Bollenbach, Jessica; Neubig, Stefan; Hein, Andreas; Keller, Robert; Krcmar, Helmut

Enabling active visitor management: local, short-term occupancy prediction at a touristic point of interest

Information Technology & Tourism

Provided in Cooperation with:
Springer Nature

Suggested Citation: Bollenbach, Jessica; Neubig, Stefan; Hein, Andreas; Keller, Robert; Krcmar, Helmut (2024) : Enabling active visitor management: local, short-term occupancy prediction at a touristic point of interest, Information Technology & Tourism, ISSN 1943-4294, Springer, Berlin, Heidelberg, Vol. 26, Iss. 3, pp. 521-552, https://doi.org/10.1007/s40558-024-00291-2

Information Technology & Tourism (2024) 26:521–552
https://doi.org/10.1007/s40558-024-00291-2

ORIGINAL RESEARCH

Enabling active visitor management: local, short-term occupancy prediction at a touristic point of interest

Jessica Bollenbach[1,2,3] · **Stefan Neubig**[4,5] · **Andreas Hein**[6] · **Robert Keller**[7] · **Helmut Krcmar**[4]

Received: 9 February 2024 / Revised: 9 April 2024 / Accepted: 7 May 2024 /
Published online: 4 June 2024

Abstract

After the temporary shock of the Covid-19 pandemic, the rapid recovery and resumed growth of the tourism sectors accelerates unsustainable tourism, resulting in local (over-)crowding, environmental damage, increased emissions, and diminished tourism acceptance. Addressing these challenges requires an active visitor management system at points of interest (POI), which requires local and timely POI-specific occupancy predictions to predict and mitigate crowding. Therefore, we present a new approach to measure visitor movement at an open-spaced, and freely accessible POI and evaluate the prediction performance of multiple occupancy and visitor count machine learning prediction models. We analyze multiple case combinations regarding spatial granularity, time granularity, and prediction time horizons. With an analysis of the SHAP values we determine the influence of the most important features on the prediction and extract transferable knowledge for similar regions lacking visitor movement data. The results underline that POI-specific prediction is achievable with a moderate relation for occupancy prediction and a strong relation for visitor count prediction. Across all cases, XGBoost and Random Forest outperform other models, with prediction accuracy increasing as the prediction time horizon shortens. For effective active visitor management, combining multiple models with different spatial aggregations and prediction time horizons provides the best information basis to identify appropriate steering measures. This innovative application of digital technologies facilitates information exchange between destination management organizations and tourists, promoting sustainable destination development and enhancing tourism experience.

Keywords Visitor management · Tourism demand · Machine learning prediction · Sustainable tourism · Overcrowding

Extended author information available on the last page of the article

Abbreviations

DMO	Destination Management Organization
DST	Data Science Trajectories
LR	Multilinear Regression
POI	Point of Interest
RF	Random Forest
SHAP	SHapley Additive explanation
SVR	Support Vector Regression

1 Introduction

After the tremendous external shock of the COVID-19 pandemic on the tourism industry, the recovery is progressing rapidly (UNWTO 2023). Recovery times, however, exhibit variations due to country-specific restrictions and global dependencies (Škare et al. 2021). While Europe and the Middle East demonstrate a fast recovery characterized by substantial intraregional demand, Asia recovers rather slowly (UNWTO 2023, 2022). This growth in Europe and the Middle East beyond 2019 levels and potential additional peaks due to a shift in tourism demand from Asia may lead to exceptionally high tourist numbers.

Despite all the benefits, such as economic growth and employment opportunities, this trend of continuous growth also drives unsustainable tourism, including overtourism and local overcrowding (Butler 2018a; Mihalic 2020). Overcrowding, characterized by the interplay between utilization and environmental damage (Monz et al. 2013), manifests in changes such as alterations in vegetation, shifts in wildlife behaviour, compromised water quality, and elevated levels of noise and air pollution (Liddle 1997; Newsome et al. 2012; Wall 2019). In addition to environmental impacts, overcrowding leads to social effects, such as cultural change and rising conflicts between different interests (Koens et al. 2018; Spenceley et al. 2015). Furthermore, the impact extends beyond local POIs, given that emission-emitting vehicles are still the predominant mode of transport. Thereby, overcrowding causes increased congestion and extended search times for parking (Paidi et al. 2022), leading to rising emissions and, thus, accelerating climate change. Regarding location, overcrowding primarily occurs at freely accessible tourist destinations and POIs without access restrictions, such as beaches or summits. The transition from crowding to overcrowding lies at the intersection of nature, social impacts, and politics and can only be determined locally for each individual POI. Therefore, without effective management measures, areas with frequent crowding may become hotspots for the detrimental effects of overtourism and overcrowding.

The United Nations already demanded in 2016 the transformation towards more sustainable tourism as part of the 17 sustainable development goals to address these issues (DESA 2016). Consequently, POIs must actively embrace transformation towards more sustainable and smart tourism (Tauber and Bausch 2022), focusing on implementing active visitor management strategies (Schmücker et al. 2022). Effective active visitor management entails the spatial distribution of visitors among nearby POIs and strategically times arrivals to prevent overcrowding

at any single location (Hall and McArthur 1996; Mason 2005). However, triggering suitable steering measures by the active visitor management system, once a problematic benchmark is met, requires a smart occupancy prediction system. Such an occupancy prediction system, enabled by continuous monitoring, predicts peak demand situations and serves as a fundamental component for making data-driven decisions and adapting to rapidly changing external factors (Archer 1987; Zelenka and Kacetl 2013). Occupancy prediction is covered by the field of tourism demand prediction, also named tourism flow or visitor monitoring, which is a well-established research topic (Sun et al. 2023; Ghalehkhondabi et al. 2019; Song et al. 2019; Song and Li 2008). However, existing literature primarily focuses on predicting tourism demand for larger regions and rarely addresses specific POIs. The studies that include a particular POI do not incorporate the actual occupancy at open-spaced, publicly freely accessible POIs, which are at high risk of overcrowding. By an open-spaced POI, we understand a site that does not include a clear boundary allowing people to move freely and widely. Further, the studies exhibit a large time granularity with, for example, weekly or monthly prediction and rarely exhibit daily or hourly prediction. Moreover, although machine learning models have gained more attention in recent years, promising gradient-boosting models are seldom considered. In summary, the current state of tourism demand prediction analysis falls short of enabling active visitor management, which requires an occupancy prediction of open-spaced and freely accessible POIs with low time granularity within one day.

The identified research gap in predicting tourism demand raises the following questions:

RQ1. *How do various prediction models perform in predicting occupancy, especially peak occupancy, at open-spaced and freely accessible POIs to enable active visitor management?*
RQ2. *What impact does search query data have on the prediction performance?*
RQ3. *How do the individual features influence the predicted value?*

We aim to address the research questions by introducing a novel method for collecting detailed visitor movement data and leveraging digital technologies to predict short-term occupancy at an open-spaced POI. Therefore, we conduct a case study focusing on beach occupancy at the Bay in Lübeck, Scharbeutz located on the Baltic Sea in northern Germany. We shed light on the optimal time and spatial granularity, as well as the optimal time horizon and compare the prediction quality of various ML models with a persistence model and an ARIMA model. Further, we compare the usability of visitor count prediction in contrast to POI occupancy prediction in the context of active visitor management. Additionally, for the most promising ML models, we examine whether incorporating search query data enhances model accuracy. Moreover, to uncover the most influential features and comprehend their impact on occupancy prediction, we utilise SHapley Additive exPlanation (SHAP) values. Our contribution to the current state of research lies in the development, implementation, and analysis of a novel

approach to locally-specific and timely fine-grained prediction models These models hold significant relevance for tourist-related businesses, policymakers, and decision-makers as they pave the way for sustainable development through digitalization (Bi et al. 2022; Rodrigues et al. 2023). By facilitating an improved flow of information between tourists and destination management organizations (DMOs), our work enables data-driven, localized active visitor management, thereby fostering sustainable destination development and offering direct recommendations for practical implementation.

The present paper is structured as follows: In Sect. 2, we provide an overview of related literature concerning overtourism, overcrowding, and active visitor management. We also discuss existing methodologies for tourist demand prediction and pinpoint research gaps. Section 3 outlines the methods employed in our study, along with a detailed account of the case study conducted, with the corresponding results presented in Sect. 4. We proceed to interpret and discuss these findings in Sect. 5. Finally, Sect. 6 offers conclusions along with limitations and suggests an outlook for future research.

2 Background and related work

In the following sections, we introduce the core concepts and related work, including the reasons, effects, and mitigation approaches of overtourism and overcrowding. Furthermore, we present related research on occupancy and tourism demand prediction as the enabler and partly solution for advanced systems for (digital) active visitor management.

2.1 Overtourism and overcrowding

The phenomenon of overtourism and overcrowding typically occurs within tourist destinations, defined as a collection of tourism related services, products, and facilities (Hu and Ritchie 1993). Tourist destinations are typically managed by DMOs, which handle the internal destination development (Sheehan and Ritchie 2005). Although this includes the development of sustainable tourism that meets the current needs of tourists without diminishing future opportunities (UNWTO 1998), DMOs have often prioritized fast economic growth resulting in an uncontrolled tourism growth (Séraphin et al. 2019; Butler and Dodds 2022). Due to insufficient implementation of sustainability-oriented tourism, the non-scientific community came up with the term *overtourism* to describe the negative impacts stemming from the constant growth of tourism (Ali 2016, 2018). While this term was quickly picked up by the scientific and economic community, research focused on urban areas rather than rural or natural areas, such as the here considered open-spaced POI of a beach (Butler and Dodds 2022; McKinsey&Company 2017; Mihalic 2020). However, nature-based tourism enjoys great popularity (Balmford et al. 2009) and can impose a heavy burden on the sensitive environment, even at seemingly manageable visitor numbers (Butler 2018b). Additionally, the global relevance of social media

can significantly highlight individual POIs and lead to overcrowding, while comparable POIs experience only moderate utilization (Gretzel 2019). Hence, even if the underlying problems of overtourism date back to the 1960s (Koens et al. 2018) the new framing allows for alternative perspectives and an increased focus on rural areas (Butler 2018b; Capocchi et al. 2020; Wall 2019). As Wall (2020, 2019) further argues, overtourism is beyond the concept of carrying capacity as it depends on individual expectations, activity types, POI infrastructure, and impact on social, environmental and economic structures. These extensive perspectives combined Mihalic (2020) in a holistic framework to connect the pillars of sustainability with accountable stakeholders to enable the identification of potential risks to POIs including overtourism. Closely related to overtourism is the concept of *overcrowding*, which refers to the temporary accumulation of people rather than the long-term problematic development of unsustainable tourism (Butler 2018a; Oklevik et al. 2019). In tourism, human crowding (i.e., limited space) and physical crowding (i.e., limited activities) are particularly relevant (Yin et al. 2020). Similar to overtourism, overcrowding is concerned with the subjective perception of the situation and its measurable impacts, rather than relying on an absolute measurement of carrying capacity (Dogru-Dastan 2022; Wall 2019). The perception of crowding depends on the travel circumstances, whereby particularly among non-guided day visitors (Kalisch and Klaphake 2007) who independently plan their trips to natural POIs known for their remoteness (Jacobsen et al. 2019; Oklevik et al. 2019), demonstrate a low tolerance for crowdedness.

