Hue University Journal of Science: Techniques and Technology; pISSN 2588-1175 | eISSN 2615-9732

Vol. 134, No. 2B, 2025, 33–50; DOI: 10.26459/hueunijtt.v134i2B.8112

An integrated approach combining neural networks and genetic algorithms for multisource time series forecasting

Ngo Van Son[1, 2*], Dao Thi Tuong Van[2], Nguyen Dang Binh[1]

[1] University of Sciences, Hue University, Hue, Vietnam
[2] School of Hospitality and Tourism - Hue University, Hue, Vietnam

Abstract. This paper proposes a data-driven framework that leverages multisource information to significantly enhance time series forecasting within the tourism domain. Specifically, we integrate a Multilayer Perceptron (MLP) neural network with the NSGA-II algorithm for efficient hyperparameter optimization. In addition to historical data on monthly international tourist arrivals to Vietnam, two exogenous data sources are incorporated: (i) sentiment indices extracted from tourist reviews on TripAdvisor, and (ii) search trend data from Google Trends. Critical hyperparameters, including input window size, learning rate, number of epochs, early stopping, and normalization methods, are simultaneously optimized using NSGA-II. Experimental results demonstrate that the proposed NSGA-II-MLP model, utilizing multisource data, consistently outperforms the standard MLP and exhibits robust performance across both the pre-COVID-19 and volatile pandemic periods. The results underscore the effectiveness of combining multisource data with evolutionaroptimization for tourism demand forecasting.

Keywords: Time series forecasting, multisource data, exogenous variables, MLP, NSGA-II

1 Introduction

Time series forecasting is one of the fundamental research directions in machine learning and artificial intelligence. Traditional approaches often rely solely on historical data, while in practice, forecasting systems must process information from multiple heterogeneous sources such as online search signals, social media data, and IoT sensors. While effectively integrating and representing these diverse data sources can substantially enhance forecasting performance, it simultaneously introduces significant challenges in information processing and fusion. In recent years, multisource data utilization has become an emerging trend, especially with the rapid growth of user-generated emotional data on digital platforms—where users continuously leave reviews, comments, and feedback about products, services, and experiences [1]. In the tourism context, TripAdvisor, a prominent online ecosystem, stands out by gathering millions of travelers' opinions on destinations, accommodations, and tourism services. These reviews not only reflect individual sentiments but also represent the collective perception of travelers regarding service quality and satisfaction trends over time. The rapid expansion of the digital

* *Corresponding:* ngovanson@hueuni.edu.vn

Received: 13–10–2025; Accepted: 10–11–2025

economy and the shift in consumer behavior toward online environments have generated massive temporal data resources, including comments, ratings, and feedback on tourism platforms. When properly processed and structurally represented, such data offer immense potential for predicting tourism trends and demand. According to [2], travelers' positive or negative sentiments can directly influence their travel intentions and level of interest in a destination. However, the integration of sentiment data into time series forecasting remains a relatively new yet promising research direction.

Tourism plays a vital role in Vietnam's economy, contributing significantly to foreign exchange earnings and employment generation. Consequently, accurate tourism demand forecasting is critical for policymakers and businesses to effectively plan, coordinate strategies, and optimize resource allocation. As user behavior evolves rapidly and big data expands, conventional forecasting models based solely on historical patterns have become inadequate. Incorporating sentiment data from digital platforms thus emerges as a timely and necessary approach in the era of data-driven decision-making. Given that tourism demand is highly seasonal and sensitive to external factors (e.g., economic fluctuations, pandemics, online behaviors), accurate forecasting plays a crucial role in policy planning, destination management, and resource allocation. Recent studies have demonstrated that online behavioral data can reflect real-time travel demand and trends, underscoring the potential of multisource data integration in forecasting applications.

In this study, we propose a multisource forecasting framework that leverages both historical data and exogenous factors, including sentiment indices extracted from online platforms and search trend data from Google Trends. By combining these data sources, the forecasting model not only learns from historical patterns but also captures supplementary behavioral and contextual signals from users and the market. We employ a Multilayer Perceptron (MLP) neural network due to its simplicity, nonlinear generalization capability, and flexibility in integrating features from heterogeneous sources. Moreover, the Non-dominated Sorting Genetic Algorithm II (NSGA-II) [3] is applied to optimize the MLP hyperparameters, forming an integrated NSGA-II-MLP model. In contrast to manual tuning or conventional GA-based methods, NSGA-II leverages two pivotal mechanisms—Pareto ranking and crowding distance—to substantially enhance search efficiency and preserve solution diversity.

The main contributions of this paper are summarized as follows:

- Analyzing and identifying optimal tourism-related keywords through Google Suggest to collect high-quality Google Trends data that represent the most frequent search terms used by international tourists visiting Vietnam.

- Extracting and analyzing sentiment information from TripAdvisor reviews using deep learning–based sentiment models, and transforming them into exogenous variables that reflect travelers' emotional dynamics and satisfaction trends over time.

- Proposing an integrated NSGA-II–MLP model for time series forecasting that effectively combines historical and exogenous data sources.

- Evaluating the model's performance during both the pre–COVID-19 and pandemic periods to assess its adaptability to volatile data conditions.

- Suggesting future extensions that employ modern deep learning architectures (e.g., Temporal Fusion Transformer [4], TSMixer [5]) to capture long-term dependencies and complex feature interactions.

- Unlike previous GA–MLP studies that use GA mainly to optimize internal network parameters (weights and biases), this study adapts NSGA-II for automated hyperparameter optimization of MLP, including input window size, learning rate, number of epochs, early stopping, and normalization. This improves optimization stability and reduces manual trial-and-error when training on multisource data.

The remainder of this paper is structured as follows: Section 2 presents related work. Section 3 describes the data, preprocessing pipeline, and proposed model. Section 4 provides experimental results and discussion. Finally, Section 5 concludes the paper and outlines directions for future research.

