Sentiment Analysis of Customer Feedback Reviews Towards Hotel's Products and Services in Labuan Bajo

Henoch Juli Christanto[1], Yerik Afrianto Singgalen[1*]

[1]Information System Department, Faculty of Engineering, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
[1*]Tourism Department, Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: [1]henoch.christanto@atmajaya.ac.id, [1*]yerik.afrianto@atmajaya.ac.id

Abstract

The feedback review column on the TripAdvisor website can process the sentiment analysis by classifying the positive and negative impressions of the consumer using machine learning, namely text mining. This study analyzes consumer sentiment for hotel products and services in Labuan Bajo based on feedback review data on the TripAdvisor website. Meanwhile, the algorithm used is Nave Bayes Classifier (NBC), Support Vector Machine (SVM), and k-Nearest Neighbor (k-NN) algorithm. In addition, SMOTE Upsampling is used as technic to balance the dataset calculated by k-Nearest Neighbor (k-NN). The classification result shows 702 negative feedback reviews and 2531 positive feedback reviews. The evaluation of algorithm performance shows the accuracy of SVM is 78,30% and NBC is 78,29% compared with k-NN with 85,24% accuracy, using SMOTE Upsampling with a class precision description of 100% prediction negative (1784 true Neg & 0 true Pos) and 77,21% prediction positive (747 true Neg & 2531 true Pos). The class recall description also shows 70,49% true Negative (1784 pred Neg & 747 pred Pos) and 100% true Positive (0 pred Neg & 2531 pred Pos). These findings indicate that the k-NN algorithm shows the best result instead of the SVM and NBC algorithm, according to the sentiment analysis result of customer feedback reviews on hotel products and services in Labuan Bajo through the TripAdvisor website.

Keywords: Sentiment Analysis, Customer, Hotel, Product, Services

1. INTRODUCTION

Hotel consumers have expectations of products and services balanced with the value paid [1]. [2] reported that product prices influence consumer expectations, especially related to product quality. In addition, [3] shows that consumers have high expectations of the product's value. It indicates that consumer preferences

for the value of a product or service will affect behavior and purchasing decisions [4]. Several previous researchers have shown that consumer expectations and satisfaction can be classified into sentiment and serve as recommendations for business development [5]. Thus, the study of consumer sentiment analysis needs to be improved to improve the quality of products and services.

This study aims to classify consumer sentiment toward hotels, resorts, and restaurants in Indonesia's super-priority tourism destination Labuan Bajo. The initiation of sentiment analysis studies at Indonesian Super Priority Tourist Destinations still needs to be optimized to be utilized by stakeholders to increase the quantity and quality of the hotel, resort, and restaurant business, and other tourism supporting facilities. The business sentiment analysis of hotels, resorts, and restaurants is used to analyze consumer preferences related to the room's characteristics, food, location, services, cleanliness, price, and entertainment [6]. Hotel managers can improve business management strategies to meet consumer expectations [7]. Several studies on sentiment analysis show that the review column on hotel websites to accommodation service reservation agents is used to identify and analyze consumers' positive-negative sentiments towards facilities, reservation systems, locations, and services [8], [9]. In addition, the study of sentiment analysis on the business of hotels, resorts, and restaurants in Indonesia, especially the super-priority tourist destination of Labuan Bajo, needs to be discussed from an empirical and theoretical perspective.

Sentiment analysis using a text mining approach in the feedback review column is essential to identify consumer perceptions as end-users on the existing conditions of products and services provided [10]. In addition, the discourse environment for hotel consumer reviews on various review platforms was reviewed by considering subjectivity, diversity, length, sentiment, and readability [11]. Each hotel has different management policies, including establishing a reservation service system for consumers [12] and human-like robot employees in the hotel industry [13]. The development of studies on sentiment analysis is not only carried out to analyze consumer perceptions. [14] also analyzed the sentiments of hotel employees related to the management system adopted by each hotel. [15] Big data analysis or text mining will provide better insight into hotel performance and guest satisfaction. It indicates the importance of the sentiment analysis of hotel consumers to evaluate the quality of products and services.

Sentiment analysis of hotel products and services in Indonesia needs to be explored in-depth, given the Covid-19 pandemic, which urges hotel management to adjust changes in the implementation of hotel guest services based on health protocols. Managing hotel accommodation services during the Covid-19 pandemic has drawn several criticisms from consumers regarding staff, service, room, cleanliness, slow booking, and hotel pandemic response [16]. The same thing is also shown by [17] that consumer confidence in the products and services

offered by the hotel during the Covid-19 pandemic guarantees health and safety through social distancing policies and optimization of cleaning services. On the other hand, [18] shows four critical dimensions in hotel management during the Covid-19 pandemic: medical preparedness, hygiene control, health communication, and self-service technology. It indicates that analyzing hotel consumer sentiment towards products and services is essential in maintaining the hotel business during the Covid-19 pandemic.

This article offers insight to discuss empirical data on consumer sentiment toward the products and services of accommodation service providers in Labuan Bajo as Indonesia's super-premium destinations. The qualification standard for the hotel and resort review dataset is limited to only ten hotels and resorts with more than 100 reviews. The data includes consumer review data on the TripAdvisor platform until May 1, 2022. Thus, the analyzed hotels and resorts are Puri Sari Beach Hotel, Plataran Komodo Resort and Spa, Bintang Flores Hotel, Ayana Komodo Resort, Sylvia Resort Komodo, Bayview Garden Hotel, Golo Hiltop Hotel and Restaurant, Laprima Hotel, Sudamala Resort Seraya, Eco Tree O'tel. Meanwhile, the source of review data comes from the TripAdvisor website. Consumer reviews of hotel and resort products and services will be classified based on positive and negative sentiments using the Naïve Bayes Classification (NBC) Support Vector Machine (SVM) algorithm, k-Nearest Neighbor (k-NN). Furthermore, the algorithm performance is evaluated based on the highest level of accuracy as a recommendation for classification methods relevant to data on consumer reviews of hotel products and services in Labuan Bajo. In addition, the consumer reviews related to employees, services, rooms, cleanliness, and hotel reservation systems are discussed with a theoretical perspective that relates to consumer expectations and satisfaction while consuming the products and services of accommodation service providers in Labuan Bajo.

