Please cite as:

Li, X., Xu,Y., Law, R., & Wang, S. (2024). Enhancing Tourism Demand Forecasting with a Transformer-based Framework, Annals of Tourism Research. Accepted.

Enhancing Tourism Demand Forecasting with a Transformer-based Framework

Xin Li[1], Yechi Xu[1], Rob Law[2] , Shouyang Wang[3, 4]

1 School of Economics and Management, University of Science and Technology Beijing, Beijing, 100083, China. drxinli@ustb.edu.cn; 593084215@qq.com

2 Asia-Pacific Academy of Economics and Management, Department of Integrated Resort and Tourism Management, Faculty of Business Administration, University of Macau, Taipa, Macau. roblaw@um.edu.mo

3 Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China; 4 School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China. sywang@amss.ac.cn

Abstract:

This study introduces an innovative framework that harnesses the most recent transformer architecture to enhance tourism demand forecasting. The proposed transformer-based model integrates the tree-structured parzen estimator for hyperparameter optimization, a robust time series decomposition approach, and a temporal fusion transformer for multivariate time series prediction. Our novel approach initially employs the decomposition method to decompose the data series to effectively mitigate the influence of outliers. The temporal fusion transformer is subsequently utilized for forecasting, and its hyperparameters are meticulously fine-tuned by a Bayesian-based algorithm, culminating in a more efficient and precise model for tourism demand forecasting. Our model surpasses existing state-of-the-art methodologies in terms of forecasting accuracy and robustness.

Keywords:

Tourism demand forecasting; time series decomposition; temporal fusion transformer; tree-structured parzen estimator

1. Introduction

Tourism has a pivotal influence on the global economy, generating employment opportunities, invigorating local economies, and promoting cultural exchange (Li, Song, & Witt, 2005; Song et al., 2019). Accurate tourism demand forecasting is essential for strategic planning, resource allocation, and decision-making processes within the tourism sector (Li, Song, & Witt, 2006). As the tourism landscape continually evolves, driven by factors such as technological advancements, shifting consumer preferences, and global socioeconomic dynamics, the need for precise and dependable tourism demand forecasting has become increasingly vital (Song, Qiu, & Park, 2023). In particular, the development of robust forecasting models enables businesses and policymakers to better anticipate fluctuations in demand, adjust their strategies accordingly, and capitalize on emerging market opportunities (Dwyer, Forsyth, & Spurr, 2004; Wu, Song, & Shen, 2017). Accurate tourism demand forecasting is not only crucial for the growth and sustainability of the tourism industry but also serves as a key contributor to the economic prosperity and well-being of communities worldwide (Chen et al., 2019; Li et al., 2019; Xu, Liu, & Jin, 2023).

Time series analysis and econometric models have been extensively employed in the realm of tourism demand forecasting, with the aim of discerning the intrinsic temporal patterns and structures present within the data. Conventional time series and econometric methodologies, such as autoregressive integrated moving average, exponential smoothing state space models, vector autoregressive models, and other classical econometric models, have been predominantly utilized for the modeling and prediction of tourism demand (Li et al., 2021; Llewellyn, Ross, & Ryan-Saha, 2023; Song & Li, 2008). However, these approaches often grapple with the challenges of addressing the intricate nonlinear patterns, seasonality, and existence of outliers in tourism demand data (Law et al., 2019). Moreover, time series analysis techniques frequently rely on assumptions regarding the underlying data structure, which may not consistently be valid in practical applications, consequently leading to potential inaccuracies in model forecasts (Yu et al., 2022).

Recently, cutting-edge artificial intelligence techniques have emerged as promising alternatives to traditional time series analysis methods, demonstrating enhanced performance in diverse forecasting tasks (Bi, Li, & Fan, 2021; Li, Zheng, & Ge, 2022; Torres et al., 2021). These techniques can autonomously learn intricate temporal patterns and structures in the data, which means they require minimal human intervention and prior knowledge of the data properties (Lara-Benítez, Carranza-García, & Riquelme, 2021). Among these techniques, recurrent neural networks, long short-term memory networks, and convolutional neural networks have been successfully applied to model and forecast tourism demand (Fotiadis, Polyzos, & Huan, 2021; Lara-Benítez et al., 2021). Yet, despite their advantages, artificial intelligence-based models exhibit certain limitations when applied to tourism demand forecasting. One primary concern is the propensity for these models to overfit, particularly when handling high-dimensional and noisy data, resulting in diminished generalization performance and reduced forecasting accuracy for new and unseen data.

Specific deep learning models, such as recurrent neural networks, may encounter difficulties in capturing long-term dependencies in time series data due to the vanishing gradient issue (Pascanu, Mikolov, & Bengio, 2013). An additional constraint is the interpretability of these models. Although machine learning and deep learning techniques can identify intricate patterns within the data, the generated models are often perceived as black boxes, thereby complicating the comprehension of underlying relationships between input features and predicted outcomes (Kucklick & Müller, 2023). This lack of transparency may impede the adoption of these models in the tourism industry, where decision-makers might require more interpretable models to guide their strategic planning and resource allocation. Consequently, while deep learning techniques have demonstrated substantial potential in tourism demand forecasting, these limitations highlight the need for continued research and development of innovative models that can overcome these challenges, thus ensuring more accurate, interpretable, and efficient forecasting solutions within the tourism sector.

The temporal fusion transformer epitomizes a state-of-the-art deep learning model for multivariate time series prediction, showcasing exceptional potential in capturing complex temporal dependencies and seasonal patterns (Lim et al., 2021). The transformer model combines the benefits of the transformer architecture with supplementary components, including adaptive gating mechanisms and skip connections, to adeptly process various types of time series data. Initially designed for natural language processing tasks, the Transformer Architecture has gained traction in time series forecasting due to its ability to model long-range dependencies and effectively manage multiple input features (Vaswani et al., 2017). The temporal fusion transformer model builds upon this architecture, incorporating specialized components tailored for time series data, such as temporal attention mechanisms and variable selection layers. These features allow the transformer model to focus on the most relevant input features and time steps, culminating in its superior forecasting performance (Wu, Wang, & Zeng, 2022).

Nonetheless, the temporal fusion transformer model also displays certain limitations. For example, a significant challenge in employing the temporal fusion transformer model is the optimal selection of hyperparameters, which can substantially impact the model’s performance. The determination of hyperparameters, such as learning rate, batch size, and the number of layers, necessitates an exhaustive search and fine-tuning, which can be computationally costly and time-consuming. In some instances, an inadequate hyperparameter selection procedure may result in suboptimal model performance or overfitting.

To address the existing limitations of the temporal fusion transformer model and enhance its predictive capacity in tourism demand forecasting, we introduce a novel transformer-based model. This model incorporates tree-structured parzen estimator for hyperparameter optimization, a robust method for time series decomposition, and temporal fusion transformer for multivariate time series prediction. Initially, our approach employs a sophisticated time series decomposition algorithm to disassemble the tourism demand time series into its trend, seasonal, and residual components, efficiently mitigating the impact of outliers (Wen et al., 2019). Subsequently, the temporal fusion transformer model is harnessed to forecast these decomposed components. The model's hyperparameters are refined through a Bayesian optimization method, offering a more efficient solution for high-dimensional hyperparameter spaces than traditional random or grid search methods (Rong et al., 2021). With the integration of advanced transformer architecture, sophisticated decomposition methods, and efficient tuning procedures, this research provides a robust and precise solution for tourism demand forecasting challenges.

The efficacy of the proposed model is demonstrated on three datasets, and the results indicate that it can surpass existing state-of-the-art methods in terms of forecasting accuracy and robustness. Our findings suggest that the proposed model presents a promising approach for improving tourism demand forecasting, thus delivering valuable insights for decision-makers in the tourism industry.

2. Literature Review

2.1 An overview of tourism demand forecasting

Tourism demand forecasting is crucial in the tourism industry as it significantly impacts stakeholders, such as policymakers, destination managers, and tourism service providers (Sun et al., 2023; Wu, Li, & Song, 2022). Traditional time series methods such as autoregressive integrated moving average and exponential smoothing state space models, which primarily rely on historical patterns, have been extensively used for forecasting tourism demand (Li, Song, & Witt, 2005; Mueller & Sobreira, 2024; Park et al., 2021). Econometric models, including vector autoregression (Gunter & Zekan, 2021), mixed data sampling approach (Hu et al., 2022; Liu et al., 2021; Wu et al., 2023), spatial temporal model (Jiao, Chen & Li, 2021; Yang & Zhang, 2019), and dynamic factor models (Li et al., 2017) incorporate various explanatory variables to account for macroeconomic factors, demographic variables, and other determinants (Li et al., 2020; Li, Hu, & Li, 2020). While these methods offer valuable insights, they may struggle to capture complex relationships, non-linear patterns, and high-dimensional data, which are prevalent in tourism demand data (Gunter & Zekan, 2021; Law et al., 2019).

In response to the limitations of traditional methods, artificial intelligence-based models that encompass machine learning and deep learning techniques have emerged as promising alternatives for tourism demand forecasting (Essien & Chukwukelu, 2022; Kulshrestha et al., 2020; Li, Gao, & Song, 2023; Xie, Qian, & Wang, 2021). These approaches can automatically learn complex temporal patterns and structures in the data and require minimal human intervention and prior knowledge of the data properties (Sun et al., 2022). Various machine learning algorithms, such as support vector machines (Bi, Li, & Fan, 2021), decision trees, and random forests, have been applied to tourism demand forecasting with varying levels of success. More recently, deep learning models such as long short-term memory networks have demonstrated improved performance in capturing nonlinear patterns and high-dimensional data (Bi et al., 2022).

Despite their potential, artificial intelligence-based models face several challenges, including overfitting, interpretability, and computational complexity (Zhang et al., 2021). Overfitting occurs when a model learns noise within the data, reducing its generalizability to unseen data (Ying, 2019). The issue of interpretability arises due to the black-box nature of artificial intelligence models, making it challenging to understand and explain their decision-making processes (Ali et al., 2023). Computational complexity is another concern, as artificial intelligence-based models often require considerable computational resources and time for training and optimization (Zhang et al., 2021).

2.2 Interpretability in artificial intelligence models

In recent years, tourism demand forecasting has witnessed significant progress with the adoption of artificial intelligence-based models, which have demonstrated remarkable performance in capturing complex patterns and high-dimensional data (Bi, Li, & Fan, 2021). However, these models face challenges related to overfitting, interpretability, and computational complexity. Interpretability in artificial intelligence models has gained considerable attention in recent years, particularly due to the prominence of deep learning techniques that exhibit exceptional performance yet lack transparent decision-making processes. In tourism demand forecasting, interpretability has become an important consideration for stakeholders to develop trust in a model’s predictions, which in turn, can lead to more informed decisions and effective resource allocation (Petropoulos et al., 2022).

In response to the challenges posed by black-box models, researchers have explored various strategies to enhance interpretability in AI models. These approaches include designing inherently interpretable models, developing post-hoc explanation techniques, and incorporating domain knowledge to guide model construction (Zhang et al., 2020; Zheng et al., 2021). Notable studies include Law et al. (2019), which introduced an attention mechanism in an long short-term memory neural network architecture to identify influential factors affecting tourism demand, and Monje et al. (2022), which employed a deep learning model with fuzzy logic for bus passenger forecasting and used local interpretable model-agnostic explanations analysis for model interpretability. Integrating such interpretability techniques can offer valuable insights into deep learning models and promote a comprehensive understanding of such models in tourism demand forecasting and other applications, thus assisting practitioners in making well-informed decisions.

2.3 Decomposition approach for tourism forecasting

Within the realm of tourism demand forecasting, a diverse array of decomposition methods has been applied to tackle the challenges arising from intricate and non-stationary time series data (Liu et al., 2023). Classical decomposition techniques, such as seasonal decomposition of time series, have been extensively utilized to partition time series data into trend, seasonal, and residual components. Apart from these methods, more sophisticated decomposition approaches, including empirical mode decomposition and variational mode decomposition (Li & Law, 2020), have been proposed to enhance the accuracy and interpretability of forecasting models (Zhang et al., 2022).

