Malaysian Journal of Sustainable Environment
Vol.13.No.1 (2026) 271-290
doi: 10.24191/18pddh71

PREDICTIVE CLASSIFICATION MODELLING OF SUSTAINABLE TOURISM PRACTICES USING ONLINE TRAVEL AGENT PLATFORM: A MALAYSIAN CASE STUDY

Nazmirul Izzad Nassir[1], Fiza Abdul Rahim[2]*,
Raja Marzyani Raja Mazlan[3] & Ramona Ramli[4]
***Corresponding Author**
[1]Faculty of Information Technology, City University Malaysia,
46100 Petaling Jaya, Selangor, MALAYSIA
[2,3]Faculty of Artificial Intelligence, Universiti Teknologi Malaysia,
54100 Kuala Lumpur, MALAYSIA
[4]School of Computer Sciences, Universiti Sains Malaysia,
11800 USM, Pulau Pinang, MALAYSIA

nazmirul.izzad@city.edu.my[1], fiza.abdulrahim@utm.my[2]*,
b-marzyani@utm.my[3], ramona@usm.my[4]
Received: 15 June 2025
Accepted: 29 November 2025
Published: 31 March 2026

ABSTRACT

Tourism is a key pillar of Malaysia's economy, recognized in the Economic Transformation Programme (ETP) as one of the 12 National Key Economic Areas (NKEA) that contribute to high-income status. However, the sector faces challenges, particularly in accommodation, due to environmental concerns and difficulties in measuring sustainable tourism. This study applies supervised machine learning classification techniques to develop a predictive model that classifies future sustainable tourism practices within Malaysia's accommodation sector. Using data from an Online Travel Agent (OTA) platform, the model addresses data gaps and provides valuable insights for informed decision-making, aligned with the environmental conservation. The research is divided into four phases: literature review, data collection via web scraping, data modelling, and model evaluation. As the outcome, the multi-layer perceptron (MLP) model with 22 sets of features outperformed others, achieving an F1-score of 58%, recall of 83%, and

Precision Recall Area Under the Curve (PRAUC) of 65%. These findings enhance both the understanding and forecasting of sustainability trends in tourist accommodation using online data.

Keywords: *Sustainable tourism, Machine learning, Online travel agent, Data scraping, Predictive model*

INTRODUCTION

Every day, around three million tourists travel across borders, and every year, more than 1.2 billion people travel abroad. Tourism has developed into a vital component of the economy, a path to success, and a force that has improved the lives of millions of people. The tourism sector has generally had a substantial impact on the socioeconomic development of the region, which helps to explain the negative impacts of COVID-19 on the economies of the region due to the collapse of their tourism sectors (Economic Research Institute for ASEAN and East Asia, 2022; Lm et al., 2024).

As travel is both highly responsive to changes in the income and has a significant impact on carbon emissions, the consumer demand for it has increased considerably more quickly than the demand for other goods and services, as the world economy has advanced (Lenzen et al., 2018). The latest Intergovernmental Panel on Climate Change (IPCC) report warns that global temperatures have risen 1.1°C above the pre-industrial levels, likely exceeding 1.5°C by 2035, intensifying heat waves, droughts, floods, wildfires, rising sea levels, species extinction, and escalating plastic pollution in oceans. (United Nation, 2023). To balance the social, economic, and environmental goals of the society, the tourism sector must be developed responsibly and sustainably (Streimikiene et al., 2021). The necessity of balancing the detrimental consequences of tourism on the economy, ecology, and society highlights the significance of sustainable tourism (Rasoolimanesh et al., 2020).

With its warm climate, diversified attractions, and worldwide connectedness, Malaysia has always relied on tourism to boost its economy. Tourism is one of Malaysia's 12 NKEA in the ETP, contributing to the nation's pursuit of high-income status (Prime Minister's Office, 2012).

Malaysia became the 14th biggest tourist destination worldwide and the 4th largest Asian country in foreign visitor arrivals after welcoming 25.7 million international tourists yearly in the preceding decade (Economic Research Institute for ASEAN and East Asia, 2022). To add, Malaysia's travel and tourism industry was estimated to have contributed RM 102 billion to the country's GDP in 2019 (Ministry of Tourism, Arts & Culture, 2020).