2. METHODS

This study uses machine learning methodology [19] as a computational technique based on the Support Vector Machine (SVM), k-Nearest Neighbor (k-NN), and Naïve Bayes Classification (NBC) algorithms. Previous studies have shown that machine learning is a relevant approach in text mining [20]. In addition, one of the essential processes in the text mining approach is the implementation of classification methods based on negative and positive sentiments. This research consists of three stages: pre-processing dataset, processing dataset, and evaluation algorithm performance. The pre-processing stage selects and determines hotels and resorts on the TripAdvisor website, as shown in figure 1 below.
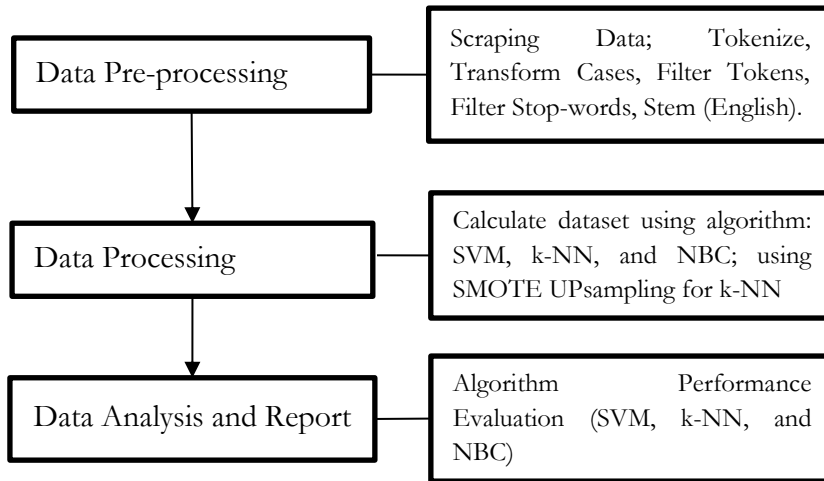
Figure 1. The flow of the Analysis Sentiment Process

The hotels and resorts dataset qualifications are limited to ten hotels or resorts with more than 100 reviews on TripAdvisor. The data includes consumer review data on the TripAdvisor platform until May 1, 2022. Thus, the dataset is selected based on the qualification standard. This research uses the dataset of consumer reviews toward products and services in hotels and resorts, such as Puri Sari Beach Hotel; Plataran Komodo Resort and Spa; Bintang Flores Hotel; Ayana Komodo Resort; Sylvia Resort Komodo; Bayview Garden Hotel; Golo Hiltop Hotel and Restaurant; Laprima Hotel; Sudamala Resort Seraya; and Eco Tree O'tel. Otherwise, the dataset crawls using the Web Scraping Software application Webharvy, shown in figure 2 below.
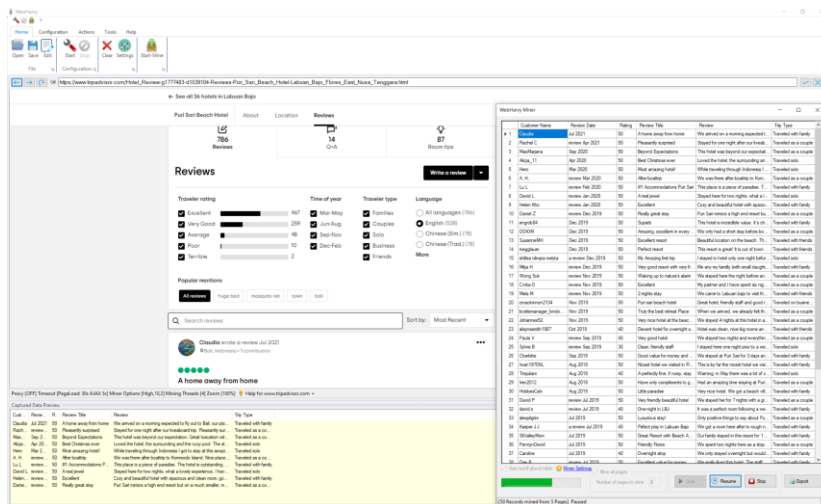

Figure 2. Web Scraping Software

Figure 2 scrapes the TripAdvisor website's review data for each hotel and resort related to location, cleanliness, service, and value. The data scraped were customer name data, review date, travel rating (excellent, very good, average, poor, terrible), time of year (Mar-May, Jun-Aug, Sep-Nov, Dec-Feb), traveler type (families, couples, solo, business, friends), and languages. Meanwhile, the limitation of this research is the review data in English. Specifically, consumer reviews used in the analysis also relate to assessments of property amenities, room features, room types, location, and services. Review data related to the hotel's products and services in Labuan Bajo, representing the quality of super-premium tourism destinations in Indonesia. Positive or negative sentiments characterize customer expectation and satisfaction analysis. Thus, the data processing should be well conducted using text mining tools, as shown in figure 3 below.
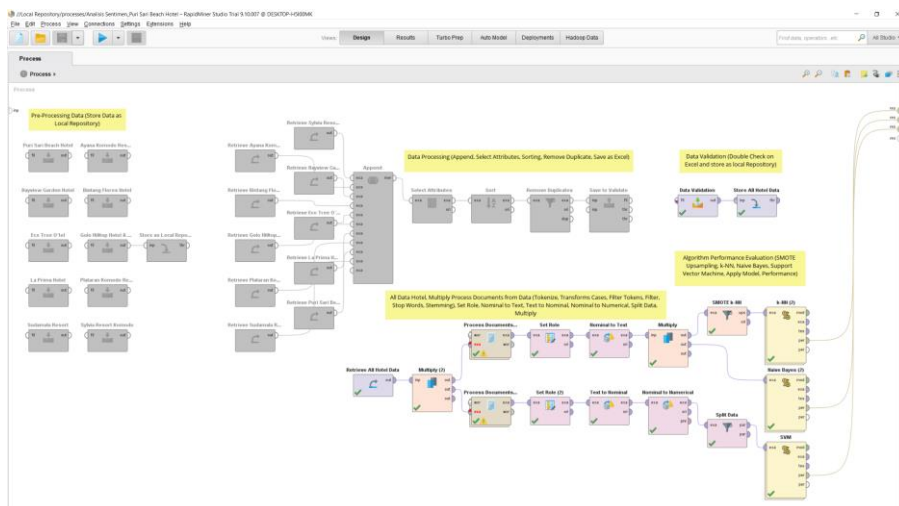