Recently, innovative decomposition methods have been proposed to address specific challenges within the tourism demand forecasting domain. For example, Grossi and Mussini (2021) introduced a decomposition technique to measure changes in the timing and magnitude of seasonality in tourist flows by decomposing the change in the Gini index, while Xie, Qian, and Wang (2020) proposed a decomposition-ensemble approach based on complete ensemble empirical mode decomposition with adaptive noise and data characteristic analysis for tourism demand forecasting. These novel methods represent advancements in the field, thus allowing for more precise and informative forecasting.

These decomposition strategies have also demonstrated their effectiveness in improving the performance of conventional forecasting models by facilitating a more precise identification of inherent patterns and dependencies within each decomposed component. Nevertheless, there remains potential for further advancement when employing sophisticated artificial intelligence models, such as the Transformer architecture.

2.4 Forecasting with a transformer framework

The Transformer architecture, initially introduced for NLP and image recognition tasks, has exhibited remarkable success in these domains (Vaswani et al., 2017). Numerous derivatives of the original Transformer model have been devised, such as BERT (Devlin et al., 2018), and Informer (Zhou et al., 2021), among others. Nevertheless, the application of these models to time series forecasting has not been straightforward due to the distinctive characteristics of time series data, necessitating the integration of sequential processing and temporal context handling mechanisms.

In response to these challenges, several augmented transformer models have been proposed for time series analysis, including the temporal fusion transformer. For example, Lim et al. (2021) employed a temporal fusion transformer model to conduct multi-horizon forecasting and demonstrated the forecasting capabilities of the transformer approach in time series forecasting. The temporal fusion transformer leverages a self-attention mechanism to discern long-term relationships across various temporal steps. This design is an adaptation of the multi-head attention found in transformer-based architectures (Vaswani et al., 2017), and is specifically tailored to enhance the model’s explainability.

Despite their potential, the application of transformer architectures in tourism demand forecasting remains limited, which may be due to the specific challenges associated with tourism data, such as seasonality, non-stationarity, and the influence of various external factors.

2.5 Rationale of the study

Tourism demand forecasting has witnessed substantial advancements, particularly with artificial intelligence-based methods. Existing deep learning models have made significant strides, leveraging sophisticated networks to analyze temporal data and generate more accurate predictions (Bi et al., 2020; Law et al., 2019). However, improvements are still needed as these approaches have shown limitations in handling complex temporal relationships, multi-step forecasting, and interpretability. Based on a cutting-edge transformer architecture, the innovations of this study can be mainly articulated in the following aspects:

The first aspect concerns differences in deep learning architecture design. In contrast to the deep learning models like those utilized by Law et al. (2019), our study employs a more advanced transformer architecture. The temporal fusion transformer inherently integrates a sophisticated attention mechanism, allowing for variable selection, temporal dependency learning, and enhanced interpretability. This enables more efficient handling of the complexities inherent in tourism demand forecasting and offers significant advantages in terms of multi-step forecasting and interpretability.

Second, from the perspective of modeling strategy and hyperparameter optimization, our approach involves a robust time series decomposition method, effectively mitigating the influence of outliers. Additionally, we integrate the tree-structured parzen estimator for precise hyperparameter optimization. This unique combination of decomposition and Bayesian-based optimization elevates our model's adaptability, flexibility, and precision over the methods described in the existing literature.

Third, with regard to forecasting performance, we conduct a comprehensive evaluation of our proposed framework against existing methodologies. Importantly, our model demonstrates improved forecasting accuracy compared to several established approaches. Through this careful evaluation, our work enhances the reliability and effectiveness of tourism demand forecasting.

In summary, by employing advanced transformer architecture, robust decomposition methods, and sophisticated optimization techniques, this study offers a more nuanced and powerful solution to the underlying challenges in the field of tourism demand forecasting. These innovations underscore the essential contributions of our work, promising a substantial impact on both theory and practice in the domain of tourism management and forecasting.

3. Methodology

3.1 Standard temporal fusion transformer

The temporal fusion transformer is proposed based on the Transformer architecture and can capture temporal dependencies and cyclical patterns (Lim et al., 2021). This model proficiently manipulates various forms of time series data by integrating the benefits of the Transformer architecture with various components, particularly the gating mechanisms, and multi-head attention. Figure 1 presents the architecture of the proposed temporal fusion transformer model.

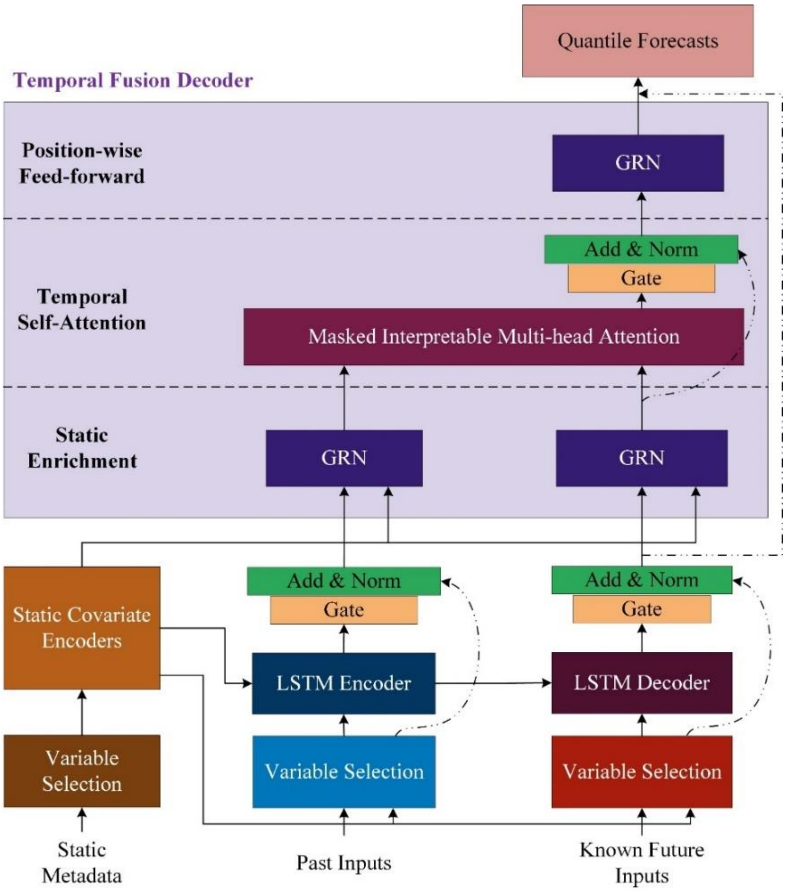


Figure 1. Overall architecture of the transformer model

In a standard temporal fusion transformer model, the gated residual networks and gated linear units serve as crucial components for effectively managing intricate associations between exogenous inputs and target variables. They can be likened to smart filters that efficiently sort and manage complex relationships between different pieces of information, which include explanatory variables and explained targets. The two components are designed to make the process of understanding and handling these complex relationships simpler and more effective.

In particular, gated residual networks enable adaptable nonlinear processing by taking a primary input and an optional context vector and then transforming them into a new vector. They use the exponential linear unit activation function, a kind of gatekeeper that lets positive values pass through but turns negative values closer to zero using an exponential calculation. Along with standard layer normalization, a process that standardizes all data to be on a comparable scale, these mechanisms ensure reliable performance across different types of input.

$$
\begin{aligned}
&GRN_{\omega}(a,c) = LayerNorm(a = GLU_{\omega}(\eta_1)) \\
&\qquad \eta_1 = W_{1,\omega}\eta_2 + b_{1,\omega} \\
&\qquad \eta_2 = ELU(W_{2,\omega}a + W_{3,\omega}c + b_{2,\omega})
\end{aligned}
\tag{1}
$$

where intermediate layers denoted by η_1 and η_2, along with weight-sharing index ω, use the exponential linear unit activation function in gated residual networks. This allows for linear or identity behavior depending on input values, while standard layer normalization maintains a consistent performance across diverse input scales.

The gated linear units work with gated residual networks to adjust the model's architecture for specific datasets by using a sigmoid activation function, which is a mathematical tool used in neural networks that helps to turn potentially complicated inputs into something simpler and more understandable. The gated linear units control input vector transformations and allow the model to manage the gated residual networks' contribution to the original input, even bypassing the layer if required.

$$GLU_{\omega}(\gamma) = \sigma(W_{4,\omega}\gamma + b_{4,\omega}) \odot (W_{5,\omega}\gamma + b_{5,\omega}) \tag{2}$$

where $\sigma(.)$ represents the sigmoid activation function, and $\odot$ is the element-wise Hadamard product, which is a specific way of multiplying two matrices together. This allows the transformer model to control the extent to which the gated residual networks contribute to the original input a, potentially skipping over the layer entirely if necessary.

Besides the gating mechanisms, the variable selection networks are designed to focus on important factors in the model. This architecture can reduce the risk of overfitting by filtering out noisy data. Variable selection weights are generated using the gated residual networks and Softmax layer, with additional nonlinear processing at each time step.

The attention mechanism allows the model to selectively focus on specific parts of the input data, much like how humans pay attention to particular details in a scene, thereby emphasizing the importance of certain information over others. Our proposed model uses multi-head attention to learn long-term relationships across time steps while improving explainability.

$$\begin{aligned} &MultiHead(Q,K,V) = [H_1,\ldots,H_{m_H}]W_H \\ &H_h = Attention(QW_Q^{(h)}, KW_K^{(h)}, VW_V^{(h)}) \end{aligned} \tag{3}$$

where $W_Q^{(h)}$, $W_K^{(h)}$, $W_V^{(h)}$ represent weights for queries, keys, and values. W_H linearly combines outputs from all heads H_h.

Attention weights do not directly indicate a feature's importance because of the different values in each head. The temporal fusion transformer mdoel addresses this by modifying multi-head attention to share values across heads and using additive aggregation of all heads. To provide an example, consider multi-head attention as a process where each head focuses on different parts of the data. These individual insights are then synthesized to provide a richer, more comprehensive comprehension of the entire data structure. This modification ensures a more coherent and aligned understanding, reflecting the cumulative importance of different features.

$$InterpretableMultiHead(Q,K,V) = \tilde{H} W_H \tag{4}$$

$$\begin{aligned} \tilde{H} &= \tilde{A}(Q,K)VW_V \\ &= \left\{ \frac{1}{m_H} \sum_{h=1}^{m_H} A(QW_Q^{(h)}, KW_K^{(h)}) \right\} VW_V \\ &= \frac{1}{m_H} \sum_{h=1}^{m_H} Attention(QW_Q^{(h)}, KW_K^{(h)}, VW_V) \end{aligned} \tag{5}$$

where W_V represents value weights shared across all heads, W_H is used for final linear mapping.

In summary, the architecture within the temporal fusion transformer model offers precise a focus on relevant temporal relationships, enabling more accurate and trustworthy forecasting, a crucial advancement for applications such as tourism demand forecasting.

3.2 The tree-structured parzen estimator

The tree-structured parzen estimator algorithm, a sequential model-based optimization method, offers an effective alternative for hyperparameter tuning (Bergstra, Yamins, & Cox, 2013). It combines the strengths of both Bayesian optimization and parzen window estimation and facilitates a more efficient exploration of the search space compared with traditional techniques. A description of the algorithm can be found in Appendix A.1. The key idea behind the algorithm is to identify the hyperparameters that maximize the expected improvement of the objective function, which can be formulated as follows:

$$EI(x) = \frac{l(x)}{g(x)} \tag{6}$$

where x represents the hyperparameters, and y denotes the corresponding performance metric for the observation history (x, y) and two models $l(x)$ and $g(x)$.

