

Original Article

Analysis of ChatGPT Application Service Quality Using Text Classification and Topic Modeling

¹Muhammad Irfan, ²Mochamad Yudha Febrianta, ³Dian Puteri Ramadhani

^{1,2,3}Faculty of Economics and Business, Telkom University, Bandung, Indonesia.

Received Date: 29 July 2024

Revised Date: 19 August 2024

Accepted Date: 23 August 2024

Published Date: 29 August 2024

Abstract: This research aims to understand the sentiments and main topics of users related to the service quality of the ChatGPT Android application by analyzing User-Generated Content (UGC) on the Google Play Store based on user experiences. The research method used in this study is sentiment analysis with the RoBERTa algorithm and topic modeling using LDA. After comparing the Naive Bayes, SVM, and RoBERTa algorithms, RoBERTa was found to have the highest accuracy, with 72% for dimension classification and 94% for sentiment classification. Positive sentiment was 86.17%, and negative sentiment was 13.83%. The dimensions used to measure service quality are Content Quality, Engagement, Reliability, Usability, and Privacy. The results of the study show that the service quality of the ChatGPT Android application is mostly positive in the dimensions of Engagement, Reliability, and Content Quality. However, the dimensions of Privacy and Usability have negative views. Positive sentiment findings include satisfaction with the information provided, the dark mode feature, the user-friendly interface, user engagement, and application reliability. On the other hand, negative sentiment highlights issues with privacy and usability. This knowledge can be used to improve service quality, especially in the dimensions of Privacy and Usability.

Keywords: ChatGPT, Sentiment Analysis, Topic Modeling, Service Quality.

I. INTRODUCTION

Artificial Intelligence (AI) is a development in information and communication technology that has emerged in the last decade (Rukmayuninda Ririh et al., 2020). One significant advancement in this field is the development of sophisticated language models, such as ChatGPT, which OpenAI developed. Many companies like Twitter, Amazon, Facebook, TikTok, OpenAI, and several others have invested large amounts of money into Artificial Intelligence (AI). In November 2022, OpenAI released ChatGPT for the first time as a web application. ChatGPT is currently the fastest-growing application ever developed in the world (Okey et al., 2023). Unlike other innovative applications such as TikTok and other chatbots, within 2 months of its launch, ChatGPT gained 100 million active users, surpassing its competitors at the time of this research (Okey et al., 2023). After the ChatGPT website was launched, OpenAI subsequently released the ChatGPT Android application, which became available on the Google Playstore on July 25, 2023.

The presence of ChatGPT on the Playstore was announced by OpenAI, the company behind ChatGPT, through their X account with the handle @OpenAI. The ChatGPT Android application is popular in Indonesia, as evidenced by its 10 million downloads and over 300,000 user reviews on the Google Playstore. According to Similarweb, the ChatGPT Android application ranks first in the Productivity category, indicating that it is one of the most popular applications on Google Playstore. Research by (Okey et al., 2023) highlights potential weaknesses of ChatGPT in the context of cybersecurity. Concerns involve several serious risks, including the ability to create malware code, engage in hacking, conduct unauthorized intelligence gathering, and launch phishing attacks. Given these vulnerabilities, it is revealed that ChatGPT has the potential to be exploited as a hacking tool that could jeopardize overall system security. From June 2022 to May 2023, ChatGPT experienced a data breach that resulted in users' personal data being leaked to the dark web. The amount of data reached 101,134 active user records from the ChatGPT application (Data Boks, 2023).

