TURISTICA
Italian Journal of Tourism

TIJT, Volume 33(1): 79–99
ISSN (print): 1974-2207
ISSN (online): 3034-9842
Received: 04.07.2024
Revised: 24.09.2024
Accepted: 02.10.2024
Published: 03.10.2024

Academic Research Paper

E-wom and territorial analyses: the use of opinion mining in tourism

Michelangelo Misuraca
Department of Business and Law (DiScAG), University of Calabria, Italy, michelangelo.misuraca@unical.it. ORCID: 0000-0002-8794-966X
Maria Spano
Department of Economics and Statistics (DiSES), University of Naples Federico II, Italy, maria.spano@unina.it. ORCID: 0000-0002-3103-2342
Luigi Celardo
Department of Social Sciences (DiSS), University of Naples Federico II, Italy, luigi.celardo@unina.it. ORCID: 0000-0001-5406-5457

Abstract: In the digital age, social media and electronic word of mouth (eWOM) have fundamentally transformed several industries, including tourism. This paper explores how opinion mining (also known as sentiment analysis) can be a strategic tool for tourism stakeholders to extract and analyse data from social media reviews and posts. The aim is to provide actionable insights to enhance customer satisfaction and destination management by assessing the sentiment expressed in online content. From a methodological viewpoint, the implemented approach integrates qualitative and quantitative data with textual data to delve into opinions embedded in eWOM. The analyses help identify patterns and trends in tourists' feedback, offering a valuable framework for understanding their behaviour and leveraging this information for the stakeholders' specific purposes. A case study on Naples (Italy) highlights the potential of sentiment analysis in tourism, suggesting future research directions. Practical applications range from improving customer satisfaction and destination reputation to guiding marketing strategies and supporting sustainable tourism development.

Keywords: *Social media, Sentiment analysis, Polarity scoring, Geo-data.*

JEL Codes: C80; C38; Z33; Z38

1. Introduction

The advent of the Internet and the proliferation of social media platforms have fundamentally transformed how people share information and experiences about everyday life aspects as well as the consumption of products and services. One significant change is the spread of information and news among the users. Platforms like Twitter and Facebook have become primary sources for many people, influencing public opinion and shaping discussions around current events (Neubaum & Krämer, 2016), demonstrating how messages on social media could influence political self-expression and information-seeking in everyday life. Furthermore, social media has transformed social interaction. While it facilitates connections, it also raises concerns regarding social comparison, self-esteem, and the fear of missing out. The constant exposure to curated online lives can impact self-perception and behaviour (Fischer & Reuber, 2011). Furthermore, social media heavily influences consumer behaviour in their purchasing decisions. Exposure to product recommendations, reviews, and influencer marketing impacts consumer choices on which products or services are consumed (Bronner & de Hoog, 2014).

This digital (r)evolution has hence given rise to the phenomenon of *electronic word of mouth* (eWOM), which refers to any statement made by potential, actual, or former customers about a product/service or a company, which is made available to a multitude of people and actors (Hennig-Thurau et al., 2004). The eWOM encompasses a variety of formats, including social media platforms, review websites, and forums, and has become a critical factor in shaping consumer perceptions and behaviours in many sectors, including tourism (Litvin et al., 2008). The impact of eWOM is particularly pronounced in the tourism sector due to the intangible and experiential nature of travel services. Potential tourists often seek online reviews and opinions to reduce uncertainty and perceived risk associated with travel-related decisions concerning destinations, accommodations, and activities (Gretzel & Yoo, 2008; Pabel & Prideaux, 2016). Research has shown that eWOM significantly affects tourists' destination choices, trust in tourism brands, and overall satisfaction with travel experiences (Filieri & McLeay, 2014).

In this framework, opinion mining has emerged as a powerful tool for extracting insights from eWOM in recent years. This approach involves the study of opinions and emotions expressed in textual data through statistical techniques, enabling researchers and practitioners to gauge the overall feeling towards a specific topic or entity (Pang & Lee, 2008). The sentiment expressed in eWOM plays a crucial role in tourism, as it can significantly affect traveller attitudes and behaviours. Positive feelings may enhance the attractiveness of a destination, while negative sentiments can discourage potential visitors (Ye et al., 2009). For instance, Liu and Park (2015) found that the sentiment expressed in online hotel reviews significantly influences booking intentions. Similarly, Wang and Fesenmaier (2004) demonstrated that the emotional content of eWOM messages sways travel planning and destination choice. By applying opinion mining to eWOM in tourism, stakeholders can better understand tourists' perceptions, identify trends and patterns in consumer feedback, and make data-driven decisions to enhance service quality and customer satisfaction.

Compared to traditional media, social media offers a cost-effective platform indeed, allowing for targeted marketing efforts and ensuring that messages reach the most relevant and engaged audiences (Effing & Spil, 2016). Moreover, this real-time data is also invaluable for understanding communities' needs, preferences, and concerns. Social media platforms have also emerged as part of a more powerful informative system for territorial management and marketing due to their ability to influence people's perceptions, gather feedback, and drive engagement (Migliaccio et al., 2023). For policymakers, the strategic use of social media can significantly enhance the promotion, development, and sustainable management of tourist destinations. Policymakers can quickly identify emerging issues by monitoring social media discussions, tracking policies' effectiveness, and adjusting strategies accordingly (Kaplan & Haenlein, 2010; Guillamón et al., 2016).

Although there is considerable literature on the use of social media in tourism analyses and their crucial role in providing valuable insight to all stakeholders, one issue that needs to be better investigated is the importance of territorial aspects. Territoriality in tourism studies refers to the spatial dimensions and geographic identity associated with tourism destinations, highlighting how places are defined, represented, and experienced by tourists and local communities. It hence plays a fundamental role in understanding the interactions between tourists and the destinations they visit, influencing at the same time destination management, marketing strategies, and sustainability practices. The perceptions of tourists and their sentiment may vary across the different part of a given place, thus it is important to analyse the opinions but also to visually map the polarity of their opinions from a spatial viewpoint. For this reason, the joint use of sentiment analysis techniques and territorial analysis may offer a powerful tool. In this sense, the present study explores the interplay between eWOM and sentiment in the tourism industry, highlighting the significance of opinion mining in deciphering traveller feedback. Additionally, here is introduced a novel approach to geo-referencing and visualise the sentiment concerning the different part of a place with tourist interest. Through a comprehensive review of the literature and analysis of empirical data, this work seeks to provide valuable insights into how sentiment derived from eWOM influences tourists' decision-making processes and the overall reputation of tourism destinations and services. The findings are expected to contribute to the growing body of knowledge on eWOM and sentiment analysis in tourism, offering a valuable tool to tourism managers and marketers to enhance their engagement strategies and to policymakers for assessing better territorial policies.

The work is structured as follows. In Section 2, the methodological framework is briefly introduced, and the reference literature in the tourism domain reviewed. In Section 3, an application of the analytical strategy focused on the city of Naples is presented to the readers, to show the effectiveness of the approach. Finally, in Section 4, the theoretical and practical implications of the approach are discussed, offering in conclusion some possible future directions of the research.

