

Elaine Scalabrini*, Jessica Ferreira, Paula Odete Fernandes, Thiago Moraes

From reviews to emotions: Analysing Bragança's tourism attractions on TripAdvisor

<https://doi.org/10.2478/ejthr-2024-0022>

received January 29, 2024; accepted April 17, 2024

Abstract: Over the past decade, sentiment analysis has emerged as a pivotal tool in tourism-related texts, driven by the sheer volume of tourist attractions and the wealth of online information. Tourists increasingly turn to travel websites to access specific information that often eludes standard evaluations of tourist attractions. Forums particularly illuminate specific information needs and their ties to potential destinations. Among these platforms, TripAdvisor has become a favoured choice for posting reviews, ratings, and facilitating online bookings. In this context, this study aims to analyse and assess sentiment in reviews sourced from the online platform TripAdvisor, focusing on tourist attractions in the northern Portuguese destination of Bragança. The research spotlights the disparity between qualitative and quantitative rankings. The study also underscores the importance of data pre-processing, including removing irrelevant information and stop words. Pre-processing was crucial in refining sentiment prediction accuracy, highlighting the differentiated roles of these words in context and meaning. Despite utilising advanced techniques such as tokenisation, TF-IDF weighting, logistic regression, and n-grams, the study's models encountered challenges in achieving high accuracy in sentiment prediction. Even the incorporation of bigrams did not yield substantial performance improvements, with the models frequently inclined to overestimate negative and positive sentiments.

Keywords: qualitative reviews; quantitative ranking; sentimental analysis; TripAdvisor, reviews.

1 Introduction

In the last decade, sentiment analyses have gained prominence in tourism-related texts. The large number of tourist attractions, coupled with a vast amount of information on the web, has made the decision-making process challenging for selection and visitation (Alaei et al., 2019; Álvarez-Carmona et al., 2021).

Travel websites are utilised by tourists to obtain specific information, often not covered in common evaluations of tourist attractions: forums reveal specific information needs and their connection to potential destinations. TripAdvisor is a popular platform for leaving reviews and ratings and making online bookings (Amplayo & Song, 2017; Tsai et al., 2020).

The definition of an attraction on TripAdvisor includes various groups and categories, such as places, tours, operators, events, and activities. Each has a geographical locator and presents different evaluations (Ray et al., 2021; An et al., 2020; Nie et al., 2020; Li et al., 2018; Spalding & Parret, 2019).

In this context, this research aims to analyse and evaluate the sentiment by reviews from the online platform TripAdvisor about the attractions in the tourist destination of Bragança, in the north of Portugal, focusing on the discrepancy in qualitative and quantitative ranking.

In this investigation, the TripAdvisor platform was deliberately selected to acquire data about the top 16 attractions in the northern region of Portugal, specifically in Bragança. Using Apify, the dataset encompassed 270 evaluations from 2020 to 2023, encompassing both ranking assessments and text-based reviews in Portuguese. Employing Python's Scikit Learn Library, a vectorisation process was applied to the evaluations,

*Corresponding author: Elaine Scalabrini, Instituto Politécnico de Bragança, Portugal, Applied Management Research Unit (UNIAG), Email: elaine@ipb.pt

Jessica Ferreira, Instituto Politécnico de Bragança, Portugal, Applied Management Research Unit (UNIAG)

Paula Odete Fernandes, Instituto Politécnico de Bragança, Portugal, Applied Management Research Unit (UNIAG)

Thiago Moraes, State University of Piauí

resulting in a dimensionality of 2,637 within a bag-of-words framework. Logistic regression modelling yielded an accuracy metric of 49%. This value was warranted despite falling short of previous research benchmarks due to the study's exploratory nature constrained by a limited sample size.

The results underscored the importance of distinguishing and understanding the emotions expressed in tourists' comments. Accurately assessing tourist satisfaction is crucial for enhancing the overall tourist experience. However, the current model, while informative, did not provide a reliable method for evaluating tourist satisfaction in the context of Bragança, highlighting the complexity of sentiment analysis in practical applications. Further studies and methodological refinements are needed to increase the accuracy of sentiment prediction in tourist reviews.

This paper is structured into five sections. Following the introduction, the theoretical framework is presented, where concepts of sentiment analysis and related studies on TripAdvisor and the tourist destination are highlighted. Next, the methods employed are introduced, followed by the analysis of results and discussion. Finally, the conclusion is presented, along with considerations and proposals for future research.

2 Literature review

The significance of tourism to the contemporary economy is unequivocal, and forecasts strongly indicate a sustained relevance for the forthcoming decades (Kontogianni & Alepis, 2020). The relationship between the performance of the tourism sector and its broader impact is intricately interwoven, with the former exerting both a direct and an indirect influence on the latter (Mucharreira *et al.*, 2019). The proliferation of information and communication technologies has left an indelible mark on tourism, largely attributed to social media and online travel platforms (OTPs), facilitating the sharing of consumer experiences (Bizirgianni & Dionysopoulou, 2013).

Within the tourism domain, numerous studies underscore the substantial transformation brought about by social media in how tourists seek information, plan their journeys, and, more crucially, share their travel experiences with peers. These diverse applications and OTPs engender a substantial corpus of quantifiable data, empowering destination marketing professionals to leverage these data effectively in decisions related to

promotion and product development (Da Mota & Pickering, 2020; Mirzaalian & Halpenny, 2021).