Malaysia's tourism sector relies heavily on accommodation, with significant contributions to overall tourist expenditure (Ab Rahman, 2023). Accommodation, food, and beverages accounted for 60.7% of tourist expenditure in 2021, reaching RM 145.0 million, surpassing transportation, local and foreign airfares, and other categories (Tourism Malaysia, 2021). As part of its transformation strategies, the country actively promotes sustainable and responsible tourism practices (Ministry of Tourism, Arts & Culture, 2020). The National Tourism Policy (NTP) underscores a commitment to sustainable and responsible tourism, with a clear focus on monitoring the tourism industry's contributions to the United Nations Sustainable Development Goals (UNSDGs) as a pivotal transformation strategy. In other words, sustainable tourism is vital for preserving Malaysia's attractions and ensuring resilient growth aligned with environmental, social, and economic factors.

A data-driven analysis is needed to measure sustainable tourism in terms of its decision-making and advancement. For tourist monitoring and management, thorough and reliable spatial-temporal data are necessary (Hoffmann et al., 2022). Despite this, the implementation of the data-driven value creation in tourism is limited to theory or a few exemplary cases, indicating a gap between theory and practice (Ardito et al., 2019). Additionally, the existing data-gathering methods are costly and fragmented (Hoffmann et al., 2022). Conventional methods of assessing tourism sustainability have also been found to be unreliable and unable to provide dependable results (Asmelash & Kumar, 2019). The assessment of progress towards the SDGs is thus, impeded by inadequate data, with only 133 out of 231 indicators having sufficient data (Economic and Social Commission for Asia and the Pacific, 2024).

This research explores the use of OTA platform data and machine learning to assess sustainability in Malaysia's accommodation sector.

It evaluates various classification models for categorizing tourist accommodations as sustainable, addressing the limitations of rule-based systems that require detailed environmental data. Far from replacing physical sustainability assessments, these models leverage correlated factors for broad applicability and cost-effective sustainability monitoring. By complementing the existing labels, they enhance real-time sustainability insights, offering a practical approach to evaluating accommodations. This research extends current methodologies, demonstrating the feasibility of OTA data in the sustainable tourism analysis and supporting informed decision-making for industry stakeholders.

RESEARCH BACKGROUND

Sustainable Tourism Framework

As more travel destinations seek to strike a balance between economic growth, environmental preservation, and social responsibility, the idea of sustainable tourism has acquired global attention. Many locations have created framework documents that express their vision and offer direction for participants in tourist development to ensure the adoption of sustainable tourism practices (Hoffmann et al., 2022).

The United Nation World Tourism Organization (UNWTO), the Global Sustainable Tourism Council (GSTC), the European Commission, the European Environmental Agency (EEA), and the Association of Southeast Asian Nations (ASEAN), among others, developed tourism-specific indicators in addition to the UN's SDGs. Specifically, Goals 8 ("Decent work and economic growth"), 12 ("Responsible consumption and production"), and 14 ("Life below water") all refer to the significance of the tourism industry as part of the SDG pillars (United Nation, 2023).

Sustainability Indicators in OTA Platform

Sustainability indicators in tourism serve as key tools for assessing and understanding the sustainability of specific areas or nations. It is significant to highlight that there are many different sustainability labels, which frequently leave consumers wondering about their precise meaning, the criteria used, and their trustworthiness considering the control systems and enforcement (United Nations Environment Programme, 2016). The ambiguity in sustainability labels confuses consumers, worsened by the varying standards and enforcement. Greater transparency and standardization are needed to ensure that trust and informed decision-making are aligned with consumer values and preferences.

According to the Booking.com survey, over 40% of travelers think it would be useful if vacation booking websites had a clear badge so customers could quickly determine if a lodging is environmentally friendly (Booking.com, 2021b). Seven in ten travelers feel overwhelmed when trying to be more sustainable, while two-thirds seek more sustainability information from lodging and transport providers, preferring listings or dedicated pages on travel sites during their trip planning (Expedia Group, 2022).

