ISSN (PRINT): 1994 - 7658 ISSN (ONLINE): 1314 - 0817

VARNA UNIVERSITY OF MANAGEMENT

European Journal of Tourism Research

http://ejtr.vumk.eu

Combining the power of artificial intelligence and mathematical modelling: A hybrid technique for enhanced forecast of tourism receipts

Ferhat Şeker [1*]

[1] Department of Tourism Management, Faculty of Business, Adana Alparslan Türkeş Science and Technology University, Adana, Türkiye. Email: fseker@atu.edu.tr

*Corresponding author

Abstract

Despite being one of the most visited countries in the world, Türkiye's share of tourism revenue does not rank among the top ten. Therefore, it would be worth researching tourist expenditures and analysing this data could provide valuable insights. This research develops a novel approach to estimating and modelling tourism receipts by analysing expenditure types. Artificial intelligence-based methods, such as machine learning, have been increasingly used in the tourism literature to improve various aspects of the industry. However, little research has been conducted using a hybrid method to model and estimate tourist expenditure. This paper is the first to combine conventional mathematical analysis, specifically first-order two-variable polynomial equations, with artificial intelligence-based machine learning algorithms in a tourism setting. The research results indicate that expenditure types such as accommodation and food & beverage significantly impact Türkiye's tourism revenue and Türkiye's total tourism revenue will not exceed 45 billion dollars by 2027. This study provides a valuable and practical contribution to improving the accuracy and efficiency of methods for managing tourism economics, particularly in European countries where the economy heavily relies on income generated by tourism. Additionally, it fills a gap in studies focused on tourists' expenditure types by combining artificial intelligence and traditional analysis, making it a unique piece of research.

Keywords: Artificial Intelligence, Machine Learning, Correlation-Based Feature Selection, Random Forest Algorithm, Tourist Expenditure Types.

Citation: Şeker, F. (2024). Combining the power of artificial intelligence and mathematical modelling: A hybrid technique for enhanced forecast of tourism receipts. *European Journal of Tourism Research*, 36, 3614. https://doi.org/10.54055/ejtr.v36i.3246.

© 2024 The Author(s)
This work is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0). To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

1. Introduction

When tourists visit a country, they typically spend money on things like food and beverage, accommodation, health, transportation, sports, education, and culture. This spending can provide a significant boost to the country's economy by creating jobs, supporting small businesses, and generating tax revenue (Fleischer and Felsenstein, 2000). Before Covid-19, travel and tourism had become one of the most important sectors in the world economy, accounting for 10% of global GDP, 7.4% of total exports, $1.466 billion in income and more than 320 million jobs worldwide (International Monetary Fund, 2020). In addition to economic benefits, tourist expenditures can also have a positive impact on other areas of development. For instance, tourism can help to preserve cultural heritage and promote environmental conservation (Wager, 1995). It can also contribute to the development of infrastructure such as transportation, hotels, medical services and other facilities (Zaei and Zaei, 2013). Overall, international tourism receipts, also known as tourism revenue, can be a crucial source of income for developing countries (Sadler and Archer, 1975) such as Türkiye.

Researching tourist expenditures is essential for various reasons (Sun *et al.*, 2022). To begin, it is beneficial to comprehend the economic impact of tourism on a specific destination or region (Pai *et al.*, 2014). Researchers can estimate the amount of revenue generated by tourism and its contribution to the local economy by analysing how much money tourists spend on things like accommodation, food and beverage, transportation, entertainment, or shopping. Second, examining tourist expenditures can provide insights into tourist behaviour and preferences (Wu *et al.*, 2012). Researchers, for instance, can determine which types of activities or attractions are most popular among tourists by analysing spending patterns, which can help destination marketers and tourism businesses tailor their offerings to better meet the needs and interests of their target market. Third, researching tourist expenditures can assist in identifying potential areas for tourism industry growth and development (Chen, 2011). Researchers can identify areas where investment is needed to improve infrastructure, services, or attractions in order to attract more tourists and increase revenue by analysing spending trends over time.

Prior to the Covid-19 pandemic, Türkiye was the sixth most visited destination in the world with 51.2 million visitors, according to the United Nations World Tourism Organization (UNWTO) in 2019. However, it did not rank among the top ten in terms of tourism receipts (UNWTO, 2020). This stands out as a significant issue as it may indicate that the country is not effectively monetizing its large number of tourists. This could lead to missed opportunities for economic growth and development. Therefore, it would be worth researching the tourist expenditures and analysing this data could provide valuable insights for policymakers, industry professionals and other stakeholders in Türkiye to improve the country's tourism industry and increase tourism revenue. Besides, accurately predicting tourist expenditures for the next years can be a vital marketing insight in developing and implementing successful strategies such as planning and budgeting, improving decision-making, identifying trends, and forecasting economic impact for the country. Hence, the research will provide a practical contribution to improving accurate and efficient approaches for tourism management in European countries such as Türkiye whose economy is dependent on tourism.

Recently, artificial intelligence-based methods such as machine learning have been increasingly used in the tourism literature to improve various aspects of the industry, such as customer service, marketing and pricing, and forecasting demand. Rice, Park, Pan, and Newman (2019) adopted machine learning to forecast campground demand in US national parks. Li *et al.* (2021) employed machine learning-based feature selection methods to forecast tourist arrivals in Beijing, China, along with hotel occupancy in the city of Charleston, South Carolina, USA. Bi, Han and Li (2022) examined international tourist arrivals in 10 European countries to forecast tourism demand with machine learning models. Although it is frequently used for demand forecasting, the application area of machine learning is quite wide. For

example, Sangkaew and Zhu (2022) tried to understand tourists' experiences at local markets in Phuket by analyzing TripAdvisor reviews using a machine learning-based algorithm. Similarly, Oh and Kim (2022) investigated emotions in online reviews of fine dining restaurants in Hong Kong and adopted semantic network analysis and a machine learning algorithm. In the field of customer service, Chen *et al.* (2021) proposed a customer purchase forecasting model for online tourism to obtain an accurate prediction result and the influence of variables as predictors. Even the machine learning method is used not only for quantitative data but also for the analysis of photos on social media. Yu and Egger (2021) used machine learning approaches to investigate the effectiveness of engagement rate regarding touristic Instagram pictures. However, most ongoing studies based on the machine learning technique do not focus on tourists' expenditure types. Considering the gap in the literature, this paper employs machine learning algorithms to gain a deeper understanding and predict inbound tourist expenditures.

The objective of this research is to develop a novel approach for estimating and modelling tourism receipts by analysing expenditure types using a hybrid method of conventional mathematical analysis and artificial intelligence-based machine learning algorithms. The study aims to gain a deeper understanding of tourist spending patterns and predict inbound tourist expenditures more accurately. By achieving this objective, the research intends to provide valuable insights for policymakers, industry professionals, and other stakeholders in countries heavily reliant on tourism income, such as Türkiye. The findings of this study can inform more targeted and effective policies and strategies for managing tourism, enhance decision-making processes, and contribute to the overall development and growth of the tourism industry. The paper contributes to the broader body of knowledge on research methods in tourism, particularly in data analysis and modelling.

2. Literature Review

2.1. Tourist Expenditure

Tourist expenditure refers to the money spent by travellers on various products and services while on vacation. This includes costs such as lodging, food and beverage, transportation, souvenirs, and entertainment. Tourist expenditure is analysed under the utility maximization theory. Based on this theory, tourism companies develop products that will maximize the utility (satisfaction) derived from tourists' time and money. They provide a package of benefits and experiences that tourists would not be able to assemble themselves at the same price. Maximizing the overall value for money impacts tourist expenditures. Understanding tourist expenditure is crucial for assessing the economic impact of tourism, analysing tourist behaviour and preferences, and developing effective strategies for destination management and marketing. Tourist spending can have a significant impact on a destination's economy, making it of interest to governments, businesses, and researchers (Alfarhan, Olya and Nusair, 2022). According to studies, tourists spend more money on trips to foreign destinations than on domestic trips, primarily due to the increased costs of transportation, accommodation, and other expenses in foreign destinations (Jiménez *et al.*, 2023; Su *et al.*, 2023; Ramos-Dominguez *et al.*, 2023; Mariani *et al.*, 2023). In addition, tourists tend to spend more in destinations that offer a high level of comfort, safety, and convenience (Brida and Scuderi, 2013). Research has also found that the length of stay is a significant determinant of tourist expenditure. Longer stays generally result in higher expenditure, as tourists have more time to explore the destination and engage in a wider range of activities. On the other hand, shorter stays may result in higher expenditure per day, as tourists try to maximize their experiences within a limited timeframe (Ramos and Murta, 2022).

Tourist expenditure patterns are also influenced by their motivations for travel. Tourists travelling for leisure, for example, tend to spend more on entertainment and recreation, whereas those travelling for business tend to spend less overall because they have less time for leisure activities (Erdem and Jiang, 2016). Finally, studies have also found that tourists' demographic characteristics, such as age, income,

and education, can influence their spending habits. For example, younger tourists tend to have a lower average expenditure than older tourists, while tourists with higher incomes tend to spend more on luxury goods and services (Jurdana and Frleta, 2017). In conclusion, the literature on tourist expenditure highlights the importance of various factors in determining the amount of money that tourists spend on their trips. However, the current research differs from the existing literature by focusing on tourists' expenditure types using artificial intelligence.