2. Methodological framework and related work

Text analytics are based on the statistical processing of textual bodies written in natural language after acquiring and pre-treating their content. Because of the text's unstructured nature, it is necessary to represent its content algebraically. The primary model used to structure the content is the so-called *vector space model* (Salton et al., 1975), in which the text chunks belonging to the analysed textual body are converted into p-dimensional vectors holding the "importance" of each single word (where p is the overall number of distinct words encompassed in all the chunks, i.e. the textual body's *vocabulary*). Formally, a collection of text chunks and each single chunk can be represented, respectively, as:

$$\mathbf{D} = (T_1, T_2, T_3, ..., T_j, ..., T_p) \tag{1}$$

$$\mathbf{d_i} = (t_{i1}, t_{i2}, t_{i3}, ..., t_{ij}, ..., t_{ip}) \tag{2}$$

where $\mathbf{D}$ is the vectorial representation of a collection of n text chunks, viewed as the distribution of its p

words (the vocabulary), with T_j total number of occurrences of the *j*-th word, and $\mathbf{d}_i$ is the vectorial representation of the *i*-th text chunk belonging to the collection, with t_{ij} total number of occurrences of the *j*-th word in the *i*-th text chunk. The encoding scheme beneath the model – known as *bag of words* (BoW) – considers the text as a multi-set of words, disregarding both grammatical and syntactic roles. The BoW scheme simplifies the computational treatment of broad textual bodies, limiting the possibility of reflecting words' contextual use. Using words' co-occurrences (i.e. the occurrence of pairs of words appearing in a chunk) partially avoids this drawback since the context may be represented by the relational structure of words appearing in the text chunks, with a granularity that can go from the entire chunk to its single clauses (Misuraca & Spano, 2020). An alternative to the vector space model is the so-called *word embedding* (Jiao & Zhang, 2021), whose primary goal is to map words as high-dimensional vectors of real numbers. Unlike BoW, word embedding models can capture words' semantic and syntactic meanings. Embedding models are typically based on neural networks (Bengio et al., 2003), with only one or two hidden layers for *prediction-based* models, following a shallow learning logic (e.g. Mikolov et al., 2013), and several hidden layers for *transformer-based* models, following a deep learning logic (e.g. Devlin et al., 2019) The latter are often fine-tuned to satisfy specific tasks like opinion mining analyses.

Depending on the chosen algebraic text representation, several strategies can be implemented to score and label text chunks by their semantic polarity. Additionally, the diverse approaches can be unsupervised or supervised, not considering prior sentiment information in the text categorisation step or leveraging the knowledge base obtained by training a proper classifier on a set of labelled examples. Unsupervised approaches based on the vector space model representation are typically lexicon-based, i.e. they use inventories of polarised words to detect the presence of negative and positive expressions. For each polarised word, it is possible to consider a binary scoring -1/+1 (e.g. Bing et al., 2005) or a more granular system based on discrete (e.g. Nielsen, 2011) or continuous values (e.g. Baccianella et al., 2010), and use a different synthesising function (Bravo-Marquez et al., 2014). The overall polarity of the text may also be computed by considering in the calculation some *valence shifters* altering the values of pivotal terms, for example, by amplifying or de-amplifying the corresponding negativity/positivity (Balbi et al., 2018). Differently, sentiment scoring and labelling can be obtained using pre-trained large language models fine-tuned for sentiment classification[1]. Following an embedding standpoint, the contextual use of words and their nuances are considered in the latter, leading, in many cases, to more accurate results outperforming lexicon-based methods (Catelli et al., 2022). Nevertheless, in both unsupervised and supervised approaches, a critical drawback concerns the linguistic resources used for the analysis, only sometimes available for specific domains such as tourism and even more only for some languages (English language resources are contemplated for all the cases above reported).

The application of opinion mining in tourism began to gain traction in the early 2000s, coinciding with the increasing use of social media and online review platforms. Initial studies focused on extracting sentiment from TripAdvisor, Booking.com and Yelp reviews. Sentiment data has been extensively used to assess the public perception of tourist destinations and brands. By analysing online reviews, social media posts, and other user-generated content, researchers and practitioners measured the overall sentiment towards a destination, identified positive and negative aspects, and understood how tourists perceive a brand or location. Hays et al. (2013) examined how destination brands are discussed on social media, using sentiment analysis to assess branding strategies' effectiveness and identify areas for improvement. Marine-Roig and Clavé (2015) explored the image of European destinations based on sentiment analysis by

[1] See, for example, https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment

analysing travel reviews, providing insights into how tourists perceive different destinations. Sentiment analysis is also a powerful tool for measuring customer satisfaction levels and identifying key factors influencing tourist experiences, helping tourist operators better understand their customers' needs and preferences, leading to improved service quality and customer satisfaction. For example, Siering et al. (2018) used sentiment analysis on online reviews to predict hotel guest satisfaction, providing actionable insights for hotel management to improve their services. Similarly, Chang et al. (2023) used sentiment analysis on luxury hotel reviews posted on Booking.com to analyse and explore marketing insights from tourists' attitudes and emotions. Sentiment data offers a valuable overview of the tourism phenomenon for market research, helping tourism actors (in the private and public sectors) understand trends, customer preferences, and competitor strengths and weaknesses. This information can inform strategic decisions, marketing campaigns, and product development. Xiang et al. (2015) demonstrated how sentiment analysis of online reviews can inform strategic decisions in hospitality management, offering insights into guest experiences and preferences. Li et al. (2023) forecasted visitor arrivals at two tourist attractions in China jointly using review data and official tourism statistics. Furthermore, sentiment analysis enables tourism actors to create personalised marketing campaigns and engage with customers more effectively, tailoring marketing messages and offers to meet specific individual needs and interests (Alaei et al., 2019). By analysing sentiment data from various sources, it is possible to identify areas where visitor experiences can be improved and enhance the quality of their services, leading to higher customer satisfaction, repeated visits, and positive word-of-mouth. In this sense, Kim et al. (2016) studied the impact of customer-generated reviews on hotel performance, showing how sentiment analysis can help hotels improve their services based on customer feedback. In a more recent work, Arici et al. (2023) examined customers' green reviews on TripAdvisor to identify environmentally friendly issues and provide critical advice to tourist operators to understand what practices customers notice and appreciate. Finally, sentiment analysis can help tourist operators and policymakers manage crises and protect the reputations of businesses and locations by monitoring and analysing public sentiment during and after a crisis, like the recent COVID-19 pandemic (Godovykh et al., 2021). The global pandemic greatly intensified interest in this area, as tourism sectors, especially hospitality, faced considerable challenges in understanding shifting customer sentiments. A study by Flores-Ruiz et al. (2021) examined the utility of opinion mining during the health crisis in Andalusia, demonstrating how sentiment analysis could provide real-time insights into tourist preferences. Similarly, Carvache-Franco et al. (2023) showed that sentiment data, combined with topic modelling, could effectively reflect tourists' preferences and behaviours during crises, helping companies and organisations' communication strategies in the tourism sector.

To the best of our knowledge, in all the inspected studies concerning the use of sentiment analysis and social media in tourism, there is a lack of attention to an essential trait of the phenomenon related to its territoriality. Tourists often seek authentic experiences deeply rooted in the territory's identity, making it an integral part of tourism research and practice. Destinations are usually marketed considering their unique territorial characteristics – such as landscapes, heritage, culture, and local traditions – which attract visitors. Territoriality also influences tourists' experiences, as the sense of place, local customs, and geographical features contribute to their perceptions and satisfaction. Nevertheless, the different parts of a place may induce different perceptions in tourists, e.g. due to the varying quality of services. An in-depth analysis of sentiment that considers the spatial dimension may offer a new and powerful tool to stakeholders, offering additional information to tourists and tourist operators. Hereafter, an example of how to integrate sentiment analysis with territorial analysis is provided, showing the significance of such a strategy in tourism studies.

3. A case study: Analysis of tourist attractions in the city of Naples

To show the capability of sentiment analysis in tourism research and its usage as a key tool both for industry and policymakers, we present the results of an analysis concerning the city of Naples' tourist attractions. Naples presents a fascinating case study, offering a unique blend of historical richness, cultural dynamism, and vibrant local life (Buonincontri et al., 2017). It boasts a 2,500-year legacy, reflected in its archaeological treasures, baroque architecture, and renowned museums. Furthermore, Naples offers an incomparable culinary experience thanks to its food tradition. While captivating for its rich history and vibrant energy, it also poses challenges for visitors seeking an experience beyond the stereotypical Italian vacation because of a perceived lack of security and some worries about public transport quality and other tourist services (De Falco & Corbino, 2023). Starting from these considerations, we decided to analyse the reviews posted by Italian speakers' users on Tripadvisor to assess the overall sentiment about the main attractions of Naples and map it as georeferenced data to directly visualise the different grades of tourist appreciation concerning the city's territory.