Therefore, there is a need to research the service quality of the ChatGPT Android application. According to Rio Sasongko (2021), the definition of service quality is a customer's evaluation that compares their expectations with the actual experience they receive from the service provided by the service provider, both partially and overall. Big data can be utilized to assess customer satisfaction with service quality (Masrury et al., 2019). In the study by Kim Ki-Hun et al. (2019), there are 5 dimensions of service quality: Content Quality, Engagement, Reliability, Usability, and Privacy. The research data was collected from User Generated Content (UGC) on the Google Playstore, consisting of reviews from users of the ChatGPT Android application. UGC tends to be honest, straightforward, enthusiastic, and often insightful (Shidqi & Yudha Febrianta, 2023). The data was taken from the "most relevant" category, ensuring that the reviews contain richer and more detailed information about the user experience. A total of 20,000 reviews were collected and stored in CSV format. Before processing the data, researchers cleaned the collected

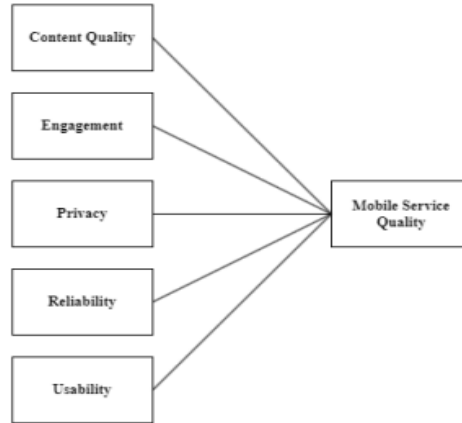


data through pre-processing. This process aims to facilitate data processing and analysis, as well as to avoid bias in the research results.

After the pre-processing stage, the researchers classified the data into 5 dimensions of mobile service quality. In the study by (Boukabous & Azizi, 2021). Natural Language Processing (NLP) refers to the processing of textual data in natural language form. For the classification process, this study compared three algorithms: Naive Bayes, RoBERTa, and Support Vector Machine. BERT requires manual pre-processing steps, such as tokenization, stopword removal, converting letters to lowercase, stemming, and lemmatization, to eliminate irrelevant elements and reduce text complexity (Alamsyah & Girawan, 2023). The next step is sentiment analysis, with data labeled into two dimensions: positive and negative. The final stage is conducting topic modeling on each positive and negative review using the Latent Dirichlet Allocation (LDA) model (Okey et al., 2023).

Evaluating mobile applications through reviews can serve as a guide in making business decisions by considering the results of user opinion analysis regarding the products or services they use. On the other hand, the combination of sentiment analysis and topic modeling can help us understand user feelings and identify topics being discussed on the platform, which can be used to design strategies for improving certain services or products (Garini et al., 2023). This sentiment analysis aims to evaluate the opinions expressed by users of the ChatGPT application, whether they are positive, negative, or neutral (Yang et al., 2020). In this context, sentiment analysis is utilized to assess the service quality of the ChatGPT application.

Figure 1: Research Model



II. LITERATURE REVIEW

A) Marketing Management

Kotler et al. (2022), marketing management is defined as the art and science of identifying target markets and developing profitable relationships with them. This requires more than just identifying target markets; it also involves acquiring, retaining, and expanding customers through the creation, delivery, and communication of superior customer value. Meanwhile, according to Ariyanto et al. (2023), marketing management is a field of management science that is crucial in all business activities. Marketing management is the most important factor for the survival of a company or business and the achievement of its desired goals. Effective marketing management can influence consumer decisions (Suryawardani et al., 2022)

