

Predictive Analysis of TripAdvisor Travelers' Choice Award Recognition Using Machine Learning Algorithms: The Case of Praia Dos Milionários, Brazil

Diego Guerrero-Vargas^{1,*}, Cristina Flores-Amador¹, Ernesto Bolaños-Rodríguez¹ and Marcelino Castillo-Nechar²

¹Universidad Autónoma del Estado de Hidalgo, Mexico

²Universidad Autónoma del Estado de México, Mexico

Abstract

The recognition of tourist destinations on digital platforms has become a key factor in their competitive positioning. This study aims to analyze the factors that influence recognition in TripAdvisor's Travelers' Choice Awards, as well as to apply machine learning (ML) algorithms to predict which tourist destinations will be included in this distinction. A mixed-methods approach is employed with a non-experimental cross-sectional research design, based on a representative sample of 368 domestic and foreign beachgoers at Praia Dos Milionários, Brazil. In the first phase, the Chi-square statistical test is applied to identify significant relationships between variables associated with perception, quality, and the tourism experience. Subsequently, machine learning models including logistic regression, decision trees, and Random Forest are implemented, and their predictive performance is evaluated. The results reveal the existence of interrelated factors that significantly influence the destination's recognition, with the Random Forest model standing out as the best performer in terms of accuracy. Furthermore, differences in perception between domestic and foreign tourists are identified, highlighting the importance of differentiated approaches in the management of tourist destinations. It is concluded that the application of ML techniques improves the prediction of tourist recognition and provides strategic tools for decision-making in destination management and promotion.

Keywords

Tourism Analytics, Machine Learning, Tourism Destination Forecasting, TripAdvisor, Tourism

1. Introduction

Tourism has established itself as one of the most significant global activities, due to its ability to influence regional development, job creation, and the international visibility of destinations [1]. Furthermore, tourism extends beyond physical travel and constitutes a complex activity involving interaction with diverse ways of life, local economies, and the enjoyment of attractions [2].

From this perspective, sun and beach tourism is recognized as a predominant activity consisting of recreation, entertainment, or relaxation on beaches due to the combined presence of water, sun, and warmth. At the same time, it includes practices such as surfing, windsurfing, diving, and recreational activities and sports [3]. Thus, Brazil's beaches constitute one of the main strategic pillars of the national tourism sector, both due to their geographical extent and their economic, sociocultural, and environmental significance. In this context, Praia Dos Milionários, located in the municipality of Ilhéus, in the state of Bahia, is a destination frequented by both domestic and international tourists [4].

However, the growth of online review and rating platforms has transformed the way travelers select their destinations, making user-shared experiences a key factor in choosing tourist attractions. In this regard, visitor perception has established itself as a fundamental variable for tourism competitiveness, directly influencing decision-making. In an analysis of research on Electronic Word of Mouth (eWOM) in

IWSAI 2026: 1st International Workshop on Systems, Applications, Artificial Intelligence and Innovation, September 16–18, 2026, Piura, Peru

*Corresponding author.

✉ gu440684@uaeh.edu.mx (D. Guerrero-Vargas)

ORCID: 0009-0004-5554-2605 (D. Guerrero-Vargas); 0000-0001-8122-3094 (C. Flores-Amador); 0000-0002-1432-7720 (E. Bolaños-Rodríguez); 0000-0003-4894-8813 (M. Castillo-Nechar)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

tourism [5], it is noted that digital reputation has become a strategic factor for destination management, as it directly influences perceptions of quality and consumer choice.

Among these platforms, TripAdvisor stands out as one of the most influential globally, as it aggregates millions of user reviews and ratings on destinations, services, and tourist attractions. Based on this information, the platform awards international recognition. These not only impact the destination's image but also influence its positioning, demand, and competitiveness.

In this context, the Travelers' Choice Awards take on particular significance. These awards recognize destinations, establishments, and experiences that stand out for the quality and quantity of positive reviews received over a 12-month evaluation period, as well as through a complementary editorial process implemented by the platform itself. Based on user experience and ratings, these destinations, establishments, and experiences rank among the best internationally [6].

