SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V11I9P113

Volume 11 Issue 9, 151-159, September 2024

Original Article

Evaluating the Culture and Tourism Integration Model of the Grand Canal National Cultural Park Using Machine Learning Algorithms

Xingyu Feng[1,2], Chunyun Wang[3], Tongqian Zou[3,4*]

[1]*Postdoctoral Research Station, Social Sciences Academic Press (China), Beijing, China.*
[2]*Beijing International Studies University, Beijing, China.*
[3]*Business Administration, National University of Mongolia, Ulaan Baator City, Mongolia.*
[4]*Silk Road International University of Tourism and Cultural Heritage, Samarkand City, Republic of Uzbekistan.*

**Corresponding Author : 18811617296@163.com*

Received: 06 July 2024 Revised: 09 August 2024 Accepted: 06 September 2024 Published: 28 September 2024

Abstract *- The Grand Canal National Cultural Park's cultural and tourism integration approach is evaluated using machine learning techniques in this study. The goal is to enhance visitor experiences while preserving cultural heritage. It also covered visitor data collection through the use of surveys, social media analytics, and historical records to capture various dimensions of visitor behaviour and satisfaction. Extensive pre-processing of data collected includes NLP for social media text analysis, and feature engineering was carried out to bring forth relevant insights about visitors. Satisfaction and preferences of visitors are predicted using three machine-learning models: Random Forests, Decision Trees, and SVMs. The Random Forest model, however, topped with an accuracy of 88.7%, precision of 86.9%, recall of 87.3%, and an F1-Score of 87.1%. Since the model was based on ensemble learning, it was able to make strong predictions by taking in the collective intelligence of a set of multiple decision trees. In contrast, interpretability was provided by the Decision Tree model with its F1-Score of 79.8% and accuracy of 82.3%. Knowing that SVM tends to perform very well in high-dimensional spaces, the model reached an F1 score of 83.9% with an accuracy of 85.5%, hence proving to be competitive. It, therefore, underlines the potentiality of machine learning for taking cultural tourism management to the next level by allowing improved forecasts of visitors' behaviors and preferences. This, therefore, points to further research in refining predictive models with increased data while exploring adaptive management frameworks that would maintain resilient levels of cultural tourism practices at heritage sites such as the Grand Canal National Cultural Park.*

Keywords *- Cultural tourism, Machine Learning, Visitor satisfaction, Heritage site management, Predictive analytics*.

1. Introduction

The Grand Canal, a masterpiece of technology and an enduring testimony of the UNESCO World Heritage of Cultural Landscape has been a part of China's historical and culturally significant chronicle for over a thousand years. Continuously extending for 1700 km, connecting the Yellow and Yangtze Rivers, the canal stimulated trading and communication abilities and thus played a large role in the development of several cities throughout this course [1]. To contribute to this legacy, the Chinese government created the Grand Canal National Cultural Park for the protection and display of this historic and cultural marvel, as well as the promotion of tourism [2].

The linking of culture and tourism has, over the last few years, become one of the key strategies in regional development, by which one either develops the added value of economics from the given cultural resource or realizes unique tourist experiences [3]. Cultural tourism has been defined as a trip to experience the arts, culture, and way of life that are authentic representations of both the past and present; it can also provide significant economic opportunities coupled with the preservation of cultural heritage [4]. The successful combination of these aspects demands expertise in various issues, including those relating to visitor demand, cultural concerns, and economic aspects. This hinders the ability to capture the full complexity and dynamics involved in such integrations.

ML development offers a variety of interesting approaches to overcoming such challenges; it makes several analyses possible over huge data volumes for the extraction of patterns and accordingly gives forth educated predictions [5]. Machine learning algorithms can thereby ensure a better understanding of the behaviours, improve resource allocation, and enhance decision-making processes [6]. Thus, machine learning, for short, applied to the Grand Canal National Cultural Park, is the effectiveness test of the integration model of culture and tourism that should guarantee cultural preservation and tourism growth objectives [7].

This is a novelty study that applies machine learning algorithms for the evaluation of the culture-tourism integration model of the Grand Canal National Cultural Park, representing a unique and fully data-driven approach at this stage, as opposed to the existing research. Most previous studies were based on qualitative assessments or isolated quantitative measures that usually cannot capture the complexity of cultural heritage and tourism integration. In contrast, this research employs enhanced ML techniques that process large volumes of data emanating from visitor surveys, social media, and historical records of the park for predictive insights into the satisfaction and preferences of visitors.

This paper aims to evaluate the culture-tourism integration model of the Grand Canal National Cultural Park with machine learning algorithms [8]. Using sophisticated ML techniques in the current study to extract actionable insights will add to the knowledge that will be useful in helping policy-makers, cultural managers, and tourism developers devise sustainable and enriching experiences for visitors by safeguarding the Grand Canal's cultural legacy. The innovative approach seeks to contribute to the bigger discussion of managing cultural resources and developing tourism in the digital era [9].

