Sustainable Tourism EXperience: a preliminary approach to restaurants recommendation systems based on sustainability

Daniel Zilio[1], Ngoc Trang Dai Vu[1] and Nicola Orio[1]

[1]*Department of Cultural Heritage, University of Padua, Piazza Capitaniato 7, Padua, 35139, Italy*

Abstract

This paper presents the initial phase of a project developing an innovative Tourism Recommender System (TRS) focused on sustainability. The proposed system, while applicable to various aspects of tourism, initially concentrates on restaurants as a case study. It utilizes three data sources: basic venue information, automated content analysis, and user-generated content. The goal is to quantify sustainability in tourism by integrating factors such as accessibility, environmental impact, and visitor perceptions. This early-stage research outlines the planned approach and addresses preliminary challenges in data collection and analysis, proposing potential solutions for automating the evaluation process. The project aims to promote more informed tourist choices and support sustainable practices across the tourism sector.

Keywords

Tourism Recommender Systems, Sustainability, User Generated Content

1. Introduction

This paper presents the initial stage of designing a recommender system for restaurants that considers sustainability factors. The methodology outlined here addresses the issue of developing a semi-automated solution to provide quantitative metrics for assessing restaurant sustainability. This approach will enable us to compare restaurants not only based on their proximity to users' needs but also by evaluating the environmental impact of each suggestion. At the base of the development of recommendation systems there is the challenge of identifying the best solutions to meet users' needs. This concept is applicable across various contexts, as evidenced by the well-known examples of Amazon and Netflix over the years [1, 2]. By employing several approaches such as collaborative filtering, content-based filtering, and hybrid systems, these platforms can suggest items that align perfectly with our current preferences [3], like recommending TV series based on our viewing behavior. Additionally, they can analyze the possibility of grouping users according to their cultural tastes [4], among many other illustrative cases. Recommender systems (RS) have greatly influenced the tourism industry (we refer to as Tourism Recommender System, TRS) by enhancing travelers' experiences. These systems aim to meet tourists' needs and preferences better. Traditionally, recommendations are generated based on user preferences gathered from historical data, such as ratings and reviews [5]. This information assists in trip planning by providing personalized suggestions for destinations, accommodations, activities, and more. The more effectively a system can achieve these goals, the more valuable it becomes. Over time, TRS must take into account various stakeholders beyond just tourists. These include host destinations and information platforms, each with its own unique needs and objectives. For example, host destinations may want to attract a large number of travelers. At the same time, information and booking platforms may focus on promoting destinations with a higher likelihood of successful transactions or better profit margins [6]. One of the main stakeholders that can no longer be overlooked is sustainability. There are various definitions of sustainable tourism. One states: "tourism that takes

full account of its current and future economic, social, and environmental impacts, addressing the needs of visitors, the industry, the environment, and host communities." [7]. The main challenge we want to face is developing a TRS that incorporates sustainability while satisfying the needs of tourists and other stakeholders. Although the reported definition embraces the essence, the primary issue is how to measure sustainability in tourism. The European Tourism Indicators System (ETIS) [8] developed by the European Commission serves as a comprehensive management tool and monitoring system designed to assist tourist destinations in measuring and enhancing their sustainability performance. It provides a set of core and supplementary indicators, along with detailed guidelines, that enable destinations to adopt a more informed approach to tourism planning. This system, which has been voluntarily implemented by over 100 destinations since 2013 through various pilot phases, offers a toolkit available in multiple languages, as well as supporting documents such as destination profiles, data sheets, and survey templates. The indicator also serves as an informational resource for policymakers, tourism enterprises, and stakeholders, complementing existing international and European methodologies. The European Commission actively promotes ETIS through conferences and awards, recognizing destinations that have successfully implemented the system. Case studies from various locations across Europe demonstrate the practical applications and benefits of ETIS in sustainable tourism management [9].

ETIS provides general guidelines for assessing tourism sustainability but does not offer specific or detailed methods for quantifying it. To create a TRS that incorporates sustainability, we need to define key elements that can be used to compare different destinations based on their sustainability practices. Additionally, we face the challenge of variability among destinations. Tourism experiences can differ significantly, whether it involves visiting a museum, relaxing on the beaches of a desert island, engaging in unique outdoor activities, or dining at a restaurant. This research focuses primarily on a specific case: restaurants. We chose this topic because it is not only an inevitable part of tourism experiences but also involves food consumption, which notably impacts the environment. This topic also requires the collection of user-generated content (UGC), and the examination of such data using text analysis, which applies to other relevant sectors under the tourism umbrella.