2 Related Work

Tourism demand forecasting is a widely explored research topic, as it plays a crucial role in both local and national economic planning. Numerous studies have investigated tourism demand across various destinations and countries, emphasizing its importance to economic growth and policy-making [6][7][8]. Traditionally, time series models have been employed to analyze and forecast tourism demand based on historical observations. In recent years, hybrid and integrated forecasting methods have gained attention as effective strategies to enhance prediction accuracy [9][10][11]. This study focuses primarily on tourism demand forecasting models that integrate multiple time series data sources, particularly Google search trends and sentiment analysis from TripAdvisor reviews.

Google is the world's most widely used search engine, accounting for over 90% of global search queries. Data from Google Trends, which have been available since January 2004, reflect user search behavior by keyword, time, category, and geographic location. Following the approach of [12], this study collects daily Google Trends data over a consecutive 30-day window and computes their average values. This method smooths fluctuations in the time series and enhances data stability. The resulting time series are normalized to the total number of search queries in each country or region, with values scaled to [0, 100].

In practice, there exists a causal relationship between travelers' information-seeking behavior and their travel decisions. Understanding this relationship helps capture tourism demand more accurately. In [13], Volchek et al. investigated methods to improve the accuracy of tourism demand forecasting at the micro level. The authors predicted visitor numbers at five London museums and compared the predictive performance of several models, including Naïve, Seasonal Naïve, Seasonal Autoregressive Moving Average (SARMA), SARMA with explanatory variables, Mixed-Frequency SARMAX, and Artificial Neural Networks. Their results expanded understanding of how various data types and forecasting algorithms perform at the attraction level. Incorporating Google Trends data into time series models improved prediction accuracy for visitor arrivals. However, no single model consistently outperformed others across all scenarios—the forecasting accuracy varied between short-term and long-term horizons. Using high-frequency search data enabled weekly predictions, which are highly valuable for destination- and attraction-level planning.

The study in [14] emphasized the business value of search engine data analytics, noting that insights derived from Google Trends can reveal market opportunities. Li et al. proposed a systematic approach for obtaining and utilizing Google Trends data to generate forecasts, using Taiwan's tourism demand as a case study. The forecasted results were compared with actual data from the Taiwan Tourism Bureau. Their findings highlighted three essential challenges in forecasting based on search query data: (1) identifying which search engines are most used by tourists; (2) determining the languages used when tourists search online; and (3) selecting appropriate keywords in Google Trends to retrieve relevant data for tourism demand prediction.

Similarly, Havránek et al. [15] examined the usefulness of Google Trends data for predicting monthly tourist arrivals and overnight stays in Prague from January 2010 to December 2016. They assessed whether including Google Trends improved predictive accuracy compared to models without search data and whether high-frequency (weekly) data provided advantages over low-frequency (monthly) data using mixed-frequency sampling techniques. The empirical results confirmed the potential of Google Trends for enhancing forecasting performance. Specifically, search information from one week and two months prior to actual arrivals proved valuable for prediction. Models using weekly Google Trends data outperformed those based on monthly data or without search data altogether.

Alongside the digitalization trend, online review platforms such as TripAdvisor have become rich sources of information reflecting tourists' experiences and emotions. Through reviews and ratings, users share not only personal opinions but also quantitative and qualitative indicators of destination, accommodation, culinary, and activity quality. Numerous studies have demonstrated that sentiment analysis of user-generated TripAdvisor data provides insights into travelers' perceptions and behaviors, offering a valuable perspective on tourism evaluation trends [16]. Machine learning and natural language processing (NLP) techniques are commonly used to extract positive, negative, or neutral sentiments from review texts, thereby generating time series representing tourists' collective sentiment dynamics.

In tourism forecasting, sentiment data can serve as exogenous variables that allow deep learning models to capture fluctuations in traveler expectations and behavior. Recent studies [17] have shown that integrating sentiment indices from TripAdvisor with online search data (e.g., Google Trends) significantly enhances forecasting accuracy, particularly during volatile periods such as the COVID-19 pandemic. Consequently, incorporating sentiment information into neural or hybrid forecasting models holds great potential for improving tourism demand predictions.

In addition to hybrid forecasting studies, several Genetic Algorithm-based MLP (GA–MLP) studies have applied genetic algorithms to optimize internal network parameters (weights and biases) during MLP training [18] [19]. However, these GA-based approaches focus on improving the learning process, not on configuring the model itself. In contrast, the present study uses NSGA-II to optimize hyperparameters, including input window size, learning rate, epochs, early stopping, and data normalization. This shifts the role of the evolutionary algorithm from parameter training to model configuration optimization, enabling automatic hyperparameter tuning and reducing manual trial-and-error effort.

Table 1. Summarizes the methodological differences compared to earlier works

Study group / Reference	**Data sources**	**Methodology /** Architecture	**Optimization target**	**Limitation / Gap**
Google Trends for forecasting tourism demand [13] [14] [15]	Google Trends (search frequency)	Time series models (Naïve, SARMA, SARMAX, ANN, Mixed-frequency models)	None (manual tuning)	Only search data; no sentiment integration; models not optimized automatically
Sentiment-based forecasting using TripAdvisor / online reviews ([16], [17])	User-generated reviews: Sentiment	Machine learning / deep learning forecasting	None (manual feature usage)	Only sentiment data; no keyword selection; no optimization of forecasting model
GA–MLP optimization studies ([18] [19])	Single-source historical time series	GA–MLP neural network	GA optimizes weights/biases	GA improves training but does not optimize hyperparameters; limited generalization
Proposed study (NSGA-II–MLP)	Multisource: Historical data + Google Trends + TripAdvisor sentiment	MLP + NSGA-II (single-objective optimization)	Hyperparameters (window size, epochs, learning rate, early stopping, normalization)	Automates hyperparameter search, improved generalization, reduced manual trial-and-error

In summary, both Google Trends and TripAdvisor sentiment analysis contribute meaningfully to the performance of hybrid forecasting models. However, challenges remain in identifying the most relevant Google Trends data, accurately analyzing sentiment from TripAdvisor reviews, and effectively integrating these external data with historical time series. Table 1 summarizes the methodological differences between prior studies and the current work, highlighting the novelty of applying NSGA-II for automated hyperparameter optimization on multisource data. The next section presents detailed procedures for collecting and processing sentiment and search trend data, the preprocessing pipeline, the MLP-based model architecture, and the evaluation of each data source's contribution to monthly forecasts of international tourist arrivals to Vietnam.