TripAdvisor (TripAdvisor.com) and Yelp (Yelp.com) are popular OTPs for posting reviews, ratings, and making online reservations. However, these platforms continue to necessitate extensive review reading by consumers to formulate personal opinions regarding establishments of interest (Amplayo & Song, 2017; Tsai *et al.*, 2020). Attributes of online platforms are recognised as pivotal determinants of customer behaviour (Rita *et al.*, 2022). Various attributes, such as the rating system, can directly influence the evaluator's conduct (Mariani & Borghi, 2018).

The efficiency of TripAdvisor as a collaborative recommendation platform fundamentally hinges on various factors: the ease with which a problem can be represented, the extent to which its solution necessitates self-motivated individuals with specific contextual knowledge, dispersed and generalised, and the degree to which its evaluation encompasses a multitude of experienced Internet users. While fake and paid online reviews may adversely affect the efficiency of TripAdvisor (Fileri *et al.*, 2015), the degree of crowd efficiency can also generate antibodies that guard against such opportunistic behaviour (Ganzaroli *et al.*, 2017).

Each OTP features a classification system that aids customers in decision-making (Israeli, 2002). However, OTPs adopt varying scales and classification systems. TripAdvisor employs a 1–5 scale, with additional dimensions that can be evaluated subsequently, with these latter dimensions not considered in the initial assessment. These classification systems significantly influence perceptions of information quality (Chen, 2017), impacting customer behaviour (Mariani & Borghi, 2018; Rita *et al.*, 2022).

Text extraction represents a specific type of data extraction involving textual content analysis to unearth concealed patterns that can be translated into actionable knowledge. The textual content encompasses documents, comments, reviews, or other related information, serving as the corpus for tasks such as text categorisation, clustering, and sentiment analysis (Srivastava & Sahami, 2009).

Data mining challenges can be addressed through supervised machine learning techniques when a target is to be modelled or unsupervised techniques when the objective is to discover relationships among instances of the problem at hand (Sharda *et al.*, 2017). For example, supervised methods can be employed to detect dissatisfied customers to prevent churn, using decision trees (Maier & Prusty, 2016), while categorisation of customer segments based on

their personal preferences regarding hotel websites through association rules represents an example of unsupervised approaches (Leung et al., 2013). However, most data mining approaches in the hospitality domain are centred around predicting tourist demand (Moro & Rita, 2016).

Text extraction, as a distinct form of data extraction, involves the analysis of textual content to reveal concealed patterns that can be translated into actionable knowledge (Fan et al., 2006). The textual content encompasses documents, comments, reviews, or other related information, serving as the corpus for tasks such as text categorisation, clustering, and sentiment analysis (Srivastava & Sahami, 2009).

Through the analysis and classification of sentiments, managers gain deeper insights into which attributes of the services offered in their establishments can significantly impact customer satisfaction (Calheiros et al., 2017).

The advent of social media on the Internet has ignited sentiment analysis, a recently developed data extraction technique capable of conducting sentiment or opinion analyses based on online reviews and comments. Its objective is to extract text from written comments about specific products or services, categorising them into positive or negative opinions based on comment polarity (Cambria et al., 2013; Casalo et al., 2015).

Several studies employ the Apify program as software for downloading comments. Apify is the platform where developers create, implement, and monitor web scraping and browser automation tools. Essentially, anything done manually in a web browser can be automated at scale with Apify (Soh, 2022; Jahanbin & Chahooki, 2023). For the analysis and stratification of the content under review, the program allows for inputting destination links, organised by various analysis categories such as Hospitality, Dining, Attractions, and Transportation.

Recent articles highlight the hotel sector as one of the most studied and assessed categories by TripAdvisor, although other variables can also be analysed and emphasised (Ray et al., 2021; An et al., 2020; Nie et al., 2020; Li et al., 2018; Rita et al., 2022), such as attractions (Scalabrini et al., 2023).

The definition of an attraction on TripAdvisor encompasses various groups and categories, including places, tours, operators, events, and activities. Each category has a geographic locator (Spalding & Parret, 2019) and has different evaluations.

Based on word frequency, prominent keywords indicative of prevailing trends include sentiment analysis, online reviews, social media, deep learning, star ratings, online travel, review utility, environmental discourse, hotel reviews, and the tourism industry, all manifesting

within the dataset (Da Mota & Pickering, 2020; Mirzaalian & Halpenny, 2021; Álvarez-Carmona et al., 2021; Tsai et al. 2020; Gour et al., 2021).

3 Methodology

To accomplish the study's aim, the TripAdvisor platform was chosen, a widely recognised platform in tourism research. TripAdvisor is considered a leading global platform for evaluating tourist experiences, as noted by various studies (Fitchett & Hoogendoorn, 2019; Taecharungroj & Mathayomchan, 2019; Nowacki & Niezgod, 2020; Scalabrini et al., 2023).

Using Apify (Nallakaruppan et al., 2023), the data were collected in September 2023 from the top 16 attractions in the northern region of Portugal, specifically in city of Bragança. The dataset consisted of 270 reviews from 2020 to 2023, including the rankings and the text-based reviews, written in Portuguese, reflecting visitors' experiences. Following previous studies (e.g., Scalabrini et al., 2023), the ranking was classified as negative (1–2), neutral (3), and positive (4–5). Following this, the database was recoded based on the sentiment classification (0=negative, 1=neutral, and 2=positive).

In the next step, the Scikit Learn Library from Python language was employed to vectorise the reviews (Bagherzadeh et al., 2021; Chen et al., 2020; Garreta et al., 2017), resulting in a bag of words, with 2.637 dimensions. Next, the database was partitioned into training and test data and accuracy was evaluated using logistic regression, yielding a score of 0.4920 (49%). Despite this accuracy being lower than that of other studies, the study moved forward as it is a preliminary study with a limited sample size. The model accuracy was improved by implementing the Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L_BFGS) method to approximate the 50 most significant words in terms of their meaning, resulting in a predictive accuracy level of 49%.