2. Related Work

One approach to measuring sustainability is presented in [10]. This study investigates the use of data from online platforms to assess sustainable tourism. It employs web-scraped data from Tripadvisor and machine learning techniques to predict which accommodations follow sustainable practices. The findings indicate that machine learning models can effectively identify sustainable accommodations based on publicly available online information. This approach provides a cost-effective and scalable method for monitoring tourism sustainability, offering high spatial and temporal granularity. This systematic review presented in [11] synthesizes research on environmental sustainability in restaurants. It identifies key stakeholders, sources of unsustainability, green initiatives, outcomes, and performance indicators. The study emphasizes the need for standardized sustainability metrics and comprehensive approaches to implementing green practices across various types of restaurants. Additionally, it highlights research gaps in areas such as the role of technology and the long-term impacts of sustainability efforts. The study [12] presents a flexible multi-criteria decision analysis method that utilizes unweighted TOPSIS to evaluate and rank alternatives based on their sustainability, without the need for precise weights for the criteria. This approach is applied to sustainable tourism in Spain, considering both client and public management perspectives. It identifies strengths and weaknesses to help guide efforts toward achieving the UN Sustainable Development Goals. A novel tourist recommender system (TRS) is designed in [13]. It utilizes deep reinforcement learning to suggest sustainable itineraries aimed at preventing overcrowding while optimizing visitor experiences. The proposed approach takes into account spatiotemporal factors, weather conditions, and predicted crowd levels to generate comprehensive tour sequences. It reported improved performance compared to traditional methods by effectively reducing wait times, increasing visit durations, and promoting diversity in recommendations. In [14] sentiment analysis is utilized on Chinese social media data to assess the quality of tourism in Spain, effectively measuring word-of-mouth

and identifying areas for improvement in tourist destinations. The research emphasizes that AI-powered sentiment analysis of user-generated content offers more nuanced and authentic insights compared to traditional survey methods. This approach enables better destination management and promotes sustainability within the tourism industry. The Tourism Sustainability Index (TSI) described in [15] measures tourism sustainability by combining open data with sentiment analysis of user-generated content. The Green Destination Recommender (GDR) [16] is a web application aimed at promoting sustainable tourism by suggesting environmentally friendly travel destinations. It combines various sustainability factors, such as transport emissions, destination popularity, and seasonal demand, into a single metric. This helps users make more responsible travel choices, addressing the increasing demand for eco-conscious tourism solutions in a market that is becoming more environmentally aware. In [17] the concept of sustainability-aware persuasive explanations in recommender systems is presented, applying Cialdini's persuasive principles to promote more sustainable choices across three product domains: books, healthy food, and cars. Through a user study with 158 participants, the research reveals that explanations based on the "authority" principle were generally most effective, while the importance of sustainability aspects varied across domains, with higher perceived relevance in food and car recommendations compared to books. Finally [18] proposes a novel approach to restaurant recommendation systems by incorporating user and venue personality features derived from eWOM, alongside topic modeling. Results demonstrate that personality-based models, particularly those using MBTI, combined with XGBoost regression, outperform traditional collaborative filtering methods in predicting user restaurant ratings.

3. Measuring the sustainability

This project aims to design a recommendation system for tourism applications that takes sustainability into account. The case study focuses on restaurants, with the primary goal of developing a system to evaluate and measure a restaurant's sustainability. This first part of the research focuses on creating a dataset of real restaurants, beginning with a chosen city, to establish a starting point for the subsequent design phases. In the initial stage of the process, we need to determine how to evaluate restaurants based on sustainability. Our assessment will focus on three main factors:

- Basic information about the restaurants
- The proposed menu
- Online User-generated content

These factors are initially gathered from the widely recognized tourism review platforms, Tripadvisor and Google Maps (see Figure 1). After data retrieval and preprocessing stages, the next task will be to compare the collected results and place each restaurant on a sustainability scale. In this paper, initial considerations of the data retrieval process are presented.

