Journal of Information Systems and Informatics
Vol. 4, No. 3, September 2022 e-ISSN: 2656-4882 p-ISSN: 2656-5935
http://journal-isi.org/index.php/isi Published By DRPM-UBD

Sentiment Analysis on Customer Perception towards Products and Services of Restaurant in Labuan Bajo

Yerik Afrianto Singgalen[1]

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

Abstract

Analysis of consumer sentiment towards restaurant products and services in the form of reviews on various digital platforms such as Tripadvisor determines business sustainability and the image of tourist destinations. This study aims to classify visitor sentiments as Tripadvisor users towards Happy Banana Komodo, MadeInItaly, Mediterraneo, and La Cucina restaurants in Labuan Bajo. The research stages are divided into three parts, namely the stages of data collection, data processing, classification and evaluation of model performance, and interpretation of data and information. At the data collection stage, Tripadvisor user reviews of products and services at each restaurant are mined using the Webharvy application based on the configuration of the customer name, date of examination, rating, review title, and review. The data mining results are cleaned and prepared to be managed using the RapidMiner application at the data processing stage. The classification and evaluation stage of model performance is the implementation and testing of classification algorithms that are relevant to the dataset, namely Decision Tree (DT), Naïve Bayes (NB), Support Vector Machine (SVM), k-Nearest Neighbor (k-NN). The findings of this study indicate the implementation of the positive and negative sentiment classification method for the comprehensive review data from the Tripadvisor website for the products and services of restaurants Happy Banana Komodo, MadeInITaly, Mediterraneo, and La Cucina Restaurants are relevant with k-Nearest Neighbor (k-NN) with accuracy value of 99.27% , a precision value of 100%, and a recall of 98.53%.

Keywords: Sentiment Analysis, k-NN, Restaurant, Labuan Bajo

1. INTRODUCTION

Consumer sentiment analysis plays a vital role as material for evaluating the quality of products and services. [1] shows that consumers express sentiment towards products and services through social media that is interpreted textually, visually, and effectively. On the other hand, [2] shows a relationship between social media sentiment analysis and the consumer confidence index. Specifically, there is a relationship between consumer sentiment and shopping behavior. The findings of previous studies have strengthened the argument that the results of processing

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

consumer review data on social media or other digital platforms can be used as a reference to evaluate the quality of products and services for the benefit of optimizing business marketing.

The popularity of sentiment analysis studies is proliferating along with the increase in users of social media applications. To obtain credible analysis results, some researchers use machine learning and text mining to process consumer data on social media by considering volume, velocity, variety, and veracity [3]. On the other hand, [4] identifies the significance of product features on customer satisfaction to recognize public sentiment polarity. It shows that the polarization of public sentiment in product and service review data on social media can be identified and analyzed as a reference for evaluating producers' quality of products and services. Thus, the sentiment analysis is essential in recommending several product and service improvement strategies.

This research develops from a recommendation of previous research findings such as methods, algorithms, and topic modeling related to sentiment analysis in the tourism sector [5], [6]. In addition, there is an in-depth niche related to restaurant consumer sentiment analysis in Indonesia's super-priority tourism destinations. Based on the recommendations of previous studies, this article aims to conduct a sentiment analysis of restaurant consumer review data regarding products and services. Meanwhile, the development of sentiment analysis studies in the tourism sector still needs to be improved to contribute to managerial aspects to optimize the quality of Indonesian tourist destinations, especially products and services produced by restaurants as an essential part of the Indonesian tourism industry.

This research used data on reviews of restaurant products and services in Labuan Bajo through the Tripadvisor website. The review data is divided into two parts: training data and test data. Data will be processed using a machine learning approach using the k-Nearest Neighbour (k-NN) algorithm, Naïve Bayes (NB), Decision Tree (DT), and Support Vector Machine (SVM). Each algorithm will be evaluated based on the best performance and comprehensively analyzed to optimize the quality of restaurant products and services in super-priority tourist destinations. [7] The Naïve Bayes method can analyze consumer sentiment towards traditional foods, considering accessibility and restaurant locations. In addition, [8] indicates that a decision tree algorithm can also be used to classify consumer sentiment. Furthermore, [9] uses k-NN and SVM to assess the performance of both algorithms in organizing consumer sentiment.

