



Feedback analysis of service quality through data mining approach

Yerik Afrianto Singgalen

¹Tourism Department, Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, Indonesia

ARTICLE INFO

ABSTRACT

Article history:

Received Nov 14, 2024

Revised Nov 22, 2024

Accepted Nov 30, 2024

Keywords:

Data Mining;
Feedback;
Hotel;
KDD;
Service Quality.

This study explores the application of data mining techniques to analyze customer feedback for improving service quality at Tanjung Lesung Beach Hotel. Utilizing the Knowledge Discovery in Databases (KDD) framework, the research systematically collected, cleaned, and analyzed 1,239 customer reviews from the Agoda platform. Through a thorough data cleaning process, 642 verified reviews were identified as authentic, providing a robust dataset for in-depth analysis. Sentiment analysis was employed to extract key insights, revealing both positive aspects of the hotel experience, such as staff friendliness and beach amenities, as well as areas requiring improvement, particularly related to room conditions and breakfast offerings. Trend analysis indicated an upward trend in customer satisfaction from 2018 to 2023, with a minor decline observed in 2024, suggesting the need for ongoing service enhancements. The findings emphasize the critical role of leveraging verified customer feedback to inform strategic improvements, ultimately aiming to enhance guest satisfaction and maintain a competitive edge in the hospitality industry. This research contributes to the field by demonstrating how data-driven approaches can optimize service quality and align organizational strategies with evolving customer expectations.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



Corresponding Author:

Yerik Afrianto Singgalen

Tourism Department, Faculty of Business Administration and Communication,

Atma Jaya Catholic University of Indonesia

Jl. Jenderal Sudirman No. 51, Jakarta, 10220, Indonesia.

Email: yerik.afrianto@atmajaya.ac.id

1. INTRODUCTION

Feedback analysis has emerged as a crucial element in evaluating service quality, mainly through the application of data mining approaches. This method leverages vast amounts of unstructured data collected from customer interactions, enabling organizations to extract meaningful patterns and insights often obscured in traditional analyses (Aakash & Gupta Aggarwal, 2022). The approach facilitates a more nuanced understanding of customer sentiments, utilizing sophisticated algorithms and data classification techniques essential for enhancing service delivery (Ho et al., 2020). Practical feedback analysis identifies specific areas for improvement and uncovers latent issues that may not be immediately evident through conventional methods. The ability to transform raw data into actionable insights signifies a shift toward a more proactive strategy in

managing service quality (Thu, 2020). Despite its advantages, this approach necessitates a robust infrastructure for handling large datasets and requires expertise in data mining to ensure accurate interpretations. When properly executed, it is an invaluable tool for organizations seeking to align their services with evolving customer expectations, fostering customer satisfaction and long-term loyalty.

This research aims to establish a data-driven framework capable of extracting and analyzing customer feedback to assess service quality effectively. By systematically applying data mining techniques, this study transforms unstructured textual data into valuable insights that inform strategic decision-making. Such an approach is intended to move beyond traditional qualitative assessments, enabling the discovery of deeper patterns and customer sentiment that may otherwise remain hidden. This level of granular analysis is particularly significant for organizations striving to align their services with ever-changing consumer expectations (Ho et al., 2020; Rita et al., 2022). Moreover, the integration of sophisticated data analysis facilitates continuous monitoring of customer satisfaction and aids in identifying potential areas for service enhancement. By harnessing this methodology, the study ultimately seeks to contribute to developing more responsive and customer-centric service strategies, thereby promoting sustainable competitive advantages.

The methodology employed in this study involves the Knowledge Discovery in Databases (KDD) process, utilizing customer reviews of Tanjung Lesung Beach Hotel sourced from the Agoda platform. This approach begins with data collection and preparation, ensuring the extracted reviews are appropriately cleansed and pre-processed to remove noise and irrelevant information (Benlahbib & Nfaoui, 2020). Through data mining algorithms within the KDD framework, valuable insights are derived from unstructured textual data, enabling a deeper analysis of customer sentiments and service quality perceptions (Nguyen & Nguyen, 2023). By systematically transforming raw reviews into structured data, the study facilitates the identification of recurring themes and potential areas for service enhancement. Such an analytical approach is particularly advantageous as it transcends mere quantitative metrics, allowing for a nuanced understanding of guest experiences that could remain overlooked. Therefore, implementing KDD in analyzing feedback optimizes data utilization and supports the strategic improvement of service standards in the hospitality sector.