Figure 3 Text Mining Tools

Figure 3 is Rapidminer software developed for machine learning, text mining, and predictive analytics. The data that has been obtained from the website will be processed using text mining operators to clean the data through the tokenize operator (regular expression and non-letters), transform cases, filter tokens (by length), stopwords filter (English), and stem (snowball). In addition, several operators such as select attributes, sort, remove duplicates, and store data in a repository. After all, data is combined and stored in the storage. The following process will classify negative and positive sentiments using the k-NN, NBC, and SVM algorithms and evaluate the performance. In addition, several studies have shown that the Tripadvisor Website is one of the providers of credible review data with a capacity control system, where the syntax on the Tripadvisor Website has set the user detection function as a reviewer and the language used in the review. If the user provides a study in an undetectable language, the system will

automatically give a notification to delete the review. The following is a link from the data source taken for analysis.

Table 1. Link of Data Source on TripAdvisor Website

Hotel & Resort	Link
Puri Sari Beach Hotel	https://www.tripadvisor.com/Hotel_Review-g1777483-d1539104-Reviews-Puri_Sari_Beach_Hotel-Labuan_Bajo_Flores_East_Nusa_Tenggara.html
Plataran Komodo Resort and Spa	https://www.tripadvisor.com/Hotel_Review-g1777483-d6218124-Reviews-Plataran_Komodo_Resort_Spa-Labuan_Bajo_Flores_East_Nusa_Tenggara.html
Bintang Flores Hotel	https://www.tripadvisor.com/Hotel_Review-g1777483-d1130294-Reviews-Bintang_Flores_Hotel-Labuan_Bajo_Flores_East_Nusa_Tenggara.html
Ayana Komodo Resort	https://www.tripadvisor.com/Hotel_Review-g1777483-d14167975-Reviews-AYANA_Komodo_Resort-Labuan_Bajo_Flores_East_Nusa_Tenggara.html
Sylvia Resort Komodo	https://www.tripadvisor.com/Hotel_Review-g1777483-d5122597-Reviews-Sylvia_Resort_Komodo-Labuan_Bajo_Flores_East_Nusa_Tenggara.html
Bayview Garden Hotel	https://www.tripadvisor.com/Hotel_Review-g1777483-d1140452-Reviews-Bayview_Gardens_Hotel-Labuan_Bajo_Flores_East_Nusa_Tenggara.html
Golo Hilltop Hotel and Restaurant	https://www.tripadvisor.com/Hotel_Review-g1777483-d636082-Reviews-Golo_Hilltop_Hotel_Restaurant-Labuan_Bajo_Flores_East_Nusa_Tenggara.html
Laprima Hotel	https://www.tripadvisor.com/Hotel_Review-g1777483-d3158351-Reviews-Laprima_Hotel-Labuan_Bajo_Flores_East_Nusa_Tenggara.html
Sudamala Resort, Seraya	https://www.tripadvisor.com/Hotel_Review-g1777483-d1775280-Reviews-Sudamala_Resort_Seraya-Labuan_Bajo_Flores_East_Nusa_Tenggara.html
Eco Tree O'tel	https://www.tripadvisor.com/Hotel_Review-g1777483-d12410420-Reviews-Eco_Tree_O_tel-Labuan_Bajo_Flores_East_Nusa_Tenggara.html

Table 1 shows the data sources used to proceed with the classification method. The Classification of positive and negative sentiments can be done manually following the rating in the review. If there is a review with a five-star rating, it can be marked with positive sentiment. If there is a review with a one-star rating, it can be marked as a negative sentiment. Determining positive and negative sentiments based on five-star and one-star ratings produces training data. Meanwhile, review data with two-star, three-star, and four-star ratings can be predicted and classified into negative and positive sentiments using the k-NN, NBC, and SVM algorithms. The importance of categorizing five clusters into two

clusters is to confirm consumer expectations and satisfaction with hotel products and services in Labuan Bajo. Thus, hotel management can optimize products and services.

Configuration for data performance testing using operators retrieving data from the repository, multiplying, processing documents from data, then setting role. For testing the k-NN and NBC algorithms, the sentiment column is used as a label and then connected to the nominal text and multiply operators. Specifically, the k-NN SMOTE Upsampling operator uses to balance the dataset. Meanwhile, the cross-validation operator processes the k-NN, NBC algorithm model. In the cross-validation operator, apply model and performance operators are needed to measure accuracy, class precision, and class recall. Furthermore, the SVM's testing process flow uses text to nominal and nominal to numerical operators. Moreover, data split operators are needed for data sharing (0.7 and 0.3) to be tested using the SVM model in operator cross-validation. The sentiment classification process's configuration, the analysis and algorithm performance evaluation, and the relevant algorithms are used in sentiment analysis and the increasing data on consumer reviews of hotels and resorts in Labuan Bajo.

3. RESULTS AND DISCUSSION

Sentiment Classification and Model Performance Evaluation

The consumer reviews of hotels and resorts in Labuan Bajo based on negative and positive sentiments indicate that the aspects most reviewed by consumers are: The strategic location of the hotel; Cleanliness in the hotel environment; The rooms used. The services provided by the hotel and the value of the product and services are in balance with consumer spending. Consumer review data is vital for developing service and product management offered by the hotel [21]. In addition, hotel consumer reviews on various digital media platforms provide an overview of the hotel's brand image and measure the performance of the hotel's marketing strategy [22]. Meanwhile, the marketing strategy of hotel products and services is implemented as a Corporate Social Responsibility program on social media and consumer brand engagement [23]. It indicates that the classification of negative and positive sentiments for hotel consumers has an empirical contribution to developing marketing strategies to increase customer loyalty.