Table 1. Spatial distribution of per municipality attractions.

Municipality	Quarters	Attractions
01	Chiaia, Posillipo, S. Ferdinando	98
02	Avvocata, Mercato, Montecalvario, Pendino, Porto, S. Giuseppe	109
03	S. Carlo all'Arena, Stella	30
04	Poggioreale, S. Lorenzo, Vicaria, Zona Industriale	82
05	Arenella, Vomero	54
06	Barra, Ponticelli, S. Giovanni a Teduccio	6
07	Miano, S. Pietro a Patierno, Secondigliano	4
08	Chiaiano, Piscinola, Scampia	2
09	Pianura, Soccavo	1
10	Bagnoli, Fuorigrotta	21

Source*: authors elaboration on TripAdvisor data*

The word `Naples` was used to query Tripadvisor, selecting every touristic point of interest different from accommodations and restaurants. Noteworthy, TripAdvisor cover a heterogeneous number of points of interest, ranging from classic cultural sites to sports complexes, bars, shopping centres, and even eno-gastronomic tours. A total of 407 "attractions" were retrieved from the website at the end of the scraping process through a Python script (Table 1). All the available metadata (e.g. address, geo-localisation, attraction type, average visit duration, number of reviews, rating) were extracted and stored. The latitude and longitude of the attractions' position were further validated with Google Maps IDs (Figure 1). We employed a hexagonal grid developed by Uber[2] on the city map to provide significant details of the reference territory. The H3 system divides the map into hexagons (see Sahr, 2011) with a given resolution (in our case, approximately 0.74 km^2).

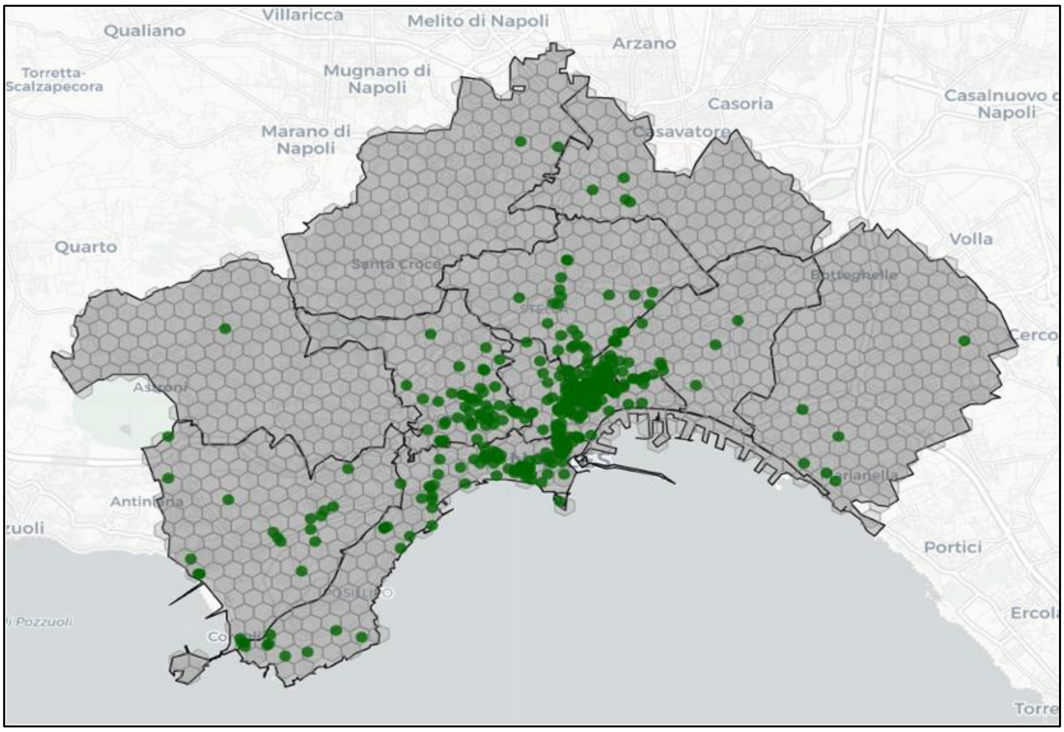
In recent years, there has been a substantial increase in tourist flows, accompanied by an equally

[2] See https://h3geo.org

significant rise in short-term rentals and a transformation of the commercial landscape. This has produced a pervasive and massive tourism that is often uncontrollable (Iovino, 2021; Esposito, 2023). As with all fast-occurring and spontaneous events without careful planning, tourism development in Naples has some drawbacks. These concerns raise significant questions about the ongoing processes' economic, social, and environmental sustainability. The sudden transformation of the landscape underscores these limitations, as it increasingly diverges from its authentic character, which currently attracts tourists from around the globe. Most peripheral neighbourhoods revealed a scarcity of POIs covered by Tripadvisor, as reflected by previous studies (e.g. La Rocca, 2021), mainly for poorer accessibility in terms of public transport, degradation of urban spaces and shortage of infrastructure and services

In addition, there need to be more policies to protect and enhance some areas of potential interest, as in the case of the municipalities of Barra, Ponticelli and S. Giovanni a Teduccio (on the right side of the map). On the contrary, in the historic centre of Naples, there is a greater concentration of tourist attractions, in correspondence with the ancient Greek and Roman plans of what is now called the *Decumani* area, producing frequent cases of over-tourism.

Figure 1. Spatial distribution of attractions in the city of Naples.



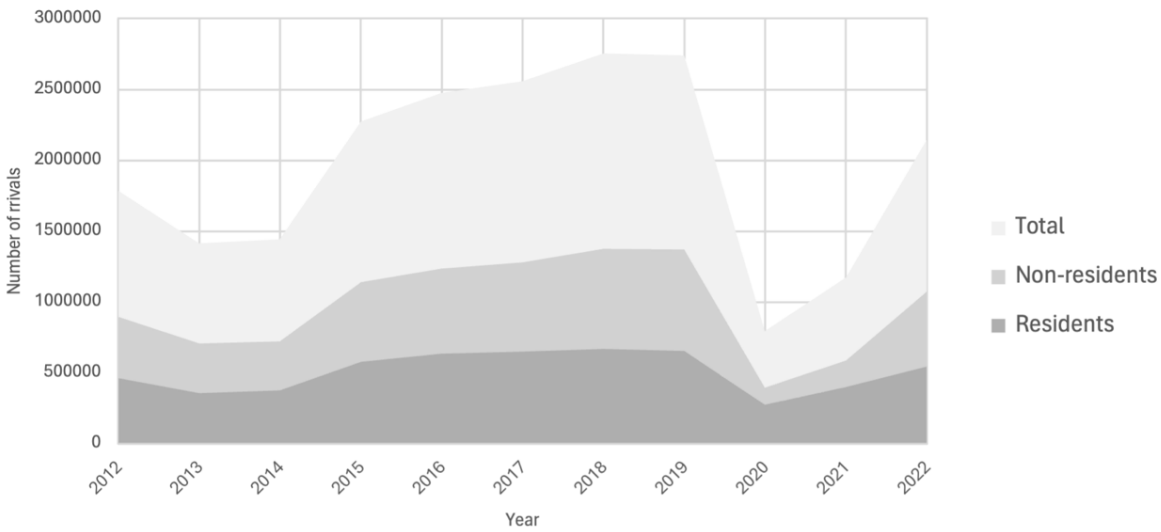
***Source**: authors elaboration on Google Maps data*

Concerning the type of attraction, we observed that 17.94% were architectural buildings, 15.24% were historical and religious sites, 8.85% were art, cultural, and natural sites, 47.17% were entertainment sites (including clubs, sports complexes, and others), and 10.81% were business activities (including stores).

