

Hybrid Deep Learning-Based Tourism Demand Forecasting Using Big Data Analytics

Dr. Amit Jain

Professor, Department of Computer Science and Engineering,

Jindal University, Raigarh, C.G

amit.jain@opju.ac.in

Abstract: *The tourism industry is increasingly relying on data-driven decision-making to gain insights into consumer behavior, manage hotel operations and enhance tourism planning. The growing volume of booking-related data offers valuable opportunities to identify demand patterns and improve prediction accuracy; however, extracting meaningful insights from complex and heterogeneous tourism data remains challenging. This study proposes an Artificial Neural Network (ANN)-based approach to predicting tourism demand using the Hotel Booking Demand Dataset. The proposed methodology analyses hotel booking information to identify significant patterns and develop an effective predictive model. Accuracy, precision, recall, and F1-score are used to assess the proposed ANN. Experimental results demonstrate that the proposed ANN achieves an accuracy of 98.91%, precision of 95.96%, recall of 99.11%, and F1-score of 98.01%. The results suggest that the ANN can effectively capture the complex relationships embedded in hotel booking data and offers reliable predictive performance. The proposed approach can therefore support accurate tourism demand forecasting and assist stakeholders in making informed, data-driven decisions for improved tourism management and planning.*

Keywords: Smart Tourism, Economic Impact, Sustainable Innovation, Predictive Analytics, Sentiment Analysis, Data Governance

I. INTRODUCTION

Tourism is an important contributor to economic growth, employment generation, foreign exchange earnings, and regional development. The rapid expansion of digital technologies has transformed tourism activities, creating new opportunities for destinations, businesses, and travelers [1][2]. With the emergence of the Internet of Things (IoT) and Artificial Intelligence (AI), the tourism sector is increasingly moving toward smart tourism, where intelligent services and data-driven decision-making are important. Accurate tourism demand forecasting is essential for resource allocation, operational planning, and improved tourist services [3][2][4].

Big Data has transformed the way tourism information is generated, collected, and analyzed. Tourism-related Big Data can originate from online travel platforms, search engines, social media, mobile devices, sensors, transaction records, and user-generated content [5]. Such information provides valuable insights into tourist preferences, travel behavior, destination popularity, and demand patterns [6][7]. Internet search trends can also reflect travelers' intentions before their actual trips, providing useful information for tourism demand forecasting [8][9]. However, data quality, privacy, infrastructure, and data heterogeneity remain important challenges [10][11].

Accurate tourism demand forecasting is therefore essential for tourism managers, hoteliers, destination authorities, and other stakeholders. Traditional approaches generally depend on historical tourist arrivals, tourism revenue, time-series information, and macroeconomic variables, but may struggle with volatile demand, nonlinear relationships, and spatial heterogeneity [12]. Tourism activities also demonstrate spatial dependencies, where changes in visitor flows at major



destinations can influence surrounding regions through attraction and spillover effects [13][14]. Therefore, forecasting models need to incorporate diverse data sources and capture complex tourism patterns [15].

Machine Learning (ML) and Deep Learning (DL) have emerged as effective approaches for analyzing large and complex tourism datasets. Conventional ML methods often depend on manual feature engineering, which may limit the representation of complex data patterns [16]. In contrast, DL models use multiple hidden layers to automatically learn high-level features, making them suitable for nonlinear and dynamic forecasting problems. Hybrid deep-learning approaches can combine complementary learning capabilities to improve prediction [17][18]. Therefore, Hybrid Deep Learning-Based Tourism Demand Forecasting Using Big Data Analytics can effectively integrate large-scale tourism data, automatically extract relevant features, and improve the accuracy and adaptability of tourism demand prediction [19].

A. Motivation and Contribution

This study has been motivated by the increasing volume and complexity of tourism data, necessitating accurate and intelligent forecasting of tourism demand. The aim of the study is to use big data and advanced deep learning methods to capture complex tourism patterns and enhance prediction performance. It also aims to facilitate effective planning, allocation of resources and decision-making in the tourism sector based on data. This research offers several key contributions as listed below:

- Developed an ANN-based tourism demand prediction model using the Hotel Booking Demand Dataset.
- Analyzed hotel booking patterns and customer-related attributes to identify meaningful tourism demand trends.
- Applied appropriate data preprocessing, normalization, and class-balancing techniques to improve data quality.
- Evaluated the proposed ANN using accuracy, precision, recall, and F1-score.
- Provided a data-driven approach that can support tourism planning, hotel management, and informed decision-making.