3.1. Basic information about the restaurant

The information gathered at this stage is crucial for analyzing the fundamental elements related to sustainability. Key considerations include the accessibility of the location via public transport, which helps avoid the use of personal vehicles, as well as the venue's hours of operation and delivery options. The primary source for obtaining this information is Google Maps. This platform also offers real-time data on the venue's crowdedness, which can significantly influence the final recommendation. Overtourism is a well-known challenge in the sustainability field, making this information particularly relevant.

Additionally, some preliminary details about the menu can be sourced in this way, such as the availability of vegetarian or vegan options. The restaurant's own description may highlight the use of local and seasonal products and the availability of regional dishes.

WORKFLOW

This is the workflow of the STEX project. It is comprised of three stages, (i) Data Retrieval, (ii) Data Preprocessing, and (iii) Data Analysis
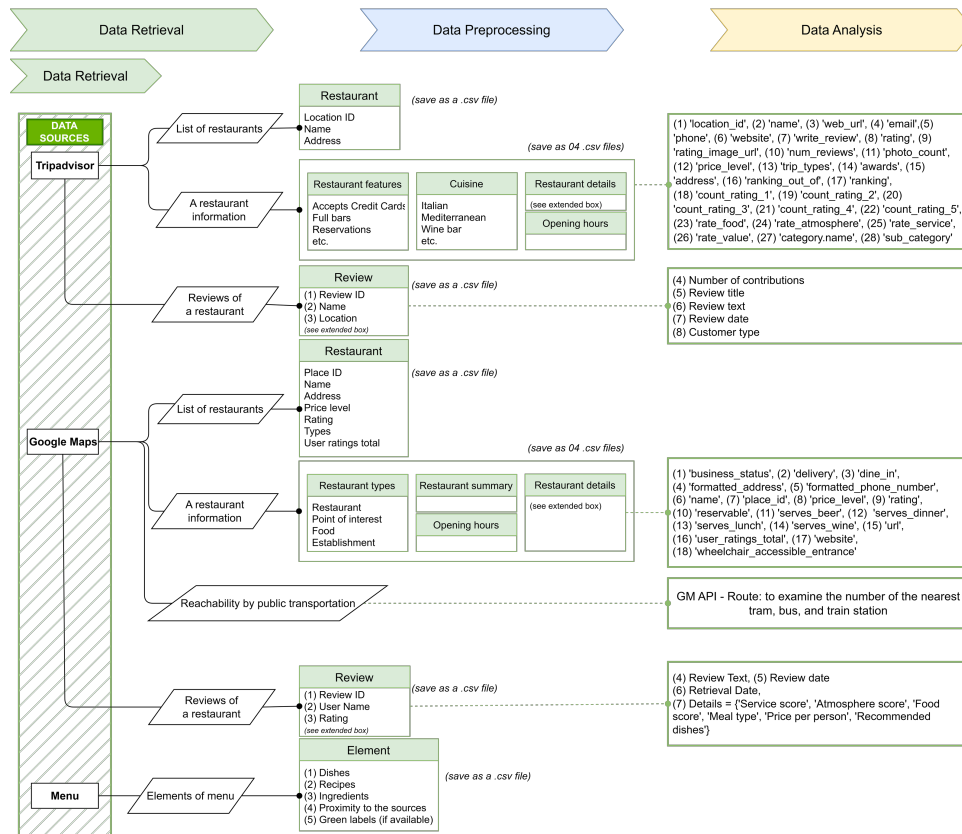
Figure 1: Workflow of STEX project - Data Retrieval stage

3.2. Automatic menu analysis

The menu offered by restaurants will be an important element for comparison. Our goal is to establish an automatic system that utilizes text analysis to examine the different available recipes and extract the various ingredients for each one. Analyzing the individual ingredients will play a crucial role in assessing their environmental impact. For instance, animal protein is considered to have a greater environmental impact than vegetable protein [19, 20]. The presence of vegan or vegetarian dishes will generally be viewed positively compared to meat-based options. In addition, we will evaluate the restaurant's use of local and seasonal foods due to their impacts on environmental, societal, and economic sustainability [21]. Ultimately, the objective is to generate a sustainability score for each element of the menu, allowing us to derive a quantitative comparative value.