Integrating data-driven methodologies in feedback analysis within the hospitality sector faces multifaceted challenges that demand strategic attention. The extensive reliance on robust technological infrastructure necessitates significant investment in hardware and software to efficiently process and analyze large volumes of customer data. This requirement is compounded by the complexities of ensuring data authenticity and relevance, where unstructured feedback must undergo rigorous cleaning and preprocessing to yield actionable insights. Furthermore, the scarcity of expertise in advanced analytical techniques presents a critical obstacle, as the practical application of data mining algorithms depends on skilled human capital. These challenges are amplified by the dynamic nature of consumer expectations, which continuously evolve and require agile adjustments in service strategies. Addressing these barriers involves aligning organizational capabilities with technological advancements, fostering interdisciplinary collaboration, and embedding analytical practices into core business processes, promoting sustainable improvements in service quality and customer satisfaction.

The urgency of this research lies in the growing need for businesses, particularly in the hospitality industry, to adapt to rapidly evolving customer expectations by leveraging data-driven insights. In an increasingly competitive market, service quality remains a key differentiator, directly influencing customer loyalty and business sustainability (Rita et al., 2022). However, traditional feedback mechanisms often fail to capture the depth and nuance of customer experiences, resulting in missed opportunities for strategic improvements. Utilizing data mining techniques to analyze customer reviews systematically offers a robust solution for extracting actionable insights from vast,

unstructured datasets (Benlahbib & Nfaoui, 2020). This method enhances the precision of service quality assessments and supports proactive decision-making, allowing organizations to respond swiftly to emerging trends. Given the dynamic nature of consumer preferences, adopting such analytical tools is imperative for maintaining a competitive edge and ensuring continuous service excellence. Thus, implementing this research is crucial for fostering resilience and adaptability in service-oriented enterprises.

This research offers significant theoretical contributions and practical implications by advancing the understanding of data-driven approaches in service quality analysis. From a theoretical perspective, it enriches the existing body of knowledge by integrating data mining techniques within the framework of service quality assessment, thus bridging a gap between traditional qualitative evaluations and modern analytical methods (Aakash & Gupta Aggarwal, 2022; Rita et al., 2022). The study highlights the potential of leveraging unstructured customer feedback to uncover nuanced insights that conventional surveys might overlook. On a practical level, this approach equips service-oriented organizations with robust tools to analyze customer sentiments in real time, thereby enabling timely interventions to enhance service delivery (Lai et al., 2021; Thu, 2020). Businesses can better align their offerings with customer expectations by systematically analyzing feedback data, fostering customer loyalty and competitive advantage. Ultimately, this dual contribution advances academic discourse and provides a tangible roadmap for practitioners aiming to optimize service strategies through data analytics.

Numerous studies have explored data mining techniques to analyze customer feedback for service quality enhancement, demonstrating the method's efficacy in various industries (Ameur et al., 2023; Peng et al., 2020; Pham, 2020; Rahmaningrum & Oktaviana, 2020). For instance, prior investigations have successfully utilized sentiment analysis and natural language processing to extract actionable insights from large volumes of unstructured customer reviews, leading to improved service delivery strategies (Ahmad & Guzmán, 2021). However, existing research often focuses predominantly on either sentiment detection or general trend analysis, potentially overlooking more nuanced aspects, such as context-specific feedback that could inform personalized service improvements (Sann & Lai, 2021). This gap suggests an opportunity for future studies to integrate deeper contextual analysis through hybrid models combining sentiment scoring with topic modeling to capture more precise customer needs. Furthermore, extending this approach across diverse service sectors would validate its robustness and scalability, ensuring broader applicability. Future research could significantly contribute to theoretical advancements and practical implementations in service quality management by addressing these areas.

2. RESEARCH METHOD

2.1 Knowledge Discovery Databases (KDD)

The Knowledge Discovery in Databases (KDD) framework represents a systematic approach for extracting valuable information from vast datasets, particularly unstructured data sources such as customer reviews. This multi-phase process begins with data selection, identifying relevant datasets, followed by data cleaning to eliminate noise and inconsistencies (Zhao et al., 2021). Once the data is pre-processed, transformation techniques are applied to convert it into a format suitable for mining. This framework's core is the application of sophisticated algorithms that uncover patterns, correlations, and trends hidden within the data (Benlahbib & Nfaoui, 2020). By leveraging these techniques, KDD enables the extraction of meaningful insights and supports predictive analysis, allowing organizations to anticipate customer needs and optimize service delivery. This structured approach is highly advantageous as it combines data management with analytical rigor, enhancing decision-making processes' precision. As industries increasingly rely on data-driven strategies, adopting the KDD framework

becomes essential for achieving a competitive edge and fostering continuous improvement in service quality.