To eliminate redundancy and establish stop words, the first word cloud was generated based on the frequency of words, without any prior treatment. For sentiment analysis, grammatical terms such as “from,” “and,” “a/an,” and “that” were ignored or removed since they carry less semantic and emotional information compared to the most relevant words (Ahuja et al., 2019; Jianqiang & Xiaolin, 2017). The words were also mapped to their base form, and to analyse word sequences, n-gram was applied. After this, punctuation was removed. Additionally, to eliminate words with similar semantics and radicals, the NLTK with the suffix RSLPStemmer was utilised (Yao, 2019). Following these steps, the results are presented in the next item.

4 Results

This study conducted a sentiment analysis of the TripAdvisor reviews for tourist attractions in Bragança, located in the Terras Trás-os-Montes region of Portugal. The reviews were categorised based on the ranking system used by TripAdvisor (1–5 points), compared with the algorithm analysis of the reviews. Considering similar studies (e.g., Scalabrini *et al.*, 2023), the rankings were classified as negative (1–2 points), neutral (=3 points), or positive (4–5 points) and, as previously mentioned, were recoded based on the sentiment classification (0=negative, 1=neutral, and 2=positive). Table 1 presents the number of reviews recoded in each sentiment.

After analysing the sentiment analysis algorithm based on the texts, discrepancies were found between the quantitative and qualitative evaluations (Table 2). This observation has been reported in previous studies (e.g., Valdivia *et al.*, 2019; Xiang *et al.*, 2018). The frequencies of negative, neutral, and positive evaluations varied significantly across different attractions. For instance, Bragança Castle had 33 negative evaluations, while the Georges Dussaud Gallery had none. On the other hand, Aldeia do Rio de Onor had 35 positive reviews, indicating a favourable reception, which was confirmed by its high mean rating, the highest among all the attractions.

Table 2 reveals a wide range in the quantitative aspects of Bragança's tourist attractions' evaluations, indicating

Table 1. Sentiments' classification.

Rating (1–5)	Sentiment	<i>n</i>
1–2	Negative	114
=3	Neutral	24
4–5	Positive	112
	Total	250

Source: Own elaboration.

that the evaluations are spread out around the mean. The data highlight the variety of perceptions on the qualitative evaluations found on the TripAdvisor platform. Notably, Aldeia do Rio de Onor stands out with consistently positive qualitative and quantitative evaluations. Conversely, Bragança Castle has the highest frequency of negative evaluations proportionally, despite having the second-highest mean rating.

When collecting data through web scraping, irrelevant information related to formatting, links, and special characters can be included. Therefore, it is necessary to pre-process the text to remove unwanted elements and standardise formatting. The initial word cloud in Figure 1 was generated without pre-processing, and after analysing it, stop words were removed. Accessing and tokenising the speech content with code was necessary, and the model's prediction accuracy was monitored. However, after this treatment, the accuracy dropped to 47%, demonstrating the importance of the terms extracted at this stage in

Table 2. Reviews classification per attraction.

Attraction	Negative	Neutral	Positive	\bar{x}	<i>s</i>
Castelo de Bragança	33	4	32	4.23	0.96
Aldeia do Rio de Onor	18	4	35	4.42	0.86
Domus Municipalis	11	1	3	3.73	1.03
Cidadela de Bragança	5	2	3	3.88	0.70
Igreja de Santa Maria	4	3	0	3.81	1.17
Centro de Arte Contemporânea Graça Morais	3	0	4	3.92	0.53
Igreja da Antiga Sé	4	3	0	3.87	0.53
Museu Ibérico da Máscara e do Traje	7	2	7	3.95	0.77
Museu do Abade de Baçal	3	1	3	3.96	1.15
Parque Natural de Montesinho	5	0	3	4.00	0.51
Centro Ciência Viva	2	1	6	4.02	1.39
Centro de Interpretação da Cultura Sefardita do Nordeste Transmontano	9	0	1	4.03	0.57
Galeria Georges Dussaud	0	0	3	4.05	0.00
Museu Militar de Bragança	9	3	9	4.04	1.20

Note: \bar{x} = mean; *s* = standard deviation.

Source: Own elaboration.



Figure 1: Word cloud in the pre-processing.

terms of context and meaning. Considering the data were collected in Portuguese, the words are presented in this language, following the method proposed.

The next step in the data processing was tokenisation, which involved calculating the frequency of each word in the bag of words as if it were an element in a list. A data frame was then created to display the frequency of the tokenised terms. Table 3 illustrates the top 10 most frequent terms found in the database.

It is important to consider that the removal of stop words, such as prepositions and conjunctions, can alter the meaning of sentences and hinder the understanding of word relationships. This is significant since it has an effect on context of the text, which is crucial for constructing meaning. Furthermore, certain stop words may hold valuable information, particularly when used in a unique way, such as negating a positive or negative sentiment with a single word. Removing such words can result in the loss of this information and negatively affect the accuracy of the model.

Table 3. Stop words’ frequencies.

Word	n
De	466
E	361
A	305
Que	184
O	169
É	163
Um	148
Uma	147
Muito	145
Com	135

Source: Own elaboration.

Upon removing stop words, a Pareto chart was generated (refer to Figure 2), which revealed that the most commonly used terms were related to places like Castle, Museum, Village, City, and Bragança. It is worth noting, however, that the accuracy of the model decreased to 49.2% after this step, as the removal of stop words not only decreases the sample size but also affects the model’s ability to interpret sentiment expressions.

