Sentiment Analysis of Hotel Reviews Using Logistic Regression and Random Forest Methods

Ni Made Hridayani Karuna[a1], I Made Agus Dwi Suarjaya[a2], Ni Putu Sutramiani[a3]
[a]Department of Information Technology, Udayana University, Indonesia
e-mail: [1]hridayanikaruna@student.unud.ac.id, [2]agussuarjaya@it.unud.ac.id, [3]sutramiani@unud.ac.id

Abstrak

Industri perhotelan adalah bagian penting dari sektor pariwisata. Perkembangan internet mengubah cara pelanggan memilih hotel, dengan ulasan online menjadi acuan utama. Analisis sentimen ulasan membantu memahami preferensi dan kepuasan pelanggan, mendukung strategi berkelanjutan. Penelitian ini membandingkan model Logistic Regression dan Random Forest menggunakan 3000 data training dari 149 hotel di 15 destinasi terbaik versi TripAdvisor versi September 2022. Hasil menunjukkan Logistic Regression lebih unggul dengan akurasi 94,33%, dibandingkan Random Forest 91,33%. Model terbaik digunakan untuk mengklasifikasikan 551.294 ulasan menjadi sentimen positif, negatif, dan netral. Tren sentimen menunjukkan penurunan selama pandemi awal 2020, namun meningkat pascapandemi 2022, didorong kampanye global dan pelestarian ikon wisata. Kuliner dan festival juga menjadi daya tarik utama. Hasil penelitian diharapkan bermanfaat bagi praktisi hotel dan pemerintah dalam menentukan layanan atau kebijakan di sektor perhotelan dan pariwisata.

Kata kunci: *analisis sentimen, ulasan, hotel, logistic regression, random forest*

Abstract

The hospitality industry plays a crucial role in tourism. The internet has changed how customers choose hotels, with online reviews becoming a key reference. Sentiment analysis of reviews helps understand customer preferences and satisfaction, supporting sustainable strategies. This study compares Logistic Regression and Random Forest models using 3,000 training data from 149 hotels in the top 15 destinations on TripAdvisor as of September 2022. Results show Logistic Regression achieved higher accuracy (94.33%) than Random Forest (91.33%), making it the preferred model for classifying 551,294 reviews into positive, negative, or neutral sentiments. Customer sentiment declined during the early 2020 pandemic but improved post-pandemic in 2022, supported by global campaigns and the preservation of tourist attractions. Culinary experiences and festivals also drew visitors. These findings aim to assist hotel practitioners and policymakers in improving services and strategies in the hospitality and tourism sectors.

Keywords: *sentiment analysis, reviews, hotel, logistic regression, random forest*

1. Introduction

The hospitality industry includes hotel businesses that focus on serving customer needs in terms of accommodation, service, food or restaurants, and various other services. The development of the internet and information, quickly allows travelers to search for hotel information before booking to reduce the risk of choosing a bad one. In addition to traditional word of mouth, electronic word of mouth is also increasingly important with the popularity of the internet [1].

Online reviews influence purchasing decisions, customer satisfaction, and customer intentions to return to a hotel. Therefore, understanding customer experiences in online reviews is necessary to develop sustainable hospitality industry strategies [2]. Online travel apps like TripAdvisor allow travelers to share their experiences widely. The American online travel app TripAdvisor offers a wealth of reviews detailing travelers' experiences that can include the performance of a hotel, restaurant, or tourist attraction [2]. With TripAdvisor review data, managers of each hotel can find out and make decisions to improve their hotel facilities or services or create new policies. However, the large amount of review data makes it difficult to understand and summarize all reviews from a hotel and assess whether all hotel reviews tend to be positive, neutral or negative sentiment.

In today's technological era, sentiment assessment can be done with a data mining approach using text classification in the form of sentiment analysis based on data obtained from website sources or social media about customer impressions after using hotel services. Based on previous related research on text classification, text classification on the BBC news dataset compares the Logistic Regression, Random Forest, and K-Nearest Neighbor algorithms. This study obtained the results of a classification comparison where the model using the Logistic Regression algorithm with TF-IDF achieved the highest accuracy of 97%. The second best is Random Forest with an accuracy of 93% [3]. In addition, in the research on classifying product review sentiment using the Bernoulli Naïve Bayes, Multinomial Naïve Bayes, Random Forest and Logistic Regression algorithms, the results showed that the highest accuracy was obtained by Random Forest at 93.17%, followed by Logistic Regression at 90.88%, Multinomial Naïve Bayes at 87.09%, and Bernoulli Naïve Bayes at 86.46% [4]. Based on the related research, it was found that the use of Logistic Regression and Random Forest algorithms obtained good accuracy in the classification of review sentiment. Therefore, this study uses a classification model by comparing the Logistic Regression method with Random Forest on an online review dataset of 149 hotels in 15 tourist destinations with positive, negative, and neutral review sentiment classes. The difference with previous studies is that it uses an online hotel review dataset that reflects the need to test the model in different contexts, with variations and unique characteristics of the online hotel review dataset. In addition, previous studies focused on sentiment classification with two classes, namely positive and negative, while this study involves three sentiment classes, namely positive, negative, and neutral.