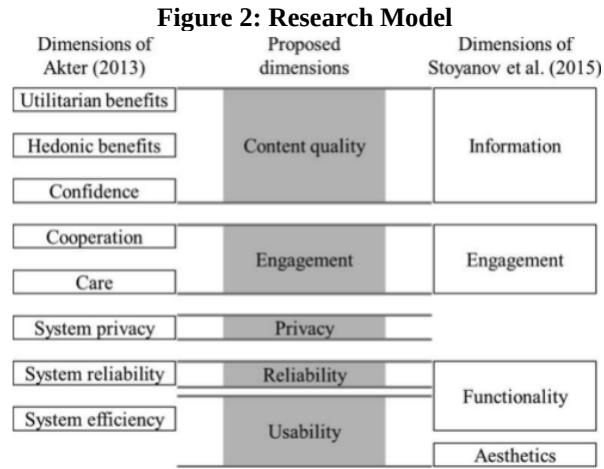
B) Service Quality

According to Zhou et al. (2021), service quality refers to the overall assessment of how well the service meets or exceeds customer expectations. If the level of service provided or received matches expectations, the service quality is considered good and satisfactory. According to Zeithaml et al. (2017), in the study by Gloria & Talavera (2020), the growth of electronic-based services has prompted the development of a service assessment dimension called ESERVQUAL. Using an approach focused on group stages and empirical data collection and analysis, there are seven critical dimensions for evaluating services. These dimensions include Efficiency, Fulfillment, System Availability, Privacy, Responsiveness, Compensation, and Contact. The quality of service provided by companies to their users must be carefully considered because good service quality will increase user satisfaction and lead to loyalty towards the application (Br. Ginting et al., 2022)

C) Mobile Service Quality

Service quality is defined as the overall impression users have of the relative inferiority/superiority of a service. According to Kim Ki-Hun et al. (2019), existing studies confirm that service quality must be multi-dimensional, hierarchical, and context-specific. In the literature, SERVQUAL and SERVPERF are used as foundational frameworks that need modification to reflect

the characteristics of the target service. Therefore, Akter et al. (2013), Stoyanov et al. (2015), and Kim Ki-Hun et al. (2019) propose five comprehensive dimensions of service quality: Content Quality, Engagement, Privacy, Reliability, and Usability.



III. RESULTS AND DISCUSSION

A) Data Classification Based on Dimensions

The results from the classification model of the five dimensions of mobile service quality using the RoBERTa algorithm with an 80:10:10 ratio achieved higher accuracy compared to the two models, Naive Bayes and Support Vector Machine. This can be seen in Table 1 below:

Table 1: Comparison of 3 Algorithm Models

Algorithm	Level of accuracy
Naive Bayes Classifier	65%
RoBERTa	72%
Support Vector Machine	62%

The results from classifying the dimensions of the dataset, totaling 18,852 entries, found that the Usability dimension is the most frequently discussed, with a proportion of 28.22%. Next, the Engagement dimension follows usability with a proportion of 27.61%, indicating the importance of features that provide active interaction to users, making them feel engaged with the application. The Reliability and Content Quality aspects also received significant attention, with proportions of 19.17% and 19.13%, respectively. This indicates that users pay attention to the reliability of the application and the quality of the content produced by ChatGPT. Meanwhile, the Privacy dimension only accounts for 5.56%, suggesting that, although privacy is important, users do not focus as much on data security.

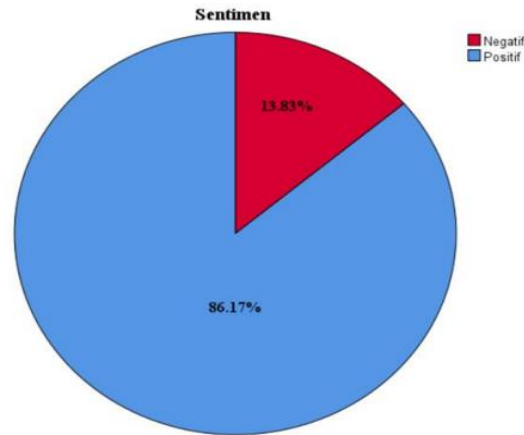
Table 2: Comparison of 3 Algorithm Models

Dimensions	Classification Results
Content Quality	19,3%
Engagement	27,91%
Reliability	19,17%
Usability	28,22%
Privacy	5,56%

Based on the classification results of mobile service quality dimensions, the dimension with the highest proportion is usability at 28.22%. In the study by Kim Ki-Hun et al. (2019), usability refers to the ease of use of the application, particularly in terms of user interface and how the application is used. With an intuitive and easy-to-use interface, users are likely to feel more comfortable and not experience difficulties interacting with the application. This not only enhances overall user satisfaction but can also increase the retention of ChatGPT users. On the other hand, the dimension with the least discussion is privacy, indicating that, although privacy is important, users are not as focused on data security.