3.1. Some descriptive statistics

Text analysis aside, social media has revolutionised data collection and analysis across various fields, including tourism (Li et al., 2021). The Web is, in fact, an incredible source of quantitative and qualitative data, together with relational data, textual data, geo-data, video and image data. Official statistics, which traditionally rely on surveys and administrative records, can complement existing methods by leveraging social media data (Daas et al., 2015). Figure 2 depicts the distribution of tourist arrivals in Naples from 2012 to 2022, collected by Istat in the "Occupancy of tourist accommodation establishments" survey, distinguishing residents and non-resident tourists (i.e. living in Italy or a foreign country). The metadata scraped from Tripadvisor may be used to estimate the volume of visitors and investigate some of their behaviours, such as period and type of visit.

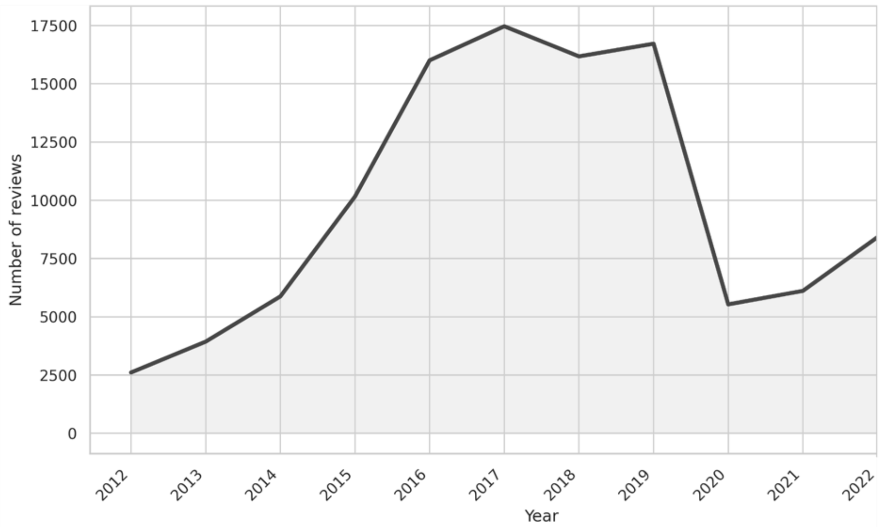
Figure 2. Yearly distribution of tourist arrivals in Naples (2012-2022).



***Source**: authors elaboration on Istat data*

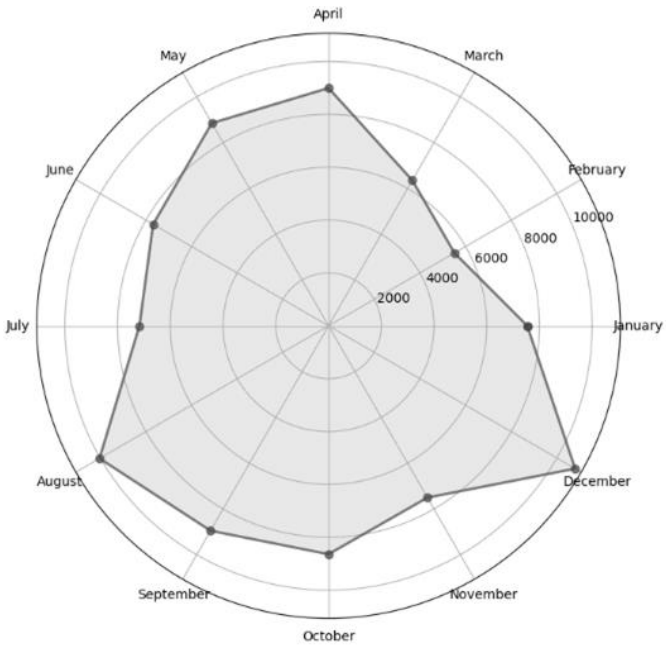
From the reviews related to Naples' attractions in 2012-2022, for example, it is possible to draw a graphical representation that can be read as a proxy of the arrival distribution (Figure 3), including the hard-to-estimate daily trips (Höpken et al., 2023). We noted that the two distributions had two inflection points, around 2014 and 2020. In the first case, there was a massive rise in reviews (and arrivals) corresponding to the increasing success of Naples as a tourism destination, also facilitated by the city's notoriety in pop culture induced by books (e.g. Elena Ferrante's books), movies (e.g. *È stata la mano di Dio* of Oscar-winner director Paolo Sorrentino) and tv series (e.g. *Un posto al sole*, *Gomorra*, *Mare Fuori*, *I bastardi di Pizzofalcone*). In the second case, the crisis caused by the COVID-19 pandemic dropped the number of visits and, hence, the reviews. Evidence shows that the two data sources have a pretty high level of coherence and could be integrated to improve the knowledge about the tourism phenomenon in a territory.

Figure 3. Yearly distribution of reviews concerning Naples' attractions (2012-2022).



Source*: authors elaboration on Tripadvisor data*

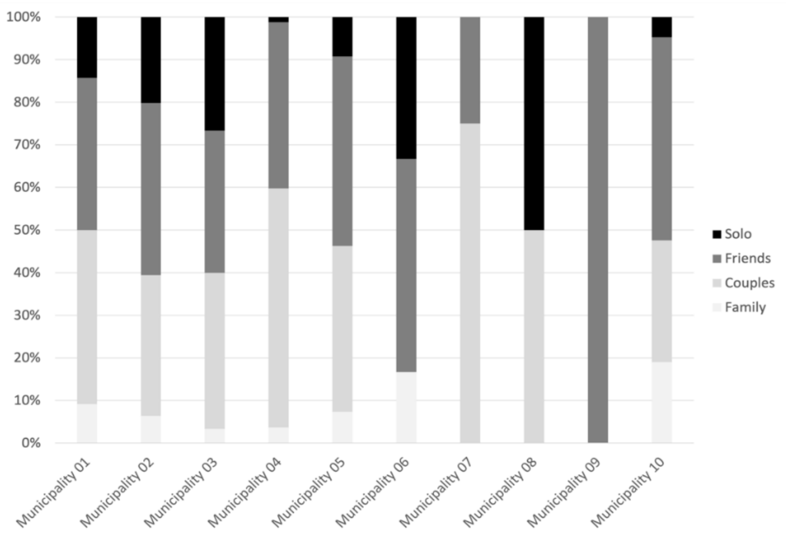
Figure 4. Monthly distribution of reviews concerning Naples' attractions (2012-2022).



Source*: authors elaboration on Tripadvisor data*

Similarly, the review date (Figure 4) allows stakeholders to overview the seasonality of the tourism phenomenon in Naples. This aspect can be valuable information in monitoring tourist flows and planning actions to avoid over-tourism and the potential shortage of quality in tourist services. As the radar chart shows, reviews decreased during winter, between January and March, with two peaks in August and December. The latter, in particular, can be explained by the Neapolitan Christmas traditions reflected in cribs' art, craftsmanship, and cuisine (Micera & Crispino, 2017).

Figure 5. Per-municipality distribution of tourist kinds in Naples (2012-2022).



***Source**: authors elaboration on Tripadvisor data*

Considering the type of visit, it is possible to explore the kind of tourists who visit the different parts of the city, offering powerful insight to tourist operators and policymakers. Figure 5 depicts the per-municipality distribution of tourist kinds, highlighting which areas of Naples individuals and groups prefer. We observed, for example, that unaccompanied tourists visit peripheral areas less often (municipalities 04, 07 and 09). Conversely, families preferred areas with the most exciting attractions for kids, like the *Interactive Science Museum* (municipality 06) and the *Railway Museum of Pietrarsa* (municipality 10). The area of Neapolitan *movida* located in the Vomero (municipality 04), an extensive neighbourhood on the hills facing the sea, was mainly preferred by couples and groups of friends instead.

