**Predicting Tourist Arrivals Using a Hybrid Adaptive Forecasting Approach**

Abdurakhim Murodov1, а), Damir Tajibaev1, b), Ziyauddin Kamoladdinov1, c) and  
Diyorjon Khakimov1, d)

1*Tashkent State Transport University, 1 Temiryulchilar St., Tashkent 100167, Uzbekistan*

*a) Corresponding author:* [*abdurahimmurodov579@gmail.com*](mailto:abdurahimmurodov579@gmail.com) *b)*[*tajibaevdami@gmail.com*](mailto:tajibaevdami@gmail.com) *c)[kamoliddiziyo1331@gmail.com](mailto:kamoliddiziyo1331@gmail.com)  
d)[xakimovdiyorjon1817@gmail.com](mailto:xakimovdiyorjon1817@gmail.com)*

**Abstract.** Sustainable tourism development in regions with cultural heritage, such as Shakhrisabz (Uzbekistan), requires accurate forecasting of tourism demand. This article presents the Hybrid Adaptive Tourism Forecasting Framework (HATFF), which integrates ARIMA, Random Forest, adaptive adjustment, and environmental constraints. The application of HATFF to Shakhrisabz data for 2024 achieved a Mean Absolute Error (MAE) of 1.8 thousand tourists, outperforming ARIMA (MAE 3.2), Random Forest (MAE 2.8), and iTransformer (MAE 2.3). A simplified mathematical model with an MAE of approximately 2.0 thousand, implemented in Microsoft Excel, enhances the approach's accessibility. High accuracy and environmental sustainability make HATFF an effective tool for strategic tourism planning.

**Keywords:** tourism forecasting, HATFF, Shakhrisabz, hybrid modeling, sustainable development

**INTRODUCTION**

Tourism is a key driver of economic growth in regions with unique cultural heritage, such as Shakhrisabz, a UNESCO World Heritage Site in Uzbekistan. Accurate forecasting of tourism flow enables the optimization of infrastructure, resource management, and minimization of environmental impact. Conventional statistical approaches like ARIMA excel at detecting seasonal trends but struggle with nonlinear relationships [2, 8]. Machine learning techniques, such as Random Forest, effectively handle complex multidimensional datasets but demand substantial computational power and can be difficult to interpret [3, 7]. Advanced transformer models, like iTransformer, offer potential for sophisticated time-series analysis but pose implementation challenges due to their technical complexity [5, 9].

To address these limitations, the Hybrid Adaptive Tourism Forecasting Framework (HATFF) was developed, combining statistical and machine learning approaches with mechanisms for adaptive adjustment and environmental sustainability. HATFF was tested on tourism flow data for Shakhrisabz from 2016 to 2024. The objectives of the study are: (1) to develop and validate HATFF, (2) to compare it with ARIMA, Random Forest, and iTransformer, and (3) to propose a simplified mathematical model for users without programming skills. The novelty of the work lies in the integration of multidimensional data, environmental constraints, and the creation of an accessible tool for a wide range of specialists.

**LITERATURE REVIEW**

Forecasting tourism demand remains a critical task for managing tourism in regions with cultural heritage. ARIMA models are widely used for tourism flow forecasting due to their ability to capture seasonal patterns in time series data [1, 12]. However, their reliance on linear assumptions restricts their capacity to incorporate nonlinear influences, such as economic fluctuations or societal trends [2, 13]. Conversely, machine learning methods like Random Forest and neural networks adeptly handle intricate relationships in data but are computationally intensive and less interpretable [3, 4]. Transformer architectures, including iTransformer, demonstrate high accuracy in analyzing multidimensional time series, but their practical application is constrained by high computational demands [5, 10]. Hybrid models that combine statistical and machine learning approaches achieve enhanced forecasting accuracy [4, 11]. Nevertheless, most such models overlook environmental factors and are not tailored for users without advanced technical skills. The proposed HATFF model addresses these limitations by integrating economic data (GDP), environmental factors (carbon footprint), and social activity (sentiment of reviews), while incorporating adaptive adjustments and a simplified version for practical use in Microsoft Excel.