Because of the close connection between overtourism and overcrowding, they have similar impacts on a tourism destination, and the two terms are often used interchangeably. According to Wall (2019), one can differentiate between scientifically verifiable physical impacts and individually perceived impacts. The physical impacts can be classified into five categories (McKinsey&Company, 2017): (1) conflict with local residents, (2) degraded tourist experience, (3) overloaded infrastructure, (4) environmental damage, and (5) threats to culture and heritage. For nature-based POIs the environmental damage due to pollution, exploitation of natural resources, waste, and disturbance of wildlife (Liddle 1997; Monz et al. 2013; Newsome et al. 2012; Wall 2019) is next to overloaded infrastructures such as waste, water and energy management the most critical factor. Turning to the perceptual consequences of overcrowding and overtourism, negative associations dominate such as stress (Popp 2012), reduced destination attractiveness (Yin et al. 2020), low revisit intentions (Oklevik et al. 2019), reduced feeling of uniqueness (Jacobsen et al. 2019), and overall decreased tourist experience (Dogru-Dastan 2022).

To mitigate the harmful effects of overtourism and overcrowding, implementing visitor management measures is crucial for distributing tourists both temporally and spatially (McKinsey&Company 2017; Zelenka and Kacetl 2013). If there are already too many visitors at the POI, hard measures are no longer sufficient. Instead, preventive measures with close monitoring and active visitor management are required specifically at open-spaced natural POIs. Schmücker et al. (2022) distinguish active visitor management from general visitor guidance by emphasizing that it should rely on soft measures. Combined with digital technologies, these soft measures such as recommendation systems provide often untapped potential to effectively prevent

overcrowding before it even occurs (Spenceley et al. 2015; Veiga et al. 2018). The utilization of digital technologies enables DMOs to gather a higher flow of information from tourists and engage in direct communication to enhance visitor experience and sustain POIs (Rodrigues et al. 2023). Effective visitor management, however, relies on local data and appropriate analytics (Neubig et al. 2022). Visitors are more likely to follow alternative suggestions, which match their interests, such as sports, recreation, or sightseeing, if they trust the occupancy predictions. Our approach predicts the occupancy at an open-spaced and freely accessible touristic POI, laying a data foundation for DMOs to determine potential overcrowding situations. Further, it enables the development of novel, digital visitor management strategies and information systems that automatically initiate steering measures before overcrowding even occurs and promote sustainable, long-term economic growth.

2.2 Occupancy prediction

Within the scope of this work, we define occupancy prediction as the task of predicting the occupancy of a certain geographic area. Such an area may be (i) open-spaced and public (i.e., freely accessible from a wide range of entries, such as beaches) or (ii) closed (i.e., only accessible from a limited number of controlled entries, such as a building). This distinction is crucial since, in closed areas, a complete measurement of occupancy is possible by employing appropriate sensors at the access points. In contrast, occupancy in open-spaced and public areas can generally not be measured this way. Notably, this distinction is not strict, as some areas can fall into both categories (e.g., the beach of Scharbeutz, which has both controlled and non-controlled entries). Within the scope of this work, we consider such border cases as open-spaced since it does not allow for fully controlled monitoring.

Research on occupancy prediction has grown rapidly since 2006 (Liu et al. 2019) and can be classified among different dimensions, including (i) data sources, (ii) prediction models, and (iii) spatio-temporal granularity. Previous work found that the number of available data sources is manifold and ranges from one-dimensional historical data that captures a certain timespan (e.g., tourist arrivals (Abu et al. 2021; Kim et al. 2021)), parking (Chawathe 2019), booking data and tickets (Phumchusri and Suwatanapongched 2021; Attanasio et al. 2022)) to the use of supplementary data, including (hotel) pricing (Tsang and Benoit 2020), weather and public holiday information (Bi et al. 2021), as well as several (behavioral) online data (e.g., Dinis et al. 2019; Önder et al. 2019; Volchek et al. 2019; Wu et al. 2017). Regarding prediction models, previous research (e.g., Jiao and Chen 2019; Song et al. 2019; Wu et al. 2017) mostly classifies occupancy prediction models into (i) time-series models, (ii) econometric models and (iii) ML models. However, as argued by Bollenbach et al. (2022), some ML models have been surprisingly rarely used in occupancy prediction tasks such as gradient boosting with XGBoost, which has shown promising results in similar applications (Bollenbach et al. 2022).

However, in the context of active visitor management, spatiotemporal granularity is arguably the most important property of occupancy prediction. Therefore, to classify existing occupancy prediction approaches among the combination of

these dimensions, we conducted a systematic literature review (Webster and Watson 2002), primarily based on the Scopus database. We included peer-reviewed literature that describes complete approaches for occupancy prediction and related dimensions of time and space (i.e., the temporal and spatial granularity of the performed predictions). On the contrary, we excluded literature that does not contain a full description and comprehensive evaluation of the underlying approach, was not written in English, or was otherwise not accessible during the review. To reflect these requirements and keep our investigation at a reasonable scope, we directly included keywords referring to time and space in our search string. Thus, to ensure that the literature does not explicitly mention 'point of interest' or 'POI,' we added popular POI categories, exemplarily derived from Outdooractive, one of Europe's largest outdoor platforms (Outdooractive 2022). Table 1 summarizes the targets and associated search strings of our literature review.

From a total of 258 results, we selected 116 articles based on our inclusion and exclusion criteria mentioned above. We further classified these papers into the granularity of their space and time dimensions and whether we considered the respective area closed or open-spaced. In our findings, we generally observed that the number of approaches is quite limited for fine temporal or spatial granularities (see Fig. 1). Regarding the time dimension, monthly and seasonal granularity are the most frequently used prediction times, whereas week is less frequently used than day. This may be because the day of the week plays an important role in tourism, as weekends are usually significantly more crowded than weekdays. Similarly, we observed that the year is a less frequently investigated prediction period, possibly due to its limited expressiveness. Regarding the spatial dimension, an interesting exception occurs for closed areas (i.e., hotels and parking lots), which have been more frequently regarded than open-spaced POIs, cities, or regions. Most likely, however, this is because closed areas are much easier to analyze due to the availability of clear measurement points and (often) booking data. Therefore, in Fig. 1, we distinguish between such closed and open-spaced POIs.

In addition to these separate perspectives, the combination of both dimensions is particularly interesting within the scope of active visitor management. For this purpose, we consider only open-space areas and illustrate our results in Table 2. This confirms what we have already obtained from the separate considerations Especially for the occupancy prediction of open-spaced POIs in temporally fine granularity, only a scarce number of literature exists.

In this work, we focus on occupancy prediction within the scope of an exemplary open-spaced POI. Here, we investigate its predictive ability and compare different models, including promising models that have only rarely been used in this context (e.g., gradient boosting). We contribute to the existing body of literature by bridging the gap between active visitor management and open-spaced POI occupancy prediction. Furthermore, we focus not only on the actual predictions but also on their explainability using SHAP values. The next section describes the underlying use case in more detail.

Table 1 Search string composition, results and selected papers

Target		Search String Composition (Title, Abstract, Keywords)
Domain		Tourism
	AND	Occupancy OR demand
	AND	Prediction OR forecasting
Time	*AND*	Second* OR minute* OR hour* OR day* OR week* OR month* OR season* OR year*
Space (General)	*AND*	"Point of interest" OR poi* OR region* OR countr* OR destination* OR town OR city OR cities OR village*
Space (POI categories)		OR accommodation OR apartment OR bar OR bay OR beach OR bistro OR brewery OR cafe OR castle OR chapel OR church OR disco OR distillery OR garden OR hostel OR hotel OR hut OR lake OR motel OR opera OR park OR parking OR pizzeria OR pub OR restaurant OR sauna OR spa OR theater OR trail OR viewpoint OR vinotheque OR winery
Total Results		258
Selected Papers		116

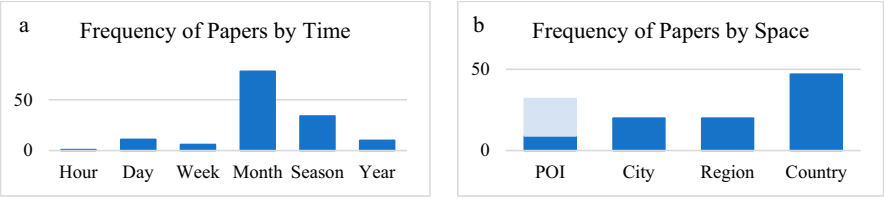

Fig. 1 Frequency of papers by time and space. Granularities smaller than 1 h are subsumed within the hour column

Table 2 Frequency of publications based on which spatial and temporal granularities they cover

	Hour	Day	Month	Season	Year
Country	0.00%	0.88%	26.55%	18.58%	2.65%
Region	0.88%	1.77%	11.50%	5.31%	2.65%
City	0.00%	2.65%	12.39%	3.54%	2.65%
POI	0.00%	0.88%	5.31%	0.88%	0.88%

3 Methods and case study

To develop, evaluate, and interpret multiple prediction models in the context of active visitor management, we use the Data Science Trajectories (DST) map by Martinez-Plumed et al. (2019). The DST-map is an evolution of the classic industrial standard CRISP-DM. It allows for a more flexible, customizable methodology with more detailed steps, while incorporating most of the classic elements of CRISP-DM (Chapman et al. 2000). Figure 2 illustrates the steps considered in

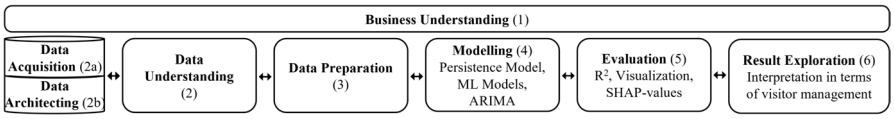

Fig. 2 Methodology of a DST-map to develop, evaluate, and interpret multiple occupancy prediction models

the DST-map for this study. Business Understanding (1) overlaps with the underlying steps, as each decision should be addressed in the specific context of the research question. In Data Understanding (2), we start with the gathering of the data, which is divided into the subparts Data Acquisition (2a) and Data Architecting (2b). Data Acquisition covers the generation of visitor movement data, and Data Architecting (2b) covers the collection and compilation of possible influencing factors. Next, in Data Preparation (3), we prepare the data for Modelling (4). Finally, we evaluate the results (5) using R^2, several visualizations, and identify the most important features using SHAP values to interpret the results in the context of local active visitor management and provide recommendations for action.