Leading OTAs are increasingly integrating sustainability into their business models, reflecting a commitment to responsible tourism practices. Booking.com stands out with its Travel Sustainable Programme, offering a Travel Sustainable Level and a flight filter emphasizing "Least CO2" options (Booking Holdings, 2022). Booking.com's Travel Sustainable badge is a global sustainability measure that empowers travelers with eco-conscious choices (Booking.com, 2021a). The Travel Sustainable badge, developed with the Travalyst Independent Advisory Group, encompasses 32 sustainability measures across five key areas. In collaboration with Sustainalize, it integrates environmental and social impact factors. Approved by the Travalyst Coalition, this framework underpins the badge. This research uses Booking.com for its robust sustainability methodology.

Leveraging Online Data for Insightful Progress and Decision-Making

One way to address the challenge of measuring sustainability indicators in tourism is through the utilization of online data. Leveraging online data offers the potential advantage of gaining real-time insights for informed and data-driven decision-making. An accurate and affordable way to track global advancements in tourism sustainability with fine spatial and temporal resolution is through the algorithmic predictions of the sustainability of accommodations using web data (Hoffmann et al., 2022). Furthermore, to improve data availability and integration, it is becoming more and more important to investigate the application of cutting-edge technology and novel data collection techniques. By incorporating new data sources, public management can adopt more transparent decision-making procedures that are in line with the good governance ideals of inclusivity and transparency (Miller & Torres-Delgado, 2023).

Likewise, numerous European and international organizations have endorsed the use of big data in official tourism statistics, indicating a favorable climate for its adoption at the municipal, regional, and national levels. The report identifies significant potential, as well as problems, related to the development of official tourist statistics, and highlights successful cases where big data has been used to gauge tourism sustainability (Guilarte & Quintáns, 2019). In addition, leveraging smart tourism, the strategy integrates big data analytics and rigorous tourism analytics, emphasizing the active monitoring of the industry's contributions to UNSDGs (Ministry of Tourism, Arts & Culture, 2020). This underscores the vital role of online data in shaping informed decisions and aligning tourism practices with global sustainability goals.

Review of Techniques in Supervised Machine Learning, Data Resampling and Feature Selection

Successful classification studies use diverse machine learning methods, leveraging feature selection to enhance model interpretability, performance, and relevance to objectives.

A sentiment analysis of reviews identifies factors influencing tourist non-revisits using Decision Tree (DT), Least Absolute Shrinkage and Selection Operator (LASSO), Support Vector Machine-Recursive Feature Elimination (SVM-RFE), Back-propagation Neural Networks (BPN), and SVM, with LASSO with SVM performing well in classifying economic hotels (Chang et al., 2019). The combination of clustering, prediction machine learning, and Higher-Order Singular Value Decomposition (HOSVD) proves superior in analysing online reviews, revealing tourists' preferences on social networking sites (Nilashi et al., 2021).

Another piece of evidence shows that with the use of Information

Gain (IG), Chi Square (Chi2), and Gini Index (GI) for feature selection, the proposed methodology outperforms the existing baseline methods, achieving a classification accuracy of 94.7% in the sentiment analysis (Saraswathi et al., 2023). A study on Super Typhoon Rai's impact on Siargao tourism analyzed the revisit intentions using Permutation Importance (PI), RFE, and LASSO, in which the research utilized Logistic Regression (LR) and Artificial Neural Network (ANN) for analysis. The LASSO-ANN model demonstrated exceptional accuracy, achieving 97.81% and simultaneously identifying key revisit factors (Cahigas et al., 2023).

Feature selection is crucial in machine learning to retain relevant features, while resampling addresses imbalanced class distributions. Combining these techniques enhances model performance, interpretability, and robustness, ensuring an effective handling of skewed data for improved predictive accuracy. An example lies in the case for tackling class imbalance in hotel cancellation prediction, where it utilizes the Synthetic Minority Oversampling Technique with Edited Nearest Neighbors (SMOTE-ENN) for resampling and filter method for feature selection. It was found that Random Forest (RF) with SMOTE-ENN outperformed other classifiers (Adil et al., 2021).

In the context of sustainable tourism, the exploration through online platform data employs Random Undersampling (RUS) and SMOTE for resampling, along with Principal Component Analysis (PCA) for feature selection. The study adopts various classifiers, with Quadratic Discriminant Analysis (QDA) excelling in F2 score and recall, while RF achieves the best Receiver Operating Characteristic- Area Under the Curve (ROC AUC) score (Hoffmann et al., 2022). The combination technique includes SMOTE for resampling and Pearson correlation for feature selection in analysing the relationship between the accommodation factors and customer reviews where the results show the model predicting guest satisfaction based on specific factors (Čumlievski et al., 2022).

