A machine learning-based approach for classifying tourists and locals using geotagged photos: the case of Tokyo

Ahmed Derdouri¹ · Toshihiro Osaragi¹

Received: 17 November 2020 / Revised: 16 June 2021 / Accepted: 27 August 2021 /

Published online: 18 September 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

In tourism-dependent cities, investigating the spatiotemporal distribution and dynamics of tourist flows is crucial for better urban planning in both steady and perturbed states. In recent years, researchers have started relying more on photo-based, geotagged social data, which offer insights about tourists, popular hotspots, and mobility patterns. However, distinguishing between tourists and locals from this data is problematic since residence information is often not provided. While previous studies rely on heuristic (e.g., period of stay) and probabilistic (Shannon entropy) approaches, this paper proposes a method for classifying tourists and residents based on machine learning (ML) algorithms and considering parameters that could explain the variability between the two (e.g., weather, mobility, and photo content). This approach was applied to Flickr users' geotagged photos taken in Tokyo's 23 special wards from July 2008 to December 2019. The results show that stacked ensemble (SE) models are superior to models based on five supervised-learning algorithms, including gradient boosting machine (GBM), generalized linear model (GLM), distributed random forest (DRF), deep learning (DL), and extremely randomized trees (XRT). Temporal entropy (TEN), mobility on workdays, and frequent visits to amusement venues and crowded places influenced how users were classified. While temporal distribution showed similar monthly/hourly patterns, spatial distribution varied. The proposed approach might pave the way for scholars to carry out future tourism research on different topics and subsequently support policymakers in the decision-making process, specifically in urban settings.

Keywords Classification · Tourists · Locals · Big data · Machine learning · Flickr · Geotagged photos · Tokyo

---

✉ Ahmed Derdouri
derdouri.a.aa@m.titech.ac.jp

¹ School of Environment and Society, Tokyo Institute of Technology, 2-12-1-M1-25 Ookayama, Meguro-ku, Tokyo 152-8550, Japan

1 Introduction

In tourism research, big data sources fall into three principal categories: (1) user-generated content (UGC), mainly consisting of online text- and photo-based social media records; (2) device data, including mobile phone and global positioning system (GPS) information; and (3) transaction data, for example, web searches and booking data (Li et al. 2018). With the rapid spread of social media tools and their popularity among travelers for documenting trips, recent literature reviews reveal that social media-based UGC is the most popular data source among researchers (Li et al. 2018; Li and Law 2020), who tend to choose social-media UGC over conventional sources of small data to analyze tourists and tourism. Major topics include examining mobility patterns by reconstructing trajectories (Paraskevopoulos and Palpanas 2018; Straumann et al. 2014; Yang et al. 2017; Yuan and Medel 2016; Zeng et al. 2012), identifying tourist landmarks and hotspots (Kim et al. 2017; Samany 2019), analyzing tourists' sentiments and behaviors (Jang and Moutinho 2019; Zhang et al. 2019, 2020), and recommending routes or planning trips (Kurashima et al. 2013; Lu et al. 2010).

The low cost of and easy access to UGC datasets are key factors behind their popularity among researchers. Unlike other data types, UGC is updated regularly and covers a long time span and large geospace, resulting in bigger datasets with rich metadata. Such data could generally be purchased from telecommunication companies at high prices, depending on the breadth of the target area and the span of the study period. An increasing number of researchers have begun conducting in-depth studies to analyze, or even predict, human mobility patterns based on low-cost, geotagged records collected from social networks (e.g., Twitter, Weibo). These data have proven useful, although records with spatial attributes only account for a small percentage of all social media data (3.33% for Twitter) (Chen et al. 2019a). In addition to text-based social networks, photo-sharing platforms (e.g., Flickr) serve as sources of geotagged data, and they are useful for analyzing tourism issues from different viewpoints. For instance, Xu et al. (2020) note that large-scale datasets with abundant metadata can be used in longitudinal studies concerning unsustainable tourism, principally resulting from the effects of long-term, accumulated behavior.

The advantages of UGC data over device and transaction data are multiple. For example, mobile phone data, a type of device data, are usually expensive to obtain, and this cost greatly depends on the spatial scale of the study area or the research period. Moreover, due to privacy concerns, mobile phone data are usually provided as macro-scale aggregated statistics instead of micro-scale samples, and such data do not include useful metadata. Thus, their application to tourism research is limited. Another form of device data is global positioning system (GPS) data. According to Li et al. (2018), two sources of GPS data are recognized in tourism studies: GPS loggers carried voluntarily by participants and GPS-enabled mobile applications owned by third parties. Despite the high accuracy and continuity of collection (Shoval et al. 2014), GPS data, when collected by volunteers, may suffer from biased results due to sample size and choice.

Distinguishing between tourists and locals is crucial because the groups are dissimilar in many ways, including size and mobility patterns (Hasnat and Hasan 2018) during steady conditions and, most importantly, during perturbed states, such as natural disasters or large events (e.g., the Olympic Games). Previous human geography-related studies have focused mainly on understanding and modeling locals' travel choices and behaviors, while ignoring those of tourists because they based their analysis on official survey data that includes only locals (e.g. Osaragi 2004; Osaragi and Hoshino 2012; Osaragi and Kudo 2019) or because they consider both groups as homogenous (e.g. Ma et al. 2020). Consequently, city planners and decision-makers know little about tourists' travel choices. However, ignoring this population group may lead to serious environmental, economic, and socioeconomic consequences, especially in cities largely reliant on tourism. Saenz-de-Miera and Rosselló (2014) simulate tourists' contribution to air pollution on the Spanish island of Mallorca, a top Mediterranean destination. They report that particulate matter (PM₁₀) concentrations could rise by up to 0.45% as a direct consequence of a 1% increase in the number of tourists. Within the same target area, Saenz-de-Miera and Rosselló (2012) also find that traffic congestion is highly associated with tourists.

Additionally, understanding mobility behaviors and travel choices during natural disasters has become more urgent in cities with extreme recurrent natural events. Kawasaki et al. (2018) report that foreigners living in Japan, who lack Japanese language proficiency, encountered challenges receiving information about recommended responses (i.e., evacuation or stay) after the Tohoku earthquake. It is reasonable to assume that tourists would face the same difficulties during a similar disaster while on a trip. Therefore, it would be beneficial, using UGC social media data, to distinguish between tourists and locals—obtaining insights into the behaviors of both groups, particularly during unstable conditions. Such information would help city planners design policies to facilitate the experiences of both tourists and locals.

Although social media-based UGC data offer excellent potential for tourism research compared to other data types, differentiating between tourists and residents remains a challenge, as social media users are not required to provide information about their places of residence. Many approaches have been proposed to distinguish tourists from locals, including heuristic methods (e.g., based on stay periods), supervised machine learning (ML) algorithms, and Shannon entropy (SHEN; Shannon 1948). SHEN is an information theory concept coined by mathematician Claude Shannon that calculates the dispersion of values of a given variable. The greater the dispersion, the greater the distinction, and the more knowledge that can be collected. Following this approach, two methods have been applied for classifying tourists and locals: (1) temporal entropy (TEN; Sun et al. 2015), based on the fact that tourists only linger for a brief time in the locations they visit and thus only take photographs for a limited time, typically not exceeding 2 months; and (2) spatial entropy (SEN; Zheng et al. 2012), based on the assumption that the complexity of tourists' and residents' travel trajectories differs, resulting in varying spatial distributions of images.

Three shortcomings have been found in these approaches. First, heuristic approaches cannot be evaluated, and their outcomes cannot be validated. Additionally, there is no consensus about an optimal way to determine the thresholds (e.g., minimum length of stay) and justify their values. Second, heuristic and

SHEN-based methods are not scalable; they can only differentiate between two groups. However, in some cases, it would be interesting to compare three or four groups (e.g., international and national tourists). Third, researchers employing ML-based methods neglect factors affecting the distinction between foreigners and locals, including weather, population density, and the content of users' posts (Sect. 3.2.2).

The present research seeks to fill these gaps by proposing an ML-based approach for distinguishing between tourists and locals based on digital traces extracted from their uploaded geotagged photos. Various explanatory variables account for the most relevant factors affecting the variations between the groups, including weather, mobility, entropy, photo content, and distances to strategic points of interest (POIs). We selected Flickr as the source of geotagged photos because of its free, maintained, and easy-to-use application programming interface (API) compared to other platforms. As a case study, we used the 23 special wards of the Tokyo Metropolitan Area (TMA) as the main target area. It should be noted that city users include four main categories of people: residents, commuters, overnight tourists, and same-day visitors or excursionists. For the sake of simplification, tourists in this study encompass nationals other than Japanese. The following questions guided the study:

- Is it possible to explain the variability between tourists and locals by considering factors such as weather conditions and mobility? For example, could variations in temperature or distances traveled help identify tourists and locals?
- If so, in addition to weather and mobility, what other factors could explain the variability between the two groups?
- Finally, what is the best ML algorithm for classifying the two groups based on these factors?