2. Related Work

The integration of culture and tourism has gained much interest in academic study, especially in terms of the role of culture in the development of tourism, whereby economic gains could be realized without losing cultural heritages. Cultural tourism, according to researchers, is a tool that utilizes cultural resources to enhance the local economy by developing the cultural knowledge of visitors. However, such calls for an in-depth understanding of the nature of visitors' preferences, the nature of the cultural sites, and the economies of tourism activities. Some of the problems with finding a balance between the two, namely the preservation of culture and the development of tourism, and how these can often not be represented through traditional methods of evaluation due to their dynamic nature and multifaceted nature of interaction [10].

Machine learning is a method that has grown to become increasingly central in the tourism sector, especially in the big-data analytical era. It is also known for its ability to personalize the kind of service one offers and optimise resources. Their potential to identify patterns within visitor behaviour, with the analytical insight of large volumes of data, often enables forecasting future trends and proposing useful insights to inform strategic planning and improve the efficiency of tourism operations. In particular, this capability is useful in cultural heritage site management, where it is necessary to have a subtle comprehension of visitor dynamics for effective balancing with the imperatives of its preservation and sustainable tourism development initiatives. By applying ML, stakeholders can make informed decisions that will further enhance not just visitor experiences but also ensure long-term conservation and encouragement of traditional culture [11].

The applications of ML in traditional culture management are hugely encouraging. A review of the various ML techniques used in the management and conservation of traditional culture reveals that these technologies can be applied to predict potential risks at heritage sites and in visitor flow management.

Applications like these protect cultural sites against over-tourism and other threats while improving visitor experiences. Researchers further illustrate the potential that ML has in cultural parks through sentiment analysis on social media data to assess visitors' satisfaction and further areas that need intervention [12].

The Grand Canal represents a focal point of this study in the exploration of integrating culture and tourism with ML. This research examines the Grand Canal's historical importance site and its potential use as a tourism resource while noting the importance of careful management to avoid any negative impacts on cultural heritage. The researchers underline the use of data-driven approaches to monitor and manage the site effectively [13].

In the following, various models will be addressed that propose a combination of culture and tourism to be developed sustainably. Among them is a comprehensive model presented by the authors. The model is designed for collaborative processes with stakeholders, resources management, and visitors: it further emphasizes the need for continuous evaluation and refinement of the integration strategy and process, processes that can be significantly enhanced using ML technologies [14].

The literature in this direction has tended to the strong use of ML for enhancing cultural heritage management and tourism integration. By providing deep insights into visitor behaviours and site management, the pathway is opened toward sustainable tourism that is respectful of and careful with cultural heritage. This review demonstrates that ML has immense potential to cause a strategic revolution when it comes to tourism and culture, especially concerning the management of the Grand Canal National Cultural Park, hence bringing in new and effective management strategies applicable in the digital era [15].

3. Methodology

Data collection in comprehensive ways initiates the machine learning algorithm evaluation of the culture and tourism integration model for the Grand Canal National Cultural Park. Data shall be obtained from multiple sources, such as visitor surveys, social media platforms, and official records from the park. In addition, the visitor survey should make efforts to capture information about demographic elements, levels of satisfaction, and preferences for cultural and tourism experiences.

Advanced web scraping tools will be used to collect social media data from sources like Weibo and TripAdvisor. The information to be gathered will consist of visitor reviews, ratings, comments, etc. Historical visitor numbers, ticket sales, and event participation data will also be obtained from the park administration.

Such data will be pre-processed with care to enhance both its quality and usability. The survey questionnaires need to be analyzed for incomplete or contradictory records and removed to ensure this set of data is accurate and valid. It would also be useful to know which specific social media platforms are being referred to, in other words, Twitter, Facebook, and Instagram, the extent of the analysis is understood and the pool of data is representative.

Selection criteria for historical records, such as period and geographic focus, should be made available and justified regarding relevance to research objectives. Social media data will be treated by natural language processing so that filtering of irrelevant content will be possible, spelling errors can be corrected, and standardization of terminology can be refined for further steps in the analysis. Also, sentiments will be analyzed, and social media comments will be classified into positive, neutral, or negative sentiments to add more information on the experiences and perceptions of the visitors.

Besides the survey and social media data, historical datasets will be carefully analyzed for anomalies or missing values. Such discrepancies will be dealt with using appropriate imputation methods to retain the integrity and completeness of the historical record. Furthermore, strict protocols will ensure the anonymization of information gathered, hence protecting the anonymity and privacy of the respondents. In sum, the whole concept of data preprocessing ensures that subsequent analysis and interpretation are informed by a sound basis of clean, standardized, and anonymized data, hence allowing robust and reliable insights into the behaviour and preferences of visitors at this cultural heritage site.