Next, is the application of the sentiment analysis of restaurant reviews using a hybrid PSO-SVM approach with multiple resampling techniques, including SMOTE, SVMSMOTE, ADASYN, and BorderlineSMOTE. The approach, evaluated against various measures, outperforms standard algorithms and recent approaches in data classification (Obiedat et al., 2022). Machine learning classifiers like RF, SVM, XGBoost, LR, DT, and NB are widely used in tourism research for predicting tourists' behavior, hotel cancellations, preferences, sustainability labels, purchases, sentiment, and post-disaster revisit intentions. Classifier selection depends on research objectives and dataset characteristics.

METHODS

Data Collection

Web scraping was conducted using Python with Playwright, Pandas, and NumPy. The algorithm starts from a specified URL, extracting data from 25 accommodations per page and capturing 40 features per listing. It iterates through 40 pages, collecting up to 1,000 accommodations per location. To maximize the data coverage, URL filtering was customized for cities across 16 states and the federal territory of Malaysia, yielding 17,866 raw samples. The extracted data were stored in a CSV file for further analysis and modelling.

Data Preprocessing

A data check found 1,997 duplicates and three samples with null name and address fields. Removing these refined the dataset to 15,263 samples. Feature expansion from 40 to 84 included imputation, engineering, and logical inference. Outliers, detected via Z-scores, were handled with winsorization to preserve data integrity and prevent skewed analysis. Winsorization involves replacing outliers with the nearest non-outlying values to reduce their impact on the analysis (Sullivan et al., 2021).

Data Resampling

To address class imbalance, this study employs Synthetic Minority Oversampling Technique (SMOTE), an oversampling technique that generates synthetic minority class instances. SMOTE enhances class balance by generating synthetic minority samples along the existing data point segments (Alija et al., 2023). Widely used in machine learning, SMOTE mitigates biases and improves model performance by enhancing the dataset balance, ensuring more reliable classification outcomes.

Feature Selection Techniques

With 84 features, the dataset risks high computational costs. To optimize efficiency, filter and wrapper methods reduce the dimensionality, removing the bias-inducing features during preprocessing. A Pearson correlation threshold was applied, eliminating one feature from each pair with a correlation above 0.9. This process reduces redundancy, ensuring that the retained features provide meaningful contributions to the model's predictive performance. Because the model would have a similar effect on model training, keeping both variables, constant could give false information about how effectively they function (Gregg et al., 2024).

Recursive Feature Elimination (RFE), a wrapper-based feature selection technique, iteratively removes less impactful features by training the model and assessing feature importance until an optimal subset is identified. By keeping the most pertinent characteristics for predictive performance, this technique improves the models' efficiency and interpretability and helps create a more efficient and powerful feature set for machine learning models (Qasim et al., 2021). RFE with cross-validation using LR was applied iteratively, from two to all features, to assess model performance across subsets. This comprehensive analysis revealed the relationship between feature count and performance, with results visualized in line plots to illustrate variations across different feature selections.

Supervised Machine Learning Classifiers

To develop a predictive model for sustainability labels in tourist accommodations, five non-linear classification algorithms were selected due to the dataset's non-linearly separable nature. The chosen classifiers—Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), K-Nearest Neighbor (KNN), and Multilayer Perceptron (MLP) were systematically fine-tuned using hyperparameter optimization. Various hyperparameter values were explored to determine the optimal configurations for each model. The finalized models, optimized with the most effective hyperparameters, were subsequently trained using a refined feature subset, and this ensures improved predictive performance.

Model Evaluation

All experimented feature selection methods and model classifiers will be evaluated using accuracy, precision, recall, F1-score, and PRAUC, leveraging the Scikit-learn library for robust analysis. These metrics provide a comprehensive assessment of classification performance, particularly in identifying the minority class. To ensure model robustness and generalizability, Repeated Stratified KFold (RSKF) cross-validation was employed with a 5-fold configuration repeated twice.