Sentiment analysis is critical as digital media for marketing products and services increases, including restaurants. In the Indonesian context, socio-economic conditions during the Covid-19 pandemic have encouraged users of digital applications to maintain the restaurant business. Digital transactions are considered the best method during the pandemic to maintain distance as

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

instructed by the Indonesian government to anticipate contracting the Covid-19 virus due to physical or direct contact. In the Covid-19 pandemic, restaurant product and service consumers have adapted using various applications to review the services and food sold in restaurants. This behavior will continue to increase and make digital data or information a valuable reference. Based on this phenomenon, it is necessary to conduct sentiment analysis using a scientific approach to produce recommendations for restaurant business people in optimizing products and services. Thus, restaurants as one of the essential parts of the tourism industry can drive the image of Indonesia's super-priority tourist destinations.

2. METHODS

This study uses the text-mining method using the classification method through the implementation of the k-Nearest Neighbour (k-NN) algorithm, Naïve Bayes (NB), Decision Tree (DT), and Support Vector Machine (SVM) using the Rapidminer application. [10] shows that the advantage of the Rapidminer application is the availability of various features that support the pre-processing and data processing stages that support the process of evaluating and reporting research results. Furthermore, [11] shows that the Rapidminer application has several features to measure and compare algorithm performance so that it can be used as a recommendation according to the needs of data processors. Considering the availability of elements in the Rapidminer application that are by the requirements of this study, the dataset pre-processing and processing stage also performance evaluation will adjust to the Rapidminer platform as shown below.

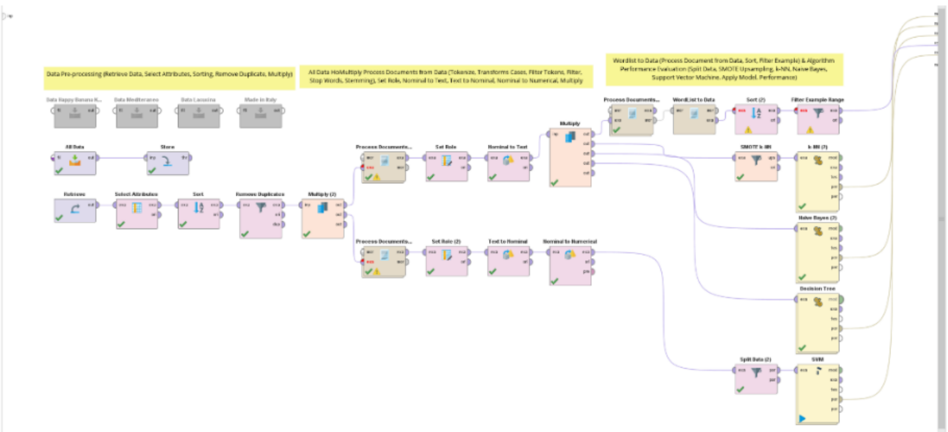

Figure 1. Data Processing Flow

Before the data collection process, the consideration for choosing a Tripadvisor Website as the primary data source was the accessibility and availability of data

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

according to research needs. Several previous 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.