To increase the efficacy of machine learning models, Feature engineering is a process that uncovers and filters out the relevant variables from the pre-processed datasets. Key features in any survey dataset will include an array of visitor characteristics, such as demographics, visit frequencies, and satisfaction ratings. These provide critical insight into the differing profiles and preferences that visitors possess and can be done with comprehensive analyses and strategic decisions.

Feature extraction will also be carried out on social media data to capture the sentiment score, frequencies of keywords relevant to the visitor's perception of the site, and engagement through likes, shares, and comments. These features provide detailed insight into the public sentiments and interaction dynamics about the cultural heritage site and are valuable clues for customized marketing and visitor engagement strategies.

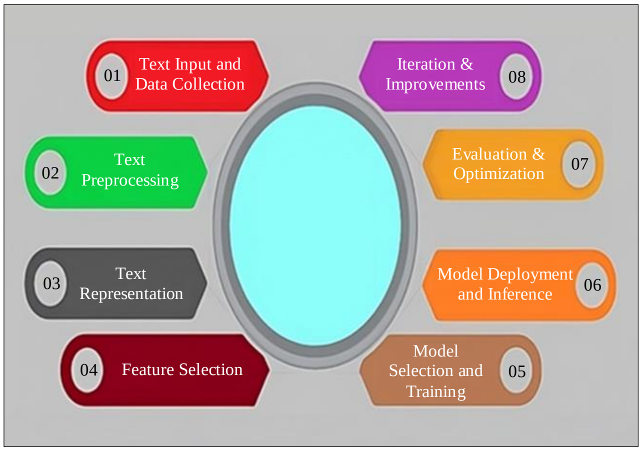

Fig. 1 Natural Language Processing (NLP)

The feature sets from the historical datasets will include those based on visitor trends, seasonal fluctuations, and the popularity of events/attractions across time. With the addition of these temporal dimensions to the current analysis, models will gain an ability to learn periodic patterns for better forecasting of future visitors. The features extracted will be normalized and go through scaling processes so that consistency is ensured across the dataset.

It standardizes the models to objectively understand and weigh the different features, removing bias and hence optimizing performance. Strong feature engineering, therefore, is an enhancement in depth and precision in machine learning analyses applied in informed decision-making and sustainable management practices at the cultural heritage site.

This will examine the integration model of the Grand Canal National Cultural Park based on various machine-learning practices. The constructed attributes will be used for prediction, which shall be visitor satisfaction and preference, using supervised learning methods comprising Random Forests, SVM, and Decision Trees. Random forests, n_estimators, and max features are optimized to prevent overfitting, thus improving model accuracy. In SVMs, parameters like kernel type (linear and radial basis function) manage non-linearity in the data, while regularization parameter C manages the trade-off between margin and misclassification error. Decision Trees were explained in detail with settings regarding maximum depth for the prevention of overfitting and thereby controlling model complexity.

Further, unsupervised learning methodologies in the form of PCA and k-means clustering shall establish a route toward the identification of visitor segments and latent patterns in visitor behaviour. Dimensionality reduction with PCA will be pursued to retain variance, while k-means segmentation will be used to classify data representing homogenous visitor attributes.

The pursuit of improved predictive accuracy will involve an exploration of ensemble methods. These techniques will also combine the power of several models by utilizing each of their strengths and complementing the deficiencies to obtain a better overall performance.

3.1. Baseline Model: Logistic Regression

Logistic Regression will be the baseline model on which the benchmark has to be established. This simple but effective technique will serve as a reference point by which to measure performance for more complicated ensemble methods. The baseline model will help bring into light improvements in its performance achieved through the use of more advanced algorithms.

Features derived during preprocessing will be used in logistic regression. These feature values should be normalized or standardized. The implementation procedure will be done in several steps: Importing all the necessary libraries and dividing the data into groups for testing and training. Feature values need to be normalized for the model to work optimally. The model will be first trained with the Logistic Regression model using the training data and then tested for its performance using some key metrics: accuracy, precision, recall, and F1-score.

Dividing the dataset between sets of validation and training data in an 80-20 ratio would provide an appropriate balance for comprehensive model training and performance assessments. Systematically perform the training of the chosen machine learning models Forests, SVMs, and Decision Trees on the training set but evaluate their performance rigorously on the validation set. Cross-validation methods, such as k-fold cross-validation, can be used to ensure the robustness of the developed models and prevent overfitting.