RESULTS AND DISCUSSION

Class Distribution

The dataset exhibits a notable class imbalance, with non-travel sustainable accommodations (12,288; 80.45%) significantly outnumbering the travel sustainable counterparts (2,978; 19.55%), resulting in a class ratio of approximately 4.13:1. This disparity underscores the challenge of accurately classifying sustainable accommodations due to the under-representation of the minority class.

Feature Selection

Pearson Correlation

Six features: 'travel_sustainable_level,' 'waste,' 'destination_and_community,' 'water,' 'nature,' and 'energy_and_greenhouse_gases' were removed to prevent model bias. These features existed only for travel-sustainable samples (1) and were absent for non-sustainable ones (0), potentially allowing the classifier to distinguish classes based on feature presence rather than meaningful patterns. Retaining them could lead to overfitting, resulting in high training accuracy but poor generalization of unseen data, altogether undermining the objective of building a robust sustainability classification model.

A threshold of 0.9 was set to identify highly correlated feature pairs. Features with a correlation coefficient above this threshold were considered redundant, leading to the removal of several highly correlated features to retain only one feature from each pair. Table 1 summarizes these results, showing high correlations among various review and count features. Consequently, these 11 features were removed to reduce redundancy and improve model efficiency and performance.

Table 1. Summary Highly Correlated Features

No	Code	Features
1	R4	review_cleanliness
2	R5	review_comfort
3	R6	review_value_for_money
4	R7	review_location
5	R9	review_facilities
6	R16	count_business travellers
7	R17	count_superb: 9+
8	R18	count_good: 7 – 9
9	R19	count_passable: 5 – 7
10	R20	count_poor: 3 – 5
11	R21	count_very poor: 1 – 3

Source: Author (2025)

Recursive Feature Elimination (RFE)

After removing irrelevant features through the previous filter methods, the focus shifted to using the wrapper method, specifically RFE, with the remaining 66 features, excluding the target variable. Since the optimal number of features was unknown, the model iterated from 2 to 66 features, incrementing by 1 in each step. Selecting the optimal number of features requires balancing various metrics, with an emphasis on F1-score and PRAUC due to class imbalance. The F1-score peaked at around 11 features, while PRAUC stabilized at 22, ensuring strong discriminatory power without any added complexity. Accuracy, precision, and recall show marginal improvements beyond this range. Thus, selecting 11 to 22 features optimally balances the performance, avoiding overfitting and excessive complexity. Figure 1 presents the model performance summary across different feature counts.

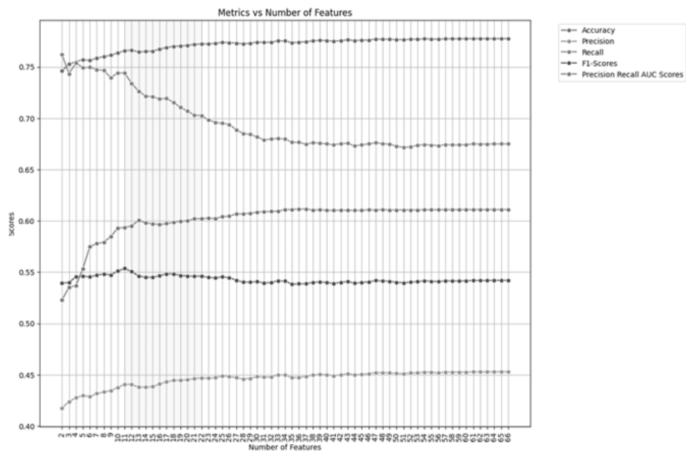

Figure 1. Classification Metrics vs Number of Features

Source: Author, (2025)

The scikit-learn for RFE function was utilized to list the names of the features based on the selected number. These finalized features will be used to train the model classifiers. Subsequently, the remaining unselected features will be dropped from the dataset before model training. The first 22 selected features are shown in Table 4.

Table 2. Selected Features using RFE

No	Code	Selected Features
1	A7	quality_rating
2	A8	preferred_partner
3	RP1	count_room_types
4	R1	overall_reviews
5	R2	count_reviews
6	SR2	count_topattractions
7	SR4	count_closestairports
8	R14	count_groups of friends
9	R22	lan_eng_proportion
10	R23	lan_malay_proportion
11	SR9	average_all_closestairports_distances
12	F3	bathroom
13	F6	business_facilities

14	F8	common_areas
15	F13	languages_spoken
16	F14	living_area
17	F16	miscellaneous
18	F18	outdoors
19	F21	reception_services
20	F23	safety_&_security
21	F24	services_&_extras
22	F26	swimming_pool

Source: Author, (2025)

Model Classifiers Result

Each model was trained using 11 to 22 selected features, ensuring a robust evaluation. The feature selection was adaptive, optimizing the predictive accuracy based on model performance. The best-performing models with their respective feature sets were identified, as summarized in Table 3, highlighting the effectiveness of the selected features for each classifier.