3 Integrated NSGA-II–MLP model for multisource data forecasting

3.1 Data collection

The dataset used in this study consists of three data sources covering the same time period from January 2008 to October 2022. Monthly international tourist arrivals were collected for a total of 178 observations, with no interpolation applied to missing or irregular values to preserve the natural volatility in the data, collected from the Viet Nam National Authority of Tourism [20]. TripAdvisor review data in the same period include 62,823 user reviews, which were preprocessed using sentiment analysis. For each month, a sentiment index was computed as the ratio of positive reviews over the sum of positive and negative reviews, and the resulting values were aggregated into a monthly time series. Google Trends data were also collected monthly over the same period (178 samples). Ten candidate tourism-related keywords were initially generated using Google Suggest, and their corresponding Google Trends time series were extracted. These keywords were then filtered based on their Pearson correlation with the tourist arrival series. Google Trends values were computed as the average over a consecutive 30-day window to smooth fluctuations and ensure consistency.

- Sentiment from hotel reviews: A publicly available dataset from Zenodo [21] was utilized, containing reviews written by international visitors to Vietnam during the same period. Sentiment classification was conducted using a pre-trained DistilBERT model [22], which categorized each review into three labels: positive, neutral, and negative. The number of positive and negative reviews was aggregated on a monthly basis to form a sentiment index time series.

- Google Trends data (www.google.com/trends): Representative search keywords were selected through a three-step process: (i) analysis of search suggestions using Google Suggest, (ii) interviews with international tourists, and(iii) consultation with tourism experts. After calculating the Pearson correlation coefficient between search volumes and international tourist arrivals, the keyword “visit to Vietnam” was selected for exhibiting the highest correlation. The monthly search index of this keyword for 2008–2022 was used as an exogenous variable in the forecasting model.

Table 2. Description of the collected data

Month	Tourist arrivals	Positive sentiment	Search volume
2008-01	399.556	2	20
2008-02	411.032	1	15
2008-03	414.332	8	13
…	…	…	…
2022-10	1.371.135	174	85

Table 2 illustrates the structure of the synchronized dataset, which consists of three main variables: international tourist arrivals, positive sentiment index, and Google search index. For model development and evaluation, the data were split chronologically into three subsets: Training set: January 2008 – June 2018; Validation set: July 2018 – June 2019; Test set: July 2019 – October 2022. This temporal partitioning ensures the realistic setting of training and forecasting, preserving the sequential nature of the time series.

3.2 Data Normalization

Before being input into the MLP model, the three variables, namely Arrival (monthly number of international tourists), Sentiment (number of positive TripAdvisor reviews), and Google (Google Trends index), were normalized to the range [0, 1] using the Min-Max scaling [23]. This normalization ensures comparability among variables of different measurement scales and prevents any single feature, such as tourist arrivals with relatively large magnitudes, from unduly dominating the training process. The procedure is particularly essential for neural network models, which lack an inherent mechanism to balance feature scales during weight optimization.

Fig. 1 illustrates the data after normalization. The Arrival series exhibits a stable upward trend before the COVID-19 pandemic, a sharp decline during 2020–2021, and a gradual recovery afterward. Meanwhile, the Sentiment series reflects the rebound of tourists' positive attitudes following periods of disruption, and the Google Trends index captures fluctuations in monthly search activity before and during the pandemic. The normalization process successfully preserves both the trend patterns and fluctuation characteristics of each data source, enabling the model to learn more effectively while enhancing the visual interpretability of multi-source temporal variations.

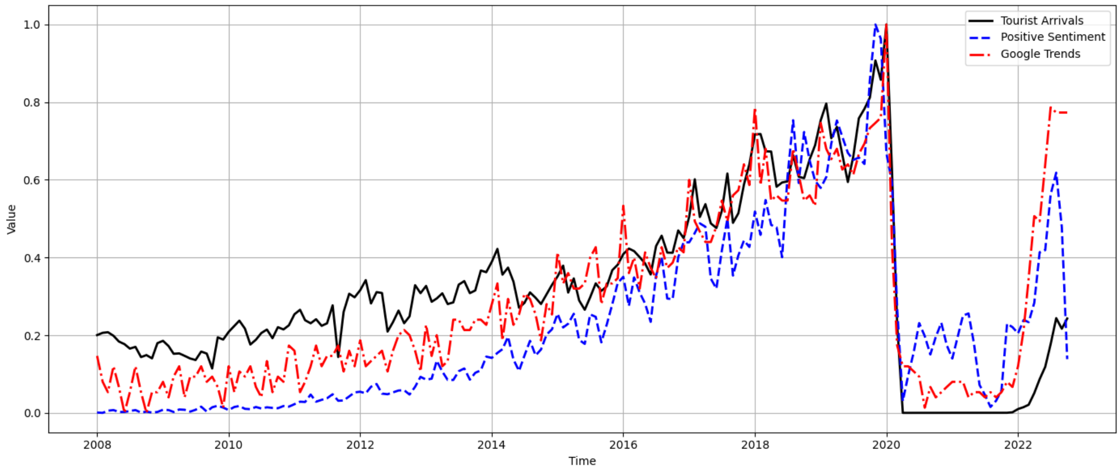

Fig. 1. Normalized multi-source time series data

3.3 Building the integrated NSGA-II-MLP model for multi-Source forecasting

The MLP network is a common type of artificial neural network architecture that consists of neurons organized into layers, including an input layer, one or more hidden layers, and an output layer. Information in an MLP network is transmitted unidirectionally from the input layer to the output layer through hidden layers, without any feedback connections. Therefore, the MLP is classified as a feedforward neural network. Fig. 2 illustrates the process of feeding input data into an MLP model to predict future tourist arrivals $\hat{a}_{t+1}$. Specifically, the tourist arrivals $a$represent historical data on the number of visitors over the previous $n-1$ time steps $(t, t-1, \dots, t-n+1)$. The time-series sentiment data $s$correspond to the number of positive sentiments expressed by tourists toward hotels in Vietnam on TripAdvisor, while $g$denotes the tourism-related search volume from Google Trends at time steps $t, t-1, \dots, t-n+1$.