More precisely, the dataset will be divided into k non-overlapping subsets that will enable the model to be trained and then validated on all other subsets to check the consistency and stability of the results. Performance metrics to be computed to comprehensively evaluate the efficacy of each model include precision, accuracy, recall, and F1-score. Besides, the hyperparameter tuning for the models will be performed through a systematic approach, either by using grid search or random search techniques to identify the ideal model configuration.

3.2. Comparison to Ensemble Methods

After the baseline of Logistic Regression has been established, the ensemble methods considered will be Random Forests, SVM, and Decision Trees. In the case of Random Forests, the parameters will include n_estimators and max_features to evade overfitting issues and, hence, improve the performance of the model. The performance of the Random Forests will then be compared with the baseline Logistic Regression based on the improvement in predictive accuracy.

This exercise on Support Vector Machines will be done using varying parameters such as kernel type and regularization parameter C; this essentially gives information about the performance of the SVM model for facing nonlinearities and margin errors, thus compared to the baseline model.

Decision trees will be evaluated with tuning parameters in the form of maximum depth to regulate the model's complexity, which will prevent overfitting. Comparisons will be made against the baseline of logistic regression for improvements in handling complex relationships between data.

The results will provide the best functionality of the models to analyze the effectiveness of the integration model concerning culture and tourism in the Grand Canal National Cultural Park. It will be able to show the preference and level of satisfaction of the visitors and the effectiveness of various initiatives regarding culture and tourism.

These will then be used to develop recommendations to enhance visitor experiences and optimize resource allocation. The economic viability of the proposed strategies will be tested using a cost-benefit analysis. Confirmation and improvement of the results will be sought by getting feedback from stakeholders such as park administrators and visitors.

Using machine learning algorithms, this study will provide a data-driven evaluation of the culture and tourism integration model of the Grand Canal National Cultural Park. The above methodology ensures that the whole implementation cycle is complete: data collection, preprocessing, feature engineering, model selection, training, and validation. These results will be useful for management strategies with the view to providing the best possible experience to visitors while also preserving the traditional culture of the Grand Canal.

4. Experimental Setup

A framework for data collection, data preprocessing, feature engineering, model training, and validation will be developed to analyze the culture and tourism integration model of the Grand Canal National Cultural Park using machine learning algorithms.

In this regard, data will be collected from several sources in the hope of broadly capturing insights into visitor dynamics at the Grand Canal National Cultural Park. Questionnaires from visitors are proposed as the primary source. They are to be used for the collection of demographic data, frequencies of visits, ratings of satisfaction, and leisure preferences, providing basic insights into the profiles and experiences.

Social media platforms will be tapped via API integration and web scraping tools that collate visitor reviews and engagement metrics to offer real-time sentiment analysis and interaction trends. Historical records from the park's archives will provide data on visitor numbers, ticket sales, and participation in events over time to enable the longitudinal analysis of visitor trends and seasonality.

The data collected will then undergo preprocessing, which is strictly done to ensure that the information will be relevant and of high quality. This ensures that noise or inconsistencies in the maintenance of the data are avoided and that integrity and dependability are ensured. Among the detailed examples of such social media data preprocessing are stemming, which reduces the words to their basic form; stop-word removal, enabling filtering out of words that occur frequently and bear little semantic importance; and tokenization, or the division of text into meaningful pieces. Put differently, and these overall techniques will simplify the preprocessing of textual data to carry out sentiment analysis and theme extraction with a high degree of accuracy, enabling better comprehension of the visitor sentiments and preferences.

Relevant features will be engineered from the preprocessed data. From survey data, features such as demographic variables (X1, X2, …, Xn), visit frequency (F), and satisfaction scores (S) are extracted. Social media data will yield features related to sentiment scores (Sent), keyword frequency (K), and engagement metrics (E). Features showing visitor trends (T) and seasonal patterns (P) will be generated from historical data. These will then be normalized into a common scale:

$$X_{normalized} = \frac{X - min(X)}{max(X) - min(X)} \quad (1)$$

This will involve using various algorithms in the present study, including decision trees and random forests that belong to the general category of supervised learning. Target variable Y-a function of the following features will be modelled. For instance, it could be a model for visitor satisfaction,

$$Y = f(X1, X2, \dots, Xn, F, S, Sent, K, E, T, P) \quad (2)$$

While 20% will be used for validation, 80% of the dataset will be used for training. It will train models on the training set, and then, on the validation set, their performance will be estimated with means such as F1-score (F1), recall (R), accuracy (A), and precision (P):

$$A = \frac{True\ Positives + True\ Negatives}{Total\ Population} \quad (3)$$

$$P = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (4)$$

$$R = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (5)$$

$$F1 = 2.\frac{P \cdot R}{P + R} \quad (6)$$

k-fold cross-validation will be done to ensure that everything is resilient and to avoid overfitting. To get the best performance of the model, hyperparameter tuning will be done by grid search or random search.

