Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

Understanding Hotel Customer Experience through User-Generated Reviews using Knowledge Discovery in Databases (KDD)

Yerik Afrianto Singgalen

Faculty of Business Administration and Communication, Tourism Study Program, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: yerik.afrianto@atmajaya.ac.id
Submitted: 02/10/2024; Accepted: 15/11/2024; Published: 15/11/2024

Abstract−This research explores the analysis of 388 hotel customer reviews to understand guest experiences, employing advanced analytical methodologies to uncover valuable insights for service quality enhancement. Utilizing the Knowledge Discovery in Databases (KDD) framework, the study applies Latent Dirichlet Allocation (LDA) for topic clustering and k-nearest Neighbors (k-NN), enhanced by the Synthetic Minority Over-sampling Technique (SMOTE) for sentiment classification. The integration of these techniques allows for the extraction of coherent thematic patterns and the accurate differentiation of sentiment categories within the reviews. The findings reveal that LDA, evaluated through metrics such as log-likelihood (-54,886.092) and coherence scores (-14.949), effectively captures the underlying themes discussed by guests, providing a clear representation of customer priorities and concerns. Additionally, applying SMOTE significantly improves the k-NN model's performance, achieving an accuracy of 91.43% and a precision of 97.26% by balancing class distributions and enhancing classification accuracy. This approach demonstrates the potential of combining topic modeling and sentiment analysis to derive actionable insights, which can be strategically utilized to optimize service delivery and elevate the overall customer experience in the hospitality industry. The study concludes that leveraging such data-driven methodologies facilitates a deeper understanding of customer feedback, ultimately supporting informed decision-making and continuous service improvement.

Keywords: Customer Experience; Hotel; User-Generated Reviews; KDD

1. INTRODUCTION

Exploring hotel customer experience through user-generated reviews leverages the robust methodology of Knowledge Discovery in Databases (KDD), offering a structured framework for extracting meaningful patterns from vast and unstructured data. As the hospitality industry increasingly shifts towards digital platforms, customer reviews' sheer volume and diversity necessitate advanced analytical techniques to discern critical insights about guest satisfaction and service quality. The application of KDD to this domain facilitates the systematic organization and transformation of raw data. It enhances identifying latent patterns and correlations that traditional methods might overlook [1]. Employing KDD, it is possible to construct a comprehensive understanding of customer sentiments and preferences, which are pivotal for improving service delivery and personalizing guest interactions [2]–[4].

Moreover, this approach supports the objective evaluation of feedback by minimizing biases associated with human interpretation, thereby ensuring a more nuanced understanding of customer expectations [5]. Such insights can subsequently inform strategic decision-making, guiding hotel management in refining service standards and fostering customer loyalty. This intersection of big data analytics and customer experience management exemplifies the potential of KDD to contribute significantly to developing more sophisticated and data-driven service enhancement strategies within the hospitality sector.

The urgency of this research is underscored by the rapid advancements in digital technology and the increasing reliance on data-driven decision-making across various industries. In today's dynamic and competitive landscape, organizations must continuously innovate to maintain relevance, necessitating a deeper understanding of complex datasets to uncover actionable insights [6]. The proliferation of data from multiple sources, such as customer feedback and market trends, demands sophisticated analytical approaches that traditional methods are ill-equipped to handle. Moreover, addressing this gap is critical for enhancing operational efficiency, improving customer satisfaction, and predicting future trends more accurately [7]. Employing advanced methodologies tailored to the context of big data, this research aims to bridge theoretical frameworks and practical applications, thereby equipping stakeholders with the knowledge required to navigate contemporary challenges. By prioritizing this study, it becomes possible to formulate strategic responses that are evidence-based and adaptable to the industry's evolving demands, ultimately contributing to long-term sustainability and competitive advantage.

The primary objective of this research is to develop an integrative framework that elucidates complex patterns within the given data context, thereby enhancing the overall understanding of the subject matter. By employing advanced analytical techniques and methodologies, the study seeks to systematically identify, extract, and interpret valuable insights that remain obscured in conventional data exploration [8]. Such a framework is crucial for addressing current limitations in knowledge, providing a more comprehensive perspective on the phenomenon under investigation [9], [10]. Additionally, this research aims to establish a foundation for subsequent studies by offering a replicable model that combines theoretical constructs with empirical validation. Ultimately, the study contributes substantively to the field by offering practical implications that can inform

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

policy, strategy, and future research, fostering a deeper engagement with the topic and paving the way for innovative solutions to emerging challenges.

The research employs the Knowledge Discovery in Databases (KDD) methodology to systematically analyze user-generated reviews from the Agoda platform, aiming to extract meaningful insights into hotel customer experiences. KDD, as a multi-step process, integrates various data preprocessing, transformation, and pattern extraction techniques, enabling the identification of hidden trends and relationships within the unstructured review data [11]. This approach is particularly suitable for addressing the complexity and volume of textual data typical of online reviews, as it ensures a structured exploration of customer sentiments and preferences [12]. Furthermore, the application of KDD in this context provides a robust framework for translating raw data into actionable knowledge, offering a comprehensive understanding of consumer perceptions and expectations. By leveraging this method, the study captures the nuanced dimensions of guest experiences. It facilitates the development of data-driven strategies that can be applied to enhance service quality and customer satisfaction within the hospitality industry.

This research's theoretical and practical contributions lie in its ability to bridge existing knowledge gaps while offering actionable insights for industry stakeholders. Theoretically, the study advances the current understanding of customer experience by integrating sophisticated data mining techniques with user-generated content analysis, thereby enriching the literature with a novel framework that captures the intricacies of consumer behavior and satisfaction [13]. This framework provides a more nuanced interpretation of how different aspects of service influence guest perceptions. It establishes a basis for further exploration into the dynamics of customer feedback in the digital era [14], [15]. The findings offer significant value to hospitality practitioners by translating complex analytical outcomes into strategic recommendations that enhance service quality, personalize guest interactions, and optimize resource allocation [16], [17]. By offering a dual contribution that marries theoretical advancements with tangible business applications, this research sets a precedent for employing data-driven methodologies in refining service delivery and improving competitive positioning within the hospitality sector.

Similar research on analyzing user-generated content has been conducted across various industries, particularly tourism and hospitality, due to the rich, real-time insights such data offers into consumer experiences. Studies employing text mining and sentiment analysis have extensively utilized reviews from TripAdvisor and Booking.com to investigate customer satisfaction and predict service quality [18], [19]. Such research has demonstrated the efficacy of advanced computational techniques in distilling large volumes of textual data into structured information, thereby uncovering patterns that are not immediately apparent through traditional qualitative methods [20]. Integrating machine learning models, such as natural language processing (NLP), with user-generated reviews has proven instrumental in understanding consumer sentiments at a granular level [21]. While these studies have successfully highlighted the potential of leveraging big data to inform strategic decisions, the present research aims to build upon this foundation by applying a more refined analytical framework, potentially expanding the scope and depth of knowledge regarding customer behavior [22], [23]. By comparing various methodological approaches and contextual applications, it becomes evident that a comprehensive, data-driven understanding of user feedback is indispensable for achieving service excellence and sustaining competitive advantage in service-oriented industries.

The limitations of this research primarily stem from the inherent challenges associated with data quality and contextual interpretation. While rich in consumer insights, user-generated reviews often exhibit language, expression, and sentiment inconsistencies that may introduce noise into the analytical process. Such variability necessitates extensive preprocessing and may limit the accuracy of sentiment classification and pattern extraction, mainly when dealing with ambiguous or context-dependent terms. Additionally, the research is constrained by the specific dataset utilized, as it relies solely on reviews from a single platform, potentially restricting the generalizability of the findings to other contexts or platforms with different user demographics and review structures. Moreover, the reliance on text-based data excludes other potential customer experience indicators, such as numerical ratings or behavioral data, which could provide a more holistic view of consumer satisfaction. Despite these constraints, the research methodology has been carefully designed to mitigate some limitations by employing advanced analytical techniques that enhance data interpretation. Nevertheless, acknowledging these limitations is crucial for guiding future studies toward a more comprehensive exploration of user-generated content and its implications for service quality assessment.

Future research endeavors should explore integrating multimodal data sources to provide a more comprehensive understanding of customer experience. Incorporating not only textual reviews but also numerical ratings, behavioral data, and multimedia content such as images or videos would enable a more nuanced analysis of consumer perceptions and preferences. Expanding the scope to include diverse datasets from multiple platforms would also enhance the robustness of findings and facilitate cross-platform comparisons, offering more profound insights into the consistency and variability of consumer behavior across different contexts. Additionally, implementing advanced machine learning models, such as deep learning algorithms, could refine sentiment analysis and topic modeling, improving the precision of identifying subtle patterns within complex data. Another promising direction is the longitudinal study of customer feedback to observe how sentiments and preferences evolve, particularly in response to changes in service delivery or market conditions. Such an

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

approach would contribute significantly to the predictive modeling of customer satisfaction trends, ultimately informing proactive strategies for enhancing service quality and fostering long-term customer loyalty in the competitive landscape of service industries.