2.2. Artificial Intelligence in Tourism

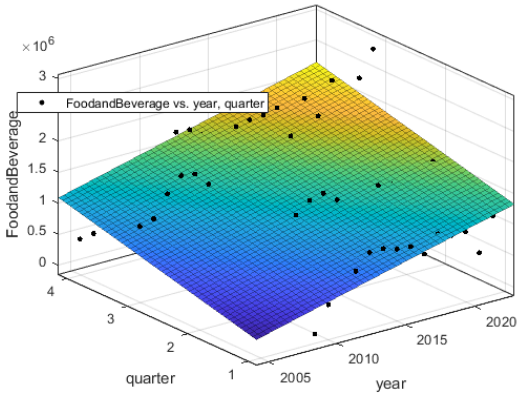
Tourism is one of the world's fastest-growing industries, and with technological advancements, Artificial Intelligence (AI) has become an essential aspect of the industry (Wong, Chau and Chan, 2023). Technology has the potential to transform the way in which tourism services are delivered, and it is being used to improve the customer experience, optimize operations, and personalize travel experiences (Ivanov and Webster, 2017). One study has shown that AI can be used to improve the customer experience by providing personal recommendations and real-time information to tourists (Zhu *et al.*, 2023). For example, AI-powered virtual assistants can help tourists find the best deals on flights, accommodations, and activities. In addition, AI can be used to analyse customer behaviour and preferences, allowing tourism companies to personalize the customer experience and improve customer satisfaction (Li *et al.*, 2023).

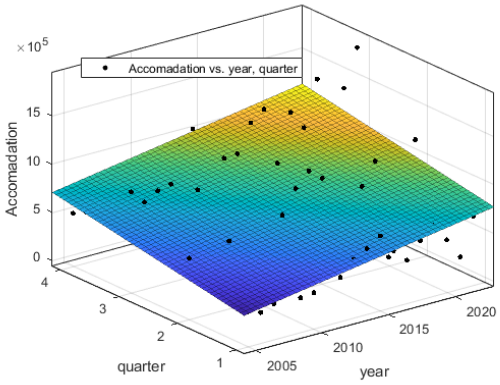
Another application of AI in the tourism industry is in operational optimization. AI can streamline processes, reduce costs, and improve efficiency in various areas, such as hotel operations, tour planning, and event management (Samara *et al.*, 2020). For example, AI can be used to optimize hotel room pricing, tour schedules, and event attendance, leading to increased revenue and improved customer experiences. Besides, AI is being used to personalize travel experiences for tourists. AI-powered systems can analyse customer preferences, travel history, and other factors to recommend personalized itineraries, activities, and destinations. This can help tourists have a more unique and enjoyable travel experience, and it can also help tourism companies increase customer loyalty and repeat business (Oday *et al.*, 2021). This study aims to contribute to the literature by modelling and estimating international tourism receipts in Türkiye according to expenditure types, using conventional mathematical analysis, namely first-order two-variable polynomial equation and artificial intelligence-based machine learning algorithms with a hybrid approach.

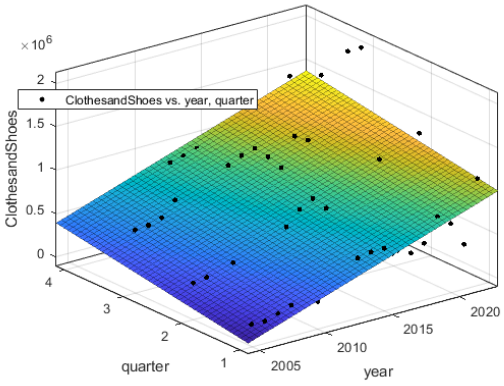
3. Methodology

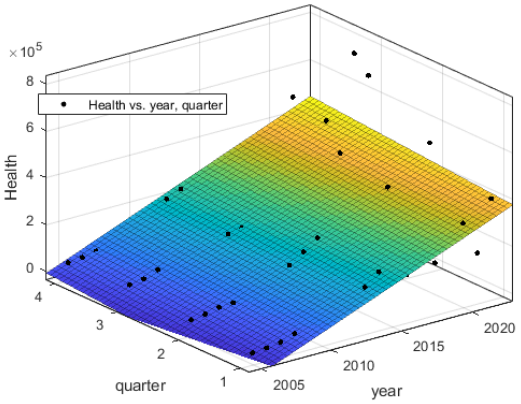
3.1. Types of Expenditures Dataset Description

The expenditures of tourists have been quarterly declared by the Turkish Statistical Institute (Türkiye İstatistik Kurumu, 2022) commonly known as TÜİK, with various types of expenditures used as parameters for these declarations. Specifically, Food-Beverage, Accommodation, Health, Transportation, Sports-Education-Culture, Tour, Clothes-Shoes, Souvenirs, and Other Expenditures are taken into consideration when determining the total revenue obtained by tourists. TÜİK is the official statistical institution in Türkiye responsible for collecting, analysing, and disseminating statistical data related to various aspects of the country's economy, society, and population. TÜİK plays a crucial role in providing reliable and comprehensive statistical information for research, policy-making, and decision-making processes. It conducts surveys, censuses, and other data collection activities to gather data from individuals, households, businesses, and other entities across Türkiye. Feature selection methods and machine learning (ML) algorithms are used to analyse and estimate tourism revenue. The proposed dataset consists of years, quarters, food and beverage, accommodation, health, transportation, sports, education, culture, tours, clothing and shoes, souvenirs, other expenditures, and the total incoming value of tourists. The detailed features with respect to the years and quarters are given in Figure 1, resulting in a dataset with 12 attributes (one of which is the class: total incoming value) and 75 instances.


×10^6
FoodandBeverage vs. year, quarter
FoodandBeverage
quarter
year


×10^5
Accomadation vs. year, quarter
Accomadation
quarter
year


×10^6
ClothesandShoes vs. year, quarter
ClothesandShoes
quarter
year


×10^5
Health vs. year, quarter
Health
quarter
year

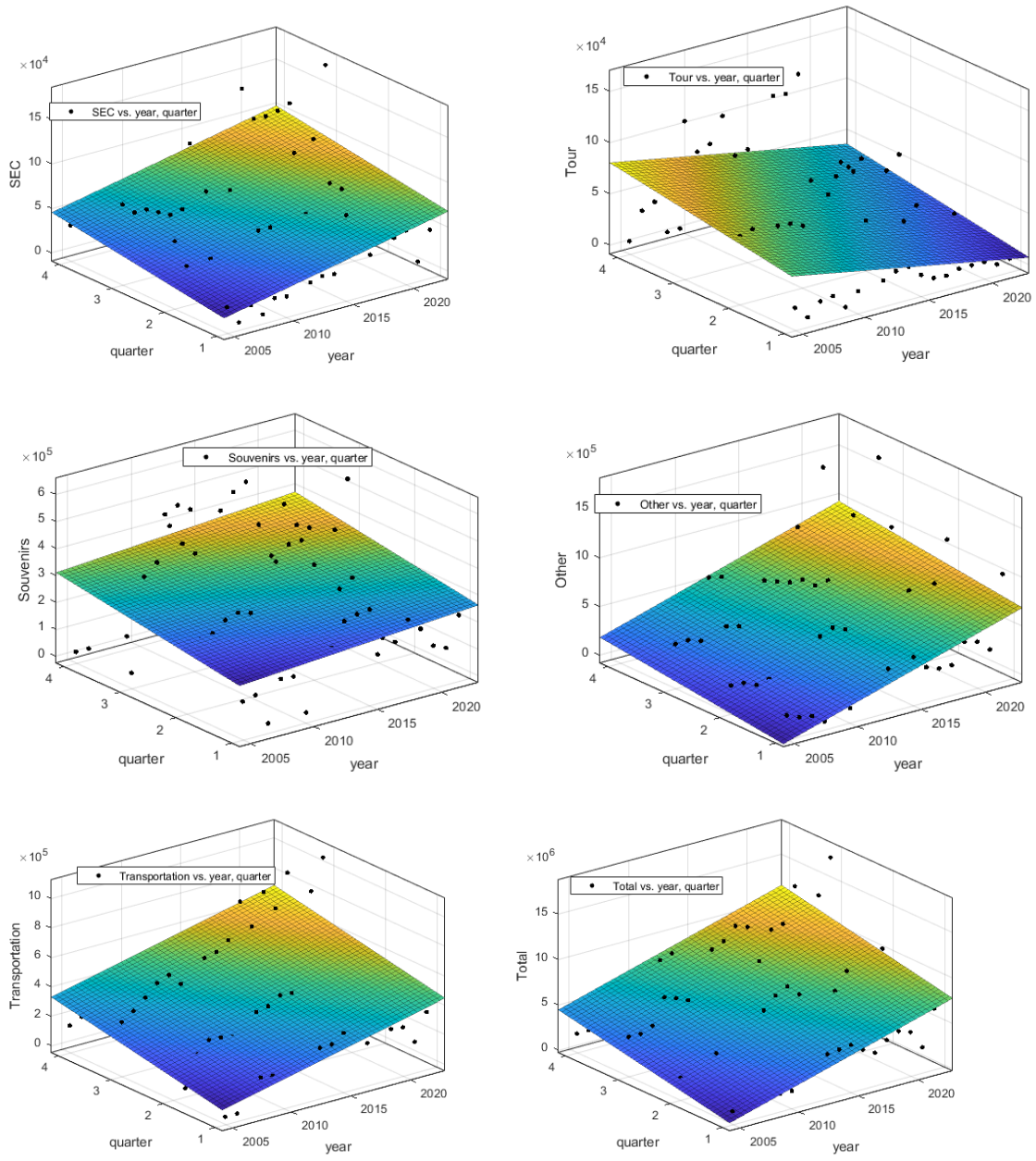


Figure 1. *The Detailed Analysis of the Expenditure Dataset*

The statistical analysis of each attribute is presented in Table 1.

3.2. Machine Learning Algorithms

There are three main types of Machine Learning (ML) algorithms: unsupervised learning, supervised learning, and reinforcement learning. Unsupervised learning is generally suitable for clustering and dimensionality reduction problems, while supervised learning is related to classification and regression analysis. Reinforcement learning, although less relevant for tourism cases, is another type of ML

algorithm. Additionally, natural language processing (NLP) is a specialized ML case, including algorithms for text classification, topic modelling, and sentiment analysis, among others.