Table 3. Best Model of Each Classifier

Models	Features	Accuracy	Precision	Recall	F1 Score	PRAUC
SVM	20	0.811	0.909	0.034	0.065	0.566
RF	22	0.848	0.630	0.534	0.578	0.627
DT	18	0.802	0.494	0.666	0.567	0.613
KNN	17	0.786	0.469	0.715	0.566	0.620
MLP	22	0.769	0.451	0.831	0.584	0.657

Source: Author, (2025)

The SVM classifier's (20 features) balance precision and PR AUC suffer from low recall, limiting its sensitivity to the minority class. The RF classifier (22 features) achieves high accuracy with moderate PR AUC, F1-score, and precision, ensuring a balanced trade-off. The DT classifier (18 features) excels in recall, F1-score, and PR AUC, effectively identifying the minority class while maintaining precision. The KNN classifier (17 features) demonstrates strong accuracy, precision, F1-score, and PR AUC, emphasizing classification reliability.

The MLP classifier (22 features) outperforms others with a recall of 0.831, the highest PR AUC (0.657), and a balanced F1-score (0.584), making it the most effective model for imbalanced data. Despite an acceptable accuracy (0.769) and moderate precision (0.451), its strength lies in detecting the minority class. However, it generates a notable number of false positives, affecting the overall prediction quality.

Based on the result, there exists the possibility of developing a cost-effective, precise measure of sustainable tourism using readily accessible online data like Booking.com. Machine learning models trained on online platform data can contribute toward the evaluation of the status of sustainable tourism in countries and regions, as evidenced by the correlation between the presence of sustainability awards and various other characteristics (Hoffmann et al., 2022). These models serve to statistically evaluate the degree of sustainable tourism in countries and regions with a high level of temporal and spatial granularity. This also underscores the potential of a large-scale analysis of online data as a valuable approach for research in sustainable tourism and for broader sustainability studies.

CONCLUSION

This research demonstrates the potential of supervised machine learning to classify sustainable tourism practices in Malaysia's accommodation sector using OTA data from Booking.com. It highlights the feasibility of OTA data in the sustainability analysis, supporting stakeholders in making informed decisions. OTA platforms serve as a supplementary data source, complementing traditional datasets to provide a more comprehensive assessment of sustainable accommodations. The MLP model emerges as the best performer, albeit with room for improvement. Despite its promising predictive capability, the study faced challenges such as limited data availability and preprocessing difficulties. Future research should enhance dataset quality, explore the advanced feature engineering, and integrate deep learning for improved accuracy. Refining these models will support sustainable tourism growth and provide valuable insights for the industry stakeholders, enabling more effective monitoring and fostering environmentally responsible practices within the accommodation sector.

ACKNOWLEDGEMENT

The author sincerely thanks Universiti Teknologi Malaysia (UTM) for its financial support and the resources provided for publishing this research.

DECLARATION OF AI-GENERATED CONTENT

AI tools (e.g, ChatGPT, Grammarly) were used for drafting, assistance, paraphrasing and grammar checking. All content was critically reviewed and edited by researcher.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS

All authors contributed to the preparation of this manuscript and approved the final version.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Ab Rahman, S. A., Bachok, S., & Mahamod, L. H. (2023). Effects of travel characteristics on tourism expenditure: A case study of Malaysian young outbound tourists. *Planning Malaysia, 21*(25). https://doi.org/10.21837/pm.v21i25.1233

Adil, M., Ansari, M. F., Alahmadi, A., Wu, J. Z., & Chakrabortty, R. K. (2021). Solving the problem of class imbalance in the prediction of hotel cancelations: A hybridized machine learning approach. *Processes, 9*(10). https://doi.org/10.3390/pr9101713.