This paper is novel for three main reasons. First, the adopted method is based on ML algorithms with solid statistical grounds and the ability to evaluate outcomes. Second, a wide range of explanatory variables is considered, including those related to weather, human mobility, and population density. These variables were extracted from various accessible and low-cost sources. Other variables were extracted from the content of UGC photos themselves via advanced deep learning (DL) techniques. Third, the proposed approach is scalable, so it can be used to classify users into specific groups beyond tourists and locals (e.g., foreign tourists, local tourists, local residents). Finally, this is the first paper to apply an ML-based approach to distinguish between tourists and locals based on their geotagged photos.

The remainder of this article is organized as follows. Section 2 provides a comprehensive literature review of previously used methods. In Sect. 3, the study area is described. In addition, the data collection and preprocessing are detailed along with the study's methodological framework. Section 4 presents the results of the analysis, and Sect. 5 includes conclusions, the results, implications, limitations, and suggestions for improvements.

2 Related work

In the extant literature, three approaches have been used to distinguish between tourists and locals: (a) heuristic approaches, based on location information provided by users, the number of posts (i.e., tags, tweets, photos, etc.), or the length of stay periods (the number of days/weeks between a user's first and last published post in the target area); (b) SHEN filtering methods; and (c) ML models based on multiple algorithms. Table 1 presents a comprehensive overview of previous studies, grouped by how they separate tourists from locals based on digital traces from textual social networks (i.e., Twitter) or photo-sharing platforms (i.e., Flickr). It is worth noting that the listed studies were collected from Scopus and Google Scholar based on their popularity (e.g., citations). Moreover, we limited the search to studies that used Twitter and Flickr because they are the most common among researchers given their free and easy-to-use API.

Heuristic approaches for distinguishing tourists from residents are based on the idea that tourists generally stay in their travel destinations for short periods. For example, Abbasi et al. (2015) consider 4 weeks to be the maximum duration of tourist stays in Sydney, Australia. Accordingly, users are considered tourists if their stay periods are < 30 days; otherwise, they are considered locals. Girardin et al. (2007) also employ a 30-day maximum duration of stay to separate tourists from locals, while Manca et al. (2017) apply the same heuristic approach to categorize Twitter users with missing location information but set their maximum stay period to 20 days. De Choudhury et al. (2010) select 21 days, and Andrienko et al. (2013) apply interesting parameters, considering Twitter users to be Seattle locals if, over a period of 60 days, they stay more than 9 days in Seattle and less than 9 days elsewhere. However, these approaches are not statistically grounded. There is no consensus about the time interval that must be considered, and the outcomes cannot be assessed or validated.

Other studies use the SHEN filtering method. For instance, Zheng et al. (2012) use SEN to discriminate between tourists and non-tourists based on the mobility entropy of their movement trajectories (H_{mob}), setting a threshold of 0.2 to identify tourists ($H_{mob} \geq 0.2$) and non-tourists. Sun et al. (2015) use TEN, and Chen et al. (2019b) and Yang et al. 2017 apply the same approach, with threshold values set to two and three, respectively. This approach offers a better solution and higher accuracies than heuristic methods (Table 1). However, the method is dependent on the data and target area used, and the accuracy cannot be improved.

ML algorithms are also used to classify tourists and residents. To the best of our knowledge, only one study applied this approach to classify Twitter users in Florida Hasnat and Hasan (2018), employing three supervised ML algorithms: Decision Tree (Safavian and Landgrebe 1991), k-Nearest Neighbors (Manning et al. 2008), and Support Vector Machine (Cristianini and Shawe-Taylor 2000). The performances of these algorithms were evaluated and compared with the results of a heuristic approach that assumes Twitter users tweet from their homes late at night (i.e., 12:00 a.m.–6:00 a.m.). Ensemble classifiers combining the abovementioned single classifiers were also developed and evaluated. However,

Table 1 Descriptive list of past studies concerning distinguishing between tourists and locals, grouped by approach: (a) heuristic approaches, (b) Shannon entropy-based probabilistic approaches, and (c) ML-based approaches

Approach	Study	Target area	Source	Main objective	Highlights	Accuracy
(a)	Abbasi et al. (2015)	Sydney, Australia	Twitter	Analysis of travel behavior of tourists and residents	A maximum stay period of 4 weeks to be identified as a tourist	Not reported
	Manca et al. (2017)	Barcelona, Spain	Twitter	Extraction of human mobility patterns of tourists and locals	A maximum stay period of 20 days to be identified as a tourist	Not reported
	De Choudhury et al. (2010)	Barcelona, Spain; Paris, France; London, UK; San Francisco and New York City, US	Flickr	Automatic generation of travel itineraries of both groups	A time span of 21 days between the first and last photo taken and at least two POIs visited in the same city to be identified as a tourist	Not reported
	Girardin et al. (2007)	Province of Florence, Italy	Flickr	Spatiotemporal analysis of tourists' distribution and patterns	A maximum stay period of 30 days to be identified as a tourist	Not reported
	Önder (2017)	Multiple cities in Austria	Flickr	Identifying and classifying multi-destination trips in Austria		Not reported
	Andrienko et al. (2013)	Greater Seattle Area, USA	Twitter	Spatiotemporal analysis of thematic patterns of Seattle residents via visual analytics	Staying more than nine days in Seattle and less than nine days elsewhere over a period of 60 days to be identified as a local	Not reported

Table 1 (continued)

Approach	Study	Target area	Source	Main objective	Highlights	Accuracy
	Liu et al. (2018)	Greater Cleveland, USA	Twitter	Comparing mobility patterns between residents and visitors	Four groups considered based on a spatial scale: local, state, national, and international. User group defined based on the number of posts, with a cutoff of 40 posts	Not reported
	Straumann et al. (2014)	Zurich, Switzerland	Flickr	Reconstructing and visualization of trajectory networks of foreign and domestic tourists	“Domestic” and “foreign” groups identified based on the <i>user location</i> attribute provided by the user	Not reported
	Yan et al. (2017)	The Central Philippines Islands Region, Philippines	Flickr	Monitoring and assessing post-disaster tourism recovery	Tourists and locals identified based on user profile location information	Not reported
	Stepchenkova and Zhan (2013)	Peru	Flickr	Comparative content analysis of images gathered from destination marketing organizations (DMO) and social data	Photos with the tags “Peru” and “travel” considered indicators of images taken by travelers rather than residents	Not reported
	Hasnat and Hasan (2018)	Florida, USA	Twitter	Classification of tourists and locals and analysis of the spatial patterns of their destinations	A heuristic approach based on users’ location information, assuming that residents likely tweet mostly from home at night	79.0%

Table 1 (continued)

Approach	Study	Target area	Source	Main objective	Highlights	Accuracy
(b)	Zheng et al. (2012)	Paris, France; London, UK; San Francisco and New York City, USA	Flickr	Reconstructing tourists' travel trajectories in relation to regions of attractions	Tourists and non-tourists identified based on the SEN filtering method	Not reported
	Sun et al. (2015)	Munich, Germany	Flickr	Building a recommendation system of popular landmarks and the best routes connecting them	Tourists and non-tourists identified based on the TEN filtering method	89.0%
	Yang et al. (2017)	Manhattan Island of New York City, USA	Flickr	Extracting tourists' patterns using the concept of travel motifs		79.2%
	Chen et al. (2019b)	Beijing, China	Flickr	Presenting several methods for processing metadata of Geotagged photos taken by inbound tourists		Not reported
(c)	Hasnat and Hasan (2018)	Florida, USA	Twitter	Classification of tourists and locals and analysis of the spatial patterns of their destinations	Three approaches applied and compared: heuristic, supervised ML algorithms, and ensemble classifiers combining those algorithms. Ensemble classifier (AdaBoost) outperformed all others	82.4%

Fig. 1 **a** Study area (in white); **b** main areas with tourist attractions (marked by colored symbols in the map legend) located along the Yamanote Line (pink). (Images downloaded from pixabay.com in April 2020 and licensed under a simplified Pixabay license)

one problem with Hasnat and Hasan's (2018) work is the limited number and nature of the explanatory variables, which were constrained to five features, mostly related to mobility measurements.

3 Study area, data, and methods

3.1 Study area: geographical settings

The 23 special wards of the TMA, hereinafter referred to as Tokyo, were selected as the study area for this analysis (Fig. 1a) because Tokyo is a top destination in Japan for foreign and domestic tourists with numerous tourist attractions (Fig. 1b). Furthermore, since this study is primarily based on geotagged photos, a large volume of such photos taken in this area is available on Flickr, allowing for reliable analysis.

Before the COVID-19 crisis, and since the 2012 introduction of the economic program nicknamed *Abenomics*, Japan's tourism sector has experienced a considerable increase in the number of incoming international tourists, with an almost four-fold increase between 2012 and 2018 (Fig. 2). Tokyo, the capital of Japan, attracts approximately half of these tourists annually. In 2018, more than 30 million international visitors entered Japan, 15 million of whom visited the capital. Although the tourism sector was hit hard by COVID-19, resulting in a 99% drop in international visitors, recovery is expected in the coming years.

Fig. 2 Yearly trends of tourists visiting Tokyo and other prefectures (2011–2018) [Statistics from (Japan National Tourism Organization [JNTO] 2020a, b)]

Fig. 3 Methodological framework depicting the four key steps of the analysis: **a** geotagged photo collection and preprocessing, **b** explanatory variables preparation, **c** training samples preparations, and **d** classification and validation process

3.2 Methodology

The methodological framework for classifying Flickr users in this study is shown in Fig. 3. The key steps are data collection and processing (Fig. 3a), extraction of the explanatory variables based on various filters (Fig. 3b), preparation of the training samples (Fig. 3c), and classification and validation of the model and the subsequent prediction of users' classifications (Fig. 3d). A detailed explanation of each step is provided in the following subsections.