From this perspective, Praia Dos Milionários' recognition can be understood as the result of a complex interplay of variables. Among these are factors associated with the visitor's profile, such as nationality and prior expectations, which shape the subjective evaluation of the destination. Added to this are intrinsic attributes of the location, such as cleanliness, safety, and infrastructure, which directly influence perceived satisfaction levels. In this context, it is pertinent to analyze beachgoers' perceptions and their relationship to the likelihood that this destination will be considered worthy of the Travelers' Choice Award.

Given this scenario, the application of Machine Learning (ML) techniques emerges as an effective tool for identifying patterns, forecasting, and classifying trends based on large volumes of data, thereby improving predictive capabilities.

Therefore, the tourism experience is shaped by the interaction between the services offered at the destination and exogenous factors, such as the seasonality of demand, which influence visitor perception and determine satisfaction levels, the generation of positive reviews, and, consequently, the likelihood of recognition on platforms such as TripAdvisor. The convergence of these elements highlights the multifactorial nature of tourism recognition, which justifies the use of statistical techniques for association analysis, as well as machine learning models aimed at identifying patterns and predicting such recognition.

Based on the foregoing, the overall objective of this research is to analyze the Travelers' Choice Awards for the case study of Praia Dos Milionários, Brazil, using the indicators established by the platform and supervised ML algorithms, with the aim of identifying the characteristics of this tourist attraction that make it eligible for inclusion in this award.

Based on this objective, the following research hypothesis is established:

The nationality of beachgoers visiting Praia Dos Milionários, whether Brazilian or foreign, is significantly associated with their perception of whether the destination deserves TripAdvisor's Travelers' Choice Award recognition.

Finally, the contributions of this study consist of the application of machine learning (ML) models to analyze and predict the potential inclusion of Praia Dos Milionários in the Travelers' Choice Awards, as well as the generation of empirical evidence regarding the relationship between visitor perception and tourism recognition on digital platforms. Furthermore, the research provides an analytical perspective on the influence of online reviews on the construction of tourist destinations' reputations.

2. Methodology

2.1. Machine Learning

Machine Learning (ML) is an application of AI used to extract knowledge from data through learning. For this reason, the algorithms used in ML are classified as supervised, unsupervised, semi-supervised, reinforcement learning, deep learning, rule-based, and hybrid. Supervised algorithms are those trained using labeled data; examples include decision trees and linear regression. On the other hand, unsupervised algorithms use unstructured data to learn patterns [7].

Supervised learning is characterized by the use of pre-labeled data, which allows for the training of models capable of performing classification and prediction tasks. This approach is widely used in studies seeking to identify relationships between independent variables and a clearly defined dependent variable [8].

In this research, supervised learning is relevant because we have structured responses from the questionnaire administered to beachgoers, as well as a target variable related to the perception of deserving the Travelers' Choice Award. Supervised ML algorithms are used because there is an objective outcome, where data has clear labels (rating from 1 to 5, number of reviews, business age, frequency of reviews, text sentiment, location, type of attraction), and learning occurs through comparison with real-world cases.

Based on the supervised learning approach, the following algorithms are selected, which are described below [9]:

Logistic regression is a statistical model widely used in binary classification problems due to its ease of interpretation and computational efficiency. This algorithm allows for estimating the probability of an event occurring based on explanatory variables.

$$LogLoss = -\frac{1}{n} \sum_{i=1}^n [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (1)$$

In this study, logistic regression is used as the base model to analyze the relationship between beachgoers' characteristics and their perception of whether Praia Dos Milionários deserves the Travelers' Choice Award.

Decision trees are predictive models that simulate decision-making processes using hierarchical structures of nodes and branches. Their main advantage lies in their interpretability, as they allow for a clear visualization of the rules that lead to a specific classification [10].

$$Entropa = - \sum_i p_i \log(p_i) \quad (2)$$

This algorithm is used to identify the factors that most influence beachgoers' perceptions, making it easier to understand the results obtained.

Finally, Random Forest is applied; this is an ensemble algorithm that combines multiple decision trees with the aim of improving the model's accuracy and reducing the risk of overfitting. This approach has demonstrated superior performance in classification tasks compared to individual models [11].