3.1 Business understanding

The case study examines the Bay of Lübeck in Scharbeutz, located on the Baltic Sea in northern Germany. The distinct climate of the Baltic Sea and its enclosed formation provide an ecological niche fostering high biodiversity worth protecting to preserve sustainability (Blicharska et al. 2019; Ducrotoy and Elliott 2008). Given that the North and Baltic Sea rank among the most popular tourist destinations for the German population, with a steadily rising trend (Arbeitsgemeinschaft Verbrauchs- und Medienanalyse 2022), the intensive use of the beaches deteriorates both flora and fauna (Schierding et al. 2011). Particularly noteworthy is the trend in tourist arrivals to Schleswig–Holstein in recent years, encompassing the challenges posed by the Covid-19 pandemic. Unlike other destinations, tourist arrivals in 2020 and 2021 nearly reached the levels of 2019 and even surpassed them in September and October. In 2022, a full recovery has already been achieved followed by further growth in 2023 (Statistisches Amt für Hamburg und Schleswig–Holstein 2023). The Baltic Sea is renowned for day trips, given its proximity to larger cities such as Hamburg, Kiel, or Lübeck, as depicted in Fig. 3. Consequently, day visitors account for 2/3 of total visitors (dwif e.V., 2016, 2019). Unlike overnight guests, available information about day travelers is relatively scarce, leading to unpredictable peaks in beach occupancy. To address the information gap concerning expected beach occupancy, laser sensors have been installed at the entrances to the beach in Scharbeutz. These entrances, illustrated in Fig. 3, are distributed along the entire beach section, with a notable concentration near the city. The installed sensors count people entering and leaving the beach, enabling the calculation of beach occupancy. Currently, this information on present occupancy is digitally disseminated to prospective tourists

Fig. 3 Overview of the location of the analyzed overcrowded Scharbeutz beach and the location of entrances with installed overhead laser sensors

(Tourismus-Agentur Lübecker Bucht 2022). Yet, beyond current occupancy data, implementing a predictive model is imperative for active visitor management, allowing the DMO to anticipate overcrowding and introduce steering measures in advance.

We compare various cases in the prediction model development to pinpoint the optimal configuration to enable active visitor management. Figure 4 illustrates the individual case elements, which we combine in multiple prediction models to identify the optimal composition of elements for active visitor management. In predicting *visitor movements*, we differentiate between two variations: the visitor count prediction, which merely accounts for the entering people; and the occupancy prediction, which considers the beach occupancy. We further compare the performance of two different temporal aggregations, called *time granularity*, with 4-h and 24-h timesteps because a higher aggregation may result in better predictions for a longer prediction time horizon (Table 6). The *spatial granularity* refers to the spatial segmentation of the POI, where we consider entrances, beach sections, and the beach. In addition, we

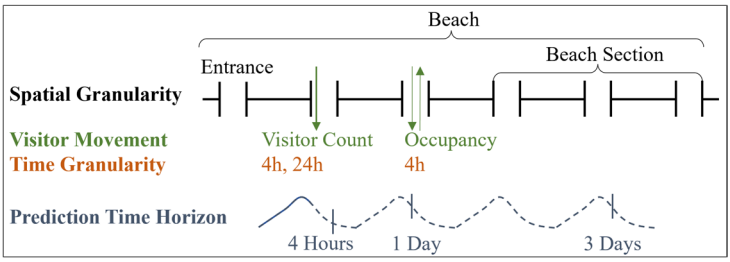

Fig. 4 Differentiation of visitor movement, spatial granularity, time granularity, and prediction time-horizon

compare three different *prediction time horizons*: four hours, one day, and three days ahead, i.e., how far in advance the visitor movements are predicted.

3.2 Data understanding and data preparation

Data Acquisition of historical visitor movement data was carried out from mid-October 2020 to April 2022 with overhead sensors at 20 beach entrances of Scharbeutz by counting the entering and leaving people in 15-min timesteps. Visitor movements indicate a yearly and daily seasonality with a higher beach occupancy during the day and in the summer months. Furthermore, irregular peaks occur throughout the year, leading to exceptionally high occupancy. For data preparation, we replace outliers and missing values since single values cannot be deleted to preserve an uninterrupted time series. We first identify outliers by a threshold value for each entry, and second, replace the outliers with the average visitor movement during the last seven days of the considered time slot and spatial granularity. Afterward, we replace missing values, which either indicate that the sensor has a defect, or no movement has taken place, and therefore, no value was recorded. We assume a sensor defect if no value is recorded for more than 16-h within one day and replace the missing value with the average visitor movement over the last four weeks of the considered time slot and spatial granularity. Below the 16-h threshold we set the missing value to zero. To obtain the beach occupancy, we subtract the leaving people from the entering per time slot and sum it up over the day. However, there is some error in the calculated occupancy: since the beach is an open-spaced area and only official entrances at the beach in Scharbeutz are equipped with overhead sensors, some people use different entrances for entering and leaving, whereby one might not have a sensor installed. Additionally, some people tend to use unofficial paths that are not equipped with sensors. To compensate for this error, the occupancy is reset to zero each night at 3 a.m. UTC, as this is the time with the lowest occupancy, including New Year's Eve. Subsequently, we aggregate the prepared entering visitor count data to a time granularity of 4-h and 24-h timesteps and the occupancy data to a time granularity of 4-h timesteps. Due to the daily seasonality, aggregation of occupancy data into 24-h timesteps would provide no information. To obtain the spatial granularities, we aggregate the visitor movement data of each time granularity to 8 beach sections and the entire beach.

When collecting the influencing factors in **Data Architecting**, it must be emphasized that each input feature of interest must be available both in the past and future within the prediction horizon. Historical values are required for training, evaluating, and testing the developed ML model, whereas future values are required for real-world deployment.

We include *time-related features* to identify the effects driven by the daily and yearly seasonality. Given the cyclic nature of time-related features, feature engineering must reflect the real-world interpretation of time within the features. In this case, cyclicity means that the distance between each timestep should be equal. However, with increasing linear values, for example, from 1 to 365 for days, the step from 364 to 365 is smaller than from 365 to 1. To solve this problem, we follow Chakraborty

and Elzarka (2019) and transform increasing linear values into cyclic sinus and cosinus values representing the coordinates of a circle. Hence, the distance between each timestep remains equal, improving the prediction performance with more accurate features.

The *holiday-related features* capture the free time of the population, as most tourist activities occur during the leisure time of the population. Besides features about the weekend, public holidays, and bridging days, school holidays are included, since a large part of the population depends on the holidays. Due to the high proportion of day tourists, Schleswig–Holstein's local school holidays are included as an explicit feature, as short travel distances lead to increased visits. To cover the demand from overnight visitors, we include the school holiday density in Germany, which reflect the proportion of federal states with holiday.

Weather greatly influences outdoor tourism activities, because roughly 80% of beach visitors check the weather forecast before starting their journey (Becken 2013; R.-Toubes et al. 2020). Further, it has already been shown that integrating weather data as features into a tourism demand prediction model can increase the predictive power of ML models (Álvarez-Díaz and Rosselló-Nadal 2010). We collect data on temperature, wind speed, precipitation intensity, and precipitation form provided by the German Weather Service for historical weather data, as well as weather forecasts for up to ten days (German Weather Service 2024). To ensure a good performance of the prediction models, we follow Studer et al. (2021) and convert continuous raw data values into categorical features. The categorization of temperature, wind speed, and precipitation intensity follows the standards of the German Weather Service

Table 3 Overview of all input features after feature engineering

	Features	Attribute	Value Ranges	# features
Historic values	Lagged historic values	Float	[0, …, max(visitor count or occupancy)]	18
Time-related	Year	Integer	[2020, 2021, 2022]	1
	Day of year sin/ day of year cos	Float	[0, …, 1]	2
	Hour sin/ hour cos	Float	[0, …, 1]	2
	Day of the week	Binary	[0, 1]	7
Holiday-related	Public holiday (bank holiday)	Binary	[0, 1]	1
	Bridging day	Binary	[0, 1]	1
	Regional school holiday (SH)	Binary	[0, 1]	1
	German school holiday density	Float	[0, …, 1]	1
Weather	Wind category	Binary	[0, 1]	13
	Rain category	Binary	[0, 1]	6
	Temperature category	Binary	[0, 1]	8
	Precipitation form	Binary	[0, 1]	3
Search Query	Google Trends (lag 1 day)	Integer	[0, …, 100]	1
	Google Trends (lag 7 days)	Integer	[0, …, 100]	1
	Google Trends (lag 14 days)	Integer	[0, …, 100]	1
				67

(German Weather Service 2023a, 2023b, 2023c). The precipitation form is created via feature engineering because the input values of the historical observation and the forecast differ and need to be combined into unified categories. Further, we transform each categorized weather feature via one hot encoding into binary features, resulting in 30 weather features (Tables 7, 8).

Lagged historical values of the considered time series are essential for the prediction since they provide up-to-date information about the visitor movement within the preceeding days. The prediction model, initially trained on seasonal data from a year ago, can be fine-tuned to current trends by incorporating lagged historical values as features. Creating the feature of lagged historical values involves determining both the lag size and the number of lagged historical values (Bi et al. 2022). The lag size depends on the prediction time horizon since only existing historical values can be used as a feature. For instance, with a prediction time horizon of one day, the first lagged historical value must be from at least one day ago. Considering the prediction time horizons in this analysis—four hours, one day, and three days—the lag size spans four hours, one day, and three days. To determine the number of lagged historical values, Liu et al. (2022) propose a higher-lag order because this is likely to improve the forecast accuracy, by which at least one complete seasonal cycle should be included (Bi et al. 2022). Given the daily seasonality, we integrate the values of the last three days as lagged historical values, starting with the previously defined lag size.