This study is justified by the increasing complexity of tourism and hotel booking data, which demand an intelligent solution to recognize the patterns of tourist demand to help in effective forecasting. The novelty of the proposed work is to employ a framework based upon an ANN to identify the complex and nonlinear relationship embedded in the hotel booking datasets for forecasting tourism demand. This has resulted in an efficient and data-based framework for the analysis of tourism demand. It may also aid in good tourism planning and informed decision-making.

B. Paper Organization

The structure of the paper is as follows: Section II presents related work, Section III describes the dataset and methodology, Section IV discusses the experimental results, and Section V concludes the study with future research directions.

II. LITERATURE REVIEW

Existing research on Big Data Analytics for Tourism Demand Prediction was thoroughly reviewed to identify relevant approaches and strengthen the study's foundation.

Kaul (2026) proposed that the performance of the traditional machine learning (ML) algorithms: Logistic Regression, Random Forest, Decision Tree, and SVM is compared to a U-Net deep learning model for classification tasks. The results indicate that while conventional models capture moderate accuracy levels, Logistic Regression is at 69 %, Random Forest is at 67 %, but U-Net is ahead of the rest with 82.5%. Precision (73.87%), recall (98.92%), and F1-score (83.75 %) indicate that the model is better at recognizing complex data patterns. Precision in the model's performance throughout the epochs indicates that the model is successful in learning and not overfitting [20].

Kamali et al. (2026) propose a spatial graph-based forecasting framework using a KD-Tree proximity network between 1,455 hotels and 452,044 tourism activities from the Travor Challenge dataset. PageRank and Degree Centrality are extracted from a bipartite graph of 8,979 nodes and 43,650 edges as predictive features for LightGBM, XGBoost, and



Random Forest under Baseline and Proposed architectures. Five-Fold Cross-Validation ensures zero label leakage, while ablation analysis compares graph features with four spatial proximity baselines. LightGBM Proposed achieves the lowest RMSE of 1.1684, with a significant 3.23% improvement over Baseline ($p=0.0103$). SHAP analysis identifies PageRank as the dominant predictor (mean SHAP=0.2273), while Degree Centrality provides zero additional predictive signal [21].

Sharma and Pandey (2025) the Adaptive Sustainability-Driven Predictive Model (ASDPM), an AI-driven system for sustainable tourism forecasting and optimization. It uses LSTM networks and reinforcement modeling to predict visitor demand, stress levels, and travel behaviors. Real-time IoT, user interactions, and ecological data enable continuous adaptation of recommendations. The model also supports sustainable path optimization, policy modeling, and cloud-based stakeholder collaboration to balance tourism growth with environmental conservation. ASDPM achieves an overall predictive accuracy of 94.9% in sustainable tourism applications [22].

Q. Ma (2025) presents a prediction model based on neural networks that identifies intricate correlations among variables such as tourist arrival, seasonal patterns, consumption, and investment in infrastructure. A feedforward neural network with multiple layers is trained on past tourism data, optimized by backpropagation and data normalization. The model's output predicts future tourism growth rates and revenue. Accuracy and Precision score are used for performance benchmarking. The neural network provides 97.2% greater accuracy than conventional models. This facilitates informed decision-making by policymakers and tourism boards for resource allocation, advertisement campaigns, and infrastructure development. The solution presents a useful instrument for tourism economics strategic planning [23].

W. A. S. M. S. Thilakarathna and I. U. Hewapathirana (2024) explore the impact of weather data on tourism demand prediction in Sri Lanka using LSTM, Random Forest, and Support Vector Regression with monthly data from January 2017 to October 2023. Models were trained using historical tourist arrivals alone and with current and historical air quality data. Support Vector Regression performed best, with MSE values of 0.01 and 0.16, respectively. However, models using only historical tourist arrival data outperformed those incorporating weather data, indicating that external weather factors did not improve forecasting accuracy [24].

Y. Kosungnoen and T. Banditwattanawong (2024) aim to predict the number of foreign tourists visiting Thailand and the revenue generated from them using the Long Short-Term Memory (LSTM) algorithm in comparison with the popular statistical method in economics, the Autoregressive Integrated Moving Average (ARIMA). They used regional and national tourism datasets collected by the Ministry of Tourism and Sports, along with performance metrics: normalized root mean square error (NRMSE) and normalized mean absolute error (NMAE). Results reveal that LSTM yielded NRMSE values better than ARIMA, ranging from 1.18 times to 19.33 times, and NMAE values better than ARIMA, ranging from 1.40 times to 21.70 times across all datasets, except for the revenue forecast using the central region dataset, where ARIMA outperformed LSTM [25].

Table I summarizes recent studies on Big Data Analytics and tourism demand prediction, highlighting the approaches adopted, datasets used, key outcomes, and identified research limitations.

