DOI: 10.30519/ahtr.1436175

Advances in Hospitality and Tourism Research (AHTR) 2024
An International Journal of Akdeniz University Tourism Faculty Vol. 12 (3)
ISSN: 2147-9100 (Print), 2148-7316 (Online) 254-286
Webpage: https://dergipark.org.tr/en/pub/ahtr

EVALUATING TOURIST DISSATISFACTION WITH ASPECT-BASED SENTIMENT ANALYSIS USING SOCIAL MEDIA DATA

Marlon Santiago VIÑÁN-LUDEÑA [1]

Escuela de Ingeniería, Universidad Católica del Norte, Coquimbo, Chile & Universidad de Granada, Spain

ORCID: 0000-0003-2692-5899

Luis de CAMPOS

Departamento de Ciencias de la Computación e Inteligencia Artificial, ETSI Informática y de Telecomunicación, CITIC-UGR, Universidad de Granada, Spain

ORCID: 0000-0001-9125-1195

ABSTRACT

Tourism satisfaction is essential for encouraging tourists to stay longer, spend more and return. However, visitor dissatisfaction can also prove useful for understanding any shortcomings of a tourist destination, and Twitter, Instagram and TripAdvisor reviews might be able to provide an insight into tourist perceptions and experiences. This study examines the major causes of tourist dissatisfaction with a tourism destination using an aspect-based sentiment analysis approach to understand the key points of negative tweets, posts or reviews. We examined 19,340 tweets, 7,712 Instagram posts and 25,483 reviews about Granada in Spain in order to evaluate the negative user's perceptions, discover management-related problems and provide feedback to destination management organizations to enable them to improve their services and operations. Our work contributes to computational methods to address tourism (dis)satisfaction with a process to identify the most important entities (places), an algorithm to identify aspects and opinions, and the use of word-trees to show the most important aspect-opinion tuples. In practical terms, we provide to tourism industry professionals and managers, as well as travelers, with methods to identify the reasons for tourist dissatisfaction from available social media data, in such a way that managerial strategies or travel plans can be improved.

Article History
Received 13 February 2024
Revised 4 June 2024
Accepted 12 June 2024
Published online 8 July 2024

Keywords
tourist dissatisfaction
aspect-based sentiment analysis
social media
tourism
BERT

[1] Address correspondence to Marlon Santiago Viñán-Ludeña, Escuela de Ingeniería, Universidad Católica del Norte, Coquimbo, Chile & Departamento de Ciencias de la Computación e Inteligencia Artificial, ETSI Informática y de Telecomunicación, CITIC-UGR, Universidad de Granada, Granada, Spain. E-mail: marsantovi@gmail.com

INTRODUCTION

Computational methods, algorithms and techniques are currently used as the baseline for understanding the customer experience and are crucial in marketing, politics and tourism, particularly because every internet-based activity leaves a digital footprint (Alaei et al., 2019). Some researchers have proposed data analysis frameworks to understand the huge amount of data, including meta data and textual data on social media, which has been created by tourists who have traveled to certain tourist destinations (Viñán-Ludeña & de Campos, 2022a).

Satisfaction is of critical importance in the tourism domain since it enables managers not only to obtain insights into the customer experience and customer satisfaction (Oh et al., 2022) but also feedback. According to Kim and Kim (2022), the main objective of corporate marketing activities is to focus on customer satisfaction management since this can increase customer loyalty, repurchase intention and can contribute to higher profitability.

In the past, tourism researchers relied on surveys to discover customer satisfaction but nowadays, we can use social media data from different platforms (i.e. TripAdvisor, Yelp, Google Travel[2], Twitter, Instagram, Facebook, etc.) to collect data from users' posts or tweets. It is fairly simple to collect data from these platforms thanks to the use of APIs[3], crawlers[4] or scrapers[5] (Viñán-Ludeña, 2019). The results of social media data analysis provide researchers and professionals with an important advantage and a cheaper way to discover the level of satisfaction or dissatisfaction with the places and services available at a tourist destination.

Many computational methods and techniques (i.e. machine learning, deep learning, soft computing, etc.) have to date been used by researchers to understand consumers' experiences (Viñán-Ludeña et al., 2022) and satisfaction. Online reviews, posts, tweets, comments and forums are important sources of data, and these can provide tourism professionals and managers with useful information about these topics. One of the most promising techniques that has been used in social media analysis to discover user satisfaction is sentiment analysis (SA) and this is based on natural language processing and other computational techniques including

[2] www.google.com/travel
[3] An API or application programming interface.
[4] Software downloads web pages.
[5] Software takes the downloaded web pages and extracts data from these pages.

machine learning techniques, computational linguistics, etc. to gather opinions about an entity.

Sentiment analysis can be used at three different levels: firstly, document-level sentiment analysis aims to classify the entire document as expressing a positive, negative or neutral opinion. Secondly, sentence-level sentiment analysis whereby a text is split into sentences to identify the emotions (i.e. positive, negative, or neutral) about each one. And finally, aspect-based sentiment analysis (ABSA), which is the most challenging task in sentiment analysis and provides us with more precise information about an opinion. In ABSA, an opinion is a quintuple comprising the sentiment (or opinion) target or entity; entity attributes or aspects; the positive, negative, or neutral orientation about entity aspects; the opinion holder; and the opinion posting time. ABSA therefore performs a detailed analysis of customer/user/traveler feedback data so that managers and service providers can learn more about their customers in order to improve their services and meet users' needs.

Visitor feedback from structured data (Viñán-Ludeña et al., 2022) and unstructured information such as reviews, tweets or posts about a tourist destination undoubtedly plays a crucial role in tourism management. However, understanding and considering not only travelers' opinions but also their complaints or negative opinions can enable tourism managers to understand, identify and reflect on how to deal with any weaknesses in order to improve their services and operations. Our main aim in this article, therefore, is to provide an approach through aspect-based sentiment analysis (ABSA) to analyze visitors' dissatisfaction with a tourism destination using Instagram posts, tweets and TripAdvisor reviews. Our study involves efficient techniques that must be considered in ABSA and these include aspect extraction, aspect clustering, opinion extraction, and summarization. Accordingly, our work makes the following research contributions: firstly, since in previous ABSA studies, the authors knew the entities in advance (e.g. hotels, laptops, cameras, etc.), we therefore propose an approach to identify entities using a semi-supervised process which includes the use of the Bidirectional Encoder Representations from Transformers (BERT) tool and to classify related terms (a tourist destination resident fine-tunes and improves previously identified entities). Secondly, we propose a rule-based approach to aspect identification (e.g. price is an attribute, or aspect, of any hotel, or entity) and the opinion assignation of the entity-aspect (positive/negative). Thirdly, we perform aspect clustering by means of word embeddings as a novel approach. Fourthly, we propose a summarization approach through visualization in order to understand customer satisfaction and experiences for a tourism destination and the services offered and this provides tourism managers and practitioners with useful information. Finally, we conduct experiments with this proposed approach using data from Twitter, TripAdvisor and Instagram posts about Granada in Spain.

This approach contributes to the existing literature because the processes of rule based aspect identification, opinion assignment and aspect clustering by means of word embeddings have not been performed before in the context of tourism (dis)satisfaction.

The remainder of this paper is organized in the following way: Background presents a theoretical overview of the different approaches in every ABSA stage. Literature review explores aspect-based sentiment analysis in tourism. Methodology examines data collection, data processing, entity extraction, aspect and opinion identification, aspect clustering and summarization techniques. Then, discusses our results. Discussion explores the relationship of the findings to the existing literature and theories on tourist (dis)satisfaction. Finally, conclusions and implications outlines the practical and research implications and presents our future lines of research.

BACKGROUND

In order to properly understand the approach proposed in this work, it is necessary to introduce some theoretical concepts about aspect-based sentiment analysis. We will proceed to summarize basic concepts about sentiment analysis and aspect-based sentiment analysis, and we will briefly review the sub-tasks that we include to perform ABSA in order to understand tourist satisfaction with a tourism destination.

Sentiment Analysis

Sentiment analysis (SA), or opinion mining, consists of identifying attitudes, moods and emotions towards entities (i.e. products, services offered by companies, events, topics, places and their attributes, or aspects). This is a very important field for understanding the social psychology of how a group or an individual (also called "influencers" on social media) might modify their beliefs, choices and perceptions of the world (Liu, 2015; Viñán-Ludeña et al., 2020).

Aspect-based Sentiment Analysis

ABSA is a sentiment analysis sub-task. In ABSA, an opinion is defined as a quintuple (*e,a,s,h,t*), whereby *e* is an entity, place or service (e.g. a monument, a neighborhood, a restaurant, etc.), *a* corresponds to one of its attributes or aspects (e.g. price, cleanliness, etc.), *s* is the sentiment about the entity or aspect (positive/negative/neutral), *h* is the opinion holder (on social media, it is easy to see which individual or user wrote a particular review or post), and *t* is the time when the opinion is given since social media platforms store and show the date of each post or comment (Liu, 2015).

We identified the following sub-tasks to perform ABSA and we also provide a brief theoretical introduction for each of these:

- *Aspect-term extraction*: this task identifies positive or negative terms or aspects in texts (e.g. In the tweet "The Alhambra is one of the most beautiful monuments in the world", the entity is the "Alhambra", the aspect or characteristic of the Alhambra is "monument", and the opinion might be recognized through a sentiment word or phrase, for example, "beautiful" or "the most beautiful").

- *Aspect categorization or clustering*: This classification task clusters the words relating to the same aspect, since one aspect might be associated with different words.

- *Summarization*: The purpose of this task is to provide or generate a structured summary from all the resources found in previous tasks and in Section 2.5, we analyze a number of ABSA summarization approaches.

Aspect-term Extraction

Luo et al. (2019), use different approaches to identify aspects:

- *Rule-based methods:* These are commonly written by hand to extract aspects from text. Luo et al. (2019) employed six rules to extract the aspects. They also built an aspect graph to narrow the aspect space, performed clustering and finally identified the most prominent aspects.