The learning process of the MLP involves adjusting the connection weights between neurons to minimize the error between predicted outputs and target values. The most common algorithm used is backpropagation combined with gradient descent optimization. The error is propagated backward from the output layer to the input layer to update the weights along the direction of the negative gradient of the loss function. Due to its ability to learn nonlinear relationships, strong generalization capability, and high flexibility, the MLP has been extensively applied in time series forecasting, particularly in complex domains such as tourism demand prediction, which is influenced by multiple exogenous factors.

The total number of input features in the model is $2 \times n$, corresponding to the temporal features from two different data sources. Integrating these sources leverages the unique advantages of each: tourist arrival data (Arrival) capture long-term trends and recurring behavioral patterns, while sentiment data over time (Sentiment) reflect tourists' immediate perceptions and expectations, potentially serving as an early indicator of changes in tourism behavior. By jointly incorporating both sources into the model, the MLP can learn complex nonlinear relationships between variables, thereby improving forecasting accuracy compared to models relying on a single data source.

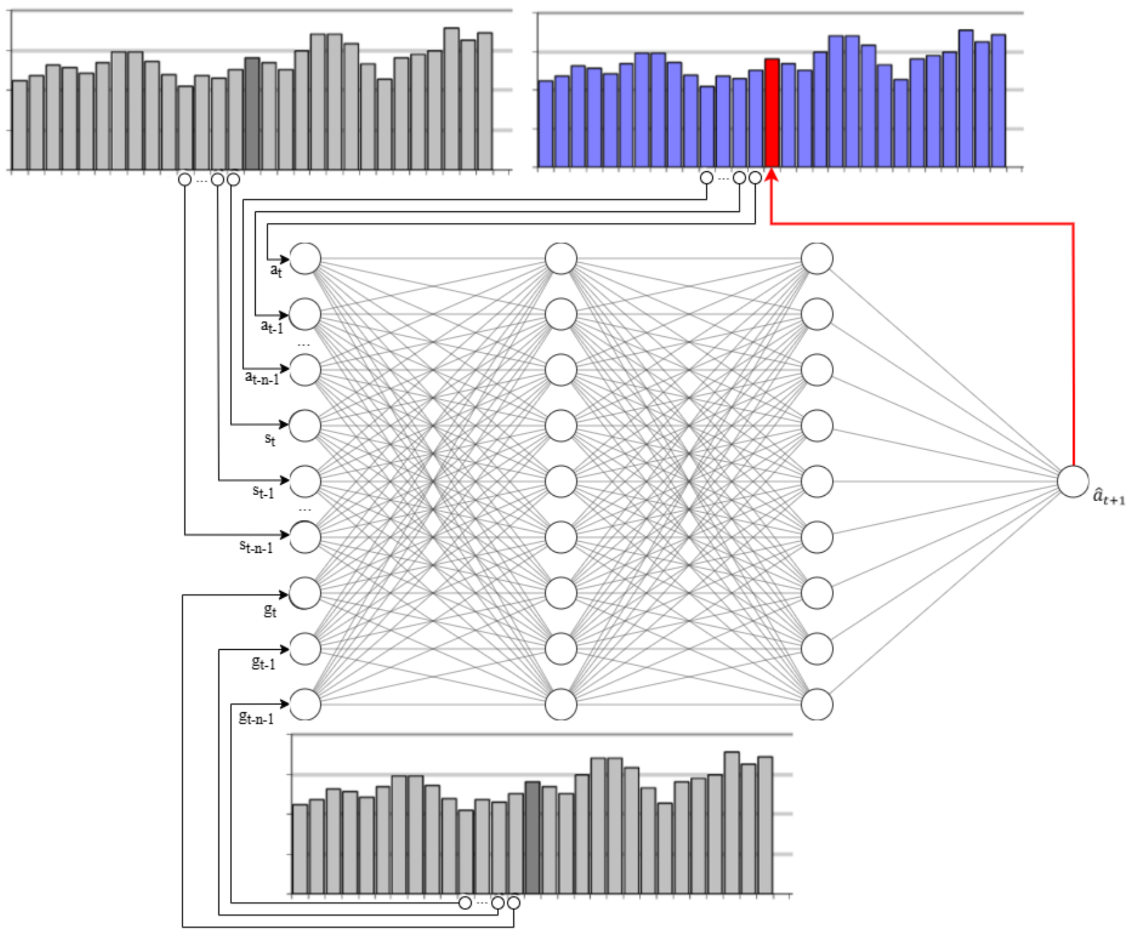

Fig. 2. MLP model for multi-source data forecasting

Building upon the MLP architecture, this study integrates the Non-dominated Sorting Genetic Algorithm II (NSGA-II) [3] to optimize the hyperparameters of the MLP model. In comparison with the traditional Genetic Algorithm (GA), NSGA-II incorporates two key techniques, namely non-dominated sorting and crowding distance, to improve optimization performance, particularly in multi-objective problems. In this study, NSGA-II is employed for single-objective optimization with the aim of minimizing the Symmetric Mean Absolute Percentage Error (sMAPE).

In the non-dominated sorting process, individuals in the population are ranked according to their dominance level. Dominant individuals are assigned higher ranks, while those that cannot be compared directly are treated as equally good and grouped into the same Pareto front.

Individuals on the same Pareto front form a Pareto curve, where ranking becomes less straightforward. The crowding distance technique is introduced to address this issue by estimating the density of individuals in the population: individuals located farther apart have larger crowding distances, indicating better diversity. This approach allows the algorithm to maintain both convergence and diversity in the solution space. The workflow of the proposed NSGA-II-MLP algorithm is illustrated in Fig. 3.