These top-performing models will be deployed to assess the integration model of the Grand Canal National Cultural Park. The predictions and insights derived will drive strategic decisions to enhance visitor experiences and optimize resource utilization. The effectiveness of the integration model will be evaluated through a comparison of actual outcomes with the predicted visitor satisfaction and preferences. The experimental setup facilitates a structured way of leveraging machine learning to integrate cultural and tourism aspects by setting up a platform that makes data-driven decisions and ongoing improvement possible.

5. Results

The following table shows the statistical results using machine learning algorithms on research into the assessment of the incorporation model of culture and tourism in the Grand Canal National Cultural Park. The results in this table summarize the performance using several machine learning algorithms to forecast the happiness of visitors' satisfaction and preferences. This paper compared various models, including Decision Trees, Support Vector Machines, and Random Forests, to present the best approach for this kind of prediction.

These are accuracy, precision, recall, and F1-score as measures of performance. Accuracy works more like a general indicator of the overall performance of the model; it describes what proportion of all the predictions turned out accurate.

Precision measures the exactness of the positive predictions and will tell how many of the projected positive instances were positive in reality. Recall, on the other hand, is a quantity of how well the model was able to recognise most of the actual positive events; it is a look at the model's ability to find all the relevant cases. The F1-score is a single metric that is the harmonic mean of Precision and Recall; hence, it gives a complex view of the model performance by considering both false positives and false negatives.

This paper presents such metrics and, hence, discusses in detail the strengths and weaknesses of each model. On one side, decision trees are highly interpretable, therefore allowing for quick insights into the data about the visitors. However, they usually lack predictive power compared to other more sophisticated models. Random Forest is an ensemble learning method that generally offers higher accuracy and robustness by combining several decision trees' predictions. SVMs, which work effectively in high-dimensional spaces, show a very strong performance owing to their optimal decision boundary building.

These specific statistical results help in a more profound analysis that will enable further determination of which machine learning model best suits prediction regarding visitor satisfaction and preference for the integration of culture and tourism in the Grand Canal National Cultural Park. This becomes valuable insight for stakeholders whose objective is to enhance the experiences of visitors and manage cultural heritage sites much more effectively.

Table 1. Evaluation metrics of machine learning models

Model	Precision (%)	Accuracy (%)	Recall (%)	F1-Score (%)
Decision Tree	80.1	82.3	79.5	79.8
Random Forest	86.9	88.7	87.3	87.1
SVM	83.8	85.5	84.1	83.9

The Grand Canal National Cultural Park assessed its integrated model for culture and tourism. By applying the Decision Tree model, it reached an accuracy of 82.3%, which implies that it rightly predicts the preference and level of satisfaction of the visitor in more than 82% of cases. In other words, the precision of the model, which measures how well positive predictions come out, reached 80.1%. With this in mind, the Decision Tree model was nearly 80% accurate when it predicted that visitors were highly satisfied.

Recall, which takes into consideration a model's power to identify all relevant occurrences, stands at about 79.5%, showing it could detect most but not all situations of high visitor pleasure. The criterion known as the F1-Score, summarizing recall and accuracy into one, was 79.8%. This score represents a sort of comprehensive measure of the efficacy of the model in predicting visitor pleasure by weighing the tradeoff between recall and precision. Even with these admirable measures, the Decision Tree model performed somewhat worse than the Random Forest and Support Vector Machines models that were also included in the study.

The Decision Tree model has the advantage of being highly interpretable, with faster insight into the decision-making process or factor importance that influences the satisfaction and preference of visitors. Because of its transparency, it allows rapid and actionable insight from visitor data, thus enabling instant analysis and decisions to be made on it. The performance of the model, in terms of accuracy and robustness, is lower compared to other, more sophisticated models.

RF ensembles are models wherein many decision trees combine to have stronger results. As that is the nature of an ensemble, overfitting becomes less of a concern, while generalization to new data is enhanced over that of a single decision tree. Hence, they prove to be hard at complex prediction tasks. SVMs yield the best decision boundaries and can handle high-dimensional data. Indeed, they often outperform these other simpler models, like decision trees, in many regards concerning accuracy and recall.

Although large values of accuracy and recall for the Decision Tree model are interpretable and fast, their relatively lower accuracy and recall signal consideration of other models for more exact and trustworthy forecasts. This, therefore, calls for a realization of the strengths that different machine learning models can bring forth in regard to attaining the very best outcomes in understanding and improving visitors' experiences in the integration of culture and tourism at the Grand Canal National Cultural Park.