These models were constructed using Parzen window estimation, a nonparametric density estimation method that assigns weights to neighboring observations based on a kernel function. The algorithm updates the observation history and models at each iteration, separating the search space into “good” and “bad” regions represented by D_l and D_g, respectively based on the median value of the performance metric.

At each iteration, the algorithm samples new hyperparameters $x^{(i+1)}$ from the distribution $l(x)$, evaluates the objective function $f(x^{(i+1)})$, updates the observation history, and models accordingly.

3.3 Robust seasonal and trend decomposition using Loess

The seasonal and trend decomposition using Loess provides a flexible and interpretable means of decomposing a time series into seasonal, trend, and residual components (Cleveland et al., 1990). In our case, the ability to explicitly model and remove both seasonal and trend patterns provided a more intuitive understanding of the underlying data dynamics. The robust seasonal and trend decomposition using Loess is designed to be more resilient to outliers and abrupt changes, as compared to traditional decomposition method (Wen et al., 2019). While other decomposition method such as singular spectrum analysis is powerful, it could introduce complexities that were not essential for our analysis. The interpretability of the algorithm is more aligned with our research objectives. This increased robustness is achieved through the use of different weighting functions and iterations designed to minimize the influence of outliers on the estimated components. A description of the algorithm can be found in Appendix A.2.

Robust seasonal and trend decomposition using Loess method consists of three main steps: trend extraction, seasonal component estimation, and remainder component estimation. In the first step, it extracts the trend component τ_t from the original time series y_t using a penalized least squares regression with a smoothness constraint. The trend extraction can be formulated as follows:

$$\min_{\tau} \sum_{t=1}^{T} \rho(y_t - \tau_t) + \lambda \sum_{t=1}^{T-2} (\tau_{t+1} - 2\tau_t + \tau_{t-1})^2 \tag{7}$$

where $\rho(\cdot)$ is a robust loss function that reduces the influence of outliers, and λ is a regularization parameter controlling the smoothness of the trend component.

After extracting the trend, the seasonal component s_t is estimated by fitting a periodic spline regression model to the detrended time series, and the remainder component r_t is computed as the residual of the original time series after subtracting the estimated trend and seasonal components:

$$\begin{aligned} &\min_{s} \sum_{t=1}^{T} \rho(y_t - \tau_t - s_t) + \gamma \sum_{t=1}^{T-P} (s_{t+P} - s_t)^2 \\ &r_t = y_t - \tau_t - s_t \end{aligned} \tag{8}$$

where P is the seasonal period, and γ is a regularization parameter that controls the smoothness of the seasonal component.

In summary, the adopted decomposition method provides a robust and efficient framework for decomposing long time series data into the trend, seasonal, and remainder components, thereby enhancing the model’s ability to handle complex data patterns and improve the overall forecasting performance.

3.4 The proposed forecasting framework

This section delineates the methodology employed in the development and assessment of the proposed framework. The comprehensive procedure is partitioned into three principal stages: data acquisition, data processing, and model preparation, and forecasting and evaluation. The workflow proceeds as follows and is illustrated in Figure 2.

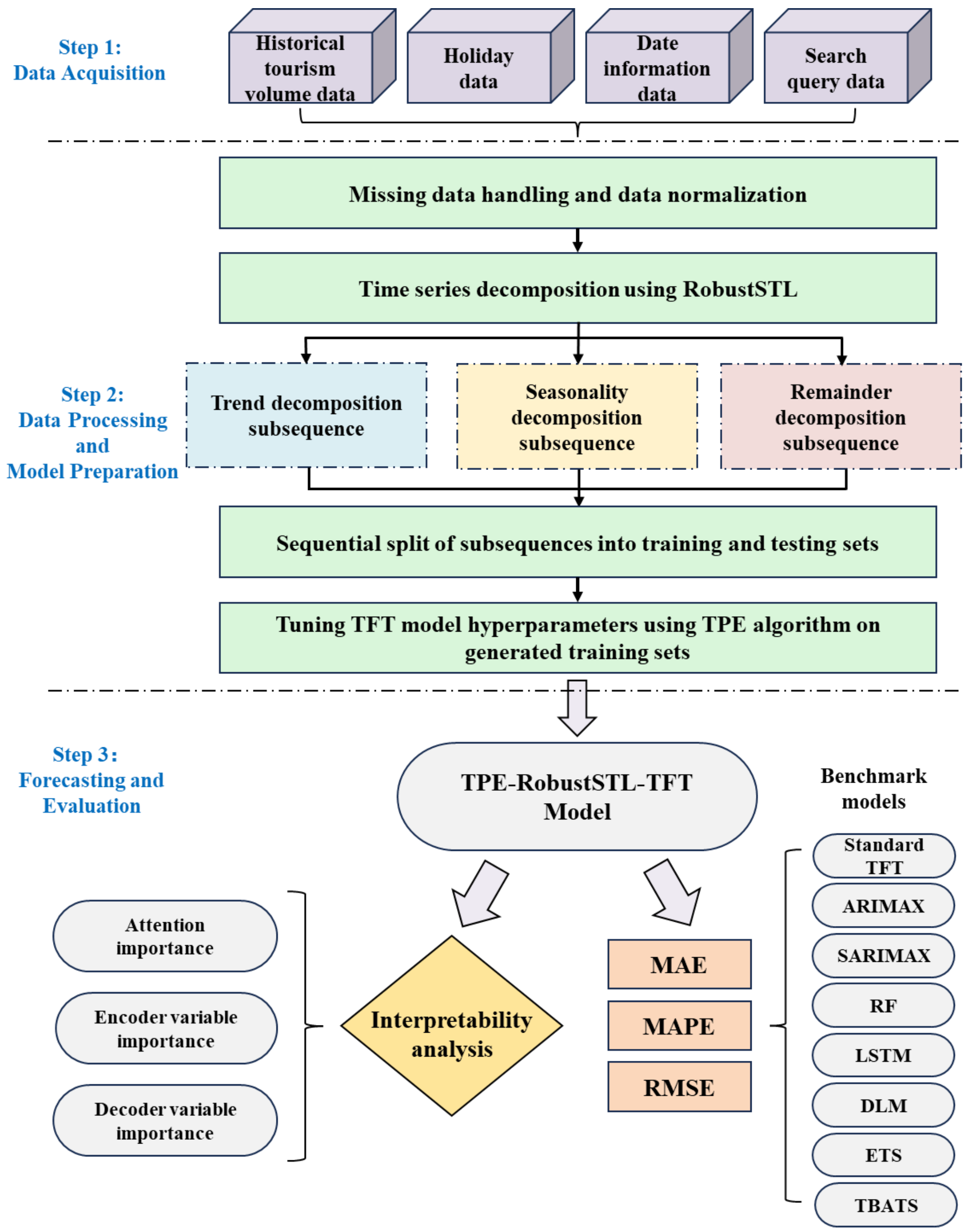


Figure 2. Overall Framework

Step 1: Data acquisition

The first step in the process involves gathering four types of data. Historical tourism volume data includes the number of tourists visiting the destination over a specific period.

Holiday data primarily focus on information regarding public holidays that have the potential to influence tourism demand. Date information includes additional temporal details, such as the day of the week, month, and year, which may be relevant for capturing seasonal patterns in the data. Baidu Index data reflect the search query volumes and tourists' interests for destinations (Law et al., 2019).

Step 2: Data processing and model preparation

This step first includes the handling of missing values and normalization of data to ensure consistent scales across varied variables. A K-nearest neighbors imputation method is utilized for missing data, replacing the absent values with the mean of the K-nearest non-missing values, factoring in the unique attributes of the time series. This method presents a more robust solution, particularly for non-linear data configurations. Following this, the data are normalized, aligning the scales across different variables to ensure uniformity.

Subsequently, the time series data are disassembled into three distinct components, namely the trend decomposition, seasonality decomposition, and remainder decomposition. Post decomposition, the components are partitioned into training and testing sets through a sequential methodology, thus maintaining the temporal sequence of the data. Additionally, the tree-structured parzen estimator employed to select the optimal parameters in the model training process, enhancing the overall efficacy of the predictive model.

Step 3: Forecasting and evaluation

This step involves forecasting and evaluation. To comprehensively evaluate the effectiveness of the proposed model, we performed a comparative analysis with several well-established benchmark models: 1) exponential smoothing state space model (ETS), 2) trigonometric seasonality, box-cox transformation, autoregressive moving average errors, trend and seasonal components (TBATS), 3) autoregressive integrated moving average with exogenous variables model (ARIMAX), 4) seasonal autoregressive integrated moving average with exogenous variables model (SARIMAX), 5) random forests (RF), 6) long short-term memory model (LSTM), 7) deep learning model (DLM) and 8) standard temporal fusion transformer model (TFT). Therefore, our model can be compared with classic time series models, machine learning models, deep learning models, and existing Transformer models.

To evaluate the performance of the proposed framework and the benchmark models, we employ three widely adopted performance evaluation criteria: mean absolute error, root mean squared error, and mean absolute percentage error.

$$MAE = \sum_{i=1}^{n}|\hat{y}_i - y_i| \Big/ n \ , RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \ , MAPE = \frac{1}{n}\frac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{y_i} \tag{9}$$

where $\hat{y}$, y_i are predicted and actual values for tourist arrivals. *n* denotes the number of observations.

4. Empirical Study

4.1 Data collection and processing

This study takes three daily datasets to verify the effectiveness of the proposed method: Hawaii, Jiuzhaigou, and Mount Siguniang, all of which are commonly used in existing research. Following the proposed methodology, the study initially handles missing values and standardizes the data, leading to the subsequent decomposition of the data into relevant components. In the Hawaii dataset, the data spanning July 1, 2009, to October 6, 2021, were allocated for training and validation, while the period from January 6, 2022, to April 5, 2022, constituted the test set. For the Mount Siguniang and Jiuzhaigou datasets, the data were divided into training and validation sets from April 1, 2020, to June 14, 2021, with a test set from June 15, 2021, to September 12, 2021. Daily tourist arrival data for Hawaii, Jiuzhaigou, and Mount Siguniang were acquired from their respective official websites, which record the total number of visitors to these tourist attractions. The data can be accessed using the following links: https://dbedt.hawaii.gov/visitor/daily-passenger-counts/ (for Hawaii), https://www.jiuzhai.com/news/number-of-tourists (for Jiuzhaigou), and https://www.sgns.cn/info/number (for Mount Siguniang).

Regarding the selection of keywords, we referred to the basic procedure in the existing literature, such as Li et al. (2017) and Bi et al. (2020). Search query data were incorporated into the forecasting of Jiuzhaigou and Mount Siguniang. To align with the travel preferences of tourists, several keywords such as "Jiuzhaigou," "Jiuzhaigou weather," and "epidemic" were selected as initial seeds and extended using the Baidu index. We found that the keywords "Jiuzhaigou," "Mount Siguniang," and "Sichuan epidemic" are highly correlated with tourist arrivals. Consequently, we selected six keywords for our analysis, which include "pc_Jiuzhaigou," "pc_Siguniang," "pc_Sichuan epidemic," "mob_Jiuzhaigou," "mob_Siguniang," and "mob_Sichuan epidemic," to encompass search queries originating from PC and mobile channels.

To provide a more comprehensive assessment of the model's ability to interpret exogenous variables, this study specifically collected Baidu Index data for the Jiuzhaigou and Mount Siguniang datasets, following the approach described in existing literature such as Zhang et al. (2022), to serve as a basis for supplementary analysis. Table 1 lists the descriptive analysis of three output variables, representing the historical tourist arrivals data for Hawaii, Jiuzhaigou, and Mount Siguniang, and some input variables, which consist of historical Baidu query data based on search keywords.