Tabel 1. Link of Data Source on TripAdvisor Website

Restaurant	Link
Happy Banana Komodo	https://www.tripadvisor.com/Restaurant_Review-g1777483-d8448989-Reviews-Happy_Banana_Komodo-Labuan_Bajo_Flores_East_Nusa_Tenggara.html
Mediterraneo	https://www.tripadvisor.com/Restaurant_Review-g1777483-d2291096-Reviews-Mediterraneo-Labuan_Bajo_Flores_East_Nusa_Tenggara.html
La Cucina	https://www.tripadvisor.com/Restaurant_Review-g1777483-d4471445-Reviews-La_Cucina-Labuan_Bajo_Flores_East_Nusa_Tenggara.html
MadeInItaly	https://www.tripadvisor.com/Restaurant_Review-g1777483-d2008791-Reviews-MadeInItaly_Ristorante_Italiano-Labuan_Bajo_Flores_East_Nusa_Tenggara.html

Table 1 is the link to the data source. There are 974 reviews in English for MadeInItaly restaurant, 664 reviews in English for La Cucina restaurant, 110 reviews in English for Mediterraneo restaurant, and 763 reviews for Happy Banana Komodo restaurant. This English review data will be prepared for analysis. The process of selecting information based on the identity of the user who provided the review is a consideration for researchers to make review data on the Tripadvisor website a data source. [12] indicates that user reviews on Tripadvisor websites influence millennials' interests. Furthermore, [13] shows that reviews on Tripadvisor websites also influence travelers' visiting decisions.

2.1. Data Collection Process

In the context of this study, the review data needed to proceed to the sentiment analysis process is data on reviews of restaurant products and services in the super-priority tourist destination of Labuan Bajo, Indonesia. Based on the process of identifying review data, users as reviewers are allowed to provide assessments in writing and ratings on food, service, and value. The evaluations in the Tripadvisor website system consist of five stages: terrible, poor, average, very good, and excellent. Meanwhile, users as reviewers are asked for information about the type of traveler, which can be categorized as follows: travel with friends, travel for business, solo travel, couple travel, and families' trip. In addition, the website

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

system also classifies the visit time based on the time of year: Mar-May, Jun-Aug, Sep-Nov, Dec-Feb. Meanwhile, the website system also provides a feature to connect reviews based on the type of language used by the user. Considering the ease of access to review data and the selection function available on TripAdvisor websites, this study focuses on review data in English with an overall time of year, traveler type, and traveler rating data as shown below.