Search query data holds the potential to reveal upcoming trends and current public interest, as many tourists rely on the internet to gather information about POIs and tourist attractions (Choi and Varian 2012). Thus, the integration of search query data may enhance the predicting performance but should be evaluated in detail for each use case (Lazer et al. 2014; Li et al. 2021; Önder et al. 2019). We integrate Google Trends data as an additional feature for the best predicting models and compare if it enhances predicting accuracy (Google Trends 2023). Google Trends data contain a search query index for given keywords based on the volume of queries within a defined geographical area (Choi and Varian 2012). The chosen keyword is 'Scharbeutz' because it showed by far the highest relative impact compared to other related keywords such as 'Scharbeutz Strand', 'Scharbeutz Tourismus' or 'Scharbeutz beach' which demonstrate the same trend but with a lower impact. For our POI prediction model we require at least daily data over the entire time horizon of the gathered visitor movement data over 1.5 years. However, raw Google Trends data do not meet this criterion as they lack daily data over an extended period, failing to capture long-term trends through aggregation (Eichenauer et al. 2022). Eichenauer et al. (2022) addressed this issue by developing a method, implemented as an R package, that provides frequency-consistent daily data over an extended period by combining available daily, weekly, and monthly Google Trends data. The obtained daily Google Trends data require further data engineering due to the lag between search time and POI visit (Choi and Varian 2012). This lag is identified by calculating the Pearson correlation between daily Google Trends data and the daily visitor count for the entire beach (Bi et al. 2020). Lags of 7, 14, and 1 day indicate the highest correlation and are integrated as a feature accordingly. Note that the Google Trends feature with one-day lag is only included in models with a prediction time horizon

equal to or below one day. An overview of all input features of the prediction models is provided in Table 3.

3.3 Modelling

We compare four different ML models with a persistence model as a benchmark to determine the best model for predicting beach occupancy to enable active visitor management. For the most important case combinations, we additionally analyze the SARIMA model as it is one of the most often used models for time-series analysis. An overview of the models is presented in Table 4.

For implementing the ML models, we rely on the Python packages of the ML scikit-learn library. Hence, to train and test the prediction models, we split the data on the first of February 2022 into a training and testing data set. To implement the LR model, we directly used the Linear Regression package of scikit-learn, which does not require further parameter tuning. To implement the SVR model, we first determined the optimal kernel by comparing the linear, poly, and rbf one by randomly selecting and comparing multiple case-combinations. We utilize the linear kernel for the final SVR model since it demonstrated the best predicting performance. The advanced ML models XGBoost and RF require additional tuning to improve the prediction quality and avoid overfitting. Thus, we adjust the maximum depth and number of decision trees by tuning the corresponding hyperparameters. To perform the hyperparameter tuning, we use grid search, including cross-validation. The possible parameters for the maximum tree depth are [3, 4, 5, 6, 7, 8, 9, 10] and [20, 30, 40, 50, 75, 90, 100] for the number of estimators. The here in the grid search included possible parameters are determined with an analysis over several case-combinations. Thus, we ensured never to reach the outer edge of the possible hyperparameter values. For cross-validation, we use the specified version of time-series cross-validation since time-series data cannot be split randomly. Time-series cross-validation ensures retaining the underlying sequential structure by continuously adding new blocks at the end of the time-series and evaluating the performance gradually. We performed a threefold cross-validation over November, December, and January and determine the optimal hyperparameters by the best average performance over all folds. The stochastic SARIMA model is implemented with the Python pmdarima package. Here, the required parameters were determined using the auto_arima function to identify the best parameters for each case-combination. We predetermined the seasonal parameter m to 6 to represent the daily seasonality with 6 periods per day à 4 h. The parameters p, P, q, Q were optimally chosen in the range from 0 to 5 and d, D in the range from 1 to 5 by the auto_arima function. We mimic the prediction time horizon of the ML models with lagged historical values as limiting input variables to ensure comparability between the models. For this purpose, we retrain the model after each prediction by adding the actual visitor movement to the training data set. To obtain the results of the different prediction time horizons, we extract the predicted values with a lag of four hours, one day, and three days from the predicted time series. In total, we analyze 1406 case and prediction

Table 4 Description of the here considered prediction models

Prediction Model	Description
Persistence model	To evaluate the perfomance of the prediction models, we create a persistence model that serves as benchmark. The persistence model takes the last available historical value of the same time slot as a prediction. Hence, depending on the prediction time horizon, the historical value is either one day or three days ago
Multilinear Regression (LR)	Multilinear Regression is a statistical model that describes the linear relationship between one dependent variable and multiple independent variables. Thus, it is an extension of Linear Regression that only describes the relationship between one dependent and one independent variable (Allison, 1999)
Support Vector Regression (SVR)	The support vector regression model is a statistical model based on a kernel, including mathematical functions. During the training process, a hyperplane is fitted to the data points. This hyperplane is located precisely in the middle of the threshold values, which contains the largest proportion of all training data points (Awad & Khanna, 2015; Vapnik, 2000)
Random Forest (RF)	The Random Forest model is an ensemble model of decision trees. Various low-correlated trees are generated by combining bagging with random feature selection. The prediction is accomplished by averaging over the results of all trees (Breiman, 2001; scikit-learn developers, 2022)
XGBoost	The XGBoost model is a system-optimized implementation of extreme gradient boosting which represents an ensemble of multiple decision trees (xgboost developers, 2022). To obtain a robust model, the ensemble of decision trees is gradually improved by adding new weak learners, reducing the loss (Friedman, 2001)
SARIMA	The SARIMA model is an extension of the ARIMA model to cover seasonal time series data. SARIMA is an extensively used model for seasonal time series data that fits the model only based on historic time series values with the Box-Jenkins method (Box et al., 2015)

model combinations to determine the optimal model for beach visitor movement prediction.

3.4 Evaluation

While the model is trained with the training dataset, the testing data set is required for the evaluation of the prediction models. To ensure a more detailed evaluation, we split the testing data set into the months February, March, and April. Afterward, we individually assess the prediction performance for each month. Following the methodology outlined by Chicco et al. (2021), we employ the coefficient of determination, denoted as R^2 (1), as the evaluation metric for regression analysis. R^2 is the recommended metric because other metrics such as root mean squared error (RMSE) or mean absolute error (MAE) lack comparability across multiple models with different underlying data, i.e., the here evaluated cases. R^2 values span from $-\infty$ to 1, with

1 corresponding to a perfect match between the prediction and the actually observed value. A negative R^2 value may occur when a model, trained on a specific portion of the dataset, is tested on a separate dataset. In the context of the training data, the calculation of R^2 follows the principle that the Explained Sum of Squares (ESS) plus the Residual Sum of Squares (RSS) equals the Total Sum of Squares (TSS), resulting in R^2 values between 0 and 1. However, when the model encounters new data, i.e., the testing dataset, the assumption that the model's errors of predictions (RSS) will remain proportionally smaller than the dataset's inherent variability (TSS) is not guaranteed, leading to unbounded negative R^2 values. An R^2 value of 0 indicates that the predictions are, on average, as accurate as a naive model that would always predict the mean value from all observations. Since an unbounded negative range hinders the interpretability of how much worse one model is compared to another and negative values of R^2 represent a worse fit than the simplest baseline model, the range from 0 to 1 is typically used for evaluation. According to Moore et al. (2013) R^2 values greater than 0.3 indicate a weak relationship, those exceeding 0.5 denote a moderate relationship, and values surpassing 0.7 signify a strong relationship. Due to the missing lower bound, negative values cannot be evaluated.

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{m} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{m} (\overline{y} - y_i)^2} \tag{1}$$

In addition to the evaluation metric R^2, we evaluate the results with a graphical analysis since it is crucial that the models identify the peak visitor movement. On top of identifying the most suitable prediction model, interpreting the underlying effects is highly important. Only if the most important influencing factors are known a successful selection of steering measures is possible. Therefore, we follow Lundberg and Lee (2017) by applying the explainable AI approach SHAP values which we implement with the Python package shap (Lundberg 2018).

4 Results

To address RQ1 and compare prediction models across various case-combinations, we determine the key components of a good prediction model for active visitor management. The model should be capable of identifying the peaks of visitor movements in advance and thus don't over- or underestimate those peaks too much. On the contrary, slight deviations from the daily variability are not problematic, because no steering measures must be initiated during a normal visitor movement. Due to the large number of 1406 cases, we present the effect of each case type individually and gradually identify the best case-combination and prediction model.

Regarding *visitor movement*, the visitor count prediction displays a better predictive performance, up to a strong relation with higher R^2 values than the occupancy prediction. The occupancy prediction is slightly worse with a moderate relation, as it is more challenging. The underlying reason is that people tend to use different

entrances for entering and leaving, as well as unofficial pathways leading to more noisy data. For the beach, with a four-hour time granularity, the R^2 measures of all case-combinations are displayed in Fig. 5.

Contrary to our expectations, aggregating the visitor count data into a *time granularity* of 24-h time steps did not lead to better prediction quality at a longer prediction time horizon. Thus, we conclude that the loss of information due to higher aggregation outweighs the reduction in timesteps, and data should only be aggregated to the most necessary.

Regarding the *spatial granularity*, the predictions of the aggregated values (beach sections, beach) display low up to strong relations, depending on the specific case-combination. However, single-entry predictions demonstrate high volatility in the R^2 measures, with occupancy predictions often indicating a weak or non-existent relationship, as evidenced by R^2 values below zero at various entrances. Also, the R^2 measures of the visitor count predictions drop below zero for cases with a longer prediction time horizon. Although single entrance predictions are not applicable in active visitor management, this analysis identify difficult-to-predict entrances whose sensor may need to be reviewed.

The *prediction time horizon* strongly influences the prediction accuracy of all case-combinations. As the prediction time horizon increases, the prediction worsens because less information about the visitor movement in the last time steps is known. This trend applies to all prediction models. However, the SARIMA models display a particularly marked deterioration since the historical time series values are the only input feature. The only exception to this downward trend, displayed in Fig. 5, is the persistence model, which in some cases, performs better at a longer prediction time horizon due to random matching by simply shifting historical values.

By *comparing the prediction models*, the intelligent prediction models perform better than our benchmark persistence model and, thus, provide higher value than the mere historical data. However, between the models there is still a remarkable

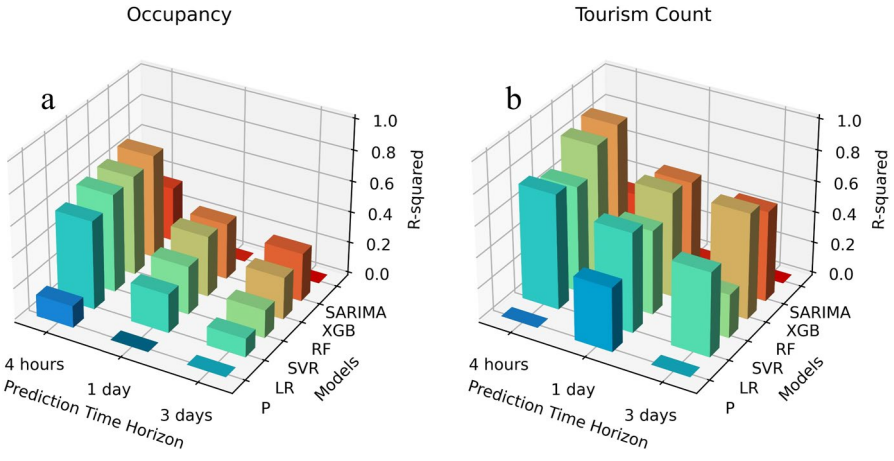

Fig. 5 R^2 measures for the visitor movement, both occupancy **a** and visitor count **b**, on the beach across all model types and prediction time horizons with a time granularity of 4 h

difference. The SARIMA model's mere reliance on historical data causes it to lag behind the actual visitor movement which worsens with an increasing prediction time horizon providing no applicable information. For instance, Fig. 6 illustrates this with a one-day prediction time horizon, where a peak on 27 March is inaccurately predicted to occur on 28 March. The SVR and Linear Regression models perform quite similarly in that both anticipate the daily trend but fail to recognize or severely underestimate visitor movement peaks. In both models, this effect can be explained by the inherent logic. The SVR model's defined hyperplanes prevent the prediction of tourism demand peaks, while the Linear Regression model's strictly linear relations fall short in explaining the relations between variables. XGBoost and Random Forest models perform comparably well, as presented in Table 5, effectively predicting peaks, such as the one on 27 March, with minor but acceptable over- or underestimation. The superiority of XGBoost or Random Forest varies with different case combinations, necessitating case-specific analysis for optimal deployment. The irrelevant negative visitor movement peaks, such as those at the end of 27 March, are not identified at a higher prediction horizon. Therefore, the R^2 measure deteriorates slightly, even if the prediction is actually good. Despite similar results, the XGBoost model impresses with a high computational speed, which is relevant for high data volumes or frequent retraining.