Following these procedures, three word clouds were produced, each representing a sentiment category in the database (refer to Figure 3).

Upon removal of the score, the accuracy was once again assessed. The model was able to attain a 49.2% accuracy rate in predicting sentiment at this stage. To showcase the word associations for each sentiment classification category in the database, three word clouds were generated. The resulting word clouds, following all processing steps executed thus far, are depicted in Figure 4.

The values calculated by applying the term frequency–inverse document frequency (TF-IDF) formula to both the training data and test data were extracted. After this, logistic regression was trained using the test data as parameters to obtain the accuracy result directly. The logistic regression showed an accuracy of 50%.

Next, n-grams were applied to the treated database, with two parameters: the first considered terms in lowercase letters, and the second parameterised n-gram as a bigram. The n-grams are resources created and frequently used as discriminative ways to classify texts and help to incorporate unknown words into the context, thus improving classification performance in tasks based on informal texts, where a large percentage of unknown words occur, such in sentiment analysis (Bojanowski et al., 2017).

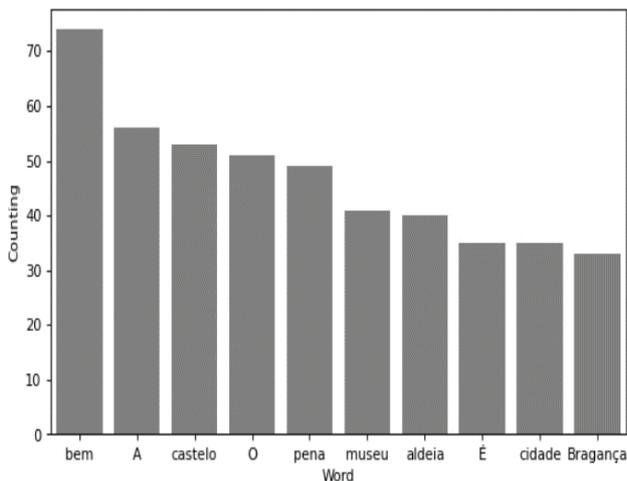


Figure 2: Pareto chart.



(a) Negative

(b) Positive



(c) Neutral

Figure 3: Figure with three subgraphs with word cloud after processing.



(a) Negative

(b) Positive



(c) Neutral

Figure 4: Figure with three subgraphs with word cloud after stemming processing.

This technique is applied to improve the accuracy of the model by capturing the context of the speeches impacted by the removal of stop words, preserving contextual information.

However, the model did not show any improvement in accuracy after applying n-grams. N-grams help to better capture the context in which the words are used, which diverts the analysis from focusing solely on the rating. It is an essential technique for determining whether an evaluation is negative, neutral, or positive. This is because, in many cases, the meaning of a word depends on the words around it. The combination of n-grams and capital letters helps to distinguish more effectively between negative and positive feelings. Capitalised terms may represent an intensification of what is intended to be expressed, making it possible to better capture linguistic and contextual nuances.

Logistic regression was applied because it allows us to analyse which words have the greatest weight in differentiating between classes and what the algorithm understands as negative, neutral, or positive words. To do this, a data frame was created in which the variables were stored with their respective sentiment classification coefficients. Table 4 shows the respective terms and their weights.

The TF-IDF weighting matrix is a powerful tool in natural language processing that helps determine the sentiment of a text. By calculating the frequency of words in a corpus and comparing it to the total number of words in the same corpus, this technique identifies words that are more important for classification. This improves the precision, retrieval, and F1-score of statistical approaches used in sentiment analysis.

Despite initially using the Multinomial Logistic Regression model, low accuracy was observed. As a result, the Multinomial Naïve Bayes classifier was employed to

re-evaluate the model for text classification. Unfortunately, this did not yield any improvement, with the model accuracy remaining at 48%. Even when a bigram prediction was performed, the model accuracy failed to improve. The accompanying metrics report is detailed in Table 5.

The confusion matrix is a useful tool for analysing a model’s accuracy. It allows data mining to measure the performance of terms based on their classification results. This method identifies the number of correctly detected negatives and positives. Additionally, it enables the calculation of accuracy, precision, and recall, which are shown in Table 6 for both single-frame and large-frame precision models. Rachman et al. (2020), Sari et al. (2023), and Steven and Wella (2020) have all highlighted the effectiveness of the confusion matrix in evaluating model performance.

The confusion matrix is a widely used tool to assess the effectiveness of a classification model. In this scenario, it is used to evaluate a model that classifies tourist reviews. The predicted class is the model’s prediction, while the actual rating is the classification given by the review itself. The negative, neutral, and positive categories are the three classes that review can be classified into, and the total of predictions and actual ratings in each category is calculated.

From this matrix, several performance metrics such as precision, F1-score, and recall can be calculated to determine how well the model is performing. It is also possible to identify some specific features. For instance, the model’s second proposal, the bigram, was used to improve the evaluation of the model. However, the model classified 46 reviews as negative, but only 8 of them were negative. This suggests that the model tends to overestimate negative reviews. The model classified 14 reviews as neutral, of which 4 were actually neutral, indicating

Table 4. Terms and their weights.

Negative		Neutral		Positive	
Algum	0.63	Fech	0.50	portug	0.48
Atra	0.50	Prox	0.47	lind	0.46
regia	0.46	Jardim	0.45	espanh	0.45
nad	0.45	Extern	0.35	otim	0.43
dent	0.43	Enriquec	0.31	bom	0.43
igreja	0.42	Muit	0.31	pezzo	0.40
portugu	0.42	Espac	0.29	bem	0.40
esper	0.40	Atenc	0.29	val	0.39
eur	0.39	Garant	0.28	lindiss	0.38
monument	0.38	Indicaca	0.28	alem	0.38

Source: Own elaboration.