Table I Comparative Analysis of the Study on Big Data Analytics Approaches for Tourism Demand Forecasting

Author	Proposed Work	Outcome	Key Findings	Limitations & Future Work
Kaul (2026)	U-Net deep learning model for hotel booking cancellation classification	Accuracy: 82.5%; Precision: 73.87%; Recall: 98.92%; F1-score: 83.75%	U-Net outperformed traditional ML models and effectively captured complex booking patterns.	Focuses mainly on cancellation classification rather than tourism demand forecasting; future work can investigate broader forecasting tasks and hybrid architectures.



Kamali et al. (2026)	Spatial graph-based tourism forecasting using KD-Tree networks and graph-derived features with ensemble models	Proposed LightGBM achieved RMSE of 1.1684, with a 3.23% improvement over baseline	PageRank provided significant predictive information, demonstrating the value of spatial and structural tourism features.	Primarily uses ensemble learning; future research could integrate deep learning and hybrid forecasting architectures.
Sharma & Pandey (2025)	Adaptive Sustainability-Driven Predictive Model using LSTM, reinforcement modeling, IoT, and ecological data	Overall accuracy: 94.9%	Demonstrated the usefulness of AI and LSTM for sustainable tourism prediction and adaptive decision-making.	Broad sustainability focus rather than direct tourism demand forecasting; future work can strengthen hybrid deep learning forecasting.
Q. Ma (2025)	Multilayer feedforward neural network for tourism prediction using historical tourism variables	Accuracy: 97.2%	Neural networks captured relationships among tourist arrivals, seasonality, consumption, and infrastructure investment.	Does not emphasize hybrid deep learning or large-scale big data analytics; future work could explore hybrid models and larger datasets.
W. A. S. M. S. Thilakarathna and I. U. Hewapathirana (2024)	Tourism-demand forecasting using LSTM, Random Forest, and SVR with historical and weather-related features	SVR achieved the best performance with MSE values of 0.01 and 0.16	Historical tourist arrival data were more effective than the added weather variables in their experiments.	Limited to Sri Lankan monthly data and does not employ a hybrid deep-learning framework; additional variables and larger datasets could be investigated.
Y. Kosungnoen and T. Banditwattanawong (2024)	LSTM-based prediction of foreign tourist arrivals and tourism revenue compared with ARIMA	LSTM generally outperformed ARIMA, with NRMSE and NMAE improvements across most datasets	Demonstrated the effectiveness of LSTM for tourism arrival and revenue forecasting.	Hybrid architectures and large-scale heterogeneous tourism data were not explored; future research could integrate multiple deep learning models and big data sources.

Research Gaps: Existing tourism demand forecasting studies have mainly focused on traditional statistical methods and individual machine learning or deep learning models, with limited integration of diverse tourism data sources. Current methods are based on past tourism or hotel booking data, which may not completely reflect nonlinear, multidimensional and dynamic hotel demand. In addition, there has been little focus on effective data preprocessing and handling class imbalance in the hotel booking prediction problem. Moreover, the exploitation of complex relationships in large-scale tourism data using intelligent neural architectures remains relatively underexplored.



Therefore, there is a need for an effective deep learning-based framework that can improve tourism demand prediction and support reliable, data-driven tourism planning.

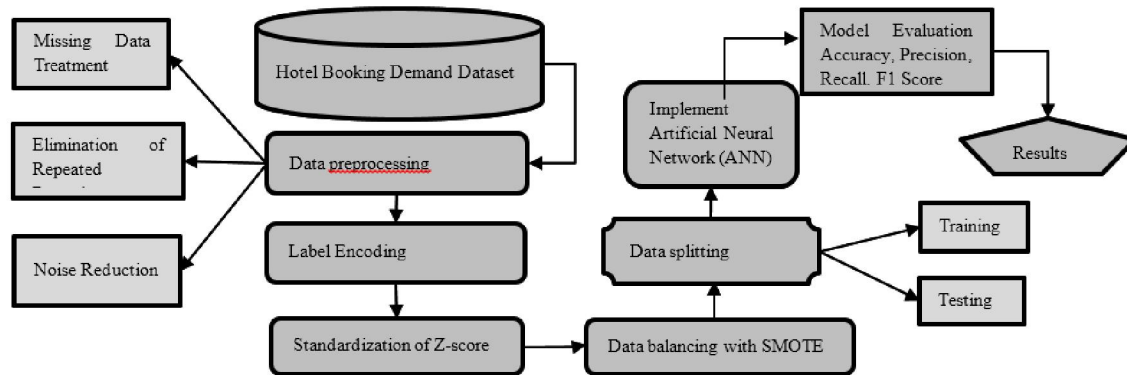


Fig1. Workflow of the Proposed Big Data Analytics-Based Tourism Demand Prediction Model

III. RESEARCH METHODOLOGY

The proposed methodology starts with data collection and analysis of the Hotel Booking Demand Dataset, followed by data cleaning, handling missing values, removing duplicates, noise reduction, label encoding, and standardization using Z-Scores. The class distribution is then balanced with SMOTE, after which a 70:30 ratio between the training and testing sets is used, with stratified sampling. An Artificial Neural Network (ANN) is trained to learn complex nonlinear relationships and predict tourism booking outcomes, and its accuracy is evaluated using precision, recall, F1 score and confusion matrix analysis. Fig. 1 presents the proposed framework for tourism demand prediction using Big Data Analytics.