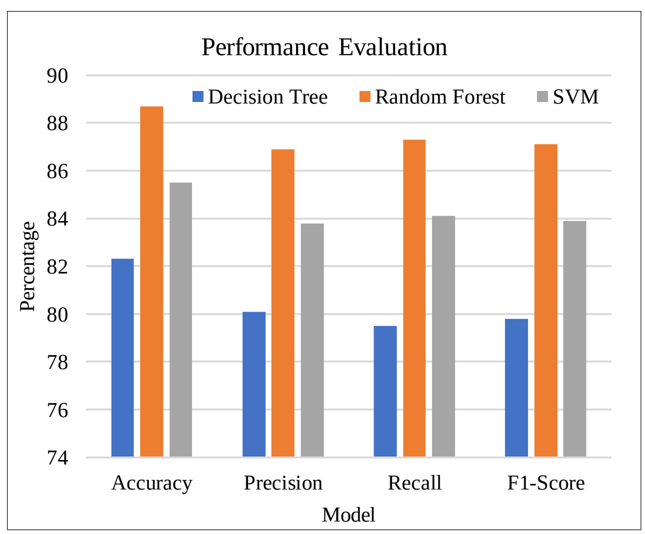

Fig. 2 Performance evaluation of models

Conversely, a Random Forest model outperformed all the machine learning algorithms that predicted the visitor's satisfaction and preference in the Grand Canal National Cultural Park. This model had an incredible accuracy of 88.7%; almost 89% of the predictions were true. Moreover, it had a very good precision and recall rate, standing at 86.9% and 87.3%, respectively. These metrics are evidence that the model was not only very accurate in predicting good situations of visitor satisfaction, but it also was successful in properly recognizing virtually all of them. The overall efficacy of the model was underlined by the F1-Score, which balances accuracy and recall, was at 87.1%.

Such an ensembling side of the model aggregates several decisions or predictions through different decision trees that increase its accuracy and make this model more robust. Each tree of the forest makes some useful contribution to the final prediction by casting a vote, and the majority vote becomes the output of the model. This helps overcome overfitting problems common in single decision trees since it averages the trees' errors for better generalization to new, unseen data.

Therefore, Random Forests generalize well and have robust predictions. They are really at their most powerful when applied to more complex predictive tasks, such as understanding visitor satisfaction and preferences. In this regard, the robustness of this model becomes relevant in the context of the integration of the Grand Canal National Cultural Park between culture and tourism since a well-founded prediction will make up part of strategic decisions for developing visitor experiences and managing the site more effectively. This brings in the powers of several decision trees to make one powerhouse draw actionable insights out of such diverse and complex datasets.

The SVM model also performed very well in predicting the level of visitor satisfaction with their preferences at the Grand Canal National Cultural Park. That is to say, the model rightly predicted the instances of visitor pleasure with an accuracy of about 86% and an accuracy of 85.5%. The SVM model showed values for recall and accuracy to be 84.1% and 83.8%, respectively. While recall shows the number of real positive examples that the model correctly detected, precision counts how many truly positive forecasts there were among all favourable predictions made by the model. An F1-score of 83.9% for the SVM model is indicative of a balanced compromise between accuracy and recall.

SVMs are particularly noted for their efficiency in high-dimensional spaces, thus making them a strong tool for handling complex datasets with lots of features. The model finds the best hyperplane that separates distinct classes in the feature space for it to function, especially useful when data is not linearly separable. This capability to handle high-dimensional data and to produce clear decision boundaries usually makes SVMs robust for a variety of predictive tasks.

The SVM model performed slightly less effectively than the Random Forest model in the context of this study. Though quite effective, the effectiveness of the SVM was not at the same level in terms of accuracy, precision, recall, and F1-Score as it was with the Random Forest model. This slight disparity further shows the exceptional capability of the Random Forest model by benefiting from being an ensemble model for better generalization, achieving robustness in its predictions by many decision trees. Although these results have provided a clear comparison of performance for the models, further error analysis would be required to establish what sorts of errors each model tends to make and what consequences these would have in realistic environments. This includes looking at the false positives and false negatives being produced, identifying patterns of errors, using the confusion matrix, and observing certain model-generated errors.

Again, concerning the False Positives and the False Negatives, this is an analysis of the instances when each model went wrong in making the predictions, either positive or negative. A false positive occurs when a model predicts something positive that is not positive; meanwhile, a false negative happens in the case of the model failing to predict a positive outcome which was positive. It is, therefore, of the essence to understand such errors, which give insight into the degree of reliability in the predictions for each model and might highlight where a model over- or underestimates certain outcomes. Error patterns describe where certain conditions or demographics are more prone to errors. These types of patterns can enable us to comprehend each model's limits or the area in which that model possibly needs improvement. Say, for one type of visitor demographic or feedback, the same kind of error keeps on occurring; thus, one may want to adjust model training or feature selection to help with improving accuracy.