In this study, Random Forest enables more robust predictions regarding the probability that Praia Dos Milionários will be deemed worthy of tourism recognition, making it the model with the best overall performance.

The most commonly used metrics to evaluate the performance of supervised ML algorithms are: Accuracy, Precision, Recall, and F1 Score. These metrics are calculated from the confusion matrix:

- Accuracy: Represents the proportion of correct predictions out of the total, useful when classes are balanced.
- Precision: Measures of how reliable positive predictions are.
- Recall: The model's ability to correctly identify positive cases.
- F1 Score: A harmonic mean between precision and sensitivity. Recommended when there is class imbalance [12].

2.2. Research Paradigm and Study Design

This study employs a mixed-methods approach, utilizing a non-experimental cross-sectional research design with a descriptive and field-based focus, to analyze trends in the recognition of Praia Dos Milionários in TripAdvisor's Travelers' Choice Awards. The research focuses on the collection and analysis of primary and secondary data to identify patterns and trends in visitor perceptions.

The study is conducted at Praia Dos Milionários, a major tourist destination on the coast of the state of Bahia, Brazil. The representative sample consists of 368 tourists, both domestic and international, selected through proportional stratified probability sampling. Data collection is conducted on-site via structured surveys administered directly to visitors.

Field surveys are administered to visitors to characterize the tourist experience and the destination's online reputation. Additionally, data on opinions, ratings, and comments available on TripAdvisor are collected, utilizing statistical analysis as well as ML with the practical application of AI techniques, through the use of supervised algorithms to complement the information obtained from the sample and provide a more comprehensive view of the perceptions and motivations that influence the choice of this destination.

Furthermore, data collection takes place directly at Praia Dos Milionários, allowing for real-time information on visitor behavior; this data is integrated into ML models to predict future trends and patterns in online reputation and to classify them, contributing to a more precise understanding of the impact of reviews and experiences on the destination's visibility and prestige.

2.3. Population, Sample and Sampling

The study focuses on the annual number of visitors to Praia Dos Milionários, including both Brazilian and foreign tourists.

According to data recorded by the National Tourism Observatory of the Brazilian Ministry of Tourism, in 2024 there were 140,721 domestic tourists and 143,605 international tourists in the state of Bahia, the state where Praia Dos Milionários is located [13].

Based on these records, a total of 284,326 potential visitors is considered as an approximate population reference for determining the sample size, taking into account the state's tourist influx as a statistical approximation parameter for the study's territorial context. This value corresponds to a population estimate based on Bahia's tourist flow, used as a reference parameter in the absence of specific records for Praia Dos Milionários.

The following formula for finite populations is used to estimate the sample size:

$$n = \frac{z_{\alpha}^2 N \cdot p \cdot q}{e^2 (N - 1) + z_{\alpha}^2 \cdot p \cdot q} \quad (3)$$

Where:

- n = sample size
- N = population size or universe (284,326 visitors corresponding to the aggregate tourist flow in the state of Bahia, used as a statistical approximation due to the lack of specific records for Praia Dos Milionários) [13].
- e = acceptable margin of error ($5\% = 0.05$)
- p = probability of occurrence (0.5)
- q = probability of non-occurrence (0.5)
- Z = critical value of the standard normal distribution for a 95% confidence level ($Z = 1.96$) [14].

Based on the application of the formula for finite populations, a representative sample size of 368 beachgoers was determined.

During the fieldwork, 368 valid questionnaires were administered, following the removal of incomplete forms, the resolution of inconsistencies in responses, and the application of established validation criteria to ensure the quality of the collected data.

The sampling method used in this study, in accordance with the research objectives, is probabilistic (random), stratified (homogeneous characteristic: nationality), and proportional.

Stratification allows the population to be segmented into two analysis groups: Brazilian tourists and foreign tourists, ensuring the representative inclusion of both strata (homogeneous characteristics) through random selection of participants.

Proportional allocation is based on the population composition reported by the National Tourism Observatory for the state of Bahia in 2024. Consequently, the effective sample consists of 182 questionnaires administered to Brazilian tourists and 186 questionnaires administered to foreign tourists, maintaining statistical proportionality between the two strata.