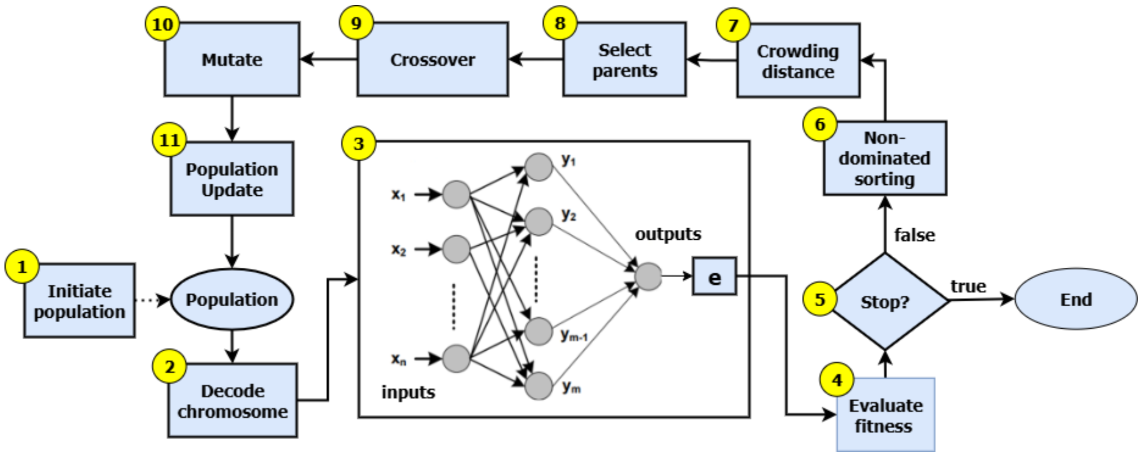

Fig. 3. Integrated NSGA-II-MLP model with hyperparameter optimization

The implementation procedure of the integrated NSGA-II-MLP model is as follows:

Step 1 (Population Initialization): An initial population of chromosomes is generated, where each chromosome represents a possible combination of MLP hyperparameters (e.g., number of hidden layers, neurons, learning rate, batch size, and activation functions). The gene values are initialized randomly within their predefined ranges.

Step 2 (Chromosome Decoding): Each chromosome is decoded to obtain its corresponding hyperparameter configuration, which is then used to construct an MLP model.

Step 3 (MLP Execution): Each MLP is trained using the prepared training dataset. The model produces forecasting results, from which the forecasting error (e.g., sMAPE, RMSE) or accuracy is recorded for evaluation.

Step 4 (Fitness Evaluation): The fitness of each chromosome is computed according to one or more objective functions. Typical objectives include minimizing forecasting error and minimizing model complexity or computation time.

Step 5 (Termination Check): A termination criterion is evaluated to decide whether the algorithm should stop or continue. Common criteria are reaching the maximum number of generations or achieving a target forecasting accuracy (or error threshold).

Step 6 (Non-dominated Sorting): All chromosomes in the current population are ranked into different Pareto fronts based on the concept of dominance. Chromosomes belonging to the first front are not dominated by any other individuals, while those in subsequent fronts are dominated by one or more individuals from the preceding fronts. This procedure assigns a Pareto rank to each chromosome, thereby ensuring elitism and promoting convergence toward the true Pareto-optimal set.

Step 7 (Crowding Distance Calculation): For each Pareto front, the crowding distance of every chromosome is calculated by measuring the average distance of its neighboring solutions in objective space. This metric estimates population density and is later used to maintain diversity—solutions in less crowded regions are preferred.

Step 8 (Parent Selection): Parent chromosomes are selected for reproduction based on Pareto rank and crowding distance. A binary tournament selection strategy is applied: the chromosome with a lower rank is preferred; if both have the same rank, the one with a higher crowding distance is chosen. This mechanism balances convergence and diversity.

Step 9 (Crossover): Pairs of selected parents undergo crossover to generate offspring. This study adopts the **uniform crossover** operator: for each gene, the offspring inherits the value from either parent with a high probability (0.9). This allows the combination of beneficial traits from both parents while maintaining genetic diversity.

Step 10 (Mutation): A **polynomial mutation** operator introduces small perturbations to the offspring's genes. Each hyperparameter has a low probability of mutation, ensuring exploration of the search space while keeping new values within feasible bounds.

Step 11 (Population Update): The parent and offspring populations are combined. Non-dominated sorting and crowding-distance assignment are applied again, and the best Nchromosomes (based on rank and diversity) are selected to form the next generation. This elitist replacement ensures that high-performing and diverse solutions are preserved.

Step 12 (Iteration): The process from Step 2 to Step 11 is repeated until the termination criterion is satisfied. The final non-dominated set represents the Pareto-optimal MLP configurations, providing trade-offs between predictive performance and model complexity.

The parameter settings of the NSGA-II algorithm followed the standard configuration proposed by Deb et al. (2002) [3]. Specifically, a population size of 100 and a maximum of 200 generations were used. The simulated binary crossover operator was applied with a crossover probability of 0.9, while the polynomial mutation operator employed a mutation probability of $1/n$, where $n = 5$ corresponds to the number of hyperparameters optimized by NSGA-II (input window size, learning rate, number of epochs, early stopping, and normalization method). This yields $p_m = 0.2$, ensuring that, on average, one gene per chromosome is mutated in each generation. These values are consistent with the original NSGA-II configuration and maintain a good balance between convergence stability and population diversity.

4 Experiments and Results

4.1 Data processing

The dataset consists of the monthly number of international tourist arrivals to Vietnam from January 2008 to October 2022 (Figure 1). It is divided into three subsets: the training set, containing data from January 2008 to June 2018, is used to learn the temporal patterns of the time series; the validation set, comprising the following 12 months from July 2018 to June 2019, is employed to monitor the training process, tune hyperparameters, and apply early stopping; and the test set, covering the period from July 2019 to October 2022, is used to evaluate the forecasting performance of the model under real-world conditions. As also illustrated in Figure 1, two types of exogenous variables are incorporated to enhance forecasting performance. The first, Sentiment, represents the historical level of positive emotions extracted from online reviews and is used as an input feature alongside the target variable's past values. This variable is included in all three subsets: training, validation, and test.