The following section presents a detailed description of each step involved in the proposed methodology:

A. Data Collection and Analysis

This study used the Hotel Booking Demand Dataset, comprising 119,390 records and 32 attributes, to support tourism demand forecasting through big data analytics. Data visualizations such as bar plots and heatmaps were used to examine booking patterns, customer characteristics, and feature correlations. These visual analyses helped identify important trends and relationships within the dataset for effective tourism demand forecasting:



Fig.2 Bar Graph of Class Distribution

Fig. 2 shows the distribution of bookings according to their cancellation status, with 63,445 non-canceled bookings (0) and 24,000 canceled bookings (1). The number shown suggests that more bookings are not canceled than are canceled, suggesting that there is a class distribution in the data set.



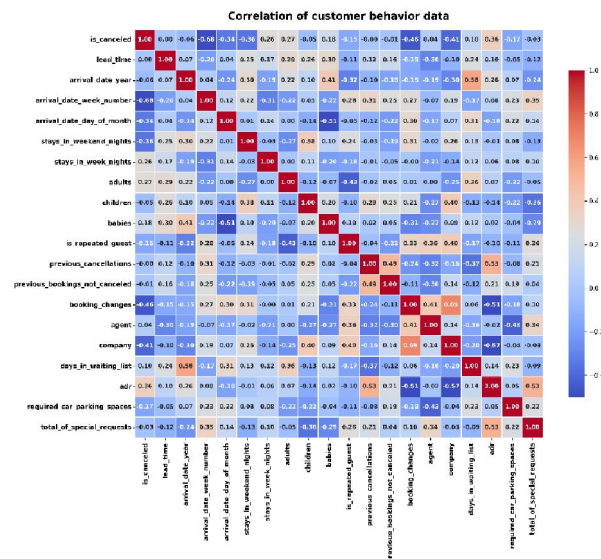


Fig3. Correlation Matrix Heatmap

The heatmap in Fig. 3 represents the correlation matrix of customer booking features, with yellow to dark blue indicating weak, negative, and strong relationships. Darker blue highlights positive correlations, while light yellow shows negative associations, such as between “weekend” and “weeknight” stay (0.56), “repeated guests” and “non-canceled bookings” (0.44), and arrival year and week number (-0.51). Grid lines and bold values make key patterns across all 20 variables easy to read.

B. Data Pre-Processing

The Hotel Booking Demand Dataset was utilized for data preparation, including data integration, cleaning, and transformation. The preprocessing phase involved identifying and handling missing values, removing duplicate and redundant records, reducing noise, and subsequently performing label encoding and normalization. These procedures improved the consistency, quality, and suitability of the dataset for subsequent analysis and model development. The key preprocessing steps are outlined below:

- **Missing Data Treatment:** Incomplete records and instances with unavailable class labels were identified and removed from the dataset. This step ensured that only complete and reliable observations were retained for subsequent analysis and model development.
- **Elimination of Repeated Records:** Duplicate observations were detected and removed to prevent the same information from being considered multiple times. This process reduced redundancy and ensured that each record contributed uniquely to the analysis.
- **Noise Reduction:** Unwanted variations, inconsistent values, and irrelevant information were identified and removed during preprocessing. Noise reduction enhanced the quality and consistency of the dataset, thereby supporting more reliable analytical and predictive outcomes[26].
- **Class Label Encoding:** Categorical attributes and class labels were transformed into numerical representations using label encoding. Each distinct category was assigned a corresponding numerical value, making the processed data suitable for machine learning algorithms.

C. Standardization of Z-score

To ensure consistency across features and prevent dominance of any single feature, the dataset underwent Z-score standardization. Each feature was rescaled by subtracting its mean and dividing by its Standard Deviation. This process



standardized all numerical attributes, improving the model's training stability and convergence. Equation (1), for Z-score standardization:

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

- Z = Standardized Value
- X = Original Value
- μ = mean of the feature
- σ = Standard Deviation of the feature

This transformation ensures that the mean of the data is zero and the Standard Deviation is one.