- *Topic-modeling-based methods:* These methods extract topics from text and aspects from topics. A supervised approach is presented in Wang et al. (2014), where the authors first use seeding aspects obtained from the product descriptions; secondly, product reviews are classified according to these seeding aspects; and finally, they propose the fine-grained LDA and unified fine-grained labeled LDA to discover aspects relating to seeding aspects. He et al. (2021) propose a Hierarchical Features-based Topic Model (HFTM) to extract aspects from online reviews and then capture specific features. Concept-LDA is presented by Ekinci and Ilhan Omurca (2020), where LDA is used to extract latent aspects by building a feature space before it is enriched with concepts and entities extracted from Babelfy[6].

- *Neural network-based methods:* These methods apply deep learning architectures. A multi-domain aspect extraction using BERT combining 15 datasets from different domains to train the model is proposed by Santos et al. (2021). The results showed a competitive alternative compared to single-domain models. Poria et al. (2016) use a deep convolutional neural network and a series of linguistic patterns to perform aspect detection. A convolutional neural network model with dynamic filters to extract the aspects in a document is presented by Zhang et al. (2021), where aspects are categorized with a neural topic model. One approach which applies linguistic patterns (single word and multi-word aspects) to label aspects and builds a dataset which is used to train the deep learning model was presented by Chauhan et al. (2020).

Aspect Clustering

Before the aspects are clustered, it is important to convert each extracted aspect into a vector representation so that the lexical relationships between aspects can be understood. Word embeddings are vector representations of the words which take into account the surrounding words. These vectors can be generated with methods such as neural networks, co-occurrence matrix probabilistic models, etc. The most common tools which are used in this task are Word2vec (Mikolov et al., 2013) and Glove (Pennington et al., 2014).

Once the vectors have been represented, it is necessary to group or cluster similar aspects. According to Ansar et al. (2021), aspects can be grouped according to their similar scores (e.g. cosine similarity). The authors applied two clustering algorithms in combination (i.e. the single-linkage clustering algorithm and the group-average-linkage clustering algorithm).

An association rule-based approach to aspect cluster detection is presented by Kumar et al. (2020), where the authors find aspect category representative words using the statistical association between review

[6] http://babelfy.org/about

words and aspect category through class-based association rules. Word embeddings are then trained on a specific domain dataset and the word embeddings are used to find the semantic association between the review words and aspect categories. Finally, class-based association rules are generated.

Summarization

According to Hu and Liu (2004), summarization has two important characteristics: firstly, it identifies the opinion targets (i.e. aspects) and their sentiments, and secondly, it is necessary to quantify how many positive and negative opinions there are about the opinion targets. Using these characteristics, this task can be presented in a bar chart, where each bar above the X-axis shows the number of positive opinions and those below the X-axis correspond to negative opinions about each aspect. Other approaches, meanwhile, apply text summaries.

Generally speaking, the summaries do not order the aspects and cannot show the most important aspect about an entity and how aspects are related to each other. These limitations are addressed by Carenini et al. (2013), who proposed an "extractive" and "abstractive" summarizer. Meanwhile, Di Fabbrizio et al. (2014) propose a hybrid method which combines natural language generation and salient sentence selection techniques. These two approaches take advantage of natural language generation (NLG) to generate new sentences from the data extracted from their corpus to generate more coherent summaries. In order to identify how the aspects and opinions are related to each other, Carenini et al. (2013) propose the application of a user-defined feature taxonomy for the aspects and a large amount of training data can also be used to this end (Di Fabbrizio et al., 2014). Finally, a framework that generates an aspect-based abstract from sentences/reviews of an entity without a feature taxonomy or training data is presented by Gerani et al. (2019), where the authors take a set of reviews about the entity (target) as the input, identify the aspects, their polarity and the strength of opinions about each aspect in each sentence, before generating the summaries with natural language generation tools using the relevance degree of the aspects in addition to the association between them.

In the following section, we will explore different studies connected with ABSA and the tourism domain.

LITERATURE REVIEW

One of the aims of ABSA is to offer fine-grained information from texts (i.e. posts, tweets, reviews, blogs, etc.) about entities (places, services, etc.), their attributes or aspects (price, cleanliness, etc.) and opinions about them (positive/neutral/negative). Researchers have proposed different ABSA approaches for the tourism domain. Moreno-Ortiz et al. (2019) validate an annotation schema for aspect-based sentiment analysis using reviews about accommodation, catering and car rental. However, they only focus on the corpus building process, which is an ABSA subtask. Afzaal et al. (2019) present a tourism mobile application whereby the authors apply a tree-based aspect extraction method and machine learning algorithms to identify aspects and perform a classification task. This application provides useful information and enables visitors to make better decisions on their journey.

A prospective design is presented in Maity et al. (2020), where a lexicon is used to identify features (aspects) from travel reviews about hotels or resorts. Stepaniuk and Sturgulewska (2021) created a methodology to analyze and visualize the emotional responses of social media users from a closed Facebook group. ABSA was used to semantically decompose 300 selected photos and the results showed the comprehension and visualization of photos (memes) as well as the emotional responses of the visual content recipient.

A methodology based on negative TripAdvisor reviews is presented by Valdivia et al. (2020), and this applies the deep learning approach presented by Poria et al. (2016). The authors use the k-means algorithm for aspect clustering and the summarization process using subgroup discovery by means of the use of description rules provides information about negative aspect reviews.

Survival prediction is crucial in tourism industry, Li et al. (2023) proposed a study to predict restaurant survival based on online reviews through ABSA using BERT, thus; authors defined aspects such as: “tastiness”, “service”, “location”, “price” and “atmosphere” to train the model.

A weighted ABSA using extended ordered weighted average operators and Word2Vec is presented by Ghosal and Jain (2023), their model considers explicit and implicit aspect segmentation for review files, incorporates the meaning of slang words and location based geospatial analysis. Another study performed emotion analysis and ABSA to study behavioral intentions of tourists (Mehra, 2023). Language interpretation in travel guidance platforms using BERT-based model is presented in Chu et al. (2022) to perform category recognition sentence-level sentiment classification and sentiment analysis. The popularity of BERT-based models has gained a lot of interest, however, the major limitation of BERT models in performing ABSA is that they require a training dataset with predefined aspects. Therefore, we propose an approach to identify automatically the most important aspects in a tourism destination.

One interesting application of ABSA is to identify service failures in the hotel sector, hotel guest satisfaction and user experiences (Sann & Lai, 2020): service failure items (aspects) were identified and grouped according to the hotel guest cycle and their corresponding operations. They also compared the expression patterns used by Asians and non-Asians in order to understand the homophily of service failure as well as their hotel experiences.

ABSA can also be used to evaluate the reputation of a tourist destination and so Ali et al. (2021) employed a technique to combine topic modeling (LDA) and lexicon-based algorithms to gather information about the reputation of a tourism destination using TripAdvisor reviews about different places and sights in the city of Marrakesh.

Finally, sentiment analysis allows a wide range of applications in the tourism industry. Polyzos et al. (2024) examine the characteristics that drive conflicting outcomes on the impact of Twitter data on tourism firm returns using financial micro data; they mine Twitter through an API and calculate sentiment analysis through a software package and emotion intensity using a lexicon approach. Gastronomy is an important part of the tourism industry and nowadays, due to climate change, green practices have been widely adopted. In Shahhosseini and Nasr (2024), the authors explore the determinants of satisfaction in green restaurants, analyzing TripAdvisor reviews; their work unveils important attributes such as value and service. In the same vein, Rauf and Pasha (2024) use textual analysis, content analysis, and sentiment analysis to examine how the Global North-South divide manifests in vlogged gastronomic tours and what responses such phenomena provoke among international audiences. Vegan and vegetarian tourists are being considered important in gastronomy. Zeng et al. (2024) examine comments about a vegetarian documentary in China; their results show that a vegetarian documentary on Chinese social media sparked resistance despite a solid cultural foundation, and flexitarian approaches may better resonate for vegetarian promotion. While other researchers applied classical deep learning techniques to understand long-stay tourist experiences and satisfaction (Kim et al., 2024). To enhance tourism satisfaction, Calderón-Fajardo et al. (2024) constructed a thesaurus enabling the measurement of sensory, affective, intellectual, and behavioral dimensions in unique and emblematic attractions, experiences, and transportation within a tourist destination, based on visitor reviews.

Researchers have published various studies covering or evaluating tourist dissatisfaction and these are summarized in Table 1.

Table 1. *Summary of studies about tourist dissatisfaction*

Author	**Approach/ Methodology**	**Dataset**	**Tourist Destination**	**Context**
(Prakash et al., 2019)	Qualitative analysis	TripAdvisor	Sri-Lanka	Wildlife tourism
(Hu et al., 2019)	Structural Topic Model	TripAdvisor	New York city	Hospitality
(Fernandes & Fernandes, 2018)	Content Analysis	TripAdvisor	Porto (Portugal)	Hospitality
(Mate et al., 2019)	Content Analysis	TripAdvisor	Cook Islands	Hospitality
(Taheri et al., 2020)	Partial Least Squares	Surveys	Iran	Airports
(Um & Kim, 2018)	Qualitative Analysis	Surveys	Korea	Medical Tourism
(Lam-González et al., 2021)	Covariance-based structural equations model	Surveys	Habana, Cuba	Cultural Tourism
(Rodrigues et al., 2020)	Qualitative Analysis	Booking.com	Thermal springs in Portugal	Spas

This study differs from the others when evaluating dissatisfaction in a number of ways: firstly, it uses data from general social networks such as Twitter, Instagram and TripAdvisor. Secondly, it provides an approach on how to use ABSA to evaluate dissatisfaction. Thirdly, a semi-supervised method is presented for entity detection by means of a BERT-based tool. Fourthly, we propose a visualization process to examine dissatisfaction factors. Finally, the results are summarized using BERT-based methods.

