

AIIT 3rd International Conference on Transport Infrastructure and Systems (TIS ROMA 2022),  
15th-16th September 2022, Rome, Italy

## A sentiment analysis approach to investigate tourist satisfaction towards transport systems: the case of Mount Etna

Juan Pineda-Jaramillo<sup>a\*</sup>, Martina Fazio<sup>b</sup>, Michela Le Pira<sup>c</sup>, Nadia Giuffrida<sup>d</sup>, Giuseppe Inturri<sup>e</sup>, Francesco Viti<sup>a</sup>, Matteo Ignaccolo<sup>c</sup>

<sup>a</sup>Department of Engineering, University of Luxembourg, 4365, Esch-sur-Alzette, Luxembourg

<sup>b</sup>University of Catania, Department of Physics and Astronomy, Via Santa Sofia 64, Catania, 95125, Italy

<sup>c</sup>University of Catania, Department of Civil Engineering and Architecture, Via Santa Sofia 64, Catania, 95125, Italy

<sup>d</sup>Polytechnic University of Bari, Via Edoardo Orabona, 4, 70126 Bari BA, Italy

<sup>e</sup>University of Catania, Department of Electrical, Electronic and Computer Engineering, Via Santa Sofia 64, Catania, 95125, Italy

---

### Abstract

This paper aims at analysing tourist satisfaction towards transport systems of the touristic destination using online reviews from TripAdvisor, the most well-known online review platform used in the tourism sector for researchers and customers. First, a web scraping method was used to retrieve online reviews available on TripAdvisor related to transport systems, then the Latent Dirichlet Allocation method was implemented to extract the topics of the online reviews, and finally an analysis was performed to find similarities and differences among positive and negative reviews for each topic in order to identify tourist trends towards transport systems. To test the methodology, the case of Mount Etna has been considered, one of the main tourist destinations in Sicily (Italy), located on the east coast of the island. According to the results of our case study, the reviews can be clustered in positive and negative predicted feelings, and, for each of them 2 topics (Topic 0 and Topic 1) are linked. Topics are related to specific tourist experience characteristics and the differences between positive and negative reviews can be ascribed to a different relationship between transport and tourism concepts. Main results show a positive attitude of tourists towards the of transport as an integrated experience of the visit of the touristic location. The proposed approach is able to provide useful information to decision-makers by identifying the possible criticalities and intervention priorities considering user judgment. Besides, this methodology can be easily applied to other tourist venues and can be used by decision-makers to provide better services with the aim of improving the engagement of tourists with the city and increasing profits.

© 2023 The Authors. Published by ELSEVIER B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the Transport Infrastructure and Systems (TIS ROMA 2022)

---

\* Corresponding author. Tel.: +352 66122239.

E-mail address: [juan.pineda@uni.lu](mailto:juan.pineda@uni.lu)

**Keywords:** Machine learning; Big data analysis; Tourist' mobility; Tourist' trends; Transport planning.

## 1. Introduction

Tourism is one of the main economic activities in the world. This pushes public administration to find increasingly innovative solutions to improve the services offered, including the quality of transport systems. Transport and tourism are strictly interwoven and influence each other: the touristic experience of a city is strictly related to the fruition of its transport system, while touristic flows affect the urban transport system (Le Pira et al., 2021). In this respect, Page (2004) identified two types of relationships between transport and tourism. “Transport for tourism” implies that transport is a means to reach a destination, without any intrinsic value, only representing a cost. “Transport as tourism” considers transport as a basis for the tourist experience and, thus, it can also be characterised by some positive attributes.