2. RESEARCH METHODOLOGY

2.1 Hotel Customer Experience

Recent trends in the study of hotel customer experience have demonstrated a significant shift towards integrating technology and data analytics to capture a more comprehensive understanding of guest perceptions and satisfaction. With the increasing prevalence of user-generated content and the evolution of digital platforms, researchers are now focusing on themes such as online reviews, sentiment analysis, and the impact of digital transformation on service quality [24], [25]. Studies have explored how these digital interactions influence customer loyalty, brand attachment, and overall satisfaction, particularly in the context of service innovation and experiential quality. Additionally, integrating concepts such as service design, emotional labor, and service failure has provided more profound insights into the nuances of the customer experience management [26], [27]. The implications of these studies suggest that harnessing big data and advanced analytics is crucial for hoteliers to respond proactively to customer expectations and enhance service standards. Given the competitive nature of the hospitality industry, this trend underscores the importance of leveraging emerging technologies and data-driven strategies to maintain a competitive edge and foster long-term customer relationships.

The novelty of this research lies in its integration of advanced data mining techniques with user-generated content analysis to develop a comprehensive framework for understanding hotel customer experiences. Unlike previous studies that often focused on isolated aspects of customer satisfaction or employed traditional qualitative methods, this research introduces a multifaceted approach that leverages the potential of big data analytics to uncover latent patterns and insights within large volumes of textual reviews [28], [29]. This study provides a more granular understanding of how customers perceive and evaluate various aspects of hotel services by employing methodologies such as Knowledge Discovery in Databases (KDD) and natural language processing (NLP). This innovative approach enables a more accurate assessment of customer sentiments, offering valuable contributions to academic literature and practical applications in service quality management. As a result, the study establishes a new benchmark for utilizing complex data in the hospitality context, presenting a replicable model that can be adapted across different service industries to optimize customer satisfaction and strategic decision-making.

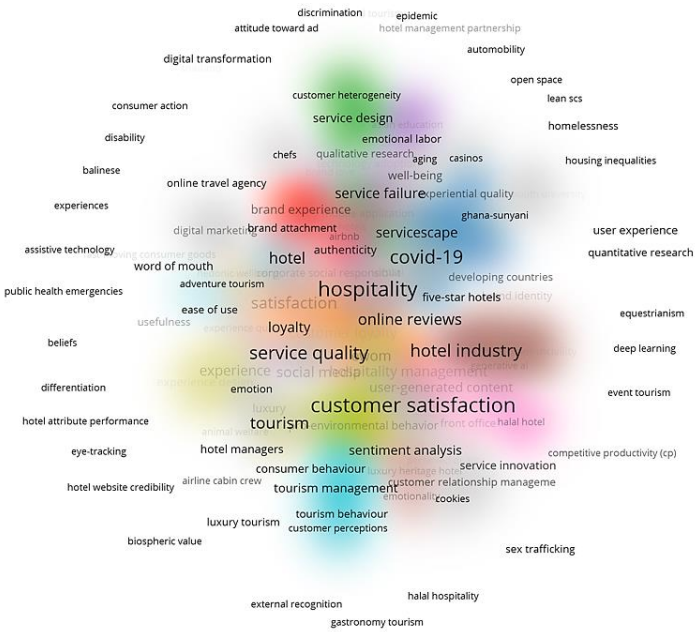


Figure 1. Conceptual Landscape Hotel Customer Experience Research (Vosviewer)

Figure 1 illustrates the conceptual landscape of hotel customer experience research, visualized using the VOSviewer software. The figure employs a clustering approach to display various research themes, with keywords such as “hospitality,” “customer satisfaction,” “service quality,” and “online reviews” appearing prominently at the center, signifying their pivotal role in this field of study. Surrounding these core concepts are subtopics like “loyalty,” “brand attachment,” and “service design,” which indicate a strong focus on understanding the emotional and experiential aspects of customer engagement. The presence of keywords such

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

as “COVID-19” and “digital transformation” suggests a contemporary shift in research interests toward exploring how external factors and technological advancements influence customer perceptions and behaviors. Furthermore, terms like “sentiment analysis” and “user-generated content” highlight the increasing utilization of advanced analytical methods to decode consumer feedback. This figure encapsulates the complexity and interdisciplinary nature of hotel customer experience research, demonstrating the field's evolution from traditional service quality assessments to a more comprehensive analysis encompassing digital interactions and evolving consumer expectations. Such visualization provides a valuable roadmap for identifying emerging trends and potential gaps in the existing literature, guiding future research directions in this dynamic and multifaceted domain.

The study of understanding hotel customer experience through user-generated reviews using Knowledge Discovery in Databases (KDD) is essential in providing deeper insights into customer perceptions and service quality. Using guest review data from platforms like Agoda, this research captures authentic customer feedback, reflecting real-world experiences and sentiments. Analyzing such data through KDD enables the extraction of valuable patterns and trends that might otherwise remain hidden in unstructured text. This method offers a robust approach to identifying key factors influencing customer satisfaction and loyalty, as it systematically processes large volumes of reviews to reveal nuanced insights into guest preferences and service shortcomings. Given the competitive nature of the hospitality industry, leveraging such findings is critical for hoteliers to tailor their services more effectively and enhance overall guest experiences. Furthermore, employing KDD in this context enriches the theoretical understanding of customer experience management and provides practical strategies for improving service quality based on data-driven evidence. This study, therefore, holds significant potential in bridging the gap between customer expectations and service delivery, ultimately contributing to the sustainable growth and competitiveness of hotels in a digitalized marketplace.

2.2 Knowledge Discovery Databases (KDD)

Knowledge Discovery in Databases (KDD) encompasses systematic stages to extract meaningful insights from complex datasets. It begins with data gathering, where relevant information is collected from various sources to ensure a comprehensive dataset. This is followed by data cleaning and preprocessing, which involves handling missing values, removing inconsistencies, and transforming data into a suitable format for analysis. The third stage, data transformation, restructures the data to facilitate more effective mining. After this, data mining is performed to identify patterns, trends, and correlations within the dataset using various analytical techniques. The penultimate stage focuses on evaluating and interpreting the mined results, ensuring that the patterns discovered are valid and valuable in the context of the research objectives. Finally, these results are leveraged to formulate strategic recommendations or improvement strategies, translating complex data patterns into actionable knowledge. By adhering to these stages, KDD provides a robust framework for converting raw data into insightful information, supporting decision-making and strategic planning across diverse fields.

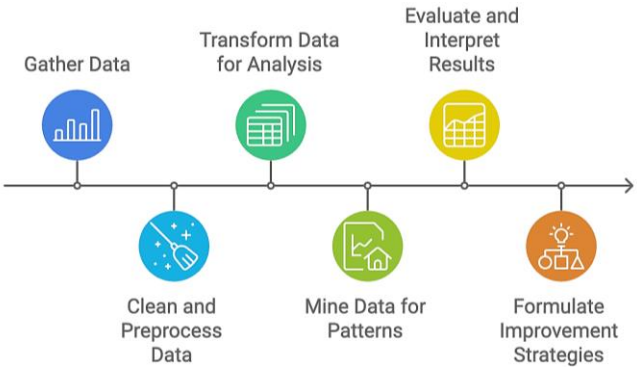


Figure 2. Knowledge Discovery Databases (KDD)

Figure 2 illustrates the sequential stages of the Knowledge Discovery in Databases (KDD) process, which systematically converts raw data into valuable insights. The process begins with data gathering, where relevant datasets are collected from various sources to ensure comprehensive coverage of the research context. Subsequently, the clean and preprocess stage addresses inconsistencies, handles missing values, and removes noise to optimize data quality. Data is restructured and formatted in the transformation phase to facilitate more efficient analysis. The data mining stage then applies various algorithms and analytical methods to identify hidden patterns, trends, and correlations within the structured dataset. Following this, the evaluation and interpretation stage ensures the mined results are meaningful and align with the research objectives. The final stage involves formulating improvement strategies based on the interpreted results, transforming the discovered knowledge into actionable strategies for enhancing decision-making or optimizing processes. This systematic framework enhances the accuracy and reliability of the insights obtained and establishes a robust foundation for data-driven decision-making across diverse applications.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

The relevance of Knowledge Discovery in Databases (KDD) to this research context lies in its ability to systematically process and analyze large volumes of unstructured data, such as user-generated reviews, to uncover hidden patterns and insights that are crucial for understanding customer experiences. By utilizing the structured approach of KDD, this study effectively transforms complex textual data into quantifiable and interpretable knowledge, facilitating a more nuanced exploration of customer sentiments, service quality perceptions, and overall satisfaction levels. The application of KDD is particularly pertinent given the dynamic nature of customer feedback in the hospitality industry, where sentiments expressed in reviews can vary widely based on numerous factors such as service quality, brand loyalty, and individual expectations. Through its multi-stage methodology—encompassing data cleaning, transformation, mining, and interpretation—KDD enhances the accuracy and depth of the analysis and ensures that the findings are robust and reflective of real-world customer perspectives. Consequently, this approach provides a solid foundation for deriving strategic insights to inform evidence-based decision-making and foster service enhancements within the hotel industry, demonstrating its critical role in advancing academic understanding and practical applications.

2.2.1 Gather Data: Local Collection Hotel in Labuan Bajo Reviews on Agoda Platform

The data utilized in this research is derived from Agoda, a platform known for its robust review features, which enable customers to provide detailed feedback on products and services during their hotel stays. The specific hotel analyzed in this study is the Loccal Collection Hotel in Labuan Bajo, a prominent destination in Indonesia's hospitality industry. Leveraging Agoda's user-generated reviews offers a distinct advantage, as the platform facilitates the collection of authentic and diverse customer opinions that reflect various dimensions of service quality and guest satisfaction. Focusing on the Local Collection Hotel provides a concentrated case study that captures the unique customer experience attributes within this context, making it an ideal subject for in-depth analysis. By examining reviews from this source, the research gains access to a wealth of information that accurately represents customer perspectives and helps identify specific areas of service excellence and potential improvement. This approach underscores the importance of utilizing comprehensive and credible data sources to gain nuanced insights into customer behavior, ultimately contributing to more targeted strategies for enhancing service delivery in the competitive hospitality market.