Table 1. *Statistics for Each Feature*

Features	Data Type	Min.	Max.	Mean ± Std. Deviation	Unit
Year	Numeric	2005	2022	2013.522±5.281	-
Quarter	Numeric	1	4	2.478±1.119	-
Food and Beverage	Numeric	4728	2922386	1204162.224±660661.601	Thousand $
Accommodation	Numeric	40353	1860528	737577.448±365104.397	Thousand $
Health	Numeric	5794	796356	185256±158701.862	Thousand $
Transportation	Numeric	9638	1074396	388211.701±234447.204	Thousand $
Sports, Education and Culture	Numeric	13860	1772278	56335.91±38987.612	Thousand $
Tour	Numeric	3279	1624340	43588.209±42083.094	Thousand $
Clothes and Shoes	Numeric	3695	2023879	695450±434309.817	Thousand $
Souvenirs	Numeric	8053	630357	286194.149±155731.186	Thousand $
Other expenditures	Numeric	11484	1735492	436715.985±327940.77	Thousand $
Total	Numeric	647474	17952361	6090317.896±3799355.361	Thousand $

There are several pros to using supervised learning algorithms like the random forest or kNN over unsupervised or reinforcement learning for estimating tourism expenditure. First of all, supervised learning uses labelled data for training. To estimate tourism expenditure, the proposed dataset should contain the actual spending amounts for tourists. As a result of labelling, supervised algorithms can learn directly from these examples to predict spending for new tourists. Unsupervised learning has no labels so it cannot make direct predictions. Reinforcement learning also requires a reward signal which may be difficult to define for this task. Secondly, supervised techniques are designed to generate predictive models. Unsupervised learning focuses on pattern detection. Reinforcement learning learns policies or actions, not predictive models. Thirdly, with supervised learning, the predictions of the model can be compared to actual labelled expenditure amounts to evaluate accuracy.

ML approaches can be classified based on the type of data and the availability of labels for the dependent variable. Supervised algorithms are used when labels are available for either continuous or discrete dependent variables, while unsupervised methods are applied when no labels are given. Hence, in this study, the most suitable algorithms take part and related to the supervised learning methodology are utilized. Because, the dataset, whose labels or classes are assisted as expenditure types obtained by the TÜİK website is suitable for supervised analysis.

3.2.1. Random Forest Algorithm

The Random Forest Algorithm was initially introduced by Breiman (2001) as a supervised method for solving regression and classification problems (Shalev-Shwartz and Ben-David, 2014). Its working principle is similar to the Decision Tree Algorithm, and the algorithm itself resembles a tree-like structure. The tree consists of leaves and test nodes, with a leaf node terminating an end node that includes classes of objects based on their characteristics. On the other hand, a test node is an internal node that compares features belonging to the desired node using the other branch. Based on the

comparison result, each test node is defined, and the final attribute is attained from the class label of the reached leaf node (Quinlan, 1986).

3.2.2. k- Nearest Neighbors (kNN) Algorithm

The kNN algorithm classifies the instances through the distance between the clusters (Sahoo and Kumar, 2012). In the kNN algorithm, distance is a hyperparameter to be determined, such as Euclidean distance, Manhattan distance, or Chebyshev distance, among others. In this paper, Euclidean distance is selected as the distance method as given in Equation (1):

$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{1}$$

Based on the above equation results, each instance attained a class which is nearest to the distance. 'k' parameter given in Equation (1) is depicted for the number of instances. Even algorithm is simple, it is a popular method in machine learning (ML) and widely used classification algorithms (Gençoğlan *et al*., 2020). During *k*NN algorithm analysis, k is selected as 1 to assign the instance effectively to the class attribute.

3.2.3. Locally Weighted Learning (LWL)

The LWL algorithm is mainly supported the approximated model among instances for analysing the features (Atkeson *et al*., 1997). This algorithm will be used for classification with the Naive mathematician approaches that are the supervised machine learning approaches and multivariate analysis is performed. Moreover, LWL is additionally categorised as a non-parametric supervised learning algorithm through density function probability.

LWL Algorithm Principle: Prediction of the inbound tourist expenditures

Dataset: INPUTs including years, quarters, food-beverage, accommodation, health, transportation, sports-education-culture, tour, clothes and shoes, souvenirs, and other expenditures are attained as query points. OUTPUT is total inbound tourist expenditures.

Given: x_q is a notation for each feature of the dataset where q is taken from 1 to 13. 'n' is the total number of the instances (the number of training sample) where n=75. Additionally; $\boldsymbol{X}$, $\boldsymbol{y}$, $\boldsymbol{D}$, $\boldsymbol{W}$, β_q and $\check{y}_q$ stand for build matrix, build vector, diagonal weight matrix, distance metric, regression coefficient and predicted output, respectively:

$$\boldsymbol{X} = (\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, \dots, \tilde{\boldsymbol{x}}_n)^T \text{ where } \tilde{\boldsymbol{x}}_i = [(\boldsymbol{x}_i - \boldsymbol{x}_q)^T 1]^T$$

$$\boldsymbol{y} = (y_1, y_2, \dots, y_n)^T$$

$$w_{i,i} = e^{\left(\frac{-1}{2}(x_i - x_q)^T \mathbf{D}(x_i - x_q)\right)}$$

$$\beta_q = (\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{y}$$

$$\check{y}_q = [x_q{}^T \quad 1]\, \beta_q$$

There are a number of terms namely correlation coefficient (CC), mean absolute error (MAE), and root mean square error (RMSE) to define the performance of the classification algorithms. Correlation coefficients are used to a degree how robust a relationship is among attributes. The correlation coefficient takes the value between -1 and 1. The obtained result, which is 1, means a strong relationship among the features of the dataset. The mean absolute error (MAE) is an average of overall absolute

errors between actual (x_i) and predicted values (y_i)($|x_i - y_i|$). The root mean square error is the average square root of the difference between forecasting and actual values. The formulas of the correlation coefficient, mean absolute error, and root mean square error are given in Equations of 2-4, respectively:

$$CC=\frac{n\sum(xy)-(\sum x)(\sum y)}{\sqrt{[n\sum x^2-(\sum x)^2]\,[n\sum y^2-(\sum y)^2]}} \quad (2)$$

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \quad (3)$$

$$RMSE=\sqrt{\sum_{i=1}^{n}\frac{(x_i-y_i)^2}{n}} \quad (4)$$

3.2.4. Feature Selection Method

The feature selection technique determines the most relevant features in the dataset and speeds up the learning process. Additionally, suitable feature extraction or selection procedures can improve the accuracy of the classification method. In other words, some of the attributes in the proposed dataset, which are expenditure types, might be redundant and irrelevant. Therefore, the irrelevant attributes may lead to lower accuracies. A myriad of feature selection techniques has been suggested for determining crucial features in the literature. For instance, Wang *et al*. (2019) used different feature selection algorithms, namely Sequential Floating Forward Selection (SFFS), Fisher Score, Sequential Forward Selection (SFS), and Fast Correlation-Based Filter Solution (FCBF), to achieve higher accuracy and efficiency during sleep stage through only one channel electroencephalograms (EEGs). Al-Batah *et al*. (2019) applied the Correlation-based Feature Selection (CbFS) algorithm to reduce dataset dimensionality and determine relevant genes for eleven microarray data. The results demonstrated that CbFS can efficiently eliminate redundant and irrelevant attributes and noise. In 2015, Begum *et al*. (2015) focused on the classification technique based on the Leukemia dataset. Initially, Consistency-based Feature Selection (CbFS), Kernelized Fuzzy Rough Set (KFRS), and Fuzzy Preference Based Rough Set (FPRS) were utilized. Then, the kNN algorithm was applied to the reduced dimension dataset. The analyzed results showed that the CbFS algorithm outperformed the others given in Begum *et al*. (2015). In this paper, Correlation-based Feature Selection (CbFS) is utilized to eliminate irrelevant and redundant attributes of expenditure data for estimating total tourist expenditures. Twelve parameters used for estimating the total income obtained by inbound tourists are examined and analyzed to determine their effect on the accuracy of the estimation performance. CbFS is a filter-based feature selection technique that considers the consistency levels of class values to investigate the effectiveness of features (Liu and Setiono, 1996).

3.3. First-Order Two-Variable Polynomial Modelling

The fundamentals of any object handled in calculus are generally described as "functions." The representation of functions can vary and includes an equation, a graph, words, or a table. Functions occur because one variable depends on the other. Dependent variables can be expressed through various independent variables. For example, the human population depends on time 't.' In other words, the population is a function of a time-dependent variable. Functions can be defined as input and output relationships, in terms of the unknown relationship or function (black box), as illustrated in Figure 2.

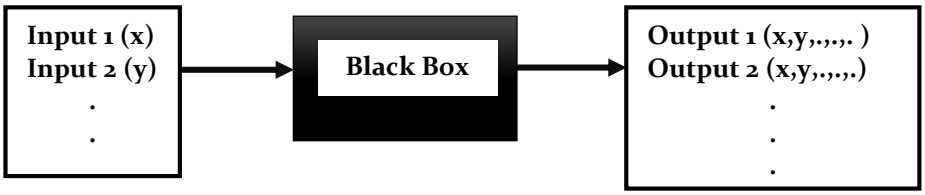


Figure 2. *Input-Output Relationship with respect to the Unknown Equation*

Graph representation of any function includes permissible information for the behaviour of the structure or any historical-based data. Based on this information, the mathematical modelling of any objects or situations might be represented as a first-order polynomial equation as given in Equation 5.:

$$Y = a + b_1X_1 + b_2X_2 + \cdots + b_mX_m + \epsilon \quad (5)$$

Where Y stands for the dependent variable, X_1, X_2, X_3, … , X_m represent the independent variables of the systems. b_1, b_2, b_3, …, b_m outline the coefficients and ϵ depicts the constant term of the proposed approach. In this study, the tourism revenue in Türkiye with respect to the various expenditure types is modelled with regard to the first-order two-variable polynomial equation to forecast via mathematical model and machine learning (ML) algorithms. In addition to machine learning, the study incorporates conventional mathematical analysis to enhance the accuracy and reliability of the predictive model. Initially, the year from 2012 to 2021 is represented in Figure 3 to visualize the first-order equation line since the historical data gives a hint for forecasting future data.