Tourism is a people-intensive phenomenon that demands high levels of mobility, and tourist behaviour is an important component. This points to the need to define *ad-hoc* policies aimed at raising awareness to promote responsible tourist behaviours (Budeanu, 2007). Public administrators should focus on analysing what tourists want in order to provide attractive and sustainable mobility solutions that can generate engagement and produce significant economic benefits, while reducing the negative transport externalities tourism can generate (Hall, 1999). The role of public transport is fundamental to this purpose. In general, increasing its service quality in order to attract more users is one of the main goals of transport operators and policy-makers (Inturri et al., 2021). However, it is rarely adapted to the needs of tourists and to accommodate additional demand from tourism, even if it implies some extra income (Albalade and Bel, 2010). Infrastructures are also important to determine public transport success. In this respect, public transport nodes should be both accessible and safe (Rossetti et al., 2020; Rossetti and Tiboni, 2020). Data are needed to monitor the actual and perceived quality of service, but the acquisition process can be costly (Le Pira et al., 2018). Inturri et al. (2021) searched for correlations between demand and supply data to reduce the need for surveys. They investigated the correlation among public transport use, user satisfaction and accessibility using a spatial and statistical approach to find useful and simple indicators for sustainable mobility planning. User perception of service quality could vary according to several variables, including the aim of the trip and the type of service (Bordagaray et al., 2014). In this respect, the difference between tourist and citizen preferences can be remarkable.

Given the complexity of data acquisition, especially for tourists, new sources of data like the ones generated by users or sensors can provide a valuable contribution to monitor important characteristics of the quality of service and give suggestions for the improvement of transport services. Fazio et al. (2020) focused on the analysis of on-board comfort based on the collection of kinematic data (accelerations along three axes), through a sensor installed inside public vehicles, showing the differences between a bus and a metro service. These types of analysis should be coupled with data generated by the users, both voluntarily and involuntarily (e.g. reviews or Google data on crowding levels).

Recent technologies make it possible to have large amounts of data available on the user experience, which can be of great help for decision-makers to improve services, even dynamically and in real time. Hence there is the need of proper methods and tools to manage and analyse such big datasets. In this respect, machine learning techniques allow to carry out sophisticated data analysis, providing an important support to several transport planning processes (Pineda-Jaramillo 2019; Giuffrida et al., 2022).

Based on this concept, the aim of this paper is to investigate the relationship between transport and tourism based on the analysis of data generated by users and related to the reviews of touristic attractions. The study draws inspiration from Pineda-Jaramillo and Pineda-Jaramillo (2021), where the Authors analysed through a machine learning approach the reviews of users of the metro systems of Paris with the final aim of predicting their travel satisfaction and creating a methodology that can be used by metro managements to deliver a better service to improve satisfaction among the tourists using their metro systems. In this paper, the case of Mount Etna has been considered, one of the main tourist destinations in Sicily (Italy), and the objective is to analyse the aspects associated with a specific tourist destination, identifying weaknesses (and strengths) related to the transport systems that can be improved to promote the tourist destination from the public administration. In the following, the method to collect and analyse data is presented. Then, the case study is introduced, and main results discussed together with future research directions.

## 2. Method

In order to infer tourist satisfaction towards transport systems, a sentiment analysis has been carried out throughout the use of the topic modelling process. Sentiment analysis is a computational analysis of feelings and opinions of users expressed in terms of “texts” and shared in WordNet (Kanakaraj and Guddeti, 2015). This type of analysis, also called opinion mining, is widely used in various fields: product pricing, election forecasting, marketing strategies (Solangi et al., 2018). It allows to gather information on the feelings of users in specific contexts and then define targeted strategic interventions to improve collective opinion. The framework of the methodology can be observed in Fig. 1 together with the main outputs that we derived from each step.

In the following, we provide a more detailed description.