Figure 3. Rating of Loccal Collection Hotel in Labuan Bajo on Agoda Platform

Figure 3 presents the rating overview of Loccal Collection Hotel in Labuan Bajo on the Agoda platform, reflecting its outstanding reputation based on 1,025 customer reviews. The hotel achieves an overall rating of 8.5, classified as "Excellent," which underscores its ability to meet or exceed guest expectations across various service dimensions. Specifically, the hotel excels in location with a score of 9.0, indicating its prime position in proximity to key attractions and amenities that enhance the guest experience. Service quality is rated at 8.8, highlighting the competence and professionalism of the hotel staff. At the same time, cleanliness receives an impressive score of 8.7, demonstrating the hotel's commitment to maintaining a hygienic and comfortable environment. The facilities are rated at 8.5, reflecting the adequacy and appeal of amenities catering to leisure and business travelers. Additionally, value for money, scored at 8.3, suggests that guests perceive the hotel's offerings as providing substantial benefits relative to the cost, making it a favorable choice for those seeking quality accommodation without compromising on budget. Collectively, these ratings establish Loccal Collection Hotel as a leading hospitality provider in Labuan Bajo, capable of delivering a comprehensive and satisfying guest experience.

On the Agoda platform, the Loccal Collection Hotel in Labuan Bajo offers a diverse selection of accommodations with four distinct room types and 16 unique room offers. The room categories available include Standard, Superior, Suites, and King Suite, each designed to cater to varying guest preferences and expectations. This variety allows the hotel to attract a broader demographic of customers, ranging from budget-conscious travelers to those seeking a more luxurious and exclusive experience. Such diverse in-room offerings indicate the hotel's strategic approach to accommodating diverse market segments, enhancing its competitive position within the hospitality industry. The availability of multiple room options ensures that the hotel can meet different demand levels while maintaining a high standard of service quality across all categories. This comprehensive range of accommodations reflects the hotel's commitment to providing personalized experiences. It positions it

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

as a versatile choice for guests visiting Labuan Bajo, thereby contributing to its reputation as a premier hospitality provider in the region.

(a) (b)

Figure 4. Local Collection Hotel Facilities in Labuan Bajo (Source: Agoda)

Figure 4 showcases the facilities of the Loccal Collection Hotel in Labuan Bajo, highlighting the architectural design and amenities that contribute to its appeal as a premium hospitality establishment. The images depict outdoor seating areas and terraces that offer guests panoramic views of the surrounding landscapes, creating an inviting and serene environment ideal for relaxation and social interaction. The use of contemporary design elements, such as clean lines, spacious layouts, and minimalist decor, complements the hotel's aesthetic, enhancing the overall guest experience. These facilities are strategically designed to cater to guests' diverse needs, providing spaces for leisure activities, dining, or simply enjoying the natural beauty of Labuan Bajo. The emphasis on spacious and well-maintained facilities also underscores the hotel's commitment to providing a comfortable and aesthetically pleasing atmosphere. This combination of functional design and aesthetic consideration positions the Loccal Collection Hotel as an attractive choice for visitors seeking comfort and visual appeal, further strengthening its competitive advantage in the hospitality market.

Three hundred eighty-eight data entries were successfully collected from the 434 verified guest reviews provided by the hotel, representing a comprehensive dataset for analyzing customer experiences. This high retention rate of usable data indicates the reliability and consistency of the review content, making it a valuable resource for gaining insights into various aspects of guest satisfaction. The discrepancy between the verified and collected reviews could be attributed to factors such as incomplete data or reviews that did not meet the criteria for inclusion, ensuring that only high-quality and relevant feedback is analyzed. Such a robust dataset enables a more nuanced exploration of patterns and trends in customer sentiments, providing a solid foundation for formulating strategic recommendations to enhance service quality and guest satisfaction. Consequently, the filtered and validated data serve as a credible basis for extracting actionable insights, contributing to the overall objective of the research in understanding and improving the hotel's service delivery and customer engagement.

2.2.2 Clean and Preprocess Data

The 388 collected review data entries can be systematically classified based on multiple attributes, including account name, country of origin, guest type, room type, length of stay, month of stay, year of stay, rating, description of rating, review content, and review date. This comprehensive classification scheme enables a structured analysis of customer feedback by organizing the data into distinct categories that facilitate the identification of patterns and correlations. For instance, segmenting reviews by guest type and room type allows for a deeper understanding of the preferences and expectations of different customer segments, while the temporal attributes, such as month and year of stay, provide insights into seasonal trends and service performance over time. Additionally, including rating descriptions and detailed review content offers a qualitative perspective that complements the quantitative ratings, enabling a more holistic evaluation of guest experiences. Such a multidimensional classification not only enhances the clarity and usability of the data but also supports more targeted and nuanced analyses, ultimately contributing to strategic decision-making processes to optimize service quality and guest satisfaction.

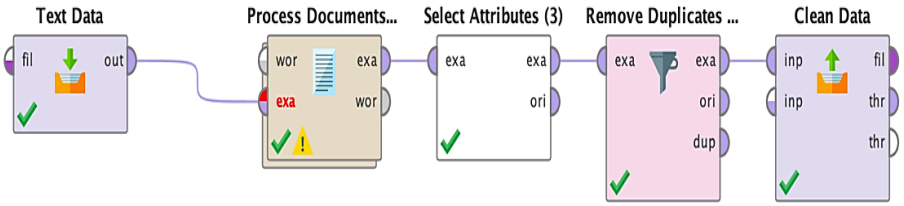


Figure 5. Data Cleaning Process (Rapidminer)

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

Figure 5 illustrates the data cleaning process using RapidMiner, a comprehensive tool for preparing and refining raw datasets for further analysis. The workflow begins with importing raw text data, which is then subjected to a series of document processing techniques designed to extract meaningful attributes and standardize the content. This stage ensures that textual data is formatted correctly and relevant features are identified. The subsequent step, attribute selection, isolates specific variables of interest, enhancing the clarity and focus of the analysis. Following this, the process removes any duplicate entries to eliminate redundancy and prevent bias in the data, thereby increasing the accuracy and reliability of the results. The final step, clean data generation, outputs a refined dataset free from inconsistencies and ready for advanced analytical procedures. This structured data-cleaning approach optimizes data quality and lays a solid foundation for extracting precise insights, highlighting the critical role of preprocessing in the overall data analysis pipeline.

2.2.3 Transform Data and Mine Data for Patterns

Transforming data using the VADER algorithm for sentiment extraction, followed by topic clustering through Latent Dirichlet Allocation (LDA), involves several sequential steps to ensure precise and meaningful data representation. Initially, the text data is preprocessed and standardized to eliminate noise and format inconsistencies, making it suitable for sentiment analysis. VADER, a rule-based model optimized for social media text, is then employed to classify sentiments into positive, negative, or neutral categories based on lexical and grammatical features. This classification enables a quantitative assessment of emotional tone within the reviews, providing insights into customer attitudes and experiences. Following sentiment extraction, the LDA algorithm is applied to identify latent topics within the text by clustering related terms and concepts. LDA assigns probability distributions to words across different topics, thus enabling the detection of thematic structures and patterns that might not be evident through traditional analysis methods. Combining sentiment extraction and topic modeling, this approach quantifies customer sentiments and contextualizes them within relevant themes, offering a comprehensive understanding of the data. Such integration of VADER and LDA enhances the analytical depth, making it possible to uncover complex relationships between sentiment and specific topics, thereby supporting more informed decision-making based on nuanced insights.

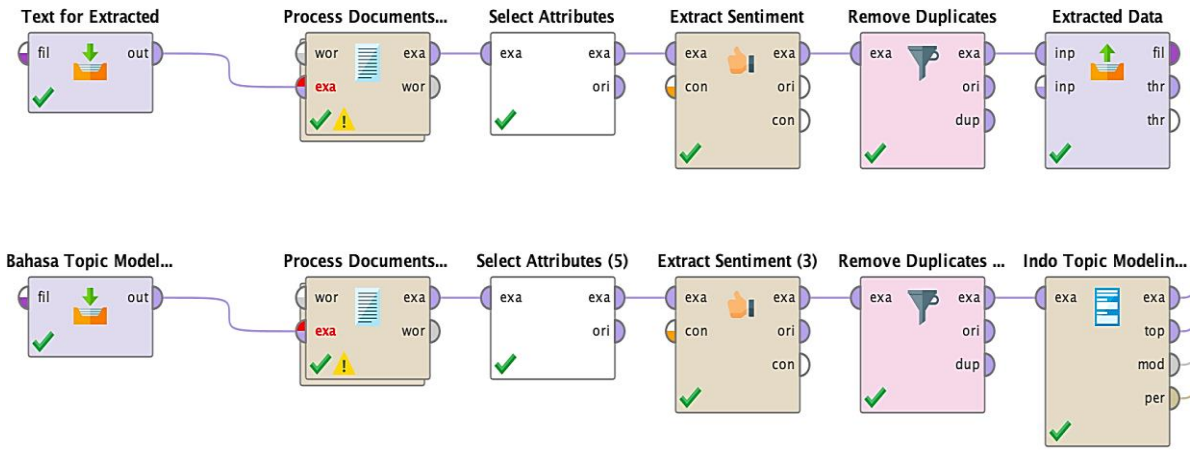


Figure 6. Implementation of Vader and LDA in Sentiment Classification and Topic Modeling (Rapidminer)