This phase presents two main challenges. The first is related to the level of automation that can be achieved. The goal is to automate all process steps, but retrieving digital menus is not always straightforward. Ideally, the restaurant would provide the menu on its website, allowing a parser to easily extract the required information. However, more often than not, the menu is available only as a PDF file, and its varied formats and layouts can complicate the automatic extraction of recipes. In the initial phase of the project, we will focus on selected cases where the menu is available online on a webpage or, at the very least, in clearly editable PDF formats. Additionally, we will explore the use of language model-based vision techniques to automate the extraction of menu text, or like in [22] directly for the dishes pictures.

The second challenge involves the assumption that certain ingredients have a greater impact on the environment and overall sustainability than others. Although there are studies on this topic, particularly on the differences between plant-based and meat-based meals, a thorough analysis of the literature is

necessary to develop a more comprehensive understanding. This will help to design a metric that is as accurate and valid as possible.

3.3. Online User-generated content

The third element to be included in our analysis is the emergence of information on sustainability trends in online user reviews. We aim to investigate related topics such as considerations regarding product quality, the restaurant's policy on allowing customers to take home leftover food, and portion sizes of dishes to assess potential food wastage. Among these topics, reducing the portion sizes, as well as plate sizes, has been suggested for buffet restaurants to lower tourism's carbon 'footprint' [23]. Additionally, we will examine information related to wait times and crowdedness in the restaurant. The volume of reviews received is also an important factor; a restaurant with a high number of reviews in a short period suggests that it attracts a substantial number of visitors, which is likely indicative of over-tourism. Meanwhile, one with fewer reviews could suffer in a recommendation system. For this part of the analysis, going beyond investigating the descriptive data related to waiting times and the number of reviews in a certain period, we will employ techniques of word frequency and topic analysis. The first is to figure out the most frequent words that reviewers leave for restaurants, which can notify the appearance of sustainability trends. The latter will capture the main themes of the reviews while highlighting the various types of information present. Given the possible low number of user reviews on both Tripadvisor and Google Maps and the short and unstructured nature of these reviews, the study will manually compile the review set and apply BERTopic [24]. Among various topic modeling techniques, including Latent Dirichlet Allocation (LDA), non-negative matrix factorization (NMF) and Top2Vec, BERTopic is more potential when extracting useful information from short text data [25].
The first phase of data collection for the project is in progress and poses some challenges as we retrieve user reviews on Google Maps and Tripadvisor. Despite their ability to offer general information about restaurants, such as restaurant features, cuisine, opening hours, restaurant types, rating scores, and so on (see *the extended box* of restaurant information in Figure 1), Tripadvisor and Google Maps only allow users to retrieve up to five reviews using their Application Programming Interface (API) on a free basis.[12] The limited number of reviews does not help to produce an adequate and meaningful topic extraction and analysis. While Tripadvisor does not provide any further information for retrieving more reviews, Google Maps allows business owners to gather all reviews of their businesses through the Business Profile APIs.[3] One way for researchers affiliated with EU-based organizations to retrieve more reviews from Google Maps is through the Google Researcher Program.[4] However, Google Maps does not specify the duration of the evaluation for a research application process. Considering that this process can take a long time and the low number of reviews retrieved from Tripadvisor, we decide to gather user reviews from these two platforms using a Python scripted program.

Also, dynamic websites that generate pages with constantly changing web elements, along with their web security techniques, challenge the data extraction process. Due to such particular technical reasons, the data collection consumes more efforts and times, and some data is missing from this process; for instance, rating scores of Tripadvisor reviewers are not scraped, or it takes several tries to retrieve all reviews on both platforms.

4. Future steps

This paper presented the initial stage of the project development, which aims to design a sustainability-based Tourism Recommender System. The primary focus will be on creating and refining the dataset, with restaurants serving as the main case study. This phase aims to address the identified technical and practical challenges related to data collection and analysis methodologies. Once the dataset is

---

[1]https://tripadvisor-content-api.readme.io/reference/overview
[2]https://developers.google.com/maps/documentation/places/web-service/details
[3]https://developers.google.com/my-business/content/review-data/#list_all_reviews
[4]https://requestrecords.google.com/researcher

successfully consolidated, the research will advance to the implementation and evaluation of various recommender system models. Subsequent research will concentrate on a thorough analysis of end-user perceptions regarding sustainability in tourism recommendations. This study will investigate how sustainability-based recommendations influence tourist decision-making processes, satisfaction levels, and long-term behavior patterns. The findings will be crucial for assessing the effectiveness of the system in promoting sustainable tourism practices.