METHODOLOGY

In order to evaluate tourist dissatisfaction, we proposed an approach which applies ABSA as a sub-task of sentiment analysis, analyzing negative opinions about places, services, events, etc. in a tourism destination. In

Figure 1, we summarize the approach used. Each stage of the proposed approach is expanded in the following subsections.

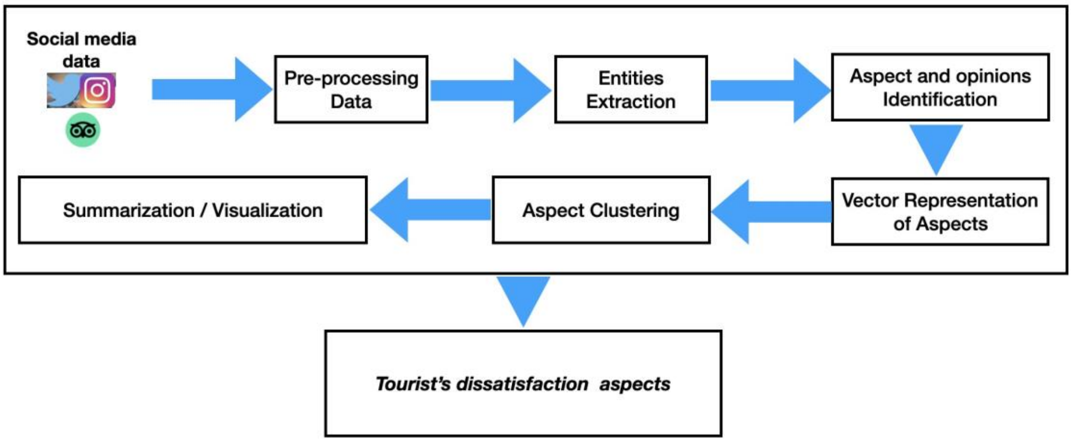


Figure 1. *Proposed ABSA approach*

Pre-processing data

In this study, we use data from Twitter and Instagram obtained in the study proposed in Viñán-Ludeña and de Campos (2022a, 2022b), in which, the python scraping tool called "Twint" were used to obtain tweets without its API; for Instagram, a Java program was necessary to obtain posts and the keywords used were: *granadaturismo*, *teenseñomigranada*, *alhambracultura*, *#alhambra*, *granada turismo*, *turismogranada*, *gastronomia granada*, *gastronomiagranada*, *hoteles granada*, *hotelesgranada*, *granadahoteles*, *restaurantes granada*, *restaurantesgranada*, *granadarestaurantes*, *#planesgranada*, *#albaicín* and *#sierranevada*. The keywords chosen for English in both platforms are *#welovegranada*, *#granadatrip*, *#granadatravel*, *granadatourism*, *granadatour*, *granadatours*, *granadatourisme*, *granadatourtravel*, *granadatravelcenter*, *granadatravels*, *travelgranada*, *granadatraveler*, *traveleringranada*, *sixsensestravelsgranada*, *granadatraveltips*, *triptogranada*, *thingstodoingranada*, *granadathingstodo*, *granadahotels*, *granadaluxuryhotels*, *cheaphotelsgranada* and *granadarestaurant*. Twint allows data collection since 2008, so we gathered tweets from 2008 to July 2020. TripAdvisor reviews were captured without date limitations, up to July 2022. Instagram posts were collected until July 2020, without any date restrictions. It is important to note that we utilized data from Twitter and Instagram until 2020 because the dataset was used in other studies (Viñán-Ludeña & de Campos, 2022a), and TripAdvisor reviews were captured until the project finished in July 2022.

The dataset corresponds to Granada, an important Spanish tourist destination. We used 19,340 tweets and 7,717 Instagram posts, all of which were written in English. Since ABSA requires that greetings, questions, compliments and farewells be discarded, we filter the tweets and posts accordingly and obtain 2,613 English tweets, 7,712 English posts (only 5 Instagram posts were removed as they do not generally contain greetings, unlike Twitter) and 25,483 TripAdvisor reviews about Granada in Spain (a python crawler was used to collect the reviews according to where they had been posted). It is essential to note that the tweets, posts, and reviews were used anonymously, and no personal user data was utilized in any analysis.

Once the input data has been loaded, every tweet, post or review is divided into sentences and the following operations are performed for each sentence:

- Removal of links using URL patterns
- Removal of user mentions
- Removal of hashtags
- Removal of non-ASCII characters (emoticons are not taken into account because the objective of this work is to find the entity attributes or aspects)
- Removal of punctuation marks
- Removal of stop words, i.e. irrelevant words included in the text

Entity extraction

Until this present day, researchers have applied ABSA with prior knowledge of the target or entity (i.e. restaurants, hotels, laptops, etc.). However, in this study we need to identify the entities that users mention on social media. We also use TripAdvisor reviews and on TripAdvisor, it is necessary to specify the entity or service in order to extract reviews about a particular entity, unlike Twitter or Instagram where the entity or service is not specified. We therefore use Twitter to identify the most important entities at a tourist destination and then apply our proposed approach with the three datasets (Twitter, Instagram and TripAdvisor). Since it is necessary to identify the tourist destination places (e.g. Alhambra[7], Albaicin[8], etc.), the service or event (e.g. Holy Week processions), we propose the following semi-automatic process for entity identification: (1) Entity candidates were identified using all the tweets by means of the

[7] The Alhambra is a palace and fortress located in Granada, Andalusia, Spain.

[8] The Albaicin is a district in the city of Granada.

approach[9] proposed by Schweter and Akbik (2020), which is a 4-class named entity recognition (NER) BERT-model for Spanish. The four classes of name identified are the person (PER), location (LOC), organization (ORG) and miscellaneous (MISC). (2) The frequency was then calculated for each identified entity. (3) Finally, a tourist destination resident grouped the entities according to place, service or event. It is important to note that we used a frequency threshold to discard irrelevant entities.

This process was performed using 2,613 English tweets and we included 21,143 Spanish tweets so that there were a representative number of tweets. We noticed that keywords which belong to specific entities might be written in Spanish and English: for example, in the English tweet "The procession won't continue its parade #SSantaGr" and the Spanish tweet "La Hermandad del Huerto tampoco sale a procesionar #SSantaGr #SemanaSanta #Granada" which announces that "the Hermandad del Huerto" canceled their procession, both tweets contain the same hashtag/keyword #SSantaGr, which refers to Holy Week. By including Spanish data, we obtained the following entities: Alhambra, Albaicin, Generalife, Semana Santa (Holy Week), Palacio de Carlos V, Patio de los Leones, Alpujarra, Guadix, Mirador de San Nicol´as (San Nicol´as Viewpoint), Sacromonte, Motril, Realejo, Parque de las Ciencias, Federico Garc´ıa Lorca, Almun˜´ecar, Valle de Lecrin, Paseo de los Tristes, Sierra Nevada. These identified entities mainly correspond to places in Granada, although there are also cultural events (e.g. Holy Week or the poet Federico Garc´ıa Lorca, who was born in Granada). As there are fewer negative tweets, posts and reviews than positive opinions, in order to perform the dissatisfaction analysis, we only used the following entities:

- Alhambra
- Albaicin
- Generalife[10]
- Sacromonte[11]

Aspect and opinion identification

This approach applies grammar rules to aspect-term identification. This is important since nouns are commonly aspects and adjectives are connected with the opinion words and so we need to recall certain rules. Firstly, the order of a basic positive sentence is Subject-Verb-Object but negative and

[9] The Flair Hugging Face website can be found at https://huggingface.co/flair/ner-spanish-large.

[10] The Generalife is a summer palace and country estate of the Nasrid rulers of the Emirate of Granada.

[11] The Sacromonte is a gypsy neighborhood and is the flamenco capital of Granada.

question sentences might have a different structure. Secondly, adjectives usually come before a noun, except when a verb separates the adjective from the noun. Thirdly, when using two or more adjectives together, the usual order is opinion-adjective + fact-adjective + noun. Although there are many grammar rules, we have selected the most important ones in order to build a proper algorithm allowing aspect and opinion identification. In this study, we use English data but these rules can be applied in other languages such as Spanish, French, etc.

These grammar rules might be converted into a representation that can be understood by a computer. "Stanford typed dependencies" provide a simple representation of grammar rules and relationships which is accessible to people without any linguistic expertise. These dependencies are all binary relations between two sentence words, arranged in a hierarchy and contain 56 grammatical relations. Since a grammatical relation holds between a *head* and a *dependent*, we must therefore identify the grammatical relation, head and dependent. For example, the sentence: "The Alhambra is the most beautiful place." has the following relations under this representation: (i) **nsubj**(place, Alhambra), (ii) **amod**(place, beautiful), (iii) **cop**(place, is), (iv) **det**(place, the), (v) **advmod**(most, beautiful). The grammatical relations (e.g. nsubj, amod, etc) were needed to build an algorithm which enables aspects-terms to be identified and these are described below.

We use the CoreNLP library[12] developed by the Stanford NLP Group in order to perform aspect and opinion identification through grammar dependencies expressed by a relation type, head and dependent[13]. Some rules were selected from (Dragoni et al., 2019) and two lexicons were used to identify word polarity: SenticNet[14] and the opinion lexicon proposed by Liu (2010):

- **Compound:** If the head and dependent are nouns, we join them using the character " _" to obtain a compound aspect. For example, in the sentence "Great sunset view from the old Arab quarter", one of the triples is (compound, sunset, view) and in this triple, both "sunset" and "view" are nouns and so the resulting aspect is sunset _view).