Table 1. Descriptive statistical analysis of the mainly used data

Variable	Mean	std	min	25%	50%	75%	max
Output variables (historical tourist arrivals data to the targeted destination)							

Hawaii_tourist	24070.04	7276.62	249	21520	24823	28481.5	41945
Jiuzhaigou_tourist	5864.06	5424.51	106	1503.5	4113.5	9092.5	30000
Siguniang_tourist	1261.76	1831.03	8	343.25	777	1521	17211
Some quantitative input variables (historical search queries for a given keyword)							
pc_Jiuzhaigou	371.70	139.43	133	257.5	357	447	930
mob_Jiuzhaigou	1045.15	325.06	550	893.25	947.5	1057.75	2919
pc_Siguniang	371.70	139.43	133	257.5	357	447	930
mob_Siguniang	1045.15	325.06	550	893.25	947.5	1057.75	2919
pc_SichuanEpidemic	263.26	414.70	64	108.5	150	218	6260
mob_SichuanEpidemic	1576.31	5215.39	256	505.25	712.5	911.25	65842

4.2 Model design

In this section, our proposed model and benchmarks are constructed, and parameters and design for each model are provided. Following the proposed methodology, the tree-structured parzen estimator algorithm is used to optimize our model and select the parameters. The process starts with a training dataset, where the encoder length is set to 30 and prediction lengths vary between 1, 3, 7, 15, and 30 steps. A corresponding validation set is then extracted to predict the final points for each series, tailored to the specific prediction length selected. For hyperparameter optimization, we employ the tree-structured parzen estimator method using the *optimize_hyperparameters* function from the PyTorch forecasting library. Appendix B provides the details of the hyperparameter optimization.

As illustrated in Section 3.4, we constructed eight benchmark models to evaluate the proposed method. Table 2 presents the inputs for the forecasting models. In our framework, the output is the tourist arrivals for the $t+1$ day; the inputs include the decomposed subseries of historical tourist arrivals data, date and holiday information, and search queries, spanning from the $t-m+1$ day to the t day. Here, m represents the maximum lag order specified by user, denoting the number of lagged days used in training the model. Essentially, we use data from the $t-m+1$ day (lag=m) and $t-m+2$ day (lag=$m-1$) to the t day (lag=1) to forecast the value on the $t+1$ day. Existing studies, such as Law et al., (2019), used a maximum order of 12 for monthly forecasts. Therefore, we set our maximum lag to 30 days ensuring optimal capture of long-term temporal patterns in a daily forecast. The determination of lag orders for other benchmark models are consistent with that of existing studies (e.g., Li et al., 2019; Law et al., 2019; Bi et al., 2021).

Table 2. A general description of inputs for the forecasting models

Category	Model inputs	Description
Historical tourist arrivals	Time series of tourist arrivals	Sequential data capturing the historical patterns of tourist arrivals.
Date and holiday	Day	Denotes the specific day within a month.

information (known inputs about the future)	Month	Numeric representation of the month (e.g., January=1).
	Year	Specifies the calendar year.
	Weekday	Indicates the specific day within a week (e.g., Monday=1).
	Time index	A sequential identifier denoting the position of the data entry within the dataset.
	Holiday	Holiday data is a dummy variable indicating holiday observance, derived from the website (https://www.officeholidays.com/countries/usa/). A value of 1 signifies a holiday, and 0 indicates otherwise.
Decomposed components	Decomposed series	Time series decomposition using the RobustSTL method, generating distinct sequences for trends, seasonal patterns, and residuals.
Historical search queries (optional)	Search engine data	Search query volume data extracted from the Baidu search engine, representing the daily search interest for specific keywords.

In our proposed model, the approach to setting lags for input variables aligns with that of recent studies that utilized deep learning models such as long short-term memory model (Bi et al., 2019), to facilitate multi-step forecasting using explanatory variables. A fundamental strength of deep learning, which is integral to our model, is its ability to identify and construct relevant features at different network layers automatically (Law et al., 2019). Our model effectively identifies and utilizes the most relevant variables and their respective lags for prediction through the application of variable selection networks. The lag selection for input variables in our model starts by defining a maximum lag, in which a comprehensive range of potential lags is set for consideration. The model's attention mechanism then dynamically evaluates these lags, focusing on the most relevant time points for the prediction task. Concurrently, the variable selection network assesses the significance of each lag within this range. This process aims to identify and focus on the most informative lags for accurate forecasting.

All models are estimated based on the following rules. The first two benchmark models are designed by minimizing the Bayesian Information Criterion. The parameters for third and fourth benchmark models are determined through a systematic process that involves testing various combinations of lag orders, including seasonal lags. The best combination is selected based on the minimization of information criteria such as the Akaike Information Criterion and the Schwarz Bayesian Information Criterion. The random forest model constructs multiple decision trees and aggregates their predictions to generate the final forecast. The important parameters of random forest are determined by the exhaustive grid search technique (Breiman, 2001). Similarly, the parameters in the last three benchmark models are obtained following the procedure in existing studies including Bi et al. (2021), Law et al. (2019), and Lim et al. (2021). Besides the first two time series models, we incorporate consistent explanatory variables into the benchmarks to ensure a fair competition.

The extended window forecasting strategy has been applied to the test sets of three datasets, to enhance the robustness and precision of our forecasting models. Taking the one-step-ahead forecast for the Jiuzhaigou dataset as an example, data from the original training set were utilized for our initial model estimation, after which a forecast for June 15, 2021 (the first data point in the test set) was generated. Subsequently, the training window was expanded to include new data points, enabling a second model estimation and a forecast for June 16, 2021. This process of extending the window continued until the forecast for September 12, 2021 (the ninetieth data point in the test set) was completed, thereby covering the entire test set. The forecasting performance of all models was thereafter evaluated.

4.3 Forecasting results

In this section, we systematically evaluate the performance of the proposed method in comparison with eight existing benchmark models across three datasets for both one-step and multi-step forecasting. Existing research, such as Bi et al. (2021) employed a three-step forecast for multi-step forecasting evaluation. This study extends the evaluation to include forecasts at 3, 7, 15, and 30 steps, ensuring a thorough assessment of our model's capabilities in multi-step forecasting. Tables 3–5 present the mean absolute error, root mean squared error, and mean absolute percentage error for the nine models in one-step and multi-step forecasting scenarios. Table 6 provides the statistical analysis of mean absolute percentage errors across all forecasting horizons for the Hawaii dataset. Our proposed model enhances the forecasting accuracy by a substantial margin compared with the other models. The specific improvements on three datasets are as follows:

First, our proposed method achieved the highest forecasting accuracy in almost all forecasting scenarios, significantly outperforming other benchmark models. Compared with another deep learning method, our method, taking the 15-step forecasting on the Hawaii dataset as an example, shows reduced forecasting error, as measured by mean absolute percentage error, on average by 64.61%. The error is also reduced on average by 36.72% in the longer-term 30-step forecasts, which further indicates the advantage of our model in multi-horizon forecasting.

Second, our forecasting method exhibits greater stability across different scenarios. Taking the Hawaii dataset as an example, our method's average mean absolute percentage error across all forecasting horizons was 0.0573, as compared to the standard temporal fusion transformer's 0.0961 and other benchmark models' errors ranging between 0.1029 and 0.1396. These figures, presented in Table 6, underscore the superior and consistent performance of our proposed approach. Moreover, our method demonstrated the smallest standard deviation (Table 6) among the models, further highlighting its consistency and robustness.

Third, a closer examination of the performance across the three datasets reveals insightful patterns. The Hawaii dataset consistently showed the most accurate forecasts for all models, while Jiuzhaigou and Mount Siguniang exhibited slightly reduced accuracy. This difference may be attributed to domestic pandemic control policies in China, introducing additional complexities into the datasets. Despite these challenges, our proposed model significantly outperformed other benchmarks across various contexts, emphasizing its adaptability and effectiveness as a valuable tool in diverse forecasting applications.

Table 3. Forecasting results across models on the Hawaii dataset

Models	TPE-RobustSTL-TFT	TFT	ARIMAX	SARIMAX	DLM	LSTM	RF	ETS	TBATS
1-step-ahead forecasting									
MAE	**595.68**	1188.01	1751.46	1744.48	1110.37	1367.77	2224.58	1536.46	1624.35
RMSE	**897.80**	1535.68	2123.63	2116.29	1131.02	1391.55	2773.58	2219.21	2338.49
MAPE	**0.0240**	0.0512	0.0727	0.0726	0.0447	0.0552	0.0907	0.0660	0.0700
3-step-ahead forecasting									
MAE	**789.48**	1585.44	2215.28	2308.34	1614.49	1853.25	2634.50	2040.86	2281.68
RMSE	**1213.71**	2112.92	2650.54	2839.59	1641.51	1884.65	3315.48	2816.34	3185.97
MAPE	**0.0329**	0.0687	0.0907	0.0953	0.0651	0.0747	0.1066	0.0869	0.0985
7-step-ahead forecasting									
MAE	**1090.24**	2599.65	2323.02	2159.63	1864.31	2356.15	3091.28	1905.41	2664.87
RMSE	**1512.53**	3479.66	2799.06	2652.59	1896.70	2396.18	3859.29	2507.56	3562.69
MAPE	**0.0452**	0.1197	0.0921	0.0876	0.0751	0.0949	0.1234	0.0802	0.1134
15-step-ahead forecasting									
MAE	**1160.89**	2225.54	3161.06	3304.29	3347.89	3836.40	3160.91	3929.91	4564.93
RMSE	**1610.23**	2769.49	3949.58	4096.80	3405.90	3899.42	3950.31	4635.70	5663.45
MAPE	**0.0477**	0.0909	0.1269	0.1332	0.1348	0.1546	0.1298	0.1577	0.1905
30-step-ahead forecasting									
MAE	**2981.63**	3299.74	4323.34	4562.75	4828.75	5086.72	4428.23	4138.11	5270.66
RMSE	**3615.83**	4114.14	4890.18	5179.91	4908.80	5171.78	5087.59	5162.25	5872.66
MAPE	**0.1365**	0.1498	0.1749	0.1831	0.1946	0.2050	0.1868	0.1734	0.2254

Table 4. Forecasting results across models on the Jiuzhaigou dataset

Models	TPE-RobustSTL-TFT	TFT	ARIMAX	SARIMAX	DLM	LSTM	RF	ETS	TBATS

1-step-ahead forecasting									
MAE	**146.08**	405.73	845.68	942.91	428.03	812.67	1589.42	1665.65	923.25
RMSE	**295.97**	673.65	1344.25	1406.58	562.81	1068.03	2086.91	2188.97	1448.38
MAPE	**0.0266**	0.0697	0.1649	0.2200	0.0554	0.1048	0.2053	0.2150	0.1524
3-step-ahead forecasting									
MAE	**355.01**	762.28	1393.09	1419.99	1011.78	1909.49	2112.73	1974.75	1603.85
RMSE	**700.70**	1375.64	2307.84	2284.51	1332.45	2531.56	2781.14	2593.68	2635.42
MAPE	**0.0718**	0.1126	0.3132	0.3360	0.1315	0.2444	0.2734	0.2547	0.2796
7-step-ahead forecasting									
MAE	**851.80**	1340.92	1711.88	1799.20	2011.45	2552.84	3429.98	2768.18	1772.86
RMSE	**1566.05**	2077.66	2942.93	3034.25	2651.48	3361.02	4516.61	3654.04	3239.08
MAPE	**0.1599**	0.2200	0.4687	0.4923	0.2593	0.3284	0.4435	0.3564	0.4209
15-step-ahead forecasting									
MAE	**930.62**	1039.47	2163.39	2338.03	4266.18	2533.12	4991.36	3432.63	4128.51
RMSE	**1467.29**	1463.23	3369.42	3508.23	5593.15	3477.42	6557.28	4425.45	6847.71
MAPE	**0.1780**	0.2678	0.2891	0.3672	0.5527	0.6201	0.6455	0.6508	0.3853
30-step-ahead forecasting									
MAE	**1779.48**	1921.79	3825.95	3889.18	5009.34	5835.69	4052.26	4239.68	3544.34
RMSE	**2781.99**	3021.77	5392.76	5705.84	6563.40	7658.30	4765.40	5549.39	4905.65
MAPE	**0.3268**	0.5333	0.4474	0.4710	0.6510	0.7573	1.5183	1.1081	0.8684