Based on the background that has been explained, this study was submitted with the intention of knowing the trend of customer sentiment towards hospitality based on sentiment analysis by comparing the Logistic Regression and Random Forest algorithms in classification. The results of the study are expected to be used as considerations by hotel practitioners or the government in determining services or policies in doing business in the hospitality and tourism sectors.

2. Research Method

The stages of the research conducted are listed in Figure 1.

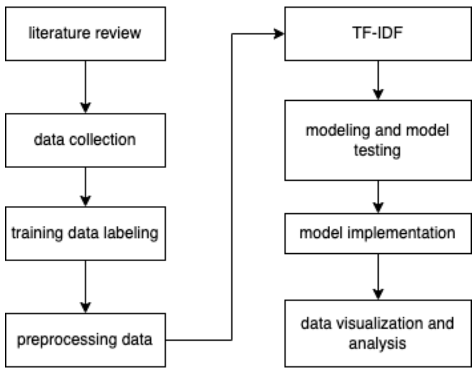

Figure 1. Research Methodology

2.1. Data Collection

The stages of collecting hotel review data were carried out on the TripAdvisor website using web scraping techniques using the Python libraries Selenium and BeautifulSoup. All data scraped was English language review data. The review data collected were ratings, review titles, review content, stay dates, reviewer countries, and trip types. The data collected were the top 10 hotels in the 14 best destinations from TripAdvisor in September 2022, and the top 9 hotels in the New Delhi destination with a total of 149 hotels and a total of 551,294 reviews as of September 2023.

2.2. Labeling Training Data

Labeling is a stage carried out to obtain training data for the classification model through manual labeling. Labeling of training data is carried out into three classification classes, namely positive with a sentiment label value of 2, neutral with a sentiment label value of 1, and negative with a sentiment label value of 0. The total number of training data labeled manually is 1000 data for each sentiment, so the total training data is 3000 data.

2.3. Preprocessing Data

The preparation process carried out on training data before performing TF-IDF and classification using a classification algorithm is called preprocessing or data preparation. Data preprocessing is carried out in five stages. The first stage is data cleansing, which is cleaning the review text from special characters, emoticons, and URLs. The next stage is case folding, which equalizes the review text format to lowercase. The third stage is tokenizing, which is changing text into tokens, followed by the stopword removal process which removes words that often appear in text that do not contain important words [5]. The final process is lemmatization, which is the process of simplifying words into dictionary form.

2.4. TF-IDF

Term Frequency - Inverse Document Frequency (TF-IDF) is an evaluation technique used to assess the significance of words in a document [6]. The research uses TF-IDF calculations in the Python library, namely Sklearn, using the TF-IDF formula calculated in Equation 1.

$$TF-IDF=tf_{(x,y)}+\log\left(\frac{Total\,Documents}{df_x}\right)+1 \qquad (1)$$

Note:
$tf_{(x,y)}$ = Frequency of word x in document y
df_x = Number of documents containing the word x

2.5. Logistic Regression

Logistic Regression is a method commonly used to make predictions by using probability as its basis [7]. Logistic Regression is an extension of Linear Regression which is used to solve classification problems. The following is the logistic function or sigmoid function which is the core function of the Logistic Regression method.

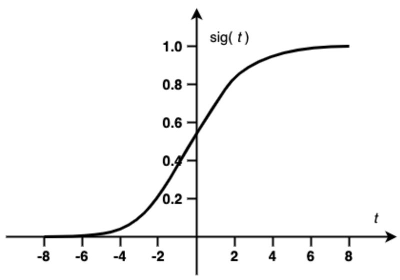

Figure 2. Sigmoid Function

The formula for the sigmoid function is defined in Equation 2.

$$Sig(t)=\frac{1}{1+e^{-t}} \qquad (2)$$

Note:
Sig(t) = The result of the sigmoid function,

t = Input
e = Natural logarithm base

Logistic Regression assigns probabilities using the Sigmoid function, which converts a numerical result into a probability between 0 and 1. Logistic Regression can make binary predictions by dividing the population into two groups, for example with a threshold of 0.5. Anything above 0.5 is considered to belong to group A, and anything below is considered to belong to group B [8]. The following is a description of the binary Logistic Regression prediction with a limit of 0.5.

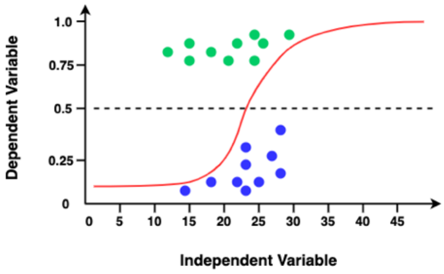

Figure 3. Logistic Regression

The equation representing Logistic Regression is calculated using Equation 3.

$$y = \frac{e^{(b_0 + b_1 X)}}{1 + e^{(b_0 + b_1 X)}} \qquad (3)$$

Note:
y = prediction of results (output)
X = Input
b_0 = bias or *intercept*
b_1 = coefficient X

In this study, we used Multinomial Logistic Regression, namely the dependent variable is more than two categories, namely positive, neutral and negative classes.