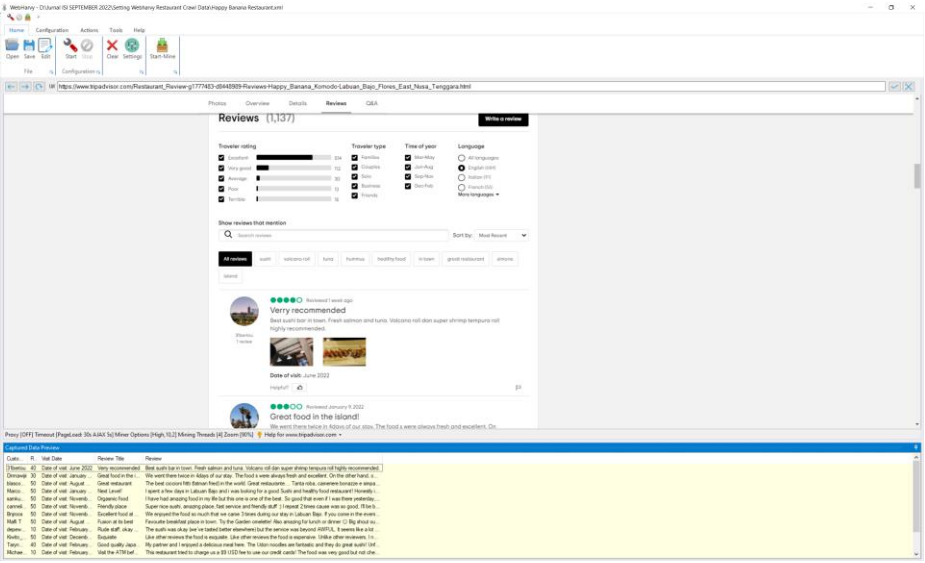

Figure 2. Crawling Data from TripAdvisor using Webharvy

Figure 2 is an interface visualization of the Webharvy application, which collects consumer review data for happy banana Komodo, Mediterraneo, La Cucina, and MadeInItaly restaurants. Selectively, the restaurant selection is carried out by considering the number of reviews related to consumer preferences such as special diets (vegan options, gluten-free options), meals (lunch, dinner, late-night, drinks), cuisines (fusion, Japanese, healthy), and features (delivery, takeout, reservation, outdoor seating, seating, parking available or street parking, serve alcohol, full bar, wine and beer, cash only, free wifi, accepts American Express, Mastercard, digital payments, Credit Cards, and table service). Based on TripAdvisor data, Happy Banana Komodo has 1,137 reviews, Mediterraneo has 1,728 reviews, La Cucina has 1,029 reviews, and MadeInItaly has 1,373 reviews. The entire review data is available in multiple languages. This research will focus more on reviewing data in English.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

2.2. Pre-Processing and Processing Dataset

Review data will be prepared for processing in the pre-processing stage using the k-NN, NB, DT, and SVM algorithms in the RapidMiner application. Individual restaurant review data stored in .csv or .xlsx format can be saved in advance into a database or imported directly from storage. Next, the data is retrieved from the database, and setting the data type and role, also sorted dataset as ascending by selecting the review attribute to be processed to remove duplicates. The pre-processing stage of the storage is to prepare review data to be classified using a predetermined algorithm. Meanwhile, the scene of data pre-processing can be seen in the following figure.

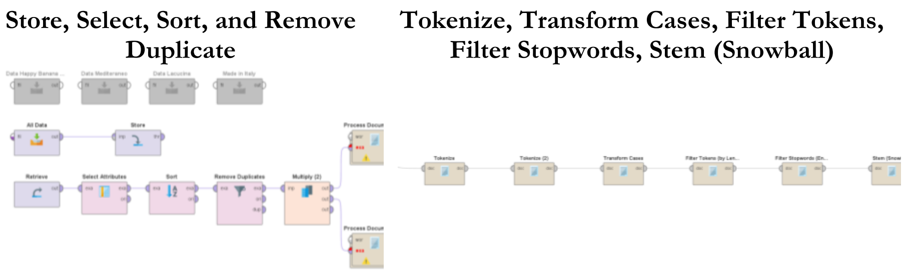

Figure 3. Data Pre-Processing Flow

In the pre-processing data stage, the process document from the data operator is used as a container to perform the first stage of the tokenizing process on review data to eliminate regular expressions ((?)[-!" #$%&'()*+,./:;<=>?@\[\\\]_`{|} ~]). Furthermore, the second stage of the tokenizing process is to remove undetectable symbols (non-letters). The system will change the word to lower case in the transforms cases process. Furthermore, the design of this study adjusts the stop words and stem process to the type of English language on the dataset. The pre-processing stage is critical before the data is classified using the k-NN, NB, DT, and SVM algorithms.

2.3. Classification and Performance Evaluation

The classification method's implementation is adapted to each algorithm's settings. Implementing the classification method using the k-NN, NB, and DT algorithms ensures the prepared dataset returns the role settings. The nominal to text operator is used. Furthermore, datasets can be directly processed using the NB and DT algorithms using cross-validation (10-fold and automatic sampling type) operators to visualize algorithm performance based on Accuracy, Area Under Curve (AUC), recall, and precision value. On the other hand, Synthetic Minority Oversampling techniques (SMOTE) operators are used to balance data inequality in quantity.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

They are linked to the k-NN algorithm model to produce better performance. Meanwhile, an overview of the classification and performance evaluation process can be seen according to the following figure.