3.2.1 Geotagged photo collection and preprocessing

The first step in this analysis is collecting and preprocessing geotagged photos (Fig. 3a). In the extant literature, three photo-sharing platforms have been used as primary sources for geotagged photos: Flickr, Panoramio, and Instagram (Li et al. 2018). Panoramio was discontinued in 2014, while Instagram has ended free access to data via its API. Therefore, Flickr has been the major source for photos in multiple tourism-related studies due to its free streaming API, which permits access to and retrieval of a large dataset of photos associated with rich metadata shared by millions of users worldwide. Accordingly, Flickr has been selected as the main source of geotagged photos for this study. First, using its API, Flickr developed a Python script to collect photo records with spatial coordinates—taken from July 1, 2008, to December 31, 2019—within a geographical bounding box covering Tokyo. Although free, the Flickr API limits the number of results to 4000 records per query to prevent the database from crashing (Flickr n.d.). Consequently, collecting all photos taken throughout a decade within a target area as large as Tokyo is impossible. Hence, the study area was divided into small grids of 27.5 km², and the authors searched for photos during a short period of time, ranging from one hour to 1 month, depending on the volume of records found in each of these grids. Each retrieved record is associated with a set of informative attributes, called metadata, which can be divided into five categories: (1) unique attributes defining the photo (e.g., *photo id* and a set of *URLs* linking to different sizes of the image); (2) temporal attributes (e.g., *taken datetime* and *uploaded datetime*); (3) geographical attributes (i.e., *latitude* and *longitude*); (4) textual attributes (e.g., *title*, *description*, and *tags*); and (5) photo owner-related attributes (e.g., *user ID*, *country*, *city*, and *hometown*). The authors limited the search to records containing spatial dimensions (*latitude* and *longitude*). However, it must be noted that Flickr does not require users to provide textual attributes for the photo or the location of the photo's owner.

As with any social network, Flickr data contain erroneous records resulting from faulty hardware (Chen et al. 2019a) or issues related to GPS accuracy. To identify and filter out erroneous records, we applied the following rules:

- Users with at least two photos were chosen. Then, we checked to see whether at least two of their photos were taken for 1 day. If not, they were excluded from the analysis. This condition was added to permit the calculation of mobility indicators (see Sect. 3.2.2).
- For each user, a single photo record was kept when multiple photos were taken continuously at the same latitude and longitude.
- “Snapshot effect” was defined as the bias that might be introduced by “active users”—that is, uploaders who contribute large numbers of photographs (Hollenstein and Purves 2010). This bias might lead to a study area dominated by the behaviors of active users (Hu et al. 2015). To minimize this effect, only one record was retained if several images were taken within the same minute.
- If a user's account indicated that the user had taken photos at the same time (i.e., within the same minute) at different geographical locations separated by at least 13 m (Merry and Bettinger 2019), those records were removed.

Table 2 Summary of the collected Flickr record counts before and after preprocessing

Year	Before preprocessing		After preprocessing		Difference (%)	
	Users (unique)	Photos	Users (unique)	Photos	Users (unique)	Photos
2008	359	9042	162	1567	54.9	82.7
2009	972	16,314	454	3799	53.3	76.7
2010	1191	18,699	599	5058	49.7	73.0
2011	1337	24,404	691	6321	48.3	74.1
2012	1177	25,085	646	10,003	45.1	60.1
2013	1113	32,312	651	16,877	41.5	47.8
2014	1060	23,203	657	12,095	38.0	47.9
2015	891	26,653	495	17,455	44.4	34.5
2016	522	45,187	297	33,248	43.1	26.4
2017	882	63,071	495	34,096	43.9	45.9
2018	391	15,981	262	6947	33.0	56.5
2019	215	8795	153	3222	28.8	63.4
Total	10,110 (6,886)	308,746	5562 (3063)	150,688	45.0 (55.5)	51.2

Table 2 summarizes the cleaning process and depicts the total number of raw records and users compared to the number of filtered records and users. After cleaning, 150,688 photos taken by 3063 unique users remained.

3.2.2 Explanatory variables and associated data

In the second step of the analysis (Fig. 3b), we determined the factors that might explain the variability between tourists and residents based on previous studies and common-sense reasoning, considering the environmental and geographic settings of the target area. In total, ten main collections of explanatory variables were considered: weather conditions, distances to different POIs, population density, human mobility measures, TEN, SEN, period of stay, photo count, average probabilities of the most visited locations (grouped into eight categories), and the number of photos taken for each category. The following sections describe the considered variables, their data sources, and how the variable-related data were preprocessed. Additionally, since human mobility changes due to various temporal factors (e.g., the day/night cycle, seasons), which could help explain the variations between tourists and locals, these factors were considered during the calculations and are referred to as “filters” hereinafter. Table 3 lists the main collections encompassing the explanatory variables and filters considered in this study. A correlation analysis was conducted to avoid multicollinearity of the variables within the same collection.

3.2.2.1 Weather conditions Unlike locals, tourists usually stay at their travel destinations for a short time. Consequently, they are eager to explore the area by going outdoors, even in weather conditions locals would consider unacceptable. According to Müller (2019), tourists can adapt to undesired weather in the short term and to cli-

Table 3 Descriptive list of variables, filters, and the adopted notation used for the analysis

Variables collection		Filters	Prefix/abbreviation	Description
1	Weather conditions	–	<i>WTHR</i>	Weather conditions of the area where photos have been taken
2	Distances to POIs	–	<i>DIST</i>	Average distances to nearest POIs
3	Population density	–	<i>STAT_POP</i>	Population density of the area where a photo has been taken
4	Human mobility	Daytime/nighttime Seasons Workdays/holidays Calm/perturbed conditions	<i>DAYTM/NIGHT</i> <i>SPRING/SUMMER/WINTER/AUTUMN</i> <i>WRKDY/HOLDY</i> <i>STEADY/PERTBD</i>	Variables related to mobility, such as the displacements accumulated by an individual calculated for different seasons and times
5	TEN	–	<i>ENTROPY_TEMPORAL</i>	Temporal entropy
6	SEN	–	<i>ENTROPY_SPATIAL</i>	Spatial entropy
7	Stay period	–	<i>STAY_PERIOD</i>	The time interval between a user's first and last uploaded images in the target area
8	Photo count	–	<i>PHOTOSCOUNT</i>	The cumulative number of photos taken by a user
9	Average probabilities of the most visited locations grouped in eight categories	–	Starts with <i>MLGC</i> and ends with <i>PROBS</i>	Type of popular places frequented by a user: nature, amusement, etc
10	Photo count per category	–	Starts with <i>MLGC</i> and ends with 50	Number of photos taken in each of these popular places

mate changes in the long term. Gössling and Hall (2006) note that a stepwise, rather than a linear, relationship exists between tourists' behavioral adaptations and weather variables. Such behavioral adaptation takes three forms (Iso-Ahola 1986): (1) spatial substitution, by heading to a different destination, (2) temporal substitution, by waiting until the bad weather improves, and (3) activity substitution, by performing an activity other than the planned one. Tourists, being free during their journeys, have time to follow at least one of these adaptation strategies, which results in different behaviors (e.g., movement patterns) compared to locals, who are often busy on weekdays. The present study considered the following weather parameters: *temperature*, *pressure*, *humidity*, *visibility*, *cloud cover*, and *wind speed*.

Hourly weather data for Tokyo were extracted from World Weather Online.¹ Historical weather observations for Tokyo are available starting from July 1, 2008, so we chose that as the start date for the study period. *Precipitation* and *snow* data were excluded because most of the collected records did not match real observations. Other variables were also excluded, including *heat index*, *wind gust*, and *wind chill*, based on the correlation analysis between all extracted variables.

3.2.2.2 Distances to POIs According to previous research, visitors' behaviors are affected by the spatial distribution of POIs. For instance, Sugimoto et al. (2019) analyze visitors' mobility in the Ueno district, one of the most popular destinations in the Japanese capital, using data collected from GPS loggers and questionnaires. They find that physical factors, including transportation hubs, commercial areas, and POI distribution, influenced the tourists' behavior. Similarly, Hauthal and Burghardt (2016) compare the activities of locals and tourists in Dresden, Germany, using Flickr photos. They find that tourists usually take pictures in the vicinity of sightseeing hotspots, while photos taken by locals are distributed across the region. In this study, we downloaded POI data from OpenStreetMap (OSM), including the following POIs: *water*, *waterways*, *places of worship (PofW)*, *natural areas*, *health-related buildings*, *accommodations*, *leisure*, *money*, *tourism destinations*, *tourism info boxes*, *railway stations*, *bus stations*, and *shopping*. For each image, we calculated its distance to every nearest considered POI using ArcMap's function "Near." Then, for every user, we calculated the average distances of their images to all POIs.

We also calculated the average distances to the *centroid* of the study area, considering its quasi-hexagonal shape (Fig. 1a). As the Yamanote Line is a train loop around the study area's centroid, connecting the most popular tourist hotspots in Tokyo, tourists unfamiliar with the area would likely prefer using this line instead of the other transportation means used by locals.

