invtur2024
conference

Proceedings of the INVTUR Conference 2024
8-10 May 2024, University of Aveiro, Portugal

Pages
623 –625

Sentiments on the Map: A Comparative Analysis of Online Tourist Reviews in Bragança's Attractions

THIAGO MORAES[1], JESSICA FERREIRA[2], ELAINE SCALABRINI[2] & PAULA ODETE FERNANDES[2]

[1]UESPI, Universidade Estadual do Piauí, [2]UNIAG, Instituto Politécnico de Bragança
Contacting author: elaine@ipb.pt

Keywords | Sentiment analysis of tourist, Tourist experience, Tourism online reviews, TripAdvisor, Google Maps

Objectives | In the preceding decade, sentiment analysis has arisen as an integral instrument in texts about tourism. Tourists are progressively resorting to travel websites to acquire precise information that frequently eludes conventional evaluations of tourist attractions. Within the domain of tourism, several studies underscore the significant transformation instigated by these tools in the way tourists seek information, organise their trips, and, notably, disseminate their travel experiences among their peers (Alharbi et al., 2022; Alzboun et al., 2023; Khairina Mohd Haris et al., 2023; Scalabrini et al., 2023). A comparative sentiment analysis based on reviews of tourist attractions on platforms such as TripAdvisor and Google Maps can provide valuable insights into tourist satisfaction and the quality of attractions in a specific destination (Viñán-Ludeña & Campos, 2022; Pineda-Jaramillo et al., 2023). This research aims to undertake a comparative sentiment analysis through reviews on the online platforms TripAdvisor and Google Maps regarding the tourist attractions in Bragança, located in the northern region of Portugal.

Methodology | The TripAdvisor is a popular platform for travel reviews (Nowacki & Niezgoda, 2020; Scalabrini et al., 2023). Google Maps, also allows for such evaluations, equally ranking attractions and other points of interest (Alharbi et al., 2022; Alzboun et al., 2023; Khairina Mohd Haris et al., 2023). Apify is a web automation and scraping platform that enables users to extract data from websites, automate various web-related tasks, and access structured data (Nallakaruppan et al., 2023). Using Apify, it was possible to compile two review databases, extracting all comments from 16 attractions in the Bragança destination between 2020-23. This database comprises 250 TripAdvisor reviews and 242 Google Maps reviews, encompassing rating assessments and text-based comments in Portuguese. Reviews were categorised according to the rating systems of both platforms (ranging from 1 to 5 points) and compared with algorithmic sentiment analysis. In line with similar studies (e.g. Scalabrini et al., 2023), the ratings were further categorised as negative (1-2 points), neutral (=3 points), or positive (4-5 points). Despite

initially using the Multinomial Naïve Bayes classifier, it was employed to re-evaluate the model for text classification.

Main Results and Contributions | The results show that some attractions received consistently positive evaluations, such as *Castelo de Bragança*; *Parque Natural de Montesinho*; *Centro Ciência Viva*; *Galeria* George Dussaud; *Museu Militar de Bragança*. Occasionally, the TripAdvisor ratings remained neutral while the Google Maps ratings shifted to positive. For instance, *Domus Municipalis*; *Cidadela de Bragança*. Conversely, the *Centro de Interpretação da Cultura Sefardita do Nordeste Transmontano* received opposite assessments, with TripAdvisor showing a positive rating while Google Maps indicated a neutral rating.

A confusion matrix was employed to classify tourist reviews (e.g. Kulkarni et al., 2020), calculating the accuracy, F1-Score, and recall, to gauge the model's performance. Specific features, such as bigrams, were used to enhance model evaluation. For TripAdvisor, classified 46 reviews as negative, while only 8 were genuinely negative, suggesting a tendency to overestimate negative reviews. The model categorised 14 reviews as neutral, of which 4 were genuinely neutral, indicating reasonable performance. Regarding positive reviews, the model classified 34 reviews as positive, while 68 were positive, indicating an overestimation of positive reviews.

Limitations | Specifically, the low accuracy value in both datasets, suggests that the quantity of data collected from the web needs to be increased. Furthermore, not all multi-label classifiers and algorithms were tested using the study's model.

Conclusions | The sentiment distribution varied significantly among different attractions in both datasets, with *Castelo de Bragança* consistently receiving positive reviews. Conversely, *Domus Municipalis, Cidadela de Bragança* and *Igreja de Santa Maria*, exhibited a disproportionately high frequency of negative comments on TripAdvisor, despite having relatively high average ratings. On the other hand, the *Centro de Interpretação da Cultura Sefardita do Nordeste Transmontano* had a disproportionately high frequency of negative comments on Google Maps. The confusion matrix analysis provided additional insights, revealing the model's tendency to overestimate specific sentiments, particularly negative and positive ratings. This highlights the need to fine-tune the model's parameters or better evaluate the quality of the training data.

References

Alharbi, B. A., Mezher, M. A., & Barakeh, A. M. (2022). Tourist reviews sentiment classification using deep learning techniques: A case study in Saudi Arabia. *International Journal of Advanced Computer Science and Applications, 13*(6).

Alzboun, G., Alhur, M., Khawaldah, H., & Alshurideh, M. (2023). Assessing gastronomic tourism using machine learning approach: The case of google review. *International Journal of Data and Network Science, 7*(3), 1131-1142.

Khairina Mohd Haris, N. A., Mutalib, S., Ab Malik, A. M., Abdul-Rahman, S., & Kamaliah Kamarudin, S. N. (2023). Sentiment classification from reviews for tourism analytics. *International Journal of Advances in Intelligent Informatics, 9*(1).

Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). Foundations of data imbalance and solutions for a data democracy. In *Data democracy* (pp. 83-106). Academic Press.

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

Nowacki, M., & Niezgoda, 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

Pineda-Jaramillo, J., Fazio, M., Le Pira, M., Giuffrida, N., Inturri, G., Viti, F., & Ignaccolo, M. (2023). A sentiment analysis approach to investigate tourist satisfaction towards transport systems: the case of Mount Etna. *Transportation Research Procedia, 69*, 400-407.

Scalabrini, E., Ferreira, J., Zekan, S. B., & Fernandes, P. O. (2023). Destination Image Through TripAdvisor´s Reviews Analysis. In *Communication Design and Branding: A Multidisciplinary Approach* (pp. 205-221). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35385-7_12

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