Acknowledgments

References

[1] B. Smith, G. Linden, Two decades of recommender systems at amazon.com, IEEE Internet Comput. 21 (2017) 12–18. URL: https://doi.org/10.1109/MIC.2017.72. doi:10.1109/MIC.2017.72.

[2] X. Amatriain, J. Basilico, Recommender systems in industry: A netflix case study, in: F. Ricci, L. Rokach, B. Shapira (Eds.), Recommender Systems Handbook, Springer, 2015, pp. 385–419. URL: https://doi.org/10.1007/978-1-4899-7637-6_11. doi:10.1007/978-1-4899-7637-6_11.

[3] F. Ricci, L. Rokach, B. Shapira (Eds.), Recommender Systems Handbook, Springer US, 2022. URL: https://doi.org/10.1007/978-1-0716-2197-4. doi:10.1007/978-1-0716-2197-4.

[4] D. Zilio, N. Orio, C. Toniolo, Tindart, an experiment on user profiling for museum applications, in: M. Ceci, S. Ferilli, A. Poggi (Eds.), Digital Libraries: The Era of Big Data and Data Science - 16th Italian Research Conference on Digital Libraries, IRCDL 2020, Bari, Italy, January 30-31, 2020, Proceedings, volume 1177 of *Communications in Computer and Information Science*, Springer, 2020, pp. 123–134. URL: https://doi.org/10.1007/978-3-030-39905-4_13. doi:10.1007/978-3-030-39905-4_13.

[5] P. Banik, A. Banerjee, W. Wörndl, Understanding user perspectives on sustainability and fairness in tourism recommender systems, in: Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2023, Limassol, Cyprus, June 26-29, 2023, ACM, 2023, pp. 241–248. URL: https://doi.org/10.1145/3563359.3597442. doi:10.1145/3563359.3597442.

[6] H. Abdollahpouri, G. Adomavicius, R. Burke, I. Guy, D. Jannach, T. Kamishima, J. Krasnodebski, L. A. Pizzato, Multistakeholder recommendation: Survey and research directions, User Model. User Adapt. Interact. 30 (2020) 127–158. URL: https://doi.org/10.1007/s11257-019-09256-1. doi:10.1007/S11257-019-09256-1.

[7] S. Gössling, Tourism, information technologies and sustainability: an exploratory review, Journal of Sustainable Tourism 25 (2017) 1024–1041. doi:10.1080/09669582.2015.1122017.

[8] C. europea, d. e. d. P. Direzione generale del Mercato interno, dell'industria, The European Tourism Indicator System : ETIS toolkit for sustainable destination management, Ufficio delle pubblicazioni, 2016. doi:doi/10.2873/983087.

[9] J. C. Carcia-Rosell, P. Hanni-Vaara, P. Iivari, E. Linna, P. Satokangas, M. Tapaninen, T. Tekoniemi-Selk"al"a, Tourism Quality and Sustainability Programmes, Labels and Criteria in the Barents Region, Technical Report, Multidimensional Tourism Institute, 2017. URL: https://core.ac.uk/download/198192458.pdf, accessed on [Insert access date].

[10] F. J. Hoffmann, F. Braesemann, T. Teubner, Measuring sustainable tourism with online platform data, EPJ Data Sci. 11 (2022) 41. URL: https://doi.org/10.1140/epjds/s13688-022-00354-6. doi:10.1140/EPJDS/S13688-022-00354-6.

[11] A. D. Arun Madanaguli, P. Kaur, S. Srivastava, G. Singh, Environmental sustainability in restaurants. a systematic review and future research agenda on restaurant adoption of green practices,

Scandinavian Journal of Hospitality and Tourism 22 (2022) 303–330. doi:`10.1080/15022250.2022.2134203`.

[12] J. Vicens-Colom, J. Holles, V. Liern, Measuring sustainability with unweighted topsis: An application to sustainable tourism in spain, Sustainability 13 (2021). URL: https://www.mdpi.com/2071-1050/13/9/5283.