3.2.2.3 Population density Previous studies have highlighted how overcrowded destination areas partially affect tourists' behaviors. Santana-Jiménez and Hernández (2011) explore the effect of population density on the choices of German and British tourists visiting the five main Canary Islands, finding that visitors' opinions

¹ <https://worldweatheronline.com>.

of overcrowding vary according to their sociological characteristics. While Britons seem less concerned about overcrowding, Germans prefer less crowded destinations. It should be noted that Santana-Jiménez and Hernández (2011) include not only residents' density (census data) but also an estimated density of all people present in the target area at the same time period, including overnight tourists, same-day visitors, and commuters. In this study, however, the authors considered only residents' density extracted from Japan's 2015 Population Census to reflect the degree of congestion in the target area. Further, since tourists are more likely to visit areas with similar cultural characteristics to their home cultures (Ng et al. 2007), the population density of foreigners living in Tokyo was added as a partial proxy for the cultural characteristics of foreign residents.

3.2.2.4 Human mobility Because locals, unlike visitors, are familiar with their home area, it is reasonable to assume that the two groups may follow different mobility patterns, specifically regarding the time spent at certain locations, the lengths of itineraries, and the mode of public transportation (PT). Kinsella and Caulfield (2011) explore the differences between resident and tourist users of PT in Dublin, Ireland. While Dubliners were more concerned about punctuality, frequency, and waiting times, tourists emphasized the availability of information and the reliability of the service. Additionally, in contrast to the mobility patterns of residents, tourists' patterns are affected by their information search behavior because they require more information when making their journeys (Thompson 2004).

Thus, two human mobility parameters were considered: the mean squared displacement (MSD) and the radius of gyration, r_g . MSD calculates the spatial extent of displacements accumulated by an individual within an area, while r_g measures how far a person goes beyond the center of mass for the visited locations. The MSD and r_g measures were calculated using the following formulas, respectively.

$$MSD(t) = \langle (r(t) - r_0)^2 \rangle = \frac{1}{n} \sum_{i=1}^n (r(t) - r_0)^2 \quad (1)$$

where r_0 is the location of the first photo taken at the beginning of the observations, and $r(t)$ is the location of a photo taken at time t .

$$r_g = \sqrt{\frac{1}{n} \sum_{i=1}^n (r_i - r_c)^2} \quad (2)$$

where n is the number of photos taken representing the visited locations; r_i is the location where photo i was taken; and r_c denotes the center of mass of the locations visited by a given person.

The distances between the two consecutive points were calculated based on their geographic coordinates using the Haversine formula (Robusto 1957):

$$d = 2r \times \arcsin \left(\sqrt{\sin^2 \left(\frac{\phi_2 - \phi_1}{2} \right) + \cos \phi_1 \cos \phi_2 \sin^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right)} \right) \quad (3)$$

where r is Earth's radius, approximately 6371 km; ϕ is the latitude; and φ is the longitude.

Because human mobility varies according to multiple factors, we calculated the mobility indicators based on the following filters: (i) daytime and nighttime, (ii) seasons, (iii) holidays and working days, and (iv) calm and perturbed weather conditions.

3.2.2.5 Temporal entropy TEN is a common metric for classifying locals and tourists (Sun et al. 2015) that has been applied successfully (Chen et al. 2019b; Yang et al. 2017). It is based on the fact that tourists stay temporarily at their destinations, and thus they take photos in a short time period, generally two consecutive months at most. Conversely, residents can take photos every month. From a probabilistic perspective, this results in higher TEN values for locals and lower ones for tourists. TEN is defined as follows:

$$TEN(k) = - \sum_i^{M(k)} P_i(k) \log P_i(k) \quad (4)$$

with $P_i(k)$ defined as:

$$P_i(k) = D_i(k) / \sum_i^{M(k)} D_i(k) \quad (5)$$

where $D_i(k)$ denotes the number of days on which person k took photos within month i in the study area, and $M(k)$ depicts the number of months in which person k took photos in the target area. Lower $TEN(k)$ values indicate that the user is more likely to be a tourist and vice versa.

3.2.2.6 Spatial entropy Similar to the probabilistic, approach-based TEN, SEN is based on the idea that the complexity of movement trajectories followed by visitors and residents varies, resulting in different spatial distributions of photos (Zheng et al. 2012). To calculate spatial entropy, the target area is first divided into $n \times m$ grids. Then, for a movement trajectory of p points, defined as $S = \langle (lat_0, lon_0, t_0), \dots, (lat_p, lon_p, t_p) \rangle$, photos taken by each user are assigned to each grid based on their coordinates. For each grid, the number of photos is counted. Lastly, for each user, SEN is calculated as follows:

$$SEN(S) = - \sum_i^n \sum_j^m P_{ij}(lat, lon) \log P_{ij}(lat, lon) \quad (6)$$

where $P_{ij}(x, y)$ denotes the spatial distribution of photos in grid (i, j) and is estimated by the number of photos in the grid (i, j) , considering the target area is divided into $n \times m$ grids. When the value of $SEN(S)$ is higher, there is a higher likelihood of the user being a tourist and vice versa.

Table 4 General categories and their descriptions for multi-labeling photos taken by Flickr users

	General category	Description
1	Nature	Greenspaces, waterbodies, and mountains
2	Culture	History-, religion-, and art-related objects and places
3	Business	Business districts
4	Amusement	Food, shopping, and entertainment
5	Infrastructure	Transportation, landmarks, and cityscape
6	Residence	Hotels and homes
7	Crowd	People
8	Other	Objects (books, dishes, etc.)

3.2.2.7 Period of stay The factor “period of stay” refers to the period between the first and last uploaded photos taken by a given user in the study area. This factor is widely used by researchers as a heuristic approach to differentiate between residents and locals. However, there is no consensus regarding an ideal threshold for period of stay in the reviewed literature, with values ranging between nine (Andrienko et al. 2013) and 30 days (Abbasi et al. 2015; Girardin et al. 2007).

3.2.2.8 Photo count The “photo count” factor denotes the total number of photos taken by a given user during the whole study period. This could be considered a simpler version of TEN that does not consider distribution over time. The logic behind selecting this factor is that locals reside in the target area and, logically, the counts of their accumulated photos, taken over a longer period of time, might be superior to those taken by tourists over shorter time periods.

3.2.2.9 Average probabilities of the most visited locations Apart from the associated textual metadata, the content of the photos themselves might contain interesting information (Li et al. 2018) that could reflect the preferences of the individual who took them. For instance, the nature of the most frequented places can be extracted by multi-labeling photos via DL techniques.

For this study, we developed a convolutional neural network (CNN)-based model to multi-label photos taken by Flickr users. The labels referred to eight general categories (Table 4). The model was based on the transfer-learning approach, using the MobileNetV2 architecture (Sandler et al. 2018). MobileNetV2 was selected for two main reasons. First, the prepared training dataset was relatively small (1416 photos), making it vulnerable to overfitting. This effect can be countered by using a smaller, yet expressive (i.e., ability to approximate functions) architecture, such as MobileNetV2. Second, the MobileNetV2 architecture optimizes memory consumption and execution time while minimizing prediction errors. Since the MobileNetV2 model had already been trained using the ImageNet dataset (Russakovsky et al. 2015), we left the trained parameters unchanged and only fine-tuned the whole network during training. The grid search technique was used for this purpose, computing the optimum values of the following hyperparameters: activation function, optimizer,

and the number of epochs. The model was trained and validated on threefold splits of a 70:30 ratio.² The splits were generated to obtain a balanced mix of samples across the eight categories. Therefore, the authors applied the stratification method proposed by Sechidis et al. (2011) and improved by Szymański and Kajdanowicz (2017).

3.2.2.10 Photo count by the nature of visited places This factor refers to the number of photos taken by a user per category (Table 4). For each user, the following steps were taken. Based on the results obtained via the multi-labeling model (see section above), predictions associated with each category for each photo were obtained. Then, the categories with predictions superior or equal to a probability threshold of 50% were replaced with one; otherwise, they were replaced with zero. Next, we counted the total number of ones referring to categories frequented by a user.

3.2.3 Preparation of training samples

The third step of the analysis involved preparing the training samples, consisting of Flickr users labeled as “tourist” or “local.” First, we selected only users with at least one specified location in their profile: country, city, or hometown. The rest were filtered out and labeled as “unknown”. Second, Algorithm 1 was applied. To determine the nationality of each user, the authors prioritized hometown (i.e., where they come from), city (i.e., city of residence), and then country, in that order. The first set of variant names of Japanese prefectures, cities, Tokyo, and its 23 wards were collected from a Massachusetts Institute of Technology (MIT)-licensed project in GitHub,³ based on official documents published by the Japanese Ministry of Internal Affairs and Communications. This set was enriched by a list of variant names for “Japan” in other languages. The second set containing variant names of foreign countries and cities was collected by removing the content of the first set containing Japanese location and region names from the locations of the users. We further checked each location among the uncategorized location names that had been misspelled by users and added them to either the first or second set.

Of the 3063 total Flickr users, 1517 (49.53%) were found to have at least one piece of location information available. After running Algorithm 1, 862 (28.14%) users were labeled as tourists and 655 (21.38%) as locals; 1546 (50.47%) were labeled as “unknown”.