Figure 1. Knowledge Discovery Databases (KDD)

Figure 1 illustrates the structured process of the Knowledge Discovery in Databases (KDD) framework, encompassing several critical phases to extract meaningful insights from raw data. Data gathering forms the foundation, collecting relevant information from diverse sources. This is followed by data cleaning and preprocessing to eliminate noise, ensuring the dataset's quality. Once the data is refined, it is transformed into a format suitable for analysis, facilitating the subsequent phase of mining patterns. In this core step, analytical algorithms are employed to uncover hidden patterns, correlations, or trends within the dataset. The results obtained are then evaluated and interpreted, enabling the extraction of actionable insights that inform strategic decisions. Finally, the findings are the basis for formulating improvement strategies, particularly in enhancing service quality or optimizing organizational processes. The systematic progression through these phases not only maximizes the utility of the data but also supports informed, data-driven decision-making, thus emphasizing the KDD framework's value in leveraging large datasets effectively.

a. Gather Data

The data collection process is the foundational stage in the Knowledge Discovery in Databases (KDD) framework, where the emphasis is placed on gathering diverse data from multiple sources to create a robust and comprehensive dataset. This step typically includes aggregating information from structured sources such as databases and surveys and unstructured inputs like customer reviews and digital interactions on various platforms (Dogan & Birant, 2021). By incorporating data from these varied origins, the process ensures that the dataset encompasses various perspectives and nuances crucial for subsequent analysis. Including diverse data types enriches the dataset and enhances the depth and reliability of its insights. A well-rounded dataset increases the likelihood of uncovering meaningful patterns and trends, which would be instrumental in making data-driven decisions. Thus, the initial focus on thorough data collection establishes a solid foundation for all subsequent stages of analysis, ultimately driving the effectiveness of the analytical process.

The data is meticulously selected and validated systematically to ensure reliability and authenticity. Initially, relevant datasets are extracted from platforms like Agoda, which provides many customer reviews. These reviews undergo a rigorous cleaning process to eliminate duplicate, irrelevant, or inconsistent entries. For instance, out of 1,239 reviews analyzed for Tanjung Lesung Beach Hotel, only 642 verified reviews were deemed suitable for further analysis. The verification process focuses on identifying genuine feedback from authenticated users, enhancing the dataset's robustness. Validation steps include cross-checking metadata such as timestamps, reviewer profiles, and consistency of review content to ensure the data accurately represents customer experiences. The credibility of the data sources is ensured by relying on established platforms like Agoda, known for its extensive user base and robust review verification mechanisms. These platforms employ algorithms and manual checks to identify spam or fraudulent reviews, ensuring the dataset reflects authentic customer sentiments.

Additionally, the analysis minimizes bias and distortion by focusing on verified reviews, providing a reliable foundation for extracting actionable insights. This rigorous data sourcing and validation approach underscores its reliability for deriving meaningful patterns and informing strategic decisions in the hospitality industry.

b. Clean and Preprocess Data

The phase of cleaning and preprocessing data plays an indispensable role in ensuring the integrity and reliability of the dataset for subsequent analysis. After the data has been collected, it is essential to remove noise, inconsistencies, and irrelevant entries that could distort the analytical outcomes. This step eliminates duplicate entries and addresses missing values, which, if left unchecked, may lead to biased or incomplete insights (Al-Turaiki & Altwaijry, 2021). Standardizing data formats is equally critical, harmonizing diverse inputs into a uniform structure and facilitating efficient processing and analysis. A meticulous approach to data cleaning enhances the dataset's quality, making it more representative and robust. Refining the dataset through these procedures makes the analysis more precise, leading to more accurate and actionable insights. Thus, this stage serves as a vital precursor to uncovering reliable patterns and trends in the data.