2.6. Random Forest

Random Forest is an algorithm that operates on the principle of using a number of Decision Trees. The following is a picture of the tree structure hierarchy of the Decision Tree.

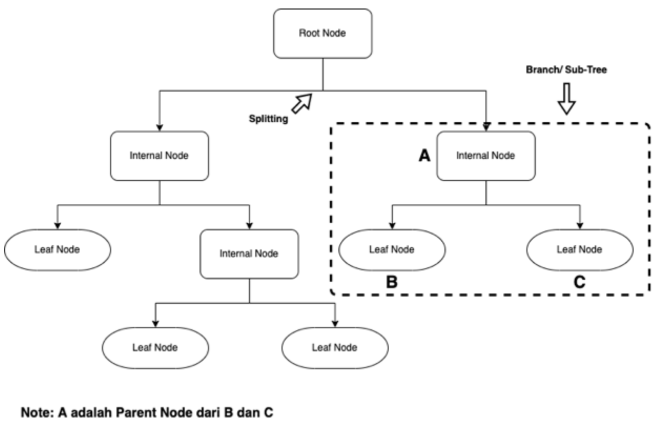

Figure 4. Decision Tree

Decision Tree starts from the root node without any previous branch input. Subsequent branches split into sub-trees and reach decision nodes for evaluation. Decision nodes produce homogeneous subsets based on features, denoted by leaf nodes that represent possible outcomes of the dataset. Decision Tree is prone to overfitting. Random Forest, as an ensemble learning, is effective in overcoming it by combining multiple algorithms to improve predictive performance [9]. Random Forest combines predictions from each Decision Tree and uses majority vote to determine the majority class and is set as the final prediction. Its characteristic is the random nature, especially in random sampling and feature selection when building the tree [10]. The overview of Random Forest is explained as in Figure 5.

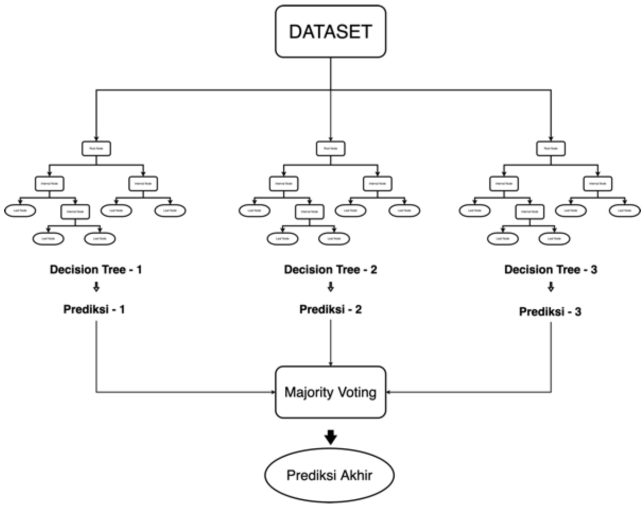

Figure 5. Random Forest

2.7. K-Fold Cross Validation

The technique of evaluating machine learning models by training several models on different subsets of input data is called Cross Validation. This technique helps detect overfitting, where the model is not effective in generalizing patterns to new data. One method of Cross Validation is K-Fold Cross Validation, where the data is divided into K equal parts to reduce bias in the dataset [11]. Training and testing data is carried out as many as K, for example in 5-Fold Cross Validation. Each fold is used as a testing set at one point in the process, which is depicted in Figure 6.

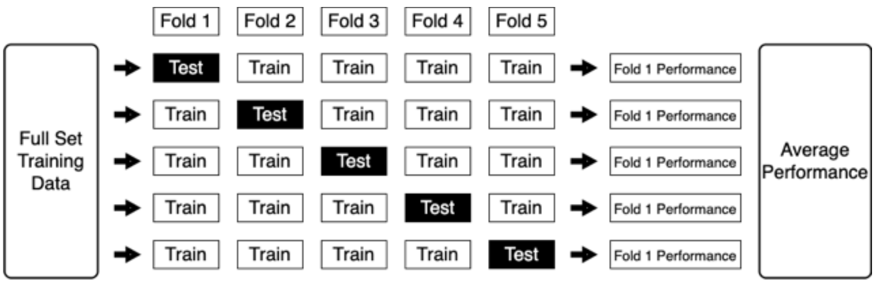

Figure 6. 5-Fold Validation

2.8. Confusion Matrix

Confusion matrix is an evaluation tool that describes the predicted results of an algorithm with the actual results [12]. Confusion matrix is a table that displays four evaluation matrices, the first is true positive, the actual value is positive, and the model predicts positive. Next is true negative, when the actual value is negative, and the model predicts negative. Third is false positive, the predicted result is positive, and the actual value is negative, so the model's predicted result is wrong. Last is false negative, the actual value is positive, and the model predicts negative. The structure of the 2x2 confusion matrix is depicted in Figure 7.