Table 5. Metrics report.

Unigram Prediction Model				
	Precision	Recall	F1-score	Support
Negative	0.48	0.50	0.49	114
Neutral	0.00	0.00	0.00	24
Positive	0.52	0.58	0.55	112
Accuracy			0.49	250
Macro average	0.33	0.36	0.35	250
Weighted average	0.45	0.49	0.47	250
Bigrated Prediction Model				
Negative	0.49	0.40	0.44	114
Neutral	0.18	0.17	0.17	24
Positive	0.51	0.61	0.55	112
Accuracy			0.47	250
Macro average	0.39	0.39	0.39	250
Weighted average	0.47	0.47	0.47	250

Source: Own elaboration.

Table 6. Calculation of accuracy, precision, and recall.

Unigram Prediction Model				
Predicted	Negative	Neutral	Positive	All
Real				
Negative	57	3	54	114
Neutral	17	0	7	24
Positive	44	3	65	112
All	118	0	126	250
Bigrated Prediction Model				
Predicted	Negative	Neutral	Positive	All
Real				
Negative	46	8	60	114
Neutral	14	4	6	24
Positive	34	10	68	112
All	94	22	134	250

Source: Own elaboration.

that the model performed reasonably well. As for positive ratings, the model defined 34 reviews as positive, even if only 10 of them were positive, showing that the model also tends to overestimate positive reviews. These findings suggest the need to adjust the parameters or re-evaluate the quality of the training data.

Despite the model's attempt to improve by bigramming the terms, its accuracy did not improve significantly, and in some classifications, it even worsened, as shown in Table 6. For example, the bigram model increased the number of positive classifications from 54 to 60 when they should have been negative. However, the same did not happen the other way round—the bigram model reduced

its negative classifications from 44 to 34 when they should have been positive.

In summary, the results in this study highlight the importance of differentiating between categorised emotions expressed in the comments of tourists. Detecting emotions related to the emotional states of tourists is crucial in evaluating their satisfaction with their experience. This can be done through two approaches—qualitative analysis through the categorisation of the texts and quantitative analysis through ratings. However, the results of the current model do not provide a reliable way of meeting the challenge of determining tourist satisfaction with their experience in the city of Bragança in this sample.

5 Discussion

The relevance of tourism to the contemporary economy is unequivocal, with forecasts indicating sustained relevance over the coming decades (Kontogianni & Alepis, 2020). The relationship between the performance of the tourism sector and its wider impact is intricately intertwined, with the former exerting a direct influence and the latter exerting an indirect influence (Mucharreira et al., 2019).

Within the tourism domain, several studies highlight the substantial transformation brought about by social media in the way tourists search for information, plan their trips, and, most crucially, share their travel experiences with their peers.

These various applications and OTPs generate a substantial corpus of quantifiable data, empowering destination marketing professionals to leverage these data effectively in decisions related to promotion and product development (Da Mota & Pickering, 2020; Mirzaalian & Halpenny, 2021).

The findings of this study reveal notable discrepancies between quantitative evaluations (based on numerical ratings) and qualitative evaluations (based on sentiment analysis of texts). This is consistent with existing literature, which highlights the complexity of measuring tourist satisfaction exclusively through numerical ratings (Valdivia et al., 2019; Xiang et al., 2018). For instance, while Bragança Castle received a high frequency of negative evaluations, Aldeia do Rio de Onor stood out with predominantly positive reviews, reflecting a favourable reception.

This variation in evaluations underscores the importance of utilising both qualitative and quantitative analyses to obtain a more comprehensive understanding of tourist satisfaction. Previous studies indicate that platforms such as TripAdvisor are essential for collecting data on tourist experiences but also point to the necessity of considering the nuances of textual reviews to truly capture tourists' perceptions (Fileri et al., 2015; Ganzaroli et al., 2017).

6 Conclusions

This study analysed and evaluated the sentiments in the TripAdvisor reviews, where discrepancies occur between the text reviews (qualitative) and the ranking (quantitative) (Valdivia et al., 2019; Rita et al., 2022). Text extraction and data extraction techniques, both supervised and unsupervised, play a pivotal role in analysing the textual

content of these platforms to extract valuable insights (Srivastava & Sahami, 2009; Sharda et al., 2017; Leung et al., 2013; Calheiros et al., 2017).

In conclusion, a comprehensive analysis of sentiment was conducted in TripAdvisor reviews of tourist attractions in Bragança, Portugal, using quantitative and qualitative approaches. The reviews were categorised based on the TripAdvisor rating system, and the sentiment was algorithmically classified as negative, neutral, or positive.

Discrepancies between the two approaches were observed, reflecting findings from previous research. Sentiment distribution varied significantly among different attractions, with Aldeia do Rio de Onor consistently receiving positive reviews. At the same time, Bragança Castle had a disproportionately high frequency of negative feedback despite having a relatively high average rating.

The study also underscored the importance of data pre-processing, including removing irrelevant information and stop words. Pre-processing was crucial in refining sentiment prediction accuracy, highlighting these words' differentiated roles in context and meaning.

Despite employing techniques such as tokenisation, TF-IDF weighting, logistic regression, and n-grams, the study's models struggled to achieve high accuracy in sentiment prediction. Even the introduction of bigrams did not substantially improve performance, with the models often overestimating negative and positive sentiments.