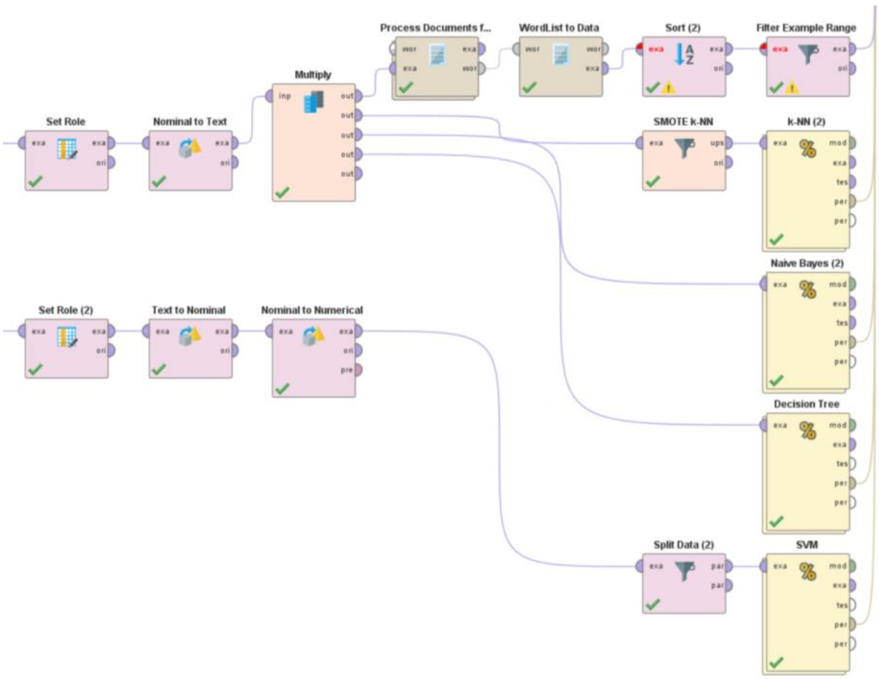

Figure 4. Classification and Performance Evaluation

Figure 4 also shows the classification method for the SVM algorithm with different processes using text to nominal and nominal to numerical operators. Furthermore, the dataset is divided into two using split data with settings 0.7 and 0.3 before being connected to the cross-validation operator using the SVM algorithm in addition to the performance evaluation of each algorithm. This study's output also features a list of 20 words that most often appear in consumer reviews for restaurant products and services. Thus, the data processing results can be analyzed comprehensively to produce constructive ideas that can recommend strategies for optimizing restaurant products and services in Labuan Bajo as a super-premium destination in Indonesia. Specifically, the discussion in this study is more dominant in the results of implementing algorithms with the highest accuracy values compared to other algorithms. Thus, the process of reporting the results of this study focuses on one algorithm relevant to the dataset.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

3. RESULTS AND DISCUSSION

Based on the results of dataset processing, the best performance was obtained from the results of the implementation of the classification method using the k-NN algorithm with an accuracy value of 99.27% when compared to the NB and DT algorithms with an accuracy value of 93.23 while dataset processing using the SVM algorithm had an accuracy value of 93.22. This shows that the most relevant k-NN algorithm is used in implementing classification methods based on consumer review datasets of Happy Banana Komodo, Mediterraneo, La Cucina, and MadeInItaly restaurants in Labuan Bajo.

Tabel 1. The Accuracy Performance of k-NN, NB, DT, and SVM

Restaurant	**NB**	**k-NN**	**SVM**	**DT**
Happy Banana Komodo	95.90 %	100 %	95.81 %	95.90 %
Mediterraneo	92.70 %	99.06 %	92.66 %	92.70 %
La Cucina	95.13 %	100 %	95.04 %	95.13 %
MadeInItaly	90.54 %	98.76 %	90.54 %	90.54 %
All Data	**93,23** %	**99.27** %	**93,22** %	**93,23** %

Table 1 shows that k-NN is the algorithm with the highest performance evaluation compared to the NB, DT, and SVM algorithms. Nonetheless, the accuracy value of k-NN can reach 99.27% because it uses the SMOTE Upsampling operator. Meanwhile, without the SMOTE Upsampling operator, the accuracy value of k-NN is 93.23%. SMOTE is a technique used to overcome class imbalance problems (CIP) by modifying unbalanced datasets and creating new synthetic data from minority classes to improve the performance of the classification method [14]. When comparing the Accuracy, AUC, Recall, and Precision values from the results of the implementation of the classification method using the k-NN algorithm with and without SMOTE, the results can be seen in the following table.

