



Tourism Sales Prediction using Machine Learning

Varsha Panvelkar*,

Komal Kesharinath Gharat**,

**(Computer Science, CKT ACS College, NewPanvel (Empowered Autonomous))*

Email: panvelkarvarsha5@gmail.com

Email: komalgharat07@gmail.com

How to Cite this Article:

Panvelkar, V. & Gharat, K. K. (2026). Tourism Sales Prediction using Machine Learning. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(04).
<https://doi.org/10.55041/ijcope.v2i4.402>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i4.402>

ABSTRACT:

Tourism plays a vital role in global economic development by contributing significantly to employment generation and gross domestic product (GDP). Tourism data are collected in large amounts and are subject to seasonal effects and complex patterns. Hence, precise forecasting is quite challenging. In this context, this study proposes a predictive model that mines historical tourism data to uncover key determinants for tourism sales. The key technique is based on the Random Forest regression algorithm, which is widely used for predictive tasks due to its superior accuracy and reliability. Moreover, this technique is highly capable of capturing non-linear relationships among various determinants such as hotel prices, transportation costs, and historical sales records. The proposed model follows a series of processes to enhance the quality of predictions based on data mining techniques. The proposed model is evaluated based on various metrics, including Mean Absolute Error (MAE) and R-squared score, which is widely used for regression-based predictive models. The proposed model is found to yield superior accuracy and is highly beneficial for decision-makers in the tourism industry. This research also focuses on the development of a user-friendly predictive system, where in the users can input the required parameters and obtain real-time sales forecasting results. The usability of the system can be improved by adding visualization tools like trend graphs and performance indicators, which will help the users

understand the results of the forecasting process better.

Keywords: Tourism Sales Prediction, Data Mining, Machine Learning, Random Forest, Predictive Analytics, Ensemble Learning, and Flask Deployment.



I. INTRODUCTION:

The tourism industry is a major contributor to the world economy, creating employment and increasing the overall rate of economic growth. Making accurate predictions about sales can help businesses improve their strategic planning, pricing, and resource allocation. However, the tourism industry's data usually exhibits seasonal variations and complex patterns, which can make forecasting a challenging task [1]. The conventional statistical approaches might fail to identify non-linear relationships, which can affect the accuracy of the forecast [2]. Data mining enables us to find meaningful patterns and hidden relationships within our data, which improves decision-making in a business context [3]. Machine learning provides us with sophisticated methods to handle complex and high-dimensional data [4]. Among these methods, Random Forest is prominent for its stability, precision, and capacity to reduce overfitting in regression problems [5]. Several researchers have applied machine learning techniques to forecast tourism demands and sales. From the studies, it is evident that machine learning techniques have better accuracy and efficiency compared to traditional forecasting techniques [6]. Time series analysis techniques have become common in improving the accuracy of forecasting in the field of tourism and related fields [7]. The incorporation of data science techniques in business forecasting has also improved accuracy and usability [8]. The present paper proposes a system for predicting tourism sale forecasts using the power of machine learning, based on data mining techniques. The system uses data mining techniques to dive into the data, retrieve important features, and make accurate predictions using Random Forest regression. The system can be evaluated using various parameters like Mean Absolute Error (MAE) and R-squared score to measure its effectiveness [9]. The system provides insights to help decision-makers improve business strategies [10].

II. RELATED WORK:

In recent years, various researchers and developers have been working towards developing intelligent systems for tourism sales prediction using data-driven techniques. In previous systems, only basic statistical methods and analysis were used for tourism sales prediction, which gave low accuracy and required more human effort. Previous

systems were not capable of handling dynamic changes in tourism trends.

With the emergence of machine learning technology, modern methods have incorporated automated prediction models that use historical tourism data to provide accurate predictions. Various methods have been used to improve the prediction accuracy of these models, such as decision trees and regression methods. However, these methods still suffer from inadequate robustness in dealing with complex non-linear interactions of various factors such as hotel price, transportation cost, and customer demand. Recent developments have emphasized the importance and efficacy of ensemble learning methods in improving the accuracy of predictions. Models based on multiple decision trees have been found to perform better than single models. The prediction errors and generalization capability of these models have been improved, making them more appropriate for tourism applications. Moreover, multiple features have been used to gather more insights into tourism sales.