Analysis of the confusion matrix revealed additional insights, showing the model's tendency to overestimate particular sentiments, particularly negative and positive ratings. This suggests the need to adjust parameters or better assess the quality of training data.

Therefore, it should be noted that people express their sentiments more openly through discourses, and sentiment analysis becomes an essential tool to monitor and understand these sentiments. However, it should not be viewed in isolation and exclusively as an automated process. Sentiment analysis can address critical issues compared to ratings, as evidenced in this study (Valdivia et al., 2019). It can help identify emerging trends, analyse competitors, and explore new markets (Ramos, 2022; Zhu et al., 2021).

The amount of data collected on the web is insufficient, and the pre-processing is not fully covered, which may impact the predicted results. Additionally, not all multi-label classifiers and algorithms were tested using the study's model. Future studies must build data banks to extract data in real time, collect more data, and improve the current model.

Acknowledgments

This work was supported by national funds through FCT/MCTES (PIDDAC): UNIAG, UIDB/004752/2020 (DOI 10.54499/UIDB/04752/2020) and UIDP/04752/2020 (DOI 10.54499/UIDP/04752/2020).

Bionotes

Elaine Cristina Borges Scalabrini has a PhD in Geography (2017) from the University of Minho, Portugal. She is a researcher and an integrated member of UNIAG (Applied Management Research Unit). She is also a university professor in Strategic Management and Marketing and Economy at the Instituto Politécnico de Bragança. She has works in the area of Tourism with a focus on residents' perceptions, Tourism Impacts, Destination Image, and Regional Development. ORCID ID: 0000-0001-7164-2471

Jessica Patrícia Miranda Ferreira has a PhD in Tourism (2022) from the University of Lisbon - IGOT. She is a PhD researcher and an integrated member of UNIAG (Applied Management Research Unit). She is also a university professor in Strategic Management and Marketing (in Portuguese and English) at the Instituto Politécnico de Bragança. She has works in the area of Tourism with a focus on Marketing, Strategic Management, and Regional Development. ORCID ID: 0000-0002-4141-6702

Paula Odete Fernandes has a PhD in Economics and Management (2005) and is a Professor of Management at the Instituto Politécnico de Bragança, Portugal (12/1992). She is a researcher and Scientific Coordinator of UNIAG (Applied Management Research Unit, since 2013). She participated in 18 international/national R&D projects (7 as Responsible Researcher) and supervised two PhD theses and several master students (over 140 master theses). She published more than 225 publications (96 with indexing to Scopus/WoS databases) in peer-reviewed scientific documents. She is a recipient of the best award scientific paper (10) in Marketing and Strategy, Tourism, and Regional Analysis. She has been head of the Management Master since 2009. ORCID ID: 000-0001-8714-4901

Thiago Assunção de Moraes has a PhD in Administration (2018) from the Federal University of Paraíba. He is Associate Professor at the State University of Piauí and a Collaborating Professor in the Master's Program in Public Management at the Federal University of Piauí. He has works in the area of Data Analysis at the Vortex Institute. He leads the research group ESCADA (Studies in Science and Data Analysis in Administration). ORCID ID: 0000-0001-9729-4858

References

- [1] Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). The Impact of Features Extraction on the Sentiment Analysis. *Procedia Computer Science*, 152(2019), 341–348. <https://doi.org/10.1016/j.procs.2019.05.008>
- [2] Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment Analysis in Tourism: Capitalizing on big data. *Journal of Travel Research*, 58(2), 175–191. <https://doi.org/10.1177/0047287517747753>
- [3] Álvarez-Carmona, M. Á., Aranda, R., Arce-Cardenas, S., Fajardo-Delgado, D., Guerrero-Rodríguez, R., López-Monroy, A. P., Martínez-Miranda, J., Pérez-Espinosa, H., & Rodríguez-González, A. Y. (2021). Overview of Rest-Mex at Eberle 2021: Recommendation System for Text Mexican Tourism. *Sociedad Española para el Procesamiento del Lenguaje Natural*, 67(2021), 163–172. <https://doi.org/10.26342/2021-67-14>
- [4] Amplayo, R. K., & Song, M. (2017). An adaptable fine-grained sentiment analysis for summarization of multiple short online reviews. *Data & Knowledge Engineering*, 110(2017), 54–67. <https://doi.org/10.1016/j.datak.2017.03.009>
- [5] An, Q., Ma, Y., Du, Q., Xiang, Z., & Fan, W. (2020). Role of user-generated photos in online hotel reviews: An analytical approach. *Journal of Hospitality and Tourism Management*, 45(2020), 633–640. <https://doi.org/10.1016/j.jhtm.2020.11.002>
- [6] Bagherzadeh, S., Shokouhyar, S., Jahani, H., & Sigala, M. (2021). A generalizable sentiment analysis method for creating a hotel dictionary: Using big data on TripAdvisor hotel reviews. *Journal of Hospitality and Tourism Technology*, 12(2), 210–238. <https://doi.org/10.1108/JHTT-02-2020-0034>
- [7] Bizirgianni, I., & Dionysopoulou, P. (2013). The Influence of Tourist Trends of Youth Tourism through Social Media (SM) & Information and Communication Technologies (ICTs). *Procedia-Social and Behavioral Sciences*, 73(2013), 652–660. <https://doi.org/10.1016/j.sbspro.2013.02.102>
- [8] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the association for computational linguistics*, 5(2017), 135–146. https://doi.org/10.1162/tacl_a_00051
- [9] Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment Classification of Consumer-Generated Online Reviews Using Topic Modeling. *Journal of Hospitality Marketing & Management*, 26(7),