3.2. The polarity detection

With a second script written in Python, 111,222 reviews in Italian were extracted for the 407 different attractions. Because of the nature of the analysed social media platform, as said above, it was not possible to collect affordable data about the socio-demographic characteristics of the different reviewers. For this reason, the subsequent sentiment analysis only considered the attractions' characteristics. For each review, we observed an average length of about 67 terms. The collection comprised 7,410,927 words with a corresponding vocabulary of 148,743 distinct terms.

Table 2. Examples of sentiment labelling performed by FEEL-IT.

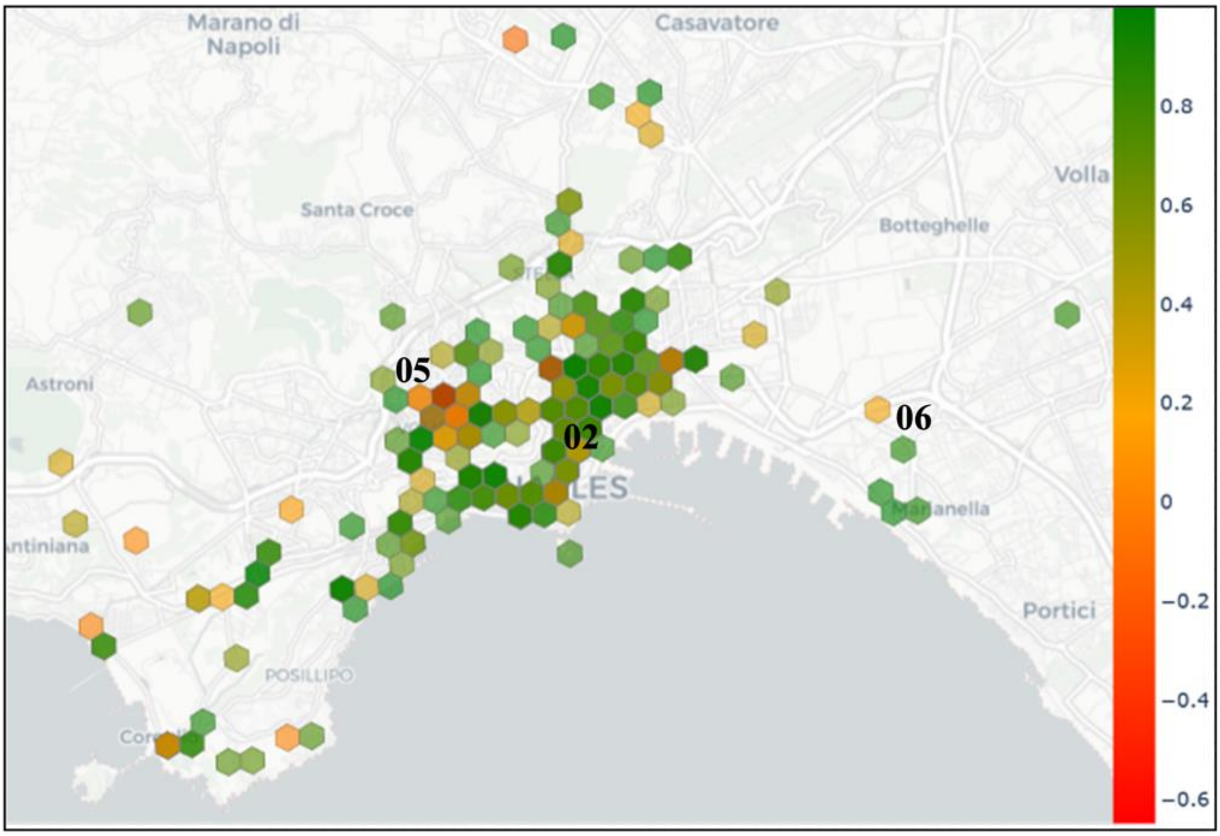
Review	Label	Probability
VISITA VELOCE SENZA LA POSSIBILITÀ DI FOTOGRAFARE. 8€ D'ENTRATA MI PARE ECCESSIVO PER QUESTA ESPERIENZA. *(Quick visit without the opportunity of taking pictures. The 8€ entrance fee seems excessive for this experience)*	NEGATIVE	0.999
BELLISSIMA!! UN MOMENTO UNICO TRA ARTE MAGIA E SCIENZA. LE STATUE UNICHE NEL LORO GENERE HANNO UN FASCINO MISTERIOSO, NON SI RIESCE A SMETTERE DI GUARDARLE. *(Beautiful!! A unique moment between art, magic, and science. The unique statues have a mysterious charm; you cannot stop looking at them)*	POSITIVE	0.999

Source*: authors elaboration on TripAdvisor data*

The polarity scoring and sentiment labelling procedure implemented in the study relied on a particular language model for sentiment analysis, known as FEEL-IT (Bianchi et al., 2021) and based on the UmBERTo model for the Italian language[3]. FEEL-IT is a collection of Italian tweets annotated with four primary emotions (*anger*, *fear*, *joy*, and *sadness*) that can be employed for emotion recognition, and by collapsing them into two dichotomous categories to express negativity/positivity in the sentiment analysis framework (Table 2). The probabilities related to negativity/positivity can be rescaled and expressed as polarity scores ranging from -1 to +1. After averaging the polarity score of reviews at a POI level, it is possible to geographically plot the sentiment and explore the spread of negativity/positivity in a spatial perspective. The polarity scores of POIs were further averaged for each hexagon in the grid guesting attractions to represent the sentiment of opinions associated with the attractions within a cell. The advantage of geo-referencing the sentiment obtained by reviews is that stakeholders have a valuable overview of the territory, discovering more appealing areas and areas with a lower level of appreciation.

[3] See https://github.com/musixmatchresearch/umberto

Figure 6. Geo-referencing of POI sentiment in the city of Naples.



Source*: authors elaboration on TripAdvisor data*

The sentiment analysis results on Neaples' attractions can be visualised on the city map in Figure 6 as a hexagonal heat map, from the most negative (in red colour) to the most positive (in green colour) level. The majority of hexagons in the city centre and along the waterfront are coloured green, indicating predominantly positive sentiment, due to their accessibility and iconic landmarks. These areas encompass popular tourist sites such as historical monuments and scenic views, contributing to a positive overall experience. A visible mix of green, yellow, and orange hexagons in the western coastal areas (near Posillipo and Coroglio) suggests a more diverse range of opinions. This variation could point to inconsistencies in service quality and environmental factors affecting the tourist experience. The hexagons towards the eastern suburbs show a slight shift towards yellow, indicating more neutral to mildly positive feedback. This could suggest that these areas are less developed as tourist hubs, possibly lacking attractions or sufficient facilities, making them less appealing than the more central regions. In some parts of the western districts, there are some orange and red hexagons, revealing that tourists have expressed negative sentiments in these areas, possibly due to unsatisfactory service. More in detail, in line with the distribution of attractions in the different neighbourhoods, the areas with positive sentiment are those located in the historic centre, covering the Avvocata, Mercato, Montecalvario, Pendino, Porto and S. Giuseppe neighbourhoods (municipality 02 on the map), rich in cultural heritage and landscape and rated positively. This area has a significant concentration of POI (e.g. the *San Gregorio Armeno street*, with hundreds of artisan workshops with colourful displays of Nativity scenes; the *Underground Naples* exhibition; the *Monumental Complex of St. Chiara*, including the church and the cloister; the *Sansevero Chapel*, showcasing baroque art and the noted *Veiled Christ* sculpture). However, in recent years, the historic centre has suffered from increasing touristification (Cerreta et al., 2020) and has experienced several episodes of over-tourism. This means that areas with a high level of tourist satisfaction must still be monitored to prevent possible future loss of image or other discomforts for the tourists and the local communities. In contrast to the poor number of attractions, the more peripheral areas of the city showed a lower positive sentiment, as in the case of Barra, Ponticelli and S. Giovanni a Teduccio (municipality 06 on the map). As highlighted by Campi (2024), the area offers a compelling example of urban regeneration, with the Pietrarsa Railway Museum showcasing the transformative impact that historical and cultural landmarks can have on the urban environment. Due to a lack of maintenance and investment, the area has faced deteriorating living conditions and unappealing landscapes. Although some progress has been made, the eastern coast still faces poor public transportation, hindering mobility for residents and tourists and limiting economic growth. The museum has the potential to revitalise the area by drawing in visitors, boosting local businesses, and contributing to the area's economic and social renewal, all while fostering sustainable development and cultural engagement. It is interesting to note that the Vomero neighbourhood (municipality 05 on the map) – although considered a high-status area – induced a negative perception in Tripadvisor users, probably due to the overabundance of clubs frequented by youngsters serving low-quality products to be more competitive. The growth of bars, eateries, and leisure spots has shifted Vomero's image to a place for "entertainment commerce" rather than cultural tourism, alienating tourists (and locals) seeking a more profound experience (D'Alessandro & Autiero, 2020). Vomero lacks a strong cultural program to attract large-scale tourist traffic, unlike other parts of Naples. While it has notable landmarks like *Sant'Elmo Castle* and the *San Martino Museum*, it is detached from the broader tourist circuits, and the neighbourhood is not effectively marketed for tourism. The area is seen as a city within a city, with a clear distinction between its residents and visitors. There is a degree of reluctance among locals to embrace events aimed at attracting outsiders, such as fairs and festivals, leading to friction and potentially limiting the area's appeal as a welcoming tourist destination. A more fine-grained analysis of the sentiment data and their visualisation on the city map – e.g. by filtering the different types of attractions – could offer several hints to the various stakeholders, helping them understand which parts of Naples deserve greater attention in terms of policies and actions for developing tourism and preserving local communities from the related possible discomfort.

