Aspect-based Sentiment Analysis for Improving Attractiveness in Shrinking Areas

Raffaele Manna[1,2,*,†], Giulia Speranza[1,2,†], Maria Pia di Buono[1,2,†] and Johanna Monti[2,†]

[1]Dahlia srl, via Duomo 219, Naples, 80139, Italy

[2]University of Naples "L'Orientale", via Chiatamone 61/62, Naples, 80121, Italy

Abstract

In this paper we present the motivations, the methodology and the data used to develop a platform aimed at improving the information about peripheral and shrinking areas in order to foster their attractiveness. As case study, we select the internal area of the Ufita Valley in Irpinia (Campania, Italy). The platform shows through maps and statistics the insights on the cultural attractions in the area of interest on the basis of an aspect-based sentiment analysis model trained on the Google reviews. The platform, addressed to local administrations, is intended as a tool for obtaining an overview of public sentiment towards cultural sites, understanding strengths and weaknesses, as well as for supporting governance and intervention policies for these sites.

Keywords

Shrinking Areas, Cultural Tourism, Local Administrations, ABSA

1. Introduction

Shrinking areas, as reported in Grasland et al. [1], are internal and rural regions affected by depopulation, demographic decline and a rise in the proportion of elderly people, remoteness of public services and scarcity of infrastructures. Shrinking areas are also characterised by geo-morphological fragility and poor accessibility which causes in some cases also economic impoverishment [2, 3]. Furthermore, the high percentage of emigration among young generations towards larger centres, the absence of job opportunities and employment, the abandonment of buildings, houses and land is also causing the disappearance of traditions, customs, local knowledge and artisan expertise. The shrinkage is caused by a multitude of natural, political and economic factors common to many similar areas around Europe [4, 5]. In order to face the problem and propose innovative and effective solutions, local governments and public administrations should work towards a common strategic plan and the establishment of a collection of actions and activities together with stakeholders, social groups, citizens, companies and organisations [4].

Compared to larger centres, governance in shrinking areas is indeed penalised by a lack of financial and institutional capacities to improve and promote cultural tourism and highly depends on external resources to cope with the different problems they may face [4, 5, 6].

In order to support public administrations and institutions in their local governance a data-driven decision making approach could prove effective [7]. Several types of data, from reports to surveys, reviews and social media, can be automatically leveraged to empower administrations to make informed choices, guide strategic decisions as well as discover new insights, identify weaknesses and strengths, to enhance public services or optimise resource allocation and investments. Effective data analysis can indeed be used to derive essential information for making prevision, statistics, anticipate trends, and discover areas of improvements thus transforming raw and disconnected data into rich knowledge to be interpreted and reused wisely.

In this paper, we present the development of the KiNESIS Project platform to support local administrations in improving attractiveness of the shrinking internal area of the Ufita Valley in Irpinia (Campania Region, Italy) with a specific focus on cultural sites and tourism. The paper is organised as follows. Section 2 offers an overview of the project; Section 3 delves into existing research on the topic, providing context for our approach. Section 4 outlines the specific methodology we applied. Following this, Section 5 discusses the conclusions and outlines potential directions for future research.

2. The KiNESIS Project

The *KNowledgE alliance for Social Innovation in Shrinking villages* Project (KiNESIS)[1] is an Erasmus+ Programme of

[1]https://www.kinesis-network.eu/homesite/1/1/home-page.html

Ital-IA 2024: 4th National Conference on Artificial Intelligence, organized by CINI, May 29-30, 2024, Naples, Italy

*Corresponding author.