To investigate and try to answer RQ2 we integrate *Google Trends* as an additional feature for the two best prediction models, XGBoost and Random Forest, and evaluate whether it improves predictive accuracy. In general, the R^2 measurements, presented in Table 5, display only a slight variation between the cases with and without Google Trends. However, Google Trends data tend to improve predictive accuracy for a longer time prediction horizon since it provides additional information about the future intention of a beach visit.

The *SHAP values* allow for a thorough analysis of feature impacts on predictions, addressing RQ3 by examining the best-performing models Random Forest

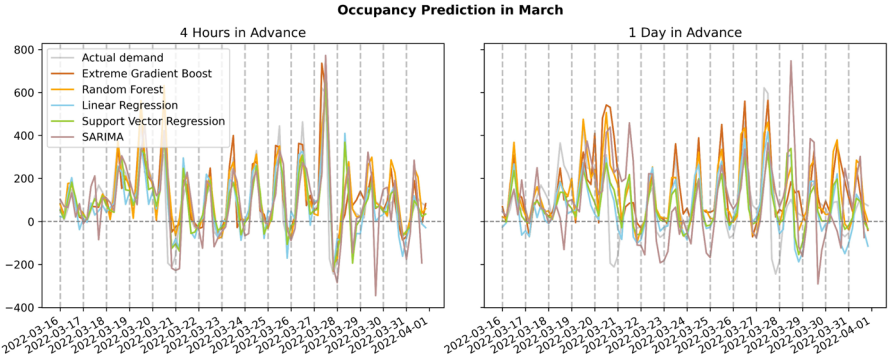

Fig. 6 Results of the occupancy prediction with a time granularity of 4 h for the beach in March for all models with prediction time horizons of 4 h and 1 day in advance. The figures display the predictions of each model in comparison to the actual demand over a period of 16 days. The x-axis labels indicate the start date of each day. The larger deviation from the actual demand in graph **b** indicates the worse prediction performance for a longer prediction time horizon

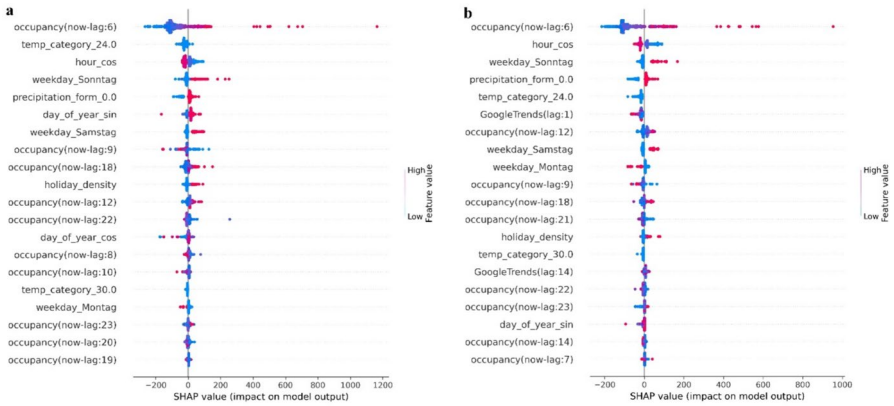

Fig. 7 SHAP values summary plots of the beach occupancy predicted one day ahead with the XGBoost model without Google Trends **a** and with Google Trends **b** respectively. The order of features indicates importance, and the colored dots display the magnitude and direction of influence on the resulting predicted value

and XGBoost. Exemplary Fig. 7 presents the SHAP value summary plot of the beach occupancy predicted one day ahead by the XGBoost model with and without Google Trends, respectively. The feature order denotes their respective levels of importance, with the most important feature at the top. The data points' distribution, frequency, and color offer additional information. The location of data points on the x-axis indicates each feature's effect on the predicted occupancy, with data points on the extremes representing strong influence. Color denotes the feature value, with red for high and blue for low values. Thus, a data point with blue color (i.e., low value) on the left side (i.e., negative impact) indicates a positive correlation between the feature and beach occupancy prediction. Regardless of the ML model or the integration of Google Trends data, the feature importance and impacts display similar results and trends, whereby, depending on the case-combination, slight variations occur. The most important feature is the last available lagged historical value explaining the reduced accuracy for longer prediction time horizon. Expectedly, a positive correlation exists and the wide distribution with several extreme positive values demonstrates that this feature allows the visitor movement peaks to be identified. Further, the time-related features hour sin and hour cos display a strong influence, whereby hour cos is more important due to its better interpretability. Low and negative values of hour cos indicate daytime between 6 and 18 o'clock, thus, having a positive impact on the visitor movement. Also important are the features that characterize the populations' free time, like the weekday reflecting the weekend or the holiday density, indicating a positive effect on occupancy. In contrast, weather categories demonstrate a mixed impact on prediction. While more straightforward categories, such as temperature or precipitation form, still tend to have a higher impact, more specific information, such as wind strength or precipitation amount, does not resolve any change in the prediction. Here, however, the limited test set with the months of February to April must be named, since only a limited amount

Table 5 R^2 measure of the best prediction models with a time granularity of 4 h

	Visitor Count					Occupancy				
	Beach Sections				Beach	Beach Sections				Beach
	13–20	14–20	21–24	25–29		13–20	14–20	21–24	25–29	
XGBoost										
1 lag	0.81	0.88	0.89	0.80	0.87	0.67	0.58	0.75	0.82	0.67
1 day	0.59	0.64	0.44	0.71	0.63	0.03	0.06	−0.2	−0.08	0.36
3 days	0.6	0.55	0.56	0.60	0.58	−0.3	−0.04	−1.26	−0.75	0.31
XGBoost, Google Trends										
1 lag	0.76	0.86	0.89	0.80	0.86	0.76	0.68	0.74	0.83	0.59
1 day	0.61	0.68	0.65	0.72	0.78	0.08	0.12	−0.17	0.42	0.34
3 days	0.61	0.63	0.59	0.67	0.70	−0.59	−0.03	−0.01	0.31	0.28
Random Forest										
1 lag	0.77	0.83	0.91	0.78	0.84	0.74	0.62	0.77	0.84	0.64
1 day	0.65	0.58	0.69	0.66	0.67	0.08	0.01	0.3	0.02	0.39
3 days	0.66	0.54	0.68	0.69	0.68	0.05	-0.12	-0.3	0.14	0.27
Random Forest, Google Trends										
1 lag	0.77	0.82	0.89	0.79	0.84	0.73	0.64	0.76	0.84	0.63
1 day	0.66	0.64	0.64	0.67	0.69	0.09	0.07	0.16	0.1	0.4
3 days	0.65	0.56	0.6	0.69	0.66	0.1	−0.05	−0.28	0.3	0.27

of the total weather categories occurs in these months. Interestingly, Google Trends features are also considered important, although they have only a marginal impact on R^2 measurement.

5 Discussion

Regarding RQ1, our analysis reveals that XGBoost and Random Forest stand out as the most suitable prediction models for visitor movement prediction to enable active visitor management. Despite the slightly weaker prediction accuracy of beach occupancy compared to visitor count, beach occupancy remains a vital and required prediction for facilitating active visitor management. While visitor count merely reflects the number of ingoing individuals, it lacks information about the duration of their stay—a critical factor influenced by external variables such as season or weather. Consequently, precise identification of crowding or overcrowding times based solely on visitor count is challenging for the DMO. For instance, a similar visitor count in summer may result in overcrowding due to prolonged stays, while in winter, people often take brief walks and cause no overcrowding. In contrast, beach occupancy encompasses both the duration of stay and the precise time of crowding and overcrowding, providing a more nuanced understanding and enabling the implementation of time-specific steering measures. This underscores the significance of occupancy prediction in active visitor management, streamlining the need for multiple threshold definitions to initiate appropriate measures. Further, the beach occupancy prediction model should be applied at a larger spatial granularity such as the entire beach. This approach ensures accuracy in occupancy calculations, even when individuals use different entrances for exiting compared to entering the beach. Additionally, our findings suggest a preference for a shorter prediction time horizon for beach occupancy, as it significantly enhances prediction quality.

Concerning the integration of Google Trends data to adress RQ2, our findings align with those of Önder et al. (2019). Due to the variability between use cases, an individual assessment of the improvement potential is required for each case. The marginal impact and modest improvement observed in the prediction models upon integrating Google Trends data can be attributed to the inherent characteristics of the data itself. Factors such as holidays or weather influence search query data, and thus, overlap with the features of our visitor movement prediction models. Consequently, the shared reliance on these influencing factors diminishes the potential for substantial improvement in the prediction models. However, despite the marginal improvement, the Google Trends features are still considered important in the SHAP value analysis, because they reflect a similar trend to the visitor movement. Hence, due to the overlaps and dependencies with the Google Trends data, they distort the importance of the other features. Therefore, the importance of the factors should be interpreted without Google Trends data.