The data cleansing process is structured and systematic to ensure the integrity and usability of the dataset for analysis. Initially, irrelevant or duplicate entries are identified and removed to reduce noise. Techniques such as text parsing and filtering are applied to eliminate non-informative elements, including excessive punctuation, memorable characters, and unrelated content. Missing values are addressed by imputation for critical fields or exclusion if the absence significantly impacts analysis reliability. Furthermore, standardization is employed to harmonize data formats, ensuring consistency across variables, while sentiment-specific preprocessing, such as stemming and lemmatization, refines textual data for sentiment analysis. While data cleansing is essential for improving dataset quality, there is an inherent risk of discarding valuable information, notably when borderline cases are excluded or oversimplified. For example, nuanced customer feedback that includes mixed sentiments may be inadvertently misclassified or removed during preprocessing. A balance between noise reduction and information retention is maintained to mitigate this risk by employing advanced algorithms that distinguish between irrelevant data and subtle variations in invaluable content. Regular audits of cleansing steps and including human oversight in ambiguous cases further minimize the likelihood of losing critical insights. This approach ensures that the process enhances data quality without compromising depth and representativeness.

c. Transform Data for Analysis

The transformation of data into an optimized structure is a critical step in preparing for practical analysis within the Knowledge Discovery in Databases (KDD) framework. After cleaning and preprocessing, the data is converted into formats more conducive to data mining techniques. This transformation often includes converting unstructured textual data into numerical representations, enabling algorithms to process qualitative information quantitatively (Ponnusamy et al., 2021). Additionally, the process may involve scaling features to ensure consistency across variables and normalizing values to prevent any undue influence of outliers on the analysis. Such adjustments are vital for improving the efficiency and accuracy of algorithmic models, allowing for a more precise identification of patterns and correlations. Structured data in a way that aligns with analytical requirements significantly enhances the robustness of subsequent data mining efforts, ultimately leading to more insightful and actionable outcomes.

d. Mine Data for Pattern

The data mining phase is pivotal in extracting meaningful insights by uncovering hidden patterns, trends, and correlations within the dataset. This stage employs sophisticated algorithms, including clustering to group similar data points, classification to categorize data, and association rule mining to identify relationships between variables (Molina-Coronado et al., 2020). These techniques allow for detecting intricate patterns often imperceptible through conventional analysis. By leveraging such algorithms, it becomes possible to transform vast amounts of raw data into actionable insights crucial for informed decision-making. The application of data mining not only enhances the depth of analysis but also enables a more precise understanding of underlying dynamics within the data. Consequently, this stage significantly contributes to converting data into a strategic resource, driving more effective and evidence-based decisions.

e. Evaluate and Interpret the Result

The evaluation and interpretation of results serve as a critical phase in validating the findings obtained from the data mining process. This stage thoroughly assesses the extracted patterns and correlations to determine their accuracy, relevance, and alignment with the study's objectives. Critically analyzing the results makes it possible to discern their implications, particularly how these insights can inform strategic decisions or improvements. This step is essential to confirm that the findings are not merely statistical anomalies but are reflective of genuine trends or behaviors within the dataset (Aurangzeb et al., 2021). Moreover, interpreting the results in the context of the original research objectives helps establish a clear connection between the analytical outcomes and the practical applications they support. Thus, this phase ensures the insights' reliability and provides a foundation for translating them into actionable strategies, thereby enhancing the value of the entire data-driven analysis.

f. Formulate Improvement Strategies

The final stage involves leveraging the insights gained from the analysis to devise targeted improvement strategies that enhance organizational processes, services, or decision-making frameworks. This step translates analytical findings into practical applications by identifying areas where efficiency or quality can be elevated (Sánchez-Franco & Rey-Tienda, 2024). Aligning these strategies with the organization's objectives ensures that data-driven insights are theoretical and integrated into actionable initiatives. This approach is precious in creating a feedback loop where continuous data analysis informs ongoing refinements, thus fostering adaptability and resilience in a dynamic market environment. The practical translation of data insights into strategic actions ultimately empowers organizations to optimize performance and sustain competitive advantages, demonstrating the tangible impact of informed decision-making.

3. RESULTS AND DISCUSSIONS

Implementing the KDD framework in the context of Tanjung Lesung Beach Hotel, rated 4 out of 5 stars, focuses on analyzing guest feedback from the Agoda platform, where a substantial dataset of 1,239 reviews, including 642 verified comments, has been collected. This diverse and comprehensive data source provides a rich foundation for uncovering insights into customer experiences and service quality perceptions. By systematically mining these reviews, it becomes possible to identify strengths and areas that require improvement, offering a more nuanced understanding of customer satisfaction. The verified nature of a significant portion of these comments adds credibility to the findings, ensuring that the derived insights are grounded in genuine customer experiences (Ho et al., 2020). Analyzing such a dataset highlights patterns in guest feedback and reveals specific factors that contribute to customer loyalty or dissatisfaction. Ultimately, the application of KDD in this scenario allows Tanjung Lesung

Beach Hotel to align its service strategies more closely with guest expectations, thereby enhancing its competitive standing in the hospitality sector.