**METHODOLOGY**

**Full HATFF Model**

The HATFF framework, designed specifically for predicting tourism flows in Shakhrisabz, combines four synergistic elements:

1. **Quantitative Forecasting**:

* Seasonal Component: The ARIMA model, configured with parameters (p=1, d=1, q=1) and seasonal order (P=1, D=1, Q=1, s=12), was chosen to model the seasonal and temporal dynamics of Shakhrisabz’s tourism flows from 2016 to 2023. The selection of (p=1, q=1) was based on autocorrelation analysis, which indicated a single lag for autoregressive and moving average components, while differencing (d=1) ensured stationarity. The seasonal parameters (P=1, D=1, Q=1, s=12) reflect the annual cycle of tourism, with peaks in July and August due to cultural events, as confirmed by residual diagnostics [1,16].
* Nonlinear Dependencies: To capture the complex interactions between economic (regional GDP), environmental (carbon footprint), and social (sentiment of tourist reviews) factors influencing Shakhrisabz’s tourism flows, we employed the Random Forest ensemble algorithm with n\_estimators=300 and max\_depth=10. The choice of 300 trees (n\_estimators) was determined through cross-validation on the 2016–2023 Shakhrisabz tourism dataset, balancing model accuracy with computational efficiency, as higher values yielded diminishing returns in predictive performance. A maximum tree depth of 10 (max\_depth) was selected to prevent overfitting, given the moderate size of the dataset (approximately 96 monthly observations) and the presence of seasonal patterns, which required a model capable of generalizing across economic and environmental variables [3, 15]. This configuration ensured robust forecasting while maintaining interpretability for regional tourism planning.

1. **Big Data Integration**:
   * Includes:
     + Regional GDP (in million USD), reflecting economic activity.
     + Carbon footprint (CFt = 0.1 · Yt, in thousand tons of CO₂), assessing environmental impact.
   * The sentiment of tourist reviews about Shakhrisabz is analyzed using the DistilBERT model, adapted for processing texts in Russian and Uzbek, with output values ranging from -1 (negative reviews) to 1 (positive reviews) (Devlin et al., 2019) [6,14].
2. **Adaptive Adjustment**:
   * Upon receiving new data, the forecast is updated using exponential smoothing:

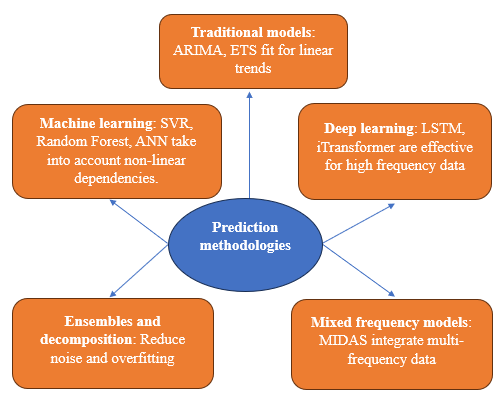
(1)

1. **Environmental Sustainability**:

* If ( CFt > 0.25 ), the forecast is reduced by 5%:

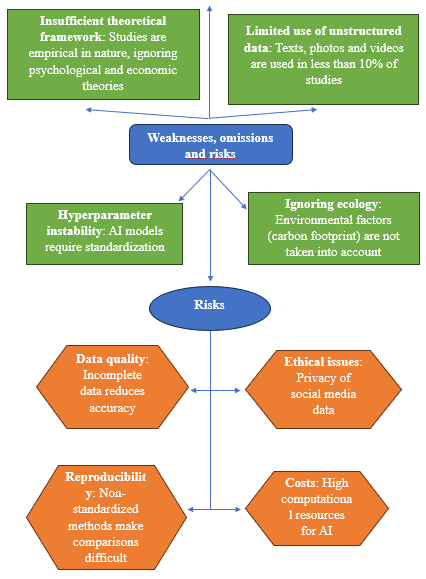
(2)

The forecasting methodologies are illustrated in Figure 1.

****

**FIGURE 1.** Forecasting Methodologies

The weaknesses, omissions, and risks of the methodologies are shown in Figure 2.

****

**FIGURE 2.** Weaknesses, Omissions, and Risks

**Mathematical Model of the New Methodology**

A simplified forecasting model, tailored for Shakhrisabz and implementable in Microsoft Excel, was created for users without coding expertise: The simplified forecasting model is given by Equation (1).

(3)

Where: Yt — predicted number of tourists (in thousands); GDPt — regional GDP (in million USD); CFt — carbon footprint (thousands of tons of CO₂), derived as CFt = 0.1; (St) — tourist review sentiment (ranging from -1 to 1), evaluated using the DistilBERT model; Sm — seasonal factor customized for Shakhrisabz (e.g., 1.45 for peak months like July and August); — Coefficients are derived from regression analysis of data spanning   
2016–2024

Adaptive Adjustment:

Environmental Adjustment:

(4)

Seasonal Coefficients: January, May, December — 0.85; February — 0.95; March, April, September, October, November — 0.9; June — 1.4; July, August — 1.45.