B) Sentiment Analysis

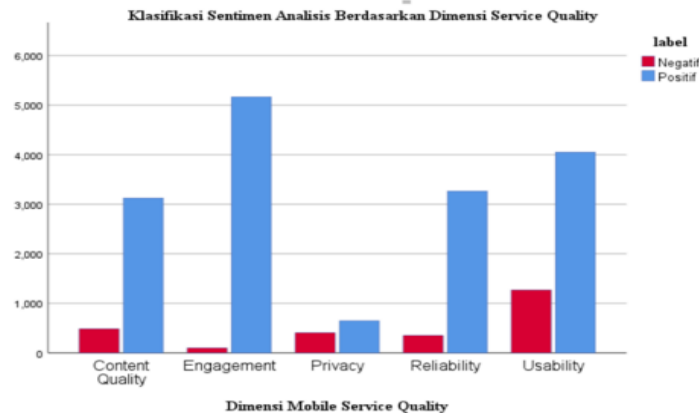
After classifying based on dimensions, the next step was sentiment analysis using the same model with an 80:10:10 ratio based on positive and negative labeling of the entire dataset. The accuracy achieved was 95%. Below is an image of the sentiment classification results:

Figure 3: Classification Sentiment Analysis

The results of the overall data show that user sentiment towards the ChatGPT Android application on the Google Playstore is predominantly positive. Positive sentiment dominates with 16,245 entries or 86% out of a total of 18,852 reviews. The remaining 14%, or 2,607 entries, reflect negative sentiment. The machine learning results show that the majority of reviews, 86% or 16,245 reviews, reflect positive sentiment, indicating that most users have had good experiences and that their satisfaction with the service quality of the ChatGPT application is met. Conversely, only 14% or 2,607 reviews reflect negative sentiment, indicating that while the majority of user satisfaction is met, there is still a small portion whose satisfaction has not been fulfilled. This suggests that there are still aspects that need improvement to enhance the service quality of the application and prevent a decline in user satisfaction with the ChatGPT application.

C) Proportion of Sentiments on Each Dimension of Service Quality

After performing sentiment analysis on the positive and negative sentiments of the entire dataset, the researchers conducted a detailed analysis to gain deeper insights into the service quality of each mobile service quality dimension. Below is a diagram showing the results of the sentiment analysis classification for each dimension of mobile service quality:

Figure 4: Sentiment Classification on Service Quality Dimensions

Based on the results of the analysis, the usability dimension is the most dominant, with the highest proportion of reviews at 28%, indicating that the ease of use of the application is a critical aspect for users. Although 76.1% of the reviews on usability are positive, 23.9% indicate dissatisfaction, signaling the need for further improvements to enhance the service quality performance of the ChatGPT application. The Engagement dimension received a very high positive sentiment of 98.1%, indicating that the application's interactive features are highly valued and effectively increase user engagement. The Reliability dimension also shows good performance, with 90.3% positive reviews and only 9.7% negative reviews, indicating that the ChatGPT application is reliable for various usage needs.

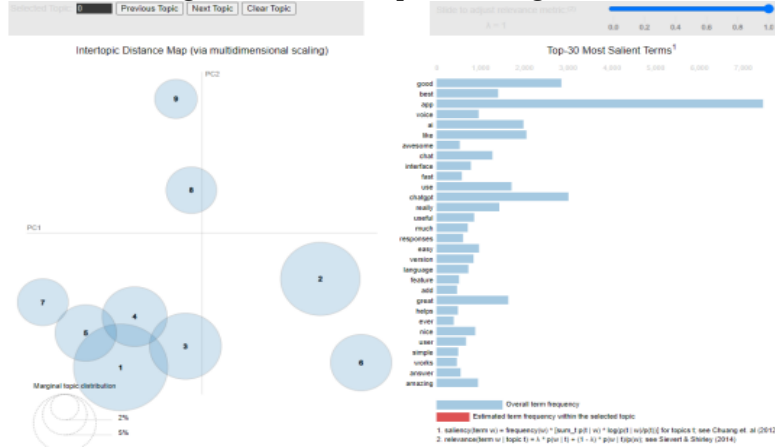
The Content Quality dimension also has a high positive review rate of 86.6%, with 13.4% negative reviews indicating that the quality of content provided by the ChatGPT application does not fully meet user needs. The Privacy dimension received the fewest reviews, with 61.4% positive reviews and 38.6% negative reviews, indicating that privacy issues remain important