2.4. Research Design and Data Collection Tools

This study employs a non-experimental, cross-sectional research design because it examines the subject of study as it occurs in reality, within its natural context, without manipulating the independent variables; it is cross-sectional because the data collection instruments are administered at a single point in time.

Two data collection instruments were designed for this study: the questionnaire and scientific observation. The questionnaire is a basic instrument consisting of a series of questions designed to gather information on specific indicators related to one or more variables of the research subject.

The designed questionnaire consists of 11 items grouped into five categories: General Information, Overall Satisfaction, Service, Quality, and Perceived Value, with multiple-choice questions on a Likert scale offering three or five response options, as well as open-ended questions to gather data on beachgoers' opinions and moods.

The questionnaire addresses the aspects considered most important in the indicators of the TripAdvisor Travelers' Choice Awards for beaches, which supports the construct validity of the instrument because it measures what it is intended to measure—in this case, the parameters established to obtain such recognition [6].

Furthermore, Cronbach's alpha coefficient is calculated using SPSS for Windows, version 27.0, to determine the questionnaire's reliability. Reliability implies stability, precision, consistency, and minimal errors; that is, if the same data collection instrument were applied to the same subjects under the same conditions, the results obtained would be similar [15].

Once the validity and reliability of the applied questionnaire were ensured, based on the representative sample of the population of 368 beachgoers, of whom 182 were Brazilian and 186 were foreign visitors to Praia Dos Milionários, a database was designed for information processing, enabling the application of the selected supervised ML algorithms; finally, the performance of each algorithm was evaluated using the established metrics.

The scientific observation guide designed in this study is based on participatory, unstructured, and real-world or natural observation. It incorporates the indicators established by the TripAdvisor Travelers' Choice Awards for beaches and describes, for each parameter, the behavior of domestic and foreign beachgoers visiting the tourist attraction, thereby enabling the triangulation of the information obtained through the questionnaire, the selected supervised machine learning models, and the theoretical frameworks for the analysis and interpretation of the results.

2.5. Application of Statistical Hypothesis Testing and Supervised ML Models

A chi-square test of independence is applied to determine whether there is a statistically significant association between the visitor's nationality (Brazilian or foreign) and their perception of whether Praia Dos Milionários deserves the TripAdvisor Travelers' Choice Award.

For this analysis, the following are considered categorical variables: a) Visitor origin (1 = Brazilian, 2 = Foreign), and b) Perception regarding whether the destination deserves the award (1 = Yes, considers it deserving; 2 = No, does not consider it deserving; 3 = is not aware of the award).

The responses obtained through the questionnaire were organized into a database and exported in CSV (Comma Separated Values) format. Subsequently, data processing and analysis were performed using the Python programming language, employing libraries specialized in statistical analysis and machine learning.

To evaluate the predictive power of the explanatory variables, three supervised ML models recommended in the specialized literature were implemented: Logistic Regression, Decision Trees, and

Random Forest. Logistic Regression allows for estimating the probability of belonging to a specific category based on independent variables; Decision Trees facilitate the identification of classification rules through hierarchical segmentation; while Random Forest, as an ensemble method, improves predictive stability and accuracy by combining multiple decision trees.

To evaluate the models' performance, standard classification metrics are used: Accuracy, Precision, Recall, and F1 Score. These metrics allow for the analysis of both overall classification capability and the balance between true positives and false positives, providing a comprehensive evaluation of each algorithm's performance [12].

Comparing these models allows us to identify which one has the greatest predictive power regarding destination recognition, thereby strengthening the study's explanatory and predictive analysis.

3. Analysis and Discussion of Results

3.1. Statistical Hypothesis Testing

3.1.1. Definition of the Statistical Problem

The statistical focus of this study is to determine whether there is a significant difference in the perceptions of domestic and foreign visitors regarding whether Praia Dos Milionários deserves the recognition awarded by TripAdvisor's Travelers' Choice Awards. To this end, the responses obtained from the administered questionnaire are analyzed, considering as categorical variables the visitor's origin (Brazilian or foreign) and their perception of whether the beach deserves the award (1 = Yes, considers it deserving; 2 = No, does not consider it deserving; 3 = is not aware of the award).