Indonesia has a variety of tourist destinations that attract investors in the hospitality industry, one of which is Labuan Bajo which is being developed as a super-premium tourist destination [24]. One of the Labuan Bajo tourism icons that attract foreign tourists is the existence of Komodo dragons, authentic local community traditions, and all stakeholders' support to develop the Labuan Bajo tourism sector [25]. On the other hand, through the Ministry of Tourism and Creative Economics, the Indonesian government develops tourism resources in

ten selected tourist destinations by relying on sustainability [26]. Thus, a study of consumer reviews of hotels and resorts in Labuan Bajo also provides an overview of the expectations and satisfaction of foreign tourists. Therefore, the output of this research contributes not only empirically but also theoretically to the field of tourism-related information systems, especially studies on machine learning and text mining.

This study uses consumer review data on the TripAdvisor platform to get an insight into the expectations and satisfaction of foreign tourists as consumers for hotel products and services in Labuan Bajo. Meanwhile, the minimum review limit and the number of hotels set in this study are ten hotels with several reviews above 100. The data includes consumer review data on the TripAdvisor platform until May 1, 2022. In the pre-processing stage in each review column, duplicated data, numbers, and expressions need to be processed to produce a dataset ready to be managed. The results of processing the review dataset are prepared to classify based on negative and positive sentiment classification process are 3233, as shown in Table 2 below.

Table 2 Sentiment Classification

Hotel & Resort	Reviews	Positive	Negative
Puri Sari Beach Hotel	528	488	40
Plataran Komodo Resort and Spa	324	306	18
Bintang Flores Hotel	359	246	113
Ayana Komodo Resort	264	196	68
Sylvia Resort Komodo	257	111	146
Bayview Garden Hotel	434	373	61
Golo Hilltop Hotel and Restaurant	393	350	43
Laprima Hotel	307	164	143
Sudamala Resort, Seraya	227	196	31
Eco Tree O'tel	140	101	39
Total	**3233**	**2531**	**702**

Data in table 2 does not refer to the rating system on website, but selective consumer reviews after the scraping process. The data in table 2 shows the difference percentage of 3233 reviews data, as follows: Puri Sari Beach Hotel (16%); Plataran Komodo Resort and Spa (10%); Bintang Flores Hotel (11%); Ayana Komodo Resort (8%); Sylvia Resort Komodo (8%); Bayview Garden Hotel (14%); Golo Hilltop Hotel and Restaurant (12%); Laprima Hotel (10%); Sudamala Resort, Seraya (7%); Eco Tree O'tel (4%). In addition, there are differences in the positive sentiment percentage of 2531 data as follows: Puri Sari Beach Hotel (19%); Plataran Komodo Resort and Spa (12%); Bintang Flores Hotel (10%); Ayana Komodo Resort (8%); Sylvia Resort Komodo (4%); Bayview Garden Hotel (15%); Golo Hilltop Hotel and Restaurant (14%); Laprima Hotel (6%); Sudamala Resort, Seraya (8%); Eco Tree O'tel (4%). Likewise, there are differences in the negative sentiment percentage of 702 data as follows: Puri Sari Beach Hotel (6%);

Plataran Komodo Resort and Spa (3%); Bintang Flores Hotel (16%); Ayana Komodo Resort (10%); Sylvia Resort Komodo (21%); Bayview Garden Hotel (9%); Golo Hilltop Hotel and Restaurant (6%); Laprima Hotel (20%); Sudamala Resort, Seraya (4%); Eco Tree O'tel (5%).

Based on the classification of positive and negative sentiments from the accumulated dataset, there are 2531 positive and 702 negative sentiments. It indicates that the image of Labuan Bajo's super-premium tourist destination for foreign tourists, especially hotel and resort consumers, is still positive. Nevertheless, the existence of 702 negative sentiments needs to be studied in depth to produce recommendations for the development of optimal hotel products and services. In the context of the Covid-19 pandemic, hotels used as accommodation facilities for Covid-19 suspects need medical assistance, control consumer hygiene, healthy communication, and self-preservation technology [18]. Complaints about services that are not optimal affect the consumer's perception that the products and services provided by the hotel or resort are not in line with expectations or even unsatisfactory. A not optimal service implies a loss of trust and a decrease in consumers [27]. Therefore, negative consumer sentiment needs to be validated as a reference for developing products and services in line with consumer expectations and satisfaction.

The sentiment classification process from the prepared dataset shows that the text mining approach is significant for processing consumer review data in the future. However, the use of the algorithm needs to be considered with a better accuracy value [28]. In addition, [29] stated that the Support Vector Machine (SVM), Artificial Neural Network (ANN), Naïve Bayes Classifier (NBC), Decision Tree (DT), C4.5, and k-Nearest Neighbor (k-NN) could classify the customer reviews. Unfortunately, this research only uses SVM, NBC, and k-NN algorithms to be evaluated. Based on the performance evaluation of the algorithm used, there are percentage differences in each hotel and resort dataset and a combination of the entire dataset, as shown in Table 3 below.

Table 3 Algorithm Performance Evaluation

Hotel & Resort	NBC	SVM	k-NN
Puri Sari Beach Hotel	92,42 %	92,43 %	50 %
Plataran Komodo Resort and Spa	94,44 %	94,27 %	50 %
Bintang Flores Hotel	68,52 %	68,52 %	50 %
Ayana Komodo Resort	74,23 %	74,04 %	50,01 %
Sylvia Resort Komodo	56,83 %	56,67 %	55,51 %
Bayview Garden Hotel	85,94 %	85,85 %	50 %
Golo Hilltop Hotel and Restaurant	89,06 %	89,09 %	50 %
Laprima Hotel	53,41 %	53,46 %	50,01 %
Sudamala Resort, Seraya	86,34 %	86,17 %	50,01 %
Eco Tree O'tel	100%	100 %	100 %
All Hotel	**78,29** %	**78,30** %	**85,24** %

Table 3 shows that each hotel dataset shows various accuracy values. The difference in accuracy values is assessed as the algorithm's relevant performance to the dataset. The accuracy value can be increased by increasing the amount of training data and test data. However, in this research, the evaluation results of algorithm performance for the classification of positive and negative sentiments with the highest accuracy value are k-NN with a total of 85, 24%. There are differences between the calculation result using SMOTE UPsampling and without SMOTE Upsampling for the k-NN algorithm, as shown in figure 4 below.