and require special attention to increase user trust. Overall, this study implies that ChatGPT should continue to strengthen the dimensions that already receive positive feedback and improve the aspects that cause user dissatisfaction to enhance overall service quality and user satisfaction.

D) Topic Modelling

For topic modeling, the data was divided into two to achieve more comprehensive results. Based on the positive sentiment analysis, 16,245 data points were used in the topic modeling. The number of clusters was determined using the Silhouette Score, identifying 9 topics as the optimal number of clusters.

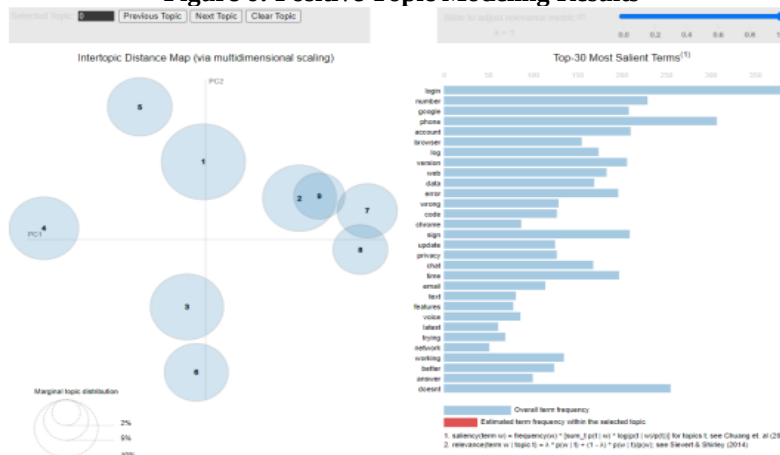
Figure 5: Positive Topic Modeling Results



Based on the topic modeling analysis of positive sentiment, we can identify the strengths of the ChatGPT application's service quality that can be used for future improvement. There are 9 topics on positive sentiment according to the Silhouette Score results. Topics 1 and 9 show a common theme that the responses provided by the application are satisfactory and very helpful to users, indicating that the content quality meets user needs. Topics 2 and 4 discuss the ease of using the ChatGPT application, specifically highlighting the dark mode feature and the clean user interface, making the application easy to use. Topics 3, 5, and 6 share a common discussion about user engagement with the ChatGPT application. For example, the features in the application are better than those in the website version. The voice feature is an important aspect that can enhance user engagement with the application and help users express emotional satisfaction with the application. Topic 7 discusses how the application helps users complete their tasks effectively, indicating that ChatGPT can be relied upon by users. Finally, Topic 8 mentions that users feel the application works very well and provides excellent privacy protection, reflecting that the application ensures user privacy.

Based on the negative sentiment results, 2,607 data points were used in the topic modeling. The number of clusters was determined using the Silhouette Score, identifying 9 topics as the optimal number of clusters.

Figure 6: Positive Topic Modeling Results



Based on the topic modeling analysis of negative sentiment, we can identify the weaknesses in the ChatGPT application's service quality that can be used for future evaluation. There are 9 optimal topics on negative sentiment according to the Silhouette Score results. Topics 1 and 7 concern users' worries about the personal data requested by the ChatGPT application, especially how personal numbers are used and how ChatGPT manages user activity history. Topics 2, 4, 5, 6, 8, and 9 share a common theme regarding issues such as bugs in the application's features, difficulties logging into the application, the inability to update the application, non-intuitive or difficult-to-use design, and application errors. These issues are related to the Usability dimension, with indicators such as visually appealing application design, ease of use, and quick access to services without technical obstacles.