D. Data balancing using SMOTE

Data balancing is an essential preprocessing step for addressing class imbalance, where some classes contain fewer samples than others. In this study, SMOTE was used to improve class distribution by generating synthetic minority-class samples. It identifies k-nearest neighbors using Euclidean distance and creates new instances by interpolating between selected samples. This increases minority-class representation, reduces class bias, and improves the model's ability to learn patterns across all classes.



Fig. 4 Bar Graph of Class Distribution After SMOTE

Fig. 4 bar graph of class distribution after SMOTE shows that the two classes, non-canceled (0) and canceled (1), are perfectly balanced. The total number of records per class is 63,445, confirming that the dataset of bookings to be canceled was balanced using SMOTE. This well-balanced distribution helps overcome class imbalance and make training machine learning models more reliable.

E. Data Splitting

The dataset was split into training and testing subsets using a stratified 70:30 split. This approach ensured that the relative class distributions in both subsets were consistent with the original dataset's class distribution.

F. Proposed Artificial Neural Network (ANN) model

ANN is a non-linear computational model widely used in machine learning, inspired by biological neural networks. It consists of interconnected neurons capable of learning, generalizing training data, and deriving conclusions from complex data. Its application in construction risk management has gained attention.

Inspired by the structure and cognitive processes of the human brain, this model provides effective pattern recognition and complex data processing. Its ability to learn non-linear relationships between features distinguishes it from linear regression and enables automatic feature transformation for precise predictions. This flexibility is valuable for the multifaceted and unpredictable nature of construction data. For cyber risk assessment in construction, a feedforward ANN is selected as the uniform model. It consists of multiple layers, with hidden layers performing automatic feature



transformation. The operation of a hidden layer with m neurons following a layer with n neurons is described by Equation (2).

$$h_j = \sigma \left(\sum_{i=1}^n w_{ij} x_i + b_j \right) \quad (2)$$

Where:

h_j —the value of the j th neuron in the hidden layer,

x_i —the value of the i th neuron in the previous layer,

w_{ij} —the weight of the i th neuron in the previous layer that connects to the j th neuron in the hidden layer,

b_j —the bias term for the j th neuron in the hidden layer,

σ —the activation function.

The proposed ANN model was configured with 100 epochs, a batch size of 32, a learning rate of 0.001, the Adam optimizer, ReLU activation in the hidden layers, sigmoid activation in the output layer, a dropout rate of 0.5, and binary cross-entropy as the loss function. The network consisted of two hidden layers with 128 and 64 neurons, respectively, to effectively learn complex nonlinear relationships in the tourism booking data.

G. Evaluation metrics

Various evaluation metrics are used to evaluate machine learning models and explore the best model among all of them. The confusion matrix represents individual instances as true positive (TP), false positive (FP), true negative (TN) and false negative (FN). The performance of the model is evaluated by different metrics: Accuracy, AUROC and F-Measure in this work, as detailed below:

Accuracy: The ratio of the number of instances correctly predicted by the trained model to the total number of instances in the dataset (input samples). It is given as (3)-

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

Precision: Precision is the proportion of positive instances successfully predicted to all positive instances predicted by the model. Precision indicates. How good the classifier is in predicting the positive classes is expressed as (4)-

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Recall: This metric, the ratio of events that were accurately predicted as positive to all instances that should have proved positive. In mathematical form it is given as (5)-

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

F1 score: It is a combination of the harmonic mean of precision and recall, that is, it helps to balance recall and precision. Its range is [0, 1]. Mathematically, it is given as (6)-

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

IV. RESULTS AND DISCUSSION

This section presents the experimental evaluation of the proposed tourism demand forecasting model and discusses its classification performance.

A. Experimental Assessment

The experiments were implemented on a high-performance computing system equipped with an Intel Core i9-13900K processor (3.0 GHz), 64 GB DDR5 RAM, and an NVIDIA RTX 4090 GPU with 24 GB VRAM, operating on Windows 11 Pro. The model was developed and evaluated using Python-based libraries, including Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn.

B. Results and analysis

The Hotel Booking Demand Dataset was utilized for model training and testing, and the performance of the Artificial Neural Network (ANN) was evaluated using accuracy, precision, recall, and F1-score. As shown in Table II, the accuracy, precision, recall, and F1-score of the proposed model were 98.91%, 95.96%, 99.11%, and 98.01%, respectively, indicating effective classification for tourism demand forecasting. The above results show that the



approach ensured reliable prediction performance and identified tourism demand patterns from the processed booking data.