Before SMOTE Upsampling   After SMOTE Upsampling
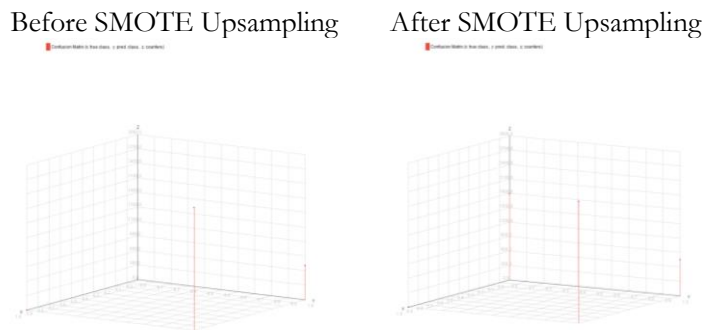
Figure 4 Confusion Matrix before and after using SMOTE UPsampling for the k-NN algorithm.

Implementing k-NN before using SMOTE UPsampling resulted in a lower accuracy and precision value of 78.29%, even though the recall value was 100%. Unlike the case with the accuracy value after using SMOTE UPsampling, which is 85.24%, where the precision value is 77.22%, and the recall value is 100%. On the other hand, the Area Under Curve (AUC) value from the calculation results using the k-NN algorithm shows a significant difference where the AUC k-NN value before using SMOTE UPsampling is 0.500 and the AUC k-NN after using SMOTE UPsampling is 0.948, as figure 5 below.

Area Under Curve (AUC)   Area Under Curve (AUC)
k-NN Before using   k-NN After using
SMOTE UPsampling   SMOTE UPsampling
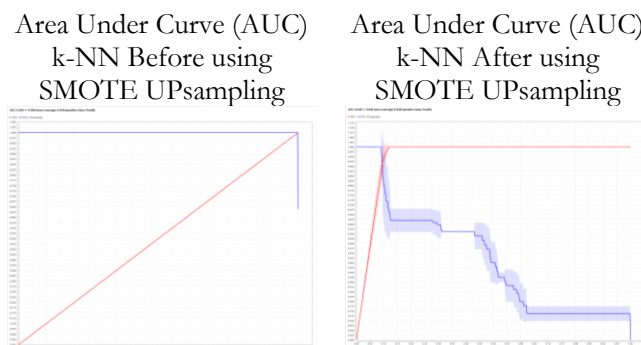
Figure 5 Area Under Curve (AUC) k-NN before and after using SMOTE UPsampling

Figure 5 is the Area Under Curve (AUC) value from the calculation results using the SMOTE UPsampling dataset balancing technique in the k-NN algorithm. [30] shows that the value of using SMOTE Upsampling greatly affects the AUC value and shows a significant difference. In this study, the difference in accuracy values before and according to SMOTE Upsampling in the k-NN algorithm was 6.95%, while the precision value was -1.07%. Meanwhile, the Recall value has not changed at all. Furthermore, [31] shows that the categories of AUC values are as follows: excellent (0.9-1.0); good (0.8-0.9); fair (0.7-0.8); poor (0.6-0.7); failure (0.5-0.6). Thus, the AUC value of k-NN after using SMOTE Upsampling shows an excellent condition with a value of 0.948. Thus, the k-NN shows the best result using SMOTE UPsampling compared to NBC (78,29%) and SVM (78,30%), as shown in the table below.

Table 4 Comparative Value of SMOTE Upsampling for NBC, SVM, k-NN

Before Using SMOTE UPsampling	Accuracy	Precision	Recall
NBC	78,29%	78,29%	100 %
SVM	78,30%	78,30%	100 %
k-NN	78,29%	78,29%	100 %
After Using SMOTE UPsampling	**Accuracy**	**Precision**	**Recall**
NBC	54,47%	90,65 %	10 %
SVM	50%	50%	100%
k-NN	85,24%	77,22%	100%

Table 4 shows the change in the values of accuracy, precision, and recall before and after using the SMOTE UPsampling value as a balancing technology for the dataset; considering this, the best accuracy, precision, and recall values are used and proceed to the data analysis stage of text mining processing results. After knowing that the results of data processing using the k-NN algorithm show the best performance, it can be continued to the stage of analysis and interpretation of data, where the results of the classification of negative sentiment and positive sentiment need to be formulated into recommendations for stakeholders to optimize product and service management in their respective hotels.

In this research, the processing of review data obtained from the TripAdvisor platform is not used to provide recommendations for each hotel but to obtain a broad picture of consumer sentiment towards hotel products and services in Labuan Bajo. Based on these considerations, partial hotel and resort data are combined into one dataset to be processed using the NBC, SVM, and k-NN classification models, then compared based on a higher accuracy[31] value. Thus, the interpretation of the review data becomes relevant to the context of consumers of hotel products and services in Labuan Bajo. In addition,

The biggest challenge of this study is to connect the results of sentiment analysis data processing with the interests of stakeholders in the hotel, resort, cottage, and

homestay businesses, especially in Labuan Bajo, to increase hotel guest satisfaction to increase consumer loyalty. The results of sentiment data processing classify the perceptions of hotel guests in the form of text, which is uploaded online on the TripAdvisor website page. However, TripAdvisor's information about tourism and travel support facilities is used as a reference before planning a trip. In the context of tourist behavior, [32], [33] show that travelers have consumption behaviors based on cognitive and affective aspects. The behavior of accessing, consuming, and distributing information on TripAdvisor websites is part of the consumption behavior of travelers. Therefore, it is essential to analyze consumer feedback reviews, especially consumer satisfaction and dissatisfaction with hotel products and services.