To shed light on RQ3, we identify the most important features with the SHAP-value analysis: the lagged historical values, time-related, and holiday-related features. The mixed importance of weather categories implies that tourists primarily focus on simple weather forecasts including temperature or precipitation form, but

do not consider more detailed information. Our findings show that the inclusion of detailed weather features only impacts visitor movement predictions when tourists consider such information during their planning process. Here, it is essential to emphasize in our discussion on enhancing prediction models, that the primary objective is not the precise prediction of visitor movement but, rather, the accurate anticipation of peak periods. Already the correct identification of peaks during for example, holidays or breeding season allows to decide about the right steering measures in an active visitor management system. Beside the potential improvement of the prediction models, we should also discuss the transferability to other public, open-spaced POIs. Generally, the presented approach serves as a blueprint for measuring and predicting visitor count and occupancy of a public, open-spaced POI to identify potential demand peaks. To transfer the approach to a new POI the first step is to define the geographic boundaries of an overcrowded and to be monitored POI. Secondly, as with the beach entrances in Scharbeutz, sensors should be installed at natural bottlenecks, where there are optimally no or unattractive alternative routes for people to pass by. Sensor selection depends on factors including the lighting conditions, bottleneck width, data protection, nature conservation regulations, and the availability of power and internet. Common choices include light or infrared barriers, video systems, or pressure sensors. Thirdly, data collection begins after installation, and visitor capacity updates may be accessed through DMO portals. After accumulating sufficient data over at least one season, the ML models XGBoost and Random Forest can be trained, evaluated, and deployed for ongoing occupancy monitoring and prediction. Additional to the universal implementation of occupancy prediction systems for individual use cases, models trained in similar environments, such as beaches, may be reused for comparable settings. Distinct POIs, such as mountainous regions geared towards hiking, necessitate developing new models tailored to specific features. However, determining the extent to which model reuse is feasible and identifying relevant features per case category remain areas for further investigation. In general, the nature of the POI determines whether we can predict only visitor count or both visitor count and occupancy. For POIs resembling the beach, where visitors tend to stay and use the same exit as the entrance, predicting occupancy is feasible. However, at POIs like mountains with diverse hiking paths, visitors often choose different routes for the outward and return journey, resulting in higher error values when calculating occupancy. Despite this specific problem for occupancy prediction, visitor count still offers valuable insights with probably high prediction quality and active visitor management remains feasible in multiple POIs. Hence, the underlying methodology is generally transferable and can be regarded as a blueprint to enable the development of an active visitor management, whereas the reuse of trained models requires additional research.

Besides interpreting the results, we aim to discuss the broader implications and theoretical contribution concerning the active visitor management system and the resultant steering measures. Various prediction models are required to facilitate local, data-driven active visitor management. Thus, we differentiate between

mid- and short-term, as well as local and temporal steering measures. Mid-term active visitor management between different POIs can be enabled by models with a long prediction time horizon and aggregated spatial granularity, such as the entire beach. Specifically, steering measures can be triggered several days before the potentially crowded day by the DMO via the active visitor management system to proactively prevent overcrowding. This involves guiding potential tourists to an alternative, less crowded POI. Thus, the steering measure should start during the planning phase of a leisure trip allowing tourists to voluntarily choose an alternative POI without coercion. Short-term active visitor management within a POI or in its immediate vicinity requires more precise prediction models regarding spatial granularity. This precision is achieved through a shorter prediction time horizon. For example, if tourists have already chosen a potentially overcrowded POI, a uniform distribution of people on the beach may slightly relax the situation. For this purpose, more intrusive short-term steering measures may guide tourists to a free parking spot to avoid parking search traffic and ensure a uniform distribution across the beach entrances as people typically take the shortest path to the beach. This, even at high occupancy, reduces the number of visitors per square meter, minimizing trampling effects on flora and fauna (Schierding et al. 2011). In addition to local tourist shifts, the fine-grained temporal prediction enables temporal steering measures. The high importance of holiday-related features suggests the need for differentiated steering measures for day tourists and overnight guests. While overnight guests may shift their POI visit to an alternative day, a day tourist can only adjust visiting hours within one day. Furthermore, the strong influence of simple weather categories implies that steering measures in the form of recommendations should be triggered and presented depending on the environmental conditions. With our occupancy prediction models as data basis, we specifically enable short- to mid-term active visitor management to avoid overcrowding. However, a long-term change with the emergence of new hot spots should be regularly reviewed, and the local responsible entity should consider installing new sensor technology.

Regarding practical application and contribution, the successful implementation of the proposed multiple prediction models in a to-be-further investigated active visitor management system, can probably mitigate crowding and overcrowding and its associated environmental impacts. For activating various steering measures both in mid- and short-term as well as across location and times, a rule-based benchmark is necessary to automatically trigger actions (Neubig et al. 2024). To identify this benachmark Wall (2020) argues that "*[t]he key questions have always been 'How many is too many?' [or] 'How can this be determined?'. […] These questions will not disappear but will remain fundamental challenges, regardless of the trajectory that tourism takes.*". Due to the multidimensional nature of overtourism (Benner 2020) and its various effects (see 2.1), a universal answer does not do justice to the problem. Instead, a detailed, case-specific analysis is crucial, assessing all potential physical and perceived impacts of overcrowding and the current state of the POI along these dimensions. When evaluating a site's capacity to handle visitors, distinct characteristics such as infrastructure must be considered. For example, it is recognized that urban areas may be able to

cope with higher visitor numbers than sensitive natural areas such as a beach (Butler 2018b). The evaluation should in addition also account for seasonal variations in the natural environment and available services. In this paper we assume that the DMO is responsible for evaluating and operating the active visitor management system, thus, this is the recipient and user of the occupancy prediction system. However, regardless of the implementing organization it is essential to stress the importance of involving all relevant stakeholders and experts for a comprehensive assessment and to prevent the negative consequences of overly optimistic benchmarks. In an optimal setting, when all stakeholders are involved and a compromise is collectively reached, the ultimate responsibility of the organization is not decisive. However, due to the high complexity and lengthy decision-making processes, a practical approach is often required. Thus, the deciding neutral organization should consider all dimensions in their evaluation and consult multiple experts, yet come to a decision in an acceptable time and remain operational. Ultimately, setting and periodically revising the benchmark is an iterative process, influenced by infrastructural developments or the success of implemented measures. By continuously adjusting this benchmark, the number of visitors can be aligned to the existing infrastructure, thereby reducing illegal parking and minimizing search traffic. Furthermore, designated trails remain adequate, as visitors refrain from creating additional paths to avoid crowded areas. The implementation of such measures not only results in fewer people per square meter and reduced beach utilization but also mitigates trampling effects, leading to a decline in the deterioration of flora and fauna. Consequently, lower emissions contribute to the overall protection of the natural environment and climate. Additionally, the prevention of overcrowding plays a crucial role in wildlife preservation, as specific sections of the POI can be temporarily closed to respect breeding seasons. Beyond environmental protection, active visitor management offers notable social benefits. Restricting the maximum number of visitors can alleviate traffic congestion, crowded public transport, and long queues. This, in turn, enhances tourism acceptance among local residents, fostering increased friendliness and openness while preserving the integrity of the local culture. In addition, other applications beyond active visitor management are conceivable. Examples include better scheduling of required labor for services related to the POI, such as: lifeguards, beach basket suppliers, restaurants, or cleaning staff. In the long term, active visitor management sets the stage for sustainable tourism by safeguarding nature, preserving social structures, and supporting steady economic growth.

6 Conclusion

In this paper, we elaborated on how the visitor movement, including occupancy and visitor count, of a freely accessible and open-spaced POI can be predicted. Our approach aims to provide a data basis that enables the development of an active visitor management system for DMOs to avoid local overcrowding. Within a case study including real-world sensor data of the Scharbeutz beach, we evaluated the best prediction model to enable active visitor management with 1406 case combinations regarding time granularity, spatial granularity, prediction time horizon, model types, and the integration of Google Trends data. Additionally, we identified the most

important features and how they influence the prediction model with a SHAP-value analysis to identify appropriate steering measures.

We demonstrate that the prediction of a freely accessible and open-spaced POI is feasible with moderate to even strong R^2 relations, depending on the case-combination. Within the tourism demand prediction research area, we aim to close the gap for a POI-specific and short-term prediction. POI occupancy prediction enables the development of data-driven active visitor management to prevent overcrowding, ensure sustainable tourism, protect natural habitats, and avoid social conflicts. Furthermore, the utility of our approach extends to residents and organizations, providing them with crucial information about potentially high occupancy rates. This knowledge allows for proactive adaptation to the situation, ultimately contributing to the enhancement or preservation of tourism acceptance and the overall tourist experience.

However, this study is also subject to some limitations. A more extensive dataset spanning over several years could enhance our ability to analyze long-term trends, particularly considering the strong influence of the Covid-19 pandemic on the data under consideration, even if the tourism numbers nearly reflect pre-pandemic years. Moreover, integrating additional influencing factors may increase the prediction quality. Beside the limitations, further research should investigate the transferability of the prediction models' quality to other locations. While the presented approach could be considered a blueprint for POI-specific occupancy prediction, exploring alternative POIs may reveal use case categories with matching feature influences. For these categories, such as the beach in this case study, it may be possible to develop a universal model to reduce the deployment time of occupancy prediction systems at further POIs. Therefore, exploring other types of touristic POIs, such as lakes, hiking trails, or bicycle paths with slightly different environmental conditions, could provide valuable insights for enabling localized active visitor management. Additionally, the active visitor management system, as outlined and described based on occupancy prediction models, requires more in-depth investigation. This includes the development of a recommender system explicitly tailored to the goal of identifying and avoiding overcrowding. In summary, our approach for visitor movement prediction enables versatile applications in active visitor management, promising positive social and environmental impacts.

Appendix

See Tables 6, 7 and 8.

Table 6 4-h time slots

Part of the day	Lower Limit	Upper Limit
0: night	$\geq 12:00AM$	$< 4:00AM$
4: early morning	$\geq 4:00AM$	$< 8:00AM$
8: morning	$\geq 8:00AM$	$< 12:00PM$
12: midday	$\geq 12:00PM$	$< 16:00PM$
16: afternoon	$\geq 16:00PM$	$< 20:00PM$
20: evening	$\geq 20:00PM$	$< 12:00AM$

Table 7 W

Temp. category	Lower Limit [°C]	Upper Limit [°C]	**Rain category**	Lower Limit $[kg/m^2]$	Upper Limit $[kg/m^2]$	**Wind category**	Lower Limit $[m/s]$	Upper Limit $[m/s]$
-10		< -10	0.5		< 0.5	0		< 0.3
0	≥ -10	< 0	2.5	≥ 0.5	< 2.5	1	≥ 0.3	< 1.6
10	≥ 0	< 10	5.0	≥ 2.5	< 5.0	2	≥ 1.6	< 3.4
18	≥ 10	< 18	10.0	≥ 5.0	< 10.0	3	≥ 3.4	< 5.5
24	≥ 18	< 24	50.0	≥ 10.0	< 50.0	4	≥ 5.5	< 8.0
30	≥ 24	< 30	51.0	≥ 50.0		5	≥ 8.0	< 10.8
35	≥ 30	< 35				6	≥ 10.8	< 13.9
36	≥ 35					7	≥ 13.9	< 17.2
						8	≥ 17.2	< 20.8
						9	≥ 20.8	< 24.4
						10	≥ 24.4	< 28.4
						11	≥ 28.4	< 32.6
						12	≥ 32.6	

Table 8 Weather categories – feature engineering of the precipitation form

Precipitation form	Historical observation precipitation form	Mosmix forecast
0: no precipitation	NaN, 0, 4, 9	Snow water equivalent = 0 & Precipitation height < 0.5
1: rain	1, 6	Precipitation height > 0.5
2: snow	7, 8	Snow water equivalent > 0

Acknowledgements We gratefully acknowledge the financial support of Project 'AIR' (67KI21005G) by the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection of Germany (BMUV) and the financial support of Project 'FEB-NAFV' (19F2198A) by the Federal Ministry for Digital and Transport (BMDV) as part of the mFUND innovation initiative.