Since the analysis involves categorical variables and seeks to determine the existence of an association or independence between them, the Chi-square test of independence is applied, which is widely used in social and tourism research to evaluate relationships between categorical variables [16].

3.2. Statistical Hypotheses

- Null hypothesis (H_0): There is no significant difference between the beachgoers' place of origin and their opinion on whether the beach deserves the TripAdvisor Award.
- Alternative hypothesis (H_1): There is a significant difference between the beachgoers' place of origin and their opinion on whether the beach deserves the TripAdvisor Award.

Table 1 shows the results obtained from the response frequencies for the item related to the question asked in the questionnaire regarding:

Do you consider that Praia Dos Milionários has characteristics that make it worthy of the Travelers' Choice Award?

Table 1
Contingency table (observed frequencies)

Source / Opinion	Yes, considers it deserving (1)	No, does not consider it deserving (2)	Is not aware of the award (3)	Total
Domestic (1)	87	29	66	182
International (2)	66	31	89	186
Total	153	60	155	368

Source: Authors' own elaboration.

Where:

- Chi-square statistic calculated using IBM SPSS Statistics for Windows, version 27.0.
- χ^2 calculated = 7.15
- Degrees of freedom (df): $(rows - 1)(columns - 1) = (2 - 1)(3 - 1) = 2$.
- Significance level: $\alpha = 0.05$
- p-value: $p = 0.028$.

Decision rule:

- If $p \leq 0.05$, reject H_0 .
- If $p > 0.05$, fail to reject H_0 .
- In this case, $p = 0.028 < 0.05$; therefore, the null hypothesis (H_0) is rejected.

Based on the results of the chi-square test of independence ($\chi^2 = 7.15$; $df = 2$; $p = 0.028$), it is concluded that there is a statistically significant relationship between the origin of beachgoers and their opinion regarding whether Praia Dos Milionários deserves the 2026 TripAdvisor Travelers' Choice Award. This result allows us to reject the null hypothesis and accept the alternative hypothesis, confirming that perceptions regarding this tourism award differ significantly between domestic and foreign beachgoers.

From an interpretive perspective, it is observed that domestic beachgoers, particularly local residents of Ilhéus and students participating in the survey, express surprise, and disbelief at the possibility that Praia Dos Milionários meets the necessary conditions to obtain such international recognition. Although they acknowledged the presence of certain positive attributes of the destination, the general trend of their responses was predominantly critical, indicating a less favorable assessment compared to other segments, which suggests a local perception that is more demanding or less aligned with the award criteria.

This trend is consistent with the official results published by TripAdvisor, since, once the lists for the corresponding year were released, Praia Dos Milionários was not included among the 25 best beaches in the world to receive the Travelers' Choice Award. This result corresponds with the findings of the research, demonstrating that the critical or skeptical perception identified, primarily among domestic beachgoers, aligns with the outcome of the tourism recognition [17].

In contrast, the perception of foreign beachgoers is relatively more favorable, highlighting differences in frames of reference, expectations, and tourist experiences between local and international visitors. This finding underscores the importance of considering local perception as a key indicator in the processes of improving and positioning tourist destinations.

Overall, the results obtained suggest that visitor ratings are a significant factor in a destination's tourism recognition and that, in the case of Praia Dos Milionários, the trend identified through statistical analysis is consistent with the results observed in the awards granted by TripAdvisor. Furthermore, these findings justify the application of machine learning (ML) models as a complementary tool for analyzing perceptual patterns and strengthening the study's predictive capacity regarding future tourism recognition.

3.3. Application of ML Algorithms

To predict whether Praia Dos Milionários is perceived as deserving of TripAdvisor's Travelers' Choice Award based on the opinions of domestic and international beachgoers, the problem can be modeled as a supervised classification task.

Table 2 presents the results obtained from the metrics used for the three supervised algorithms applied. These results were obtained by applying 5-fold cross-validation. We removed missing data instead of imputing, to avoid biasing the performance of prediction models.

To complement the evaluation of model performance, confusion matrix analysis is incorporated, which allows for the identification of the number of true positives, false positives, true negatives, and false negatives generated by each algorithm. See Tables 3, 4, and 5.