4. Discussion, conclusions, and future implications

The rise of social media and the development of opinion mining techniques have significantly revolutionised how tourists make choices and decisions.

The new information and communication technologies offer a great variety of information that can guide travellers in selecting destinations, accommodations, activities, and dining options during their trips. Social media platforms are indeed rich sources of first-hand experiences shared by individuals. Tourists frequently turn to these sources to read about others' experiences before making decisions about their own choices. Nevertheless, travel planning can be overwhelming due to the many available options. Opinion mining helps reduce uncertainty by providing aggregated insights from a large volume of user feedback. This data assists tourists in feeling more confident in their decisions, knowing they are based on the collective experiences of others, helping to refine the tremendous amount of information into comprehensible insights. For instance, sentiment analysis can reveal overall satisfaction levels for diverse destinations, highlighting popular attractions and identifying common issues. For example, a potential guest might use sentiment analysis data to determine which hotels consistently receive high marks for cleanliness, customer service, or amenities. When prospective tourists read positive reviews and see captivating photos or videos of a destination, they are more likely to be influenced in their decision to visit. Contrarily, negative feedback can deter them from choosing a particular location, highlighting the importance of managing online reputations effectively.

At the same time, social media has significantly transformed the landscape of tourism management from the offer side. Opinion mining can also provide valuable insights for tourism stakeholders, enhance public engagement, and facilitate effective marketing and communication. Analysing social media data allows tourist operators and policymakers to monitor tourists' sentiments in real time, providing an immediate understanding of the public perception of various tourism-related aspects. By systematically monitoring these sentiments, tourist operators can quickly address negative feedback, improving overall tourist satisfaction. Similarly, policymakers can use these insights to identify and mitigate potential problems before they escalate, ensuring that the destination maintains a positive image and attracts more visitors. Detailed insights into tourists' preferences, behaviours, and experiences enable tourism operators and policymakers to make data-driven decisions. This can lead to developing targeted marketing campaigns, enhancing existing attractions, and creating new tourist products that align with visitors' expectations and preferences. Additionally, personalised marketing strategies informed by sentiment analysis can create a more engaging and tailored experience for tourists. If analysis reveals a growing interest in eco-tourism or cultural experiences, for example, local administrations can plan and promote relevant activities and attractions, thereby aligning their offerings with market trends. This not only enhances the tourist experience but also fosters sustainable tourism practices.

Furthermore, the insights gained from social media and opinion mining can also directly enhance the quality of services provided to tourists. Tourist operators can use feedback to improve their offerings continuously. For example, hotels and restaurants can adjust their services based on specific criticisms or suggestions mentioned in online reviews, leading to higher customer satisfaction and repeat visits. Similarly, by leveraging data derived from social media platforms, local administrations can better understand and respond to the needs of their communities, promote their territories, and manage potential crises more effectively. The strategic use of social media in territorial management and marketing can lead to more informed decision-making, more robust community relations, and greater overall success in achieving policy objectives. Integrating opinion mining and social media analytics into tourism management also benefits the local community. By understanding tourists' sentiments, local authorities can plan infrastructure improvements and public services that supply tourists, improving, at the same time, residents' quality of life. For instance, if tourists frequently complain about transportation issues, local governments can prioritise investments in public transport systems, which will benefit tourists and locals. Data-driven tourism development can lead to more equitable economic benefits for the local community. Policymakers can disperse tourist traffic more evenly across the territory by identifying and promoting lesser-known attractions based on positive social media feedback. This can help alleviate pressure on over-visited sites and broadly distribute economic gains, supporting local businesses and communities.

This paper showed how sentiment analysis and territorial analysis can be integrated to provide stakeholders a powerful tool to collect information and to monitor an area of tourist interest, taking into account the opinions and the experiences of tourists. The study is not without limitations.

First of all, data collection requires an articulated system based on web scraping, often not available or not directly implementable in open-source software and platforms. Moreover, as reported by several authors (e.g. Pagallo & Ciani Sciolla, 2023), the collection of data without an API (not ever free of costs) or a specific authorisation in the social media terms of services may lead to ethical and legal problems. Nevertheless, the proposed strategy may be used also on primary data collected through traditional surveys.

Another possible limitation concerns the linguistic resources required for sentiment analysis. As discussed in the paper, it is necessary to have lexicons or large linguistic models in the language of the analysed texts. It is very easy to find resources for the English language and for the other main languages (e.g., French or Spanish), but not equally easy to find effective resources for the Italian language. Additionally, main resources are for general purpose and not useful in specific domains like tourism. Clearly, the choice of a not suitable linguistic resource affects all the subsequent analysis, from polarity scoring to sentiment classification. In an economic perspective, it means that the results obtained from such analyses may be deceptive and lead to unproductive decisions, both in a private and public context. Finally, there is a dark side of using social media as a data source. The presence of fake reviews represents a significant challenge, undermining the trustworthiness of online platforms and potentially leading to misguided decisions. Fake reviews may be "artificially" generated comments or ratings intended to mislead consumers. These can be positive reviews aimed at boosting the reputation of a business or a location or negative ones intended to harm competitors. The tourism industry, which includes hotels, restaurants, attractions, and travel services, is particularly susceptible to this issue. Several studies found that a significant percentage of reviews on popular platforms may be fake (e.g. Choi et al., 2017; Tuomi, 2021). Several measures are being taken to address the problem. Review platforms like TripAdvisor, Yelp, and Google have implemented algorithms and manual review processes to detect and remove fraudulent content. These systems analyse patterns in review submissions, such as the frequency of reviews, the timing, and the reviewer's profile, to identify suspicious activity. Additionally, legal frameworks are being strengthened to combat fake reviews. For instance, in some jurisdictions, posting fake reviews is considered a form of deceptive marketing and can result in legal penalties for involved businesses and individuals. The progress in artificial intelligence is playing a crucial role. These technologies can analyse vast amounts of data to identify anomalies and patterns indicating fraudulent behaviour.