- 675–693. <https://doi.org/10.1080/19368623.2017.1310075>
- [10] Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intelligent Systems*, 28(2), 15–21. <http://doi.org/10.1109/MIS.2013.30>
- [11] Casalo, L. V., Flavián, C., Guinalú, M., & Ekinci, Y. (2015). Avoiding the dark side of positive online consumer reviews: Enhancing reviews' usefulness for high risk-averse travelers. *Journal of Business Research*, 68(9), 1829–1835. <http://doi.org/10.1016/j.jbusres.2015.01.010>
- [12] Chen, C.-W. (2017). Five-star or thumbs-up? The influence of rating system types on users' perceptions of information quality, cognitive effort, enjoyment and continuance intention. *Internet Research*, 27(3), 478–494. <https://doi.org/10.1108/IntR-08-2016-0243>
- [13] Chen, W., Xu, Z., Zheng, X., Yu, Q., & Luo, Y. (2020). Research on sentiment classification of online travel review text. *Applied Sciences*, 10(15), 1–21. <https://doi.org/10.3390/app1015275>
- [14] Da Mota, V. T., & Pickering, C. (2020). Using social media to assess nature-based tourism: Current research and future trends. *Journal of Outdoor Recreation and Tourism*, 30(2020), 1–11. <https://doi.org/10.1016/j.jort.2020.100295>
- [15] Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, 49(9), 76–82. <https://doi.org/10.1145/1151030.1151032>
- [16] Filieri, R., Alguezaui, S., & McLeay, F. (2015). Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management*, 51(2015), 174–185. <http://dx.doi.org/10.1016/j.tourman.2015.05.007>
- [17] Fitchett, J. M., & Hoogendoorn, G. (2019). Exploring the climate sensitivity of tourists to South Africa through TripAdvisor reviews. *South African Geographical Journal*, 101(1), 91–109. <https://doi.org/10.1080/03736245.2018.1541022>
- [18] Ganzaroli, A., De Noni, I., & van Baalen, P. (2017). Vicious advice: Analyzing the impact of TripAdvisor on the quality of restaurants as part of the cultural heritage of Venice. *Tourism Management*, 61(2017), 501–510. <http://dx.doi.org/10.1016/j.tourman.2017.03.019>
- [19] Garreta, R., Moncecchi, G., Hauck, T., & Hackeling, G. (2017). *scikit-learn: Machine Learning Simplified: Implement scikit-learn into every step of the data science pipeline*. Packt Publishing
- [20] Gour, A., Aggarwal, S., & Erdem, M. (2021). Reading between the lines: analyzing online reviews by using a multi-method Web-analytics approach. *International Journal of Contemporary Hospitality Management*, 33(2), 490–512. <https://doi.org/10.1108/IJCHM-07-2020-0760>
- [21] Israeli, A. A. (2002). Star rating and corporate affiliation: their influence on room price and performance of hotels in Israel. *International Journal of Hospitality Management*, 21(4), 405–424. [https://doi.org/10.1016/S0278-4319\(02\)00037-3](https://doi.org/10.1016/S0278-4319(02)00037-3)
- [22] Jahanbin, K., & Chahooki, M. A. Z. (2023). *Database of Twitter Influencers in Cryptocurrency (2021-2023) with Sentiments*. <https://doi.org/10.21203/rs.3.rs-3192598/v1>
- [23] Jianqiang, Z., & Xiaolin, G. (2017). Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis. *IEEE Access*, 5(2017), 2870–2879. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7862202>
- [24] Kontogianni, A., & Alepis, E. (2020). Smart tourism: State of the art and literature review for the last six years. *Array*, 6(2020), 1–14. <https://doi.org/10.1016/j.array.2020.100020>
- [25] Leung, R., Rong, J., Li, G., & Law, R. (2013). Personality Differences and Hotel Web Design Study Using Targeted Positive and Negative Association Rule Mining. *Journal of Hospitality Marketing & Management*, 22(7), 701–727. <http://doi.org/10.1080/19368623.2013.723995>
- [26] Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68(2018), 301–323. <https://doi.org/10.1016/j.tourman.2018.03.009>
- [27] Maier, T. A., & Prusty, S. (2016). Managing Customer Retention in Private Clubs Using Churn Analysis: Some Empirical Findings. *Journal of Hospitality Marketing & Management*, 25(7), 797–819. <http://doi.org/10.1080/19368623.2016.1113904>
- [28] Mariani, M. M., & Borghi, M. (2018). Effects of the Booking.com rating system: Bringing hotel class into the picture. *Tourism Management*, 66(2018), 47–52. <https://doi.org/10.1016/j.tourman.2017.11.006>
- [29] Mirzaalian, F., & Halpenny, E. (2021). Exploring destination loyalty: Application of social media analytics in a nature-based tourism setting. *Journal of Destination Marketing & Management*, 20(2021), 1–15. <https://doi.org/10.1016/j.jdmm.2021.100598>
- [30] Moro, S., & Rita, P. (2016). Forecasting tomorrow's tourist. *Worldwide Hospitality and Tourism*