This tool is essential for understanding the model's behavior beyond aggregate metrics, as it allows for a more detailed analysis of classification errors and the model's ability to correctly distinguish between the categories of interest [18].

Table 2

Performance metrics for supervised algorithms

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression (LR)	0.55	0.53	0.54	0.51
Decision Tree (DT)	0.58	0.59	0.58	0.56
Random Forest (RF)	0.62	0.63	0.62	0.60

Source: Authors' own elaboration.

Table 3

Logistic regression confusion matrix (supervised classification)

Actual class	Predicted class		
	I Don't Know	No	Yes
I Don't Know	123	7	24
No	44	2	13
Yes	72	0	73

Source: Authors' own elaboration.

Table 4

Decision tree confusion matrix (supervised classification)

Actual class	Predicted class		
	I Don't Know	No	Yes
I Don't Know	124	14	16
No	44	10	5
Yes	65	8	72

Source: Authors' own elaboration.

The analysis of the confusion matrices confirms that the Random Forest model presents the best balance between predictive capability and classification error reduction.

The accuracy values obtained, ranging from 0.55 to 0.62, indicate moderate performance, which is expected in social and tourism-related prediction problems, where user behavior is inherently complex and influenced by multiple subjective factors.

The F1-score is particularly relevant, as it balances precision (avoiding false recognitions) and recall (preventing the omission of destinations that could genuinely receive recognition). This means that the best model (Random Forest) is unlikely to overlook actual recognition cases. However, its relatively low precision (0.60) suggests that it generates a considerable number of false positives, predicting recognition in situations where the award would ultimately not be granted.

Random Forest is therefore appropriate when the primary objective is to maximize the detection of potential recognitions, despite its lower overall accuracy.

Regarding the Decision Tree (DT) model, the results reflect a better balance between precision and recall compared to Logistic Regression. This indicates that the model achieves a more balanced

Table 5

Random Forest confusion matrix (supervised classification)

Actual class	Predicted class		
	I Don't Know	No	Yes
I Don't Know	123	6	25
No	53	12	10
Yes	53	4	88

Source: Authors' own elaboration.

classification performance by correctly identifying recognitions while reducing false positives.

However, its slightly lower F1-score (0.56) suggests that, although more balanced, it loses some capacity to detect all positive cases compared to Logistic Regression.

Finally, the Random Forest (RF) model demonstrates the best overall performance among the three models. Its higher accuracy (0.62) and highest F1-score (0.60) indicate a greater ability to generalize and correctly classify both recognition and non-recognition cases.

The combination of higher precision (0.63) and adequate recall (0.62) shows that the model achieves an optimal balance between correctly identifying actual award cases and minimizing false recognitions.

Random Forest is therefore the most robust model for predicting the recognition of Praia Dos Milionários in the Travelers' Choice Awards, as it captures non-linear relationships and interactions among multiple factors associated with the user experience on TripAdvisor.

The superior performance of the Random Forest model can be explained by its ability to capture non-linear relationships and complex interactions among multiple variables. Unlike Logistic Regression, which assumes linear relationships, and individual Decision Trees, which may be prone to overfitting, Random Forest reduces model variance through the aggregation of multiple decision trees trained on different data subsets, resulting in more stable and accurate predictions [11].

4. Conclusion

The analysis of supervised ML algorithms applied to the case of Praia Dos Milionários provides empirical evidence consistent with the fundamentals of electronic word-of-mouth (eWOM), showing that variables associated with user interaction and evaluation have predictive power regarding recognition awarded by the Travelers' Choice Awards.

Furthermore, the sensitivity observed in the models suggests that a favorable digital perception may be associated with a higher probability of institutional recognition, which supports the relevance of electronic word-of-mouth in tourism positioning processes. In this sense, the Travelers' Choice Awards can be understood as an aggregate expression of the reputational consensus generated by users through their ratings and experiences shared on the platform.

The results of the Chi-square test of independence reveal statistically significant differences between the perceptions of Brazilian and foreign beachgoers regarding whether Praia Dos Milionários deserves TripAdvisor's Travelers' Choice recognition. In particular, domestic visitors exhibit a more critical assessment, while foreign tourists present a relatively more favorable perception, suggesting the influence of the sociocultural context on the evaluation of the destination.