Figure 6 illustrates the implementation of the VADER algorithm for sentiment classification and the Latent Dirichlet Allocation (LDA) method for topic modeling, executed within the RapidMiner environment. The process initiates with text data preparation, where raw textual information undergoes preprocessing to standardize and eliminate irrelevant elements. Following this, VADER performs sentiment extraction, categorizing each review based on its polarity—positive, negative, or neutral—using a rule-based approach that accounts for semantic nuances and grammatical structures. This step provides a quantitative measure of customer sentiment, crucial for understanding overall customer perceptions. After the sentiment analysis, LDA is applied to the preprocessed text data to uncover latent topics within the reviews by clustering words with high probabilities of co-occurrence into coherent themes. This probabilistic model identifies hidden patterns and structures within the textual data, thereby revealing key topics that represent the most prevalent themes discussed by customers. Combining VADER for sentiment classification with LDA for topic modeling offers a comprehensive analytical framework, allowing emotional tone and thematic content extraction from large-scale textual data. This integrated approach enhances the interpretability of complex datasets, making it possible to derive more profound insights into customer feedback that inform strategic decision-making and service improvements.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

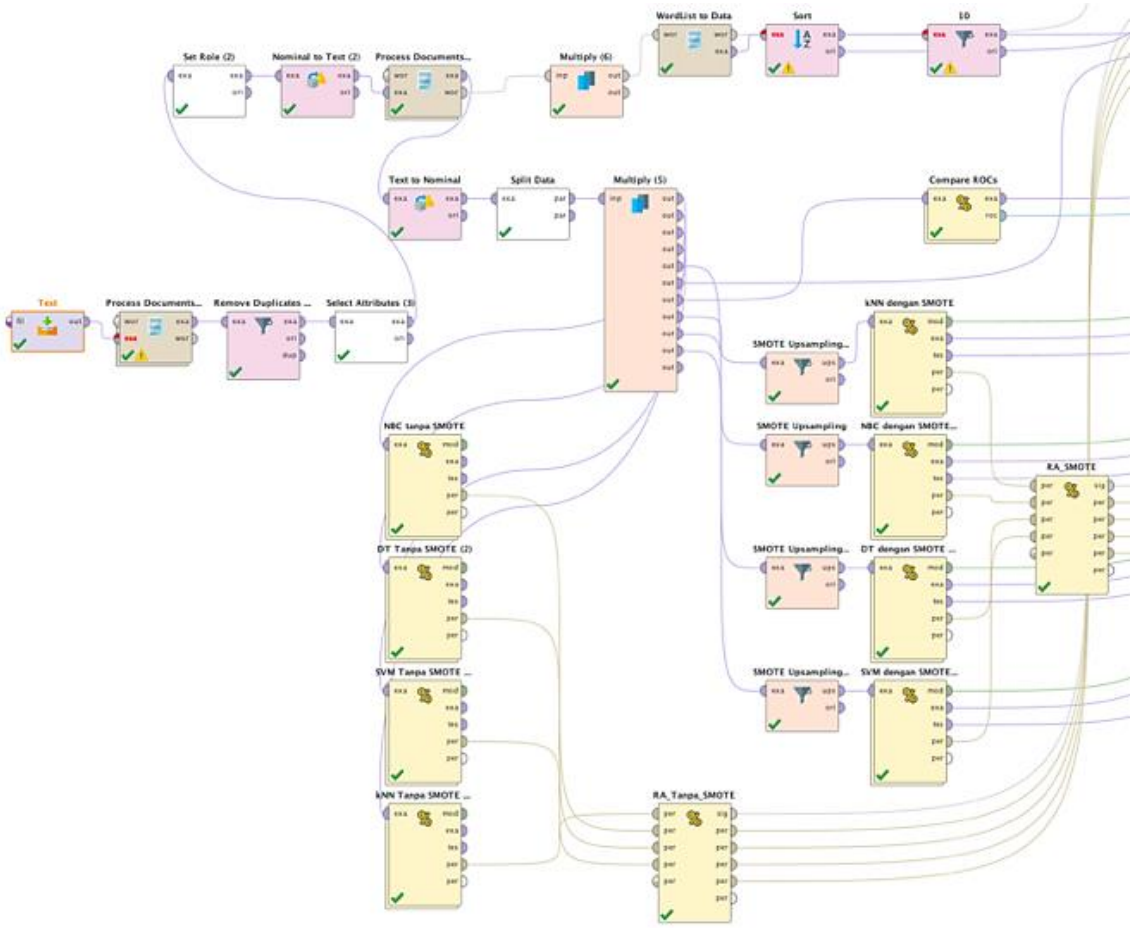


Figure 7. Classification Algorithm Evaluation

Figure 7 illustrates the evaluation process of various classification algorithms within a comprehensive analytical framework designed to assess their performance and accuracy. The workflow consists of multiple stages, beginning with the preprocessing of input data and feature selection, followed by applying diverse classification models, such as Decision Tree, Naive Bayes, and Support Vector Machine (SVM). Each model undergoes a series of iterations to optimize its parameters and enhance predictive performance. The evaluation metrics employed, including precision, recall, F1-score, and accuracy, are systematically compared to determine the efficacy of each algorithm in categorizing the data. This comparative analysis not only identifies the most suitable model for the given context but also highlights the strengths and limitations of each algorithm based on their performance across different datasets. By visualizing the outcomes, the figure effectively demonstrates how varying classification methods yield distinct results, providing valuable insights into selecting the optimal approach for specific analytical objectives. Such a structured evaluation process is critical for ensuring the chosen model is robust and reliable, supporting informed decision-making in data-driven research and applications.

After completing the transform data and mine data for patterns stages, the subsequent phase involves evaluating and interpreting the results to ensure that the identified patterns align with the research objectives and provide actionable insights. This transition is critical, as it verifies the relevance and accuracy of the extracted data, allowing for a comprehensive understanding of the underlying trends and relationships within the dataset. By systematically assessing these outcomes, it becomes possible to identify areas that require further refinement or additional exploration. Once the evaluation confirms the validity of the findings, the process proceeds to formulate improvement strategies, where the insights gained are translated into practical recommendations. These strategies may include optimizing service delivery, enhancing customer engagement, or addressing identified gaps in performance. This final stage consolidates the research findings and serves as a bridge between theoretical analysis and real-world application, facilitating the implementation of data-driven strategies that can significantly contribute to organizational growth and customer satisfaction.

2.2.4 Evaluate and Interpret Results and Formulate Improvement Strategies

The evaluation stage involves assessing the effectiveness of the Latent Dirichlet Allocation (LDA) results as well as the performance of the k-Nearest Neighbors (k-NN) algorithm in classifying consumer reviews into positive and negative sentiments. This process is crucial for validating the coherence and interpretability of the topics

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

generated by LDA, ensuring that the identified themes accurately represent the underlying patterns within the textual data. Additionally, the performance of the k-NN algorithm is measured using standard evaluation metrics such as accuracy, precision, recall, and F1-score, which provide a quantitative understanding of its capability to differentiate between sentiment categories. Such an evaluation not only highlights the strengths and weaknesses of the classification model but also offers insights into potential areas of improvement, such as fine-tuning parameters or optimizing feature selection. By systematically analyzing these aspects, the evaluation stage facilitates a deeper understanding of the model's reliability and effectiveness, ultimately informing more refined strategies for sentiment classification and topic analysis in consumer review datasets.

LDA Performance Vector

```
PerformanceVector:
LogLikelihood: -54886.092
Perplexity: 665.673
Avg(tokens): 738.900
Avg(document_entropy): 3.009
Avg(word-length): 4.960
Avg(coherence): -14.949
Avg(uniform_dist): 2.518
Avg(corpus_dist): 2.449
Avg(eff_num_words): 151.735
Avg(token-doc-diff): 0.013
Avg(rank_1_docs): 0.475
Avg(allocation_count): 0.665
Avg(exclusivity): 0.656
AlphaSum: 0.358
Beta: 0.162
BetaSum: 416.467
```

(a)

k-NN Performance Vector enhanced by SMOTE

```
PerformanceVector:
accuracy: 91.43% +/- 3.13% (micro average: 91.43%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       273     41
Positive:       7       239
AUC (optimistic): 0.989 +/- 0.007 (micro average: 0.989) (positive class: Positive)
AUC: 0.947 +/- 0.031 (micro average: 0.947) (positive class: Positive)
AUC (pessimistic): 0.904 +/- 0.056 (micro average: 0.904) (positive class: Positive)
precision: 97.26% +/- 3.67% (micro average: 97.15%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       273     41
Positive:       7       239
recall: 85.36% +/- 4.89% (micro average: 85.36%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       273     41
Positive:       7       239
f_measure: 90.83% +/- 3.41% (micro average: 90.87%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       273     41
Positive:       7       239
```

(b)

Figure 8. LDA and k-NN Performance Vector Evaluation (a and b)

Figure 8 presents the performance evaluation of both the Latent Dirichlet Allocation (LDA) and k-nearest Neighbors (k-NN) algorithms, with the latter enhanced using the Synthetic Minority Over-sampling Technique (SMOTE). The LDA performance vector metrics, such as log-likelihood, perplexity, and coherence scores, indicate the model's effectiveness in generating coherent topics that accurately reflect the underlying structure of the text data. Lower perplexity and higher coherence values suggest that the model successfully captured the semantic relationships between words and topics. In contrast, the k-NN performance vector, evaluated using accuracy, precision, recall, and F1-score, demonstrates the algorithm's ability to classify consumer reviews into positive and negative sentiments with reasonable accuracy. Using SMOTE in the k-NN model enhances its performance by balancing the class distribution, thereby mitigating the effects of data imbalance and improving classification outcomes for the minority class. The comparative evaluation of these models provides insights into their respective strengths: LDA excels in identifying latent thematic patterns, while k-NN, combined with SMOTE, proves effective in sentiment classification by addressing class imbalances and enhancing predictive reliability. Such an integrated approach offers a robust framework for analyzing complex textual data, enabling a comprehensive understanding of consumer opinions and sentiments.