- Themes*, 8(6), 643–653. <http://doi.org/10.1108/WHATT-09-2016-0046>
- [31] Mucharreira, P. R., Antunes, M. G., Abranja, N., Justino, M. R. T., & Quirós, J. T. (2019). The relevance of tourism in financial sustainability of hotels. *European Research on Management and Business Economics*, 25(3), 165–174. <https://doi.org/10.1016/j.iiedeen.2019.07.002>
- [32] Nallakaruppan, M. K., Srivastava, G., Gadekallu, T. R., Reddy, P. K., Krishnan, S., & Polap, D. (2023). *Child Tracking and Prediction of Violence on Children In Social Media Using Natural Language Processing and Machine Learning*. In L. Rutkowski, R. Scherer, M. Korytkowski, W. Pedrycz, R. Tadeusiewicz, & J. M. Zurada (Eds.), *Artificial Intelligence and Soft Computing* (pp. 560–569). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-42505-9_47
- [33] Nie, R. -X., Tian, Z. -P., Wang, J. -Q., & Chin, K. S. (2020). Hotel selection driven by online textual reviews: Applying a semantic partitioned sentiment dictionary and evidence theory. *International Journal of Hospitality Management*, 88(2020), 1–16. <https://doi.org/10.1016/j.ijhm.2020.102495>
- [34] Nowacki, M., & Niezgodna, A. (2020). Identifying unique features of the image of selected cities based on reviews by TripAdvisor portal users. *Scandinavian Journal of Hospitality and Tourism*, 20(5), 503–519. <https://doi.org/10.1080/15022250.2020.1833362>
- [35] Rachman, F. F., & Pramana, S. (2020). Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang vaksin COVID-19 pada Media Sosial Twitter [Sentiment analysis of the pros and cons of the Indonesian people about the COVID-19 vaccine on social media Twitter]. *Indonesian of Health Information Management Journal*, 8(2), 100–109.
- [36] Ramos, C. M. Q. (2022). Business Intelligence Approach and Sentiment Analysis as a Management Strategy Applied to Study Customer Satisfaction in the Hospitality Sector. In J. V. D. Carvalho, P. Liberato, & A. Peña (Eds.), *Advances in Tourism, Technology and Systems* (pp. 537–547). Springer Nature Singapore. https://doi.org/10.1007/978-981-16-9701-2_44
- [37] Ray, B., Garain, A., & Sarkar, R. (2021). An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews. *Applied Soft Computing*, 98(2021), 1–18. <https://doi.org/10.1016/j.asoc.2020.106935>
- [38] Rita, P., Ramos, R., Borges-Tiago, M. T., & Rodrigues, D. (2022). Impact of the rating system on sentiment and tone of voice: A Booking.com and TripAdvisor comparison study. *International Journal of Hospitality Management*, 104(2022), 1–12. <https://doi.org/10.1016/j.ijhm.2022.103245>
- [39] Sari, A. W., Hermanto, T. I., & Defriani, M. (2023). Sentiment Analysis Of Tourist Reviews Using K-Nearest Neighbors Algorithm And Support Vector Machine. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, 7(3), 1366–1378. <https://doi.org/10.33395/sinkron.v8i3.12447>
- [40] Scalabrini, E., Ferreira, J., Zekan, S. B., & Fernandes, P. O. (2023). Destination Image Through TripAdvisor's Reviews Analysis. In N. Martins & D. Raposo (Eds.), *Communication Design and Branding: A Multidisciplinary Approach* (pp. 205–221). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35385-7_12
- [41] Sharda, R., Delen, D., & Turban, E. (2017). *Business intelligence, analytics, and data science: a managerial perspective*. Pearson
- [42] Soh, C. W. (2022). *Sentiment Analysis and Fake News on social media* [Doctoral dissertation, Tunku Abdul Rahman University College]. TAR UMT Institutional Repository. <https://eprints.tarc.edu.my/id/eprint/22468>
- [43] Spalding, M., & Parrett, C. L. (2019). Global patterns in mangrove recreation and tourism. *Marine Policy*, 110(2019), 1–8. <https://doi.org/10.1016/j.marpol.2019.103540>
- [44] Srivastava, A. N., & Sahami, M. (2009). *Text Mining: Classification, Clustering, and Applications*. CRC Press
- [45] Steven, C., & Wella, W. (2020). The Right Sentiment Analysis Method of Indonesian Tourism in Social Media Twitter. *International Journal of New Media Technology*, 7(2), 102–110. <https://doi.org/10.31937/ijnmt.v7i2.1732>
- [46] Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*, 75(2019), 550–568. <https://doi.org/10.1016/j.tourman.2019.06.020>
- [47] Tsai, C. -F., Chen, K., Hu, Y. -H., & Chen, W. -K. (2020). Improving text summarization of online hotel reviews with review helpfulness and sentiment. *Tourism Management*, 80(2020), 1–13. <https://doi.org/10.1016/j.tourman.2020.104122>

- [48] Valdivia, A., Hrabova, E., Chaturvedi, I., Luzón, M. V., Troiano, L., Cambria, E., & Herrera, F. (2019). Inconsistencies on TripAdvisor reviews: A unified index between users and Sentiment Analysis Methods. *Neurocomputing*, 353(2019), 3–16. <https://doi.org/10.1016/j.neucom.2018.09.096>
- [49] Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2018). Assessing reliability of social media data: Lessons from mining TripAdvisor hotel reviews. *Information Technology & Tourism*, 18(1), 43–59.
- [50] Yao, J. (2019). Automated Sentiment Analysis of Text Data with NLTK. *Journal of Physics: Conference Series*, 1187(5), 052020. <https://iopscience.iop.org/article/10.1088/1742-6596/1187/5/052020/meta>
- [51] Zhu, J. J., Chang, Y.-C., Ku, C.-H., Li, S. Y., & Chen, C.-J. (2021). Online critical review classification in response strategy and service provider rating: Algorithms from heuristic processing, sentiment analysis to deep learning. *Journal of Business Research*, 129(2021), 860–877. <https://doi.org/10.1016/j.jbusres.2020.11.007>