Table 1. Confusion Matrix

	Predicted Value	
	Negative	Positive

Actual Values	Negative	*True Negative* (TN)	*False Negative* (FN)
	Positive	*False Positive* (FP)	*True Positive* (TP)

- Accuracy is a metric that measures the accuracy of a model in making correct predictions on all test data.

$$accuracy=\frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

- Precision measures the ratio of correct predictions of predicted labels to the total amount of labelled data. Precision determines the reliability of the model, especially if false positive errors are a major concern.

$$precision=\frac{TP}{TP+FP} \tag{5}$$

- Recall is a method for measuring the extent to which a model is successful in correctly predicting the actual class.

$$recall=\frac{TP}{TP+FN} \tag{6}$$

- F1-Score is a calculation used to compare the average precision and recall values.

$$f1-score=\frac{2*precision*recall}{precision+recall} \tag{7}$$

3. Results and Discussion

3.1. Training Data Labelling

The training data was manually labelled and resulted in labels of 1000 positive reviews, 1000 neutral reviews, and 1000 negative reviews. Label information can be seen in Table 2.

Table 2. Training Data Labelling

review	sentiment label
The view and environment is outstanding. The activities are enjoyable like golf, archery, aerial yoga, jogging, and others. The staff are so friendly, like Mr. Agus who teached me golf and archery with kindness and patience.	2
Location is excellent. Food is also good.Corridors are not as good. There's a smell of smoke all around the corridors. Lobby is acoustically poor and hence there's lot of noise from the restaurant adjacent to lobby. Gymnasium is very good. Room is OK type.	1
AWFUL experience, this is NOT a 5 star resort. Traveling with my family of 12 people and 3 toddlers— we ALL experienced the same problems. This place is absolutely HORRIFIC. STAY AWAY! Dirty and rude	0

3.2. Preprocessing

The text preprocessing steps performed include data cleansing, case folding, tokenizing, stopword removal, and lemmatization. In data cleansing, the review text is cleaned from URLs, emoticons, and special characters. The data cleansing process is described in Table 3.

Table 3. Cleansing Data Process

Before Cleansing Data	After Cleansing Data
Never stay here again **!!!!!!!!!** The staff here don**'**t understand basic English**,** breakfast tables were not cleaned **.** The A**/**c the the rooms were not cool **,** house keeping were not thorough**,** food ordered through room service was below standard **.....**😡😡 I will never stay here again **....**	Never stay here again The staff here dont understand basic English breakfast tables were not cleaned The Ac the the rooms were not cool house keeping were not thorough food ordered through room service was below standard I will never stay here again

The second stage is case folding, which is changing capital letters to lower case letters, thus creating uniformity in writing words [5]. The case folding process is depicted in Table 4.

Table 4. Case Folding Process

After Cleansing Data	After Case Folding
Never stay here again The staff here dont understand basic **English** breakfast tables were not cleaned **The Ac** the the rooms were not cool house keeping were not thorough food ordered through room service was below standard **I** will never stay here again	never stay here again the staff here dont understand basic english breakfast tables were not cleaned the ac the the rooms were not cool house keeping were not thorough food ordered through room service was below standard i will never stay here again

The third step is tokenizing, where the review sentences are divided into smaller units and separating characters such as spaces, new lines, and so on are removed [13]. The tokenizing process is depicted in Table 5.

Table 5. Tokenizing Process

After Case Folding	After Tokenizing
never stay here again the staff here dont understand basic english breakfast tables were not cleaned the ac the the rooms were not cool house keeping were not thorough food ordered through room service was below standard i will never stay here again	['never', 'stay', 'here', 'again', 'the', 'staff', 'here', 'dont', 'understand', 'basic', 'english', 'breakfast', 'tables', 'were', 'not', 'cleaned', 'the', 'ac', 'the', 'the', 'rooms', 'were', 'not', 'cool', 'house', 'keeping', 'were', 'not', 'thorough', 'food', 'ordered', 'through', 'room', 'service', 'was', 'below', 'standard', 'i', 'will', 'never', 'stay', 'here', 'again']

The fourth stage is stopword removal, which is removing words that do not have an important contribution [5]. These words were removed because they were considered to have no relevance or important information in the text. The stopword removal process is illustrated in Table 6.

Table 6. Stopword Removal Process

After Tokenizing	After Stopword Removal
['never', 'stay', **'here'**, **'again'**, **'the'**, 'staff', **'here'**, 'dont', 'understand', 'basic', 'english', 'breakfast', 'tables', **'were'**, **'not'**, 'cleaned', **'the'**, 'ac', **'the'**, **'the'**, 'rooms', **'were'**, **'not'**, 'cool', 'house', 'keeping', **'were'**, **'not'**, 'thorough', 'food', 'ordered', **'through'**, 'room', 'service', **'was'**, **'below'**, 'standard', **'i'**, **'will'**, 'never', 'stay', **'here'**, **'again'**]	['never', 'stay', 'staff', 'dont', 'understand', 'basic', 'english', 'breakfast', 'tables', 'cleaned', 'ac', 'rooms', 'cool', 'house', 'keeping', 'thorough', 'food', 'ordered', 'room', 'service', 'standard', 'never', 'stay']

The final stage is lemmatization, which is a process that aims to return words to their basic form or dictionary form to provide a consistent representation of words [14]. The lemmatization process is depicted in table 7.