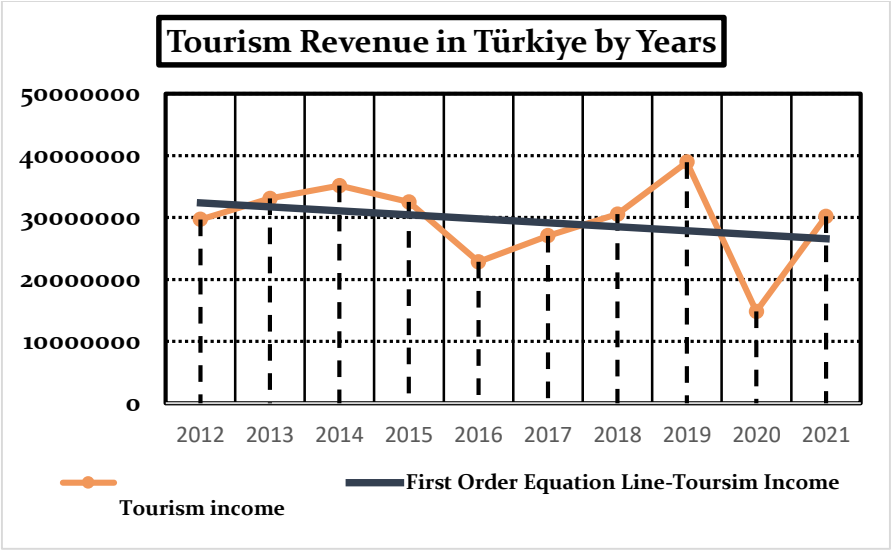


Figure 3. *The Total Tourism Revenue with respect to Years from 2012 to 2021*

Then, the total tourism revenue is detailed in Table 2 as well to illustrate the total tourism revenue in terms of a table as mentioned before.

Table 2. *Total Tourism Revenue in Türkiye from the Year 2012 to 2021 (Thousand $)*

Year	Tourism Revenue
2012	29689249
2013	33073502
2014	35137949
2015	32492212
2016	22839468
2017	27044542
2018	30545924
2019	38930474
2020	14817273
2021	30173587

4. Findings

The ML algorithms discussed in the previous section have been applied to the proposed dataset for estimating the total revenue obtained by inbound tourists. A 10-fold cross-validation procedure has been utilized for testing and training the data. In other words, out of 75 instances, 68 instances (90% of the dataset) have been used for training the model, and 7 instances (10% of the dataset) have been used to test the model via iterations. The results of the algorithms are listed in Table 3. According to the performance analysis of the algorithms, Random Forest outperforms the kNN and LWL algorithms in terms of a higher correlation coefficient. The CC for these algorithms is determined as 0.98, 0.96, and 0.84, respectively. These results reveal that the total revenue obtained by tourists can be estimated using the 12 attributes used in this study with a correlation coefficient higher than 85%.

Table 3. *Performance Comparisons of Machine Learning (ML) Algorithms for Estimation*

Algorithms	CC	MAE	RMSE
Random Forest	0.98	528.038	822.216
***k*NN**	0.96	788.611	1.067.718
LWL	0.84	1.685.289	2.064.462

When CbFS is applied, the most relevant and crucial features are determined as the year, food-beverage, and accommodation. The reduced dataset through CbFS is tested again with machine learning algorithms for the estimation of the total revenue obtained from incoming tourists. Therefore, in this case, the number of ML models has three inputs rather than twelve. However, the number of the model output remains the same. The results with the reduced dataset, including the most relevant features that occurred by feature selection methods, are given in Table 4. The reduced dataset results are clearly given in Table 4, where the CbFS method is applied to three algorithms, namely Random Forest, kNN, and LWL.

Comparing the results before and after feature selection, the Random Forest algorithm had a high correlation coefficient (CC) of 0.98, indicating a strong positive correlation between the predicted and actual values. The mean absolute error (MAE) was relatively low at 528.038, suggesting small average errors in the predictions. The root mean square error (RMSE) was 822.216, indicating overall small errors in the predictions. After feature selection: The CC decreased slightly to 0.97, indicating a slightly weaker but still strong correlation between the predicted and actual values. The MAE increased to 579.722,

suggesting slightly larger average errors in the predictions. The RMSE also increased to 895.159, indicating slightly larger overall errors in the predictions.

Feature selection led to a minor decrease in performance for the Random Forest algorithm, as indicated by a slightly lower correlation coefficient and slightly larger errors in both MAE and RMSE. However, the algorithm still performs well overall. Before feature selection, the kNN algorithm had a high CC of 0.96, indicating a strong positive correlation between the predicted and actual values. The MAE was 788.611, suggesting larger average errors in the predictions compared to Random Forest. The RMSE was 1.067.718, indicating larger overall errors in the predictions compared to Random Forest. After feature selection, the CC decreased to 0.95, indicating a slightly weaker but still strong correlation between the predicted and actual values. The MAE increased to 843.735, suggesting slightly larger average errors in the predictions. The RMSE increased to 1.221.101, indicating slightly larger overall errors in the predictions. Similar to Random Forest, feature selection resulted in a minor decrease in performance for kNN. The correlation coefficient decreased slightly, and both MAE and RMSE increased, indicating slightly larger errors in the predictions. The LWL algorithm had a CC of 0.84, indicating a weaker correlation between the predicted and actual values compared to Random Forest and kNN. The MAE was 1.685.289, suggesting larger average errors in the predictions. The RMSE was 2.064.462, indicating larger overall errors in the predictions. After feature selection, the CC remained the same at 0.84, indicating no change in the correlation between the predicted and actual values. The MAE increased to 1.713.489, suggesting slightly larger average errors in the predictions. The RMSE increased to 2.084.604, indicating slightly larger overall errors in the predictions. Feature selection did not significantly impact the performance of the LWL algorithm. The correlation coefficient remained the same, and there were slight increases in both MAE and RMSE, indicating slightly larger errors in the predictions.

After feature selection, all three algorithms experienced some degree of decrease in performance, as evidenced by decreases in the correlation coefficient and increases in both MAE and RMSE. However, it's important to note that the changes were relatively small, and the overall ranking of the algorithms remained the same, with Random Forest performing the best, followed by kNN, and then LWL.

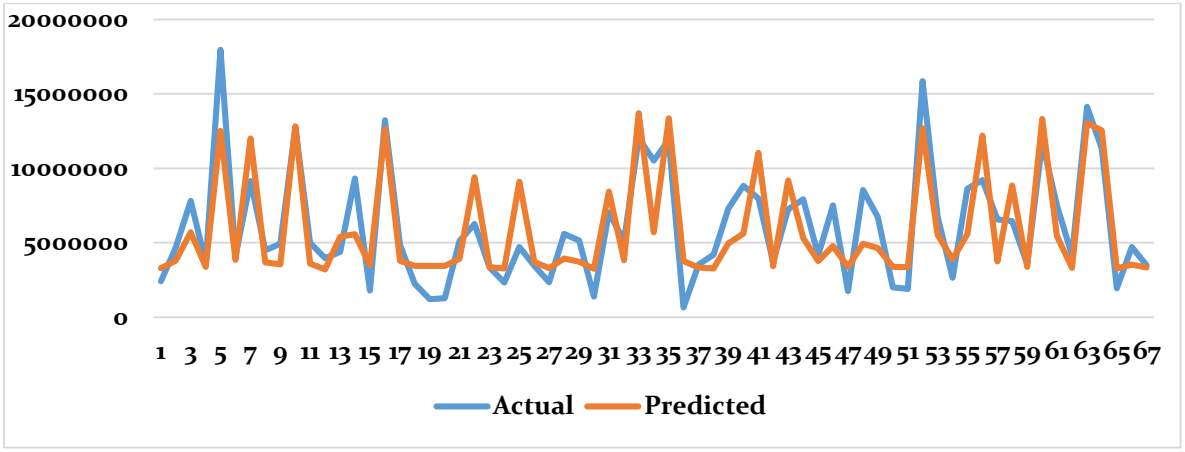
Table 4. *Performance Analysis of Reduced Dataset*

Algorithms	CC	MAE	RMSE
Random Forest	0.97	579.722	895.159
kNN	0.95	843.735	1.221.101
LWL	0.84	1.713.489	2.084.604