Table 5. Forecasting results across models on the Mount Siguniang dataset

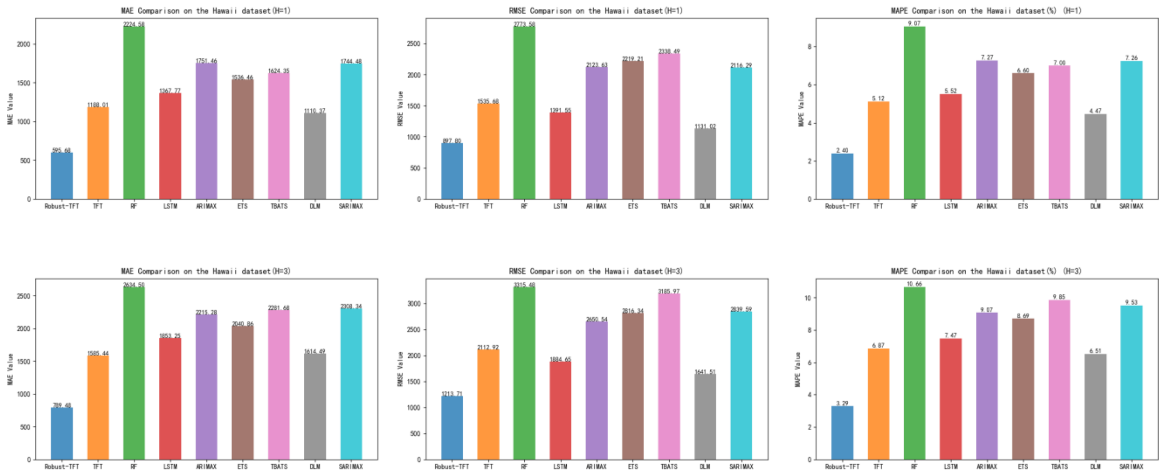
Models	TPE-RobustSTL-TFT	TFT	ARIMAX	SARIMAX	DLM	LSTM	RF	ETS	TBATS
1-step-ahead forecasting									
MAE	**83.21**	265.15	468.33	430.87	234.74	563.19	435.84	386.45	384.74
RMSE	**264.36**	393.70	798.76	765.29	271.62	730.17	600.05	510.37	529.20
MAPE	**0.0377**	0.1591	0.3390	0.2961	0.1351	0.3773	0.2627	0.3021	0.3150
3-step-ahead forecasting									
MAE	**131.26**	469.18	615.00	555.63	365.68	784.24	512.70	469.35	481.48
RMSE	**223.16**	790.23	972.17	958.23	423.65	1194.96	714.57	606.08	675.25
MAPE	**0.0992**	0.2604	0.4368	0.3780	0.2098	0.4988	0.2908	0.3662	0.4137
7-step-ahead forecasting									
MAE	**274.68**	380.11	630.80	587.11	606.14	921.34	544.00	661.21	610.15
RMSE	**470.19**	590.96	1005.37	976.32	707.87	1269.43	754.86	1008.88	918.67
MAPE	**0.1866**	0.2167	0.4404	0.4087	0.3485	0.6458	0.3063	0.6591	0.5610
15-step-ahead forecasting									
MAE	**333.05**	508.92	697.81	659.65	785.84	1015.80	552.29	486.27	644.31
RMSE	**461.99**	679.85	1072.64	1045.93	912.49	1395.31	740.16	653.73	862.132
MAPE	**0.2281**	0.3230	0.5054	0.4598	0.4497	0.7635	0.3073	0.3837	0.5685

30-step-ahead forecasting									
MAE	**342.30**	549.74	661.77	626.68	606.55	921.34	558.78	404.32	723.14
RMSE	**498.87**	760.11	1056.38	1031.97	709.22	1269.43	752.36	570.97	1052.45
MAPE	**0.2173**	0.3362	0.4314	0.4048	0.3472	0.6458	0.3041	0.3476	0.7784

Table 6. Statistical analysis of mean absolute percentage errors across all forecasting horizons on the Hawaii dataset

Forecasting models	Mean	Minimum	Maximum	Standard Deviation
TPE-RobustSTL-TFT	0.0573	0.0240	0.1365	0.0405
TFT	0.0961	0.0512	0.1498	0.0353
ARIMAX	0.1115	0.0727	0.1749	0.0363
SARIMAX	0.1144	0.0726	0.1831	0.0398
DLM	0.1029	0.0447	0.1946	0.0548
LSTM	0.1169	0.0552	0.2050	0.0552
RF	0.1275	0.0907	0.1868	0.0326
ETS	0.1128	0.0660	0.1734	0.0438
TBATS	0.1396	0.0700	0.2254	0.0586

Figures 3, 4, and 5 offer a more intuitive understanding of the performance by presenting visualizations of the evaluation metrics. These visualizations further emphasize the superiority of the proposed model over other benchmark models, demonstrating a clear advantage across all forecasting horizons.



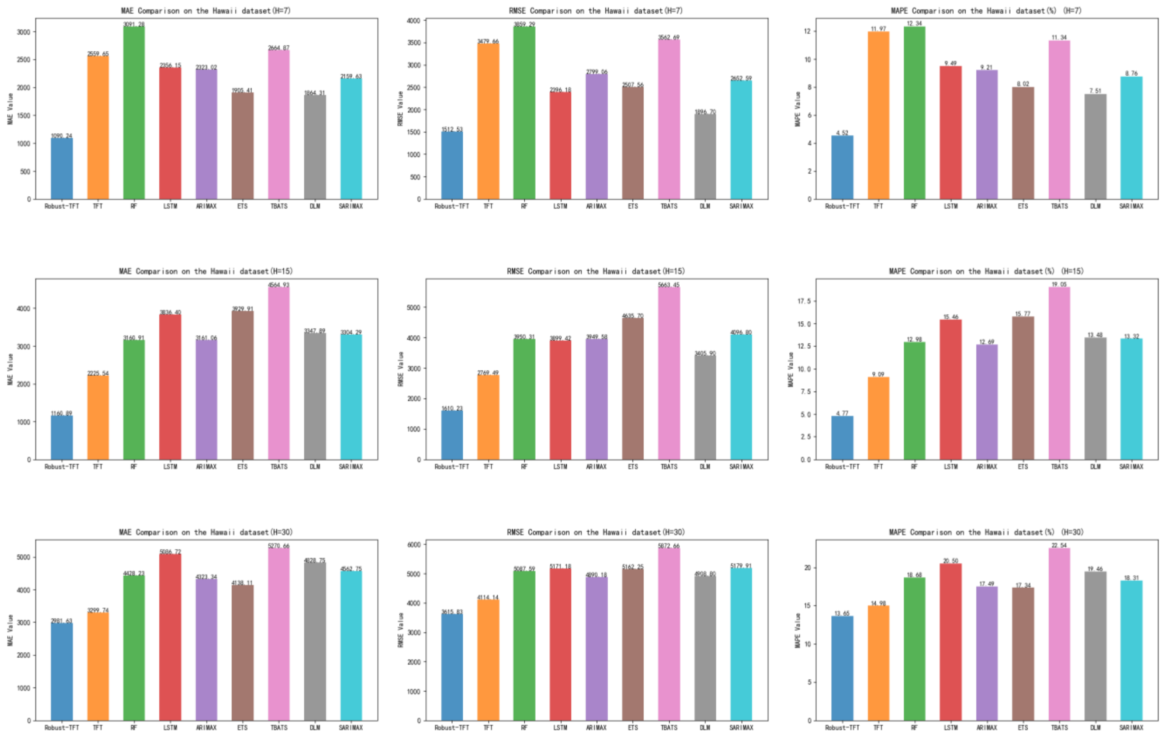
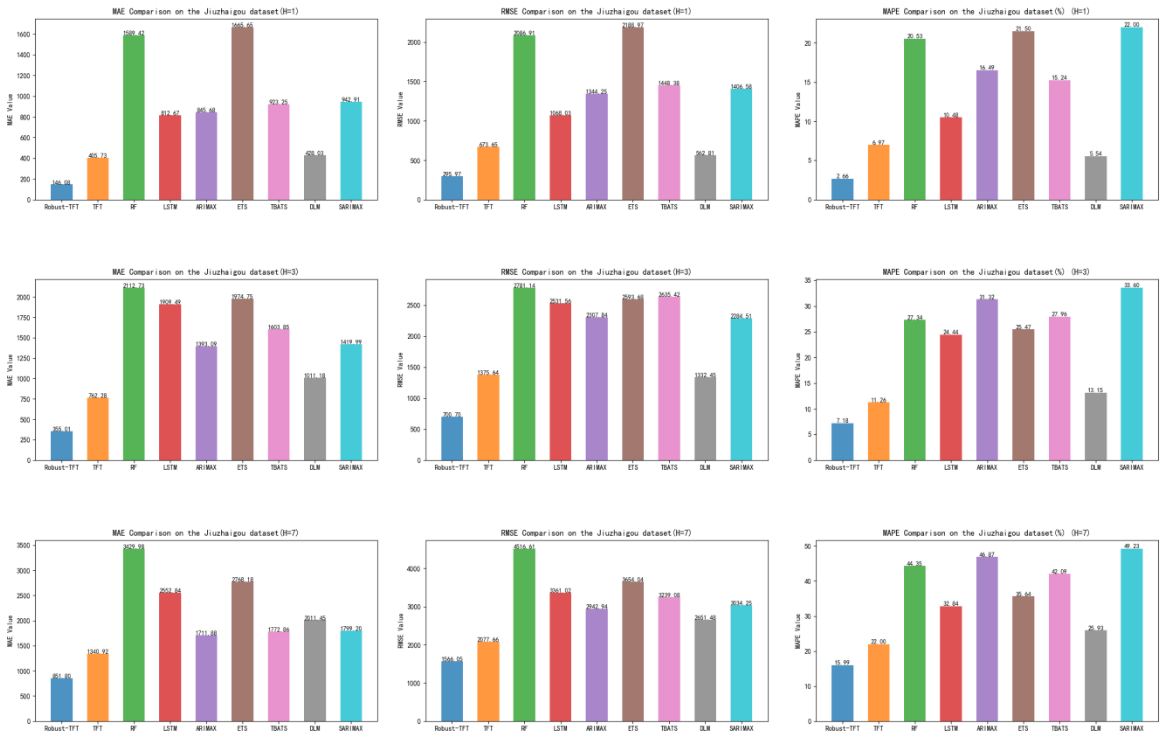


Figure 3. Visualization of the forecasting accuracy of all models on the Hawaii dataset

Note: The comparison charts for the various models, arranged from left to right, display the contrasts among the three forecasting errors. Here, h (3, 5, 7, 15, 30) represents the forecast horizons.