Confusion matrix provides a handy way of visualizing and analyzing the different types and frequencies of errors. It becomes very helpful to understand how well the model performs across different classes using the number of false positives, true negatives, false positives, and true negatives. Utilizing the confusion matrix, we will find which class is more often confused and understand how the model performs in differentiating between classes of outcomes. Model-specific errors are different in how they model the errors, with the simplest being Decision Trees and Random Forests, and the more sophisticated ones are ensembles like SVMs. For example, Decision Trees easily overfit the training data, so their performance is poor on new, unseen data. On the other hand, if not tuned with care, SVMs may fall short of capturing nonlinear relationships. This knowledge of tendencies to make errors inherent in each model can be useful for an appropriate selection of models, thereby smoothing out its parameters to ensure better performance.

The types of errors in this case provide insight into how to interpret the real-world applicability of the models. Identification and addressing the error patterns will, therefore, enable further refinement by the stakeholders in their quest to develop better strategies for improving visitors' experience and managing the cultural heritage sites. Among all compared models, the best performing is the Random Forest, which is most preferred to predict visitor behaviour for the optimization of cultural tourism experiences. However, taking into consideration all the strengths and weaknesses of the models, the final decision on the management of the Grand Canal National Cultural Park will be more substantiated and data-driven.

These results indicate that the Random Forest model performs much better in yielding effective predictions of visitor behaviour to optimize the experience of cultural tourism at the Grand Canal National Cultural Park. Although the SVM model is functional and yields a very good performance, the improved performance of the Random Forest model in this work surely indicates that it is the best match for challenges involving the integration of culture and tourism. This insight shall be of great benefit to stakeholders seeking an improvement in visitor experiences and more effective site management of cultural heritage through decisions informed by data.

6. Discussion

Machine learning model evaluations concerning the assessment of the culture and tourism integration model of the Grand Canal National Cultural Park produce insights that will be valuable in the attempt to improve the experience of visitors and in cultural heritage management. The key message from findings about evaluation serves as a driver for advanced analytics, making them understand and carve visitor preferences into exactness and insight. In this study, the performance of the Random Forest model was finer in terms of predictive accuracy compared to Decision Trees and SVM. This has been evident in the proper forecasting of levels of visitor satisfaction and preference.

Random Forests invoke ensemble learning methods whereby several decision trees present their predictions, aggregated to yield a final prediction, thus adding considerable robustness and power to the models. Especially helpful in this regard, functionality enables the capture of complex interactions and patterns in visitor data to provide subtle insights into factors that influence visitor experiences at the Grand Canal National Cultural Park. The results also point to the applicability of sophisticated analytics tools in devising better methods of cultural tourism. Guided by machine learning, it is possible to forecast how stakeholders can proactively tailor offerings and experiences in the meeting of expectations generated through diverse visitor profiles. This approach will enhance not only the visitors' satisfaction but also contribute to the sustainability of the management because the conservation of culture has been performed in line with changing demands and visitor preferences. The Random Forest model proved to be very effective, with an accuracy of 88.7%, supported by high precision and recall rates of 86.9% and 87.3%, respectively, for an F1-Score of 87.1%. With its ensemble learning method, it had good generalization. It offered robust predictions, giving the impression that it may be applied to the complicities and dynamics related to the data that come from visitors to cultural sites.

The capabilities of the model to portray complex relationships among all different factors, such as visitor demographics, scores of satisfaction, and historical trends, enable park administrators to apply specific treatments aimed at improving visitor experiences and maintaining cultural heritage. In contrast, the Decision Tree model, while very interpretable and offering fast insights, showed overall performance with a lower F1-score of 79.8% and an accuracy of 82.3%. This perhaps means that while Decision Trees are efficient and hence used as an exploratory tool in the initial stages, their predictive power suffers in front of more robust models like Random Forests.

The SVM model has been noted for its capacity to handle high-dimensional data spaces with ease, whereas in this particular application, it showed an F1 score of 83.9% and an accuracy of 85.5%. These very credible metrics notwithstanding, it fell short of the Random Forest model in terms of predictive accuracy and overall robustness. This comparative study underlines the subtlety of model selection and the necessity of aligning the machine learning algorithms to the specific characteristics and requirements of the dataset.

The competitive performance registered by the SVM underlines its efficiency in cases where data relationships and decision boundaries are complex. This ability to detect patterns within high-dimensional feature spaces explains its real value for different predictive analytics tasks, including visitor satisfaction and preference forecasting at cultural heritage destinations. However, the slight underperformance compared with the Random Forest model underlines the need to consider further alternatives that could better grasp the complex dynamics involved in the visitors' behaviour and preferences. The results of this study underline the importance of using techniques from advanced machine learning to drive decision-making in managing cultural tourism. The Random Forest model makes very accurate predictions of visitor behaviour and preference, therefore providing the means for stakeholders to intervene in focused ways to optimize resource allocation, mitigate risks of over-tourism, and reinforce cultural experiences. Other factors, such as seasonal changes and socio-economic influences, ought to be taken into account in future research to further enhance the predictive model and arrive at better sustainable cultural tourism practices in heritage sites such as the Grand Canal National Cultural Park.