†These authors contributed equally.

rmanna@dahliasrl.it (R. Manna); gsperanza@dahliasrl.it (G. Speranza); mpdibuono@unior.it (M. P. di Buono); jmonti@unior.it (J. Monti)

0009-0006-6285-8557 (R. Manna); 0000-0002-9249-492X (G. Speranza); 0000-0001-7284-6556 (M. P. di Buono); 0000-0002-4563-5988 (J. Monti)

the European Union co-founded project coordinated and managed by the University of Naples "L'Orientale" which gathers together many academic institutions, stakeholders and organisations across Europe, in particular Italy, Spain, Germany, The Netherlands and Estonia. The KiNESIS Project addresses the topic of international cooperation focused on shrinking areas with the aim of promoting and fostering ideas, developing and sharing best practices, projects, workforce, productivity and attractiveness. The Project's objectives are to revitalise depopulated, shrinking and marginalised areas by stimulating entrepreneurship and entrepreneurial skills; to create local living laboratories to promote social inclusion and entrepreneurial development; to experimenting new, innovative and multidisciplinary approaches in teaching and learning; to facilitate the exchange, flow and co-creation of knowledge at a local and global level. During the years, several activities have been carried out as part of the KiNESIS Project such as co-participation tables, workshops and conferences, training sessions and summer schools, internships and Erasmus+ students exchanges, Hackathon and fairies, publication of handbooks, reports, scientific documents and best practices, the creation and dissemination of promotion materials.

3. Related Works

The cultural tourism sector is increasingly driven by the use of data-driven approaches [8, 9]. In this context, data from user-generated content platforms allows for the collection of increasingly up-to-date and real-time insights capable of identifying trends to guide decisions around economic activities [10]. Specifically with regard to tourism-related economic activities, in recent years, methods using Natural Language Processing (NLP) techniques have been applied to hotel reviews to extract from user-generated content the sentiment and perceptions of users in relation to various categories relevant to the structure and the services it offers [11, 12]. In addition, NLP and topic modeling methods have been applied to user-generated content in reference to quality dimensions in the museum field [13]. Additionally, other studies have investigated the attractiveness of Italian cities using user-generated content. Specifically, users' behaviour has been measured to identify the annual trend of photographic activity in cities [14]. User-generated content (UGC) on social media and review platforms related to tourist attractions represents a valuable source of information for guiding decisions towards more informed economic growth in regions that potentially benefit from cultural tourism [15]. However, information from UGC has often been analysed by considering only user ratings or focusing on large tourist hubs such as Italian art cities [14].

In the following sections, we propose a study related to the analysis and extraction of information related to specific categories in relation to different cultural sites in areas at risk of depopulation. Specifically, we show the development of a model capable of extracting different dimensions of intervention regarding cultural sites to guide public administrations in potential investments for the maintenance and enhancement of cultural attractions present in the territory. In this context, some studies show how investments in the tourism sector can be beneficial for the growth of inland areas at risk of depopulation [16].

4. Methodology

Within the KiNESIS Project, with the aim of supporting the local administrations and institutions in the improvement of the attractiveness in the shrinking internal area of the Ufita Valley (Italy), we developed a user-friendly visualisation platform trained on an Aspect-Based Sentiment Analysis classification system for cultural attractions and cultural sites. In this section, the methodology adopted for training the Aspect-Based Sentiment Analysis (ABSA) classification system for cultural attractions and potentially of tourist interest in the Irpinia area will be discussed. Specifically, in Section 4.1, the data collected and on which the ABSA model was trained will be described; in Section 4.2, the model architecture and the elicited outputs will be presented; and finally, the integration into the data analysis visualisation platform will be presented.

4.1. Data and Exploratory Data Analysis

The textual data used for training the ABSA model were collected from the Google Maps platform. Specifically, data on reviews of cultural attractions in Irpinia were collected. The names of the attractions were collected using the resources offered by *Sistema Irpinia*[2]. Sistema Irpinia is an interactive platform that promotes sites of historical, artistic, architectural, cultural, environmental and food and wine heritage of Irpinia. This platform contains 408 cultural sites.

Therefore, the reviews related to the cultural attractions present on Sistema Irpinia were extracted from Google Maps. A total of 9504 reviews were extracted. Of these, about 4% are represented by reviews that do not convey any textual information, showing only the user rating expressed on Google Maps through the assignment of stars. For ABSA model training purposes, these latter reviews were removed. Before training the ABSA model, an exploratory analysis was conducted on the information conveyed by the data extracted in the

[2]https://sistemairpinia.provincia.avellino.it/it/node/4

Type	#
Churches	160
Historical buildings	65
Castles	48
Other places	36
Archeological area	24
Religious complexes	24
Castles - historic palaces	8
Total	**365**