Tabel 2. The Performance of k-NN with and without SMOTE

Performance	**k-NN with SMOTE**	**k-NN without SMOTE**
Accuracy	99.27%	93.23%
Precision	100 %	100 %
Recall	98.53 %	98.53 %

Table 2 shows the difference in performance accuracy values of the k-NN algorithm when connected with the SMOTE Upsampling operator, which offers a difference of 6.04%. This indicates that using SMOTE can increase the accuracy value of k-NN performance in classifying datasets based on negative and positive sentiments. [15] shows that SMOTE utilizes nearest neighbors and an adjustable

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

amount of oversampling to provide an overfitting solution. On the other hand, as shown below, there is an AUC value of k-NN with SMOTE and without SMOTE.

AUC of k-NN with SMOTE **AUC of k-NN without SMOTE**

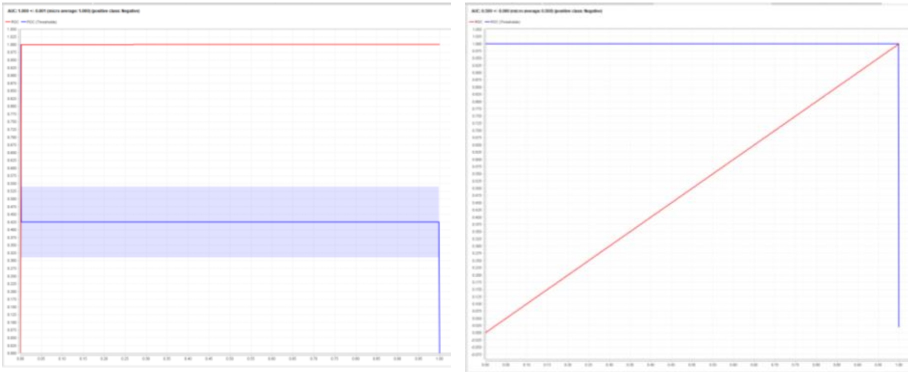

Figure 5. Difference between k-NN with and without SMOTE

Figure 5 shows that the AUC value of k-NN using SMOTE obtained a better accuracy value compared to not using SMOTE. [16] argues that data imbalances can occur when the objects of one data class are more numerous in quantity compared to other courses. The dominant data is referred to as major, while subordinate information is called minor. Meanwhile, an imbalance occurs when the dominant data ignores subordinate data. On the other hand, [17] argues that SMOTE can be integrated with different algorithms to obtain the best classification results. Thus, it can be known that this study's data processing results show the best accuracy values obtained from the integration of SMOTE with the k-NN classification algorithm.

Based on the data processing results, it was found that consumers of restaurant products and services in Labuan Bajo were more on positive sentiment (love, recommend, good, delict, nice, great) compared to negative sentiment. Based on the results of the visualization of the 20 words that most often appear in consumer reviews, cuisine, products (food and beverages), price, services (restaurant services), trip type, and interior and exterior design of the restaurant, and the location of the restaurant are the main highlights of consumers in reviews on the TripAdvisor website, as shown below.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Figure 6. Wordcloud of 20 Top Word on Customer Review

Figure 6 shows that the classification based on the word that most often appears in consumer reviews is cuisine (Italian), products (pizza, pasta, food), services (staff, service, order, time), trip type (friend), price of the product, interior and exterior design (place, restaurant), location (labuan, bajo). Meanwhile, the consumer sentiment expressed in the word is great, love, recommend, good, excellent, and delici. This shows that restaurant business managers can pay attention to essential aspects of cuisine, products, services, price, restaurant conditions, and business locations that match consumer preferences and traveler types.