Consumer Feedback Reviews: Satisfaction and Dissatisfaction of Hotels Product and Services

The hotel industry plays an essential role in the development of tourism. However, [34] stated that hotels also suffer from the economic effects of tourism seasonality. Furthermore, [35] points to hotel management changes in response to the crisis caused by the Covid-19 pandemic. The condition of the Covid-19 pandemic with the limitation of access mobility to various tourist destinations has increased technology users as a medium of communication and electronic transactions [36]. In response to these changes, the management adjusted the needs of consumers by providing electronic services to mobilize users' lack of product information and e-services applied in the Covid-19 pandemic [27]. The intensity of the use of technology also encourages the development of feedback review services on travel platforms. Also, used by hotel management to increase users' satisfaction, engagement, and loyalty [37]. Meanwhile, the problems faced by the hotel industry in various tourist destinations cannot be generalized or contextual. Therefore, studying consumer sentiment analysis toward hotel products and services is necessary.

Consumer perceptions of the hospitality industry are very complex to predict, so it is necessary to conduct comprehensive research to get an idea of the disappointment and satisfaction of consumers with hotel products and services [38]. On the other hand, [39] stated the consequences of consumer dissatisfaction that affect business sustainability. Consumer dissatisfaction also affects the behavior of guests in the future. The hotel management must provide excellent service to every guest who comes to stay and gives a positive impression to the guests. In addition, [40] argues that some hotels have adopted sophisticated technology in serving customers for the reservation process, but this can also negatively impact them when it does not meet consumer expectations. It shows that the effectiveness and efficiency in service to hotel guests are not only supported by human skills and technology but, more importantly, the link to customer expectation.

Several studies on cross-hotel and cross-country show that consumer segmentation based on culture influences expectations of hospitality service attributes [41]. Furthermore, [42] argues that an in-depth analysis of the complaint resolution process is necessary to measure consumer satisfaction or dissatisfaction. There is a critical dimension: compensation received and attributes of the retailer's representative. Meanwhile, [43] stated that social media had expanded word of mouth (WOM). It is no massive means of online communication. So that consumers have had a significant impact on the business's reputation disseminated externally in virtual media, it shows that negative sentiment is an incident that must be taken seriously by considering the demographic background of consumers and the class of the hotel.

If explored in-depth, each hotel and resort studied has different products and services. Meanwhile, the number of visits to the hotel also significantly differs and is related to the hotel's location. Likewise, review data on TripAdvisor shows that Puri Sari Beach Hotel and Bayview Garden Hotel have a higher quantity of guest reviews than Eco Tree O'tel and Sudamala Resort, Seraya. In addition, based on the classification of negative and positive sentiments, the Bintang Flores and Laprima Hotel have several positive and negative reviews that are not much different. Unlike the case with Sylvia Resort Komodo, which needs to evaluate the performance of product and service management to increase positive reviews on the TripAdvisor platform. Although not all guests intend to fill out review data on the TripAdvisor platform, complaints about incidents experienced by consumers tend to review on the TripAdvisor platform. This condition affects the image of the hotel or resort in the competitive super premium tourism market of Labuan Bajo.

The challenge for the hotel is the inability to control consumer perceptions to give a positive rating to hotel or resort products and services. [44] stated that the challenges and opportunities for hotels to obtain positive ratings as a form of customer satisfaction is by transferring knowledge about hotel management and maintenance processes to maintain and improve product and service quality. [45] emphasized that the hotel business is a business that is very vulnerable to decline due to consumer perceptions, expectations, and dissatisfaction. Therefore, hotel management must always be innovative in adopting an effective and efficient managerial approach and meeting consumer expectations. Likewise, [16] shows that the skills of staff who deal directly with consumers in the service process will determine the image of the hotel's business. In Labuan Bajo's context, employees hired to serve guests must understand consumer characteristics based on demographic, geographic, and psychographic segmentation. Consumers will give positive reviews if the service is imposing and continue improving the hotel's image or resort on the TripAdvisor platform.

4. CONCLUSION

The results of this study indicate a theoretical and empirical contribution. The practical assistance of this research emphasizes the study of tourism and information systems through machine learning and text mining approaches. This study's in-depth concern was consumer sentiment towards the products and services of hotels and resorts in the super-premium tourist destination of Labuan Bajo. The findings of this study indicate that the positive sentiment of hotel and resort consumers in Labuan Bajo is still dominant, thus maintaining a positive image of Labuan Bajo's super-premium tourism. However, negative consumer sentiment towards hotel products and services is relatively high. It requires hotel management to update services and products to align with consumer expectations and increase customer satisfaction and loyalty. In addition, in dataset processing, the classification model algorithm ideally used to sort negative and positive sentiments is k-NN, with the highest accuracy value compared to NBC and SVM. Furthermore, the accuracy value will change along with the processing of review data obtained from the TripAdvisor platform. This research is still limited because it only uses one outlet in the sentiment analysis process. Therefore, a further research recommendation is to analyze consumer sentiment for hotels, resorts, cottages, and restaurants in super-premium tourist destinations using datasets from various platforms.

ACKNOWLEDGEMENT

REFERENCES

[1]　S. Sleenhoff and P. Osseweijer, "Consumer choice: Linking consumer intentions to actual purchase of GM labeled food products.," *GM Crops Food*, vol. 4, no. 3, pp. 166–171, 2013, doi: 10.4161/gmcr.26519.

[2]　U. R. Orth, D. Campana, and K. Malkewitz, "Formation of consumer price expectation based on package design: Attractive and quality routes," *J. Mark. Theory Pract.*, vol. 18, no. 1, pp. 23–40, 2010, doi: 10.2753/MTP1069-6679180102.

[3]　U. Golob, M. Lah, and Z. Jančič, "Value orientations and consumer expectations of Corporate Social Responsibility," *J. Mark. Commun.*, vol. 14, no. 2, pp. 83–96, 2008, doi: 10.1080/13527260701856525.