[12] https://stanfordnlp.github.io/CoreNLP/

[13] https://downloads.cs.stanford.edu/nlp/software/dependencies

- **Adjectival modifier "amod"**: This rule is applied if the head is an aspect (noun) and the dependent is an adjective with a polarity value. For example, in the sentence "Very poor online booking system", one of the triples is (amod, booking _system, poor), and in this triple, we have a compound aspect "booking system" as the head, and a negative adjective "poor" as the dependent.

- **Nominal subject "nsubj"**: The head should have a polarity value to apply this rule: for example, in the case of the triple (nsubj, fantastic, view), we have a positive opinion (fantastic) about the aspect (view).

- **Conjunction "conj"**: If the head and dependent are aspects, then if one of them is present in the "amod" rule, the other aspect should be assigned the same adjective. For example, in the sentence "This place has great music and decoration.", we have the triples (amod, music, great) and (conj, music, decoration), and so we can generate the triple (amod, decoration, great). By way of contrast, if the head and dependent are adjectives or polarity words, then if one of them is present in a "nsubj" rule, the other adjective should be assigned the same aspect. For example, in the sentence "This service is bad and expensive", we have the triples (nsubj, bad, service) and (conj, bad, expensive) and so we can generate the triple (nsubj, expensive, service)

- **Negation**: If the words such as not, never, neither, nor, can't, etc. are present in the text, the polarity of the opinion words is modified.

Algorithm 1. *Aspect and opinion identification*

```
entities ← entities_identification_process()
for t in tweet/post/review do
    if t == review then
        important_entity ← Null
    else
        pub_entities ← get_entities(t, entities)
        important_entity ← get_entity(pub_entities, entities)
    end if
    aspect_opinion_rules ← [amod, nsub, dobj, compound, conjunction, negation]
    sentences ← split_text(t)
    for s in sentences do
        rules ← get_rules(s)
        result ← get_aspects_opinions(rules, aspect_opinion_rules)
    end for
end for
```

In summary, we first identify entities from the text for every tweet, post or review (with the exception of TripAdvisor reviews since we already know the entity). Secondly, we split the text into sentences, and for each sentence, we identify the most important entity according to the frequency and associate that entity with the sentence. For example, if a text has the two entities of Alhambra (2952) and Albaicin (197), we select Alhambra as the important entity in this sentence for the tweet or post (the process of identifying the most important entity is excluded for TripAdvisor reviews). Thirdly, we identify the rules and parts of speech for each word. Finally, we apply the previously described rules to aspect and opinion identification. This procedure is summarized in Algorithm 1.

Vector representation of aspects

The process of transforming words into vector representations is an important step in natural language processing. These representations can be used for aspect categorization or clustering. After completing the aspect-term extraction, we transform each aspect into a vector representation of finite dimension using ConceptNet (CN) Numberbatch (Speer et al., 2017). This set of semantic vectors was selected because it is built using a set which combines ConceptNet data[13], word2vec, GloVe, and OpenSubtitles 2016[15]. The word embedding or numeric vector representation of text used in this study through pre-trained CN Numberbatch thereby enables us to maintain the semantic and contextual relationships within the aspects in our dataset so that they can be clustered.

Aspect clustering

Since word embedding is necessary for aspect clustering, we apply word embedding to each aspect. This enables the information to be summarized through visualizations or summaries in order to understand traveler satisfaction with the tourism destination.

Once each aspect has been represented as a vector, aspect clustering is performed. Clustering is the process whereby sets of objects are grouped into classes on account of the fact that objects in the same group are more similar to each other than to those in another group. We therefore need to identify which aspects are similar to each other: for example, the aspects "customer service, reservation system" and "service" are similar and might be grouped into one cluster whereas the aspect "tour experience" should be grouped in another cluster. We use k-means to perform aspects clustering

[13] https://conceptnet.io
[14] https://www.opensubtitles.org/en/search/subs

with cosine similarity. Additionally, we use the Silhouette metric to determine the optimal number of clusters (Rousseeuw, 1987).

Visualization / Summarization

Once we have identified places and events in a tourism destination, the next stages must be performed. The visualization or word-tree shows the combinations entitycluster-aspect-opinion word(s) and comprises four levels: level 0 (root node) corresponds to the entity (e.g. Alhambra); level 1 contains the numbers which correspond to the clusters found with the k-means algorithm; level 2 corresponds to the aspects, which are commonly nouns (e.g. audio system, beauty, place and entrance, etc.); and level 3 corresponds to the sentiment words, which are commonly adjectives (e.g. awful, dull, sad, etc.). Every word-tree was created using the D3.js library[16]. Once we had obtained the final dataset, we used word trees to examine the opinions where different aspects appear. The chart gives us a general idea about the satisfaction of the entity.

Lastly, we performed the summarization process and this provide us with all the relevant and most important information from the tweets, posts and reviews without having to read every social media post. There are two summarization categories: *extractive text summarization*, which extracts the significant sentences from the text and *abstractive text summarization*, which is an advanced method to identify the important sections of the text, interpret the context and compile a summary with the core information in a different way. In this study, we used the first category because we have a series of posts, tweets and reviews rather than a large document and this is the best option for summarizing social media content.

In order to perform the summaries, we selected two tools which were constructed using BERT technology. The first of these is BART and was proposed by a Facebook team (Lewis et al., 2019), whereby the authors pre-trained their model using the English language and fine-tuned and improved it with the CNN Dailymail dataset[17] (which contains over 300,000 unique news articles written by CNN and Daily Mail journalists). This tool achieves good results in terms of abstractive dialogue, question answering and summarization tasks.

[16] This is a Javascript library for producing interactive visualizations and further information can be found at https://d3js.org/.

[17] More information about this dataset can be found at https://huggingface.co/datasets/cnn dailymail.

The second tool was built using a transformer created by a Google team (Raffel et al., 2019). This transformer is called "T5" which stands for Text-To-Text Transfer Transformer[18] and was fine-tuned with 4,515 examples of news articles from The Hindu, The Indian times and The Guardian[19].

It is possible to generate summaries about entities by taking into account every aspect belonging to a specific entity, or selecting one cluster of an entity that includes any reviews, tweets or posts relating to aspects belonging to that cluster. Summaries can also be generated using sentences belonging to two or more clusters.

RESULTS

We will now proceed to analyze user dissatisfaction perceptions for each entity grouping data from Twitter, Instagram and TripAdvisor to perform the visualization and summarization.

The Alhambra

TripAdvisor reviews are subjective and so they provide opinion perceptions about a place or service. We apply the same tools in both Twitter and Instagram to identify aspects and opinions. Although users generally rate their TripAdvisor reviews as positive or negative, the aim of this study is to find aspects and the positive or negative opinions in these reviews. We therefore use every review regardless of its rating since positive reviews also contain negative aspects.

As the Alhambra is the most visited place in Granada, there were a lot of references to it (16,116 TripAdvisor reviews, 891 Instagram posts and 499 tweets). Figure 2 shows the more important aspects and the respective negative opinions in the TripAdvisor reviews, tweets and Instagram posts. For the sake of completeness, the word-tree including all the aspects for the Alhambra is displayed in the Appendix.

Although we have detected many clusters about the Alhambra, we have only selected Cluster 6 to show the automatic summaries:

- **BART Tool Summary**: "Once you are inside there is little or no information on signboards so unless you purchase their terrible audio guide

[18] https://ai.googleblog.com/2020/02/exploring-transfer-learning-with-t5.html.

[19] https://huggingface.co/mrm8488/t5-base-finetuned-summarize-news.

or have prebooked a tour you are walking blind. Long queues, badly organised and very rude staff. Very poor online booking system, ticketmaster has a very poor ticketing system."

- **T5-base fine-tuned Tool Summary**: "ticket office was ruined by the stupid ticket machine operated by ticketmaster I planned my night visit to Alhambra, and got the ticket at the ticket machine beside the ticket office. My visit was ruined by the stupid ticket system operated by ticketmaster. Long queues, badly organised and very rude staff. Worst part is the ridiculous ticket office. Unless you book your tour or audio guide, you are walking blind. You will likely be ruined by the poor online booking system."

