Bernatapi, "Evolving Customer Experience Management in Internet Service Provider Company using Text Analytics," in ***International Conference on ICT for Smart Society (ICISS)***, 2019. doi: https://doi.org/10.1109/ICISS48059.2019.8969828.

[8] N. Nurul, S. Putri, A. Alamsyah, and S. Widiyanesti, "Fulfillment and Responsiveness on Online Travel Agencies Using Multiclass Classification," in ***2020 8th International Conference on Information and Communication Technology (ICoICT),*** IEEE Xplore, Jun. 2020, pp. 1–6.

[9] P. K. Sari, A. Alamsyah, and S. Wibowo, "Measuring e-Commerce service quality from online customer review using sentiment analysis," in ***Journal of Physics: Conference Series***, Institute of Physics Publishing, Apr. 2018. doi: 10.1088/1742-6596/971/1/012053.

[10] M. Siva, D. P. Nayak, and K. A. Narayan, "Strengths and Weaknesses of Online Surveys**," *IOSR Journal of Humanities and Social Sciences (IOSR-JHSS***, vol. 24, no. 5, pp. 32–38, May 2019.

[11] F. P. Hasian, R. W. Damayanti, and E. Pujiyanto, "Literature Review of the Internet Provider Service Quality based on Sentiment Analysis," in ***Proceedings of the 3rd Asia Pacific International Conference on Industrial Engineering and Operations Management,*** Malaysia: IEOM Society International, Sep. 2022, pp. 283–290.

[12] W. Duan, Q. Cao, Y. Yu, and S. Levy, "Mining online user-generated content: Using sentiment analysis technique to study hotel service quality," in ***Proceedings of the Annual Hawaii International Conference on System Sciences,*** 2013, pp. 3119–3128. doi: 10.1109/HICSS.2013.400.

[13] Dhini and D. A. Kusumaningrum, "Sentiment Analysis of Airport Customer Reviews," in *Proceedings of the 2018 IEEE IEEM*, 2018, pp. 502–506.

[14] R. A. Masrury, Fannisa, and A. Alamsyah, "Analyzing Tourism Mobile Applications Perceived Quality using Sentiment Analysis and Topic Modeling," in ***7th International Conference on Information and Communication Technology (ICoICT)***, 2019. doi: 10.1109/icoict.2019.8835255.