Topic 3 discusses how the information or responses provided by the application are irrelevant to the users' questions. This falls under the Content Quality dimension, with indicators such as the usefulness and relevance of the information provided. However, since these discussions are in the negative sentiment category, the analysis reflects negative values for each indicator.

IV. CONCLUSION

A) Conclusion

Based on the research conducted on the service quality of the ChatGPT Android application from user reviews on Google Playstore using sentiment analysis and topic modeling, the following conclusions can be drawn:

1. The classification results of the mobile service quality dimensions related to the service quality of the ChatGPT application show that the most reviewed dimension is usability. This is reinforced by comparing classification algorithms, with RoBERTa producing the best accuracy at 72% compared to other algorithms such as Naive Bayes and Support Vector Machine.
2. Based on the sentiment analysis results on service quality, it is evident that the proportion of positive sentiment reaches 86%, while negative sentiment is only 14%, indicating that the service quality of the ChatGPT application is quite satisfactory and successfully meets user expectations.
3. Based on the proportion of positive and negative sentiment in each dimension of mobile service quality, the dimensions with the highest positive proportions are Engagement, Reliability, and Content Quality. Conversely, the dimensions with the highest negative proportions are Privacy and Usability.
4. The topic modeling analysis reveals positive and negative sentiment reviews on the service quality of the ChatGPT application, resulting in 9 optimal topics each. Positive sentiment reviews indicate that the ChatGPT application has various advantages that users appreciate, including content quality, ease of use, available features, and the application's ability to assist users with their tasks. On the other hand, negative sentiment reviews highlight users' concerns about personal data privacy, such as the use of phone numbers and activity history, bugs in application features, login difficulties, problematic application updates, non-intuitive design, and errors when prompts are run.

B) Recommendations

The results of this study are expected to serve as considerations for ChatGPT, particularly regarding the Usability and Privacy dimensions.

1. Based on these findings, these two dimensions should be the focus for improving service quality for users.
2. Specifically, transparency regarding the use of users' personal data and the ease of using the application are crucial.
3. Issues that need attention include login failures, application bugs, problematic updates, non-intuitive design, errors during prompt execution, and features that do not work, such as an unresponsive keyboard, which limits the application's usability.
4. It is recommended that more diverse data from various customer review sources be used to enhance the sentiment analysis results of the RoBERTa model.
5. In addition to Google Playstore, data from other platforms such as Instagram, Twitter, and other sources can also be utilized.

Future research could explore different objects and dimensions of service quality for the ChatGPT Android application using the RoBERTa model, providing deeper insights.

V. REFERENCES

- [1] Alamsyah, A., & Girawan, N. D. (2023). Improving Clothing Product Quality and Reducing Waste Based on Consumer Review Using RoBERTa and BERTopic Language Model. *Big Data and Cognitive Computing*, 7(4), 168. <https://doi.org/10.3390/bdcc7040168>
- [2] Ariyanto, A., Bangun, R., Rifqi, M., Indillah, M., Ferlina, A., Trenggana, M., Sholihah, R., Ariyanti, M., Widiati, E., Irawan, P., Ratih, S. D., Suryanti Ismail, R., Putra, S., Mulia Utama, A., Syahputra, J., & Budiman, B. (2023). MANAJEMEN PEMASARAN. www.freepik.com
- [3] Boukabous, M., & Azizi, M. (2021). A comparative study of deep learning-based language representation learning models. *Indonesian Journal of Electrical Engineering and Computer Science*, 22(2), 1032–1040. <https://doi.org/10.11591/ijeecs.v22.i2.pp1032-1040>
- [4] Br.Ginting, D. Y., Suyanto, A., & Mochamad, Y. F. (2022). Analysis Of The Effect Of E-Service Quality On E-Customer Satisfaction And E-Customer Loyalty On Shopee. *Asian Journal of Research in Business and Management*. <https://doi.org/10.55057/ajrbm.2022.4.3.35>
- [5] Data Boks. (2023, July). Data Pengguna ChatGPT Bocor di “Dark Web”, Indonesia Masuk Daftar Kebocoran 10 Besar. *DataBoks*.