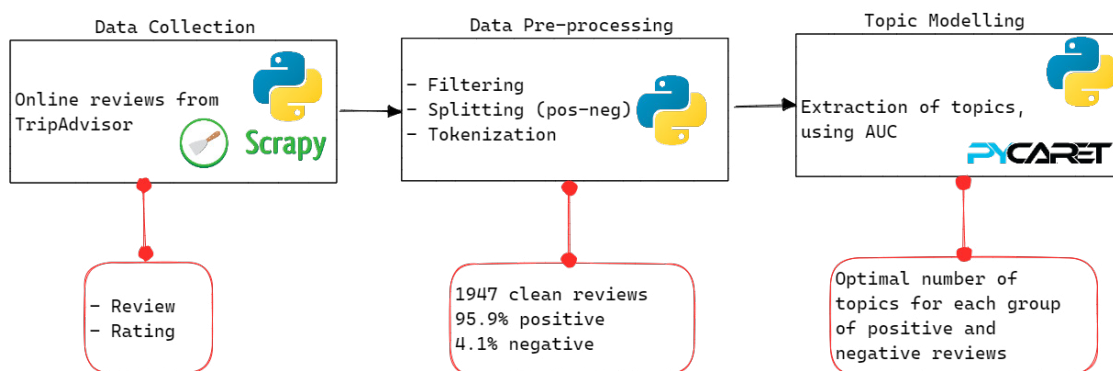


Fig. 1. Framework of the methodology.

### 2.1. Data Collection

The collection and extraction of the data was carried out through web scraping. Commonly, data displayed by majority of websites can be seen through a web browser and a direct access to the data is not provided. In this respect, web scraping is a methodology that allows to obtain raw information from a website through a computer software program.

Data collection from a website can be done manually, through an Application Programming Interface (API), or through a web scraper, i.e. an algorithm that collect data by sending requests to the Webserver. In addition, it can transform unstructured data into structured data by using several automation methods known as parsing, which implements techniques such as regular expressions, XPath, and CSS selectors, achieving the data to be extracted and stored in database in a structured way such as JSON or CSV files (Adhinugroho et al., 2020). In this study, an algorithm was developed to perform web scraping using Python and Scrapy, allowing to obtain the reviews and the rating given by the users on TripAdvisor. Online reviews posted on TripAdvisor are a combination of user-generated content and a rating system based on the number of “stars” (from 1 to 5) assigned by the user (Simeon et al., 2017).

### 2.2. Sentiment Analysis

After performing the pre-processing of the data, the sentiment analysis approach was carried out to understand users’ satisfaction. Sentiment analysis is a technique of the so-called Natural Language Processing (NLP), a field of computer science that deals with the extraction of meaning from text.

In order to determine the sentiment analysis, the Latent Dirichlet Allocation (LDA) algorithm is used to identify topics in the collected text through a process known as “topic modelling”. The reviews are classified in positive and negative and then the reviews can be clustered according to the belonging to a topic. Regarding the classification, “neutral” reviews (i.e., 3 of 5) are not included, since the objective of this study is to find the weaknesses in which

public administrations could intervene in order to improve the transport services associated with this tourist site. Finally, the determination of the feelings of the reviews can be inferred (Taecharungroj & Mathayomchan, 2019).

The area under the curve (AUC) of the receiver operating characteristic (ROC) curve was adopted to determine the optimal number of topics in both groups of positive and negative reviews, where the LDA method allows to estimate the likelihood of a review of belonging to a specific topic. The final output of this analysis is a word-cloud for each positive/negative topic.

### 3. Case study and Data

For this study, the case of Mount Etna has been investigated. Mount Etna is the tallest active volcano in Europe (currently 3,329 m high)<sup>†</sup> and one of the main Sicilian tourist destinations, located on the eastern coast of Sicily (Fig. 2). It is suitable for different types of excursion and travel experiences: using cable car, jeep, hiking, climbing, mountaineering, trekking and skiing.

The aim of this study is to understand tourist satisfaction towards transport options for the Mount Etna. From TripAdvisor, total of 3,419 online reviews in English and related to the Mount Etna were extracted. To perform the sentiment analysis, three main steps were involved: (i) filtering the reviews using transport-related keywords (i.e., transport, car, cable, rail, bus, transportation, mobility bike); (ii) splitting the data into positive and negative reviews to further analyse the sentiment of the reviews; (iii) performing the tokenization of the reviews, which involved techniques such as stemming, lemmatization, lowercasing, and punctuation/stop words removal (Pineda-Jaramillo & Pineda-Jaramillo, 2021). After these steps, a final dataset of 1,947 reviews was obtained. Then, the LDA method was applied using a word matching on reviews in order to identify topics within the text. These topics are the latent constructs distributed over a vocabulary of words which tourists use to describe their sentiments towards the transport systems in Mount Etna. The sentiment analysis performed returned two topics (Topic 0 and Topic 1) for both positive and negative feelings, for a total of 4 word-clouds. In the next section, a description of the results is presented.