Author contributions J.B. conceptualized the research with valuable inputs from S.N. and R.K.; J.B. and R.K. developed the prediction models and investigated the results, whereby J.B. wrote the original draft and did part of the literature review; S.N. did part of the literature review and wrote the original draft of this part; A.H. supported by editing and reviewing the text; R.K. and H.K. supported the study by reviewing and supervision.

Funding Open Access funding enabled and organized by Projekt DEAL. The authors declare they have no financial interests.

Data availability The visitor movement data for the case study were collected as part of the 'AIR' project and are accessible in an aggregated form on the open-data platform 'Schleswig–Holstein' (https://opendata.schleswig-holstein.de/dataset?groups=tran). The authors declare they could freely dispose of the data and its processing. The weather data are available from the German Weather Service (https://opendata.dwd.de/). The search query data are available from Google Trends (https://trends.google.de/trends/).

Code Availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interests Stefan Neubig works for Outdooractive AG, a company that, among other things, offers visitor management services. The remaining authors declare no competing interests.

Consent for publication During the preparation of this work the authors used Grammarly and DeepL to improve readability and language. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Abu N, Syahidah WN, Afif MM, Nordin SZ (2021) SARIMA and exponential smoothing model for forecasting ecotourism demand: a case study in National Park Kuala Tahan. Pahang J Phys Conf Ser 1988:012118. https://doi.org/10.1088/1742-6596/1988/1/012118

Ali R (2016) Exploring the coming perils of overtourism [WWW Document]. Skift. URL https://skift.com/2016/08/23/exploring-the-coming-perils-of-overtourism/ (accessed 8.5.23)

Ali R (2018) The genesis of overtourism: why we came up with the term and what's happened since [WWW Document]. Skift. URL https://skift.com/2018/08/14/the-genesis-of-overtourism-why-we-came-up-with-the-term-and-whats-happened-since/ (accessed 8.5.23)

Álvarez-Díaz M, Rosselló-Nadal J (2010) Forecasting British tourist arrivals in the Balearic Islands using meteorological variables. Tour Econ 16:153–168. https://doi.org/10.5367/000000010790872079

Arbeitsgemeinschaft Verbrauchs- und Medienanalyse (2022) Verbrauchs- und Medienanalyse - VuMA 2022: Freizeit, Urlaub, Reisen.

Archer B (1987) Demand forecasting and estimation. Demand Forecast. Estim. 77–85

Attanasio A, Maravalle M, Muccini H, Rossi F, Scatena G, Tarquini F (2022) Visitors flow management at Uffizi gallery in florence. Italy Inf Technol Tour 24:409–434. https://doi.org/10.1007/s40558-022-00231-y

Balmford A, Beresford J, Green J, Naidoo R, Walpole M, Manica A (2009) A global perspective on trends in nature-based tourism. PLoS Biol 7:e1000144. https://doi.org/10.1371/journal.pbio.1000144

Becken S (2013) Measuring the effect of weather on tourism. J Travel Res 52:156–167. https://doi.org/10.1177/0047287512461569

Benner M (2020) Overcoming overtourism in Europe: towards an institutional-behavioral research agenda. Z Für Wirtsch 64:74–87. https://doi.org/10.1515/zfw-2019-0016

Bi J-W, Liu Y, Li H (2020) Daily tourism volume forecasting for tourist attractions. Ann Tour Res 83:102923. https://doi.org/10.1016/j.annals.2020.102923

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. Tour Econ 28:621–645. https://doi.org/10.1177/1354816620976954

Bi JW, Li C, Xu H, Li H (2021) Forecasting daily tourism demand for tourist attractions with big data: an ensemble deep learning method. J Travel Res. https://doi.org/10.1177/00472875211040569

Blicharska M, Smithers RJ, Mikusiński G, Rönnbäck P, Harrison PA, Nilsson M, Sutherland WJ (2019) Biodiversity's contributions to sustainable development. Nat Sustain 2:1083–1093. https://doi.org/10.1038/s41893-019-0417-9

Bollenbach J, Neubig S, Hein A, Keller R, Krcmar H (2022) Using machine learning to predict poi occupancy to reduce overcrowding presented at the Informatik, Gesellschaft für Informatik. Bonn. https://doi.org/10.18420/INF2022_34

Butler R (2018a) Challenges and opportunities. Worldw Hosp Tour Themes 10:635–641. https://doi.org/10.1108/WHATT-07-2018-0042

Butler R (2018b) Sustainable tourism in sensitive environments: a wolf in sheep's clothing? Sustainability 10:1789. https://doi.org/10.3390/su10061789

Butler RW, Dodds R (2022) Overcoming overtourism: a review of failure. Tour Rev 77:35–53. https://doi.org/10.1108/TR-04-2021-0215

Capocchi A, Vallone C, Amaduzzi A, Pierotti M (2020) Is 'overtourism' a new issue in tourism development or just a new term for an already known phenomenon? Curr Issues Tour 23:2235–2239. https://doi.org/10.1080/13683500.2019.1638353

Chakraborty D, Elzarka H (2019) Advanced machine learning techniques for building performance simulation: a comparative analysis. J Build Perform Simul 12:193–207. https://doi.org/10.1080/19401493.2018.1498538

Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R (2000) CRISP-DM 1.0 step-by-step data mining guide

Chawathe SS (2019) Using historical data to predict parking occupancy, in: 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE, New York City, NY, USA, pp. 0534–0540. https://doi.org/10.1109/UEMCON47517.2019.8993084

Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 7:623. https://doi.org/10.7717/peerj-cs.623

Choi H, Varian HAL (2012) Predicting the present with google trends. Econ Rec 88:2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

DESA UN (2016) Transforming our world: The 2030 agenda for sustainable development

Dinis G, Breda Z, Costa C, Pacheco O (2019) Google Trends in tourism and hospitality research: a systematic literature review. J Hosp Tour Technol 10:747763. https://doi.org/10.1108/JHTT-08-2018-0086

Dogru-Dastan H (2022) A chronological review on perceptions of crowding in tourism and recreation. Tour Recreat Res 47:190–210. https://doi.org/10.1080/02508281.2020.1841373

Ducrotoy JP, Elliott M (2008) The science and management of the North Sea and the Baltic Sea: natural history, present threats and future challenges. Mar. Pollut. Bull 57:8–21. https://doi.org/10.1016/j.marpolbul.2008.04.030

dwif e.V (2016) Wirtschaftsfaktor Tourismus für das Reisegebiet Ostsee (Schleswig-Holstein).

dwif e.V. (2019.)Wirtschaftsfaktor Tourismus für das Reisegebiet Ostsee 2019.

Eichenauer VZ, Indergand R, Martínez IZ, Sax C (2022) Obtaining consistent time series from Google Trends. Econ Inq 60:694–705. https://doi.org/10.1111/ecin.13049

German Weather Service, 2023a. Formulations of the Weather Elements [WWW Document]. URL https://www.dwd.de/DE/service/lexikon/begriffe/W/Wetterelementeformulierungen_pdf.pdf?__blob=publicationFile&v=3

German Weather Service 2023b. Weather and climate - Deutscher Wetterdienst - Glossary - N - Precipitation intensity [WWW Document]. URL https://www.dwd.de/DE/service/lexikon/Functions/glossar.html?lv2=101812&lv3=101906

German Weather Service, 2023c. Weather and climate - German Meteorological Service - Glossary - B - Beaufort scale [WWW Document]. URL https://www.dwd.de/DE/service/lexikon/Functions/glossar.html?lv2=100310&lv3=100390

German Weather Service, 2024. Climate Data Center (CDC) [WWW Document]. URL https://opendata.dwd.de/climate_environment/CDC/

Ghalehkhondabi I, Ardjmand E, Young WA, Weckman GR (2019) A review of demand forecasting models and methodological developments within tourism and passenger transportation industry. J Tour Futur 5:75–93. https://doi.org/10.1108/JTF-10-2018-0061

Gretzel U (2019) 5. The role of social media in creating and addressing overtourism, in: Overtourism. De Gruyter Oldenbourg, pp. 62–75. https://doi.org/10.1515/9783110607369-005

Google Trends (2023) Google Trends [WWW Document]. Google Trends. URL https://trends.google.de/trends/?geo=DE (accessed 1.10.23).

Hall CM, McArthur S (1996) Visitor management: Principles and practice. Heritage management in Australia and New Zealand. Oxford University Press, Melbourne, pp 37–54

Hu Y, Ritchie JRB (1993) Measuring destination attractiveness: a contextual approach. J Travel Res 32:25–34. https://doi.org/10.1177/004728759303200204

Jacobsen J, Kr S, Iversen NM, Hem LE (2019) Hotspot crowding and over-tourism: antecedents of destination attractiveness. Ann Tour Res 76:53–66. https://doi.org/10.1016/j.annals.2019.02.011

Jiao EX, Chen JL (2019) Tourism forecasting: a review of methodological developments over the last decade. Tour Econ 25:469492. https://doi.org/10.1177/1354816618812588

Kalisch D, Klaphake A (2007) Visitors' satisfaction and perception of crowding in a German National Park: a case study on the island of Hallig Hooge, in: Forest Snow and Landscape Research. pp. 109–122.