[13] A. D. Vecchia, S. Migliorini, E. Quintarelli, M. Gambini, A. Belussi, Promoting sustainable tourism by recommending sequences of attractions with deep reinforcement learning, Information Technology & Tourism 26 (2024) 449–484. URL: https://api.semanticscholar.org/CorpusID:269382951.

[14] F. Borrajo-Millán, M.-d.-M. Alonso-Almeida, M. Escat-Cortes, L. Yi, Sentiment analysis to measure quality and build sustainability in tourism destinations, Sustainability 13 (2021). URL: https://www.mdpi.com/2071-1050/13/11/6015. doi:`10.3390/su13116015`.

[15] D. De Marchi, R. Becarelli, L. Di Sarli, Tourism sustainability index: Measuring tourism sustainability based on the etis toolkit, by exploring tourist satisfaction via sentiment analysis, Sustainability 14 (2022). URL: https://www.mdpi.com/2071-1050/14/13/8049. doi:`10.3390/su14138049`.

[16] A. Banerjee, T. Mahmudov, W. Wörndl, Green destination recommender: A web application to encourage responsible city trip recommendations, UMAP Adjunct '24, Association for Computing Machinery, New York, NY, USA, 2024, p. 486–490. URL: https://doi.org/10.1145/3631700.3664909. doi:`10.1145/3631700.3664909`.

[17] T. N. T. Tran, S. Polat Erdeniz, A. Felfernig, S. Lubos, M. El Mansi, V.-M. Le, Less is more: Towards sustainability-aware persuasive explanations in recommender systems, RecSys '24, Association for Computing Machinery, New York, NY, USA, 2024, p. 1108–1112. URL: https://doi.org/10.1145/3640457.3691708. doi:`10.1145/3640457.3691708`.

[18] E. Christodoulou, A. Gregoriades, H. Herodotou, M. Pampaka, Combination of user and venue personality with topic modelling in restaurant recommender systems, in: J. Neidhardt, W. Wörndl, T. Kuflik, D. Goldenberg, M. Zanker (Eds.), Proceedings of the Workshop on Recommenders in Tourism (RecTour 2022) co-located with the 16th ACM Conference on Recommender Systems (RecSys 2022), Seattle, WA, USA and Online, September 22, 2022, volume 3219 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 21–36. URL: https://ceur-ws.org/Vol-3219/paper2.pdf.

[19] L. Ferrari, S.-A. Panaite, A. Bertazzo, F. Visioli, Animal- and plant-based protein sources: A scoping review of human health outcomes and environmental impact, Nutrients 14 (2021). doi:`https://doi.org/10.3390/nu14235115`.

[20] J. Sabaté, S. Soret, Sustainability of plant-based diets: back to the future, The American Journal of Clinical Nutrition 100 (2014) 476S–482S. URL: https://www.sciencedirect.com/science/article/pii/S0002916523048992. doi:`https://doi.org/10.3945/ajcn.113.071522`.

[21] A. M. Vargas, A. P. de Moura, R. Deliza, L. M. Cunha, The role of local seasonal foods in enhancing sustainable food consumption: A systematic literature review, Foods 10 (2021). doi:`https://doi.org/10.3390/foods10092206`.

[22] J.-H. Kim, N.-H. Kim, D. Jo, C. S. Won, Multimodal food image classification with large language models, Electronics 13 (2024). URL: https://www.mdpi.com/2079-9292/13/22/4552. doi:`10.3390/electronics13224552`.

[23] S. Gössling, B. Garrod, C. Aall, J. Hille, P. Peeters, Food management in tourism: Reducing tourism's carbon 'foodprint', Tourism Management 32 (2011) 534–543. URL: https://www.sciencedirect.com/science/article/pii/S0261517710000701. doi:`https://doi.org/10.1016/j.tourman.2010.04.006`.

[24] M. Grootendorst, Bertopic: Neural topic modeling with a class-based tf-idf procedure, arXiv preprint arXiv:2203.05794 (2022).

[25] R. Egger, J. Yu, A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts, Frontiers in Sociology 7 (2022). URL: https://api.semanticscholar.org/CorpusID:248530058.