Table 1
Type of cultural sites

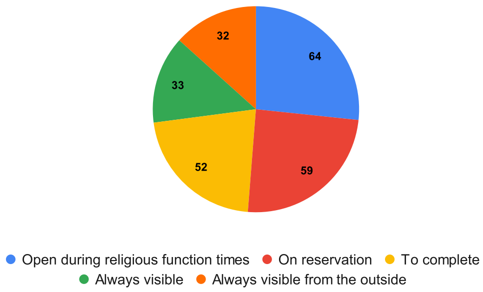

Figure 1: Availability of visits for cultural sites.

manner previously described. This information focused on: 1) the number of cultural sites present in each municipality in the Irpinia area; 2) the type of cultural attraction; and 3) the accessibility to the type of cultural site. The largest number of cultural sites present and represented on the Google Maps platform belongs to two municipalities (*Rocca San Felice* and *Bonito*) in the Ufita Valley, a consortium of municipalities participating in the KiNESIS Project. Specifically, the municipality of Rocca San Felice is represented by 14 cultural sites, while Bonito has 13 cultural sites.

Table 1 shows the number of types of cultural sites extracted from the *Sistema Irpinia* platform and consequently for which reviews were found. The most represented type is related to churches (160), followed by historical buildings (65), castles (48), and the 'other place' type (36), which represents noble residences. The least represented, although the most extensive in terms of spatial extent, are archaeological areas and religious complexes.

For each cultural site that falls into one of these types, categories/aspects were assigned to the reviews using the distant supervision method [17, 18] in combination with the information from the overall rating score given by the review stars. Specifically, this method allows to build an annotated dataset without or with little human intervention. In fact, rule-based heuristics are used in order to produce labeled data and on these labeled data produced being then used to train a model. The rule-based heuristic consists of lists of words, primarily adjectives and adverbs, that can signal positive and negative characteristics related to the aspects identified as salient for describing the conditions of a cultural site. In the case of the ABSA model, four categories/aspects were identified that are able to describe the conditions of cultural sites. The aspects related to cultural sites are the following: 1) accessibility; 2) signage; 3) appearance of the place and 4) overall score.

Specifically, 'Accessibility' refers both to the availability of opening hours for the public and to the provision of equipment to make the visit more accessible to a wide range of visitors, while 'Signage' refers both to road signs that allow the site to be reached and to the presence of any explanatory totems within cultural sites. For example, in the following review extracted from Google Maps in reference to the Goleto Abbey located in Sant'Angelo dei Lombardi:

> L'abbazia del Goleto è ancora CHIUSA PER LAVORI DI RISTRUTTURAZIONE che dovrebbero terminare il 4 giugno 2024. Spero venga rispettata la data di consegna dei lavori perché è sempre un piacere visitare il complesso. L'abbazia è spettacolare e mi aspetto che al termine della ristrutturazione, lo sarà ancor di più. Complimenti alla ditta dell'architetto X. (*The Goleto Abbey is still CLOSED FOR RESTORATION WORK which should be completed on June 4, 2024. I hope the deadline for the work will be respected as it is always a pleasure to visit the complex. The abbey is spectacular and I expect it to be even more so after the renovation is complete. Congratulations to the firm of architect X.*)

The extracted aspects are related to 'Accessibility' and 'Appearance of the place', as well as an overall score that shows the ratio for all aspects. As far as the 'Accessibility' aspect is concerned, the identified text portion is *'CLOSED FOR RESTORATION WORK'*, while for the 'Appearance of the place' aspect the identified text portion is *'The abbey is spectacular'*.

The application of the distant supervision method allowed us to annotate the data at our disposal with minimal human intervention. Subsequently, an exploratory analysis of the annotated data at our disposal was carried out. In Figure 1, for example, the frequencies related to the accessibility of the cultural sites present in the dataset are shown. In addition, as previously mentioned, the 'Accessibility of cultural sites' refers to two different use functions. In the case of Figure 1, accessibility is shown in terms of availability for visits. For example, in Figure

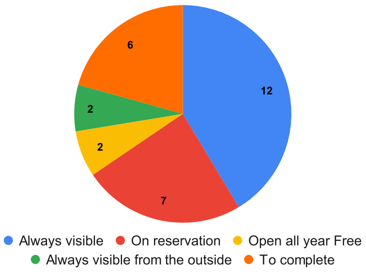

Figure 2: Availability of Archeological Areas.