Based on data extracted from the Agoda platform, the Tanjung Lesung Beach Hotel has earned an overall rating of 8.2, classified as "Excellent," from 1,239 reviews. Analyzing the individual indicators, the highest scores are observed in the categories of service (8.6) and facilities (8.3), suggesting that guests highly value the attentiveness of staff and the quality of amenities provided. Location and value for money score 8.2, indicating a solid perception of the hotel's strategic placement and pricing. Cleanliness receives a commendable score of 8.1, reflecting a satisfactory level of hygiene maintenance. However, the room comfort and quality category trails behind with a score of 7.3, revealing a potential area for improvement. This relatively lower rating highlights a discrepancy that might influence overall guest satisfaction, suggesting that targeted enhancements in room conditions could significantly elevate the hotel's standing. Addressing this aspect would improve comfort and enhance the guest experience, bolstering the hotel's reputation in the competitive hospitality market.



Figure 2. Hotel Guest based on Country of Origin (426 Accounts)

Figure 2 presents a graphical representation of hotel guests categorized by their country of origin, highlighting the diversity of the clientele at the Tanjung Lesung Beach Hotel. The data indicates that most guests come from a dominant country, as evidenced by the extensive bar segment representing a significant proportion of the total. In contrast, visitors from other countries are represented by much smaller segments, suggesting that international tourism is less prominent in comparison. This distribution may reflect the hotel's current market focus or its strategic positioning, which appears to attract a substantial number of guests from a specific region. Such a concentration could indicate strong brand recognition or targeted marketing efforts in that particular market. However, the relatively limited presence of guests from other countries implies potential opportunities for expanding the hotel's reach to a more global audience. By diversifying marketing strategies to attract guests from a broader range of countries, the hotel could enhance its occupancy rates and foster a more internationally diverse clientele.

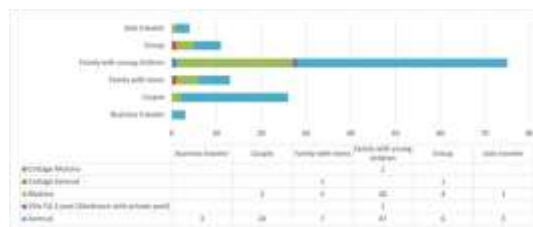


Figure 3. Popular Room based on Visitor Type (132 Account)

Figure 3 provides a detailed breakdown of room preferences based on different types of visitors, reflecting varied accommodation choices among different demographics at Tanjung Lesung Beach Hotel. The data reveals that families with young children prefer the "2 Bed (Bedroom with Private Pool)" option, suggesting a desire for spaciousness and privacy during their stay. In contrast, couples and groups appear to opt more frequently for the "Cottage Mutiara" and "Cottage Samudra" rooms, possibly due to their suitability

for intimate or shared experiences. On the other hand, business travelers and solo visitors exhibit a relatively lower demand across all room categories, indicating that the hotel's current offerings are more tailored to family and group-oriented vacations rather than business stays. This pattern underscores the hotel's positioning as a family-friendly destination, highlighting an opportunity to diversify its room types or marketing efforts to attract a broader range of guests, including solo and business travelers. The hotel could enhance occupancy rates and broaden its appeal by aligning room amenities with the specific needs of these underrepresented visitor types.



Figure 4. Popular Room Based on Year of Visit (233 Accounts)

Figure 4 illustrates the trends in room popularity at Tanjung Lesung Beach Hotel over the years, based on data from 233 guest accounts. The graph indicates a noticeable increase in the preference for the "Villa 2 Bed (Bedroom with Private Pool)" option, particularly from 2020 onward. This surge may be attributed to the growing demand for private and spacious accommodations, especially during the post-pandemic era when travelers valued privacy and social distancing more. In earlier years, the "Cottage Samudra" and "Cottage Mutiara" options were relatively popular, suggesting that guests initially favored traditional cottage settings. However, as guest preferences evolved, the villas with private pools have steadily gained dominance, likely due to their luxurious amenities and enhanced privacy. The data also reflects a consistent rise in overall bookings from 2020 to 2024, which may correlate with the gradual recovery of the tourism industry. These patterns highlight the importance of adapting accommodation offerings to align with shifting guest preferences, suggesting that continued investment in premium and private lodging options could further enhance the hotel's attractiveness and competitiveness in the market.