The accuracy of the prediction, modern systems have also emphasized the need for the overall user experience through the integration of interactive interfaces. The inclusion of dashboards, graphs, and trend analysis has enabled the users to better comprehend the prediction results. However, the majority of the prediction systems lack the integration of accuracy, usability, and real-time prediction within the same platform.

In order to overcome these limitations, the proposed project includes a tourism sales prediction system using a combination of machine learning and a user-friendly web interface. This system uses a random forest-based approach for accurate predictions and includes features like trend graphs and performance indicators. This not only provides accurate predictions but also makes the system more user-friendly

III. RESEARCH GAP:

Despite the advancements in tourism sales prediction using statistical and machine learning techniques, several limitations still exist in current research. Many traditional and earlier machine learning models are not capable of



effectively capturing complex non-linear relationships among key influencing factors such as pricing, seasonality, and customer demand. Furthermore, most existing studies focus primarily on model accuracy without developing integrated systems that support real-time prediction. There is also a lack of user-friendly interfaces, as previous systems do not adequately incorporate interactive visualization tools for better understanding of results.

IV. METHODOLOGY:

In this section, the methodology for designing and developing the tourism sales prediction system is discussed. The proposed system for tourism sales prediction is based on a structured approach to data collection, pre-processing, feature selection, training, prediction, and result visualization. The proposed system is based on machine learning for accurate and precise sales prediction.

A. Data Collection

The data is collected from various sources related to tourism and includes historical sales data along with other features such as hotel prices, transportation costs, number of customers, and seasonality. The collected data is used to train and test the prediction model. The data collected includes actual trends and patterns related to tourism, which helps to increase the reliability of the system.

B. Data Pre-processing

In this step, the raw data is processed and cleaned to prepare it for further analysis. In this step, missing values in the data are handled, and inconsistent data is removed to increase data quality. Numerical data is normalized to maintain uniformity. Data pre-processing is performed to increase the efficiency and accuracy of the machine learning model.

C. Feature Selection

The relevant features that have an impact on tourism sales are identified and selected. Some of the features considered in this case include previous sales, prices, and seasonal changes. Feature selection is essential in making the Model more efficient and ensuring that only relevant features are used. Feature selection is essential in ensuring that the model is more accurate and efficient.

D. Model Selection and Training

The Random Forest regression algorithm is chosen for the prediction process based on its precision and reliability. The model is trained using the pre-processed data, and multiple

decision trees are created during this process. The trees are combined to form a single prediction, which ensures that overfitting is avoided. The model learns from the features and makes a prediction based on the features and sales.

E. Prediction Process

The trained model is used to predict tourism sales based on user input. The parameters, such as price and previous sales values, are passed to the model. The parameters are processed, and a predicted sales value is returned. The prediction is efficient and quick.

F. Model Evaluation

The model is evaluated based on parameters such as Mean Absolute Error and R-squared score. The parameters measure the accuracy and reliability of the predictions. The lower the error, the better the model performs. The model is evaluated to ensure that it is producing accurate and precise results.

G. System Implementation

The system uses a framework for its implementation, which is a web-based framework. The backend of the system uses the Flask framework, while the frontend uses HTML, CSS, and JavaScript for its design. The system enables users to add data, make predictions, and analyse the results by using the interactive system. The system combines machine learning and web technology to create a complete system.

H. Data Visualization

The system uses visualization to display the results of the prediction in a form that can be easily understood by the user. The system uses graphs and charts to display the results of the prediction, which can be easily interpreted by the user. The system improves the user experience by ensuring that the results of the prediction can be easily interpreted by the user.

V. RESULTS AND DISCUSSION:

This section discusses the results obtained from the developed tourism sales prediction system, as well as the evaluation of the implemented machine learning model. The system was tested using inputs provided by the users, including the price of the hotel, price of the flight, marketing spend, weather index, festival indicator, past sales, month, and year, which were then processed by the implemented



machine learning model to obtain accurate tourism sales predictions.