Table II. Experimental Classification Performance of the Proposed Tourism Demand Forecasting Model

Matrix	Artificial Neural Network (ANN)
Accuracy	98.91
Precision	95.96
Recall	99.11
F1-score	98.01

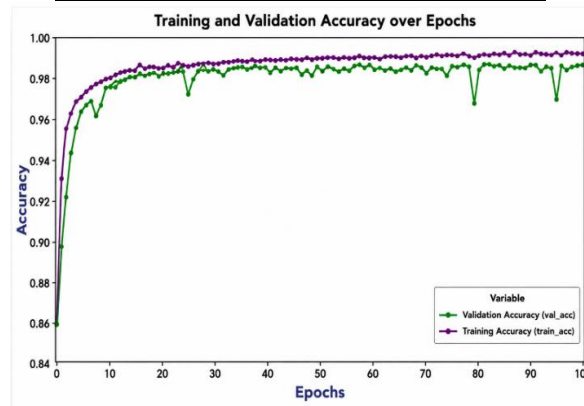


Fig. 5 Training and Validation Accuracy Curve for ANN

Fig. 5 shows the training and validation accuracy of the Artificial Neural Network (ANN) over 100 epochs. Both curves increase rapidly during the initial epochs and then stabilize at a high accuracy level, indicating effective learning and stable model convergence. The small fluctuations in validation accuracy demonstrate consistent generalization performance without significant overfitting.

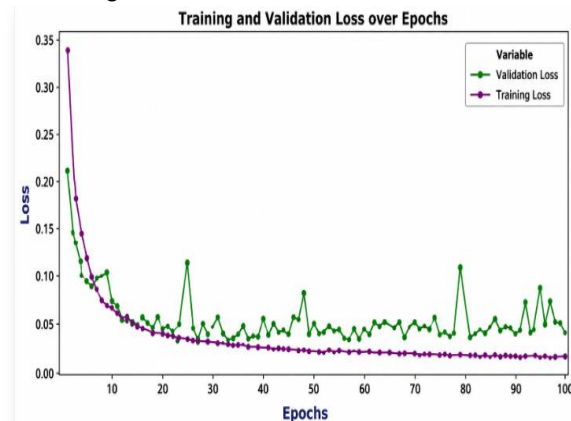


Fig. 6 Training and Validation Loss Curve for ANN

Fig. 6 illustrates the training and validation loss of the Artificial Neural Network (ANN) across 100 epochs. The training loss decreases steadily, while the validation loss shows minor fluctuations but remains relatively low, indicating effective learning and good model convergence.



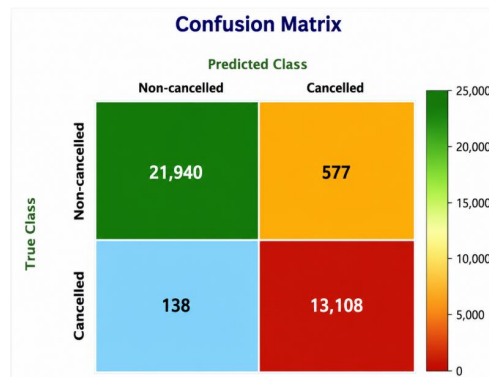


Fig. 7 Confusion Matrix for ANN

Fig. 7 illustrates the classification performance of the Artificial Neural Network (ANN) for non-cancelled and cancelled bookings. The model correctly classifies 21,940 non-cancelled and 13,108 cancelled instances, with only 577 and 138 misclassified cases, respectively. These results demonstrate that the ANN provides highly accurate and reliable classification performance with a low number of prediction errors.

C. Comparative Performance Analysis

To assess the effectiveness of the proposed model, its predictive performance was compared with several existing machine learning and deep learning approaches, as presented in Table II. The comparison was conducted using accuracy, precision, recall, and F1-score as the primary evaluation measures. Among the baseline models, Logistic Regression achieved an accuracy of 96.89%, while XGBoost and Random Forest obtained accuracies of 78.34% and 76.27%, respectively. In contrast, the proposed Artificial Neural Network (ANN) achieved the highest accuracy of 98.91%, along with precision of 95.96%, recall of 99.11%, and an F1-score of 98.01%. These results demonstrate that the proposed ANN provides superior and consistent classification performance for tourism demand forecasting compared with the selected existing models.

Table III. Performance Comparison of Various Machine Learning and Deep Learning Approaches for Tourism Demand Forecasting

Model	Accuracy	Precision	Recall	F1-score
RF[27]	76.27	-	-	72.81
XGBoost [28]	78.34	80.63	92.26	86.05
LR[29]	96.89	96.20	99.06	97.61
Proposed ANN	98.91	95.96	99.11	98.01

D. Discussion and Analysis

The proposed ANN achieved a high classification accuracy of 98.91%, demonstrating its strong predictive capability for tourism demand forecasting. This superior accuracy indicates that the model can effectively learn complex patterns within the tourism booking data. Compared with the existing models, the ANN provides more reliable and accurate demand prediction. Therefore, the proposed ANN offers a significant advantage in improving the overall accuracy of tourism demand forecasting.