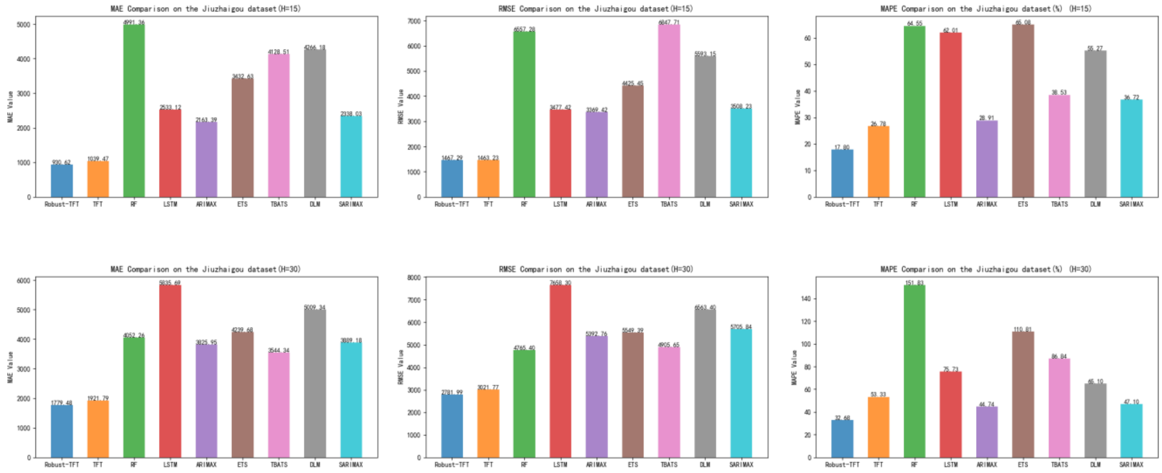
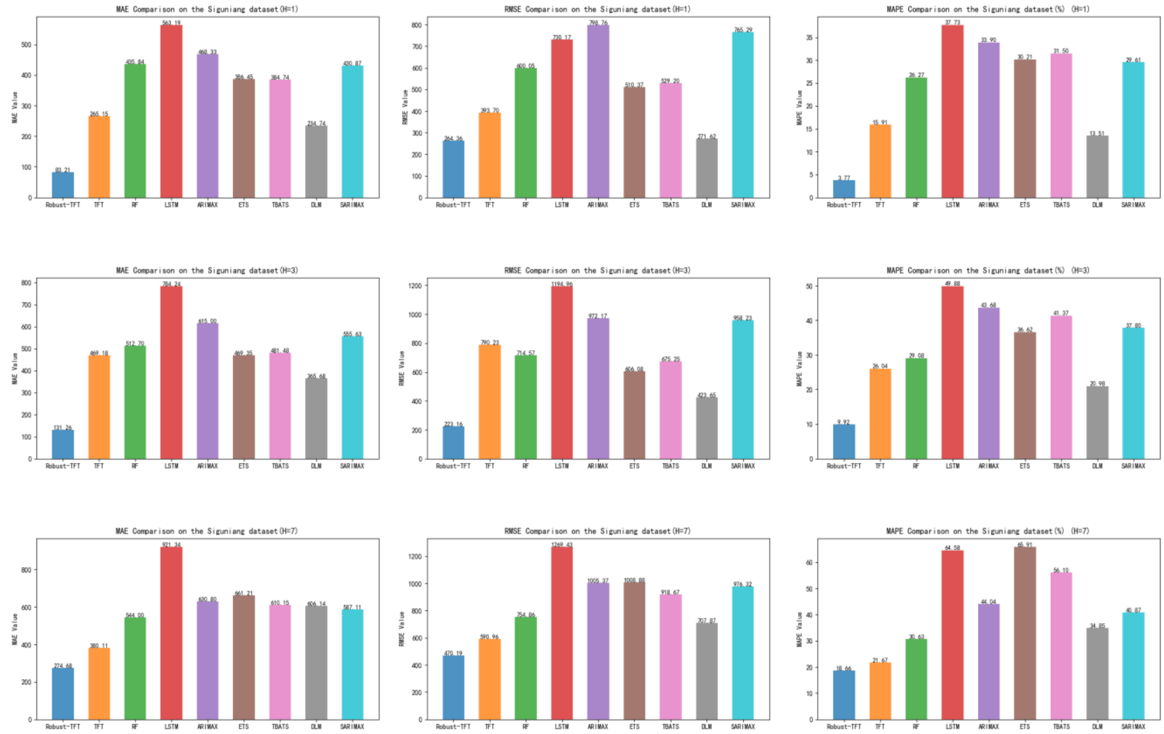


Figure 4. Visualization of the forecasting accuracy of all models on the Jiuzhaigou dataset



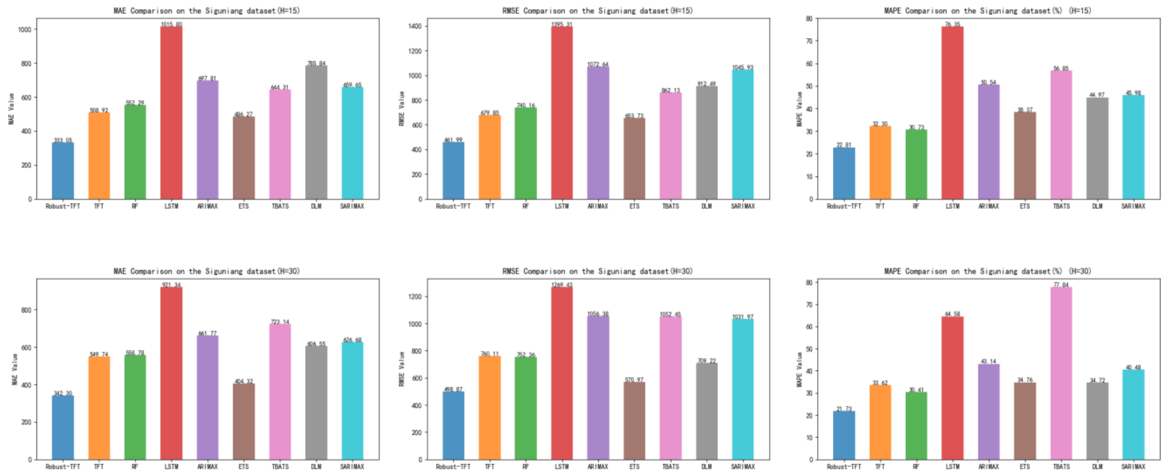


Figure 5. Visualization of the forecasting accuracy of all models on the Mount Siguniang dataset

Moreover, using the three-step forecast results from three datasets on the test set as examples, we have generated fitted plots for the forecasted results of each model, as shown in Figures 6, 7, and 8 below. These plots provide a more direct comparison of the proposed model's predictions against those of the other benchmark models. The plots also highlight the robustness and accuracy of our proposed model in capturing the underlying patterns and dynamics of tourist arrivals.

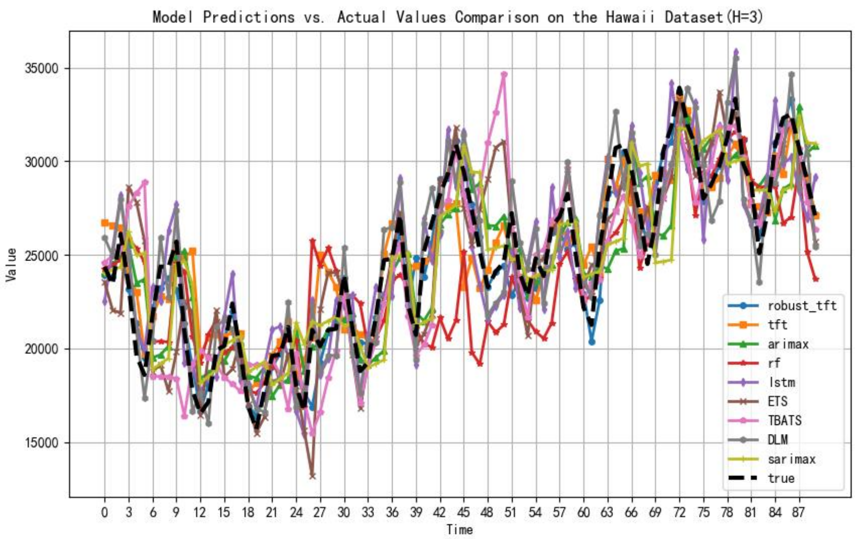


Figure 6. Predicted results on the Hawaii test set from January 6, 2022, to April 5, 2022

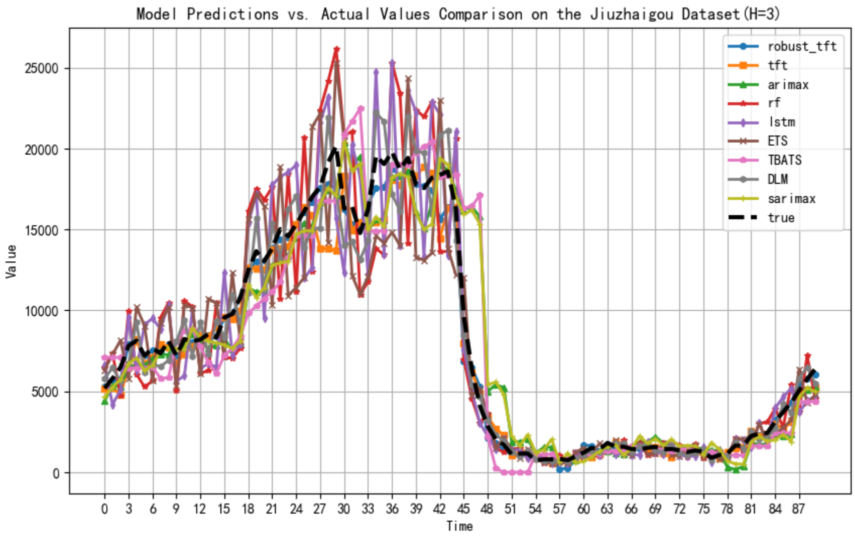


Figure 7. Predicted results on the Jiuzhaigou test set from June 15, 2021, to September 12, 2021

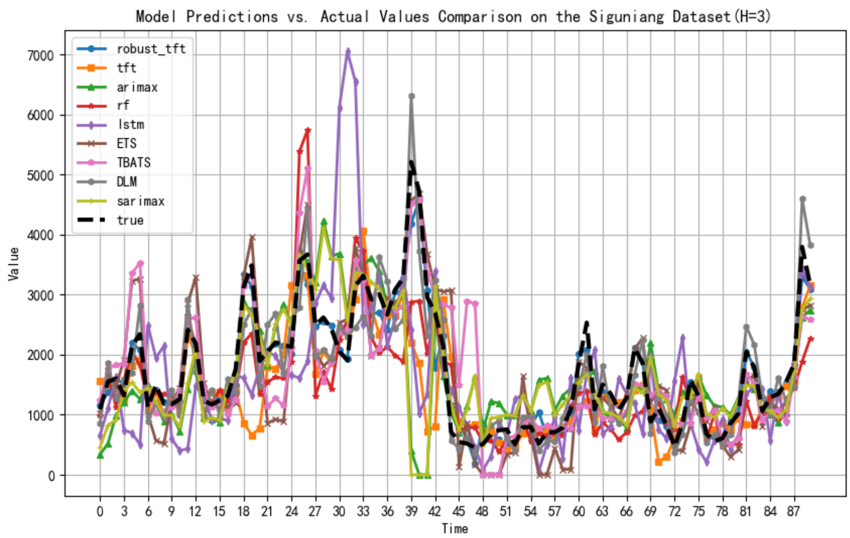


Figure 8. Predicted results on the Mount Siguniang test set from June 15, 2021, to September 12, 2021

In summary, the proposed model exhibits remarkable forecasting performance, outperforming other benchmark models across various time horizons and evaluation metrics. This superior performance highlights the potential of the proposed model as a valuable tool that can be used by tourism researchers and practitioners in predicting daily tourist arrivals with greater accuracy and reliability.

4.4 Robust check

To further substantiate the robustness of our proposed model, we conducted a Diebold–Mariano test, a widely accepted method for comparing the forecasting accuracy of different models (Sun et al., 2019). The Diebold–Mariano test provides a statistical evaluation of the null hypothesis, which assumes that the two models being compared have equal forecast accuracy. A significant Diebold–Mariano test statistic indicates that one model is superior to the other in terms of forecasting performance.

The results of the Diebold–Mariano test for the three datasets are visualized using heatmaps shown in Figures 9, 10, and 11. The heatmaps demonstrate that our proposed model consistently outperforms the other eight benchmark models in terms of the Diebold–Mariano statistic. As the forecasting horizon extends, the p-values become increasingly significant. By the time the forecasting horizon reaches 30 days, the p-values become highly significant, suggesting that the proposed model's superior performance is statistically meaningful.