Following identifying topic clusters and guest sentiments, strategies for enhancing the immersive hotel customer experience can be formulated specifically for the context of Loccal Collection Hotel's facilities in Labuan Bajo. Analyzing these clusters provides a nuanced understanding of guests' expectations, preferences, and areas of concern, which are pivotal for developing targeted improvement strategies. For instance, positive sentiments associated with aesthetics and amenities can be further leveraged by introducing personalized guest experiences, such as curated local tours or thematic dining options that align with the hotel's unique character. Conversely, any negative feedback on service efficiency or facility maintenance can be addressed by implementing staff training programs and scheduling regular facility inspections to uphold service quality standards. Integrating these insights into the hotel's operational strategy enhances guest satisfaction and fosters a deeper connection between the hotel's offerings and customer preferences. Such a data-driven approach enables Loccal Collection Hotel to optimize its service delivery, create memorable experiences, and position itself as a top-tier destination in Labuan Bajo, ultimately contributing to sustained competitive advantage and guest loyalty.

3. RESULT AND DISCUSSION

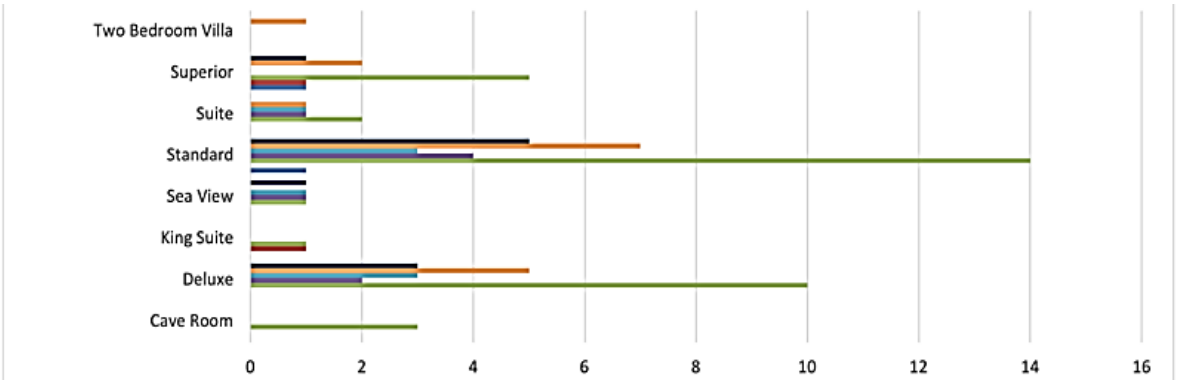
The discussion in this research is divided into two key sections: 3.1 Customer Experience Analysis Based on Reviews Data and 3.2 Topic Clustering and Sentiment Classification Performance. The first section focuses on a comprehensive analysis of customer experiences derived from user-generated reviews, highlighting key themes and patterns that reflect guest perceptions and satisfaction levels. This analysis provides an empirical foundation

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

for understanding how various aspects of hotel services influence customer sentiments and overall experiences. The second section delves into the performance evaluation of topic clustering and sentiment classification techniques used in the study. By applying algorithms such as Latent Dirichlet Allocation (LDA) for topic modeling and k-nearest Neighbors (k-NN) for sentiment analysis, this segment assesses the effectiveness and accuracy of the models in categorizing and interpreting review data. The discussion synthesizes these findings, emphasizing the alignment between identified topics and customer sentiment, thus validating the analytical approach. Ultimately, these two sections offer a holistic view of how customer experiences can be systematically examined and leveraged to formulate strategic recommendations for service enhancement and customer satisfaction optimization within the hospitality context.

3.1 Customer Experience Analysis Based on Reviews Data: Local Collection Hotel in Labuan Bajo

The analysis of customer experience based on review data for Loccal Collection Hotel provides valuable insights into guest perceptions and satisfaction levels, offering a detailed understanding of service quality from the consumers' perspective. The reviews highlight room quality, service efficiency, and overall ambiance, critical determinants of a positive guest experience. Positive feedback often emphasizes the aesthetic appeal of the hotel's design, the comfort of its accommodations, and the friendliness of the staff, indicating that these elements significantly contribute to guest satisfaction. On the other hand, occasional negative reviews focus on service response time and facility maintenance issues, suggesting areas where improvement could enhance the overall guest experience. This analysis not only uncovers the primary factors influencing guest sentiment but also allows for the identification of specific strengths and weaknesses in service delivery. By addressing these key areas, Local Collection Hotel can refine its operational strategies to align more closely with customer expectations, fostering higher satisfaction levels and encouraging repeat patronage. Such data-driven insights are essential for achieving sustainable competitive advantage in the hospitality industry.



	Cave Room	Deluxe	King Suite	Sea View	Standard	Suite	Superior	Two Bedroom Villa
Solo traveler		3		1	5		1	
Group		5			7	1	2	1
Family with young children		3		1	3	1		
Family with teens		2		1	4	1		
Couple	3	10	1	1	14	2	5	
Cave Room			1				1	
Business traveler					1		1	

Figure 9. Guest and Room Type (82 Account)

Figure 9 illustrates the distribution of guest types across different room categories based on data from 82 accounts at the Local Collection Hotel. The Standard room is the most popular choice, accommodating 14 solo travelers, seven couples, and three families with young children, indicating its broad appeal across various guest segments. The Deluxe and Sea View rooms, which house 10 and 8 guests, respectively, are predominantly selected by families and groups, suggesting that these categories are preferred for their enhanced amenities and spaciousness. The King Suite, with five bookings, shows a higher preference among business travelers, likely due to its conducive environment for work and relaxation. Interestingly, with only three bookings, the Cave Room caters to a niche market due to its unique design that appeals to guests seeking a distinctive experience. This distribution analysis reveals the alignment between guest profiles and room preferences. It provides actionable insights for optimizing room offerings and tailoring marketing strategies to target specific guest demographics more effectively. Such targeted strategies can enhance the overall guest experience, ensuring that each room type caters precisely to the expectations and needs of its intended audience.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

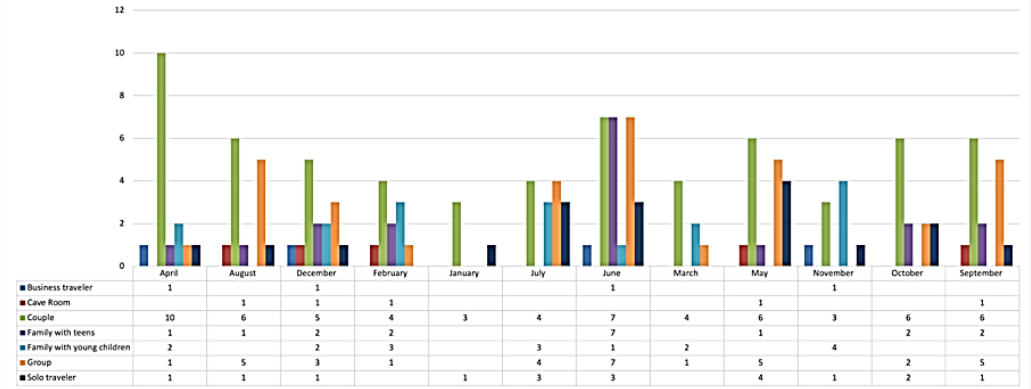


Figure 10. Length of Stay based on Guest Type Classification (160 Accounts)

Figure 10 displays the distribution of the length of stay across different guest types based on data from 160 accounts at the Local Collection Hotel. The graph indicates significant variation in the duration of stay among guest categories, with couples and solo travelers showing a higher frequency of shorter stays, mainly concentrated in April, June, and December. This trend suggests that these segments primarily include leisure travelers seeking quick getaways. In contrast, family groups and business travelers tend to have more extended stays, often exceeding four days, with a noticeable increase in bookings during the mid-year period, particularly in June and July, which might coincide with vacation seasons and business events. The data also reveals that group travelers demonstrate a balanced distribution throughout the year, reflecting diverse travel purposes and flexibility in travel timing. This analysis underscores the importance of aligning hotel services and promotional strategies with each guest type's unique preferences and needs, optimizing offerings for short-stay leisure travelers while ensuring comprehensive facilities and services for long-stay guests such as families and business travelers. The hotel can enhance its capacity management and marketing efforts by understanding these patterns, ultimately improving guest satisfaction and overall business performance.