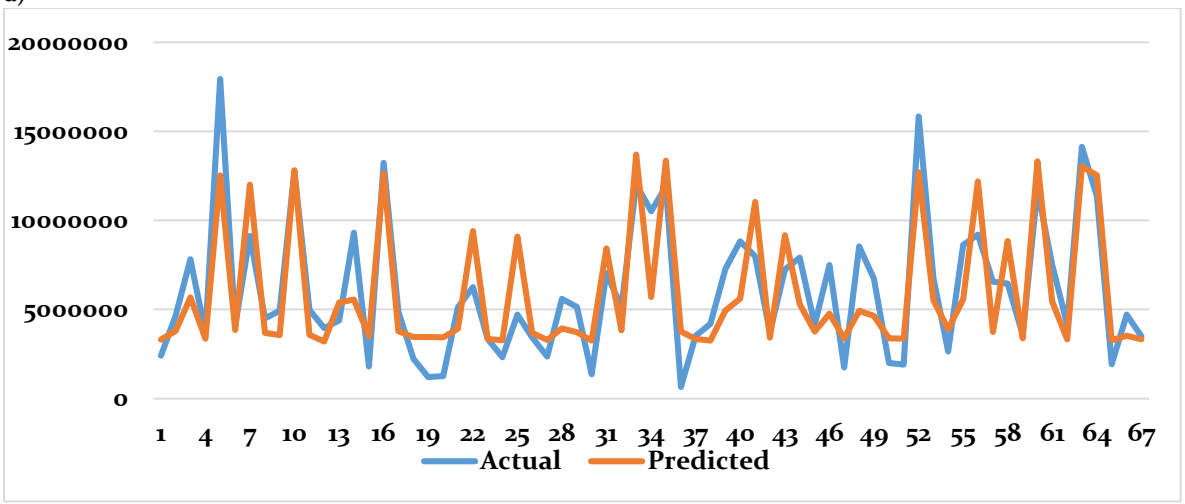
When the real and predicted lines are compared in Figure 4, it is clear that the random forest produces the most overlapping lines. The tourist expenditure types are modelled through the first-order two-variable polynomial equation. The closed-form mathematical expression is expressed by Equation 6:

$$f(x,y) = p_{00} + p_{10}*x + p_{01}*y \quad (6)$$

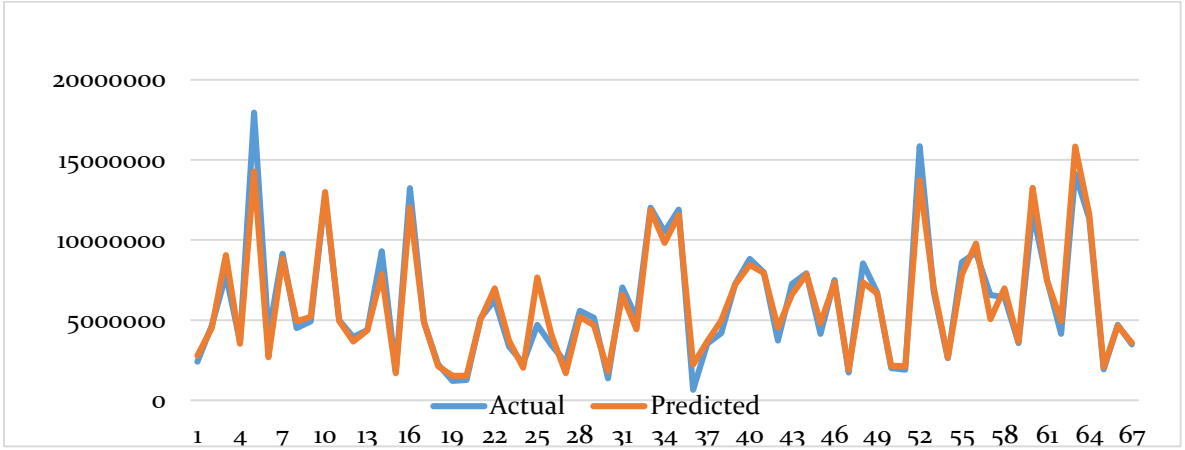
Where x and y stand for the year and quarter variables. Each type has different coefficients of eleven tourist expenditure types calculated by past data from 2005 to 2022. According to the past data, Table 5 outlines the polynomial coefficients of each expenditure type.



a)



b)



c)

Figure 4. *Actual and Predicted Values of the total tourist income via machine learning algorithms of a) kNN, b) LWL, and c) Random Forest*

Table 5. *Polynomial Coefficients of Expenditure Types*

Types	Coefficients (with 95% confidence bounds)
1. Accommodation	p_{00} =-1.119e+08(-1.622e+08, -6.161e+07) p_{10} = 5.587e+04 (3.09e+04, 8.083e+04) p_{01} = 2.467e+05 (1.289e+05, 3.645e+05)
2. Clothes and Shoes	p_{00} = -1.037e+08 (-1.34e+08, -7.343e+07) p_{10} = 5.171e+04 (3.668e+04, 6.674e+04) p_{01} = 1.106e+05 (3.967e+04, 1.815e+05)
3. Food and Beverage	p_{00} =-1.119e+08 (-1.622e+08, -6.161e+07) p_{10} = 5.587e+04 (3.09e+04, 8.083e+04) p_{01} = 2.467e+05 (1.289e+05, 3.645e+05)
4. Health	p_{00} = -4.808e+07 (-5.706e+07, -3.91e+07) p_{10} = 2.395e+04 (1.949e+04, 2.84e+04) p_{01} = 2.003e+04 (-994.9, 4.106e+04)
5. SEC	p_{00} = -5.486e+06 (-8.814e+06, -2.158e+06) p_{10} = 2742 (1089, 4394) p_{01} = 8920 (1125, 1.672e+04)
6. Tour	p_{00} = 4.781e+06 (1.076e+06, 8.485e+06) p_{10} = -2363 (-4203, -523.9) p_{01} = 8732 (54.6, 1.741e+04)
7. Transportation	p_{00} = -3.714e+07 (-5.613e+07, -1.814e+07) p_{10} = 1.855e+04 (9118, 2.798e+04) p_{01} =6.984e+04 (2.535e+04, 1.143e+05)
8. Souvenirs	p_{00} = -6.086e+06 (-2.045e+07, 8.278e+06) p_{10} = 3125 (-4007, 1.026e+04) p_{01} = 3.211e+04 (-1538, 6.575e+04)
9. Others	p_{00} = -7.825e+07 (-1.016e+08, -5.486e+07) p_{10} = 3.899e+04 (2.738e+04, 5.061e+04) p_{01} =6.854e+04 (1.375e+04, 1.233e+05)
10. Total Tourism Income	p_{00} = -7.536e+08 (-1.04e+09, -4.674e+08) p_{10} = 3.759e+05 (2.338e+05, 5.18e+05) p_{01} =1.143e+06 (4.728e+05, 1.814e+06)

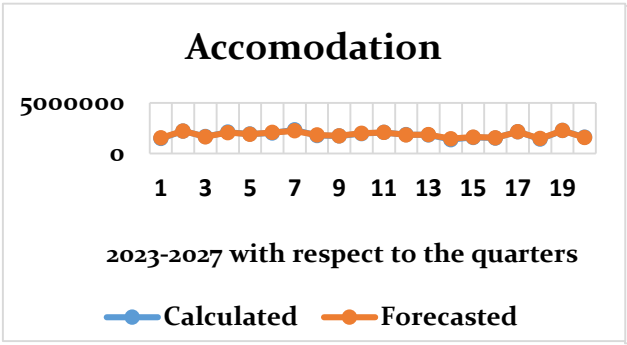
Based on the coefficients obtained and given in Table 5, tourist expenditure for each expenditure type is calculated with respect to year (from 2023 to 2027) and quarters (from 1 to 4). The Random Forest algorithm, which outperforms the other two algorithms in this study, is again evaluated to emphasize the mathematical modeling approach. By evaluating the algorithm, the tourist expenditure types for the next five years are forecasted. The calculated and forecasted values of each expenditure type are outlined in Table 6. It is clear from Table 6 that the total tourism income in the next five years is expected to reach up to 44,827,200 (thousand dollars) calculated or 43,543,949 (thousand dollars)

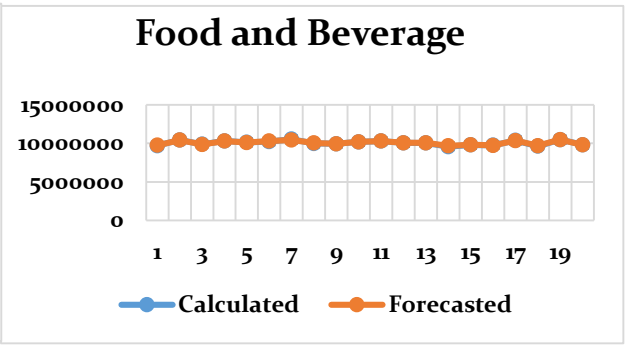
forecasted. In addition, the most relevant and crucial types of tourist expenditure are deduced to be accommodation and food & beverage due to their higher income compared to the other types.

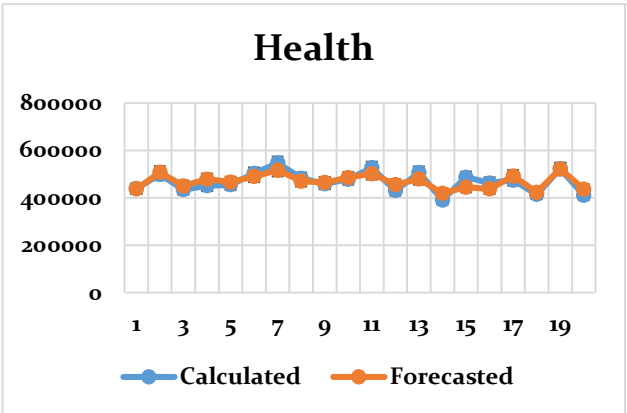
Table 6. *Each Pattern Analysis in terms of the future years from 2023 to 2027 (Quarterly)*