7. Conclusion

In other words, the evaluation of the machine learning model in the application of the culture and tourism integration model for the Grand Canal National Cultural Park showcases leveraging roles of data-driven approaches in the management of heritage sites. This study revealed that the Random Forest model stood out as the best for the prediction of the satisfaction and preference of visitors, with an F1-Score as high as 87.1% and a precision rate of 88.7%. It is effective in optimizing the visitors' experience without compromising cultural integrity since it generalizes well and robustly predicts. These findings illustrate the importance of the use of advanced analytics in strategic decisions relating to cultural tourism. The administrators and policymakers can apply appropriate interventions to the use of machine learning insights in their quest for a balance between the demand of tourists and cultural preservation imperatives. Such an approach will have impacts not only in terms of enhancing visitor satisfaction and engagement but also in ensuring sustainable management practices that are attractive to both tourists and the local communities. In this respect, going forward, continued research towards refining predictive models and real-time integration of data analytics is vital for engendering resilient and inclusive practices in cultural tourism at iconic heritage sites like the Grand Canal National Cultural Park.

References

[1] UNESCO, "The Grand Canal," 2016. [Online]. Available: https://whc.unesco.org/en/list/1443/

[2] Zhao Yun, The Grand Canal and its National Cultural Park, City of Architecture and Heritage, 2024. [Online]. Available: https://www.citedelarchitecture.fr/fr/article/grand-canal-and-its-national-cultural-park#:~:text=Relying%20on%20important%20resources%20to%20establish%20the,a%20symbol%20of%20China's%20diversity%20and%20integration.

[3] Muyan Tang, and Hongzhang Xu, "Cultural Integration and Rural Tourism Development: A Scoping Literature Review," *Tourism and Hospitality*, vol. 4, no. 1, pp. 75-90, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Roman Egger, "Machine Learning in Tourism: A Brief Overview: Generation of Knowledge from Experience," *Applied Data Science in Tourism, Generation of Knowledge from Experience*, pp. 85-107, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Ying Chen, "The Hot Spots and Frontiers of Research on the Grand Canal Culture Belt in China: Literature and Academic Trends," *Humanities and Social Sciences Communications*, vol. 9, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Ioanna Farsari, "Exploring the Nexus between Sustainable Tourism Governance, Resilience and Complexity Research," *Tourism Recreation Research*, vol. 48, no. 3, pp. 352-367, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Rohit Agrawal et al., "Big Data Analytics and Sustainable Tourism: A Comprehensive Review and Network-Based Analysis for Potential Future Research," *International Journal of Information Management Data Insights*, vol. 2, no. 2, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Zohreh Doborjeh et al., "Artificial Intelligence: A Systematic Review of Methods and Applications in Hospitality and Tourism," *International Journal of Contemporary Hospitality Management*, vol. 34, no. 3, pp. 1154-1176, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Yuan Li et al., "A Review of the Tools and Techniques Used in the Digital Preservation of Architectural Heritage within Disaster Cycles," *Heritage Science*, vol. 11, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Fuad Mehraliyev, Irene Cheng Chu Chan, and Andrei Petrovich Kirilenko, "Sentiment Analysis in Hospitality and Tourism: A Thematic and Methodological Review," *International Journal of Contemporary Hospitality Management*, vol. 34, no. 1, pp. 46-77, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Elliot Mbunge, and Benhildah Muchemwa, "Deep Learning and Machine Learning Techniques for Analyzing Travellers' Online Reviews: A Review," *Optimizing Digital Solutions for Hyper-Personalization in Tourism and Hospitality*, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Li Ying et al., "The Development of Digital Tourism in China," *Cogent Social Sciences*, vol. 10, no. 1, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Wen Zhen Li, and Hong Zhong, "Development of a Smart Tourism Integration Model to Preserve the Cultural Heritage of Ancient Villages in Northern Guangxi," *Heritage Science*, vol. 10, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] Yiwen Li, and Bing Qiu, "Using Deep Learning Approaches to Quantify Landscape Preference of the Chinese Grand Canal: An Empirical Case Study of the Yangzhou Ancient Canal," *Sustainability*, vol. 16, no. 9, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Yinghua He, and Lei Wu, "Analysis on Spatial Development Mode of Eco-Sports Tourism in Grand Canal Landscape Environment Culture Belt," *Environmental Monitoring and Assessment*, vol. 194, 2022. [CrossRef] [Google Scholar] [Publisher Link]