Table 1. Country of Origin and Guest Type (157 Account)

Country	Business traveler	Couple	Family with teens	Family with young children	Group	Solo traveler
Australia		3	2	1	2	1
China						1
France						1
Germany			1			
India					1	
Indonesia	4	42	12	15	24	13
Italy						1
Malaysia		5	1		3	1
Mexico				1	1	
Nigeria		1				
Philippines					1	
Portugal		1				
Singapore		4	2		2	
South Africa		1				
Spain		1				
Taiwan		1				
Thailand		1				
United Arab Emirates		1				
United Kingdom		1				
United States		1			1	
Uruguay		1				
Vietnam		1				

Table 1 provides an overview of the distribution of guest types at Loccal Collection Hotel based on the country of origin, reflecting data from 157 accounts. The table reveals a high concentration of Indonesian guests across various categories, including 42 couples, 12 families with teens, 15 with young children, 24 groups, and 13 with solo travelers. This dominance indicates the hotel's strong appeal to domestic travelers, potentially due to its location and familiarity. Malaysia and Australia follow, with a noticeable presence of couples and groups, suggesting that the hotel is also a popular destination for international tourists from neighboring countries.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

Interestingly, the business traveler segment is more diverse, with guests from countries such as Indonesia, Germany, South Africa, and Singapore, pointing to the hotel's ability to attract professionals from various regions. The varied guest demographics across countries highlight Loccal Collection Hotel's capability to cater to a diverse clientele, each with distinct travel purposes and preferences. Such data can be strategically used to tailor marketing efforts and service offerings according to the preferences of dominant nationalities and guest segments, further enhancing the hotel's appeal and customer satisfaction.

The findings from the customer experience analysis for the Loccal Collection Hotel reveal significant insights into guest satisfaction and preferences based on various criteria. The data suggests that room quality, service efficiency, and overall ambiance influence positive guest sentiment. In contrast, occasional issues related to service response and facility maintenance highlight areas for potential improvement. Figure 9 illustrates that the Standard room type, accommodating the highest number of solo travelers and couples, is the most popular choice among the 82 accounts analyzed, indicating its widespread appeal across diverse guest segments. The Deluxe and Sea View rooms are favored by family groups, reflecting their suitability for guests seeking more spacious and amenity-rich accommodations. Figure 10 demonstrates the variation in length of stay, with shorter stays preferred by leisure travelers. At the same time, families and business guests tend to book longer durations, particularly in peak seasons like June and December. Table 1 underscores the hotel's domestic solid appeal, with most guests originating from Indonesia, while Malaysia and Australia represent significant portions of international visitors. The comprehensive understanding of these patterns enables the formulation of targeted strategies to enhance service offerings, optimize room configurations, and tailor promotional efforts based on the preferences of each guest demographic. This approach ultimately supports the hotel in refining its operational strategies to achieve higher customer satisfaction and strengthen its position as a preferred destination for diverse traveler types.

3.2 Topic Clustering and Sentiment Classification Performance: LDA and k-NN enhanced by SMOTE

The performance of topic clustering and sentiment classification in this study is evaluated using Latent Dirichlet Allocation (LDA) and k-nearest Neighbors (k-NN) algorithms, with the latter being enhanced by the Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance issues. LDA effectively identifies thematic structures within the text data by categorizing words into topics based on their probabilistic distributions, revealing underlying patterns and common themes customers discuss. Meanwhile, the k-NN algorithm is utilized to classify the reviews into positive and negative sentiments, with SMOTE employed to balance the dataset and improve the classifier's performance. Including SMOTE ensures that majority and minority classes are represented proportionately, thus accurately enhancing the model's ability to differentiate between sentiment categories. The evaluation metrics, such as accuracy, precision, and recall, demonstrate the reliability of the k-NN model, while coherence scores validate the LDA's capability to generate meaningful topics. This integrated approach leverages the strengths of both algorithms, providing a comprehensive understanding of customer feedback and enabling more informed decision-making for service improvement. Such a methodology maximizes the interpretability of thematic patterns and ensures robust sentiment classification, ultimately contributing to a deeper and more accurate analysis of consumer perceptions.

The analysis of topic clustering using Latent Dirichlet Allocation (LDA) is evaluated through several performance metrics, such as log-likelihood, perplexity, and coherence scores, which provide insight into the quality and interpretability of the generated topics. The log-likelihood value of -54886.092 suggests the model's overall fit, while the perplexity score of 665.673 indicates the model's ability to predict word distribution within the documents. Lower perplexity values generally signify better performance, reflecting the model's capability to capture the semantic relationships between words. Although negative, the average coherence score of -14.949 is still helpful for comparing topic coherence to other models, as higher absolute coherence values indicate more interpretable topics. The entropy and exclusivity metrics, with average values of 3.009 and 0.656, further demonstrate the corpus's distinctiveness and separation of topics. A high exclusivity score means that the terms within each topic are more unique and less shared across other topics, enhancing the clarity of each topic cluster. These findings suggest that the LDA model effectively categorizes textual data into coherent themes, enabling a more profound exploration of underlying patterns and insights within the dataset, which can be instrumental in guiding strategic decision-making and refining content analysis.

The performance analysis of the k-Nearest Neighbors (k-NN) algorithm, enhanced by the Synthetic Minority Over-sampling Technique (SMOTE) for sentiment classification, demonstrates its effectiveness in distinguishing between positive and negative sentiments within the dataset. The model achieves an overall accuracy of 91.43%, indicating high precision in its predictions. The area under the curve (AUC) scores, including optimistic (0.989), realistic (0.947), and pessimistic (0.904) estimates, reflecting the model's discriminatory solid capability across various thresholds, which signifies its robustness in differentiating between sentiment classes. Moreover, the precision of 97.26% and recall of 85.36% suggest that the model is highly effective in correctly identifying positive sentiment, with minimal false positives. The F1-score of 90.83% further supports the model's balanced performance in terms of precision and recall, demonstrating its reliability in handling both majority and minority classes. The application of SMOTE proves instrumental in addressing class imbalance, as evidenced by the enhanced detection rates for the positive sentiment category. This

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

comprehensive evaluation validates the suitability of the k-NN model, in conjunction with SMOTE, for sentiment analysis tasks, making it a valuable tool for understanding customer feedback and facilitating data-driven decision-making in service quality assessments.

The performance of the k-Nearest Neighbors (k-NN) algorithm without applying SMOTE reveals a relatively high overall accuracy of 91.49%. Still, there are notable limitations when dealing with imbalanced class distributions. The model achieves a micro-averaged precision of 92.95% and recall of 98.22%, indicating strong performance in identifying positive class instances. However, the Area Under the Curve (AUC) metrics, with an optimistic AUC of 0.858, a realistic AUC of 0.580, and a pessimistic AUC of 0.303, suggest that the model's ability to differentiate between positive and negative classes diminishes significantly under varying conditions. This inconsistency in AUC values points to challenges in maintaining stable classification boundaries, particularly for the minority (hostile) class, as reflected by the confusion matrix showing a high misclassification rate for negative samples (5 true negatives vs. five false negatives). While the f-measure of 95.47% highlights the overall balance between precision and recall, the disparity in AUC metrics indicates that the model struggles with generalization across different thresholds. Such performance suggests that although the k-NN algorithm can effectively handle the majority class, it fails to classify minority instances accurately. It emphasizes the need for techniques like SMOTE to enhance its robustness and achieve a more balanced classification outcome.

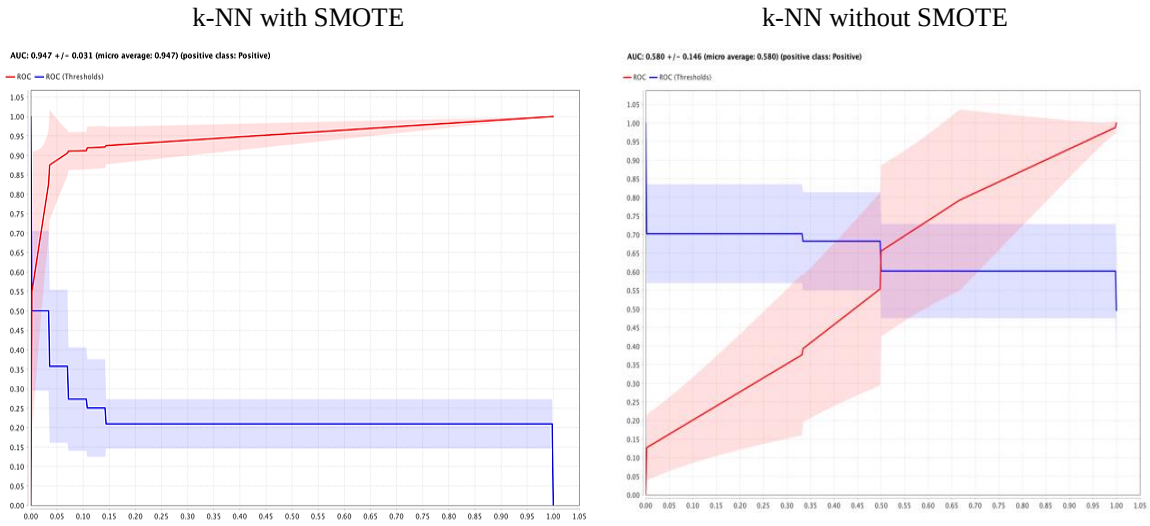


Figure 11. Area Under Curve of k-NN with and without SMOTE

Figure 11 illustrates the comparison of the Area Under the Curve (AUC) performance for the k-nearest Neighbors (k-NN) algorithm with and without the application of the Synthetic Minority Over-sampling Technique (SMOTE). The graph on the left depicts the AUC for k-NN without SMOTE, where the curve demonstrates lower discriminatory capability, particularly for the minority class, as indicated by a steeper drop and more significant variance. This suggests that the model struggles to maintain consistent classification performance due to the imbalanced distribution of classes within the dataset. In contrast, the graph on the right shows the AUC after implementing SMOTE, where the curve exhibits a significantly improved trajectory with a smoother and more stable pattern, indicating enhanced classification accuracy for both majority and minority classes. The increase in the AUC value, combined with the reduced variance, reflects the effectiveness of SMOTE in mitigating the class imbalance issue, thereby enabling the k-NN model to achieve more reliable and robust performance. This comparison underscores the importance of employing resampling techniques like SMOTE to optimize model accuracy and generalization, particularly in skewed class distributions.