V. CONCLUSION AND FUTURE STUDY

Tourism is associated with rest and vacation for people who spend most of their day working, adapting to changing lifestyles and global standards. With the continued growth in tourism demand, its contribution to the global economy is increasing significantly, making accurate prediction of tourism demand essential for effective planning and decision-



making. In this study, the Hotel Booking Demand Dataset was analyzed, and an Artificial Neural Network (ANN) model was developed for tourism demand prediction. The proposed ANN achieved an accuracy of 98.91%, outperforming RF (76.27%), XGBoost (78.34%), and LR (96.89%). These findings demonstrate that the proposed ANN can effectively capture complex patterns in hotel booking data and provide reliable prediction results. Therefore, the proposed approach offers a promising solution for improving tourism demand forecasting and supporting data-driven decision-making in the tourism sector. One drawback of this work is the use of a single hotel booking dataset, which could limit the approach's generalizability to other datasets and tourism regions. Future work can incorporate real-time social media, weather, and destination-related data with advanced hybrid deep learning models to improve the system's adaptability and forecasting capabilities.

REFERENCES

1. H. Song, R. T. R. Qiu, and J. Park, "A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting," *Ann. Tour. Res.*, vol. 75, pp. 338–362, Mar. 2019, doi: 10.1016/j.annals.2018.12.001.
2. H. Li, M. Hu, and G. Li, "Forecasting tourism demand with multisource big data," *Ann. Tour. Res.*, vol. 83, no. March, p. 102912, 2020, doi: 10.1016/j.annals.2020.102912.
3. A. Georgakis, V. Profilidis, and G. Botzoris, "Forecasting tourism demand using search engine data," in *ICTR2021 – 10th International Congress on Transportation Research*, Rhodes, Greece: ICTR, 2021, pp. 1–13, September.
4. F. Xu, "Research on Tourism Demand Forecasting Model Based on Machine Learning Algorithm," in *2024 International Conference on Internet of Things, Robotics and Distributed Computing (ICIRDC)*, 2024, pp. 82–86. doi: 10.1109/ICIRDC65564.2024.00021.
5. V. Misal, V. P. Janeja, S. C. Pallaprolu, Y. Yesha, and R. Chintalapati, "Iterative unified clustering in big data," in *2016 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA: IEEE, 2016, pp. 3412–3421, December. doi: 10.1109/BigData.2016.7841002.
6. H. Song and G. Li, "Tourism demand modeling and forecasting—A review of recent research," *Tour. Manag.*, vol. 29, no. 2, pp. 203–220, Apr. 2008, doi: 10.1016/j.tourman.2007.07.016.
7. S. Shen, G. Li, and H. Song, "An Assessment of Combining Tourism Demand Forecasts over Different Time Horizons," *J. Travel Res.*, vol. 47, no. 2, pp. 197–207, Nov. 2008, doi: 10.1177/0047287508321199.
8. S. S. Padhi and R. K. Pati, "Quantifying potential tourist behavior in choice of destination using Google Trends," *Tour. Manag. Perspect.*, vol. 24, pp. 34–47, Oct. 2017, doi: 10.1016/j.tmp.2017.07.001.
9. X. Li, Q. Wu, G. Peng, and B. Lv, "Tourism forecasting by search engine data with noise-processing," *African J. Bus. Manag.*, vol. 10, no. 6, pp. 114–130, Mar. 2016, doi: 10.5897/AJBM2015.7945.
10. [10] S. Li, T. Chen, L. Wang, and C. Ming, "Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index," *Tour. Manag.*, vol. 68, pp. 116–126, Oct. 2018, doi: 10.1016/j.tourman.2018.03.006.
11. M. Atreya, V. B. Meruga, R. K. Kanneganti, S. R. Puli, P. Kanungo, and S. Chintala, "Federated AI Orchestration in Heterogeneous Edge Environments," in *2025 IEEE International Conference on Emerging Trends in Computing and Communication (ETCOM)*, Mangalore, India: IEEE, 2026, pp. 1–7, November. doi: 10.1109/ETCOM66606.2025.11436271.
12. K. Prasanthi, K. Mounish, B. Venkateswarlu, A. Mallikarjun, S. Pusalra, and B. V. Ramana, "Comparative Evaluation of Ensemble Machine Learning Models for Hotel Booking Cancellation Prediction," in *2026 Third International Conference on Networking and Communications (ICNWC)*, 2026, pp. 1–6. doi: 10.1109/ICNWC68145.2026.11518531.
13. G. Li, H. Song, and S. F. Witt, "Recent Developments in Econometric Modeling and Forecasting," *J. Travel Res.*, vol. 44, no. 1, pp. 82–99, Aug. 2005, doi: 10.1177/0047287505276594.