Fig. 2. Mount Etna location.

### 4. Results

The 4 word-clouds resulting from the sentiment analysis are reported in Fig. 3 and 4.

Fig. 3 can be considered as representative of the “Transport as tourism” concept, and it shows some positive attitude towards it. More in detail, Fig. 3a includes terms related to place-time considerations (Topic 0), e.g. “high”, “hour”, “place”, “park”, “jeep”, “reach”, “road”. Fig. 3b includes visit-related words (Topic 1), e.g. “eruption”, “active”,

<sup>†</sup><https://earth.esa.int/web/earth-watching/image-of-the-week/content/-/article/mount-etna-volcano-sicily-italy/>



“area”, “place”, “climb”, “hike”, “stop”. The adventure experience on site emerges from both topics, and the main transport references are related to adventure and physical activity (e.g. “jeep”, “hike”, “climb”).

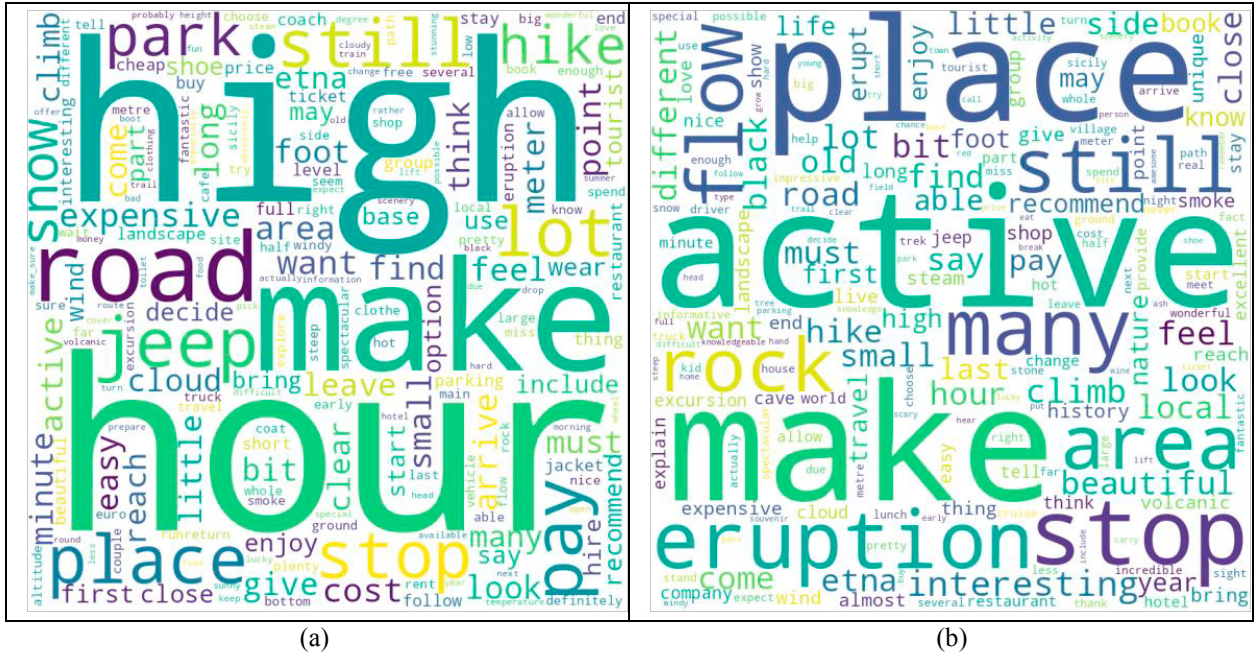


Fig. 3. (a) Reviews positive Topic 0; (b) Topic 1.