² The results of this process yielded the best model, with a mean accuracy of 83.46% on the three splits (standard deviation = 0.00182), a micro-averaged area under the curve (AUC) of 0.87, and a micro-averaged area under the precision-recall curve (AUPRC) of 0.78.

³ <https://github.com/nobuf/list-of-cities-in-japan>.

Algorithm 1: Identification of tourists and locals from their Flickr location attributes

---

Input:

- Set of Flickr users with at least one known location (country, city, or hometown)
- Variant names of Japan, prefectures, cities, Tokyo, and its 23 special wards in English and Japanese
- Variant names of foreign countries and cities

Output:

- Set of users identified as tourists and residents
-

For any user in known-location users set:

If hometown is set:

If hometown is in variant names of Japan, Tokyo, and its 23 special wards: user is “local”
else: user is “tourist”

Else:

If city is set:

If city is in variant names of Japan, Tokyo, and its 23 special wards: user is “local”
Else: user is “tourist”

Else:

If country is set:

If country is in variant names of Japan, Tokyo, and its 23 special wards: user is “local”
Else: user is “tourist”

3.2.4 Classification and validation processes

The last step of the analysis was the classification process (Fig. 3c). The Python version of the H2O package (H2O.ai 2020b) (version 3.30.1.1) was employed to train models based on the prepared training sample labeled using Algorithm 1 detailed in Sect. 3.2.3. The H2O package offers an AutoML algorithm (H2O.ai 2020a), which automates building and comparing a set of candidate models based on selected ML algorithms. Five supervised ML algorithms were considered: (1) gradient boosting machines (GBM), (2) the generalized linear model (GLM), (3) distributed random forest (DRF), (4) a DL algorithm specifically for neural nets, and (5) extremely randomized trees (XRT). Additionally, two stacked ensemble (SE)-based models were generated, one based on the stacking of all developed models and one based solely on the best-performing model of each ML family.

To validate the models, tenfold cross-validation was used to avoid a unique validation split that might result in “lucky” or biased results. To assess and compare the performances of the trained models, we used the following statistical metrics:

1. Accuracy: the percentage of correct predictions, defined as follows:

$$accuracy = \frac{\text{number of correct predictions}}{\text{number of samples}} \quad (7)$$

2. Logarithmic loss (*logloss*): metric assessing how close the values predicted by a given model are to the true values. It measures how strongly a model distinguishes between different classes. Its value can be greater than or equal to zero. A value of zero denotes a model with the perfect capability to assign probabilities of 0% and 100%. The *logloss* for binary classification is defined as:

$$logloss = -\frac{1}{n} \sum_{i=1}^n \omega_i (y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)) \quad (8)$$

where n is the total number of samples, ω is the pre-defined weight of each sample ($\omega = 1$), and y and p refer to the predicted and target values, respectively.

3. F1 score: a measure of a given model's ability to classify positive cases (given a threshold value), defined as:

$$F1 = 2 \left(\frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \right) \quad (9)$$

where precision and recall are:

$$\text{precision} = \frac{TP}{TP + FP} \quad (10)$$

where TP and FP are the numbers of true and false positives predicted by the model, respectively. Recall was defined as:

Fig. 4 Performance boxplots for the best-performing models categorized by ML families

$$\text{recall} = \frac{TP}{TP + FN} \quad (11)$$

where FN refers to the number of false negative samples predicted by the model.

4. AUC (Area Under the ROC Curve): used to assess a given model's ability to distinguish between true positives (TP) and false positives (FP). Values range from zero to one. An AUC value of one indicates a perfect model, while values below or equal to 0.5 indicate a mediocre model.
5. AUPRC (Area Under the Precision-Recall Curve): evaluates a given model's ability to distinguish between precision–recall pairs.
6. Gini coefficient: an index for measuring the inequality across values of a frequency distribution based on the Lorentz curve.

4 Results

This section presents the evaluation results of all the classification models, identifying the best-performing model. Then, we explore the importance and dependence of the most important explanatory variables for this model. Next, the classification results based on this model are compared with those of the TEN method. Finally, a spatiotemporal analysis of tourists and locals classified using the best-performing model are presented and discussed.

Fig. 5 Performance of classification models plotted as dots based on the values of logloss (x-axis), AUC (y-axis), and training time (dot size)

4.1 Evaluation and comparison of all models

The prediction performance of the 200 best-performing models, based on ten-fold cross-validation data, was assessed in terms of six metrics: *accuracy*, *logloss*, *F1-score*, *AUC*, *AUPRC*, and *Gini score* (Fig. 4). It is visible that SE models outperformed other methods' models in terms of all metrics except AUPRC. Furthermore, it can be observed that DL values of all metrics tended to be less robust due to higher standard deviations compared to GBM models.

To visualize the differences between the best-performing models, the authors plotted the AUC scores as a function of the logloss scores (Fig. 5). The size of each dot indicates the required training time.

The main takeaways from evaluating and comparing the models' performance are summarized as follows:

- The two SE models scored higher for all considered metrics except AUPRC. The top-performing SE model (hereinafter named the best SE model) was built based on the predictions of all models, while the second was built based on the predictions of the best-performing model in each ML family.
- GBM models rated second in terms of performance metrics, with small variations in the standard deviation of the mean. They also consumed less time and fewer resources.
- DL models showed high score fluctuations, which might be attributed to the changes in the values of the hyperparameters used for fine-tuning. Moreover, these models require more training time, depending on the hyperparameter values.
- The unique models of XRT, GLM, and DRF scored lower for all metrics. Yet, their accuracies were still relatively high.

Table 5 Best model's performance metrics and confusion matrix reported by ten-fold cross-validation data

Logloss		0.524	MPCE		0.267
AUC		0.815	AUPRC		0.753
AIC		0.815	Gini		0.630
Confusion matrix					
	Tourists	Locals	Error	Error rate	
Tourists	395	213	0.350		213/608
Locals	73	391	0.157		73/464
Total	468	604	0.267		286/1072

Fig. 6 SHAP summary plots showing the global variable importance (left) and local explanation summary (right) for the top 12 variables impacting the predictions of tourists

4.2 An overview of the best SE model's performance

Next, we look more closely at the performance of the best SE model (i.e., confusion matrix; variables' importance and dependence). Additionally, for reference, the model's accuracy is compared against that of the TEN method. Finally, we present the results of the spatiotemporal analysis of locals and tourists classified using this model.

4.2.1 Confusion matrix

Table 5 lists the metrics and the resulting confusion matrix of the best SE model. The logloss (0.602) suggests that the model is decisive when assigning greater values, 0% or 100%, to the predicted value of each class. The AUC, indicating the model's ability to distinguish between TPs and FPs (1: perfect; 0.5: poor) was 0.815, indicating a good classifier. The mean per class error (the average error rates of the two classes) indicates that approximately 27% of the samples were misclassified. The confusion matrix gives a closer look at the error rates for each class. The model misclassified about 35% of tourist users as local users. This high percentage might be attributed to the fact that foreign tourists include foreign residents, whose behaviors might be similar to locals.

4.2.2 Variables' importance and dependence

To determine the variables that greatly impacted the predictions, we employed the Shapley additive explanations (SHAP) method (Lundberg and Lee 2017), which is a unified approach for explaining the output of any ML model. The variables' importance is measured by calculating SHAP values, which are an extended version of the game-theoretic Shapely values describing the influence of variables on the output of a given model.

This method was applied to a test dataset of 445 Flickr users to determine whether a user was a tourist or a local. Figure 6 shows the global feature importance (left) and the local explanation summary (right) of the top 12 variables impacting the predictions—ranked based on their average impact on the model

Fig. 7 Dependence plots of the variable “ENTROPY_TEMPORAL” influencing the model predictions of tourists versus its SHAP value

prediction. Each dot in the right plot corresponds to a person, and the dot’s color represents the impact of a given variable on the model’s output for that individual. The x-axis value of each dot refers to the impact on the prediction results. When a set of dots had the same impact, they were combined, forming a density.

For the best SE model, the following variables had the highest impact on the prediction results: “ENTROPY_TEMPORAL,” referring to TEN; “WKDY_MOB_RG_MEAN,” representing how far a user moves around the centroid of the visited locations during weekdays; and *MLGC_AMUSEMENT_PROBS* and *MLGC_AMUSEMENT_PROBS*, depicting whether photos were taken often within or around crowded places and amusement venues, respectively. TEN impacted the prediction negatively; as TEN increased, there was a better chance that a given user was a local. This supports the TEN-based approach proposed by Sun et al. (2015) and applied in other studies (Chen et al. 2019b; Yang et al. 2017), suggesting that TEN is an effective parameter for distinguishing between tourists and locals. The extent of the area covered by tourists and locals on workdays is another influential factor. This finding is consistent with common sense, as locals are committed to job obligations on weekdays, and consequently, they seldom travel long distances and take fewer photos (Hauthal and Burghardt 2016). Other variables with less influence on the model predictions included population density (*STAT_POP*), distance to the centroid of the study area (*DIST_CENTROID*), and distance to bus stations (*DIST_BUS_STATION*). While weather parameters did not greatly influence the predictions, the nature of the frequently visited places did. This information is extracted from the content of photos themselves, highlighting the advantage of using photo-based rather than text-based UGC data to explore tourists’ travel choices.