This perceptual trend is supported by comparing the analysis results with TripAdvisor's official listings, where Praia Dos Milionários was not included among the world's top 25 beaches awarded the Travelers' Choice Award. This result corresponds with the findings obtained in the research and demonstrates empirical consistency between the observed perception, particularly among domestic beachgoers and the outcome of tourism recognition, thereby strengthening the study's external validity.

The results obtained using the three supervised models suggest, with moderate predictive perfor-

mance, a consistent trend toward a limited probability of Praia Dos Milionários being included in TripAdvisor's list of the 25 best beaches in the world. This trend is consistent with the official lists subsequently published by the platform. Among the models evaluated, Random Forest performs best at capturing nonlinear relationships among the analyzed variables, while logistic regression and decision trees complement the interpretation of the phenomenon.

From a methodological perspective, integrating the Chi-square test with supervised machine learning models strengthens the analysis by combining explanatory and predictive power. This approach allows for the identification of relevant patterns in visitor perception and provides evidence of the utility of these tools in the quantitative analysis of tourism recognition.

The findings demonstrate that visitor perception constitutes a strategic factor for the competitiveness and digital positioning of tourist destinations. In the case of Praia Dos Milionários, the results provide useful insights for the management and improvement of the destination, highlighting the need to incorporate local and national perceptions into tourism evaluation and enhancement processes. Furthermore, the study opens a line of research focused on the use of machine learning techniques for the predictive analysis of digital recognition and the competitiveness of tourist destinations.

Declaration on Generative AI

During the preparation of this work, the authors used Claude (Anthropic) for assistance in manuscript drafting, language polishing, structural review, and verification of statistical workflow code. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

- [1] World Tourism Organization, Glossary of tourism terms, 2024.
- [2] C. Amoiradis, E. Velissariou, M. Stankova, Tourism as a socio-cultural phenomenon: A critical analysis, *Journal of Social and Political Sciences* 4 (2021).
- [3] Ministério do Turismo and Inter-American Development Bank, Sustainable innovation in tourism: Climate resilience and sustainable practices, 2025.
- [4] Embassy of Brazil in Mexico, Information on the state of bahia, 2025.
- [5] C. F. Osorio-Andrade, C. A. Arango Pastrana, A. Rodríguez Orejuela, Evolution of scientific research on electronic word of mouth in the tourism industry: A bibliometric analysis, *Estudios Gerenciales* 39 (2023) 110–123.
- [6] TripAdvisor, Tripadvisor travelers, 2025.
- [7] E. B. Rodríguez, J. Islas, O. L. Ortega, E. L. León, Application of machine learning for the prediction of atmospheric corrosion in the metropolitan area of mexico city, *ITEGAM-JETIA* 10 (2024) 36–42.
- [8] E. Alpaydin, Machine learning, MIT press, 2021.
- [9] M. Hakimi, Comparative analysis of machine learning models for data classification: an in-depth exploration, *Journal of Computer Science and Technology Studies* (2023).
- [10] S. B. Kotsiantis, Decision trees: a recent overview, *Artificial Intelligence Review* 39 (2013) 261–283.
- [11] L. Breiman, Random forests, *Machine learning* 45 (2001) 5–32.
- [12] S. Manoochehri, M. Zamani, M. Afshari, A. R. Soltanian, Z. Manoochehri, Evaluating the performance of different machine learning algorithms based on smote in predicting musculoskeletal disorders in elementary school students, *BMC Medical Research Methodology* 25 (2025) 227.
- [13] Observatório Nacional do Turismo, Estimates of international tourist arrivals in brazil, 2024.
- [14] R. D. MASON, D. A. LIND, Statistical techniques in business and economics, 1994.
- [15] D. Wangzom, The validity and reliability of perception of the traffic safety survey questionnaire for active school travel: A pilot study, *Green Health* 1 (2025) 25.
- [16] A. Agresti, et al., An introduction to categorical data analysis, volume 2, Wiley-Interscience, Hoboken, NJ, 2007.

- [17] TripAdvisor, Travelers' choice awards: Best of the best beaches,, 2026.
- [18] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, *Information processing & management* 45 (2009) 427–437.