Fig. 4 can be related to the “Transport for tourism” concept and shows a negative attitude of the reviews. In particular, Fig. 4a is related to cost-time considerations (Topic 0), with the emergence of terms as “hour”, “minute”, “ticket”, “pay”, “arrive”, “stop”, while Fig. 4b only to cost considerations (Topic 1), including terms as “pay”, “money”, “cost”, “expensive”, “Jeep”.

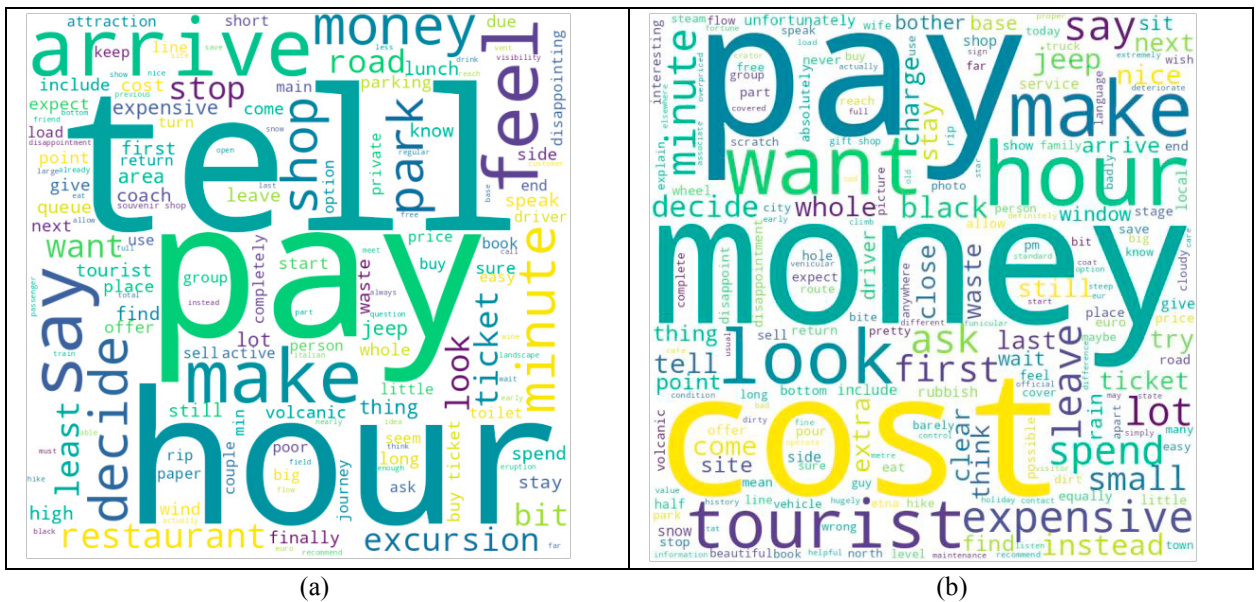


Fig. 4. (a) Reviews negative Topic 0; (b) Topic 1.

Fig. 3 and Fig. 4 show the attitudes of users towards two different ways of experiencing transport systems. Both positive and negative word-clouds have similarities especially related to cost-time considerations.

Table 1 shows the words (1-gram) and sequence of most frequent words (bigram and trigram) for each word cloud. Combination of more words were not considered since they would not have been significant for the interpretation of the results.

Table 1. 1-gram, bigram and trigram for each word cloud.

n-gram	positive_topic0		positive_topic1		negative_topic0		negative_topic1	
	Word	Number	Word	Number	Word	Number	Word	Number
1-gram	hour	383	eruption	82	tell	16	money	20
	high	315	active	80	pay	15	pay	20
	pay	312	make	78	arrive	14	cost	18
	Lot	264	place	76	hour	14	hour	15
	road	261	stop	75	ticket	13	want	15
	Make	259	many	71	shop	13	make	13
bigram	buy ticket	52	recent eruption	8	buy ticket	5	waste money	4
	parking lot	35	vantage point	6	souvenir shop	4	gist shop	3
	must sicily	28	must sicily	6	decide first	2	shop sell	3
	hiking boot	28	year old	6	feel short	2	make sure	3
	next level	27	force nature	5	min arrive	2	window scratch	2
	pay parking	26	still active	5	star excursion	2	next jeep	2
parking area	25	small town	5	ticket queue	2	jeep leave	2	
trigram	due high wind	6	eruption last year	2	black ash scree	2	NA	/
	still lot snow	5	soil support extensive	2	shop bar restaurant	2	NA	/
	speak several language	5	sun set dinner	2	NA	/	NA	/
	hire boot jacket	5	little bit extensive	2	NA	/	NA	/
	pay include jeep	4	volcanic soil support	2	NA	/	NA	/
	wear comfortable shoe	4	pass small town	2	NA	/	NA	/
	buy full ticket	4	last big eruption	2	NA	/	NA	/