[4]　J. Savard, C. Hickerton, S. A. Metcalfe, C. Gaff, A. Middleton, and A. J. Newson, "From Expectations to Experiences: Consumer Autonomy and Choice in Personal Genomic Testing," *AJOB Empir. Bioeth.*, vol. 11, no. 1, pp. 63–76, 2020, doi: 10.1080/23294515.2019.1701583.

[5] K. Kaivanto and P. Zhang, "Is business formation driven by sentiment or fundamentals?," *Eur. J. Financ.*, 2022, doi: 10.1080/1351847X.2022.2038648.

[6] M. Gharzouli, A. K. Hamama, and Z. Khattabi, "Topic-based sentiment analysis of hotel reviews," *Curr. Issues Tour.*, vol. 0, no. 0, pp. 1–8, 2021, doi: 10.1080/13683500.2021.1940107.

[7] E. Park, J. Kang, D. Choi, and J. Han, "Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews," *Curr. Issues Tour.*, vol. 23, no. 5, pp. 605–611, 2020, doi: 10.1080/13683500.2018.1549025.

[8] J. Luo, S. Huang, and R. Wang, "A fine-grained sentiment analysis of online guest reviews of economy hotels in China," *J. Hosp. Mark. Manag.*, vol. 30, no. 1, pp. 71–95, 2021, doi: 10.1080/19368623.2020.1772163.

[9] S. Shayaa *et al.*, "Linking consumer confidence index and social media sentiment analysis," *Cogent Bus. Manag.*, vol. 5, no. 1, pp. 1–12, 2018, doi: 10.1080/23311975.2018.1509424.

[10] M. Gharzouli, A. K. Hamama, and Z. Khattabi, "Topic-based sentiment analysis of hotel reviews," *Curr. Issues Tour.*, 2021, doi: 10.1080/13683500.2021.1940107.

[11] 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.

[12] H. N. Io and C. B. Lee, "Social Media Comments about Hotel Robots," *J. China Tour. Res.*, vol. 16, no. 4, pp. 606–625, 2020, doi: 10.1080/19388160.2020.1769785.

[13] C. E. Yu, "Humanlike robots as employees in the hotel industry: Thematic content analysis of online reviews," *J. Hosp. Mark. Manag.*, vol. 29, no. 1, pp. 22–38, 2020, doi: 10.1080/19368623.2019.1592733.

[14] V. Chittiprolu, S. Singh, R. S. Bellamkonda, and S. Vanka, "A text mining analysis of online reviews of Indian hotel employees," *Anatolia*, vol. 32, no. 2, pp. 232–245, 2021, doi: 10.1080/13032917.2020.1856157.

[15] A. Aakash and A. Gupta Aggarwal, "Assessment of Hotel Performance and Guest Satisfaction through eWOM: Big Data for Better Insights," *Int. J. Hosp. Tour. Adm.*, vol. 23, no. 2, pp. 317–346, 2022, doi: 10.1080/15256480.2020.1746218.

[16] M. P. Mehta, G. Kumar, and M. Ramkumar, "Customer expectations in the hotel industry during the COVID-19 pandemic: a global perspective using sentiment analysis," *Tour. Recreat. Res.*, vol. 0, no. 0, pp. 1–18, 2021, doi: 10.1080/02508281.2021.1894692.

[17] Y. C. Lin and C. M. Chen, "How do hotel characteristics moderate the impact of COVID-19 on hotel performance? Evidence from Taiwan," *Curr. Issues Tour.*, vol. 25, no. 8, pp. 1192–1197, 2022, doi: 10.1080/13683500.2021.1910213.

[18] H. A. Atadil and Q. Lu, "An investigation of underlying dimensions of

customers' perceptions of a safe hotel in the COVID-19 era: effects of those perceptions on hotel selection behavior," *J. Hosp. Mark. Manag.*, vol. 30, no. 6, pp. 655–672, 2021, doi: 10.1080/19368623.2021.1877588.

[19]   M. Ustaszewski, "Towards a machine learning approach to the analysis of indirect translation," *Transl. Stud.*, vol. 14, no. 3, pp. 313–331, 2021, doi: 10.1080/14781700.2021.1894226.

[20]   T. H. Le, C. Arcodia, M. A. Novais, and A. Kralj, "Proposing a systematic approach for integrating traditional research methods into machine learning in text analytics in tourism and hospitality," *Curr. Issues Tour.*, vol. 24, no. 12, pp. 1640–1655, 2021, doi: 10.1080/13683500.2020.1829568.

[21]   C. Makanyeza, T. D. Svotwa, and O. Jaiyeoba, "The effect of consumer rights awareness on attitude and purchase intention in the hotel industry: Moderating role of demographic characteristics," *Cogent Bus. Manag.*, vol. 8, no. 1, 2021, doi: 10.1080/23311975.2021.1898301.

[22]   U. Chakraborty and S. K. Biswal, "Impact of Online Reviews on Consumer's Hotel Booking Intentions: Does Brand Image Mediate?," *J. Promot. Manag.*, vol. 26, no. 7, pp. 943–963, 2020, doi: 10.1080/10496491.2020.1746465.

[23]   C. D. Chung, L. Gao, and D. Leung, "Corporate Social Responsibility Communications on Social Media and Consumers' Brand Engagement: A Case Study of Hotels in Hong Kong," *J. China Tour. Res.*, vol. 16, no. 4, pp. 547–565, 2020, doi: 10.1080/19388160.2019.1643434.

[24]   S. P. Dhakal and S. P. Tjokro, "Tourism enterprises in Indonesia and the fourth industrial revolution–are they ready?," *Tour. Recreat. Res.*, vol. 0, no. 0, pp. 1–6, 2021, doi: 10.1080/02508281.2021.1996687.

[25]   A. H. Lasso and H. Dahles, "A community perspective on local ecotourism development: lessons from Komodo National Park," *Tour. Geogr.*, vol. 0, no. 0, pp. 1–21, 2021, doi: 10.1080/14616688.2021.1953123.