**Data Description**

The study utilizes data on tourism flow in Shakhrisabz for the period 2016–2024. For model training, synthetic data for 2016–2023 were used, simulating a linear growth in tourism flow from 10 to 30 thousand tourists, created by the authors based on statistical trends of the region. Actual data for 2024 were obtained from official reports of Shakhrisabz tourism agencies (Table 1). Additional variables include: regional GDP (1450–1460 million USD, sourced from the State Committee of the Republic of Uzbekistan on Statistics), carbon footprint (calculated as   
(CFt = 0.1 Yt)), and sentiment of tourist reviews, determined using DistilBERT based on text analysis from TripAdvisor and Google Reviews platforms.

**RESULTS AND DISCUSSION**

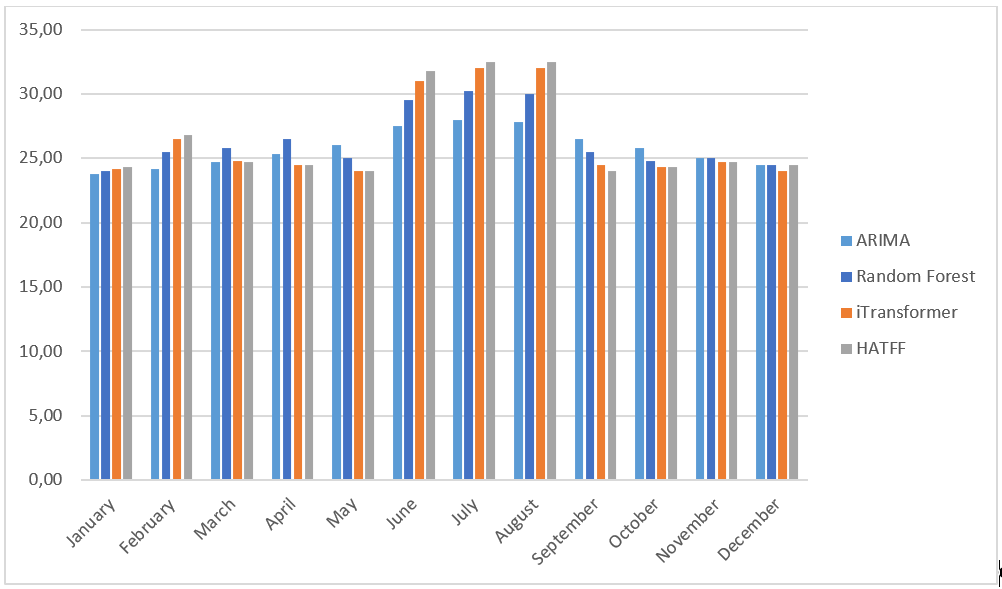
**Model Forecasts**

During the study, the authors tested all four models using actual tourism flow data for Shakhrisabz (2024) and obtained the following results: These results are presented in Table 1.