Alija, S., Beqiri, E., Gaafar, A. S., & Hamoud, A. K. (2023). Predicting Students Performance Using Supervised Machine Learning Based on Imbalanced Dataset and Wrapper Feature Selection. *Informatica (Slovenia), 47*(1), 11–20. https://doi.org/10.31449/INF.V47I1.4519.

Ardito, L., Cerchione, R., Del Vecchio, P., & Raguseo, E. (2019). *Big data in smart tourism: challenges, issues and opportunities.* In Current Issues in Tourism (Vol. 22, Issue 15, pp. 1805–1809). Routledge. https://doi.org/10.1080/13683500.2019.1612860.

Asmelash, A. G., & Kumar, S. (2019). Assessing progress of tourism sustainability: Developing and validating sustainability indicators. *Tourism Management, 71,* 67–83. https://doi.org/10.1016/j.tourman.2018.09.020.

Booking Holdings. (2022). *Sustainability Report.* https://www.bookingholdings.com/wp-content/uploads/2023/04/BKNG-Sustainability-Report-2022.pdf.

Booking.com. (2021a). *Booking.com Launches First-of-its-Kind Travel Sustainable Badge to Lead Industry in Showcasing a Wider Variety of Sustainable Stays.* https://globalnews.booking.com/bookingcom-launches-first-of-its-kind-travel-sustainable-badge/.

Booking.com. (2021b). *Sustainable Travel Report 2021*. https://5bf1d946-e0b5-464d-82d0-.

Cahigas, M. M. L., Ong, A. K. S., & Prasetyo, Y. T. (2023). Super Typhoon Rai's Impacts on Siargao Tourism: Deciphering Tourists' Revisit Intentions through Machine-Learning Algorithms. *Sustainability (Switzerland), 15*(11). https://doi.org/10.3390/su15118463.

Chang, J. R., Chen, M. Y., Chen, L. S., & Tseng, S. C. (2019). *Why Customers Don't Revisit in Tourism and Hospitality Industry?* IEEE Access, 7, 146588–146606. https://doi.org/10.1109/ACCESS.2019.2946168.

Čumlievski, N., Brkić Bakarić, M., & Matetić, M. (2022). A Smart Tourism Case Study: Classification of Accommodation Using Machine Learning

Models Based on Accommodation Characteristics and Online Guest Reviews. *Electronics (Switzerland), 11*(6). https://doi.org/10.3390/electronics11060913.

Economic and Social Commission for Asia and the Pacific. (2024). *Asia and The Pacific SDG Progress Report 2024: Showcasing Transformative Actions*. UNITED NATIONS.

Economic Research Institute for ASEAN and East Asia. (2022). *Study to develop a framework on Sustainable Tourism Development in ASEAN in the Post COVID-19 Era*. https://asean.org/wp-content/uploads/2023/01/Study-to-Develop-a-Framework-on-Sustainable-Tourism-Development-in-ASEAN.pdf.

Expedia Group. (2022). *Sustainable Travel Study Consumer Attitudes, Values, and Motivations in Making Conscientious Choices*. https://go2.advertising.expedia.com/rs/185-EIA-216/images/April_2022-Sustainable-Travel-Study-PDF-No-URL.pdf.

Gregg, A., Blasco-Puyuelo, J., Jordá-Muñoz, R., Martín Martínez, I., Burrieza-Galán, J., & Cantú Ros, O. G. (2024). Airport accessibility surveys and mobile phone records data fusion for the analysis of air travel behaviour. *Transportation Research Procedia, 76,* 269–282. https://doi.org/10.1016/j.trpro.2023.12.054.

Guilarte, Y. P., & Quintáns, D. B. (2019). Using big data to measure tourist sustainability: Myth or reality? *Sustainability, 11*(20), Article 5641. https://doi.org/10.3390/su11205641

Hoffmann, F. J., Braesemann, F., & Teubner, T. (2022). Measuring sustainable tourism with online platform data. *EPJ Data Science, 11*(1), 1–21. https://doi.org/10.1140/EPJDS/S13688-022-00354-6.

Lm, L. L., Tilaki, M. J. M., Abdul Rahim, A., & Marzbali, M. H. (2024). Covid-19's impact on food stall workers: Night markets' resilience in Penang, Malaysia. *Planning Malaysia, 22*(34). https://doi.org/10.21837/pm.v22i34.1604.