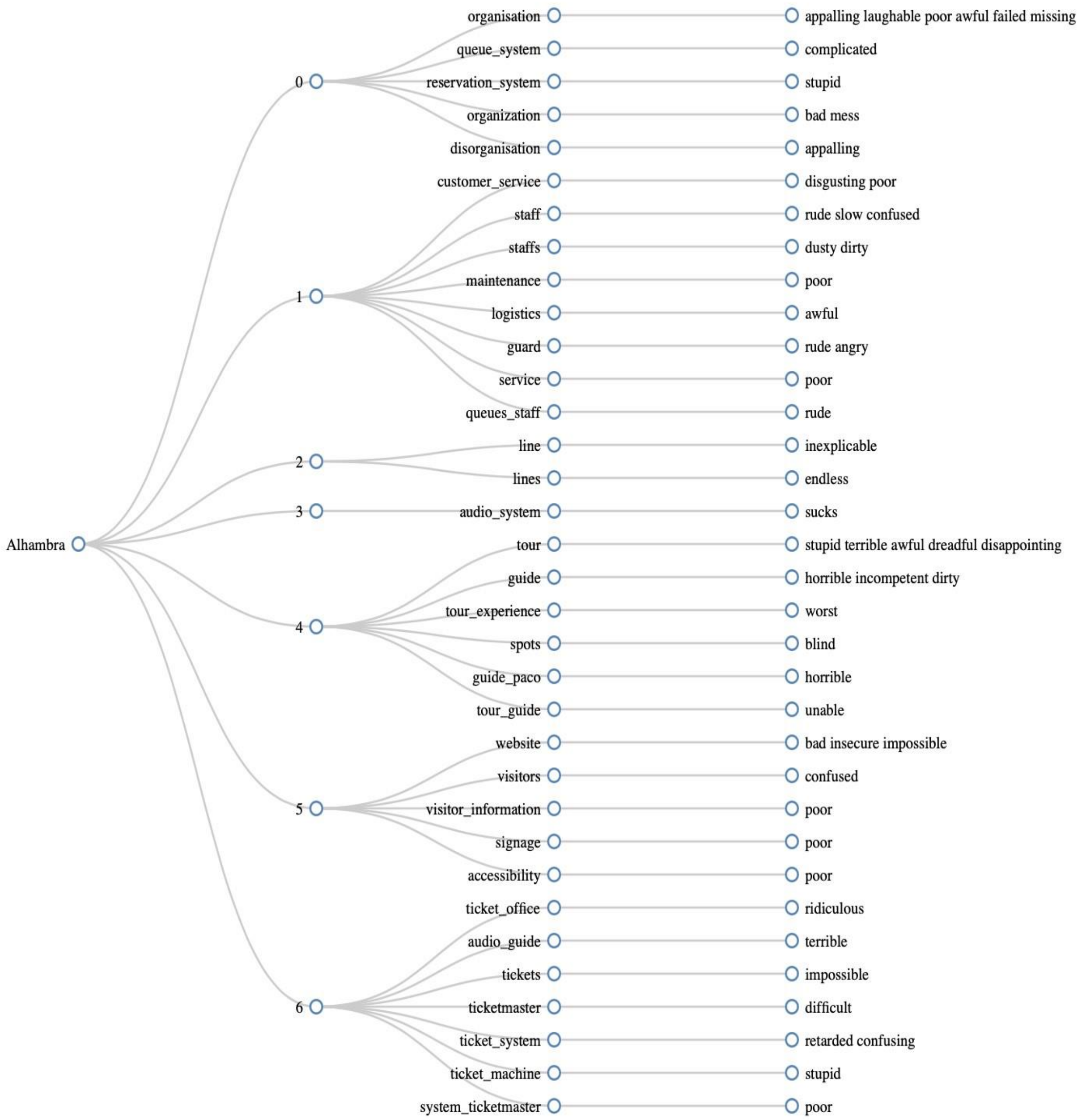


Figure 2. *Negative perceptions about the Alhambra*

The most important points that can be highlighted from the summaries corresponding to the Alhambra are problems with the ticket system, as well as with audio guides and signboards.

If the results are analyzed for each social media platform, the most important entity, according to Twitter data is the "Alhambra" as this is the most visited monument in Granada. However, the most important negative aspects and opinions or tourist dissatisfaction concern the ticket office, the online ticket booking system and the audio guides which narrate the history of the palaces, etc. These clues are essential to improve the services at this fortified palace complex. TripAdvisor data also shows that the most important entity is the "Alhambra" and tourists were most dissatisfied with the staff and organization, queues, ticket booking system, customer service, logistics, insecure website, visitor information, audio guide, ticket system, etc. Along the same lines, Instagram data reflects the fact that the most important entity is the "Alhambra" and some of the most important aspects and opinions include bad impressions, empty halls, annoying people, etc.

The Albaicin

The Albaicin is another important place in Granada. We obtained 2,456 reviews and the sentences express positive or negative opinions about the aspects. We obtained only 5 sentences with positive opinions and 1 with a negative opinion with twitter data and 16 Instagram posts and of these, 4 were negative. Figure 3 shows the word-tree about the negative aspects and opinions.

The following summaries were obtained from the tools used in this study:

- **BART Tool Summary**: "Small winding streets but dwellings mainly hidden behind high walls. Would be charming but for amount of dog faeces and dodgy loiterers. Full of totally useless souvenir shops that looks cool only in their context but when you actually buy it youll feel almost scammed. Poor signposting and poorer tourist maps mean it is easy to get lost."

- **T5-base fine-tuned Tool Summary**: "We had a fixed walk with an awful guide called diego. Would be charming but for amount of dog faeces and dodgy loiterers. Poor signposting and poorer tourist maps mean that it is easy to get lost and end up in rough looking areas full of people and traditional shops. The whole atmosphere is fake and almost a tourist cash grab trap."

The summaries about the Albaicin highlight the lack of cleanliness, poor signage and how easy it is to get lost in its labyrinth-like streets. According to TripAdvisor data, tourists expressed their dissatisfaction with the souvenir shops, guides, poor signposting, etc. On the other hand, Twitter and Instagram data does not contain enough negative aspects to draw conclusions about this entity.

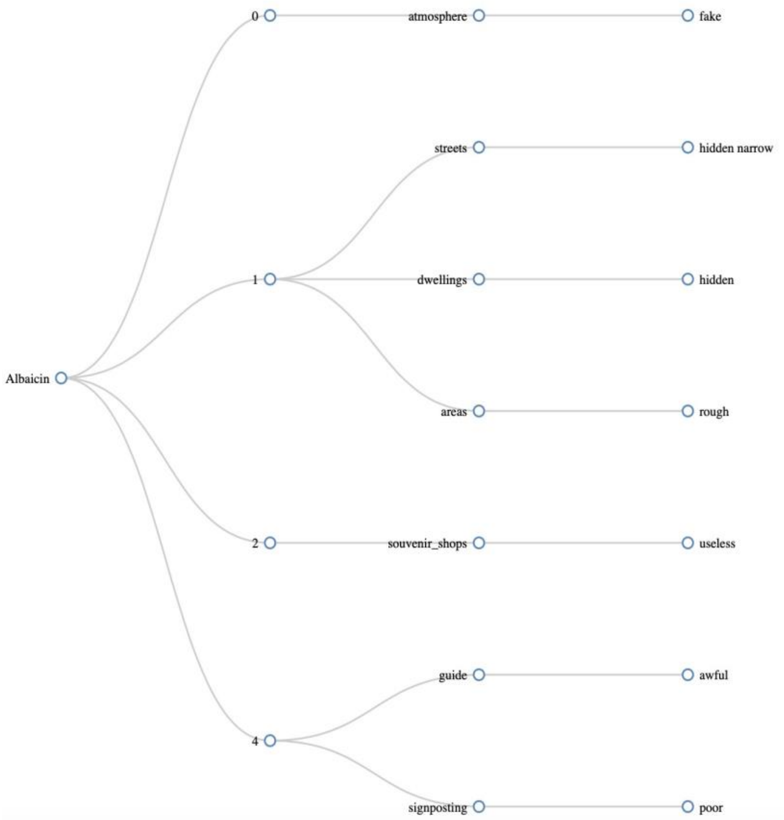


Figure 3. *Negative perceptions about Albaicin*

The Generalife

The Generalife is an important place to visit in Granada. We obtained 3,895 reviews and we employed the proposed approach to obtain only a few negative opinions and aspects about this entity. There are not negative opinions in Twitter and Instagram data about this entity. Figure 4 shows the aspects for the Generalife.

The following summaries were obtained from the tools used in this study:

- **BART Tool Summary**: "The gardens were in a mess and the plants on display were not the type of plants that would have been used so the character of the gardens was nothing like the original. The gardens were very tired with little or no consideration for autumn planting or colour. The internet site was very confusing you don't get any idea what you are buying."

- **T5-base fine-tuned Tool Summary**: "Alhambra itself was nice however the tour was poor. The gardens were in a mess and the plants on display were not the type that would have been used. The staff at the venue need to do more to fulfill the hype and plant natives or give the effect of what the plantings were like originally."

In general, the negative aspects and opinions mention poor organization, poor tour, confusing website, the messy gardens, the lack of plants native to the site, etc.

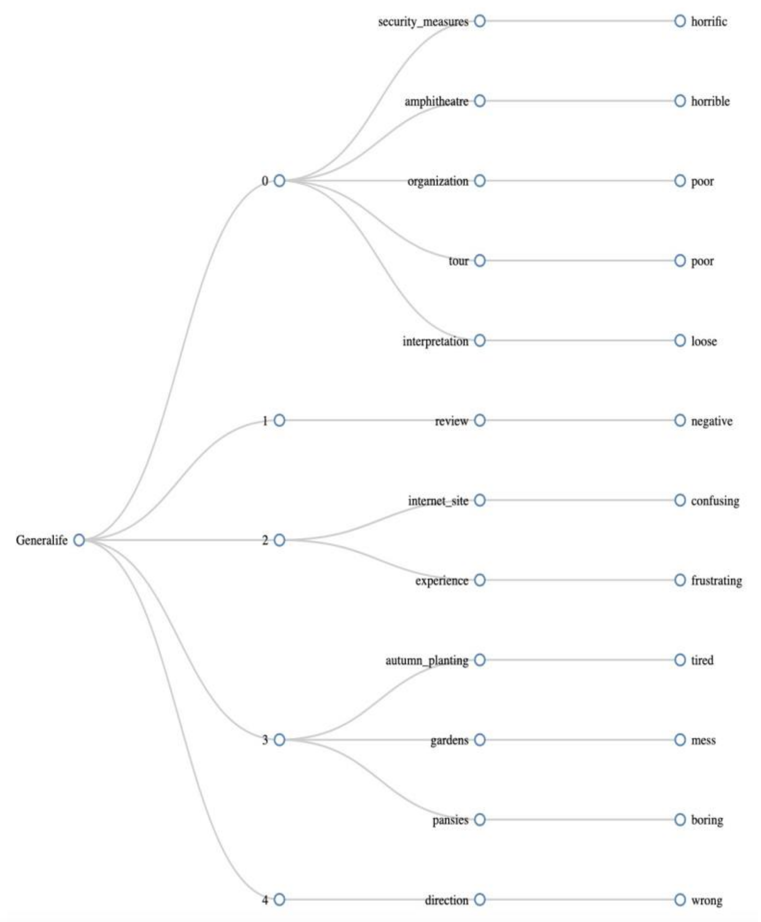


Figure 4. *Negative perceptions about the Generalife*

Sacromonte

Sacromonte is another important place to visit in Granada and we obtained 570 reviews and without opinions from Twitter and Instagram data about it. However, once the proposed approach had been employed, very few aspects were obtained. Figure 5 shows the negative opinions and aspects about this entity.