Overall, in the product context, it can be known that the number of words "food" detected in consumer reviews was 2401, but 2195 were classified as positive sentiments, and 206 were classified as negative sentiments. Furthermore, the word “pizza” detected as many as 1418; there were 1316 words classified as positive sentiment and 102 words classified as negative sentiment. In the context of services, the number of words "servic" appears to be as many as 883 words, with 818 words classified as positive and 65 classified as negative. The same is also shown by the wordlist of the SMOTE and k-NN classification results, which shows that restaurant products and services are the most reviewed by restaurant consumers. This indicates that the products and services of restaurants in Labuan Bajo are reaping positive sentiments from consumers that need to be maintained for business continuity.

The contribution of this research to the development of science emphasizes the intensity of sentiment analysis studies in the tourism industry. Text mining and machine learning have become essential topics popular among researchers these days. This shows that the era of big data demands the adaptive ability of researchers to utilize technology as an effective medium to support the research process. In the context of this study, TripAdvisor website user review data is information that cannot be ignored, responding to the disruptive digital age. Restaurant consumer review data can improve marketing strategies, including efforts to optimize food and beverage services and products in restaurants. So far, researchers have looked more at the popularity of tourism destinations from the perspective of specific businesses such as Micro, Small, and Medium Enterprises (SMEs) without paying attention to micro-scale business processes that are operational centers with different competitiveness. Through a specific study in the restaurant business, this study could decipher the perceptions of restaurant consumers, who are an essential part of the Indonesian tourism industry.

The limitation in this study is that the amount of consumer review data based on restaurants that meet the standards and qualifications are still limited to Happy Banana Komodo, Mediterraneo, La Cucina, and MadeInItaly restaurants. Based on the results of information searches on the TripAdvisor website, restaurants with the number of reviews above 1000 are limited to these four restaurants. In addition, the context of the discussion about restaurants is limited to the popularity of the word in the review, expressed in the form of the word cloud. Recommendations for the development of further research are datasets that can be collected from other sources with the scope of restaurant sentiment analysis in ten priority tourist destinations by emphasizing culinary tourism activities.

4. CONCLUSION

The results of this study show that restaurant consumers have a more dominant positive sentiment towards cuisine, products, services, price, restaurant conditions, and business locations. However, restaurant business managers must consider consumer preferences based on type (travel type). On the other hand, the processing of consumer review data that integrates the SMOTE Upsampling operator with the k-NN algorithm can produce the best accuracy value of 99.27%. The results of the dataset processing also show that overall, the integration of SEMOTE with k-NN results in better accuracy values when compared to NB, DT, and SVM algorithms. Thus, it can be seen that the output of this study can be used as a recommendation for restaurant business managers in Labuan Bajo to optimize products and services to improve the positive image of Labuan Bajo restaurants as an essential part that supports the idea of Indonesia's super-premium tourist destinations.

ACKNOWLEDGEMENT

Thanks to the Faculty of Business Administration and Communication (FIABIKOM), Research Institution and Community Services, Atma Jaya Catholic University of Indonesia, and LLDIKTI III.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

REFERENCES

[1] H. Kennedy, "Perspectives on Sentiment Analysis," J. Broadcast. Electron., vol. 56, no. 4, pp. 435–450, 2012, doi: 10.1080/08838151.2012.732141.

[2] 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.

[3] W. van Atteveldt, M. A. C. G. van der Velden, and M. Boukes, "The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms," Commun. Methods Meas., vol. 15, no. 2, pp. 121–140, 2021, doi: 10.1080/19312458.2020.1869198.