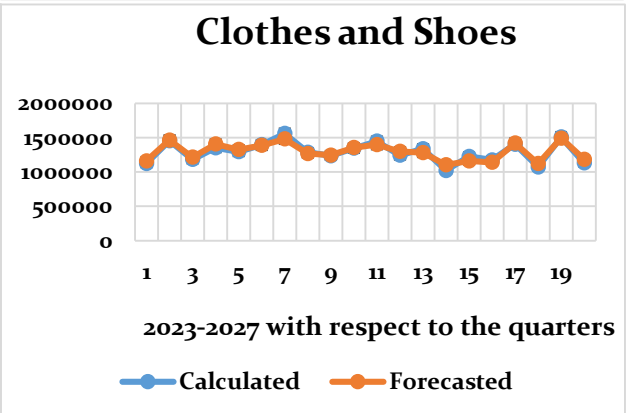
Patterns		Mean ± Std Dev	Median	Min.	Max.
Accommodation	**Calculated**	1853500 ± 286912	1853500	286912	2335290
	Forecasted	1862214 ± 260738	1855223	260738	2270902
Clothes and Shoes	**Calculated**	1289250 ± 143660	1289250	1019930	1558570
	Forecasted	1294806 ± 123074	1289828	1101899	1491068
Food and Beverage	**Calculated**	10053500 ± 286912	10053500	9571710	10535290
	Forecasted	10060031 ± 261825	10055641	9673154	10468202
Health	**Calculated**	468825 ± 109430	468825	390880	546770
	Forecasted	469057 ± 29830	468367	419410	521605
Sports, Culture and Education	**Calculated**	88850 ± 10700	88850	69986	107714
	Forecasted	89268 ± 9703	88820	74415	104566
Tour	**Calculated**	17755 ± 10318	17755	3279	35579
	Forecasted	17547 ± 8327	18163	4986	30195
Transportation	**Calculated**	598350 ± 82372	598350	456490	740210
	Forecasted	601483 ± 75351	599463	486546	719281
Souvenirs	**Calculated**	322400 ± 36171	322400	267985	274644
	Forecasted	323071 ± 33156	320540	274644	373008
Others	**Calculated**	876100 ± 94406	876100	695310	754472
	Forecasted	878792 ± 80105	874415	754472	1012889
Total	**Calculated**	10455000 ± 1384074	10455000	7988700	12921300
	Forecasted	10513781 ± 1253011	10458967	8599277	12495572

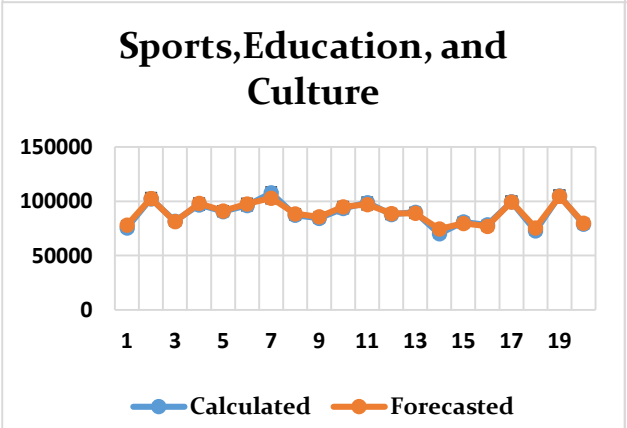
Figure 5 visually presents the calculated and forecasted values for each tourist expenditure type.

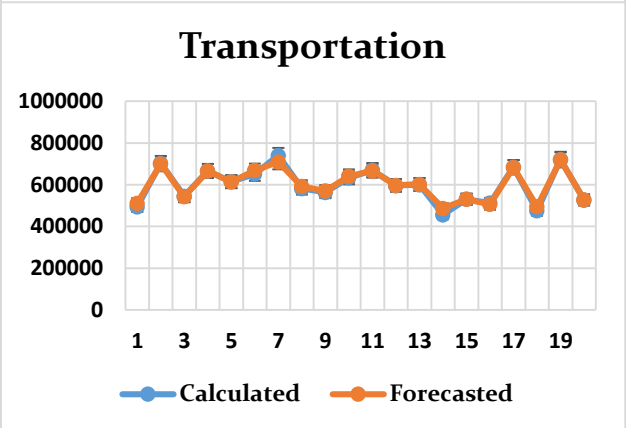

Accomodation
5000000
0
1 3 5 7 9 11 13 15 17 19
2023-2027 with respect to the quarters
Calculated
Forecasted

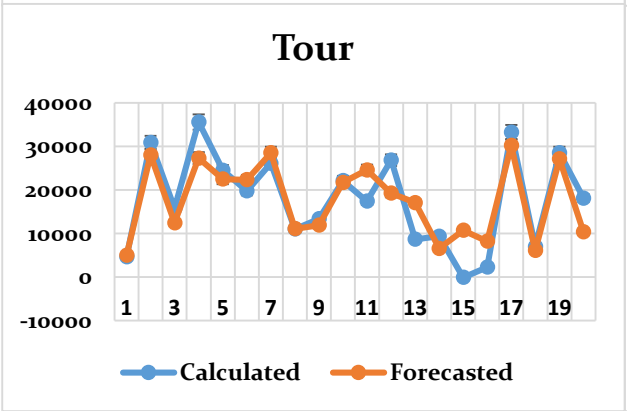

Food and Beverage
15000000
10000000
5000000
0
1 3 5 7 9 11 13 15 17 19
Calculated
Forecasted

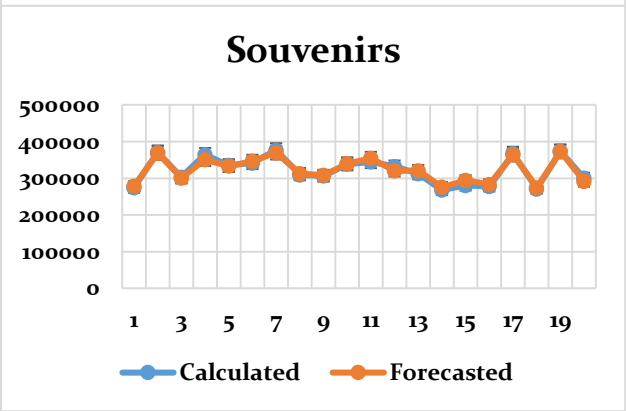

Health
800000
600000
400000
200000
0
1 3 5 7 9 11 13 15 17 19
Calculated
Forecasted


Clothes and Shoes
2000000
1500000
1000000
500000
0
1 3 5 7 9 11 13 15 17 19
2023-2027 with respect to the quarters
Calculated
Forecasted


Sports,Education, and Culture
150000
100000
50000
0
1 3 5 7 9 11 13 15 17 19
Calculated
Forecasted


Transportation
1000000
800000
600000
400000
200000
0
1 3 5 7 9 11 13 15 17 19
Calculated
Forecasted


Tour
40000
30000
20000
10000
0
-10000
1 3 5 7 9 11 13 15 17 19
Calculated
Forecasted


Souvenirs
500000
400000
300000
200000
100000
0
1 3 5 7 9 11 13 15 17 19
Calculated
Forecasted

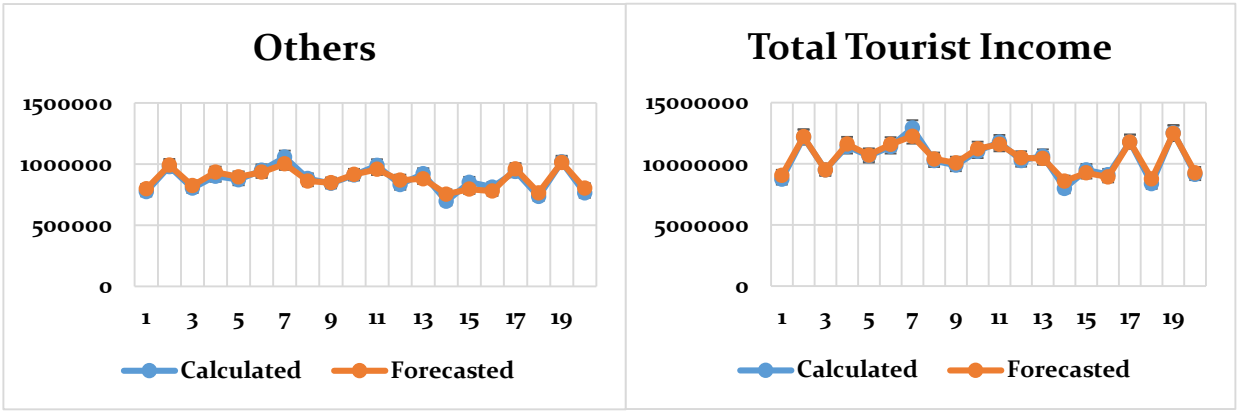


Figure 5. *Calculated and Forecasted Values for Each Tourist Expenditure Type*

5. Discussion and Recommendations

This paper aims to model and estimate international tourism receipts in Türkiye based on expenditure types using conventional mathematical analysis namely first-order two-variable polynomial equation and artificial intelligence-based machine learning algorithms. A hybrid approach was preferred in the study, which employed both conventional and cutting-edge methods. Tourism is undoubtedly an essential sector of the Turkish economy. Türkiye is a natural tourism centre due to its location, geographical features, and historical heritage on the lands it occupies. Nevertheless, this is not the reason for choosing Türkiye as the case in this paper which examines tourist expenditure types through artificial intelligence-based machine learning algorithms. Despite being one of the most visited countries in the world (ranked sixth with 51 million tourists in 2019), Türkiye's share of tourism revenue (under 40 billion dollars in 2019 and not in the top ten) is not significant. Therefore, the reason for Türkiye being considered a "cheap" country where a tourist can return without spending even a thousand dollars has been thoroughly examined, and artificial intelligence has been used to predict whether this situation will change in the future. The reason for taking into account 2019 is that the World Tourism Organization declared 2020 as the worst year in tourism history, claiming that it would take four years to return to 2019 levels and fully recover from the destructive effects of the Covid19 pandemic.