- [6] Garini, J. G., Hidayanto, A. N., & Fina, A. (2023). Using machine learning to improve a telco self-service mobile application in Indonesia. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 12(4), 1947. <https://doi.org/10.11591/ijai.v12.i4.pp1947-1959>
- [7] Gloria, M., & Talavera, V. (2020). Measuring Service Quality in Philippine Banks: An Exploratory Study Using SERVQUAL and Q-Methodology. In *Philippine Management Review* (Vol. 27).
- [8] Kim Ki-Hun, Kim Kwang-Jae, Lee Dae-Ho, & Kim Min-Geun. (2019). Identification of critical quality dimensions for continuance intention in. *International Journal of Information Management*.
- [9] Kotler, P., Keller, K. L., & Chernev Alexander. (2022). *Marketing Management* (16 Edition) (16th ed.). Pearson Education.
- [10] Masrury, R. A., Fannisa, & Alamsyah, A. (2019, July 1). Analyzing tourism mobile applications perceived quality using sentiment analysis and topic modeling. 2019 7th International Conference on Information and Communication Technology, ICoICT 2019. <https://doi.org/10.1109/ICoICT.2019.8835255>
- [11] Okey, O. D., Udo, E. U., Rosa, R. L., Rodríguez, D. Z., & Kleinschmidt, J. H. (2023). Investigating ChatGPT and cybersecurity: A perspective on topic modeling and sentiment analysis. *Computers and Security*, 135. <https://doi.org/10.1016/j.cose.2023.103476>
- [12] Rio Sasongko, S. (2021). FAKTOR-FAKTOR KEPUASAN PELANGGAN DAN LOYALITAS PELANGGAN (LITERATURE REVIEW MANAJEMEN PEMASARAN). 3(1). <https://doi.org/10.31933/jimt.v3i1>
- [13] Rukmayuninda Ririh, K., Laili, N., Wicaksono, A., & Tsurayya, S. (2020). STUDI KOMPARASI DAN ANALISIS SWOT PADA IMPLEMENTASI KECERDASAN BUATAN (ARTIFICIAL INTELLIGENCE) DI INDONESIA. In *Jurnal Teknik Industri* (Vol. 15, Issue 2).
- [14] Shidqi, F., & Yudha Febrianta, M. (2023). Analisis Kualitas Layanan Internet Service Provider Menggunakan Metode Analisis Sentimen Dan Topic Modelling. *SEIKO : Journal of Management & Business*, 6(2), 439–450.
- [15] Suryawardani, B., Wulandari, A., Satriya, G. B., Marcelino, D., Prawita, F. N., Febrianta, M. Y., & Tantra, T. (2022). Digital Tools for Co-working Space Batununggal Using Digital-Ads, AR-Based Socmed Content, and Enterprise Resource Planning. *Jurnal Penyuluhan*, 19(01), 12–21. <https://doi.org/10.25015/19202341777>
- [16] Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning. *IEEE Access*, 8, 23522–23530. <https://doi.org/10.1109/ACCESS.2020.2969854>
- [17] Zeithaml, V. A., Bitner, M. J., & Gremler, D. D. (2017). *Services marketing : integrating customer focus across the firm*.
- [18] Zhou, Q., Lim, F. J., Yu, H., Xu, G., Ren, X., Liu, D., Wang, X., Mai, X., & Xu, H. (2021). A study on factors affecting service quality and loyalty intention in mobile banking. *Journal of Retailing and Consumer Services*, 60. <https://doi.org/10.1016/j.jretconser.2020.102424>