[4] M. N. Imtiaz and M. K. Ben Islam, "Identifying Significance of Product Features on Customer Satisfaction Recognizing Public Sentiment Polarity: Analysis of Smart Phone Industry Using Machine-Learning Approaches," Appl. Artif. Intell., vol. 34, no. 11, pp. 832–848, 2020, doi: 10.1080/08839514.2020.1787676.

[5] Y. A. Singgalen, "Analisis Sentimen dan Pemodelan Topik dalam Optimalisasi Pemasaran Destinasi Pariwisata Prioritas di Indonesia," J. Inf. Syst. Informatics, vol. 4, no. 1, pp. 459–470, 2021, [Online]. Available: http://journal-isi.org/index.php/isi/article/view/171

[6] Y. A. Singgalen, "Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Sistematic Literature Review," J. Inf. Syst. Informatics, vol. 3, no. 2, pp. 278–302, 2021.

[7] N. D. Pratama, Y. A. Sari, and P. P. Adikara, "Analisis Sentimen Pada Review Konsumen Menggunakan Metode Naive Bayes Dengan Seleksi Fitur Chi Square Untuk Rekomendasi Lokasi Makanan Tradisional," J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya, vol. 2, no. 9, pp. 2982–2988, 2018.

[8] P. Sun, D. A. Cárdenas, and R. Harrill, "Chinese Customers' Evaluation of Travel Website Quality: A Decision-Tree Analysis," J. Hosp. Mark. Manag., vol. 25, no. 4, pp. 476–497, 2016, doi: 10.1080/19368623.2015.1037977.

[9] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," J. Inform., vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.

[10] F. Ariani, Amir, N. Alam, and K. Rizal, "Klasifikasi Penetapan Status Karyawan Dengan Menggunakan Metode Naive Bayes," J. Paradig., vol. 20, no. 2, pp. 33–38, 2018, doi: 10.31294/p.v.

[11] Z. Muhammad, R. Rahmadhani, H. Rizqifaluthi, and M. A. Yaqin, "Process Mining Akademik Sekolah menggunakan RapidMiner," Matics, vol. 10, no. 2, pp. 47–51, 2019, doi: 10.18860/mat.v10i2.5158.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[12] D. I. Angkiriwang, K. C. Susanto, and S. Thio, "Pengaruh Ulasan Online Di Tripadvisor Terhadap Minat Menginap Generasi Milenial Di Surabaya," J. Hosp. dan Manaj. Jasa, vol. 6, no. 2, pp. 452–468, 2018, [Online]. Available: http://publication.petra.ac.id/index.php/manajemen-perhotelan/article/view/7517

[13] S. Herayanti, P. A. Wikanatha Sagita, and N. P. Eka Mahadewi, "Faktor-Faktor Yang Mempengaruhi Keputusan Wisatawan Menggunakan Situs Tripadvisor Di Kawasan Pariwisata Kuta Selatan-Bali," J. IPTA, vol. 6, no. 2, pp. 149–155, 2018, doi: 10.24843/ipta.2018.v06.i02.p09.

[14] 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.

[15] Y. E. Kurniawati, "Class Imbalanced Learning Menggunakan Algoritma Synthetic Minority Over-sampling Technique – Nominal (SMOTE-N) pada Dataset Tuberculosis Anak," J. Buana Inform., vol. 10, no. 2, pp. 134–143, 2019, doi: 10.24002/jbi.v10i2.2441.

[16] R. Azmatul Barro, I. D. Sulvianti, and M. Afendi, "Penerapan Synthetic Minority Oversampling Technique (Smote) Terhadap Data Tidak Seimbang Pada Pembuatan Model Komposisi Jamu," Xplore J. Stat., vol. 1, no. 1, pp. 1–6, 2013.

[17] A. N. Rais, "Integrasi SMOTE Dan Ensemble AdaBoost Untuk Mengatasi Imbalance Class Pada Data Bank Direct Marketing," J. Inform., vol. 6, no. 2, pp. 278–285, 2019, doi: 10.31311/ji.v6i2.6186.