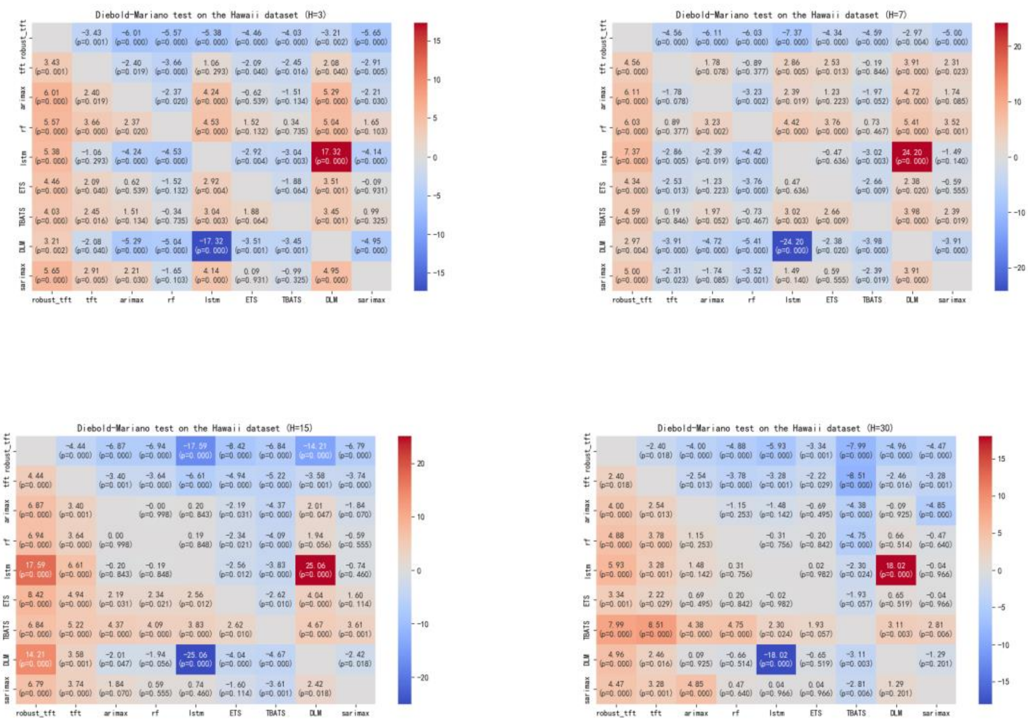


Figure 9. Results of the Diebold–Mariano test on the Hawaii dataset

Note: The number located at the top of each square represents the Diebold–Mariano statistic, which indicates the performance difference between the models. The number below denotes the significance level p, and it is generally considered that a p-value less than 0.05 indicates statistical significance.

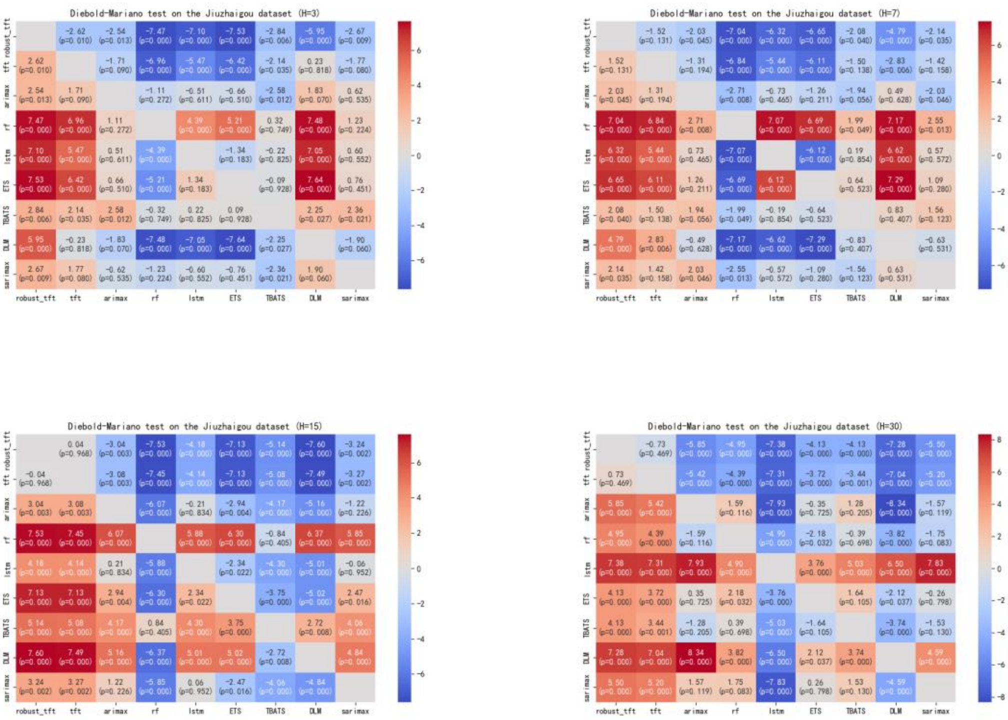


Figure 10. Results of the Diebold–Mariano test on the Jiuzhaigou dataset

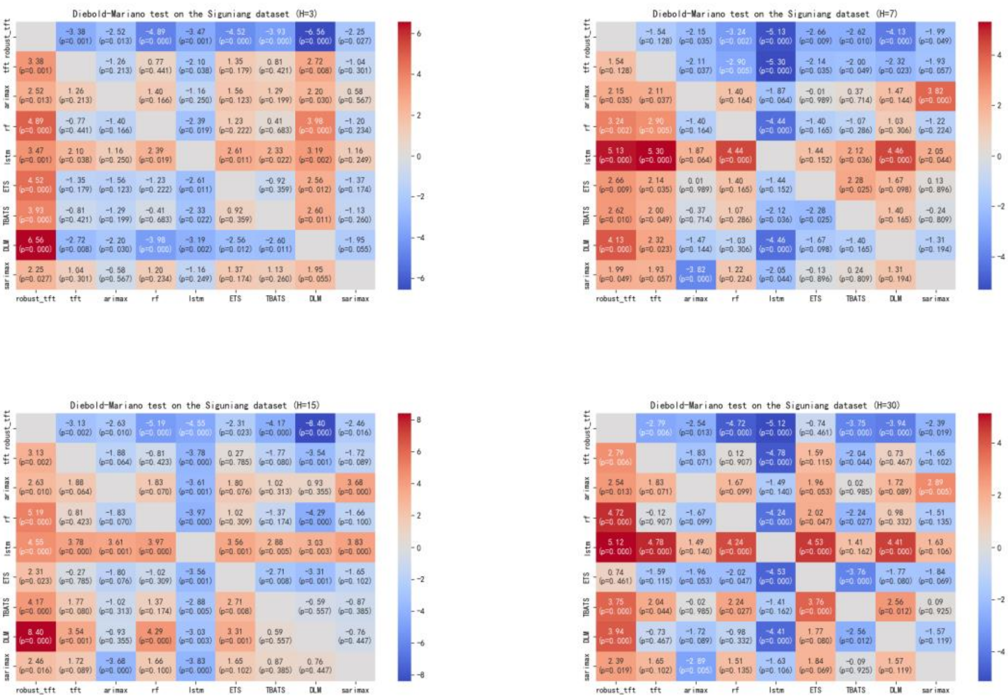


Figure 11. Results of the Diebold–Mariano test on the Mount Siguniang dataset

4.5 Analysis of results interpretability

In this section, we provide an interpretability analysis obtained from the model's attention weights and variable importance, focusing on the Jiuzhaigou dataset as a case study. This analysis was conducted based on the 30-day forecast results for the Jiuzhaigou tourist arrivals, to explore the underlying patterns and relationships between the variables and their impacts on the forecasting of the trend, seasonal, and residual components of the time series.

Figure 12 indicates the attention weights for the trend, seasonal, and residual components computed using the proposed Transformer-based model. The varying attention weight values across different days can be observed from the figure.

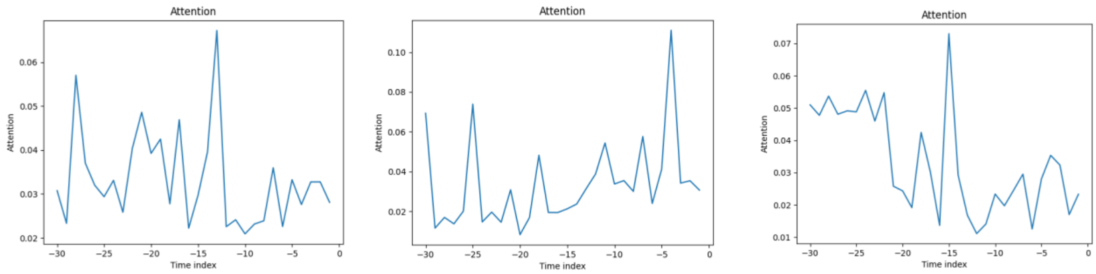


Figure 12. Attention weights of the decomposed components (from left to right): trend, seasonal, and residual

First, the higher attention weights on specific days are used to capture essential changes in the trend component. For instance, if there is a noticeable increase or decrease in the trend component on a particular day, the model might assign a higher attention weight to that day to better understand and predict the trend. The attention weights can fluctuate, reflecting the model's adaptability to various influencing factors, such as local events, pandemics, or major tourism developments.

Second, for the seasonal component, attention weights vary similarly to the trend component but focus on capturing recurring seasonal patterns. Higher attention weights might be assigned to holidays or weather patterns to recognize the seasonal demand. These weights may not always align with the trend component's weights, as they focus on periodic fluctuations rather than long-term changes.

Third, the attention weights for the residual component are distinct from trend and seasonal components, emphasizing the unexplained or random variations in the data. The model allocates higher weights to irregular patterns, possibly due to noise or random events. Given that the residual component focuses on capturing the unexplained variations that are not associated with the long-term trends or seasonal patterns, the attention weights for the residual component might not align with those in the trend or seasonal components.

In the proposed model, the encoder-decoder variable importance facilitates interpretability by quantifying the significance of different features within the forecasting process. Appendix C offers detailed analysis of the importances of variables. In summary, the findings regarding the interpretability of the forecasting framework can be summarized as follows:

(1) The analysis of the attention weights shows the model's adaptability in identifying key aspects across different days in the trend, seasonal, and residual components. This adaptability highlights the need to consider various factors, such as local epidemic situations, major events, and key developments in the tourism sector. Such consideration enhances the comprehensive understanding of tourism demand dynamics.

(2) The encoder variable importance assessment demonstrates the complex interplay among different factors influencing the trend, seasonal, and residual components. This finding highlights the necessity for tourism stakeholders to consider a diverse set of features to capture the multifaceted nature of the tourism industry, thereby enabling a more accurate forecasting and effective decision-making.

(3) The examination of decoder variable importance provides valuable insights into the factors that drive overall trends, seasonal patterns, and residual variations in tourism demand. By recognizing and understanding the impact of these factors, tourism authorities and businesses can devise well-informed policies, marketing strategies, and responsive measures. In turn, these can optimize resource allocation, improve service provision, and address unforeseen challenges, thus fostering a resilient and thriving tourism ecosystem.

5. Conclusions and implications

This study presents a novel approach to tourism demand forecasting, which synergistically combines the strengths of a Bayesian optimization method used for hyperparameter tuning, transformer architecture, and a novel robust decomposition approach into the proposed transformer model. Our analysis shows that this model outperforms conventional and advanced forecasting methods, including the most recent deep learning model, thereby demonstrating the potential of combining advanced machine learning techniques with domain-specific approaches. This work also utilized the tree-structured parzen estimator algorithm for hyperparameter selection to improve forecasting accuracy and reliability.

Our study provides valuable insights into the factors that shape tourism demand and the underlying patterns in the data, revealing the significance of input variables, such as Baidu Index keywords. The proposed model showcases proficiency in handling nonlinear patterns, multiple input variables, and complex interactions, which further highlights its utility in addressing the challenges of forecasting tourism demand. Our findings have important implications for various stakeholders in the tourism industry, such as policymakers, marketers, and service providers, who can use this knowledge to identify key factors that drive tourism demand and devise targeted interventions to enhance the industry's competitiveness and sustain its growth.