Table 7. Lemmatization Process

After Stopword Removal	After Lemmatization
['never', 'stay', 'staff', 'dont', 'understand', 'basic', 'english', 'breakfast', **'tables'**, **'cleaned'**, 'ac', **'rooms'**, 'cool', 'house', **'keeping'**, 'thorough', 'food', **'ordered'**, 'room', 'service', 'standard', 'never', 'stay']	['never', 'stay', 'staff', 'dont', 'understand', 'basic', 'english', 'breakfast', 'table', 'clean', 'ac', 'room', 'cool', 'house', 'keep', 'thorough', 'food', 'order', 'room', 'service', 'standard', 'never', 'stay']

3.3. Model Testing

The testing of the Logistic Regression and Random Forest classification algorithm models was carried out through the K-Fold Cross Validation process and four testing scenarios, namely the train data and test data splitting process. Figures 7 and 8 are the output of the K-Fold Cross Validation test with K = 5 on 3000 training data.

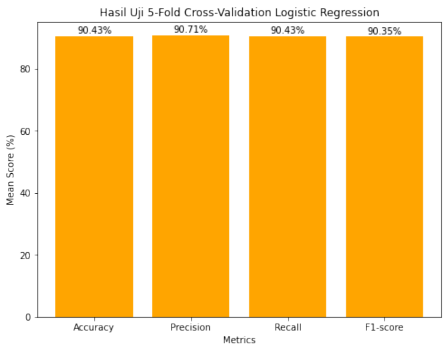

Figure 7. 5-Fold Cross Validation Logistic Regression Test Results

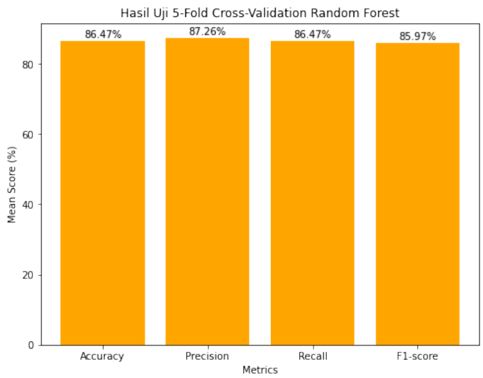

Figure 8. 5-Fold Cross Validation Random Forest Test Results

Comparison of the test results of the Logistic Regression (LR) algorithm model with the Random Forest (RF) algorithm model is shown in Table 2. Based on Table 2, the labeling of 553,152 review data will use the Logistic Regression model which is determined as the best classification model.

Table 8. Comparison of Algorithm Testing Results

	skenario	***accuracy***	***precision***	***recall***	***f1-score***
LR	90:10	94.33%	94.35%	94.33%	94.32%
	80:20	92.67%	92.72%	92.67%	92.68%
	70:30	92.78%	92.78%	92.78%	92.78%
	60:40	93.25%	93.29%	93.25%	93.26%
RF	90:10	91.33%	91.60%	91.33%	91.27%
	80:20	89.50%	89.59%	89.50%	89.38%
	70:30	89.67%	89.95%	89.67%	89.54%
	60:40	88.50%	88.72%	88.50%	88.25%

The 3x3 Confusion Matrix test was run on Logistic Regression and Random Forest with a 90:10 scenario (90% training data, 10% testing data), with 300 testing data reviews and 2700 training data. Figure 9 shows the results of the Confusion Matrix test on Logistic Regression and Random Forest.

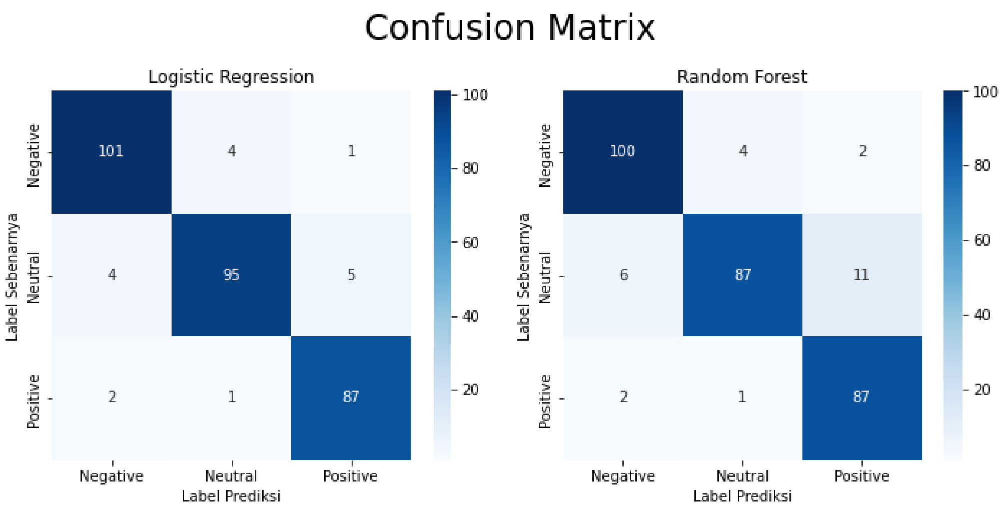

Figure 9. Confusion Matrix Comparison Scenario 90:10

In terms of prediction accuracy, the model using Logistic Regression has better accuracy when predicting neutral and negative label classes and is equally good when predicting Negative labels with the Random Forest model.