To further understand the model output, we drew SHAP dependence plots, which revealed the variability patterns of a given variable regarding its impact on the predictions. The x-axis of each plot represents the values of the variable in question, while the y-axis reflects the corresponding SHAP values for each user represented by a dot. The dots’ color corresponds to the degree of interaction with another variable, with red and blue representing high and low interactions, respectively.

Figure 7 illustrates the SHAP dependence plots of “ENTROPY_TEMPORAL” with the “DIST_CENTROID,” “WKDY_MOB_RG_MEAN,” and “SPRING_MOB_RG_MEAN.” Plot 1 (left) shows the dependence of TEN on the distance to the

Fig. 8 Classification accuracy of user types with different thresholds using the TEN-based approach

Fig. 9 Temporal distribution of tourists (green curves) and locals (red curves) per month (left) and per hour (right). Sunrise and sunset (right) are shown by green dots and arrows, pointing both right and left to reflect variations in these times. Bars represent the standard deviation of the mean

centroid of the study area. When this distance is above 7 km, the SHAP value for TEN diminishes, resulting in a lower probability that a person is a local. This suggests that locals visit a larger area than tourists. Plots 2 (middle) and 3 (right) show the dependence of TEN on how far a user moves around the centroid of the visited locations during weekdays and during the spring season, respectively. This suggests again that the average spatial extent of locals' movement (> 2 km) is wider, especially during working days and spring.

4.2.3 Comparative analysis with TEN-based results

For reference, the classification results of the commonly used TEN-based approach were evaluated. $TEN(k)$ was calculated for different values of $Mon(k)$, referring to the number of months using Eqs. (4) and (5) described in Sect. 3.2.2.5. A range of k

values was selected between 2 weeks ($k = 0.5$) and 5 months ($k = 5$). This approach was applied to the whole training sample set. The prediction accuracy results are illustrated in Fig. 8. The best accuracy (71.64%) was obtained for an optimal value of $k = 2.5$. For comparison, Sun et al. (2015) and Chen et al. (2019a) report accuracies of 89% ($k = 5$) and 79.2% ($k = 4$), respectively, while Yang et al. (2017) report an optimal threshold ε_t of 0.7 ($k = 1.5$).

Using the proposed approach in this study, the best SE model (Fig. 4) showed slightly better accuracies (ranging between 73 and 76%) than those achieved using the TEN-based method. The performance of the ML methods might be further enhanced by adding other explanatory variables. Moreover, the TEN approach cannot be extended when classifying users into more detailed types other than two groups (e.g., local tourists, foreign tourists, foreign locals). In contrast, the proposed model can be adjusted to consider multiple user types.

4.2.4 Spatiotemporal distribution of tourists and locals

Using the best SE model, the authors classified all Flickr users who took photos during the study period (2008–2019). To visualize the differences between the populations of the groups, their temporal and spatial distributions were analyzed.

According to the temporal distribution differences, Fig. 9 shows the statistics for the average, minimum, and maximum numbers of local and tourist users who visited the study area during every month (left) and hour (right) from 2008 to 2019. The monthly distribution (Fig. 9, left) shows that the number of locals was generally twice that of tourists. Both groups followed a similar distribution pattern. The spring months (i.e., March, April, and May) were the most popular among both groups, with the winter months (i.e., December, January, and February) coming in second, especially among locals, with January reaching a maximum number of 125 users, more than any other month. The least popular season was summer, especially the months of July and August. These fluctuations might be attributed to the study area's weather conditions, as Tokyo has a humid subtropical climate (Peel et al. 2007) with hot and humid summers and mild to cool winters with sporadic cold waves. Another reason for the drop in users beginning in July could be the typhoon season, which usually starts in July and ends in October, with the most powerful typhoons occurring in August and September. The hourly distribution of locals and tourists (Fig. 9, right) is similar in both groups, with afternoons (12:00 p.m.–6:00 p.m.) being the preferred period to take photos. The number of users began to decrease around 5:00 p.m.–7:00 p.m., corresponding to sunset.

Regarding spatial differences, Fig. 10 maps the kernel densities for the spatial distribution of photos taken by tourists (top) and locals (bottom) during the whole study period (left) and during the winters, springs, summers, and autumns (right). Photos taken by both groups were concentrated mainly in the vicinity of the Yamanote Line: Ueno, Akihabara, Tokyo station, Asakusa, Ikebukuro, Shibuya, and Shinjuku. Areas in the south, southeast, and southwest were not reached by Flickr users of either group, mainly because of the lack of tourist attractions.

Fig. 10 Maps of the seasonal spatial density of photos taken by tourists and locals from 2008 to 2019. Circles represent buffer zones of 2, 4, 6, and 8 km

Locals visited a wider area across Tokyo, with high and very high concentrations in a buffer zone of 8 km from the centroid of the study area. These highly dense clusters diminished beyond the 8 km buffer zone. These areas are on the north side of the study area, far from the Yamanote Line. Meanwhile, photos taken by tourists were mainly concentrated in an area 2–6 km from the centroid. These clusters were not evenly distributed across this zone and were mainly located near the Yamanote Line stations close to tourist attractions.

The seasonal spatial distributions of photo densities taken by tourists and locals show similar patterns for all seasons, except autumns. Although the number of users during summers was roughly equal to that recorded during autumns (Fig. 9, left), the

number of photos taken was different, as indicated by the lack of medium-, high-, and very high-density clusters. This can be explained by the perturbed weather conditions, given that August and September are the peak typhoon season in Japan, influencing the mobility of both tourists and locals.

In summary, the places visited by tourists in Tokyo, which is notorious for its maze-like train network, are limited and mainly concentrated around attractions in the vicinity of the Yamanote Line stations (e.g., Tokyo Station, Shibuya, and Akihabara) or areas close to it, such as Asakusa. This is most likely attributable to the simplicity of the Yamanote Line and the difficulties hampering tourists' use of other PT means, such as buses, particularly the language barrier. Moreover, the geographical settings of a destination influence tourists' spatial patterns (Hwang and Fesenmaier 2003). The choice of transportation mode depends on the transportation infrastructure (Thrasher et al. 2000) and, perhaps in Tokyo, how easy it is to use. Additionally, areas popular among tourists are known for amusement parks and crowded places, while locals, who are familiar with the area, tend to visit other locations—mainly those on the northern side of the area, known for temples, shrines, and natural landmarks.

5 Conclusions and discussion

This study proposed an ML-based approach for distinguishing between tourists and locals based on their digital traces extracted from geotagged photos they shared online. Numerous variables affecting this classification were considered, including those related to weather conditions, human mobility, entropy, and information extracted from photos. Mobility variables were recalculated, considering criteria including the time of the day and seasons, to reflect the fluctuations in human mobility during different environmental states. Photos taken by users were multi-labeled into eight general categories using a CNN-based model. The proposed approach was applied to Flickr users who took photos in the 23 special wards of Tokyo between July 2008 and December 2019.

Six supervised ML algorithms were considered: GBM, GLM, DRF, XRT, DL, and SE. The authors generated and compared 200 of the best-performing models across different ML families. The performance of each model was assessed based on tenfold cross-validation data using six metrics: *accuracy*, *logloss*, *F1-score*, *AUC*, *AUPRC*, and *Gini score*. The results showed that SE-based models were superior (accuracy of 75.5%). For reference, the authors also applied the widely used TEN-based approach for classifying tourists and locals, and the highest accuracy was 71.7%. Most of the built models of all families, except for the DL family, returned slightly better accuracies. Using the SHAP method, different variables influencing whether a user is a local or a tourist were evaluated. TEN was found to have a greater influence on this classification, proving the validity of the TEN-based approach. However, the findings also showed that other variables impact the classification process—mainly the mobility of individuals on weekdays and the nature of frequented places extracted from the photo content.

While the proposed approach scored better accuracy compared to the TEN-based method, its value remains modest because it depends on two factors: first, the level of accuracy in the formulation of some variables. For instance, the nature of the most frequented places is deemed to have an important influence on the distinction between locals and tourists (especially amusement-related venues and crowded places). However, we only considered eight categories for multi-labeling photos, including nature, culture, business, amusement, infrastructure, residence, crowd, and others (i.e., objects). Some “relatively broad” categories might have compromised the overall accuracy of the proposed approach, considering that “amusement”, for example, could be categorized into more specific aspects such as food (e.g., western-style restaurants, Japanese-style restaurants such as Izakayas), which could improve the results. Moreover, relying on residents’ density only to assess overcrowding might lead to compromised outcomes as well. Second, the lack of data associated with critical variables such as precipitation. The results suggest that cloud cover is a relatively important factor in the classification process. So, it is reasonable to assume that “precipitation measurements” would have similar or even higher importance considering that tourists do not have much time to explore the area in contrast to locals who might cancel or postpone going out when the weather is rainy.

Although, on average, there were twice as many locals as tourists, the temporal distribution of their average number followed a quasi-similar pattern throughout the year and throughout the day, with the most popular seasons being spring and winter and the preferable time of day being afternoons. However, the spatial distribution of photos showed major differences. Tourists tended to visit limited areas around stations connected with the easy-to-use Yamanote Line connecting Tokyo’s main attractions, while locals tended to visit a wider region, mainly places in the north of the study area.