Khan, M. R., Khan, H. U. R., Lim, C. K., Tan, K. L., & Ahmed, M. F. (2021). Sustainable tourism policy, destination management and sustainable tourism development: A moderated-mediation model. *Sustainability (Switzerland), 13*(21). https://doi.org/10.3390/su132112156.

Lenzen, M., Sun, Y. Y., Faturay, F., Ting, Y. P., Geschke, A., & Malik, A. (2018). The carbon footprint of global tourism. *Nature Climate Change, 8*(6), 522–528. https://doi.org/10.1038/s41558-018-0141-x.

Miller, G., & Torres-Delgado, A. (2023). Measuring sustainable tourism: a state of the art review of sustainable tourism indicators. *Journal of Sustainable Tourism.* https://doi.org/10.1080/09669582.2023.2213859.

Ministry of Tourism, Arts & Culture. (2020). *National Tourism Policy 2020 - 2030. In 2020.* https://www.tourism.gov.my/files/uploads/Executive_Summary.pdf.

Nilashi, M., Samad, S., Ahani, A., Ahmadi, H., Alsolami, E., Mahmoud, M., Majeed, H. D., & Abdulsalam Alarood, A. (2021). Travellers' decision making through preferences learning: A case on Malaysian spa hotels in TripAdvisor. *Computers & Industrial Engineering, 158*, Article 107348. https://doi.org/10.1016/j.cie.2021.107348

Obiedat, R., Qaddoura, R., Al-Zoubi, A. M., Al-Qaisi, L., Harfoushi, O., Alrefai, M., & Faris, H. (2022). *Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution. IEEE Access,* 10, 22260–22273. https://doi.org/10.1109/ACCESS.2022.3149482.

Prime Minister's Office. (2012). *Economic Transformation Programme Annual Report 2012.* https://www.pmo.gov.my/dokumenattached/Eng_ETP2012_Full.pdf.

Qasim, H. M., Ata, O., Ansari, M. A., Alomary, M. N., Alghamdi, S., & Almehmadi, M. (2021). Hybrid feature selection framework for the Parkinson imbalanced dataset prediction problem. *Medicina, 57(*11), Article 1217. https://doi.org/10.3390/medicina57111217.

Rasoolimanesh, S. M., Ramakrishna, S., Hall, C. M., Esfandiar, K., & Seyfi, S. (2020). A systematic scoping review of sustainable tourism indicators in relation to the sustainable development goals. *Journal of Sustainable Tourism*. https://doi.org/10.1080/09669582.2020.1775621

Saraswathi, N., Sasi Rooba, T., & Chakaravarthi, S. (2023). Improving the accuracy of sentiment analysis using a linguistic rule-based feature selection method in tourism reviews. *Measurement: Sensors, 29*. https://doi.org/10.1016/j.measen.2023.100888.

Streimikiene, D., Svagzdiene, B., Jasinskas, E., & Simanavicius, A. (2021). Sustainable tourism development and competitiveness: The systematic literature review. *Sustainable Development, 29*(1), 259–271. https://doi.org/10.1002/sd.2133.

Sullivan, J. H., Warkentin, M., & Wallace, L. (2021). So many ways for assessing outliers: What really works and does it matter? *Journal of Business Research, 132,* 530–543. https://doi.org/10.1016/j.jbusres.2021.03.066.

Tourism Malaysia. (2021). *Malaysia Tourism Key Performance Indicators 2021.* https://mytourismdata.tourism.gov.my/wp-content/uploads/2023/01/KEY PERFORMANCE-INDICATORS-2021.pdf.

United Nation. (2023). *The Sustainable Development Goals Report 2023.* https://unstats.un.org/sdgs/report/2023/The-Sustainable-Development-Goals-Report-2023.pdf.

United Nations Environment Programme. (2016). *Consumer Information Tools and Climate Change Facilitating low-carbon choices in Tourism, Buildings and Food Systems Guidance for Policy Makers and Business Leaders.* http://www.oneplanetnetwork.org/consumer-information-scp/,orcontact:ciscp@un.org.

Weidenfeld, A. (2018). Tourism diversification and its implications for smart specialisation. *Sustainability (Switzerland), 10*(2). https://doi.org/10.3390/su10020319.