Firstly, using the feature selection method, the expenditure types that have the greatest impact on Türkiye's tourism revenue were identified, and it was discovered that accommodation and food & beverage were prominent. The results of the research indicate that expenditure types such as accommodation and food & beverage have a significant impact on Türkiye's tourism revenue. This finding is consistent with previous studies that have shown that tourists tend to spend more on basic needs such as lodging and food. For example, a study by Fleissig (2021) found that accommodation and food & beverage expenditures were inelastic and necessary needs. Gómez-Déniz *et al.* (2020) found that the most significant expense was expenditure on accommodation. Accommodation and food & beverage were also highlighted as the top two expenditure categories in Northern Italy's tourism sector by Disegna and Osti (2016). Similarly, a study by Alegre *et al.* (2011) on the analysing the influence of tourist motivations on tourist expenditure found that accommodation was a significant variable in explaining total expenditure levels and expenditure in the country of origin. Jingwen and Mingzhu (2018), on the other hand, found accommodation and food & beverage to have a limited pulling function on visitor expenditure when compared to other commodities. Since these types of expenditures are the basic needs of every tourist, it is obvious why Türkiye's tourism income does not increase in proportion to

the number of tourists. This finding of the study shows that most tourists visiting Türkiye hardly ever leave their hotels. They keep their extra expenses low, and this may be due to the all-inclusive system applied in most holiday destinations in the country. The all-inclusive system is a popular pricing strategy in the tourism industry, especially in resort destinations. Under this system, tourists pay a fixed price for their accommodation, food, and some additional services such as drinks, snacks, and entertainment. The popularity of this pricing model has increased in recent years, with more and more hotels and resorts adopting it (Bladh & Holm, 2013).

Many all-inclusive resorts offer a limited number of excursions or activities. Besides, all-inclusive resorts can limit tourists' interactions with local people and businesses, which can negatively impact local economies and the cultural exchange experience. Alegre and Pou (2008) have suggested that the all-inclusive system can have both positive and negative impacts on the local economy and the tourism industry. On the one hand, it can attract more tourists by providing them with a hassle-free and predictable holiday experience, and thus increase the revenue for the local businesses. On the other hand, it can discourage tourists from exploring the local attractions and spending money on local businesses outside the resort, resulting in a leakage of tourism revenue from the local economy. Additionally, the all-inclusive system can lead to environmental problems, such as excessive waste generation and water usage, if not managed properly. Therefore, the necessity to call for a more sustainable and responsible approach to the all-inclusive system in tourism, such as promoting local food and culture, reducing waste, and involving local communities in the decision-making process has emerged. Originally, the all-inclusive system was designed to prevent tourists from going out in countries having security problems. However, it has turned into a model focused on the hotels' profit margins and has led to very low per capita expenditures, so it is recommended that Türkiye should abandon this system in the course of time.

The study then forecasted Türkiye's tourism revenue for the next few years using a machine learning (ML) algorithm with a 97% accuracy rate. According to the calculations, Türkiye's total tourism revenue will not exceed 45 billion dollars by 2027. Based on this finding, it is possible to say that Türkiye's tourism income will also remain low in the future. Türkiye's rival countries in Europe such as Spain, Italy and France, which are all Mediterranean countries like itself, had already exceeded 45 billion dollars even in 2019. In contrast, some studies offer a more optimistic outlook for Türkiye's tourism industry. For instance, a study by Mengu (2018) predicts that Türkiye's tourism revenue will increase in the future due to the country's strategic location, diverse natural and cultural resources, and improved tourism infrastructure. Another study highlights the potential of health tourism in Türkiye and its contribution to the growth of the country's tourism industry (Uygun & Ekiz, 2016). Türkiye should embark on a transformation process without wasting too much time and focus on tourism which has high expenditure levels but relatively low environmental damage and costs. To achieve this, encouraging sustainable tourism and promoting the country's unique culture, history, and natural beauty can be recommended because high-income level tourists often seek out destinations with rich cultural heritage, such as art, music, cuisine, and history.

It is also crucial to diversify the tourism service offered in this process. The diversified types of tourism implied cover a wide spectrum. For example, congress or fair tourism can take place in big cities. For business tourism such as conferences, trade fairs, and meetings, it is necessary to develop and promote the country. This can attract high-income level tourists who are also business executives and entrepreneurs. High-income level tourists may seek out destinations that offer sustainable and environmentally friendly tourism options. Developing and promoting eco-tourism activities such as wildlife safaris, bird watching, and nature walks can attract this type of tourist, so rural tourism should take place in rural areas. It is important to develop and promote the country as a destination for medical tourism, such as cosmetic surgery, dental treatments, and other health-related services. This can attract

high-income level tourists who are looking for these types of services. Therefore, health tourism, medical tourism or thermal tourism should be developed in regions where health centres are concentrated. These types of tourism, which are generally tailored to individuals or small groups, can be attended by high-income people that will both generate more revenue for the country and have low societal costs. Last but not least, high-income level tourists are often looking for unique and exclusive experiences such as escapism in casinos. In fact, a controlled reopening of casinos may be on the agenda by taking the United States of America, which has the highest tourism income with 214 billion dollars, as an example. These recommendations are not only applicable to Türkiye but also to other European countries whose economies rely on tourism yet cannot receive a sufficient share of tourism income worldwide.

The paper's theoretical framework draws on optimization theory to compare and select the best machine learning algorithm for estimating tourist expenditures. Optimization theory is a subfield of mathematics that deals with finding the best solution for a given problem by minimizing or maximizing an objective function subject to constraints. In the context of machine learning, optimization theory is used to find the best parameters for a model that minimize a cost function. This paper contributes to the international body of knowledge by presenting a novel approach to estimating and modelling tourism receipts. The study provides a more accurate and efficient method for estimating tourism revenues by combining traditional mathematical analysis and artificial intelligence-based machine learning algorithms. The study contributes to tourism theory by identifying the most significant types of expenditure affecting tourism revenue, which can then be used to inform tourism strategies and policies. Furthermore, the use of machine learning algorithms in estimating future tourism revenue in the study provides a valuable methodological approach that can be replicated in other studies. As a whole, this paper offers a useful and innovative contribution to the field of tourism research.

Besides, the study's forecasting of Türkiye's tourism revenue using machine learning algorithms can provide valuable insights for tourism policymakers and industry practitioners in planning and budgeting for the future. Knowing that the country's total tourism revenue may not exceed 45 billion dollars by 2027 can help them make informed decisions about where to allocate resources and how to develop and promote tourism services. As a result, the findings of this research can contribute to more effective and efficient tourism management in Türkiye and other European countries that heavily rely on tourism income, ultimately leading to economic growth and development. While this study provides useful awareness of the impact of tourist expenditure types on a country's tourism revenue and offers a practical approach to estimating future revenue, there is still much to be explored. Researchers should be encouraged to pursue further investigation in this area, particularly in terms of expanding the scope of the study and testing the validity of the assumptions made in the analysis. Using machine learning is a nish area in the tourism literature. In the future, different datasets linked to tourism can be integrated, and alternative machine learning algorithms can be employed. Lastly, Türkiye has been chosen as the case in this study, other European countries can be selected in the future and tourism receipts can be handled with artificial intelligence-based machine learning techniques.

References

Al-Batah, M.S. Zaqaibeh, B.M., Alomari, S.A., and Alzboon, M.S. (2019). Gene microarray cancer classification using correlation-based feature selection algorithm and rules classifiers. *International Journal of Online and Biomedical Engineering*, 15 (8), 62-73. DOI: https://doi.org/10.3991/ijoe.v15i08.10617.

Alfarhan, U. F., Olya, H., and Nusair, K. (2022). How do prosperity and aspiration underlie leisure tourism expenditure patterns? *Tourism Economics*. DOI: https://doi.org/10.1177_13548166211064215.

Alegre, J., Cladera, M., & Sard, M. (2011). Analysing the Influence of Tourist Motivations on Tourist Expenditure at a Sun-And-Sand Destination. *Tourism Economics*, 17(4), 813–832. DOI: https://doi.org/10.5367/te.2011.0063.

Alegre, J., & Pou, L. (2008). Research Note: Tourism Expenditure and all-Inclusive Packages – the case of a Mature Mediterranean Destination. *Tourism Economics*, 14(3), 645–655. DOI: https://doi.org/10.5367/000000008785633631.

Atkeson, C.G., Moore, A.W., & Schaal, S. (1997). Locally weighted learning, *Artificial Intelligence Review,* 11, 11-73. DOI: https://doi.org/10.1023/A:1006559212014.

Begum, S., Chakraborty, D., & Sarkar, R. (2015). Data classification using feature selection and kNN machine learning approach," in International Conf. on Computational Intelligence and Communication Networks (CICN), IEEE, pp. 811–814.

Bladh, C., & Holm, H. J. (2013). Can Economics Explain Where All-Inclusive Deals are Offered? *Tourism Economics*, 19(2), 339–348. DOI: https://doi.org/10.5367/te.2013.0209.

Bi, J.-W., Han, T.-Y., & Li, H. (2022). International tourism demand forecasting with machine learning models: The power of the number of lagged inputs. *Tourism Economics*, 28(3), 621–645. DOI: https://doi.org/10.1177/1354816620976954.

Breiman, L. (2001). Random Forests. *Machine Learning* 45, 5–32. DOI: https://doi.org/10.1023/A:1010933404324.

Brida, J., & Scuderi, R. (2013). Determinants of tourist expenditure: A review of microeconometric models. *Tourism Management Perspectives*, 6, 28-40. DOI: https://doi.org/10.1016/j.tmp.2012.10.006.

Chen, K. (2011). Combining linear and nonlinear model in forecasting tourism demand. *Expert Systems with Applications*, 38, 8, 10368-10376. DOI: https://doi.org/10.1016/j.eswa.2011.02.049.

Chen, S., Wang, X., Zhang, H., Wang, J., & Peng, J. (2021). Customer purchase forecasting for online tourism: A data-driven method with multiplex behavior data, *Tourism Management*, 87, 104357. DOI: https://doi.org/10.1016/j.tourman.2021.104357.

Disegna, M., & Osti, L. (2016). Tourists' Expenditure Behaviour: The Influence of Satisfaction and the Dependence of Spending Categories. *Tourism Economics*, 22(1), 5–30. DOI: https://doi.org/10.5367/te.2014.0410.