14. R. Law, G. Li, D. K. C. Fong, and X. Han, "Tourism demand forecasting: A deep learning approach," *Ann. Tour. Res.*, vol. 75, pp. 410–423, Mar. 2019, doi: 10.1016/j.annals.2019.01.014.
15. G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006, doi: 10.1016/j.neucom.2005.12.126.
16. T. Dergiades, E. Mavragani, and B. Pan, "Google Trends and tourists' arrivals: Emerging biases and proposed corrections," *Tour. Manag.*, vol. 66, pp. 108–120, Jun. 2018, doi: 10.1016/j.tourman.2017.10.014.
17. G. I. Crouch, "Effect of income and price on international tourism," *Ann. Tour. Res.*, vol. 19, no. 4, pp. 643–664, Jan. 1992, doi: 10.1016/0160-7383(92)90059-X.
18. Y. Li, Z. Lin, and S. Xiao, "Using social media big data for tourist demand forecasting: A new machine learning analytical approach," *J. Digit. Econ.*, vol. 1, no. 1, pp. 32–43, 2022, doi: 10.1016/j.jdec.2022.08.006.
19. K. Dixit, "Artificial Intelligence-Driven Prediction of Hotel Booking Demand for Revenue Optimization," *Int. J. Curr. Eng. Technol.*, vol. 14, no. 6, pp. 267–274, Nov/Dec, 2024, doi: 10.14741/ijcet/v.14.6.20.
20. D. Kaul, "U-Net Deep Learning Model to Forecast Guest Cancellations in Hotel Bookings," in 2026 13th International Conference on Computing for Sustainable Global Development (INDIACom), 2026, pp. 1–8. doi: 10.23919/INDIACom70271.2026.11525479.
21. [M. A. Kamali, M. Angello Qadosy Riyadi, S. M. Al Mujahidin, and N. Alifia Rustan, "Spatial Graph-Enhanced Accommodation Demand Forecasting Using KD-Tree: Algorithm-Agnostic Validation with Explainable AI," in 2026 14th International Conference on Information and Communication Technology (ICoICT), 2026, pp. 1–6. doi: 10.1109/ICoICT71118.2026.11664757.
22. S. Sharma and N. Pandey, "Enhancing Sustainable Tourism with AI-Powered Cloud-Based Predictive Models for Intelligent Travel Destinations," in 2025 2nd International Conference On Multidisciplinary Research and Innovations in Engineering (MRIE), 2025, pp. 529–534. doi: 10.1109/MRIE66930.2025.11156218.
23. Q. Ma, "A Novel Prediction of Tourism Economic Development Trends based on Machine Learning," in 2025 International Conference on Intelligent Computing and Knowledge Extraction (ICICKE), 2025, pp. 1–7. doi: 10.1109/ICICKE65317.2025.11136217.
24. W. A. S. M. S. Thilakarathna and I. U. Hewapathirana, "Predictive Modeling for Tourist Arrivals: Assessing the Impact of Weather Data Using Machine Learning Models," in 2024 International Conference on Advances in Technology and Computing (ICATC), 2024, pp. 1–6. doi: 10.1109/ICATC64549.2024.11025220.
25. Y. Kosungnoen and T. Banditwattanawong, "Comparative Analysis of Deep Learning and Statistical Algorithms for Forecasting Inbound Foreign Tourism in Thailand," in 2024 8th International Conference on Information Technology (InCIT), 2024, pp. 769–774. doi: 10.1109/InCIT63192.2024.10810650.
26. B. Kantheti and C. A. P. De Araujo, "Decoupling Privacy Noise from Optimization in Transformer Forecasting," *Mach. Learn. Knowl. Extr.*, vol. 8, no. 6, pp. 156, May, 2026, doi: 10.3390/make8060156.
27. S. Satu, "Performance Analysis of Machine Learning Techniques to Predict Hotel Booking Cancellations in Hospitality Industry," pp. 19–21, 2020.
28. R. Chatziladas, "Predicting Hotel Booking Demand And Cancellations Using Machine Learning And Comparison Of Feature Importance," *Dr. Diss. Dr. Diss. Tilbg. Univ.*, 2024.
29. I. Gómez-talal, M. Azizoltani, P. Talón-Ballester, and A. Singh, "Machine Learning in Hospitality : Interpretable Forecasting of Booking Cancellations," 2025, doi: 10.1109/ACCESS.2025.3536094.