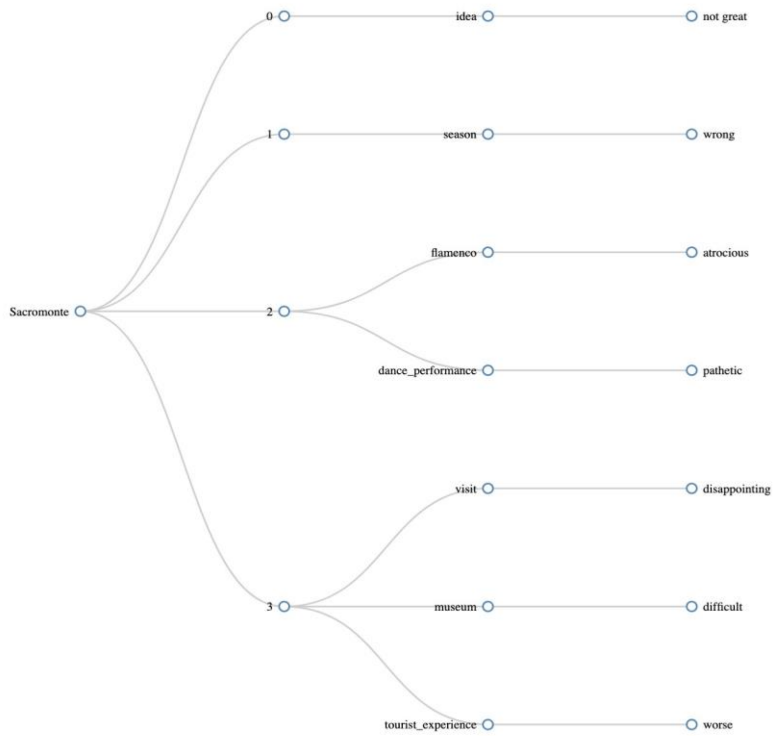


Figure 5. *Negative perceptions about Sacromonte*

The following summaries were obtained from the tools used in this study:

- **BART Tool Summary**: "A long uphill walk to visit caves but the museum was closed even though it said opening hours were from 10:00 am. The museum of rural life at sacramonte is difficult to reach without a car and involves climbing several flights of steps from the road. The intention was to see a touristytype flamenco ie focus on the woman dancer with all her polka dotted finery.. the show itself was pretty much a disaster."

- **T5-base fine-tuned Tool Summary**: "I was accosted by a very aggressive young man who threatened me with bodily injury while walking through the old part of sacromonte. The museum of rural life at sacromonte is difficult to reach without a car and involves climbing several flights of steps from the road. It is the worse tourist experience i have ever had."

The summaries that correspond to the Sacromonte highlight the difficult access to the "museum of rural life", a disastrous flamenco show, awful dance performances and how dangerous it is to walk through the old part of this neighborhood.

DISCUSSION

Our approach uses APIs, crawlers or scrapers for data acquisition, which allows to easily automate this process in the tourism (dis)satisfaction literature. Furthermore, this study uses aspect-based sentiment analysis with grammar rules (i.e. no training data is needed) and thus it can be easily automated and implemented in a mobile/web application. In contrast, other studies use face-to-face semi-structured interviews to capture the (dis)satisfaction opinions of multifaceted spectrum of travelers (Bianchi, 2016). On the other hand, Taheri et al. (2020) use surveys and the data are analyzed with Partial Least Squares (PLS) and multi-group analysis (MGA) and focus on airport traveler dissatisfaction. Ma et al. (2022) also use surveys to explore the relationships between tourists' perceived deception, dissatisfaction, revisit intention and negative word of mouth. In view of the above, data collection is expensive and hard to replicate, and the automation is not possible with classical statistical methods.

One of the studies that should be highlighted is (Prakash et al., 2019) where the authors examined the major causes of visitor dissatisfaction during wildlife tourism through TripAdvisor reviews. Their results were obtained through content analysis and are related with management problems and sustainable tourism. We propose a simple method for obtaining the causes of tourism dissatisfaction, not only for wildlife tourism or another specific case but to tourism in general, and reaffirm the fact that tourism data from social media platforms are reliable to analyze visitor feedback information on travel destinations and services.

Tang and Zeng (2021) propose a method for calculating tourism e-commerce user satisfaction. They propose factors such as management service, tour guide service, travel support service, attractions tour service and contract travel service. To each of these factors, they assign a value corresponding to satisfaction. However, there may be more factors and categories and therefore, a more general method is more useful. In this sense, the proposed aspect clustering by means of word embedding is more suitable for finding the user perception factors posted on social media.

Kuhzady and Ghasemi (2019) propose a text mining method to show the positive and negative opinions through a word cloud. This approach does not show in a clear way which are the causes of dissatisfaction. While, Park et al. (2021) use a regression method to explore the dissatisfaction at luxury hotels.

It is important to note that there is no model that incorporates all the possible attributes that allow the evaluation of tourist dissatisfaction. Alegre and Garau (2010) mention 13 characteristics that were qualified in terms of dissatisfaction, such as destruction of landscape, too much building, too many people, expensive, and so on. However, having a general model that can involve the characteristics of tourist dissatisfaction is unfeasible, because destination management organizations incorporate new technologies and gradually becoming involved in industry 4.0 that incorporates a series of devices and sensors that are part of smart-tourism; this leads to incorporate new features to the dissatisfaction model, as is the case of a ticket system, digital twins for customer service, etc. In addition, a tourist destination that is located on the beach does not have the same dissatisfaction characteristics as mountain or adventure tourism. Therefore, our study does not propose a new model, nor does it incorporate new characteristics to the existing model; rather, it identifies the characteristics or complaints through social media data as a practical and inexpensive approach for the evaluation of tourist dissatisfaction, allowing to apply this study in any tourist destination of any type. Furthermore, none of the previously analyzed studies uses platforms such as Twitter or Instagram to explore tourists' complaints. Therefore, this study contributes to the existing literature with an automated approach that can be applied to any social media platform and furthermore it can be used in the field of accommodation, gastronomy and tourism in general.

Researchers base their research on specific types of tourism to evaluate dissatisfaction, such as medical tourism (Um & Kim, 2018), tourists' dissatisfaction caused by failures in tourist and cultural services (Jiang et al., 2022; Lam-González et al., 2021), dissatisfaction in thermal and mineral spas (Rodrigues et al., 2020), evaluating service failure situations at a restaurant (Jang et al., 2013), cultural heritage tourism destination (Thanyasunthornsakun, 2016) or airport servicescape (Taheri et al., 2020). Our study is based on the perceptions that travelers have about a destination, identifying the attributes of tourist dissatisfaction that allow us to evaluate the problems of a given tourist destination without a model based on surveys or traditional statistical methods, allowing managers to find the causes of dissatisfaction with the aim of improving the management of places, events or services.

Finally, the most significant contributions of this study lie in the psychological variables of motivations and attitudes related to tourist satisfaction and dissatisfaction. Unlike other researchers who have focused on designing and validating annotation schemas for aspect-based sentiment analysis (ABSA) (Moreno-Ortiz et al., 2019) or methodological aspects (Ghosal & Jain, 2023), this study aims to identify the variables that motivate dissatisfaction in the tourism destination, in this case Granada, Spain. Additionally, the study highlights the importance of destination perception in determining tourist satisfaction. To achieve this, we first identify the main entities in a tourism destination and then evaluate the satisfaction and dissatisfaction of the principal attractions within that destination. In contrast, other studies have focused on specific types of reviews, such as restaurant reviews (Li et al., 2023) or hotel reviews (Ozen & Ozgul Katlav, 2023; Shahhosseini & Nasr, 2024; Zhang et al., 2023).

Herzberg's Motivator and Hygiene Factor Theory (Chan & Baum, 2007) is applied to identify the factors that satisfy and dissatisfy tourists, including hygiene factors such as cleanliness and amenities. Our study found important aspects related to the theory, such as organizations, staff, and customer service. Furthermore, we identify the factors that contribute to dissatisfaction as a crucial theoretical aspect in destination management and, we emphasize that the use of satisfaction as a driving force for loyalty and the development of marketing strategies that cater to the needs and wants of tourists is essential. Our approach can be used to identify problems in tourism destinations and their services, providing insights on enhancing tourist satisfaction.

CONCLUSIONS AND IMPLICATIONS

User-generated social media content plays a fundamental role in monitoring tourist satisfaction or dissatisfaction. We have analyzed Twitter, TripAdvisor and Instagram data using an aspect-based sentiment analysis approach based on rules and this is presented in this study. This approach includes: a semi-supervised algorithm to identify the most important entities (places or events) at a tourist destination; a rule-based algorithm for aspect identification and opinion calculation (and although this can be applied using data in other languages, we only use English tweets, posts and reviews in our study); a visualization process, which includes vector representation of aspects through ConceptNet Numberbatch in order to keep the semantic and contextual relationships within the aspects and a clustering process using k-means and word-trees to better understand the aspect-opinion or dissatisfaction factors. This final stage of our approach applies two different summarizer tools which generate short summaries. Although these might be generated using whole word-tree or by cluster, if there are many aspects or opinions, it is advisable to apply these tools for each cluster. The information provided enables managers and other operators to improve their services.

We examined 19,340 tweets, 7,712 Instagram posts and 25,483 TripAdvisor reviews of Granada (an important tourist destination in Spain). The entities analyzed were the Alhambra, Albaicin, Generalife and Sacromonte.