In conclusion, opinion mining and social media have become indispensable tools for supporting tourists' choices and decisions and tourism stakeholders' actions. By providing aggregated insights from user-generated content, these technologies offer valuable recommendations, enhance the travel planning process, and reduce uncertainty, leading to more satisfying travel experiences. As these technologies evolve, their impact on tourism and related decision-making processes will strengthen. By fostering greater transparency and awareness, the tourism industry and the local governments can work towards maintaining the integrity of the review systems, ensuring that tourists can make well-informed decisions. From a methodological viewpoint, integrating sentiment and territorial analysis may represent a stimulating research front, as shown in some recent studies (Celardo et al., 2024).

The proposed strategy can be developed both methodologically and practically. From a technical viewpoint, an interesting improvement is represented by the integration of sentiment with keywords characterising the different areas, leading to interactive maps that can provide additional insight concerning the various aspects that induce negative or positive tourist experiences. Furthermore, a dynamic analysis involving the temporal dimension may help in understanding if the sentiment modifies according to corrective actions pursued by local policymakers or marketers to promote the territory. From an applicative viewpoint, extending the strategy to multiple territories can help scholars and practitioners recognise the strengths and weaknesses of an area of tourist interest in the logic of territories' competitiveness analysis and regional studies, providing a more comprehensive framework to all the stakeholders.

Acknowledgements

We would like to thank Antonella Bianchino (ISTAT) and the ISTAT Territorial Office for Southern Area for organising the meeting *Tourism, Overtourism, and Territories: Statistics for Public Decision-Makers*, held in Reggio Calabria (Italy) on 4 June 2024, where this study was initially presented.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

Alaei, A.R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: Capitalizing on big data. *Journal of Travel Research*, *58*(2):175-191; DOI: https://doi.org/10.1177/0047287517747753

Arici, H.E., Cakmakoglu Arıcı, N. & Altinay, L. (2022). The use of big data analytics to discover customers' perceptions of and satisfaction with green hotel service quality. *Current Issues in Tourism*, 26(2):270-288; DOI: https://doi.org/10.1080/13683500.2022.2029832

Baccianella, S., Esuli, A. & Sebastiani, F. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In: Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M. & Tapias., D. (eds.) *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, ELRA, 2200-2204

Balbi, S., Misuraca, M. & Scepi, G. (2018). Combining different evaluation systems on social media for measuring user satisfaction. *Information Processing & Management*, *54*(4):674-685; DOI: https://doi.org/10.1016/j.ipm.2018.04.009

Bengio, Y., Ducharme, R., Vincent, P. & Janvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, *3*:1137-1155

Bianchi, F., Nozza, D. & Hovy D. (2021). FEEL-IT: Emotion and Sentiment Classification for the Italian Language. In: De Clercq, O., Balahur, A., Sedoc, J., Barriere, V., Tafreshi, S., Buechel, S. & Hoste V. (eds.) *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, ACL, 76–83

Bing, L., Minqing, H. & Junsheng C. (2005). Opinion observer: analyzing and comparing opinions on the Web. In: *Proceedings of the 14th international conference on World Wide Web*, ACM, New York, 342-351. DOI: https://doi.org/10.1145/1060745.1060797

Bravo-Marquez, F., Mendoza, M. & Poblete, B. (2014). Meta-level sentiment models for big social data analysis. *Knowledge-Based Systems*, *69*:86-99; DOI: https://doi.org/10.1016/j.knosys.2014.05.016

Bronner, F. & de Hoog, R. (2014). Social media and consumer choice. *International Journal of Market Research*, *56*(1):51-71; DOI: https://doi.org/10.2501/IJMR-2013-053

Buonincontri, P., Morvillo, A., Okumus, F. & van Niekerk, M. (2017). Managing the experience co-creation process in tourism destinations: Empirical findings from Naples, *Tourism Management*, *62*:264-277; DOI: https://doi.org/10.1016/j.tourman.2017.04.014

Campi, M. (2024). The regenerative potential of Point of Interest (POIs) in urban renewal: the Pietrarsa Railway Museum for the coast of San Giovanni a Teduccio. Industrial Archaeology. In: Capone, M., Galván Desvaux, N., Agustin-Hernandez, L. & Fernádez-Trapa, L. (eds.) *European approach to recovery productive memory*, FedOAPress, Napoli, 297-316

Carvache-Franco, O., Carvache-Franco, M., Carvache-Franco, W. & Iturralde, K. (2023). Topic and sentiment analysis of crisis communications about the COVID-19 pandemic in Twitter's tourism hashtags. *Tourism and Hospitality Research*, *23*(1):44-59; DOI: https://doi.org/10.1177/14673584221085470

Catelli, R., Pelosi, S. & Esposito, M. (2022). Lexicon-Based vs. Bert-Based Sentiment Analysis: A Comparative Study in Italian. *Electronics*, *11*(3):374; DOI: https://doi.org/10.3390/electronics11030374

Celardo, L., Misuraca, M. & Spano, M. (2024). Geo-referenced sentiment analysis for tourists' points of interest: the case of Matera European Capital of Culture, *Rivista di Economia, Demografia e Statistica*, *78*(2):103-114

Cerreta, M., Mura, F.D. & Poli, G. (2020). Assessing the Interstitial Rent: The Effects of Touristification on the Historic Center of Naples (Italy). In: Gervasi, O., Murgante, B., Misra, S., Garau, C., Blečić, I., Taniar, D., Apduhan, B.O., Rocha, A.M.A.C., Tarantino, E., Torre, C.M. & Karaca Y. (eds.) *Computational Science and Its Applications – ICCSA 2020*, Springer, Cham, 952-967. DOI: https://doi.org/10.1007/978-3-030-58808-3_68

Chang, V., Liu, L., Xu, Q., Li, T. & Hsu, C-H. (2023). An improved model for sentiment analysis on luxury hotel review. *Expert Systems*, *40*(2):e12580; DOI: https://doi.org/10.1111/exsy.12580

Choi, S., Mattila, A.S., Van Hoof, H.B. & Quadri-Felitti, D. (2017). The Role of Power and Incentives in Inducing Fake Reviews in the Tourism Industry. *Journal of Travel Research*, *56*(8):975-987; DOI: https://doi.org/10.1177/0047287516677168

D'Alessandro, L. & Autiero, A. (2020). Retail Changes and Consumption Practices in the Neighbourhood of Vomero, Naples. *Bollettino della Società Geografica Italiana*, *14*(3):25-35; DOI: https://doi.org/10.36253/bsgi-1008

Daas, P.J.H., Puts, M.J., Buelens, B. & van den Hurk, P.A.M. (2015). Big Data as a Source for Official Statistics. *Journal of Official Statistics*, *31*(2):249-262; DOI: https://doi.org/10.1515/jos-2015-0016

De Falco, S. & Corbino, A. (2023). Naples and tourism: conflicts of a dream realised? Analysis of a fast-changing urban landscape. *AIMS Geosciences*, *9*(4):754-768; DOI: https://doi.org/10.3934/geosci.2023040

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT - Pre-training of deep bidirectional transformers for language understanding. In: Burstein, J., Doran, C., & Solorio, T. (eds.) *Proceedings of NAACL-HLT 2019*, ACL, Minneapolis, 4171-4186; DOI: https://doi.org/10.18653/v1/N19-1423

Effing, R., & Spil, T.A.M. (2016). The social strategy cone: Towards a framework for evaluating social media strategies. *International Journal of Information Management*, *36*(1):1-8; DOI: https://doi.org/10.1016/j.ijinfomgt.2015.07.009

Esposito, A. (2023). Tourism-driven displacement in Naples, Italy. *Land Use Policy*, *134*:106919; DOI: https://doi.org/10.1016/j.landusepol.2023.106919

Filieri, R., & McLeay, F. (2014). E-WOM and accommodation: An analysis of the factors that influence travelers' adoption of information from online reviews. *Journal of Travel Research*, *53*(1):44-57; DOI: https://doi.org/10.1177/004728751348127