3.4. Data Visualization and Analysis

Trend visualization using line charts is effective in showing changes in variable values over time. Visualization helps analyze review sentiment patterns and guest travel types. The result is a visualization of sentiment trends for 149 hotels in TripAdvisor's top 15 destinations. Figure 10 shows a visualization of sentiment trends for the top 10 hotels in Dubai, the top favorite destination for reviewers in September 2022 on the TripAdvisor website.

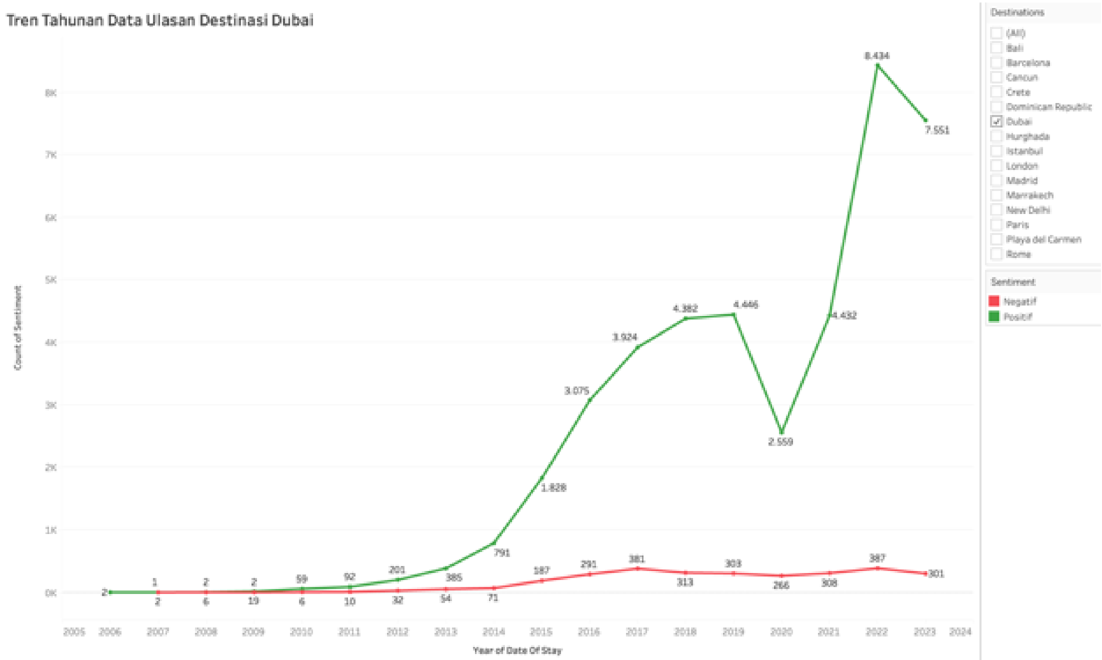

Figure 10. Top 10 Dubai Hotel Reviews Data Trend Visualization

Based on the annual sentiment trend, Dubai saw a decline in reviews in 2020 due to the Covid-19 pandemic. However, after the new normal in 2021, positive reviews increased and peaked in 2022. Factors such as travel safety, Expo 2020 Dubai, a rich culinary sector, cruise tourism, and global campaigns are the reasons for the increase in positive sentiment. Dubai also launched several new attractions, including the Museum of The Future, which attracted tourists. Information on the peak and low season months of reviewers who stayed in the top 15 destinations can be seen in Table 9.

Table 9. Reviewer's High and Low Seasons in Top 15 Destinations

Destination	Peak and Low Season
Dubai	Peak season months: July - December Low season months: January - June
Bali	Peak season months: June - December Low season months: January - May
London	Peak season months: August - October Low season months: March - May
Rome	Peak season months: May, July - October Low season months: November - February
Paris	Peak season months: April - October Low season months: November - March
Cancun	Peak season months: November - March Low season months: June - September
Crete	Peak season months: April - October Low season months: November - March
Marrakech	Peak season months: May - June, September - October Low season months: November - February
Dominican Republic	Peak season months: March, June - July, November - December Low season months: September - October
Istanbul	Peak season months: May - December Low season months: January - April
Playa del	Peak season months: December - August

Carmen	Low season months: September - November
Barcelona	Peak season months: May - September Low season months: November - February
New Delhi	Peak season months: August - February Low season months: March - June
Hurghada	Peak season months: March, July - October Low season months: May, December - January
Madrid	Peak season months: September - October Low season months: November - February

Information on the types of guests with certain types of travel who stayed at the top 15 destinations can be seen in Table 10. There are five types of travel, namely, TFM (Traveled with family), TC (Traveled as a couple), TB (Traveled on business), TFR (Traveled with friends), and TS (Traveled solo).