As the results demonstrated, TripAdvisor and Twitter provide more subjective information than Instagram. In terms of the number of aspects displayed, TripAdvisor might reflect a greater use to spread complaints about a certain resource, Twitter has a moderate number of dissatisfaction factors while Instagram seems less useful, with users sharing news and leisure activities (i.e. sad stories, tragic love stories, bad movies, terrible companies, etc); in addition, Instagram users usually share their experiences at specific points of their trip to show sights and landscapes from the tourist destination whereby, they also tend to show their meals and drinks (i.e. beer cold, happyhour michelada cold) and, there is also a tendency for these posts to be written in a poetic or metaphorical way. This study can be used either to evaluate the current situation of the places, events and services relating to a specific tourist destination or event or to perform an independent audit to identify problems so that any necessary action can be taken and improvements made.

The present study has some limitations. Classifying the posts/reviews into domestic tourism and international tourism would be of great importance to compare how dissatisfaction attributes differ between these two segments. On the other hand, the classification by age or gender of the posts/reviews has not been taken into account and would be very useful to focus improvement plans to a specific population. In addition, fake news and bots are not taken into account, which can have a negative impact on the results.

In the future, we will train a BERT model to improve aspect-opinion identification, which is an important stage in aspect-based sentiment analysis. The construction of a recommender system which incorporates dissatisfaction factors and management strategies and monitors their effectiveness may be useful for tourism sector professionals and managers. Furthermore, summarization is an important feature to understand tourism (dis)satisfaction with social media data, so that managers can obtain insights about user perceptions through few sentences. However, we utilized only two tools to perform this process and these can be improved. Thus, we will propose, train and evaluate a BERT-based model to summarize social media text with special focus in tourism (dis)satisfaction. We will also evaluate our novel aspect clustering by means of word embeddings with another approaches, in order to use it with other algorithms and applied in other domains. In addition, since the results of this study can be potentially very useful to managers of a tourist destination, in the future we plan to conduct a dedicated study for users' evaluation with the active participation of the administrators of the places, events or tourist services.

Finally, our work has important implications not only for professionals and managers but also for researchers.

Research Implications

Smart-Tourism refers to the interaction and/or combination of communication networks, internet, sensors, internet of things and tourism. Therefore, the contribution made in this study to the field of Smart-Tourism is important, due to the fact that the data that were analyzed come from social media that tourists use while they are visiting some place and thus, communication networks, their mobile devices and internet are necessary. Therefore, this study contributes to the area of smart tourism by using an interesting approach based on computational methods, considering that previous studies about complaints in the tourism domain use descriptive statistical methods and mostly focus on hotel complaints. Our findings, therefore, supplement the limited amount of knowledge and research into tourist complaints and opinions on social media and they establish the baseline for future studies and analysis based on traveler satisfaction with a tourist destination.

Practical Implications

For tourism industry professionals and managers, it is extremely important to be aware of what is happening on social media where travelers voluntarily share their experiences about a tourist destination. Analysis of this data affords us with an excellent opportunity to identify the causes or reasons for tourist dissatisfaction so that managerial strategies can be improved. The approach proposed in this study can be easily applied to construct a software or application to re-evaluate tourist services and strategies every so often since it is fully available on Github for further research and practical purposes.

ACKNOWLEDGMENT

This work has been funded by the Spanish Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación under project PID2019-106758GB-C31, and the European Regional Development Fund (ERDF-FEDER).

REFERENCES

Afzaal, M., Usman, M., & Fong, A. (2019). Tourism mobile app with aspect-based sentiment classification framework for tourist reviews. *IEEE Transactions on Consumer Electronics*, *65*(2), 233-242.

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: Capitalizing on big data. *Journal of Travel Research*, *58*(2), 175–191.

Alegre, J., & Garau, J. (2010). Tourist satisfaction and dissatisfaction. *Annals of Tourism Research*, *37*(1), 52-73.

Ali, T., Marc, B., Omar, B., Soulaimane, K., & Larbi, S. (2021). Exploring destination's negative e-reputation using aspect based sentiment analysis approach: Case of Marrakech destination on Tripadvisor. *Tourism Management Perspectives*, *40*, 100892.

Ansar, W., Goswami, S., Chakrabarti, A., & Chakraborty, B. (2021). An efficient methodology for aspect-based sentiment analysis using BERT through refined aspect extraction. *Journal of Intelligent & Fuzzy Systems*, *40*(5), 9627–9644.

Bianchi, C. (2016). Solo holiday travellers: Motivators and drivers of satisfaction and dissatisfaction. *International Journal of Tourism Research*, *18*(2), 197-208.

Calderón-Fajardo, V., Anaya-Sánchez, R., & Molinillo, S. (2024). Understanding destination brand experience through data mining and machine learning. *Journal of Destination Marketing & Management*, *31*, 100862

Carenini, G., Cheung, J. C. K., & Pauls, A. (2013). Multi-document summarization of evaluative text. *Computational Intelligence*, *29*(4), 545–576.

Chan, J. K. L., & Baum, T. (2007). Determination of satisfiers and dissatisfiers using Herzberg's motivator and hygiene factor theory: An exploratory study. *Tourism Culture & Communication*, *7*(2), 117–131.

Chauhan, G. S., Kumar Meena, Y., Gopalani, D., & Nahta, R. (2020). A two-step hybrid unsupervised model with attention mechanism for aspect extraction. *Expert Systems with Applications*, *161*, 113673.

Chu, M., Chen, Y., Yang, L., & Wang, J. (2022). Language interpretation in travel guidance platform: Text mining and sentiment analysis of Tripadvisor reviews. *Frontiers in Psychology*, *13*, 1029945.

Di Fabbrizio, G., Stent, A., & Gaizauskas, R. (2014). A hybrid approach to multi-document summarization of opinions in reviews. In *Proceedings of the 8th international natural language generation conference (inlg)* (pp. 54–63).

Dragoni, M., Federici, M., & Rexha, A. (2019). An unsupervised aspect extraction strategy for monitoring real-time reviews stream. *Information Processing & Management*, *56*(3), 1103-1118.

Ekinci, E., & Ilhan Omurca, S. (2020). Concept-lda: Incorporating babelfy into lda for aspect˙ extraction. *Journal of Information Science*, *46*(3), 406-418.

Fernandes, T., & Fernandes, F. (2018). Sharing dissatisfaction online: Analyzing the nature and predictors of hotel guests negative reviews. *Journal of Hospitality Marketing & Management*, *27*(2), 127-150.

Gerani, S., Carenini, G., & Ng, R. T. (2019). Modeling content and structure for abstractive review summarization. *Computer Speech & Language*, *53*, 302-331. https://doi.org/10.1016/j.csl.2016.06.005

Ghosal, S., & Jain, A. (2023). Weighted aspect based sentiment analysis using extended OWA operators and Word2Vec for tourism. *Multimedia Tools and Applications*, *82*(12), 18353–18380.

He, J., Li, L., Wang, Y., & Wu, X. (2021). Hierarchical features-based targeted aspect extraction from online reviews. *Intelligent Data Analysis*, *25*(1), 205-223.

Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. In *Aaai* (Vol. 4, pp. 755–760).

Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? text analysis using structural topic model. *Tourism Management*, *72*, 417-426.

Jang, Y. J., Cho, S.-B., & Kim, W. G. (2013). Effect of restaurant patrons' regret and disappointment on dissatisfaction and behavioral intention. *Journal of Travel & Tourism Marketing*, *30*(5), 431-444.

Jiang, S., Moyle, B., Yung, R., Tao, L., & Scott, N. (2022). Augmented reality and the enhancement of memorable tourism experiences at heritage sites. *Current Issues in Tourism*, *26*(2), 242-257.

Kim, S. S., Shin, W., & Kim, H.-W. (2024). Unravelling long-stay tourist experiences and satisfaction: text mining and deep learning approaches. *Current Issues in Tourism,* 1–19. https://doi.org/10.1080/13683500.2024.2327840

Kim, Y.-J., & Kim, H.-S. (2022). The impact of hotel customer experience on customer satisfaction through online reviews. *Sustainability*, *14*(2), 848.

Kuhzady, S., & Ghasemi, V. (2019). Factors influencing customers' satisfaction and dissatisfaction with hotels: A text-mining approach. *Tourism Analysis*, *24*(1), 69-79.

Kumar, A., Saini, M., & Sharan, A. (2020). Aspect category detection using statistical and semantic association. *Computational Intelligence*, *36*(3), 1161-1182.

Lam-González, Y. E., Clouet, R., Cruz Sosa, N., & de Le´on, J. (2021). Dissatisfaction responses of tourists in the Havana world heritage site. *Sustainability*, *13*(19), 11015.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., et al. (2019). BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *CoRR*, *abs/1910.13461*. https://doi.org/10.48550/arXiv.1910.13461

Li, H., Yu, B. X., Li, G., & Gao, H. (2023). Restaurant survival prediction using customer generated content: An aspect-based sentiment analysis of online reviews. *Tourism Management*, *96*, 104707.

Liu, B. (2010). Sentiment analysis and subjectivity. In *Handbook of natural language processing* (2nd edition) (pp. 627-666). Taylor and Francis Group.

Liu, B. (2015). Preface. In *Sentiment analysis* (pp. xi–xiv). Cambridge: Cambridge University Press.

Luo, Z., Huang, S., & Zhu, K. Q. (2019). Knowledge empowered prominent aspect extraction from product reviews. *Information Processing & Management*, *56*(3), 408-423.

Ma, J., Li, F. S., & Shang, Y. (2022). Tourist scams, moral emotions and behaviors: impacts on moral emotions, dissatisfaction, revisit intention and negative word of mouth. *Tourism Review*, *77*(5), 1299–1321.

Maity, A., Ghosh, S., Karfa, S., Mukhopadhyay, M., Pal, S., & Pramanik, P. K. D. (2020). Sentiment analysis from travelers' reviews using enhanced conjunction rule-based approach for feature-specific evaluation of hotels. *Journal of Statistics and Management Systems*, *23*(6), 983-997.

Mate, M. J., Trupp, A., & Pratt, S. (2019). Managing negative online accommodation reviews: evidence from the Cook Islands. *Journal of Travel & Tourism Marketing*, *36*(5), 627644.