Fischer, E. & Reuber, A.R. (2011). Social interaction via new social media: (How) can interactions on Twitter affect effectual thinking and behavior? *Journal of Business Venturing*, *26*(1):1-18; DOI: https://doi.org/10.1016/j.jbusvent.2010.09.002

Flores-Ruiz, D., Elizondo-Salto, A. & Barroso-González, M. de la O. (2021). Using Social Media in Tourist Sentiment Analysis: A Case Study of Andalusia during the COVID-19 Pandemic. *Sustainability*, *13*(7):3836; DOI: https://doi.org/10.3390/su13073836

Gretzel, U., & Yoo, K.H. (2008). Use and impact of online travel reviews. In: O'Connor, P., Höpken, W., & Gretzel, U. (eds.) *Information and Communication Technologies in Tourism 2008*, Springer, Vienna, 35-46; DOI: https://doi.org/10.1007/978-3-211-77280-5_4

Godovykh, M., Ridderstaat, J., Baker, C. & Fyall, A. (2021). COVID-19 and Tourism: Analyzing the Effects of COVID-19 Statistics and Media Coverage on Attitudes toward Tourism. *Forecasting*, *3*(4):870-883; DOI: https://doi.org/10.3390/forecast3040053

Guillamón, M.-D., Ríos, A.-M., Gesuele, B. & Metallo, C. (2016). Factors influencing social media use in local governments: The case of Italy and Spain, *Government Information Quarterly*, *33*(3):460-471; DOI https://doi.org/10.1016/j.giq.2016.06.005

Hays, S., Page, S. J., & Buhalis, D. (2013). Social media as a destination marketing tool: Its use by national tourism organizations. *Current Issues in Tourism*, *16*(3):211-239; DOI: https://doi.org/10.1080/13683500.2012.662215

Hennig-Thurau, T., Gwinner, K.P., Walsh, G., & Gremler, D.D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, *18*(1):38-52; DOI: https://doi.org/10.1002/dir.10073

Höpken, W., Regitz, D., Liedtke, N. & Fuchs, M. (2023). Estimating Tourist Arrivals by User Generated Content Volume in Periods of Extraordinary Demand Fluctuations. In: Ferrer-Rosell, B., Massimo, D. & Berezina, K. (eds) *Information and Communication Technologies in Tourism 2023*, Springer, Cham, 231-242; DOI: https://doi.org/10.1007/978-3-031-25752-0_25

Iovino, G. (2021). Affitti turistici a breve termine nelle città d'arte. La "airificazione" di Napoli e i suoi impatti. *Bollettino dell'Associazione Italiana di Cartografia*, *172*:4-19

Jiao, Q., & Zhang, S. (2021). A Brief Survey of Word Embedding and Its Recent Development. In: *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, IEEE, Chongqing, 1697-1701; DOI: https://doi.org/10.1109/IAEAC50856.2021.9390956

Kaplan, A.M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, *53*(1):59-68; DOI: https://doi.org/10.1016/j.bushor.2009.09.003

Kim, W.G., Li, J.J., & Brymer, R.A. (2016). The impact of social media reviews on restaurant performance: The moderating role of excellence certificate. *International Journal of Hospitality Management*, *55*:41-51; DOI: https://doi.org/10.1016/j.ijhm.2016.03.001

La Rocca R.A. (2021). Urban Accessibility and Tourist Activity: An Application to the Metropolitan City of Naples. In: La Rosa, D. & Privitera, R. (eds.) *Innovation in Urban and Regional Planning*, Springer, Cham, 583-591; DOI: https://doi.org/10.1007/978-3-030-68824-0_62

Li, H., Gao, H. & Song, H. (2023). Tourism forecasting with granular sentiment analysis. *Annals of Tourism Research*, *103*:103667; DOI: https://doi.org/10.1016/j.annals.2023.103667

Li, X., Law, R., Xie, G. & Wang, S. (2021). Review of tourism forecasting research with internet data, *Tourism Management*, *83*:104245; DOI: https://doi.org/10.1016/j.tourman.2020.104245

Litvin, S.W., Goldsmith, R.E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, *29*(3):458-468; DOI: https://doi.org/10.1016/j.tourman.2007.05.011

Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, *47*:140-151; DOI: https://doi.org/10.1016/j.tourman.2014.09.020

Marine-Roig, E., & Clavé, S.A. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing & Management*, *4*(3):162-172; DOI: https://doi.org/10.1016/j.jdmm.2015.06.004

Micera, R. & Crispino, R. (2017). Destination web reputation as "smart tool" for image building: the case analysis of Naples city-destination. *International Journal of Tourism Cities*, *3*(4):406-423; DOI: https://doi.org/10.1108/IJTC-11-2016-0048

Mikolov, T., Sutskever, I, Chen, K, Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *arXiv*; DOI: https://doi.org/10.48550/arXiv.1310.4546

Migliaccio, G., Lausi, C. & Baldi, M. (2023). Digital tourism and sustainable development of rural and inland areas with food and wine tourism: the role of web portals. *Turistica - Italian Journal of Tourism*, *32*(3):79-107

Misuraca, M. & Spano, M. (2020). Unsupervised Analytic Strategies to Explore Large Document Collections. In: Iezzi, D.F., Mayaffre, D. & Misuraca, M. (eds.) *Text Analytics. Advances and Challenges*, Springer, Cham, 17-28; DOI: https://doi.org/10.1007/978-3-030-52680-1_2

Neubaum, G., &Krämer, N.C. (2016). Monitoring the Opinion of the Crowd: Psychological Mechanisms Underlying Public Opinion Perceptions on Social Media. *Media Psychology*, *20*(3):502-531; DOI: https://doi.org/10.1080/15213269.2016.1211539

Nielsen, F.Å. (2011). A new anew: Evaluation of a word list for sentiment analysis in microblogs. In: Rowe, M., Stankovic, M., Dadzie, A.-S. & Hardey, M. (eds.) *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages*. CEUR Workshop Proceedings, *718*:93-98

Pabel, A. & Prideaux, B. (2016). Social media use in pre-trip planning by tourists visiting a small regional leisure destination. *Journal of Vacation Marketing*, *22*(4):335-348; DOI: https://doi.org/10.1177/1356766715618998

Pagallo, U. & Ciani Sciolla, J. (2023). Anatomy of web data scraping: ethics, standards, and the troubles of the law. *European Journal of Privacy Law & Technologies*, 2023(2):1-19; DOI: https://doi.org/10.57230/EJPLT232PS

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1-2):1-135; DOI: http://dx.doi.org/10.1561/1500000011

Sahr, K. (2011). Hexagonal discrete global grid systems for geospatial computing. *Archives of Photogrammetry, Cartography and Remote Sensing*, *22*:363-376

Salton, G., Wong, A., & Yang, C.S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, *18*(11):613-620; DOI: https://doi.org/10.1145/361219.361220

Siering, M., Muntermann, J., & Rajagopalan, B. (2018). Explaining and predicting online review helpfulness: The role of content and reviewer-related signals. *Decision Support Systems*, *108*:1-12; DOI: https://doi.org/10.1016/j.dss.2018.01.004

Tuomi, A. (2021). Deepfake consumer reviews in tourism: Preliminary findings. *Annals of Tourism Research Empirical Insights*, *2*(2):100027; DOI: https://doi.org/10.1016/j.annale.2021.100027

Wang, Y., & Fesenmaier, D.R. (2004). Modeling participation in an online travel community. *Journal of Travel Research*, *42*(3):261-270; DOI: https://doi.org/10.1177/0047287503258824

Xiang, Z., Schwartz, Z., Gerdes, J.H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, *44*:120-130; DOI: https://doi.org/10.1016/j.ijhm.2014.10.013

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, *28*(1):180-182; DOI: https://doi.org/10.1016/j.ijhm.2008.06.011