Moreover, our study highlights the importance of considering unforeseen factors and external variables in tourism demand forecasting, such as short-term changes in search interest or regional events, and the incorporation of external variables, such as data related to the covid-19 pandemic, to capture more effectively the influence of external events on tourism demand. Integrating these variables into the proposed transformer model facilitates precise predictions and a deeper comprehension of the factors that shape tourism demand. Our findings suggest that the incorporation of external variables can significantly enhance the accuracy and robustness of tourism demand forecasting models, particularly during times of significant disruptions.

In conclusion, the practical implications derived from this study can influence decision-making processes. They assist industry stakeholders in creating well-defined strategies that enhance the positive aspects of tourism while simultaneously reducing its negative effects. The potential of the proposed model for precise and dependable tourism demand forecasting emphasizes the significance of input variable selection and the promise of transformer-based methodologies in predicting tourism demand. In addition, integrating external variables into tourism demand forecasting models may contribute to addressing external events and fostering more sustainable, resilient tourism practices.

While this study makes significant contributions, it also has certain limitations that point to future research directions. First, our investigation represents an initial foray into applying the state-of-the-art transformer architecture to tourism demand forecasting, and we propose an innovative method to improve the predictive performance of transformers. We have analyzed the forecasting scenarios for three daily tourist attractions, and demonstrated the improved forecasting accuracy. Nonetheless, future research should extend this evaluation to assess whether transformer-based deep learning models and their variants are applicable to a wider array of tourism demand forecasting contexts, including monthly tourism destination demand predictions. Researchers can also evaluate the performance of our framework using diverse sources, such as other search engine data and multimodal social media data. Moreover, the feature selection network structure of the temporal fusion transformer model facilitates an initial exploration into the interpretability of deep learning models by quantifying the degree to which various explanatory variables influence tourist flow. In future research, it would be beneficial to integrate recent advancements in explainable machine learning, such as post-hoc interpretability techniques like Shapley additive explanations, to further augment the interpretability of deep learning models in tourism demand forecasting.

References

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Herrera, F. (2023). Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion, 99*, 101805.

Bergstra, J., Yamins, D., & Cox, D. (2013). Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. *Proceedings of the 30th International Conference on International Conference on Machine Learning (PMLR)* (pp. 115-123). Atlanta GA, USA.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

Bi, J. W., Li, C., Xu, H., & Li, H. (2022). Forecasting daily tourism demand for tourist attractions with big data: An ensemble deep learning method. *Journal of Travel Research, 61*(8), 1719-1737.

Bi, J. W., Li, H., & Fan, Z. P. (2021). Tourism demand forecasting with time series imaging: A deep learning model. *Annals of Tourism Research, 90*, 103255.

Bi, J. W., Liu, Y., & Li, H. (2020). Daily tourism volume forecasting for tourist attractions. *Annals of Tourism Research, 83*, 102923.

Chen, J. L., Li, G., Wu, D. C., & Shen, S. (2019). Forecasting seasonal tourism demand using a multiseries structural time series method. *Journal of Travel Research, 58*(1), 92-103.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. *Journal of Official Statistics, 6*(1), 3-73.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Dwyer, L., Forsyth, P., & Spurr, R. (2004). Evaluating tourism's economic effects: new and old approaches. *Tourism Management*, *25*(3), 307-317.

Essien, A., & Chukwukelu, G. (2022). Deep learning in hospitality and tourism: a research framework agenda for future research. *International Journal of Contemporary Hospitality Management*, *34*(12), 4480-4515.

Fotiadis, A., Polyzos, S., & Huan, T.-C. T. (2021). The good, the bad and the ugly on COVID-19 tourism recovery. *Annals of Tourism Research, 87*, 103117.

Grossi, L., & Mussini, M. (2021). Seasonality in tourist flows: Decomposing and testing changes in seasonal concentration. *Tourism Management, 84*, 104289.

Gunter, U., & Zekan, B. (2021). Forecasting air passenger numbers with a GVAR model. *Annals of Tourism Research, 89*, 103252.

Hu, M., Li, H., Song, H., Li, X., & Law, R. (2022). Tourism demand forecasting using tourist-generated online review data. *Tourism Management, 90*, 104490.

Jiao, X., Chen, J. L., & Li, G. (2021). Forecasting tourism demand: Developing a general nesting spatiotemporal model. *Annals of Tourism Research, 90*, 103277.

Kucklick, J.-P., & Müller, O. (2023). Tackling the accuracy-interpretability trade-off: Interpretable deep learning models for satellite image-based real estate appraisal. *ACM Transactions on Management Information Systems, 14*(1), 1-24.

Kulshrestha, A., Krishnaswamy, V., & Sharma, M. (2020). Bayesian BILSTM approach for tourism demand forecasting. *Annals of Tourism Research*, 83, 102925.

Lara-Benítez, P., Carranza-García, M., & Riquelme, J. C. (2021). An experimental review on deep learning architectures for time series forecasting. *International Journal of Neural Systems, 31*(03), 2130001.

Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research, 75*, 410-423.

Li, C., Zheng, W., & Ge, P. (2022). Tourism demand forecasting with spatiotemporal features. *Annals of Tourism Research, 94*, 103384.

Li, C., Ge, P., Liu, Z., & Zheng, W. (2020). Forecasting tourist arrivals using denoising and potential factors. *Annals of Tourism Research, 83*, 102943.

Li, G., Song, H., & Witt, S. F. (2005). Recent developments in econometric modeling and forecasting. *Journal of Travel Research, 44*(1), 82-99.

Li, G., Song, H., & Witt, S. F. (2006). Time varying parameter and fixed parameter linear AIDS: An application to tourism demand forecasting. *International Journal of Forecasting, 22*(1), 57-71.

Li, G., Wu, D. C., Zhou, M., & Liu, A. (2019). The combination of interval forecasts in tourism. *Annals of Tourism Research, 75*, 363-378.

Li, H., Hu, M., & Li, G. (2020). Forecasting tourism demand with multisource big data. *Annals of Tourism Research, 83*, 102912.

Li, H., Gao, H., & Song, H. (2023). Tourism forecasting with granular sentiment analysis. Annals of Tourism Research, 103, 103667.

Li, X., Law, R., Xie, G., & Wang, S. (2021). Review of tourism forecasting research with internet data. *Tourism Management, 83*, 104245.

Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism Management, 59*, 57-66.

Li, X., & Law, R. (2020). Forecasting tourism demand with decomposed search cycles. *Journal of Travel Research, 59*(1), 52-68.

Llewellyn, M., Ross, G., & Ryan-Saha, J. (2023). COVID-era forecasting: Google trends and window and model averaging. *Annals of Tourism Research*, 103, 103660.

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting, 37*(4), 1748-1764.

Liu, X., Liu, A., Chen, J. L., & Li, G. (2023). Impact of decomposition on time series bagging forecasting performance. *Tourism Management, 97*, 104725.

Liu, H., Liu, Y., Li, G., & Wen, L. (2021). Tourism demand nowcasting using a LASSO-MIDAS model. *International Journal of Contemporary Hospitality Management*, *33*(6), 1922-1949.

Monje, L., Carrasco, R. A., Rosado, C., & Sánchez-Montañés, M. (2022). Deep learning XAI for bus passenger forecasting: A use case in Spain. *Mathematics, 10*(9), 1428.

Mueller, R., & Sobreira, N. (2024). Tourism forecasts after COVID-19: Evidence of Portugal. *Annals of Tourism Research Empirical Insights, 5(1), 100127.*

Park, E., Park, J., & Hu, M. (2021). Tourism demand forecasting with online news data mining. *Annals of Tourism Research, 90*, 103273.

Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training recurrent neural networks. *Proceedings of the 30th International Conference on Machine Learning (ICML)* (pp. 1310-1318). Atlanta GA, USA.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Ziel, F. (2022). Forecasting: theory and practice. *International Journal of Forecasting, 38*(3), 705-871.

Rong, G., Li, K., Su, Y., Tong, Z., Liu, X., Zhang, J., Li, T. (2021). Comparison of tree-structured parzen estimator optimization in three typical neural network models for landslide susceptibility assessment. *Remote Sensing, 13*(22), 4694.

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. *Tourism Management, 29*(2), 203-220.

Song, H., Qiu, R. T., & Park, J. (2023). Progress in tourism demand research: Theory and empirics. *Tourism Management, 94*, 104655.

Song, H., Qiu, T., & Park, J. (2019). A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research, 75,* 338-362.

Sun, S., Wei, Y., Tsui, K. L., & Wang, S. (2019). Forecasting tourist arrivals with machine learning and internet search index. *Tourism Management, 70*, 1-10.

Sun, S., Li, Y., Guo, J.-e., & Wang, S. (2022). Tourism demand forecasting: An ensemble deep learning approach. *Tourism Economics, 28*(8), 2021-2049.

Sun, Y., Zhang, J., Li, X., & Wang, S. (2023). Forecasting tourism demand with a new time-varying forecast averaging approach. *Journal of Travel Research*, 62(2), 305-323.

Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., & Troncoso, A. (2021). Deep learning for time series forecasting: a survey. *Big Data, 9*(1), 3-21.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems 30 (NIPS)* (pp. 6000-6010). Long Beach, CA.

Wen, Q., Gao, J., Song, X., Sun, L., Xu, H., & Zhu, S. (2019). RobustSTL: A robust seasonal-trend decomposition algorithm for long time series. *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)* (pp. 5409-5416). Honolulu, Hawaii, USA.

Wu, B., Wang, L., & Zeng, Y.-R. (2022). Interpretable wind speed prediction with multivariate time series and temporal fusion transformers. *Energy, 252*, 123990.

Wu, D. C., Song, H., & Shen, S. (2017). New developments in tourism and hotel demand modeling and forecasting. *International Journal of Contemporary Hospitality Management, 29*, 507-529.

Wu, J., Li, M., Zhao, E., Sun, S., & Wang, S. (2023). Can multi-source heterogeneous data improve the forecasting performance of tourist arrivals amid COVID-19? Mixed-data sampling approach. *Tourism Management, 98*, 104759.

Wu, D. C., Li, G., & Song, H. (2022). *Econometric Modelling and Forecasting of Tourism Demand.* Taylor & Francis. London.

Xie, G., Qian, Y., & Wang, S. (2020). A decomposition-ensemble approach for tourism forecasting. *Annals of Tourism Research, 81*, 102891.

Xie, G., Qian, Y., & Wang, S. (2021). Forecasting Chinese cruise tourism demand with big data: An optimized machine learning approach. *Tourism Management, 82*, 104208.

Xu, S., Liu, Y., & Jin, C. (2023). Forecasting daily tourism demand with multiple factors. *Annals of Tourism Research*, 103, 103675.

Yang, Y., & Zhang, H. (2019). Spatial-temporal forecasting of tourism demand. *Annals of Tourism Research, 75*, 106-119.

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, *1168*(2), 022022.

Yu, X., Zhang, D., Zhu, T., & Jiang, X. (2022). Novel hybrid multi-head self-attention and multifractal algorithm for non-stationary time series prediction. *Information Sciences, 613*, 541-555.

Zhang, C., Li, M., Sun, S., Tang, L., & Wang, S. (2022). Decomposition methods for tourism demand forecasting: A comparative study. *Journal of Travel Research, 61*(7), 1682-1699.

Zhang, Y., Li, G., Muskat, B., & Law, R. (2021). Tourism demand forecasting: A decomposed deep learning approach. *Journal of Travel Research, 60*(5), 981-997.

Zhang, Y., Li, G., Muskat, B., Law, R., & Yang, Y. (2020). Group pooling for deep tourism demand forecasting. *Annals of Tourism Research*, 82, 102899.

Zheng, W., Huang, L., & Lin, Z. (2021). Multi-attraction, hourly tourism demand forecasting. *Annals of Tourism Research*, 90, 103271.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)* (pp. 11106-11115). Washington, D.C., USA.