The system has successfully predicted the tourism sales in real time using the Random Forest regression model. In this case, the model has generated the predicted tourism sales value of 1205 using the given input values. This has then been analysed with the previous sales value of 1300. The system has also calculated the percentage change and has indicated the decrease in the value by -7.31%. This has also proved the effectiveness of the model in not only predicting the values but also analysing the sales trends.

In order to improve the understanding of the data, the system also offers the option of displaying the previous sales and the predicted sales graphically, compared to each other, through the use of a bar chart. The difference between the two values is clearly depicted through the chart, making it easy for the users to interpret the data, whether the sales are increasing or decreasing. The inclusion of the option for changing the chart type, such as the bar chart and line chart, also improves the usability of the interactive system.

The inclusion of interactive features such as the download of the chart and the changing of the graph makes the system more usable. The inclusion of the visualization tools also makes the system more practical and usable compared to the prediction model.

The web-based interface is helpful in entering data and generating instant predictions with minimal technical difficulties. This simple and organized interface helps in improving the user experience and interacting with the system in the best way. The integration of machine learning prediction and visualization makes the system highly efficient in tourism sales forecasting.

Overall, the results show that the proposed system is capable of making accurate predictions and offering useful insights and an interactive experience for users. The proposed system is effective in capturing relationships between input features and sales output. This shows that the proposed system is reliable and efficient and can be used in real-world applications in the tourism industry.

VI. CONCLUSION:

In the present paper, the concept of an accurate machine learning-based tourism sales prediction system, which uses data mining techniques for precise predictions, has been proposed. The proposed system effectively integrates the Random Forest regression algorithm with the web interface, thus ensuring an efficient solution for the analysis of tourism

sales. The proposed model considers various parameters such as hotel price, flight price, marketing spend, weather index, festival, and previous sales, thus effectively understanding the complex relationships.

The developed system is capable of predicting sales for tourism and offers further insights through percentage change analysis to show an increase or decrease in sales. The inclusion of graphical visualization techniques such as bar charts and line charts makes the system easier to understand and allows for easy comparisons to be made for previous and predicted sales.

From the results, it is clear that the proposed system is effective and accurate. The integration of machine learning and other technologies makes the system ready for real-world applications. The system is effective in decision-making and helps in optimizing business strategies in the tourism industry.

VII. FUTURE WORK:

The Tourism Sales Prediction system can be further enhanced by integrating real-time data from social media, travel platforms, and local events to anticipate sudden changes in tourist behaviour. Incorporating advanced machine learning models like XGBoost or LSTM can improve long-term accuracy and capture complex patterns in tourism demand.

Future improvements could also include personalized recommendations for hotels and travel agencies, such as optimized pricing, marketing strategies, and promotions based on predicted trends. Expanding the system to cover multi-city or international tourism would help stakeholders identify emerging destinations and seasonal patterns, supporting strategic planning and investment.

Improving the user interface with interactive dashboards, mobile access, and AI explanations would make the system more intuitive and actionable. Finally, integrating sustainability metrics could help businesses plan eco-friendly tourism strategies while maintaining profitability. With these enhancements, the system has the potential to evolve into a comprehensive, intelligent platform for forecasting, planning, and optimizing tourism operations effectively.



VIII. REFERENCE:

1. Han, J., Kamber, M., & Mining, D. (2006). Concepts and techniques. *Morgan Kaufmann*, 340(1), 94104-103205.
2. Frank, E., & Hall, M. A. (2011). Data mining: practical machine learning tools and Techniques. Morgan Kaufmann.
3. Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
4. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning.
5. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*, 13(3), e0194889.
6. Şeker, F. (2023). Evolution of machine learning in tourism: A comprehensive review of seminal research. *Journal of Artificial Intelligence and Data Science*, 3(2), 54-79.
7. Box, G., & Jenkins, G. M. (1976). Analysis: Forecasting and control. *San Francisco*, 10.
8. Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. a. "O'Reilly Media, Inc."
9. Hao, J., & Ho, T. K. (2019). Machine learning made easy: a review of scikit-learn package in python programming language. *Journal of educational and behavioural statistics*, 44(3), 348-361.
10. Kuhn, M., & Johnson, K. (2013). *Applied predictive modelling* (Vol. 26, p. 13). New York: Springer.