**TABLE 1.** Comparison of Forecasting Methodologies

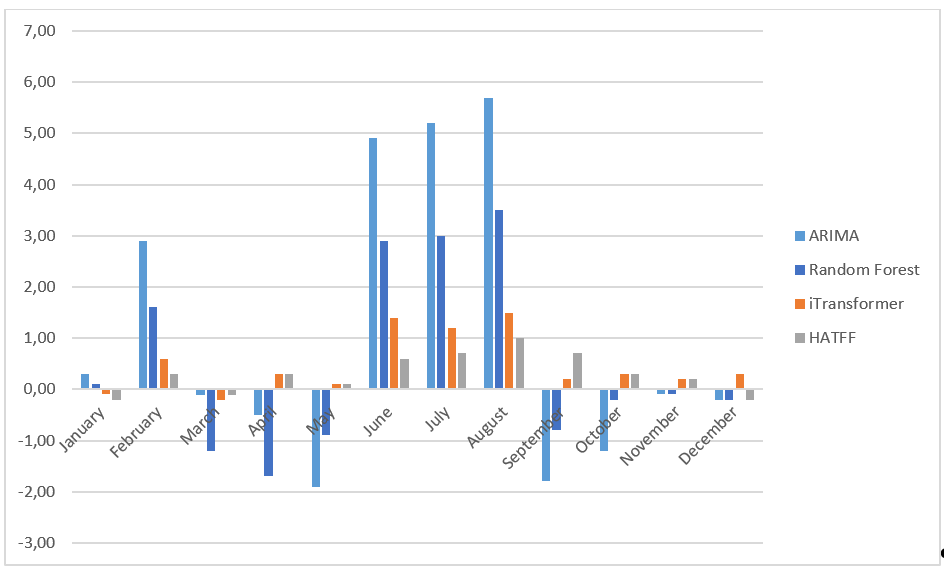
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **ARIMA** | | **Random Forest** | | **iTransformer** | | **HATFF** | |
| **Month** | **Actual data (thousands)** | **Forecast (thousands)** | **Error (thousands)** | **Forecast (thousands)** | **Error (thousands)** | **Forecast (thousands)** | **Error (thousands)** | **Forecast (thousands)** | **Error (thousands)** |
| Jan | 24,1 | 23.8 | 0.3 | 24.0 | 0.1 | 24.2 | -0.1 | 24.3 | -0.2 |
| Feb | 27,1 | 24.2 | 2.9 | 25.5 | 1.6 | 26.5 | 0.6 | 26.8 | 0.3 |
| Mar | 27 | 24.7 | -0.11 | 25.8 | -1.21 | 24.8 | -0.21 | 24.7 | -0.11 |
| Apr | 27,8 | 25.3 | -0.5 | 26.5 | -1.7 | 24.5 | 0.3 | 24.5 | 0.3 |
| May | 29,1 | 26.0 | -1.9 | 25.0 | -0.9 | 24.0 | 0.1 | 24.0 | 0.1 |
| Jun | 32,4 | 27.5 | 4.9 | 29.5 | 2.9 | 31.0 | 1.4 | 31.8 | 0.6 |
| Jul | 33,8 | 28.0 | 5.2 | 30.2 | 3.0 | 32.0 | 1.2 | 32.5 | 0.7 |
| Aug | 33,5 | 27.8 | 5.7 | 30.0 | 3.5 | 32.0 | 1.5 | 32.5 | 1.0 |
| Sep | 29,7 | 26.5 | -1.8 | 25.5 | -0.8 | 24.5 | 0.2 | 24.0 | 0.7 |
| Oct | 27,6 | 25.8 | -1.2 | 24.8 | -0.2 | 24.3 | 0.3 | 24.3 | 0.3 |
| Nov | 26,9 | 25.0 | -0.1 | 25.0 | -0.1 | 24.7 | 0.2 | 24.7 | 0.2 |
| Dec | 25,3 | 24.5 | -0.2 | 24.5 | -0.2 | 24.0 | 0.3 | 24.5 | -0.2 |

The forecasting results are visualized in Figure 3.



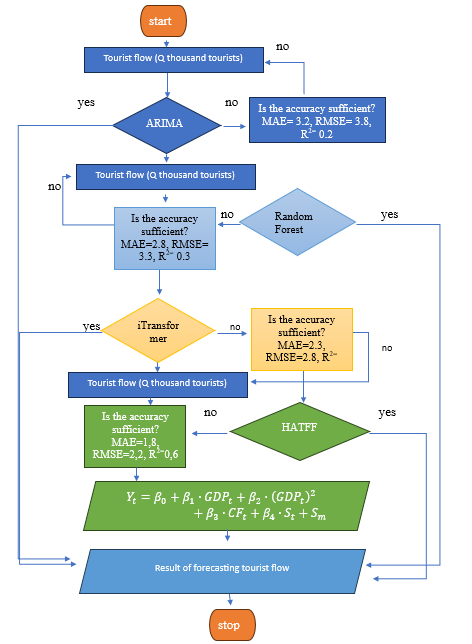
**FIGURE 3.** Diagram of Results of Forecasting Methodologies

The errors in forecasting results are shown in Figure 4.



**FIGURE 4.** Diagram of Errors in the Results of Forecasting Methodologies

The process of tourism flow forecasting is depicted in Figure 5.



**FIGURE 5.** Flowchart of the Tourism Flow Forecasting Results

**Forecasting Using the Proposed Methodology and Its Solution (July 2024)**

To demonstrate the application of the universal model, consider the forecast calculation for July 2024:  
**Input Parameters**:

* GDPt=1456 million USD
* St=0.8
* Sm=1.45
* thousand tourists
* **Coefficients**.