These insights are helpful for scholars and city planners alike. For academic researchers, the proposed approach may be applied to other target areas as an alternative to currently used methods based on stay period or Shannon entropy. Furthermore, the method is scalable at three levels: First, instead of distinguishing between locals and tourists based on nationality alone, other criteria can be considered, such as residency status (i.e., foreign locals, foreign tourists, national tourists). Second, other ML algorithms could be employed (e.g., SVM). Third, depending on the area of study and data availability, other explanatory variables might be added to increase the accuracy of the outcome, which would allow scholars to ensure the credibility of their studies. Nevertheless, it should be mentioned that explanatory variables need to be selected while taking into consideration the target area’s characteristics. For example, in this study we used only residents’ density to assess overcrowding. However, this is not applicable in many European cities where historical centers have low resident densities compared to their outskirts. Likewise, using foreigners’ density to approximate international tourists’ movements is only valid in cities of North America or Asia where there are specific city districts such as Little Italy, Koreatown, and Chinatown. This is not necessarily valid in the case of European cities, for example.

City planners would also be interested in such reliable outcomes to understand how tourists and locals differ and behave, considering, for example, the length of

their displacements, the spatial extent of their activities, and the type of attractions each group is attracted to. Thus, planners, in collaboration with tourism administrators and authorities, could prepare guidelines, depending on the season or perturbation status, designed to help residents and tourists avoid congested areas (e.g., train stations and attractions during peak times).

This analysis has several limitations. First, we considered tourists as a homogeneous group, but this is not the case, as their characteristics vary depending on their (i) period of stay (i.e. overnight and same-day tourists) affecting the way they visit the city in terms of space and time and subsequently the way they take photos; and (ii) nationalities (e.g. domestic and international tourists) and/or cultural backgrounds. Koo et al. (2012), for instance, report that tourists' spatial dispersal depends significantly on their nationalities; Western (Germany, the United Kingdom, and the United States) and Chinese tourists visiting Australia were more dispersed than tourists from New Zealand, Korea, and Japan. Although they worked on a country-level scale, their results could also be true for a metropolitan area. Second, the training samples were based on the location information provided in the users' profiles. This information might not be correct or up to date. Furthermore, that process might incorrectly label foreign residents as tourists, which could influence the results, given that the mobility behavior of long-term foreigners could differ from that of foreigners visiting for a temporary period. Additionally, the hourly measurements of the weather variables in this study were considered to be equal across the study area, which was not the case. Finally, although deemed important for classifying locals and tourists, precipitation and snow data were not included because of data quality issues.

This research could be expanded in a variety of ways. Geotagged records from other social media, in addition to Flickr, might be considered, either to supplement the Flickr data or to compare and analyze biases in the analysis results based on various data sources. Moreover, during the filtering process, we disregarded the environments in which photos were taken (i.e., indoor or outdoor). Filtering out indoor photos or considering the environment where photos were taken as a new explanatory variable could help improve the results.

Acknowledgements This work is part of the research outcomes funded by JSPS Grant-in-Aid for Challenging Exploratory Research KAKENHI (Grant Number: 17H00843). The authors would like to thank the editor and the three anonymous reviewers for their valuable comments that greatly improved the final version of this article. We also thank Dr. Eng. Takuya Oki for his comments at the early stages of this research. Ahmed Derdouri (the corresponding author) would like to extend his sincere gratitude to the Yoshida Scholarship Foundation for the funding received to undertake his PhD.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Code availability The Python code used to produce the paper's results is available from the authors upon request.

References

- Abbasi A, Rashidi TH, Maghrebi M et al (2015) Utilising location based social media in travel survey methods: bringing Twitter data into the play. In: Proceedings of the 8th ACM SIGSPATIAL international workshop on location-based social networks, Bellevue, WA, USA, 3 November 2015, pp 1–9. LBSN'15. Association for Computing Machinery. <https://doi.org/10.1145/2830657.2830660>
- Andrienko G, Andrienko N, Bosch H et al (2013) Thematic patterns in georeferenced tweets through space-time visual analytics. *Comput Sci Eng* 15(3):72–82. <https://doi.org/10.1109/MCSE.2013.70>
- Chen M, Arribas-Bel D, Singleton A (2019a) Understanding the dynamics of urban areas of interest through volunteered geographic information. *J Geogr Syst* 21(1):89–109. <https://doi.org/10.1007/s10109-018-0284-3>
- Chen W, Xu Z, Zheng X et al (2019b) Geo-tagged photo metadata processing method for beijing inbound tourism flow. *ISPRS Int J Geo-Inf* 8(12):556. <https://doi.org/10.3390/ijgi8120556>
- Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511801389>
- De Choudhury M, Feldman M, Amer-Yahia S et al (2010) Automatic construction of travel itineraries using social breadcrumbs. In: Proceedings of the 21st ACM conference on Hypertext and hypermedia, New York, NY, USA, 13 June 2010, pp 35–44. HT '10. Association for Computing Machinery. <https://doi.org/10.1145/1810617.1810626>
- Flickr (n.d.) Flickr Services: Flickr API: flickr.photos.search. <https://www.flickr.com/services/api/flickr.photos.search.html> (Accessed 8 Sept 2020)
- Girardin F, Blat J, Ratti C (2007) Understanding of tourist dynamics from explicitly disclosed location information. In: International symposium on LBS & TeleCartography
- Gössling S, Hall CM (2006) Uncertainties in predicting tourist flows under scenarios of climate change. *Clim Change* 79(3):163–173. <https://doi.org/10.1007/s10584-006-9081-y>
- H2O.ai (2020a) H2O AutoML. <https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html>
- H2O.ai (2020b) Python Interface for H2O. <https://github.com/h2oai/h2o-3>
- Hasnat MM, Hasan S (2018) Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data. *Transp Res C Emerg Technol* 96:38–54. <https://doi.org/10.1016/j.trc.2018.09.006>
- Hauthal E, Burghardt D (2016) Using VGI for analyzing activities and emotions of locals and tourists. In: Proceedings of AGILE 2016, Helsinki
- Hollenstein L, Purves R (2010) Exploring place through user-generated content: using Flickr tags to describe city cores. *J Spat Inf Sci* 2010(1):21–48. <https://doi.org/10.5311/JOSIS.2010.1.13>
- Hu Y, Gao S, Janowicz K et al (2015) Extracting and understanding urban areas of interest using geotagged photos. *Comput Environ Urban Syst* 54:240–254. <https://doi.org/10.1016/j.compenvurb.2015.09.001>
- Hwang Y-H, Fesenmaier DR (2003) Multidestination pleasure travel patterns: empirical evidence from the American Travel Survey. *J Travel Res* 42(2):166–171. <https://doi.org/10.1177/0047287503253936>
- Iso-Ahola SE (1986) A theory of substitutability of leisure behavior. *Leisure Sci* 8(4):367–389. <https://doi.org/10.1080/01490408609513081>
- Jang S, Moutinho L (2019) Do price promotions drive consumer spending on luxury hotel services? The moderating roles of room price and user-generated content. *Int J Hosp Manag* 78:27–35. <https://doi.org/10.1016/j.ijhm.2018.11.010>
- Kawasaki A, Henry M, Meguro K (2018) Media preference, information needs, and the language proficiency of foreigners in Japan after the 2011 Great East Japan Earthquake. *Int J Disaster Risk Sci* 9(1):1–15. <https://doi.org/10.1007/s13753-018-0159-8>
- Kim J, Vasardani M, Winter S (2017) Landmark extraction from web-harvested place descriptions. *KI-Kunstliche Intelligenz* 31(2):151–159. <https://doi.org/10.1007/s13218-016-0467-3>
- Kinsella J, Caulfield B (2011) An examination of the quality and ease of use of public transport in Dublin from a newcomer's perspective. *J Public Transp*. <https://doi.org/10.5038/2375-0901.14.1.4>
- Koo TTR, Wu C-L, Dwyer L (2012) Dispersal of visitors within destinations: descriptive measures and underlying drivers. *Tour Manag* 33(5):1209–1219. <https://doi.org/10.1016/j.tourman.2011.11.010>