Table 10. Top 5 Types of Customer Journeys in Each Destination

Destination	Top 5 Types of Customer Journeys
Dubai	TFM, TC, TB, TFR, TS
Bali	TC, TFM, TFR, TS, TB
London	TC, TFM, TB, TFR, TS
Rome	TC, TFM, TFR, TS, TB
Paris	TC, TFM, TFR, TS, TB
Cancun	TC, TFM, TFR, TB, TS
Crete	TC, TFM, TFR, TB, TS
Marrakech	TFM, TC, TFR, TS, TB
Dominican Republic	TC, TFM, TFR, TB, TS
Istanbul	TC, TFM, TB, TFR, TS
Playa del Carmen	TC, TFM, TFR, TB, TS
Barcelona	TC, TFM, TFR, TB, TS
New Delhi	TB, TFM, TC, TFR, TS
Hurghada	TFM, TC, TFR, TS, TB
Madrid	TC, TFM, TB, TFR, TS

Information regarding the top 15 reviewer countries that stayed in the top 15 destinations can be seen in Table 11.

Table 11. Top 15 Reviewer Countries in Top 15 Destinations

Destination	Top 15 Negara Pengulas
Dubai	UK, UAE, USA, Australia, India, Saudi Arabia, Germany, Switzerland, Canada, France, Kuwait, Netherlands, Italy, South Africa, Ireland
Bali	Australia, Indonesia, UK, USA, Singapore, India, New Zealand, Canada, China, UAE, Malaysia, Germany, Netherlands, South Korea, dan Switzerland
London	UK, USA, Australia, Canada, Ireland, Spain, Italy, Germany, India, France, Switzerland, Singapore, Netherlands, China, UAE
Rome	UK, Australia, Canada, Ireland, Israel, Italy, Singapore, Netherlands, Belgium, New Zealand, France, Germany, India, Norway
Paris	USA, UK, Australia, Canada, France, Israel, Germany, Ireland, India, Netherlands, Switzerland, Belgium, Italy, Singapore, Spain
Cancun	USA, Canada, UK, Mexico, Australia, Ireland, Germany, Spain, France, Switzerland, Italy, Israel, Colombia, Netherlands, Sweden
Crete	UK, Israel, USA, Greece, Germany, Belgium, Switzerland, Romania, France,

	Netherlands, Poland, Ireland, Italy, Czech Republic, Canada
Marrakech	UK, USA, Morocco, Ireland, Canada, Australia, Spain, France, Netherlands, Belgium, Germany, Portugal, Switzerland, Italy, Denmark
Dominican Republic	USA, Canada, UK, Dominican Republic, Mexico, Germany, Spain, Switzerland, Colombia, France, Ireland, Belgium, Italy, Portugal, Netherlands
Istanbul	USA, UK, Turkey, Australia, Canada, UAE, Saudi Arabia, Iran, India, Germany, South Africa, France, Romania, Russia, Netherlands
Playa del Carmen	USA, Canada, UK, Mexico, Ireland, Australia, Germany, Spain, Switzerland, France, Netherlands, Italy, Belgium, Portugal, Argentina
Barcelona	USA, UK, Australia, Canada, Spain, Ireland, Israel, Netherlands, Norway, Switzerland, France, Belgium, Germany, Sweden, Singapore
New Delhi	India, UK, USA, Australia, Canada, UAE, Singapore, Germany, New Zealand, Italy, France, Malaysia, Thailand, Spain, South Africa
Hurghada	UK, Egypt, Belgium, Romania, USA, Netherlands, Serbia, Germany, Czech Republic, Poland, Switzerland, UAE, Denmark, Saudi Arabia, Canada
Madrid	USA, UK, Australia, Spain, Canada, Israel, Ireland, France, Switzerland, Italy, Netherlands, Germany, UAE, Belgium, China

Negative sentiment word cloud information in 15 destinations, frequently discussed words are ‘hotel’, ‘room’, ‘service’, ‘staff’, ‘night’, ‘bed’, ‘book’, ‘reception’, ‘food’, ‘charge’, ‘location’, ‘manager’, and other words listed in each word cloud. Negative reviews for hotels due to small rooms or dirty bathrooms, smelly, lack of repairs, hard mattresses that are uncomfortable, disappointing service, unfriendly and poor service from staff such as service by reception staff for check-in or check-out that wastes time, takes time when calling service staff, inappropriate hotel bookings, noisy guests, loud music that disturbs other guests to sleep at night, breakfast or food at the hotel restaurant that is not hygienic, not varied, or not tasty, the hotel location is not strategic such as far from tourist attractions, supermarkets, airports, or metro, hotel managers who do not help hotel customer problems, and other things that disappoint customers that influence the giving of negative sentiment. An illustration of the negative sentiment word cloud for the Barcelona and Hurghada destinations is depicted in Figure 11.