As can be seen from Table 1, positive transport-related judgments rely on all those activities that are part of the visit, such as “hiking” (15.5% of the reviews) and “jeep” (15.3% of the reviews); while the negative feelings refer to costs and how the destination is reached (e.g., “ticket” with 12.5% of the reviews, and “arrive” with 11.6% of the reviews).

In summary, the word-clouds show a positive attitude of users towards the “Transport as a tourism” experience, while a negative attitude towards the “Transport for tourism” one. The results drawn from the world cloud analysis are in line with the actual transport supply to reach the Mount Etna. In fact, there is only a single morning ride on the public transport line that connects the tourist destination with the nearest city (Catania) and takes about an hour to reach the only final stop; conversely, there are several private tours that connect the city with Mount Etna, but obviously this type of trip involves a fairly high cost. Once at their destination, however, users can take advantage of various services, also related to transport (e.g. bike rental, ropeway) and often organized by the private tour operators themselves, which ensure a pleasant tourist experience. The results of the analysis can be of help for decision-makers and transport planners to improve the accessibility of the tourist destination: to make Mount Etna more accessible, it could be considered to increase public transport runs and the number of stops for reach the destination.

## 5. Conclusions

Tourist satisfaction towards transport systems is often overlooked by decision-makers. In this paper, we addressed this issue through a sentiment analysis. Sentiment analysis is a field of NLP analysis, which is one of the well-known machine learning approaches. To test the proposed methodology, the case of Mount Etna volcano (Sicily) was considered and the relative reviews from TripAdvisor were retrieved.

On the base of tourists' experience and their predicted feelings, a good level of satisfaction emerged linked to the on-site experience. In this case, transport systems can be associated to the "Transport as tourism" concept. On the other hand, the negative predicted feelings highlighted a clear need to improve the level of transport connection to reach the destination. In this respect, we referred to the "Transport for tourism" concept.

In conclusion, this paper proposes a machine learning approach able to provide useful insights regards tourist experience related not only to the experience on site, but also to their "transport" experience, and identify the priority of intervention with the aim of improving tourist satisfaction.

Within the limitations of the present study, it should be highlighted that the use of a single platform to collect comments can lead to platform bias (Xiang et al., 2017), even though the methodology presented in this study can be easily applied to comments that could be collected from different online platforms. Another recurring limitation in this type of study is the fact that there may be a bias in the average appreciation between the average tourist, and the tourist who usually writes this type of online reviews on platforms such as TripAdvisor. Future research could include conducting surveys to complement this study, where key aspects of problems related to "transport for tourism" can be further analysed.

## Acknowledgements

The work was partially supported by the project "WEAKI-TRANSIT: WEAK-demand areas Innovative TRANsport Shared services for Italian Towns" (unique project code: E44I17000050001) under the programme "PRIN 2017", by the project of M. Le Pira "AIM Linea di Attività 3 – Mobilità sostenibile: Trasporti" (unique project CODE CUP E66C180013890007) under the programme "PON Ricerca e Innovazione 2014-2020–Fondo Sociale Europeo", Azione 1.2 'Attrazione e mobilità internazionale dei ricercatori', and by the project "ADDRESS" under the University of Catania programme "PIACERI Linea 2".