Kim DK, Shyn SK, Kim D, Jang S, Kim K (2021) A daily tourism demand prediction framework based on multi-head attention CNN: the case of the foreign entrant in South Korea, in: 2021 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, Orlando, FL, USA, pp. 01–10. https://doi.org/10.1109/SSCI50451.2021.9659950

Koens K, Postma A, Papp B (2018) Is overtourism overused? understanding the impact of tourism in a city context. Sustainability 10:4384. https://doi.org/10.3390/su10124384

Lazer D, Kennedy R, King G, Vespignani A (2014) The parable of google flu: traps in big data analysis. Science 343:1203–1205. https://doi.org/10.1126/science.1248506

Li X, Li H, Pan B, Law R (2021) Machine learning in internet search query selection for tourism forecasting. J Travel Res 60:1213–1231. https://doi.org/10.1177/0047287520934871

Liddle M (1997) Recreation ecology: the ecological impact of outdoor recreation and ecotourism. Chapman & Hall Ltd, UK
Liu H, Liu Y, Wang Y, Pan C (2019) Hot topics and emerging trends in tourism forecasting research: a scientometric review. Tour Econ 25:448–468. https://doi.org/10.1177/1354816618810564
Liu A, Lin VS, Li G, Song H (2022) Ex ante tourism forecasting assessment. J Travel Res 61:64–75. https://doi.org/10.1177/0047287520974456
Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions, in: advances in neural information processing systems. Curran Associates, Inc.
Lundberg SM (2018) SHAP documentation [WWW Document]. URL https://shap.readthedocs.io/en/latest/index.html (accessed 10.1.22)
Martinez-Plumed F, Contreras-Ochando L, Ferri C, Hernandez-Orallo J, Kull M, Lachiche N, Ramirez-Quintana MJ, Flach P (2019) CRISP-DM Twenty Years later: from data mining processes to data science trajectories. IEEE Trans Knowl Data Eng 33:3048–3061. https://doi.org/10.1109/TKDE.2019.2962680
Mason P (2005) Visitor management in protected areas: from 'hard' to 'soft' approaches? Curr Issues Tour 8:181–194. https://doi.org/10.1080/13683500508668213
McKinsey&Company (2017) Coping with success - managing overcrowding in tourism destinations.
Mihalic T (2020) Conceptualising overtourism: sustainability approach. Ann Tour Res 84:103025. https://doi.org/10.1016/j.annals.2020.103025
Monz CA, Pickering CM, Hadwen WL (2013) Recent advances in recreation ecology and the implications of different relationships between recreation use and ecological impacts. Front Ecol Environ 11:441–446. https://doi.org/10.1890/120358
Moore DS, Notz W, Fligner MA (2013) The basic practice of statistics, 6th edn. W.H. Freeman and Co., New York
Neubig S, Baecker J, Gehring N, Hein A, Weking J, Krcmar H (2022) Data-driven initiatives of destinations supporting sustainable tourism, in: Americas Conference on Information Systems (AMCIS) 2022 (Accepted)
Neubig S, Bečevová M, Brosta F, Loges R, Hein A, Keller R, Krcmar H (2024) Beyond sensors: a rule-based approach for cost-effective visitor guidance. Presented at the ENTER 24, Izmir, Turkey
Newsome D, Moore SA, Dowling RK (2012) Natural Area Tourism: Ecology. Multilingual Matters, Impacts and Management
Oklevik O, Gössling S, Hall CM, Steen Jacobsen JK, Grøtte IP, McCabe S (2019) Overtourism, optimisation, and destination performance indicators: a case study of activities in Fjord Norway. J Sustain Tour 27:1804–1824. https://doi.org/10.1080/09669582.2018.1533020
Önder I, Gunter U, Scharl A (2019) Forecasting tourist arrivals with the help of web sentiment: a mixed-frequency modeling approach for big data. Tour Anal 24:437452. https://doi.org/10.3727/108354219X15652651367442
Outdooractive, 2022. Outdooractive Data Base.
Paidi V, Håkansson J, Fleyeh H, Nyberg RG (2022) CO2 Emissions induced by vehicles cruising for empty parking spaces in an open parking lot. Sustainability 14:3742. https://doi.org/10.3390/su14073742
Phumchusri N, Suwatanapongched P (2021) Forecasting hotel daily room demand with transformed data using time series methods. J Revenue Pricing Manag. https://doi.org/10.1057/s41272-021-00363-6
Popp M (2012) Positive and negative urban tourist crowding: florence Italy. Tour Geogr 14:50–72. https://doi.org/10.1080/14616688.2011.597421
Rodrigues V, Eusébio C, Breda Z (2023) Enhancing sustainable development through tourism digitalisation: a systematic literature review. Inf Technol Tour 25:13–45. https://doi.org/10.1007/s40558-022-00241-w
Schierding M, Vahder S, Dau L, Irmler U (2011) Impacts on biodiversity at Baltic Sea beaches. Biodivers Conserv 20:1973–1985. https://doi.org/10.1007/s10531-011-0069-1
Schmücker D, Keller R, Reif J, Schubert J, Sommer G (2022) Digitales Besuchermanagement im Tourismus: Konzeptioneller Rahmen und Gestaltungsmöglichkeiten. In: Gardini MA, Sommer G (eds) Digital Leadership Im Tourismus: Digitalisierung Und Künstliche Intelligenz Als Wettbewerbsfaktoren Der Zukunft. Springer, Wiesbaden
Séraphin H, Zaman M, Olver S, Bourliataux-Lajoinie S, Dosquet F (2019) Destination branding and overtourism. J Hosp Tour Manag 38:1–4. https://doi.org/10.1016/j.jhtm.2018.11.003

Sheehan L, Ritchie JRB (2005) Towards a model of the roles and activities of destination management organizations. J. Hosp. Tour. Leis. Sci.
Škare M, Soriano DR, Porada-Rochoń M (2021) Impact of COVID-19 on the travel and tourism industry. Technol Forecast Soc Change 163:120469. https://doi.org/10.1016/j.techfore.2020.120469
Song H, Li G (2008) Tourism demand modelling and forecasting—A review of recent research. Tour Manag 29:203–220. https://doi.org/10.1016/j.tourman.2007.07.016
Song H, Qiu RTR, Park J (2019) A review of research on tourism demand forecasting: launching the annals of tourism research curated collection on tourism demand forecasting. Ann Tour Res 75:338–362. https://doi.org/10.1016/j.annals.2018.12.001
Spenceley A, Kohl J, McArthur S, Myles P, Notarianni M, Paleczny D, Pickering C, Worboys GL (2015) Visitor management. In: Worboys GL, Lockwood M, Kothari A, Feary S, Pulsford I (eds) Protected area governance and management. ANU Press, Canberra, pp 715–750
Statistisches Amt für Hamburg und Schleswig-Holstein (2023) Handel, Tourismus, Dienstleistungen - Statistikamt Nord [WWW Document]. URL https://www.statistik-nord.de/zahlen-fakten/handel-tourismus-dienstleistungen (accessed 1.4.24)
Studer S, Bui TB, Drescher C, Hanuschkin A, Winkler L, Peters S, Müller K-R (2021) Towards CRISP-ML(Q): a machine learning process model with quality assurance methodology. Mach Learn Knowl Extr 3:392–413. https://doi.org/10.3390/make3020020
Sun H, Yang Y, Chen Y, Liu X, Wang J (2023) Tourism demand forecasting of multi-attractions with spatiotemporal grid: a convolutional block attention module model. Inf Technol Tour 25:205–233. https://doi.org/10.1007/s40558-023-00247-y
Tauber V, Bausch T (2022) Will COVID-19 boost sustainable tourism: wishful thinking or reality? Sustainability 14:1686. https://doi.org/10.3390/su14031686
Tourismus-Agentur Lübecker Bucht (2022) Strandticker Lübecker Bucht [WWW Document]. URL https://www.luebecker-bucht-ostsee.de/strandticker
Toubes RD, Araújo-Vila N, Fraiz-Brea JA (2020) Influence of weather on the behaviour of tourists in a beach destination. Atmosphere 11:121. https://doi.org/10.3390/atmos11010121
Tsang WK, Benoit DF (2020) Gaussian processes for daily demand prediction in tourism planning. J Forecast 39:551–568. https://doi.org/10.1002/for.2644
UNWTO (1998) Guide for local authorities on developing sustainable tourism. A tourism and the environment publication, Egraf, Madrid
UNWTO (2022) Tourism Recovery Gains Momentum as Restrictions Ease and Confidence Returns [WWW Document]. URL https://www.unwto.org/news/tourism-recovery-gains-momentum-as-restrictions-ease-and-confidence-returns (accessed 1.4.24).
UNWTO (2023) Tourism on Track for Full Recovery as New Data Shows Strong Start to 2023 [WWW Document]. URL https://www.unwto.org/news/tourism-on-track-for-full-recovery-as-new-data-shows-strong-start-to-2023 (accessed 11.30.23)
Veiga C, Santos MC, Águas P, Santos JAC (2018) Sustainability as a key driver to address challenges. Worldw Hosp Tour Themes 10:662–673. https://doi.org/10.1108/WHATT-08-2018-0054
Volchek K, Liu A, Song H, Buhalis D (2019) Forecasting tourist arrivals at attractions: search engine empowered methodologies. Tour Econ 25:425–447. https://doi.org/10.1177/1354816618811558
Wall G (2020) From carrying capacity to overtourism: a perspective article. Tour Rev 75:212–215. https://doi.org/10.1108/TR-08-2019-0356
Wall G (2019) Perspectives on the environment and overtourism, in: overtourism - issues, realities and solutions. De Gruyter Berlin, pp. 27–43.
Webster J, Watson RT (2002) Analyzing the past to prepare for the future: writing a literature review. MIS Q 26, xiii–xxiii.
Wu DC, Song H, Shen S (2017) New developments in tourism and hotel demand modeling and forecasting. Int J Contemp Hosp Manag 29:507–529. https://doi.org/10.1108/IJCHM-05-2015-0249
Yin J, Cheng Y, Bi Y, Ni Y (2020) Tourists perceived crowding and destination attractiveness: The moderating effects of perceived risk and experience quality. J Destin Mark Manag 18:100489. https://doi.org/10.1016/j.jdmm.2020.100489
Zelenka J, Kacetl J (2013) Visitor management in protected areas. Czech J Tour 2:5–18. https://doi.org/10.2478/cjot-2013-0001

Authors and Affiliations

Jessica Bollenbach[1,2,3] · Stefan Neubig[4,5] · Andreas Hein[6] · Robert Keller[7] · Helmut Krcmar[4]

✉ Jessica Bollenbach
jessica.bollenbach@fit.fraunhofer.de

1 FIM Research Center for Information Management, Alter Postweg 101, 86159 Augsburg, Germany

2 University of Bayreuth, Wittelsbacherring 10, 95444 Bayreuth, Germany

3 Branch Business and Information Systems Engineering of the Fraunhofer FIT, Alter Postweg 101, 86159 Augsburg, Germany

4 Chair for Information Systems and Business Process Management (KrcmarLab), Technical University of Munich, Boltzmannstr. 3, 85478 Garching, Germany

5 Outdooractive AG, Missener Str. 18, 87509 Immenstadt, Germany

6 Institute of Information Systems and Digital Business, University of St. Gallen, Müller-Friedberg-Strasse 8, 9000 St. Gallen, Switzerland

7 University of Applied Sciences Kempten, INIT, Bahnhofstraße 61, 87435 Kempten (Allgäu), Germany