4.2 Comparison of forecasting models before and after integrating COVID-19 data

In this study, the key hyperparameters of the MLP model were optimized using the NSGA-II algorithm to simultaneously minimize three error metrics: sMAPE, MAE, and RMSE.

$$sMAPE = \frac{100\%}{n}\sum_{t=1}^{n}\frac{|F_t - A_t|}{\frac{|A_t|+|F_t|}{2}} \tag{1}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|F_t - A_t| \tag{2}$$

$$\mathrm{RMS}E = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(F_t - A_t)^2} \tag{3}$$

where n is the number of observations in the test sets, A_t and F_t are the actual and predicted values at step t, respectively. According to the above formulas, the smaller the MAPE, sMAPE, MAE, and RMSE, the better the model's forecasting performance.

Unlike the traditional GA, NSGA-II enables the search for a Pareto-optimal solution set, thereby allowing the model to achieve a balance between forecasting accuracy and stability. The selected hyperparameters include:

- *input_size:* the number of past time steps (look-back window) used for forecasting future values, determining the temporal context length the model learns from;

• *learning_rate:* the learning speed, directly affecting the convergence behavior and the model's ability to avoid overfitting or underfitting;

• *max_steps:* the maximum number of training iterations;

• *early_stop:* an early stopping mechanism that prevents overtraining and conserves computational resources;

• *scaler_type:* the normalization method applied to the input data to ensure consistent feature distributions and mitigate potential bias during training.

The search space for each hyperparameter was defined based on preliminary experiments with the MLP model (Table 3). Through this optimization process, NSGA-II enables the MLP model to flexibly adapt to varying data conditions, particularly between stable periods (pre-COVID) and highly volatile periods (during COVID).

Table 3. Optimization settings of the NSGA-II-MLP model with A: Arrival; AS: Arrival + Sentiment; AG: Arrival + Google; G: Google; ASG: Arrival + Sentiment + Google

Hyper-parameter	Search Space	Pre-COVID-19				During COVID-19			
		A	AS	AG	ASG	A	AR	AG	ASG
input_size	[12, 36]	12	12	13	13	12	19	12	35
learning_rate	[1e-4, 1e-2]	0.004156	0.004703	0.000374	0.06541	0.004	0.0007	0.003212	0.008182
max_steps	[100, 500]	363	488	367	164	363	231	281	115
early_stop	[2, 10]	8	9	10	5	8	8	4	7
scaler_type	['identity', 'robust', 'standard']	standard	robust	standard	robust	standard	standard	robust	standard

After performing the optimization and obtaining the best set of hyperparameters, we compared the integrated NSGA-II-MLP model with the baseline MLP model, as shown in Table 4, across two periods: before COVID-19 (Pre-COVID-19) and during COVID-19 (During COVID-19), using different combinations of input variables. The results indicate that the NSGA-II-MLP model outperforms the MLP model in both periods, as evidenced by significantly lower sMAPE, MAE, and RMSE values. This demonstrates the effectiveness of hyperparameter optimization using the NSGA-II algorithm, which enhances the learning and generalization ability of the MLP network. When analyzing the impact of exogenous variables, a clear pattern emerges between the two periods:

- Pre-COVID-19 period: The inclusion of sentiment data and Google search trends helps reduce forecasting errors, especially when only one exogenous source is added (Arrivals + Sentiment or Arrivals + Google). However, when both sources are combined (Arrivals +

Sentiment + Google), the error slightly increases, possibly due to overlapping information or noise between the two sources during this relatively stable period.

- During-COVID-19 period: In contrast, under highly volatile conditions, combining both exogenous variables enables the model to better capture behavioral and psychological signals from travelers, thereby improving forecasting performance compared to using only one additional variable. This highlights the crucial role of sentiment and online search data in reflecting sudden shifts in tourism demand during the pandemic.

Table 4. Comparison results between the NSGA-II-MLP and MLP models

Input	Period	NSGA-II-MLP			MLP		
		sMAPE	MAE	RMSE	sMAPE	MAE	RMSE
Arrival + Sentiment + Google	During Covid-19	127.10	0.26	0.55	182.30	5.55	8.36
Arrival + Sentiment	During Covid-19	112.91	0.21	0.38	176.26	0.64	0.78
Arrival + Google	During Covid-19	103.70	0.34	1.13	147.82	2.42	4.51
Arrival	During Covid-19	138.10	0.17	0.24	173.64	0.34	0.40
Arrival + Sentiment + Google	Pre-Covid-19	8.18	0.06	0.07	17.31	0.11	0.13
Arrival + Sentiment	Pre-Covid-19	6.60	0.05	0.07	12.44	0.09	0.10
Arrival + Google	Pre-Covid-19	7.57	0.06	0.08	15.28	0.10	0.12
Arrival	Pre-Covid-19	8.80	0.06	0.08	10.44	0.07	0.09

A noteworthy observation is that the incorporation of exogenous variables derived from sentiment data substantially enhances the forecasting performance of the NSGA-II-MLP model across both examined periods. During the COVID-19 period, the sMAPE decreased from 138.094 (using only Arrivals) to 112.912 (Arrivals + Sentiment), indicating that the model effectively captures additional informational cues related to tourists' emotional dynamics. In the pre-COVID-19 period, where the time series exhibited greater stability, both models yielded satisfactory results; however, the NSGA-II-MLP consistently outperformed the baseline MLP, achieving an sMAPE of 6.596 compared to 12.44 when incorporating sentiment data. Even in the absence of exogenous variables, the NSGA-II-MLP (sMAPE = 8.797) still surpassed the MLP (sMAPE = 10.441). These findings underscore the robustness and generalization capability of the NSGA-II-MLP framework, highlighting the effectiveness of single-objective hyperparameter optimization in improving predictive accuracy, even under relatively stable conditions.

The results illustrated in Fig. 4 (pre-COVID-19 period) indicate that the NSGA-II-MLP model demonstrates a superior ability to replicate the fluctuations in tourist arrivals compared to the baseline MLP model. The forecasted trajectories of NSGA-II-MLP (blue dashed lines) closely follow the actual observations (black line) across most input data configurations. The sMAPE values for NSGA-II-MLP range from 6.6% to 8.8%, significantly lower than those of MLP (10.4%

to 17.3%), highlighting the substantial benefits of hyperparameter optimization through the NSGA-II algorithm.