Figure 11. Word Cloud Negative Sentiment Destination Barcelona & Hurghada

4. Conclusion

This study predicts sentiment and analyses 551,294 review data on 149 hotels in 15 destinations from the TripAdvisor website. Review data was scraped until September 2023 using python and the Logistic Regression classification model built obtained an evaluation result of 94.33% accuracy, 94.35% precision, 94.33% recall, and 94.32% F1-Score which is the best model compared to Random Forest with 91.33% accuracy, 91.60% precision, 91.33% recall, and 91.27% F1-Score. All destinations excelled in positive sentiment and sentiment trends in 15 destinations were concluded to have decreased during the pandemic in early 2020. The majority of destinations experienced an increase in positive sentiment after the pandemic in 2022 due to conducting global campaigns in collaboration with regional and international celebrities, influencers, and community leaders highlighting the unique attractions of Dubai, the easing of various policies such as quarantine-free in Bali, and the maintenance of iconic places, monuments, museums, beaches, to maintain their beauty. In addition, culinary and festivals are also attractions for tourists from each destination. The types of travel of tourists in 15 destinations were concluded, Dubai, Marrakech, Hurghada, and Madrid were dominated by tourists traveling with families. Bali, London, Rome, Paris, Cancun, Crete, Dominican Republic, Istanbul, Playa del Carmen, and Barcelona were dominated by tourists traveling with couples, while New Delhi was dominated by tourists traveling for business. Further research can perform more optimal labelling of training data in positive, neutral, and negative sentiments to improve model accuracy and explore hyperparameters of each classification algorithm used. In addition, further research can use more varied datasets such as variations in review languages and variations in data sources not only from TripAdvisor website review data but also adding variations in data from other websites such as Traveloka, or Google Reviews.

Daftar Pustaka

[1] J. Park, H. Hyun, and T. Thavisay, 'A study of antecedents and outcomes of social media WOM towards luxury brand purchase intention', *Journal of Retailing and Consumer Services*, vol. 58, p. 102272, Jan. 2021, doi: 10.1016/j.jretconser.2020.102272.

[2] Y. J. Kim and H. S. Kim, 'The Impact of Hotel Customer Experience on Customer Satisfaction through Online Reviews', *Sustainability (Switzerland)*, vol. 14, no. 2, Jan. 2022, doi: 10.3390/su14020848.

[3] K. Shah, H. Patel, D. Sanghvi, and M. Shah, 'A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification', *Augmented Human Research*, vol. 5, no. 1, p. 12, Dec. 2020, doi: 10.1007/s41133-020-00032-0.

[4] B. K. Shah, A. K. Jaiswal, A. Shroff, A. K. Dixit, O. N. Kushwaha, and N. K. Shah, 'Sentiments Detection for Amazon Product Review', in *2021 International Conference on Computer Communication and Informatics, ICCCI 2021*, Institute of Electrical and Electronics Engineers Inc., Jan. 2021. doi: 10.1109/ICCCI50826.2021.9402414.

[5] I. Kurniawan and A. Susanto, 'Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019', *Eksplora Informatika*, vol. 9, no. 1, pp. 1–10, Sep. 2019, doi: 10.30864/eksplora.v9i1.237.

[6] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, 'The impact of features extraction on the sentiment analysis', in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 341–348. doi: 10.1016/j.procs.2019.05.008.

[7] H. R. Alhakiem and E. B. Setiawan, 'Aspect - Based Sentiment Analysis on Twitter Using Logistic Regression with FastText Feature Expansion', *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 5, pp. 840–846, Nov. 2022, doi: 10.29207/resti.v6i5.4429.

[8] S. Jessica, 'How Does Logistic Regression Work?', KDnuggets. Accessed: Feb. 03, 2024. [Online]. Available: https://www.kdnuggets.com/2022/07/logistic-regression-work.html

[9] N. S. Chauhan, 'Decision Tree Algorithm, Explained', KDnuggets. Accessed: Feb. 02, 2024. [Online]. Available: https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html

[10] G. Khanvilkar and Prof. Deepali Vora, 'Sentiment Analysis for Product Recommendation Using Random Forest', *International Journal of Engineering & Technology*, vol. 7, no. 3.3, p. 87, Jun. 2018, doi: 10.14419/ijet.v7i3.3.14492.

[11] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, 'Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung', *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 81–86, Aug. 2020, doi: 10.33096/ilkom.v12i2.507.81-86.

[12] D. Valero-Carreras, J. Alcaraz, and M. Landete, ‘Comparing two SVM models through different metrics based on the confusion matrix’, *Comput Oper Res*, vol. 152, p. 106131, Apr. 2023, doi: 10.1016/j.cor.2022.106131.
[13] R. I. Pristiyanti, M. A. Fauzi, and L. Muflikhah, ‘Sentiment Analysis Peringkasan Review Film Menggunakan Metode Information Gain dan K-Nearest Neighbor’, 2018. [Online]. Available: http://j-ptiik.ub.ac.id
[14] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotae, ‘Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset’, *Expert Syst Appl*, vol. 212, p. 118715, Feb. 2023, doi: 10.1016/j.eswa.2022.118715.