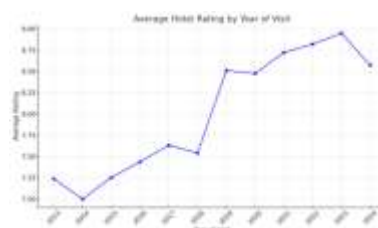


Figure 5. Average Hotel Rating by Year of Visit

Figure 5 illustrates the trend of average hotel ratings over the years, providing insights into the fluctuations in customer satisfaction. Between 2013 and 2017, ratings remained relatively stable, fluctuating within the 7.0 to 7.5 range, indicating moderate guest satisfaction. However, starting in 2018, a clear upward trajectory emerged, suggesting significant improvements in the hotel's services or facilities that positively impacted guest experiences. This positive trend peaked between 2020 and 2023, with average ratings consistently exceeding 8.5, likely reflecting successful enhancements in service quality, customer engagement, or facility upgrades. The peak scores in these years may also indicate an adequate adaptation to changing customer expectations or a successful response to post-pandemic travel demands. Nonetheless, a slight decline in ratings is observed in 2024, which may signal the emergence of new challenges or

shifting customer preferences. Addressing these factors promptly could be crucial for the hotel to sustain high levels of customer satisfaction and maintain its competitive edge in the market.

Analyzing service quality feedback using a data mining approach provides a nuanced understanding of customer perceptions, enabling organizations to refine their service offerings (Jeong et al., 2023; Sann & Lai, 2020; Sharma & Aggarwal, 2021; Shu et al., 2023). By systematically extracting insights from unstructured data, such as online reviews, it becomes possible to identify patterns and sentiments that traditional surveys may overlook. This method reveals general satisfaction trends and highlights specific areas of service that require improvement. For instance, the upward trend in ratings between 2018 and 2023, as shown in Figure 5, suggests that implementing targeted enhancements in service delivery likely contributed to improved customer experiences. However, the slight decline observed in 2024 may indicate evolving customer expectations or emerging service gaps that warrant attention. Therefore, leveraging feedback analysis through data mining is a strategic tool for continuously monitoring and adapting to customer needs, ensuring that service standards are maintained and progressively aligned with changing market demands.



Figure 6. Top 10 Words in Positive and Negative Reviews

Figure 6 visually represents the most frequently mentioned words in positive and negative reviews about Tanjung Lesung Beach Hotel service quality. The analysis highlights that positive feedback is predominantly associated with terms such as "staff," "beach," "good," and "nice," indicating guest satisfaction with the hospitality, scenic environment, and overall pleasant experiences. On the other hand, the negative feedback word cloud emphasizes concerns around "room," "water," and "breakfast." The prominence of words like "kurang" (less/insufficient) and "bad" suggests specific areas where guest expectations were not fully met, particularly concerning room conditions and service quality related to water amenities and dining options. These findings underscore the hotel's customer service and ambiance strengths while pointing to potential areas for targeted improvements to address recurring concerns. By focusing on these insights, the hotel management could further enhance guest satisfaction and overall service quality.

To enhance service quality, it is essential to prioritize targeted improvements based on customer feedback analysis. Addressing common concerns highlighted in negative reviews, such as room conditions, water quality, and breakfast offerings, should be a primary focus. Enhancing the maintenance and cleanliness of rooms and ensuring a consistent water supply would directly elevate the guest experience (Brzezińska-Wójcik, 2022; Sann et al., 2022; Tanrisevdi et al., 2022). Refining the breakfast menu to include a wider variety of high-quality options may address dissatisfaction in this area. Given the upward trend in customer ratings observed between 2018 and 2023, as illustrated in Figure 5, continuous investment in service enhancements influences guest satisfaction positively. However, the slight decline in ratings noted in 2024 indicates that evolving customer expectations must be proactively managed. By implementing these strategies, the hotel can sustain high service standards, fostering greater customer loyalty and improving overall competitiveness in the hospitality market.