[26]   R. Westoby, S. Gardiner, R. W. Carter, and N. Scott, "Sustainable livelihoods from tourism in the '10 New Balis' in Indonesia," *Asia Pacific J. Tour. Res.*, vol. 26, no. 6, pp. 702–716, 2021, doi: 10.1080/10941665.2021.1908386.

[27]   N. T. VO, M. Chovancová, and H. T. Tri, "The Impact of E-service Quality on the Customer Satisfaction and Consumer Engagement Behaviors Toward Luxury Hotels," *J. Qual. Assur. Hosp. Tour.*, vol. 21, no. 5, pp. 499–523, 2020, doi: 10.1080/1528008X.2019.1695701.

[28]   H. N. T. Thu, "Measuring guest satisfaction from online reviews: Envidence in Vietnam," *Cogent Soc. Sci.*, vol. 6, no. 1, 2020, doi: 10.1080/23311886.2020.1801117.

[29]   B. Noori, "Classification of Customer Reviews Using Machine Learning Algorithms," *Appl. Artif. Intell.*, vol. 35, no. 8, pp. 567–588, 2021, doi: 10.1080/08839514.2021.1922843.

[30]   C. Cahyaningtyas, Y. Nataliani, and I. R. Widiasari, "Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree

Berbasis SMOTE," *Aiti*, vol. 18, no. 2, pp. 173–184, 2021, doi: 10.24246/aiti.v18i2.173-184.

[31]　M. I. Putri and I. Kharisudin, "Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Analisis Sentimen Data Review Pengguna Aplikasi Marketplace Tokopedia," in *PRISMA, Prosiding Seminar Nasional Matematika*, 2022, vol. 5, pp. 759–766. [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/

[32]　E. M. Anggraini and N. Wibisono, "Analisis Faktor yang Mempengaruhi Niat Wisatawan Untuk Berkunjung Kembali Ke Destinasi Wisata : Studi Kasus Dago Dream Park Bandung," in *The 13th Industrial Research Workshop and National Seminar*, 2022, pp. 1017–1023.

[33]　V. K. Adhi, J. Gunawan, and S. Hakim, "Investigasi Karakteristik Demografi dan Perilaku Bepergian Wisatawan Domsestik di Masa pandemi Covid-19," *J. Sains dan Seni ITS*, vol. 11, no. 2, pp. 8–13, 2022.

[34]　A. Y. Chiriko, "How hotels suffer from and deal with the economic effects of tourism seasonality: A case study of Aksum, Ethiopia," *Res. Hosp. Manag.*, vol. 11, no. 2, pp. 137–143, 2021, doi: 10.1080/22243534.2021.1917751.

[35]　L. Zhong, S. Sun, R. Law, and X. Li, "Tourism crisis management: evidence from COVID-19," *Curr. Issues Tour.*, vol. 24, no. 19, pp. 2671–2682, 2021, doi: 10.1080/13683500.2021.1901866.

[36]　A. M. Elshaer and A. M. Marzouk, "Memorable tourist experiences: the role of smart tourism technologies and hotel innovations," *Tour. Recreat. Res.*, vol. 0, no. 0, pp. 1–13, 2022, doi: 10.1080/02508281.2022.2027203.

[37]　N. V. Thi, V. V. Hung, Z. Tuckova, N. T. Pham, and L. H. L. Nguyen, "Guest Online Review: An Extraordinary Focus on Hotel Users' Satisfaction, Engagement, and Loyalty," *J. Qual. Assur. Hosp. Tour.*, vol. 00, no. 00, pp. 1–32, 2021, doi: 10.1080/1528008X.2021.1920550.

[38]　P. D. Modica, L. Altinay, A. Farmaki, D. Gursoy, and M. Zenga, "Consumer perceptions towards sustainable supply chain practices in the hospitality industry," *Curr. Issues Tour.*, vol. 23, no. 3, pp. 358–375, 2020, doi: 10.1080/13683500.2018.1526258.

[39]　B. Kim, S. (Sam) Kim, and C. Y. Heo, "Consequences of Customer Dissatisfaction in Upscale and Budget Hotels: Focusing on Dissatisfied Customers' Attitude Toward a Hotel," *Int. J. Hosp. Tour. Adm.*, vol. 20, no. 1, pp. 15–46, 2019, doi: 10.1080/15256480.2017.1359728.

[40]　A. Fan, L. Wu, L. Miao, and A. S. Mattila, "When does technology anthropomorphism help alleviate customer dissatisfaction after a service failure?–The moderating role of consumer technology self-efficacy and interdependent self-construal," *J. Hosp. Mark. Manag.*, vol. 29, no. 3, pp. 269–290, 2020, doi: 10.1080/19368623.2019.1639095.

[41]　G. Bodet, V. Anaba, and P. Bouchet, "Hotel Attributes and Consumer Satisfaction: A Cross-Country and Cross-Hotel Study," *J. Travel Tour.*

Mark., vol. 34, no. 1, pp. 52–69, 2017, doi: 10.1080/10548408.2015.1130109.

[42]   M. D. Richard and D. D. Hicks, "The HASSLE scale: Measuring consumer satisfaction/dissatisfaction with the complaint resolution process," *J. Cust. Serv. Mark. Manag.*, vol. 3, no. 4, pp. 61–73, 1997, doi: 10.1300/J127v03n04_05.

[43]   T. Fernandes and F. Fernandes, "Sharing Dissatisfaction Online: Analyzing the Nature and Predictors of Hotel Guests Negative Reviews," *J. Hosp. Mark. Manag.*, vol. 27, no. 2, pp. 127–150, 2018, doi: 10.1080/19368623.2017.1337540.

[44]   P. Longart, "Understanding Hotel Maintenance Management," *J. Qual. Assur. Hosp. Tour.*, vol. 21, no. 3, pp. 267–296, 2020, doi: 10.1080/1528008X.2019.1658148.

[45]   M. Li, A. Hsiao, E. J. Ma, and S.-J. Li, "Hotel service convergence innovation," *J. Glob. Sch. Mark. Sci.*, vol. 00, no. 00, pp. 1–17, 2021, doi: 10.1080/21639159.2020.1808848.