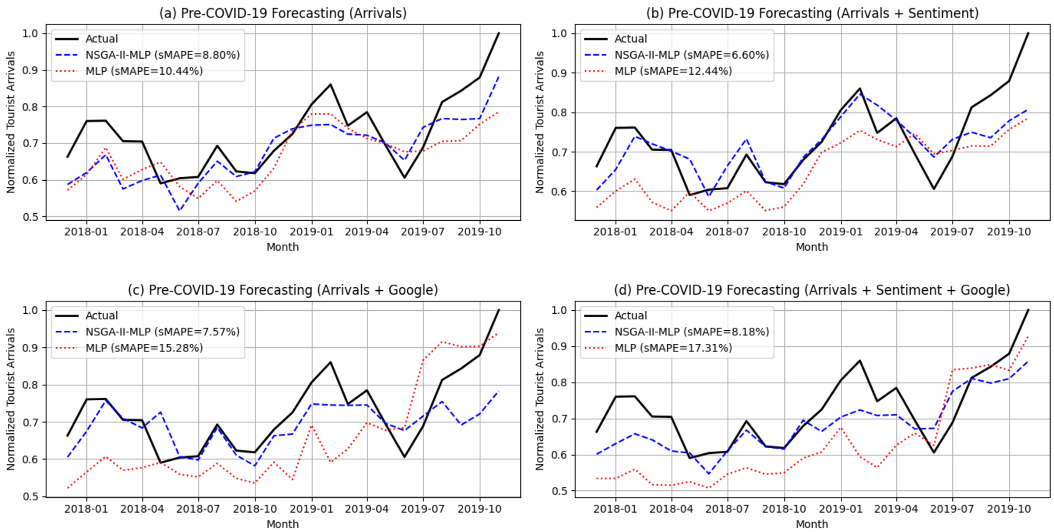

Fig. 4. Twelve-month demand forecasting results of NSGA-II-MLP and MLP in the pre-Covid-19 period

However, when the input variables are expanded from Arrivals to Arrivals + Sentiment + Google Trends, both models exhibit a slight increase in sMAPE. This suggests that, during the pre-pandemic period, exogenous factors such as online sentiment and search trends provided limited additional predictive power. The relatively stable tourism environment likely made historical arrival data sufficiently informative to capture the overall trend, thereby reducing the marginal benefit of incorporating external variables.

The results illustrated in Fig. 5 (During-COVID-19 period) reveal substantial changes in the forecasting patterns. Under the influence of economic shocks and travel restrictions, both models struggled to capture abrupt fluctuations, leading to high sMAPE values (up to 127.1% for NSGA-II-MLP and 182.3% for MLP). However, the NSGA-II-MLP model demonstrated greater stability and adaptability, as its forecast curve exhibited fewer oscillations and more reasonable responses to the post-pandemic recovery trend. The inclusion of Sentiment and Google Trends variables enhanced the model's sensitivity to changes in travelers' attitudes and search behaviors, despite the increased noise inherent in crisis-period data.

These findings suggest that exogenous variables play a more critical role in highly volatile contexts such as the pandemic, where historical data alone fail to capture future trends accurately. This result reinforces the rationale for integrating sentiment and online search data into modern tourism forecasting models.

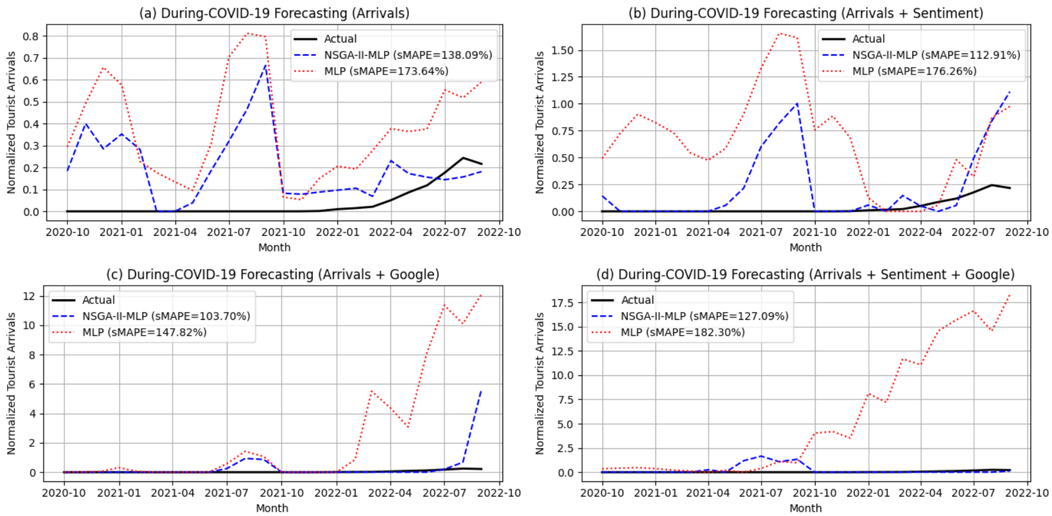

Fig. 5. Twelve-month demand forecasting results of NSGA-II-MLP and MLP during the Covid-19 period

5 Conclusion

This study analyzed the performance of the NSGA-II-MLP model for time series forecasting of monthly international tourist arrivals to Vietnam, incorporating additional exogenous variables such as positive sentiment indices extracted from TripAdvisor and search trend data from Google Trends. The model was evaluated across two real-world periods, namely the pre-COVID-19 and during-COVID-19 phases, using a fixed forecasting horizon of 12 months (h = 12). Experimental results demonstrate that hyperparameter optimization using the NSGA-II genetic algorithm significantly improves forecasting accuracy compared to the conventional MLP model. Specifically, NSGA-II-MLP consistently achieved lower error metrics (sMAPE, MAE, RMSE), while the inclusion of exogenous variables as covariates further reduced errors in highly volatile conditions. However, during the COVID-19 period, both models exhibited elevatedsMAPE values (over 100%), indicating that although NSGA-II enhances stability, the inherently standard MLP still struggles with unexpected shocks and nonlinear shifts in travel behavior. Consequently, future research should focus on advanced deep learning architectures such as TSMixer [5] or the Temporal Fusion Transformer, which can capture nonlinear relationships, long-term dependencies, and complex cross-source interactions to improve forecasting performance under high uncertainty and dynamic conditions.