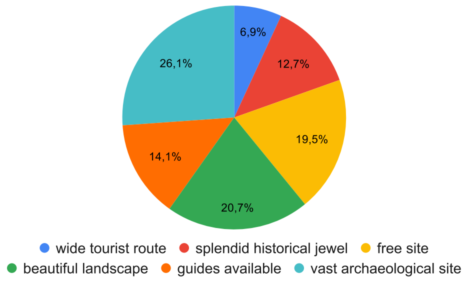

Figure 3: Overall score - Positive.

1, 64 sites are available only during the hours dedicated to religious functions. Instead, the 'To complete' label with 52 cases refers to closed cultural sites that are not available for visits or are undergoing renovation work.

As previously mentioned, the dataset also includes information concerning the type of cultural site among its characteristics. This information allows for more granular information in relation to the different aspects identified. For example, Figure 2 shows the frequencies of availability for visits only for cultural sites of the 'Archaeological Areas' type. In this case, for example, we can note that in 2 cases it was reported that the archaeological area is open all year round.

In addition, the dataset also includes user reviews of cultural sites in the form of star ratings. These were used in conjunction with the identified aspects to balance the overall score. Specifically, the textual spans related to the aspects of 'Accessibility', 'Appearance of the Place', and 'Signage' were identified. A sentiment and emotion lexicon[3] was then applied to these spans to add a score associated with each aspect. The score from the lexicon was used in conjunction with the user's rating score in the review. These scores from the lexicon for each aspect identified in the text, together with the scores assigned by the user, were used as supervision labels to train the ABSA model. Figure 3 shows the textual spans extracted for different aspects in relation to the positive overall score.

Specifically, Figure 3 shows textual spans related to: Accessibility with the span 'free site'; Signage with 'guides available'; and Appearance of the place with textual spans related to the beauty and spaciousness of the cultural sites.

[3]The sentiment and emotion lexicon used comes from this repository: https://saifmohammad.com/WebPages/nrc-vad.html and the lexicon for the Italian language was used. Specifically, the NRC Valence, Arousal, and Dominance (VAD) Lexicon was used. This resource includes a list of more than 20,000 words and their valence, arousal, and dominance scores. These scores represent the emotional qualities of the words.

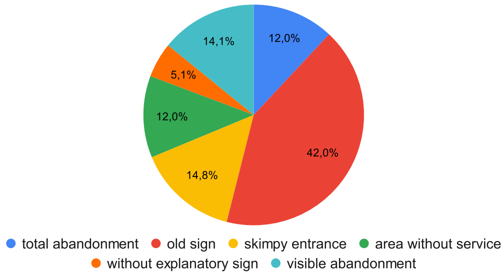

Figure 4: Overall score - Negative.

In Figure 4 are shown the negative textual spans about the 'Appearance of the place' with judgements around the state of neglect of some cultural sites; about the 'Signage' with the reporting of the complete absence of explanatory panels or damaged and in relation to 'Accessibility' with comments on the lack of services or structural deficiencies.

4.2. ABSA Model and Platform

This section outlines the fine-tuning process of the chosen model and the implementation of the platform prototype to be made available to public administrations for effectively directing policies towards cultural sites with potential appeal for cultural tourism. The platform's objective is to provide insights to inform decisions on which aspects of a specific cultural site to focus on in order to prepare it for the influx of tourists.

In this context, we applied ABSA to analyze and classify user reviews of cultural sites and attractions in the Ufita Valley. We employed the XLM-Roberta-base model[4], a multilingual pre-trained transformer-based language model[19], for the ABSA task. The model was fine-tuned using a dataset of user reviews described in section 4.1.