- Kurashima T, Iwata T, Irie G et al (2013) Travel route recommendation using geotagged photos. *Knowl Inf Syst* 37(1):37–60. <https://doi.org/10.1007/s10115-012-0580-z>
- Li X, Law R (2020) Network analysis of big data research in tourism. *Tour Manag Perspect* 33:100608. <https://doi.org/10.1016/j.tmp.2019.100608>
- Li J, Xu L, Tang L et al (2018) Big data in tourism research: a literature review. *Tour Manag* 68:301–323. <https://doi.org/10.1016/j.tourman.2018.03.009>
- Liu Q, Wang Z, Ye X (2018) Comparing mobility patterns between residents and visitors using geotagged social media data. *Trans GIS* 22(6):1372–1389. <https://doi.org/10.1111/tgis.12478>
- Lu X, Wang C, Yang J-M et al (2010) Photo2Trip: generating travel routes from geo-tagged photos for trip planning. In: MM'10—proceedings of the ACM multimedia 2010 international conference, pp 143–152. <https://doi.org/10.1145/1873951.1873972>
- Lundberg SM, Lee S-I (2017) A Unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S et al (eds) *Advances in Neural Information Processing Systems* 30. Curran Associates, Inc., pp 4765–4774. <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf> (Accessed 28 Sept 2020)
- Ma D, Osaragi T, Oki T et al (2020) Exploring the heterogeneity of human urban movements using geo-tagged tweets. *Int J Geogr Inf Sci*. <https://doi.org/10.1080/13658816.2020.1718153>
- Manca M, Boratto L, Morell Roman V et al (2017) Using social media to characterize urban mobility patterns: state-of-the-art survey and case-study. *Online Soc Netw Media* 1:56–69. <https://doi.org/10.1016/j.osnem.2017.04.002>
- Manning CD, Raghavan P, Schütze H (2008) *Introduction to information retrieval*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511809071>
- Merry K, Bettinger P (2019) Smartphone GPS accuracy study in an urban environment. *PLoS One* 14(7):e0219890. <https://doi.org/10.1371/journal.pone.0219890>
- Müller DK (2019) A research agenda for tourism geographies. Edward Elgar Publishing. <https://www.elgaronline.com/view/edcoll/9781786439307/9781786439307.xml> (Accessed 9 Sept 2020)
- Japan National Tourism Organization (2020a) Visitor Arrivals. Available at: <https://statistics.jnto.go.jp/en/graph/#graph--inbound--travelers--transition>. Accessed 5 Sept 2021
- Japan National Tourism Organization (2020b) Visits to Regions of Japan. Available at: <https://statistics.jnto.go.jp/en/graph/>. Accessed 5 Sept 2021
- Ng SI, Lee JA, Soutar GN (2007) Tourists' intention to visit a country: The impact of cultural distance. *Tour Manag* 28(6):1497–1506. <https://doi.org/10.1016/j.tourman.2006.11.005>
- Önder I (2017) Classifying multi-destination trips in Austria with big data. *Tour Manag Perspect* 21:54–58. <https://doi.org/10.1016/j.tmp.2016.11.002>
- Osaragi T (2004) Modeling of pedestrian behavior and its applications to spatial evaluation. In: *Proceedings of the third international joint conference on autonomous agents and multiagent systems, AAMAS 2004*, New York City, New York, USA, July 2004, pp 836–843. <https://doi.org/10.1109/AAMAS.2004.10119>
- Osaragi T, Hoshino T (2012) Predicting spatiotemporal distribution of transient occupants in urban areas. In: Gensel J, Josselin D, Vandenbroucke D (eds) *Bridging the geographic information sciences: international AGILE'2012 conference*, Avignon (France), April, 24–27, 2012. *Lecture Notes in Geoinformation and Cartography*. Springer, Berlin, pp 307–325. https://doi.org/10.1007/978-3-642-29063-3_17
- Osaragi T, Kudo R (2019) Enhancing the use of population statistics derived from mobile phone users by considering building-use dependent purpose of stay. In: Kyriakidis P, Hadjimitsis D, Skarlatos D et al (eds) *Geospatial technologies for local and regional development*. *Lecture Notes in Geoinformation and Cartography*. Springer International Publishing, Cham, pp 185–203. https://doi.org/10.1007/978-3-030-14745-7_11
- Paraskevopoulos P, Palpanas T (2018) What do geotagged tweets reveal about mobility behavior? *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10731:36–53. https://doi.org/10.1007/978-3-319-73521-4_3
- Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Köppen–Geiger climate classification. *Hydrol Earth Syst Sci* 11(5):1633–1644. <https://doi.org/10.5194/hess-11-1633-2007>
- Robusto CC (1957) The Cosine–Haversine formula. *Am Math Mon* 64(1):38–40. <https://doi.org/10.2307/2309088>
- Russakovsky O, Deng J, Su H et al (2015) ImageNet large scale visual recognition challenge. *Int J Comput vis* 115(3):211–252. <https://doi.org/10.1007/s11263-015-0816-y>

- Saenz-de-Miera O, Rosselló J (2012) The responsibility of tourism in traffic congestion and hyper-congestion: a case study from Mallorca, Spain. *Tour Manag* 33(2):466–479. <https://doi.org/10.1016/j.tourman.2011.06.015>
- Saenz-de-Miera O, Rosselló J (2014) Modeling tourism impacts on air pollution: the case study of PM10 in Mallorca. *Tour Manag* 40:273–281. <https://doi.org/10.1016/j.tourman.2013.06.012>
- Safavian SR, Landgrebe D (1991) A survey of decision tree classifier methodology. *IEEE Trans Syst Man Cybern* 21(3):660–674. <https://doi.org/10.1109/21.97458>
- Samany NN (2019) Automatic landmark extraction from geo-tagged social media photos using deep neural network. *Cities* 93:1–12. <https://doi.org/10.1016/j.cities.2019.04.012>
- Sandler M, Howard A, Zhu M et al (2018) MobileNetV2: inverted residuals and linear bottlenecks. In: 2018 IEEE/CVF conference on computer vision and pattern recognition, Salt Lake City, UT, June 2018. IEEE, pp 4510–4520. <https://doi.org/10.1109/CVPR.2018.00474>
- Santana-Jiménez Y, Hernández JM (2011) Estimating the effect of overcrowding on tourist attraction: the case of Canary Islands. *Tour Manag* 32(2):415–425. <https://doi.org/10.1016/j.tourman.2010.03.013>
- Sechidis K, Tsoumakas G, Vlahavas I (2011) On the stratification of multi-label data. In: Gunopulos D, Hofmann T, Malerba D et al (eds) *Machine learning and knowledge discovery in databases. Lecture Notes in Computer Science*. Springer, Berlin, pp. 145–158. https://doi.org/10.1007/978-3-642-23808-6_10
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Shoval N, Isaacson M, Chhetri P (2014) GPS, smartphones, and the future of tourism research. In: *The Wiley Blackwell companion to tourism*. Wiley, pp 251–261. <https://doi.org/10.1002/9781118474648.ch20>
- Stepchenkova S, Zhan F (2013) Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. *Tour Manag* 36:590–601. <https://doi.org/10.1016/j.tourman.2012.08.006>
- Straumann RK, Çöltekin A, Andrienko G (2014) Towards (re)constructing narratives from georeferenced photographs through visual analytics. *Cartogr J* 51(2):152–165. <https://doi.org/10.1179/1743277414Y.0000000079>
- Sugimoto K, Ota K, Suzuki S (2019) Visitor mobility and spatial structure in a local urban tourism destination: GPS tracking and network analysis. *Sustainability* 11(3):919. <https://doi.org/10.3390/su11030919>
- Sun Y, Fan H, Bakillah M et al (2015) Road-based travel recommendation using geo-tagged images. *Comput Environ Urban Syst* 53(Special Issue on Volunteered Geographic Information):110–122. <https://doi.org/10.1016/j.compenvurbysys.2013.07.006>
- Szymański P, Kajdanowicz T (2017) A network perspective on stratification of multi-label data. [arXiv:1704.08756](https://arxiv.org/abs/1704.08756) [cs, stat] (Accessed 15 June 2020)
- Thompson K (2004) Tourists' use of public transport information: what they need and what they get. In: *World transit research*, Strasbourg, France, 1 October 2004. <https://www.worldtransitresearch.info/research/3237>
- Thrasher SA, Hickey TR, Hudome RJ (2000) Enhancing transit circulation in resort areas: operational and design strategies. *Transp Res Rec* 1735(1):79–83. <https://doi.org/10.3141/1735-10>
- Xu F, Nash N, Whitmarsh L (2020) Big data or small data? A methodological review of sustainable tourism. *J Sustain Tour* 28(2):144–163. <https://doi.org/10.1080/09669582.2019.1631318>
- Yan Y, Eckle M, Kuo C-L et al (2017) Monitoring and assessing post-disaster tourism recovery using geotagged social media data. *ISPRS Int J Geo-Inf* 6(5):144. <https://doi.org/10.3390/ijgi6050144>
- Yang L, Wu L, Liu Y et al (2017) Quantifying tourist behavior patterns by travel motifs and geo-tagged photos from Flickr. *ISPRS Int J Geo Inf* 6(11):345. <https://doi.org/10.3390/ijgi6110345>
- Yuan Y, Medel M (2016) Characterizing international travel behavior from geotagged photos: a case study of Flickr. *PLoS One* 11(5):e0154885. <https://doi.org/10.1371/journal.pone.0154885>
- Zeng Z, Zhang R, Liu X et al (2012) Generating tourism path from trajectories and geo-photos. In: Wang XS, Cruz I, Delis A et al (eds) *Web information systems engineering—WISE 2012. Lecture Notes in Computer Science*. Springer, Berlin, pp 199–212. https://doi.org/10.1007/978-3-642-35063-4_15
- Zhang K, Chen Y, Li C (2019) Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: the case of Beijing. *Tour Manag* 75:595–608. <https://doi.org/10.1016/j.tourman.2019.07.002>

- Zhang K, Chen D, Li C (2020) How are tourists different? Reading geo-tagged photos through a deep learning model. *J Qual Assur Hosp Tour* 21(2):234–243. <https://doi.org/10.1080/1528008X.2019.1653243>
- Zheng Y-T, Zha Z-J, Chua T-S (2012) Mining travel patterns from geotagged photos. *ACM Trans Intell Syst Technol* 3(3):561–5618. <https://doi.org/10.1145/2168752.2168770>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.