4. CONCLUSION

The research utilized a structured methodology centered around the Knowledge Discovery in Databases (KDD) framework to ensure the rigor and accuracy of the analysis. This process began with data collection, gathering 1,239 customer reviews from the Agoda platform. Following this, an intensive data cleaning phase was conducted, which filtered out irrelevant and duplicate entries, resulting in a refined dataset. Out of the original pool, 642 reviews were verified as authentic hotel guest feedback and deemed suitable for analysis. By focusing on verified reviews, the study ensured the reliability of the data, thereby enhancing the validity of the insights drawn. Sentiment analysis and data mining techniques were applied to identify customer satisfaction patterns and service improvement areas. As illustrated in Figure 5, the trend analysis revealed a significant increase in customer ratings between 2018 and 2023, followed by a slight decrease in 2024. These findings underscore the importance of continuous service enhancements to align with evolving customer expectations. By employing a rigorous methodological approach, this study provides actionable recommendations to bolster service quality and sustain competitive advantage in the hospitality sector.

ACKNOWLEDGEMENTS

Thanks to the Tourism Study Program, Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, LPPM, and Center of Digital Transformation and Tourism Development for the support in this research publication.

REFERENCES

- Aakash, A., & Gupta Aggarwal, A. (2022). Assessment of Hotel Performance and Guest Satisfaction through eWOM: Big Data for Better Insights. *International Journal of Hospitality and Tourism Administration*, 23(2), 317–346. <https://doi.org/10.1080/15256480.2020.1746218>
- Ahmad, F., & Guzmán, F. (2021). Negative online reviews, brand equity and emotional contagion. *European Journal of Marketing*, 55(11), 2825–2870. <https://doi.org/10.1108/EJM-10-2019-0820>
- Al-Turaiki, I., & Altwaijry, N. (2021). A Convolutional Neural Network for Improved Anomaly-Based Network Intrusion Detection. *Big Data*, 9(3), 233–252. <https://doi.org/10.1089/big.2020.0263>
- Ameur, A., Hamdi, S., & Ben Yahia, S. (2023). Sentiment Analysis for Hotel Reviews: A Systematic Literature Review. *ACM Computing Surveys*, 56(2). <https://doi.org/10.1145/3605152>
- Aurangzeb, K., Ayub, N., & Alhoussein, M. (2021). Aspect Based Multi-Labeling Using SVM Based Ensemble. *IEEE Access*, 9, 26026–26040. <https://doi.org/10.1109/ACCESS.2021.3055768>
- Benlahbib, A., & Nfaoui, E. H. (2020). Aggregating customer review attributes for online reputation generation. *IEEE Access*, 8, 96550–96564. <https://doi.org/10.1109/ACCESS.2020.2996805>
- Brzezińska-Wójcik, T. (2022). Service Quality Evaluation in Four-Star Hotels in Lublin Using the Contextual Analysis Method. *Annales Universitatis Mariae Curie-Sklodowska. Sectio B*, 77, 149–176. <https://doi.org/10.17951/B.2022.77.0.149-176>
- Dogan, A., & Birant, D. (2021). Machine learning and data mining in manufacturing. In *Expert Systems with Applications* (Vol. 166). <https://doi.org/10.1016/j.eswa.2020.114060>
- Ho, R. C., Withanage, M. S., & Khong, K. W. (2020). Sentiment drivers of hotel customers: a hybrid approach using unstructured data from online reviews. *Asia-Pacific Journal of Business Administration*, 12(3–4), 237–250. <https://doi.org/10.1108/APJBA-09-2019-0192>
- Jeong, M., Shin, H. H., Lee, M., & Lee, J. (2023). Assessing brand performance consistency from consumer-generated media: the US hotel industry. *International Journal of Contemporary Hospitality Management*, 35(6), 2056–2083. <https://doi.org/10.1108/IJCHM-12-2021-1516>
- Lai, X., Wang, F., & Wang, X. (2021). Asymmetric relationship between customer sentiment and online hotel ratings: the moderating effects of review characteristics. *International Journal of Contemporary Hospitality Management*, 33(6), 2137–2156. <https://doi.org/10.1108/IJCHM-07-2020-0708>
- Molina-Coronado, B., Mori, U., Mendiburu, A., & Miguel-Alonso, J. (2020). Survey of Network