[4]https://huggingface.co/FacebookAI/xlm-roberta-base

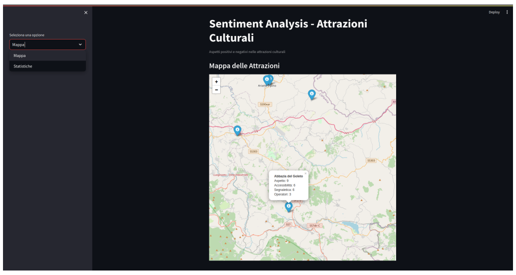

Figure 5: ABSA Platform - Map.

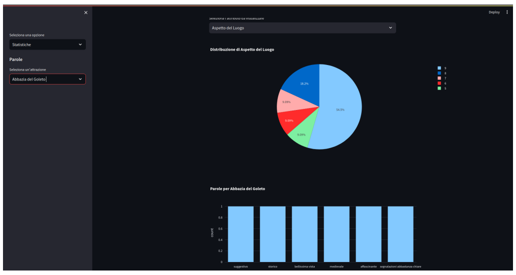

Figure 6: ABSA Platform - Statistics.

The dataset was divided into training (60%) and testing (40%) sets. The fine-tuning process involved optimising the model parameters to minimise the loss function, which measured the discrepancy between the predicted and actual sentiment labels. We employed a multi-task learning approach, training the model to simultaneously perform two tasks:

- **Aspect Category Classification**: Classifying each textual span into its corresponding aspect category (e.g., 'Accessibility', 'Signage', 'Appearance').
- **Overall Sentiment Score Assignment**: Assigning a sentiment score (ranging from 1 to 10) to each textual span based on the sentiment expressed towards the corresponding aspect.

The overall sentiment score for each review was calculated as the average of the individual aspect scores, weighted by the number of mentions of each aspect. Additionally, the overall score was adjusted based on the user's overall judgement of the review (positive, negative, or neutral). The fine-tuned model's performance achieves an F1-score of 78% in extracting the correct textual spans for each aspect and assigning the correct score. This performance represents the average value across the model's performance. The platform developed to provide public administrations with insights into cultural sites features two main data visualisation modes: a map and cultural site-specific statistics, as illustrated in Figure 5.

The map is populated with markers corresponding to the coordinates of cultural sites and displays the weighted average of the scores for each identified aspect of the cultural site. This approach visualises and considers the frequencies of the different scores (ranging from 1 to 7, negative to positive) for each review of the cultural site. The map provides a quick overview of the sentiment associated with the identified aspects for each cultural site. Additionally, the platform offers a second data visualisation tool that focuses more on the specific cultural site. For the statistics, you can select the name of the cultural site and the aspect of interest and view the corresponding information. For example, Figure 6 shows the information related to Abbazia del Goleto in relation to the 'Appearance of the place'.

In this case, when selecting the aspect to view for a specific cultural site, the user will find the information extracted using two methods: a pie chart and a bar chart. The first shows the sentiment scores associated with the selected aspect, while the second shows the top-n words (sorted by frequency of use) extracted from the reviews and associated with a particular aspect. Indeed, in Figure 6, we can observe that the majority of reviews express a very favourable opinion of the Goleto Abbey, as more than half (54%) have a very high 'Appearance of the place' score (9) and use words such as "evocative," "historical," "charming," and "beautiful view."

5. Conclusion and Future Works

In this paper, we present the implementation of a platform capable of extracting and classifying sentiment indices for various aspects from online reviews of cultural sites. Specifically, the ABSA model and the platform prototype were implemented within the KiNESIS project, which aims to investigate methods for mitigating the effects of ongoing depopulation in rural areas across Europe. In this context, the platform is proposed to public administrations as a tool to support their policies regarding cultural tourism sites of potential interest. Indeed, the platform can be a valuable tool for obtaining an overview of public sentiment towards cultural sites, as well as a tool for directing active intervention policies for the maintenance of specific aspects related to cultural sites. The platform is currently under development and is only being tested for the Ufita Valley area and for Italian-language reviews. Additionally, thanks to the KiNESIS Project, activities and data collection have already begun to extend the training and testing phase to the Oldambt region (The Netherlands) in collaboration with the Dutch partner of the KiNESIS Project. In this context, therefore, the ABSA model and platform will be tuned and made available in other languages to analyse data from other European regions.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the Erasmus+ Programme of the European Union for the KiNESIS Project (Grant Agreement 621651-EPP-1-2020-1-ITEPPKA2-KA).