## References

- Adhinugroho, Y., Putra, A. P., Luqman, M., Ermawan, G. Y., Mariyah, S., & Pramana, S. (2020). Development of online travel Web scraping for tourism statistics in Indonesia.
- Albalade, D., & Bel, G. (2010). Tourism and urban public transport: Holding demand pressure under supply constraints. *Tourism Management*, 31(3), 425-433.
- Bordagaray, M.; dell'Olio, L.; Ibeas, A.; Cecin, P. 2014. Modelling user perception of bus transit quality considering user and service heterogeneity, *Transportmetrica A: Transport Science*, 10(8): 705-721.
- Budeanu, A. (2007). Sustainable tourist behaviour—a discussion of opportunities for change. *International Journal of Consumer Studies*, 31(5), 499-508.
- Fazio, M., Le Pira, M., Inturri, G., & Ignaccolo, M. (2020). Bus Rapid Transit vs. Metro. Monitoring on-board comfort of competing transit services via sensors. In *2020 forum on integrated and sustainable transportation systems (FISTS)* (pp. 292-297). IEEE.
- Giuffrida, N., Fajardo-Calderin, J., Masegosa, A. D., Werner, F., Steudter, M., & Pilla, F. (2022). Optimization and Machine Learning Applied to Last-Mile Logistics: A Review. *Sustainability*, 14(9), 5329.
- Hall, C. M., & Ram, Y. (2019). Measuring the relationship between tourism and walkability? Walk Score and English tourist attractions. *Journal of Sustainable Tourism*, 27(2), 223-240.
- Inturri, G., Giuffrida, N., Le Pira, M., Fazio, M., & Ignaccolo, M. (2021). Linking public transport user satisfaction with service accessibility for sustainable mobility planning. *ISPRS International Journal of Geo-Information*, 10(4), 235.
- Kanakaraj, M., & Guddeti, R. M. R. (2015, February). Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques. In *Proceedings of the 2015 IEEE 9th international conference on semantic computing (IEEE ICSC 2015)* (pp. 169-170). IEEE.
- Le Pira, M., Marcucci, E., & Gatta, V. (2021). Roman holiday: Tourist heterogeneous preferences for walking path elements. *Research in Transportation Economics*, 90, 101106.

- Le Pira, M.; Inturri, G.; Ignaccolo, M. (2018). Public transport and proactive real-time information: A survey on users' preferences and attitudes toward multimodal traveler information services. In Proceedings of the International Conference on Traffic and Transport Engineering (ICTTE 2018), Belgrade, Serbia, 27–28 September 2018; ISBN 978-86-916153-4-5.
- Page, S. (2004). Transport and tourism: Global perspectives. 2nd edition. Pearson Education.
- Pineda-Jaramillo, J. D. (2019). A review of Machine Learning (ML) algorithms used for modeling travel mode choice. *Dyna*, 86(211), 32-41.
- Pineda-Jaramillo, J., & Pineda-Jaramillo, D. (2021). Analysing travel satisfaction of tourists towards a metro system from unstructured data. *Research in Transportation Business & Management*, 100746.
- Rossetti, S., & Tiboni, M. (2020). In Field Assessment of Safety, Security, Comfort and Accessibility of Bus Stops: A Planning Perspective. *Eur. Transp./Trasporti Europei*, 80, paper 8, <https://doi.org/10.48295/ET.2020.80.8>
- Rossetti, S., Tiboni, M., Vetturi, D., Zazzi, M., & Caselli, B. (2020). Measuring pedestrian accessibility to public transport in urban areas: A GIS-based discretisation approach. *Eur. Transp. Trasp. Eur*, 76, paper 2.
- Simeon, M. I., Buonincontri, P., Cinquegrani, F., & Martone, A. (2017). Exploring tourists' cultural experiences in Naples through online reviews. *Journal of Hospitality and Tourism Technology*.
- Solangi, Y. A., Solangi, Z. A., Aarain, S., Abro, A., Mallah, G. A., & Shah, A. (2018, November). Review on natural language processing (NLP) and its toolkits for opinion mining and sentiment analysis. In *2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS)* (pp. 1-4). IEEE.
- Taecharunroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*, 75, 550-568.
- Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51–65. <https://doi.org/10.1016/j.tourman.2016.10.001>