The comprehensive evaluation of topic clustering and sentiment classification using LDA and k-NN algorithms, with the latter enhanced by SMOTE, indicates a high degree of effectiveness in categorizing textual data and accurately classifying customer sentiments. The LDA model successfully identifies thematic structures within the text, as evidenced by metrics such as log-likelihood and coherence scores, highlighting the generated topics' clarity and interpretability. Meanwhile, the k-NN algorithm, particularly when enhanced by SMOTE, achieves an overall accuracy of 91.43% in sentiment classification, demonstrating strong performance distinguishing between positive and negative sentiments. The integration of SMOTE addresses class imbalance by balancing the distribution of positive and negative instances, as reflected in the significant improvement in AUC scores and a reduction in variance. This adjustment enhances the model's robustness, allowing for more reliable predictions across diverse thresholds.

In contrast, the k-NN model without SMOTE, although exhibiting a high accuracy of 91.49%, struggles with maintaining stable classification boundaries, particularly for the minority class, as indicated by the lower

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

AUC scores and higher misclassification rates. These findings suggest that employing SMOTE improves the model's overall performance and ensures more consistent and balanced classification outcomes. Consequently, the combined use of LDA for topic clustering and SMOTE-enhanced k-NN for sentiment classification provides a comprehensive analytical framework that effectively captures the complexity of customer feedback, thereby facilitating data-driven decision-making for service quality enhancement.

3.3 Discussion: Strategic Improvement for Hotel Customer Experience

The strategic improvement of hotel customer experience can be effectively guided by insights derived from topic clustering, as this method identifies key themes that influence guest satisfaction. Management can pinpoint specific areas that require attention by categorizing reviews into distinct clusters, such as service quality, room amenities, cleanliness, and staff responsiveness [30]. For example, if a cluster reveals frequent mentions of delays in service delivery, targeted strategies such as staff training or process optimization can be implemented to enhance service efficiency. Similarly, resource allocation for facility upgrades or regular inspections can be prioritized if a cluster indicates dissatisfaction with room maintenance. The ability to discern positive and negative feedback patterns within these clusters allows for a balanced approach to decision-making, enabling the hotel to reinforce its strengths while addressing weaknesses [31]. This data-driven strategy improves operational effectiveness and aligns the service offerings with customer expectations, ultimately fostering a more personalized and immersive guest experience. Such an approach ensures that improvement initiatives are not based on isolated incidents but are instead reflective of broader trends, thereby maximizing the impact of strategic decisions on overall customer satisfaction and loyalty.

Strategic improvement for hotel customer experience can be effectively achieved through sentiment classification, which offers a nuanced understanding of guest emotions and satisfaction levels. By classifying reviews into positive, negative, or neutral sentiments, management gains valuable insights into which aspects of the hotel are perceived favorably and which are sources of dissatisfaction [32]. For instance, if positive sentiments predominantly highlight the friendliness and professionalism of the staff, these strengths can be further leveraged in marketing communications to enhance brand positioning. Conversely, if negative sentiments frequently mention service response times or room cleanliness issues, targeted interventions such as process reengineering and housekeeping protocol adjustments can be prioritized [33]. Sentiment classification also facilitates trend analysis over time, allowing management to monitor the effectiveness of implemented strategies and adjust them accordingly. This approach enables hotels to move beyond a reactive stance, proactively addressing recurring concerns and refining the guest experience. Ultimately, leveraging sentiment classification supports a data-driven, customer-centric strategy that addresses immediate concerns and fosters long-term improvements, enhancing guest satisfaction and encouraging repeat business.

Strategic customer experience improvement at the Loccal Collection Hotel in Labuan Bajo can be achieved by focusing on insights derived from guest feedback and operational analysis. A review of guest sentiments reveals critical strengths in aesthetic appeal and staff friendliness. This suggests that these aspects should be continuously emphasized to differentiate the hotel in a competitive market [34]. However, recurring concerns related to service response times and room maintenance indicate opportunities for operational enhancements. A more structured training program for staff to ensure prompt and efficient service, combined with periodic maintenance checks, would address these issues and elevate service quality [35].

Additionally, investing in digital tools such as guest feedback platforms can facilitate real-time monitoring of guest satisfaction and enable immediate corrective actions. Enhancing amenities that guests frequently highlight, such as upgrading room facilities and introducing more personalized services, would contribute to a positive guest experience. These targeted strategies, informed by specific guest experiences, improve operational efficiency and align service delivery with customer expectations, fostering higher guest satisfaction and loyalty.

Further strategic improvement for the Loccal Collection Hotel in Labuan Bajo should focus on enhancing personalized guest experiences and leveraging technological innovations to optimize service delivery. One recommendation is to implement a data-driven approach for understanding guest preferences, utilizing advanced customer relationship management (CRM) systems that track individual preferences and tailor services accordingly, such as customized room setups or personalized dining experiences. Additionally, incorporating real-time digital feedback mechanisms would enable the hotel to address issues as they arise, thereby improving service response times and ensuring immediate guest satisfaction. Another critical area for improvement is expanding eco-friendly initiatives, which could appeal to environmentally conscious travelers, such as introducing sustainable in-room amenities and energy-efficient practices. Regularly upgrading facilities, especially those related to room maintenance and common areas, would also enhance guest comfort and overall experience. Lastly, fostering partnerships with local tourism operators to offer exclusive packages or curated experiences would further elevate the hotel's value proposition, positioning it as a preferred destination in Labuan Bajo. When implemented cohesively, these strategies will increase operational efficiency, deepen guest loyalty, and strengthen the hotel's competitive edge in a growing tourism market.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

4. CONCLUSION

The research findings, utilizing Knowledge Discovery in Databases (KDD) alongside the performance evaluation of Latent Dirichlet Allocation (LDA) and k-Nearest Neighbors (k-NN), indicate a robust framework for extracting and classifying insights from 388 customer reviews. The LDA performance vector metrics, including a log-likelihood of -54,886.092 and an average coherence score of -14.949, demonstrate the model's efficacy in capturing coherent thematic structures within the dataset, thereby enabling a deeper understanding of customer feedback patterns. Meanwhile, the k-NN performance vector, both with and without applying the Synthetic Minority Over-sampling Technique (SMOTE), highlights the model's strengths and limitations in sentiment classification. The incorporation of SMOTE proved instrumental in balancing class distribution, significantly improving the classification accuracy to 91.43% and AUC scores ranging from 0.904 (pessimistic) to 0.989 (optimistic), which reflect the model's enhanced ability to distinguish between positive and negative sentiments. Integrating KDD methodologies with advanced text mining and classification algorithms, this comprehensive approach effectively translates complex textual data into actionable insights. The results validate the applicability of combining LDA and k-NN for analyzing large-scale customer reviews, providing a solid foundation for strategic decision-making and service quality optimization. Consequently, this research demonstrates the effectiveness of these analytical techniques and underscores their potential to support data-driven enhancements in hospitality management and customer experience.

REFERENCES

[1] G. D. Mendonça, S. R. de M. Oliveira, O. F. Lima, and P. T. V. de Resende, "Intelligent algorithms applied to the prediction of air freight transportation delays," *Int. J. Phys. Distrib. Logist. Manag.*, vol. 54, no. 1, pp. 61–91, Jan. 2024, doi: 10.1108/IJPDLM-10-2022-0328.

[2] O. A. George and C. M. Q. Ramos, "Sentiment analysis applied to tourism: exploring tourist-generated content in the case of a wellness tourism destination," *Int. J. Spa Wellness*, vol. 7, no. 2, pp. 139–161, 2024, doi: 10.1080/24721735.2024.2352979.

[3] N. Amat-Lefort, F. Barravecchia, and L. Mastrogiacomo, "Quality 4.0: big data analytics to explore service quality attributes and their relation to user sentiment in Airbnb reviews," *Int. J. Qual. Reliab. Manag.*, vol. 40, no. 4, pp. 990–1008, Jan. 2023, doi: 10.1108/IJQRM-01-2022-0024.

[4] İ. A. Özen and E. Özgül Katlav, "Aspect-based sentiment analysis on online customer reviews: a case study of technology-supported hotels," *J. Hosp. Tour. Technol.*, vol. 14, no. 2, pp. 102–120, Jan. 2023, doi: 10.1108/JHTT-12-2020-0319.

[5] S. Bagherzadeh, S. Shokouhyar, H. Jahani, and M. Sigala, "A generalizable sentiment analysis method for creating a hotel dictionary: using big data on TripAdvisor hotel reviews," *J. Hosp. Tour. Technol.*, vol. 12, no. 2, pp. 210–238, Jan. 2021, doi: 10.1108/JHTT-02-2020-0034.