Acknowledgment

The authors thank the School of Hospitality and Tourism – Hue University, Vietnam, for funding this project (Code DHH2025-10-45).

References

1. A. Fronzetti Colladon, B. Guardabascio, and R. Innarella, "Using social network and semantic analysis to analyze online travel forums and forecast tourism demand," *Decis Support Syst*, vol. 123, p. 113075, Aug. 2019, doi: 10.1016/J.DSS.2019.113075.

2. W. Xiong, M. Huang, B. Okumus, S. Chen, and F. Fan, "The predictive role of tourist-generated content on travel intentions: emotional mechanisms as mediators," *Asia Pacific Journal of Tourism Research*, vol. 27, no. 5, pp. 443–456, May 2022, doi: 10.1080/10941665.2022.2075778.

3. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, Apr. 2002, doi: 10.1109/4235.996017.

4. B. Lim, S. Arık, N. Loeff, and T. Pfister, "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting," *Int J Forecast*, vol. 37, no. 4, pp. 1748–1764, Oct. 2021, doi: 10.1016/J.IJFORECAST.2021.03.012.

5. S.-A. Chen, C.-L. Li, S. O. Arik, N. C. Yoder, and T. Pfister, "{TSM}ixer: An All-{MLP} Architecture for Time Series Forecast-ing," *Transactions on Machine Learning Research*, 2023, [Online]. Available: https://openreview.net/forum?id=wbpxTuXgm0

6. C. Lim and M. McAleer, "Forecasting tourist arrivals," *Ann Tour Res*, vol. 28, no. 4, pp. 965–977, 2001, doi: 10.1016/S0160-7383(01)00006-8.

7. V. Cho, "A comparison of three different approaches to tourist arrival forecasting," *Tour Manag*, vol. 24, no. 3, pp. 323–330, 2003, doi: 10.1016/S0261-5177(02)00068-7.

8. H. Hassani, E. S. Silva, N. Antonakakis, G. Filis, and R. Gupta, "Forecasting accuracy evaluation of tourist arrivals," *Ann Tour Res*, vol. 63, pp. 112–127, 2017, doi: 10.1016/j.annals.2017.01.008.

9. M. E. Nor, A. I. M. Nurul, and M. S. Rusiman, "A Hybrid Approach on Tourism Demand Forecasting," *J Phys Conf Ser*, vol. 995, no. 1, pp. 1–11, Apr. 2018, doi: 10.1088/1742-6596/995/1/012034.

10. M. L. Shen, H. H. Liu, Y. H. Lien, C. F. Lee, and C. H. Yang, "Hybrid approach for forecasting tourist arrivals," *ACM International Conference Proceeding Series*, vol. Part F1479, pp. 392–396, 2019, doi: 10.1145/3316615.3316628.

11. Purwanto, Sunardi, F. T. Julfia, and A. Paramananda, "Hybrid model of ARIMA-linear trend model for tourist arrivals prediction model in Surakarta City, Indonesia," *AIP Conf Proc*, vol. 2114, no. June, 2019, doi: 10.1063/1.5112481.

12. Y. Carrière-Swallow and F. Labbé, "Nowcasting with Google trends in an emerging market," *J Forecast*, vol. 32, no. 4, pp. 289–298, 2013, doi: 10.1002/for.1252.

13. K. Volchek, A. Liu, H. Song, and D. Buhalis, "Forecasting tourist arrivals at attractions: Search engine empowered methodologies," *Tourism Economics*, vol. 25, no. 3, pp. 425–447, 2019, doi: 10.1177/1354816618811558.

14. K. H. Huarng and T. H. K. Yu, "Application of Google trends to forecast tourism demand," *Journal of Internet Technology*, vol. 20, no. 4, pp. 1273–1280, 2019, doi: 10.3966/160792642019072004025.

15. T. Havranek and A. Zeynalov, "Forecasting tourist arrivals: Google Trends meets mixed-frequency data," *Tourism Economics*, vol. 27, no. 1, pp. 129–148, 2021, doi: 10.1177/1354816619879584.

16. M. Mariani and M. Borghi, “Customers’ evaluation of mechanical artificial intelligence in hospitality services: a study using online reviews analytics,” *International Journal of Contemporary Hospitality Management*, vol. 33, no. 11, pp. 3956–3976, Nov. 2021, doi: 10.1108/IJCHM-06-2020-0622.

17. D. C. Wu, S. Zhong, R. T. R. Qiu, and J. Wu, “Are customer reviews just reviews? Hotel forecasting using sentiment analysis,” *Tourism Economics*, vol. 28, no. 3, pp. 795–816, May 2022, doi: 10.1177/13548166211049865.

18. R. V. Savangouder and J. C. Patra, “GA-MLP-based Inverse Modeling Technique for Prediction of Process Parameters and Cost Optimization,” *Conf Proc IEEE Int Conf Syst Man Cybern*, pp. 510–515, 2021, doi: 10.1109/SMC52423.2021.9658929.

19. M. J. Er and F. Liu, “Genetic algorithms for MLP neural network parameters optimization,” *2009 Chinese Control and Decision Conference, CCDC 2009*, pp. 3653–3658, 2009, doi: 10.1109/CCDC.2009.5192353.

20. “Viet Nam National Authority of Tourism.” Accessed: Nov. 07, 2025. [Online]. Available: https://vietnamtourism.gov.vn/en/statistic/international

21. A. D. Van, T. T. T. Kieu, H. T. N. Ngoc, A. N. T. Linh, and T. H. N. Thanh, “TripAdvisor Vietnam Hotel Reviews,” May 2023, *Zenodo*. doi: 10.5281/zenodo.7967494.

22. “lxyuan/distilbert-base-multilingual-cased-sentiments-student · Hugging Face.” Accessed: Jun. 25, 2025. [Online]. Available: https://huggingface.co/lxyuan/distilbert-base-multilingual-cased-sentiments-student

23. A. Géron, “Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow (2019, O’reilly),” *Hands-On Machine Learning with R*, p. 510, 2017.