Erdem, M., & Jiang, L. (2016). An overview of hotel revenue management research and emerging key patterns in the third millennium. *Journal of Hospitality and Tourism Technology*, 7 (3), 300-312. DOI: https://doi.org/10.1108/JHTT-10-2014-0058.

Fleischer, A., & Felsenstein, D. (2000). Support for rural tourism: Does it make a difference? *Annals of Tourism Research*, 27 (4), 1007-1024. DOI: https://doi.org/10.1016/S0160-7383(99)00126-7.

Fleissig, A. R. (2021). Expenditure and price elasticities for tourism sub-industries from the Fourier flexible form. *Tourism Economics*, 27(8), 1692–1706. DOI: https://doi.org/10.1177/1354816620938170.

Gençoğlan, D.N., Arslan, M.T., Çolak, Ş., & Yildirim, E. (2020). Ultra-Wideband (UWB) characteristic estimation of elliptic patch antenna based on machine learning techniques. *Frequenz*, 74 (9-10), 351-358. DOI: https://doi.org/10.1515/freq-2019-0210.

Gómez-Déniz, E., Pérez-Rodríguez, J. V., & Boza-Chirino, J. (2020). Modelling tourist expenditure at origin and destination. *Tourism Economics*, 26(3), 437–460. DOI: https://doi.org/10.1177/1354816619840845.

Ivanov, S., & Webster, C. (2017). Adoption of Robots, Artificial Intelligence and Service Automation by Travel, Tourism and Hospitality Companies – A Cost-Benefit Analysis. Prepared for the International Scientific Conference "Contemporary Tourism – Traditions and Innovations", Sofia University, 19-21 October 2017, Available at SSRN: https://ssrn.com/abstract=3007577.

Jingwen, W., & Mingzhu, L. (2018). Characteristics of visitor expenditure in Macao and their impact on its economic growth. *Tourism Economics*, 24(2), 218–233. DOI: https://doi.org/10.1177/1354816617749352.

Jiménez, J., Valido, J., & Pellicer, A. (2023). Evaluating the effect of air transport resident subsidies on non-residents tourists' expenditure. *Journal of Air Transport Management*, 106, 102319, DOI: https://doi.org/10.1016/j.jairtraman.2022.102319.

Jurdana, D., & Frleta, D. (2017). Satisfaction as a determinant of tourist expenditure. *Current Issues in Tourism*, 20 (7), 691-704, DOI: 10.1080/13683500.2016.1175420.

International Monetary Fund (IMF). (2020). Finance and Development, Wish You Were Here, Adam Behsudi, December 2020. Available online: https://www.imf.org/en/Publications/fandd/issues/2020/12/impact-of-the-pandemic-on-tourism-behsudi (accessed on 20 December 2022).

Li, X., Li, H., Pan, B., and Law, R. (2021). Machine Learning in Internet Search Query Selection for Tourism Forecasting. *Journal of Travel Research*, 60(6), 1213–1231. DOI: https://doi.org/10.1177/0047287520934871.

Li, N., Yang, X., Wong, I.A., Law, R., Xu, J.Y., & Zhang, B. (2023). Automating tourism online reviews: a neural network based aspect-oriented sentiment classification. *Journal of Hospitality and Tourism Technology*, 14 (1), 1-20. DOI: https://doi.org/10.1108/JHTT-03-2021-0099.

Liu, H., & Setiono, R. (1996). A probabilistic approach to feature selection: a filter solution, Proceed. Thirteenth Intern, *Conf. Machine Learn.*, 96, 319–327.

Mariani, M., Platanakis, E., Stafylas, D., & Sutcliffe, C. (2023). Identifying a destination's optimal tourist market mix: Does a superior portfolio model exist? *Tourism Management*, 96, 104722. DOI: https://doi.org/10.1016/j.tourman.2023.104722.

Mengu, C. (2018). Approaches on The Future of Turkey Tourism in 2023. *European Journal of Educational & Social Sciences*, 3 (2), 164-184.

Oday, A., Ozturen, A., Ilkan, M., & Abubakar, A.M. (2021). Do eReferral, eWOM, familiarity and cultural distance predict enrollment intention? An application of an artificial intelligence technique. *Journal of Hospitality and Tourism Technology*, 12 (3), 471-488. DOI: https://doi.org/10.1108/JHTT-01-2020-0007.

Oh, M., & Kim, S. (2022). Role of Emotions in Fine Dining Restaurant Online Reviews: The Applications of Semantic Network Analysis and a Machine Learning Algorithm. *International Journal of Hospitality and Tourism Administration*, 23 (5), 875-903. DOI: https://doi.org/10.1080/15256480.2021.1881938.

Pai, P., Hung, K., & Lin, K. (2014). Tourism demand forecasting using novel hybrid system. *Expert Systems with Applications*, 41, 8, 3691-3702. DOI: https://doi.org/10.1016/j.eswa.2013.12.007.

Quinlan, J.R. (1986). Induction of decision trees, *Machine learning*, 1, 81-106.

Ramos, L.M., & Sol Murta, F. (2022). Tourism seasonality management strategies – what can we learn from payment data. *Journal of Hospitality and Tourism Insights*, 6 (3), 1205-1221. DOI: https://doi.org/10.1108/JHTI-12-2021-0337.

Ramos-Domínguez, A., Hernández-Martín, R., & Padrón-Ávila, H. (2023). How does the country of origin affect tourist expenditure? An extended quantile regression analysis. *Anatolia*. DOI: 10.1080/13032917.2023.2165128.

Rice, W.L., Park, S.Y., Pan, B., & Newman, P. (2019). Forecasting campground demand in US national parks, *Annals of Tourism Research*, 75, 424-438. DOI: https://doi.org/10.1016/j.annals.2019.01.013.

Sadler, P.G., and Archer, B.H. (1975). The economic impact of tourism in developing countries, *Annals of Tourism Research*, 3 (1), 15-32. https://doi.org/10.1016/0160-7383(75)90015-8.

Sahoo, G. and Kumar, Y. (2012). Analysis of parametric and non-parametric classifiers for classification technique using WEKA. *International Journal of Information Technology and Computer Science (IJITCS)*, 4 (7).

Samara, D., Magnisalis, I., & Peristeras, V. (2020). Artificial intelligence and big data in tourism: a systematic literature review. *Journal of Hospitality and Tourism Technology*, 11 (2), 343-367. DOI: https://doi.org/10.1108/JHTT-12-2018-0118.

Sangkaew, N., & Zhu, H. (2022). Understanding Tourists' Experiences at Local Markets in Phuket: An Analysis of TripAdvisor Reviews. *Journal of Quality Assurance in Hospitality and Tourism*, 23 (1), 89-114. DOI: https://doi.org/10.1080/1528008X.2020.1848747.

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press.

Su, C., Umar, M., & Chang, T. (2023). How consumer confidence is reshaping the outbound tourism expenditure in China? A lesson for strategy makers! *Economic Research-Ekonomska Istraživanja*, 36 (1), 708-726. DOI: 10.1080/1331677X.2022.2106266.

Sun, S., Du, Z., Zhang, C., & Wang, S. (2022). Improving multi-step ahead tourism demand forecasting: A strategy-driven approach. *Expert Systems with Applications*, 30, 118465. DOI: https://doi.org/10.1016/j.eswa.2022.118465.

Turkish Statistical Institute. Statistical Tables, Tourism Income by Type of Expenditure. Available online: https://data.tuik.gov.tr (accessed on 30 December 2022).

UNWTO. (2020). World Tourism Barometer and Statistical Annex, December 2020. Available online: www.e-unwto.org/doi/abs/10.18111/wtobarometereng.2020.18.1.7 (accessed on 10 December 2022).

Uygun, M., & Ekiz, E. (2016). An Overview of Health Tourism Within The Context of Turkey's Tourism Strategy (2023). *Journal of Hospitality and Tourism*, 14 (1), 18-26.

Wager, J. (1995). Developing a strategy for the Angkor World Heritage Site, *Tourism Management*, 16 (7), 515-523. DOI: https://doi.org/10.1016/0261-5177(95)00069-Z.

Wang, Q., Zhao, D., Wang, Y., & Hou, X. (2019). Ensemble learning algorithm based on multi-parameters for sleep staging. *Medical & Biological Engineering & Computing*, 57 (8), 1693-1707. DOI: https://doi.org/10.1007/s11517-019-01978-z.

Wong, I.A., Chau, K.F., & Chan, H.U. (2023). An empirical study on customers' gambling intention in AI-supported casinos. *Journal of Hospitality and Tourism Technology*, Vol. ahead-of-print No. ahead-of-print. DOI: https://doi.org/10.1108/JHTT-08-2021-0225.

Wu, Q., Law, R., & Xu, X. (2012). A sparse Gaussian process regression model for tourism demand forecasting in Hong Kong. *Expert Systems with Applications*, 39, 5, 4769-4774. DOI: https://doi.org/10.1016/j.eswa.2011.09.159.

Yu, J., & Egger, R. (2021). Color and engagement in touristic Instagram pictures: A machine learning approach. *Annals of Tourism Research*, 89, 103204. DOI: https://doi.org/10.1016/j.annals.2021.103204.

Zaei, M. E., & Zaei, M. E. (2013). The Impacts of Tourism Industry on Host Community. *European Journal of Tourism Hospitality and Research*, 1 (2), 12-21.

Zhu, Y., Zhang, R.(R)., Zou, Y., & Jin, D. (2023). Investigating customers' responses to artificial intelligence chatbots in online travel agencies: the moderating role of product familiarity. *Journal of Hospitality and Tourism Technology*, Vol. ahead-of-print No. ahead-of-print. DOI: https://doi.org/10.1108/JHTT-02-2022-0041.

Received: 16/04/2023
Accepted: 11/07/2023
Coordinating editor: Alicia Orea-Giner