[6] M. Gao, J. Wang, and O. Liu, "Is UGC sentiment helpful for recommendation? An application of sentiment-based recommendation model," *Ind. Manag. Data Syst.*, vol. 124, no. 4, pp. 1356–1384, Jan. 2024, doi: 10.1108/IMDS-05-2023-0335.

[7] G. Rasool and A. Pathania, "Reading between the lines: untwining online user-generated content using sentiment analysis," *J. Res. Interact. Mark.*, vol. 15, no. 3, pp. 401–418, Jan. 2021, doi: 10.1108/JRIM-03-2020-0045.

[8] R. C. Ho, M. S. Withanage, and K. W. Khong, "Sentiment drivers of hotel customers: a hybrid approach using unstructured data from online reviews," *Asia-Pacific J. Bus. Adm.*, vol. 12, no. 3–4, pp. 237–250, Jan. 2020, doi: 10.1108/APJBA-09-2019-0192.

[9] H. M. Zolbanin and D. Wynn, "From star rating to sentiment rating: using textual content of online reviews to develop more effective reputation systems for peer-to-peer accommodation platforms," *J. Bus. Anal.*, vol. 6, no. 2, pp. 127–139, Apr. 2023, doi: 10.1080/2573234X.2022.2122880.

[10] C. Kaveski Peres and E. Pacheco Paladini, "Exploring the attributes of hotel service quality in Florianópolis-SC, Brazil: An analysis of tripAdvisor reviews," *Cogent Bus. Manag.*, vol. 8, no. 1, p. 1926211, Jan. 2021, doi: 10.1080/23311975.2021.1926211.

[11] F. Leal, B. Malheiro, B. Veloso, and J. C. Burguillo, "Responsible processing of crowdsourced tourism data," *J. Sustain. Tour.*, vol. 29, no. 5, pp. 1–21, 2020, doi: 10.1080/09669582.2020.1778011.

[12] A. Arabameri *et al.*, "Flood susceptibility mapping using meta-heuristic algorithms," *Geomatics, Nat. Hazards Risk*, vol. 13, no. 1, pp. 949–974, 2022, doi: 10.1080/19475705.2022.2060138.

[13] Z. Z. Zarezadeh, R. Rastegar, and Z. Xiang, "Big data analytics and hotel guest experience: a critical analysis of the literature," *Int. J. Contemp. Hosp. Manag.*, vol. 34, no. 6, pp. 2320–2336, 2022, doi: 10.1108/IJCHM-10-2021-1293.

[14] R. Iloranta and R. Komppula, "Service providers' perspective on the luxury tourist experience as a product," *Scand. J. Hosp. Tour.*, vol. 22, no. 1, pp. 39–57, 2022, doi: 10.1080/15022250.2021.1946845.

[15] N. N. Quang and D. C. Thuy, "Mindfulness affecting loyalty with mediating role of customer experience in the context of adventure tourism in Vietnam," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi:

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 48-63
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6014

10.1080/23311886.2024.2312651.
[16] S. Girija, D. R. Sharma, and V. Kaushal, "Exploring dimensions of the customer experience at budget hotels during the COVID-19 pandemic: a netnography approach," *Qual. Mark. Res.*, vol. 26, no. 4, pp. 320–344, Jan. 2023, doi: 10.1108/QMR-03-2022-0039.
[17] C. H. Lee, Q. Li, Y. C. Lee, and C. W. Shih, "Service design for intelligent exhibition guidance service based on dynamic customer experience," *Ind. Manag. Data Syst.*, vol. 121, no. 6, pp. 1237–1267, Jan. 2020, doi: 10.1108/IMDS-06-2020-0356.
[18] M. S. Viñán-Ludeña and L. M. de Campos, "Analyzing tourist data on Twitter: a case study in the province of Granada at Spain," *J. Hosp. Tour. Insights*, vol. 5, no. 2, pp. 435–464, Jan. 2022, doi: 10.1108/JHTI-11-2020-0209.
[19] T. Falatouri, P. Brandtner, M. Nasseri, and F. Darbanian, "Service quality dimensions in Austrian food retailing–a text mining approach for physical retail stores," *Int. Rev. Retail. Distrib. Consum. Res.*, vol. 00, no. 00, pp. 1–36, 2024, doi: 10.1080/09593969.2024.2371456.
[20] Y. Wu, J. Wang, Y. Xia, Q. Li, and Y. Pan, "Sensing hotel customers distribution and their sentiment variations using online travel agent data: a case of Shanghai star-rated hotels," *Ann. GIS*, vol. 30, no. 3, pp. 323–343, 2024, doi: 10.1080/19475683.2024.2335976.
[21] M. Mariani and M. Borghi, "Environmental discourse in hotel online reviews: a big data analysis," *J. Sustain. Tour.*, vol. 29, no. 5, pp. 829–848, 2020, doi: 10.1080/09669582.2020.1858303.
[22] T. Albayrak, A. Dursun-Cengizci, L. H. N. Fong, and M. Caber, "The changing role of hotel attributes in destination competitiveness throughout a crisis," *Int. J. Contemp. Hosp. Manag.*, vol. 36, no. 10, pp. 3264–3282, Jan. 2024, doi: 10.1108/IJCHM-06-2023-0779.
[23] R. Rahimi, M. Thelwall, F. Okumus, and A. Bilgihan, "Know your guests' preferences before they arrive at your hotel: evidence from TripAdvisor," *Consum. Behav. Tour. Hosp.*, vol. 17, no. 1, pp. 89–106, Jan. 2022, doi: 10.1108/CBTH-06-2021-0148.
[24] H. Xu, L. T. O. Cheung, J. Lovett, X. Duan, Q. Pei, and D. Liang, "Understanding the influence of user-generated content on tourist loyalty behavior in a cultural World Heritage Site," *Tour. Recreat. Res.*, vol. 48, no. 2, pp. 173–187, 2023, doi: 10.1080/02508281.2021.1913022.
[25] T. D. Quang, N. M. P. Tran, E. Sthapit, and B. Garrod, "Exploring Guests' Satisfaction and Dissatisfaction with Homestay Experiences: A Netnographic Study of a Rural Tourism Destination in Vietnam," *Int. J. Hosp. Tour. Adm.*, vol. 00, no. 00, pp. 1–25, 2024, doi: 10.1080/15256480.2024.2350005.
[26] D. D'Acunto, S. Volo, and R. Filieri, "'Most Americans like their privacy.' Exploring privacy concerns through US guests' reviews," *Int. J. Contemp. Hosp. Manag.*, vol. 33, no. 8, pp. 2773–2798, Jan. 2021, doi: 10.1108/IJCHM-11-2020-1329.
[27] F. Hu, R. Trivedi, and T. Teichert, "Using hotel reviews to assess hotel frontline employees' roles and performances," *Int. J. Contemp. Hosp. Manag.*, vol. 34, no. 5, pp. 1796–1822, Jan. 2022, doi: 10.1108/IJCHM-04-2021-0491.
[28] M. J. Sánchez-Franco and S. Rey-Tienda, "The role of user-generated content in tourism decision-making: an exemplary study of Andalusia, Spain," *Manag. Decis.*, vol. 62, no. 7, pp. 2292–2328, Jan. 2024, doi: 10.1108/MD-06-2023-0966.
[29] "User-generated online content and hospitality firms: Identifying appropriate response strategies," *Strateg. Dir.*, vol. 36, no. 9, pp. 49–52, Jan. 2020, doi: 10.1108/SD-07-2020-0131.
[30] V. O. Olorunsola, M. B. Saydam, T. T. Lasisi, and K. K. Eluwole, "Customer experience management in capsule hotels: a content analysis of guest online review," *J. Hosp. Tour. Insights*, vol. 6, no. 5, pp. 2462–2483, 2023, doi: 10.1108/JHTI-03-2022-0113.
[31] S. Bharwani and D. Mathews, "Techno-business strategies for enhancing guest experience in luxury hotels: a managerial perspective," *Worldw. Hosp. Tour. Themes*, vol. 13, no. 2, pp. 168–185, 2021, doi: 10.1108/WHATT-09-2020-0121.
[32] A. Yucel, M. Caglar, H. Ahady Dolatsara, B. George, and A. Dag, "Predicting hotel reviews from sentiment: a multinomial classification framework," *J. Model. Manag.*, vol. 17, no. 2, pp. 697–714, Jan. 2022, doi: 10.1108/JM2-09-2020-0255.
[33] A. Tanrısevdi, G. Öztürk, and A. C. Öztürk, "A supervised data mining approach for predicting comment card ratings," *Int. J. Contemp. Hosp. Manag.*, vol. 34, no. 5, pp. 1823–1853, 2022, doi: 10.1108/IJCHM-05-2021-0675.
[34] A. Kayumov, Y. joo Ahn, K. Kiatkawsin, I. Sutherland, and S. Zielinski, "Service quality and customer loyalty in halal ethnic restaurants amid the COVID-19 pandemic: a study of halal Uzbekistan restaurants in South Korea," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2024.2301814.
[35] M. V. Ciasullo, R. Montera, and R. Palumbo, "Online content responsiveness strategies in the hospitality context: exploratory insights and a research agenda," *TQM J.*, vol. 36, no. 9, pp. 234–254, Jan. 2020, doi: 10.1108/TQM-12-2019-0299.