- Intrusion Detection Methods from the Perspective of the Knowledge Discovery in Databases Process. *IEEE Transactions on Network and Service Management*, 17(4), 2451–2479. <https://doi.org/10.1109/TNSM.2020.3016246>
- Nguyen, H. T. T., & Nguyen, T. X. (2023). Understanding customer experience with Vietnamese hotels by analyzing online reviews. *Humanities and Social Sciences Communications*, 10(1). <https://doi.org/10.1057/s41599-023-02098-8>
- Peng, Y. C., Cheng, M. S., & Hsu, P. Y. (2020). Constructing an Integrated Classifiers for Identifying Authenticity and Sentiment Analysis- A Case of Hotel Reviews. In *Proceedings - 2020 International Computer Symposium, ICS 2020* (pp. 319–324). <https://doi.org/10.1109/ICS51289.2020.00070>
- Pham, D. H. (2020). A latent aspect mining framework from textual reviews. *Indian Journal of Computer Science and Engineering*, 11(4), 347–359. <https://doi.org/10.21817/indjcse/2020/v11i4/201104209>
- Ponnusamy, V., Humayun, M., Jhanjhi, N. Z., Yichiet, A., & Almufareh, M. F. (2021). Intrusion Detection Systems in Internet of Things and Mobile Ad-Hoc Networks. *Computer Systems Science and Engineering*, 40(3), 1199–1215. <https://doi.org/10.32604/CSSE.2022.018518>
- Rahmaningrum, S. A., & Oktaviana, P. P. (2020). Sentiment classification of hotel service review on traveloka sites using naïve bayes classifier (NBC) and binary logistic regression. In *Journal of Physics: Conference Series* (Vol. 1490, Issue 1). <https://doi.org/10.1088/1742-6596/1490/1/012065>
- Rita, P., Ramos, R., Borges-Tiago, M. T., & Rodrigues, D. (2022). Impact of the rating system on sentiment and tone of voice: A Booking.com and TripAdvisor comparison study. *International Journal of Hospitality Management*, 104(2), 1–12. <https://doi.org/10.1016/j.ijhm.2022.103245>
- Sánchez-Franco, M. J., & Rey-Tienda, S. (2024). The role of user-generated content in tourism decision-making: an exemplary study of Andalusia, Spain. *Management Decision*, 62(7), 2292–2328. <https://doi.org/10.1108/MD-06-2023-0966>
- Sann, R., & Lai, P. C. (2020). Understanding homophily of service failure within the hotel guest cycle: Applying NLP-aspect-based sentiment analysis to the hospitality industry. *International Journal of Hospitality Management*, 91. <https://doi.org/10.1016/j.ijhm.2020.102678>
- Sann, R., & Lai, P. C. (2021). Do expectations towards Thai hospitality differ? The views of English vs Chinese speaking travelers. *International Journal of Culture, Tourism, and Hospitality Research*, 15(1), 43–58. <https://doi.org/10.1108/IJCTHR-01-2020-0010>
- Sann, R., Lai, P. C., Liaw, S. Y., & Chen, C. T. (2022). Predicting Online Complaining Behavior in the Hospitality Industry: Application of Big Data Analytics to Online Reviews. *Sustainability (Switzerland)*, 14(3). <https://doi.org/10.3390/su14031800>
- Sharma, H., & Aggarwal, A. G. (2021). The influence of user generated content on hotel sales: an Indian perspective. *Journal of Modelling in Management*, 16(4), 1358–1375. <https://doi.org/10.1108/JM2-07-2020-0201>
- Shu, Z., Torralba, M. H., Carrasco, R. A., & López, M. F. B. (2023). Assessing customer satisfaction of London luxury hotels with the AHP method and the SERVPERF scale: a case study of customer reviews on TripAdvisor. In *Procedia Computer Science* (Vol. 221, pp. 73–80). <https://doi.org/10.1016/j.procs.2023.07.011>
- Tanrısevdi, A., Öztürk, G., & Öztürk, A. C. (2022). A supervised data mining approach for predicting comment card ratings. *International Journal of Contemporary Hospitality Management*, 34(5), 1823–1853. <https://doi.org/10.1108/IJCHM-05-2021-0675>
- Thu, H. N. T. (2020). Measuring guest satisfaction from online reviews: Evidence in Vietnam. *Cogent Social Sciences*, 6(1). <https://doi.org/10.1080/23311886.2020.1801117>
- Zhao, N., Gao, H., Wen, X., & Li, H. (2021). Combination of convolutional neural network and gated recurrent unit for aspect-based sentiment analysis. *IEEE Access*, 9, 15561–15569. <https://doi.org/10.1109/ACCESS.2021.3052937>