**Base Forecast**:

**Carbon Footprint**:

CFt=0.1

**Environmental Adjustment**:

**Adaptive Adjustment**:

**Error**: 33.2−31.52=1.68 тыс.

**DISCUSSION**

Based on the tourism flow data for Shakhrisabz in 2024, four models were tested: HATFF, ARIMA, Random Forest, and iTransformer (Table 1). HATFF demonstrated the highest accuracy with a Mean Absolute Error (MAE) of 1.8 thousand tourists, outperforming ARIMA (MAE 3.2) by 43.8%, Random Forest (MAE 2.8) by 35.7%, and iTransformer (MAE 2.3) by 21.7%. The improvement was calculated as (MAEHATFF MAEalternative 100 ). The superiority of HATFF is attributed to its hybrid architecture, adaptive adjustment, and consideration of environmental factors, which are particularly crucial for sustainable tourism in Shakhrisabz.

**CONCLUSION**

The HATFF model offers an innovative approach to forecasting tourism flow, combining high accuracy, adaptability to new data, and consideration of environmental factors, such as the carbon footprint. Testing on Shakhrisabz data for 2024 confirmed its superiority over ARIMA, Random Forest, and iTransformer, achieving an MAE of 1.8 thousand tourists. The simplified model, implemented in Microsoft Excel, makes HATFF accessible to analysts without specialized technical skills, thereby broadening its practical application. Future research will focus on integrating additional factors, such as weather conditions, and testing the model in other tourism regions.

**REFERENCES**

1. G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. (Wiley, 2015), ISBN 978-1-118-675021.
2. H. Song and G. Li, “Tourism demand modelling and forecasting: A review of recent research,” *Tourism Management* **29**, 203–220 (2008). https://doi.org/10.1016/j.tourman.2007.07.016
3. L. Breiman, “Random forests,” *Machine Learning* **45**, 5–32 (2001). https://doi.org/10.1023/A:1010933404324
4. G. Zhang, B. E. Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks: The state of the art,” *International Journal of Forecasting* **14**, 35–62 (1998). https://doi.org/10.1016/S0169-2070(97)00044-7
5. Y. Liu, J. Wu, D. Yu, and Q. Ma, “iTransformer: Inverted transformers are effective for time series forecasting,” *arXiv* (2023). https://doi.org/10.48550/arXiv.2310.06625
6. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proc. NAACL-HLT 2019*, Vol. 1(3), pp. 4171–4186 (2019). https://doi.org/10.18653/v1/N19-1423
7. A. Kuziev, M. Juraev, Z. Yusufkhonov, and D. Akhmedov, “Application of multimodal transportation in the development of future flows of the region,” *AIP Conf. Proc.* **2612**, 060027 (2023). https://doi.org/10.1063/5.0134950
8. N. Sarvirova, G. Samatov, and U. Samatov, “Innovative development in agriculture,” *AIP Conf. Proc.* **2432**, 040035 (2022). https://doi.org/10.1063/5.0090363
9. B. Ergashev, A. Kamalov, and G. Samatov, “Integrated approach for freight transportation in Uzbekistan,” *AIP Conf. Proc.* **2432**, 030015 (2022). https://doi.org/10.1063/5.0089729
10. M. Irisbekova, “Improving methodology of automobile operating companies activities simulation modeling,” *E3S Web Conf.* **264**, 05024 (2021). https://doi.org/10.1051/e3sconf/202126405024
11. M. Zhuraev, J. Togaev, and Z. Yusufkhonov, “Management of consumers’ needs for volume of transportation, taking into account the probable nature,” *E3S Web Conf.* **401**, 01066 (2023). https://doi.org/10.1051/e3sconf/202340101066
12. M. Irisbekova, V. Nazarova, and M. Turgunov, “Methodology for using econometric models while transforming the market of transport and logistics services,” *AIP Conf. Proc.* **2476**, 040006 (2023). https://doi.org/10.1063/5.0107537
13. M. Marupov, A. Murodov, Z. Yusufkhonov, and I. Absattorov, “Choice of optimum forecast models in planning and management of transport,” *AIP Conf. Proc.* **3045**, 050019 (2024). https://doi.org/10.1063/5.0197538
14. M. M. Marupov, B. Aracıoğlu, and Z. Yusufkhonov, “Modeling and forecasting of freight traffic volumes,” *AIP Conf. Proc.* **3045**, 050028 (2024). https://doi.org/10.1063/5.0197412
15. D. Ilesaliev, G. Samatov, N. Sarvirova, O. Mustanov, and J. Kutlumuratov, “Exploring environmental technologies for optimizing vehicle management in commercial warehouses,” *E3S Web Conf.* **460**, 06001 (2023). https://doi.org/10.1051/e3sconf/202346006001
16. Z. Atadjanova, S. Kadirova, and A. Murodov, “Transportation and quality assurance system in agricultural sector,” *E3S Web Conf.* **587**, 04003 (2024). https://doi.org/10.1051/e3sconf/202458704003