References

[1] C. Grasland, R. Ysebaert, B. Corminboeuf, N. Gaubert, N. Lambert, I. Salmon, M. Baron, S. Baudet-Michel, E. Ducom, D. Rivière, et al., Shrinking regions: a paradigm shift in demography and territorial development, Ph.D. thesis, Parlement Européen; Direction Générale des politiques internes de l'Union ..., 2008.

[2] T. Amodio, Territories at risk of abandonment in italy and hypothesis of repopulation, Belgeo. Revue belge de géographie (2022).

[3] S. De Rubertis, Dinamiche insediative in italia: spopolamento dei comuni rurali, Perspectives on rural development 2019 (2019) 71–96.

[4] A. Haase, G.-J. Hospers, S. Pekelsma, D. Rink, Shrinking areas: Front-runners in innovative citizen participation, The Hague: European Urban Knowledge Network, 2012.

[5] D. Rink, P. Rumpel, O. Slach, C. Cortese, A. Violante, P. C. Bini, A. Haase, V. Mykhnenko, B. Nadolu, C. Couch, et al., Governance of shrinkage: Lessons learnt from analysis for urban planning and policy, Leipzig: Helmholtz Centre for Environmental Research (2012).

[6] D. Rink, A. Haase, K. Großmann, M. Bernt, C. Couch, M. Cocks, A. Violante, C. Cortese, P. Calza Bini, How shrinkage and local governance are interrelated across urban europe: a comparative view, 2011.

[7] R. Matheus, M. Janssen, D. Maheshwari, Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities, Government Information Quarterly 37 (2020) 101284.

[8] T. Kalvet, M. Olesk, M. Tiits, J. Raun, Innovative tools for tourism and cultural tourism impact assessment, Sustainability 12 (2020) 7470.

[9] M. T. Cuomo, D. Tortora, P. Foroudi, A. Giordano, G. Festa, G. Metallo, Digital transformation and tourist experience co-design: Big social data for planning cultural tourism, Technological Forecasting and Social Change 162 (2021) 120345.

[10] M. P. A. Austin, P. Austin, M. M. Marini, A. Sanchez, C. Simpson-Bell, J. Tebrake, Using the Google Places API and Google Trends data to develop high frequency indicators of economic activity, International Monetary Fund, 2021.

[11] Y. Liu, T. Teichert, M. Rossi, H. Li, F. Hu, Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews, Tourism Management 59 (2017) 554–563.

[12] S. U. S. Chebolu, F. Dernoncourt, N. Lipka, T. Solorio, Survey of aspect-based sentiment analysis datasets, arXiv preprint arXiv:2204.05232 (2022).

[13] D. Agostino, M. Brambilla, S. Pavanetto, P. Riva, The contribution of online reviews for quality evaluation of cultural tourism offers: The experience of italian museums, Sustainability 13 (2021) 13340.

[14] S. Giglio, F. Bertacchini, E. Bilotta, P. Pantano, Using social media to identify tourism attractiveness in six italian cities, Tourism management 72 (2019) 306–312.

[15] A. Torre, H. Scarborough, Reconsidering the estimation of the economic impact of cultural tourism, Tourism Management 59 (2017) 621–629.

[16] M. H. Guimarães, L. C. Nunes, A. P. Barreira, T. Panagopoulos, Residents' preferred policy actions for shrinking cities, Policy Studies 37 (2016) 254–273.

[17] A. Go, R. Bhayani, L. Huang, Twitter sentiment classification using distant supervision, CS224N project report, Stanford 1 (2009) 2009.

[18] A. Giannakopoulos, D. Antognini, C. Musat, A. Hossmann, M. Baeriswyl, Dataset construction via attention for aspect term extraction with distant supervision, in: 2017 IEEE International Conference on Data Mining Workshops (ICDMW), IEEE, 2017, pp. 373–380.

[19] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, CoRR abs/1911.02116 (2019). URL: http://arxiv.org/abs/1911.02116. arXiv:1911.02116.