Mehra, P. (2023). Unexpected surprise: Emotion analysis and aspect based sentiment analysis (ABSA) of user generated comments to study behavioral intentions of tourists. *Tourism Management Perspectives*, *45*, 101063.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *CoRR*, *abs/1301.3781*. https://doi.org/10.48550/arXiv.1301.3781

Moreno-Ortiz, A., Salles-Bernal, S., & Orrequia-Barea, A. (2019). Design and validation of annotation schemas for aspect-based sentiment analysis in the tourism sector. *Information Technology & Tourism*, *21*(4), 535–557.

Oh, S., Ji, H., Kim, J., Park, E., & del Pobil, A. P. (2022). Deep learning model based on expectation-confirmation theory to predict customer satisfaction in hospitality service. *Information Technology & Tourism*, *24*(1), 109–126.

Ozen, I. A., & Ozgul Katlav, E. (2023). Aspect-based sentiment analysis on online customer reviews: a case study of technology-supported hotels. *Journal of Hospitality and Tourism Technology, 14*(2), 102–120

Park, H., Lee, M., & Back, K.-J. (2021). Exploring the roles of hotel wellness attributes in customer satisfaction and dissatisfaction: application of Kano model through mixed methods. *International Journal of Contemporary Hospitality Management*, *33*(1), 263– 285.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Empirical methods in natural language processing (EMNLP)* (pp. 1532–1543).

Polyzos, E., Fotiadis, A., & Huan, T.-C. (2024). The asymmetric impact of twitter sentiment and emotions: Impulse response analysis on European tourism firms using micro-data. *Tourism Management, 104*, 104909.

Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, *108*, 42-49.

Prakash, S. L., Perera, P., Newsome, D., Kusuminda, T., & Walker, O. (2019). Reasons for visitor dissatisfaction with wildlife tourism experiences at highly visited national parks in Sri Lanka. *Journal of Outdoor Recreation and Tourism*, *25*, 102-112.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... Liu, P. J. (2019). *Exploring the limits of transfer learning with a unified text-to-text transformer.* arXiv. Retrieved from https://arxiv.org/abs/1910.10683

Rauf, A. A., & Pasha, F. M. (2024). Vlogging gastronomic tourism: understanding global north-south dynamics in YouTube videos and their audiences' feedback. *Tourism Geographies, 26*(3), 407–431.

Rodrigues, H., Brochado, A., & Troilo, M. (2020). Listening to the murmur of water: Essential satisfaction and dissatisfaction attributes of thermal and mineral spas. *Journal of Travel & Tourism Marketing*, *37*(5), 649-661.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53-65.

Sann, R., & Lai, P.-C. (2020). Understanding homophily of service failure within the hotel guest cycle: Applying nlp-aspect-based sentiment analysis to the hospitality industry. *International Journal of Hospitality Management*, *91*, 102678.

Santos, B. N. D., Marcacini, R. M., & Rezende, S. O. (2021). Multi-domain aspect extraction using bidirectional encoder representations from transformers. *IEEE Access*, *9*, 9160491613.

Schweter, S., & Akbik, A. (2020). *Flert: Document-level features for named entity recognition.*

Shahhosseini, M., & Nasr, A. K. (2024). What attributes affect customer satisfaction in green restaurants? an aspect-based sentiment analysis approach. *Journal of Travel & Tourism Marketing, 41*(4), 472–490

Speer, R., Chin, J., & Havasi, C. (2017). ConceptNet 5.5: An open multilingual graph of general knowledge. In Proceedings of the *AAAI Conference on Artificial Intelligence* (Vol. 31, No. 1) (pp. 4444–4451).

Stepaniuk, K., & Sturgulewska, A. (2021). Hitchhiking experiences and perception of affective label polarity in social networking sites—potential memetic implications for digital visual content management. *Sustainability, 13*(1), 223.

Taheri, B., Olya, H., Ali, F., & Gannon, M. J. (2020). Understanding the influence of airport servicescape on traveler dissatisfaction and misbehavior. *Journal of Travel Research*, *59*(6), 1008-1028.

Tang, G., & Zeng, H. (2021). Evaluation of tourism e-commerce user satisfaction. *Journal of Organizational and End User Computing*, *33*(5), 25-41.

Thanyasunthornsakun, K. (2016). An evaluation of cultural heritage tourism destination attributes for delighting visitors: A case study of the Ban Chiang archaeological site. *Pertanika Journal of Social Science and Humanities*, *24*(SI), 95-114.

Um, K.-H., & Kim, S.-M. (2018). Application of fairness theory to medical tourists' dissatisfaction and complaint behaviors: The moderating role of patient participation in medical tourism. *Journal of Social Service Research*, *44*(2), 191-208.

Valdivia, A., Martínez-Cámara, E., Chaturvedi, I., Luzón, M. V., Cambria, E., Ong, Y.-S., & Herrera, F. (2020). What do people think about this monument? understanding negative reviews via deep learning, clustering and descriptive rules. *Journal of Ambient Intelligence and Humanized Computing, 11*(1), 39–52.

Viñan-Ludeña, M.-S. (2019). A systematic literature review on social media analytics and smart tourism. In *Smart tourism as a driver for culture and sustainability* (pp. 357–374). Springer, Cham.

Viñán-Ludeña, M. S., de Campos, L. M., Jacome-Galarza, L. R., & Sinche-Freire, J. (2020). Social media influence: A comprehensive review in general and in tourism domain. In *Smart innovation, systems and technologies* (Vol. 171, pp. 25–35). Springer.

Viñán-Ludeña, M. S., & de Campos, L. M. (2022a). Analyzing tourist data on twitter: A case study in the province of Granada at Spain. *Journal of Hospitality and Tourism Insights*, *5*(2), 435–464.

Viñán-Ludeña, M. S., & de Campos, L. M. (2022b). Discovering a tourism destination with social media data: BERT-based sentiment analysis. *Journal of Hospitality and Tourism Technology*, *13*(5), 907–921.

Viñán-Ludeña, M. S., Mora-J´acome, V., Viñán-Merecí, C. S., & Sánchez-Cevallos, E. (2022). Exploratory data analysis of the tourist profile: Case study in Loja-Ecuador. In A. Abreu, D. Liberato, & J. C. Garcia Ojeda (Eds.), *Advances in tourism, technology and systems* (pp. 351–360). Singapore: Springer Nature Singapore.

Wang, T., Cai, Y., Leung, H.-F., Lau, R. Y., Li, Q., & Min, H. (2014). Product aspect extraction supervised with online domain knowledge. *Knowledge-Based Systems*, *71*, 86-100.

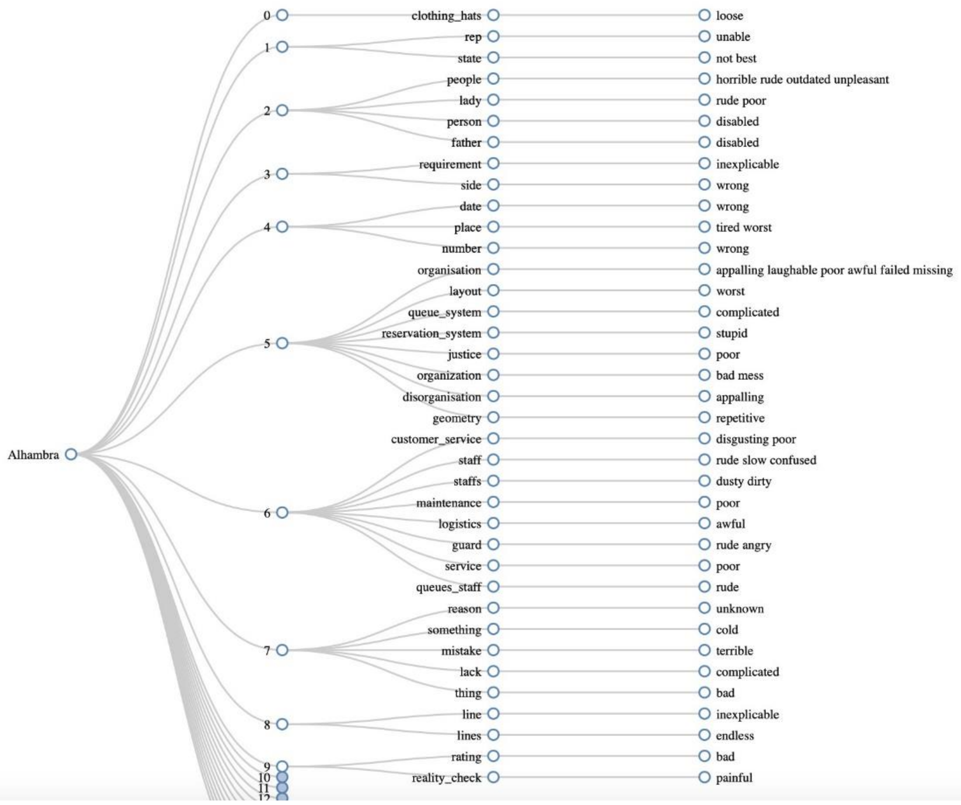
Zeng, G., Chen, Z., & Zhong, S. (2024). “we Chinese just want meat!” an analysis of Chinese netizens’ reactions to vegetarian advocacy. *Food Quality and Preference*, *115*, 105128.

Zhang, F., Seshadri, K., Pattupogula, V. P. D., Badrinath, C., & Liu, S. (2023). Visitors’ satisfaction towards indoor environmental quality in Australian hotels and serviced apartments. *Building and Environment*, *244*, 110819

Zhang, Z., Rao, Y., Lai, H., Wang, J., & Yin, J. (2021). Tadc: A topic-aware dynamic convolutional neural network for aspect extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 1-13.

Appendices

Appendix 1. *Negative perceptions about the Alhambra-part1*



Appendix 2. *Negative perceptions about the Alhambra-part2*