

Sentiment Analysis as Feedback for Improving Sustainable Management in the Northern Subcarpathians of Oltenia Destination

Truțescu Marius-Nicolae¹, Nicolaie Daniela¹, Matei Elena²

1-Simion Mehedinți Doctoral School, Faculty of Geography, University of Bucharest

2-University of Bucharest, Faculty of Geography, 1 Nicolae Balcescu Blvd, 010041, Romania

1. INTRODUCTION

Tourists' option for a destination/product is influenced, conventionally, by the recommendations of friends and relatives, travel agencies or marketing policy (Teodorescu et al., 2014). In the framework of nowadays development of technology, the final decision on holiday destination is, now, clearly influenced by the online environment, specifically by the online reviews (González-Rodríguez et al. 2016).

There are numerous systems on the market that operate globally, regionally or nationally, with facilitation tools by translating into different languages, shaping on customers' profile and demand, online payment systems, etc. Romania, the country with one of the best online navigation systems in the world (Matei et al., 2021) has experienced a fast adjustment of the tourism industry to online promotion (Booking.com, Tripadvisor.com, Agoda.com or Infoturism.ro, etc.).

These systems represent Big data platforms used by different stakeholders (Rodríguez-Díaz & Espino-Rodríguez, 2018) and/or researchers (Li et al., 2018). One of the latest research topic is sentiment analysis (SA) (Serrano-Guerrero et al., 2015) which processes and analyzes large amounts of data based on machine learning techniques. SA is a concept that includes several tasks, mainly processing, mining or extracting information from various textual sources and classifying feelings. Opinions, feelings, appreciations, attitudes and emotions are the central assessed elements of the analysis of feelings (Liu, 2012) in order to obtain the specific features of an action, product, service, matter or subject (Ainin et al., 2020) and they signify qualitative data.

4. RESULTS AND DISCUSSIONS

4.1. Sentiment Analysis (SA)

Tourists' reviews analysis regarding Oltenia Subcarpathians shows that predominant emotions (over 80%) are positive, on the spectrum of trust, joy, prediction, and less negative emotions including anger, disgust, fear, and sadness. Results shows that reviews on Turistinfo.ro are slightly higher in terms of positive emotions and sentiments than the other two international platforms (Booking.com, Tripadvisor.com). Also, the evolution of a feeling shows that negative emotions such as anger and disgust are stronger in 2020 (affected by COVID-19 pandemic), while positive emotions such as joy and trust are stronger in 2019 and 2018.

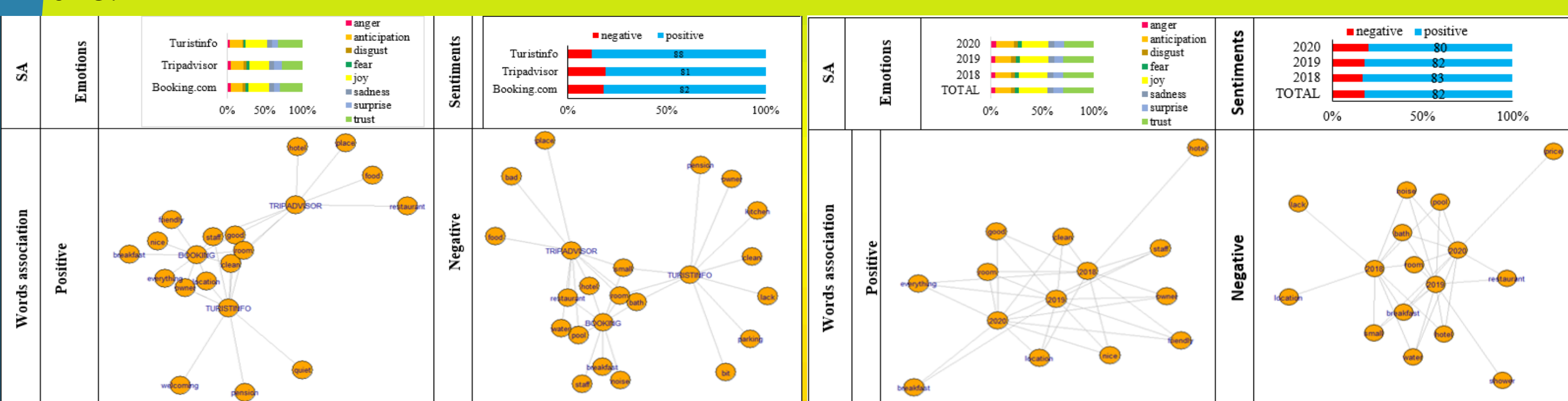


Fig. 2. SA results and words association on the three platforms, 2018-2020 (R)

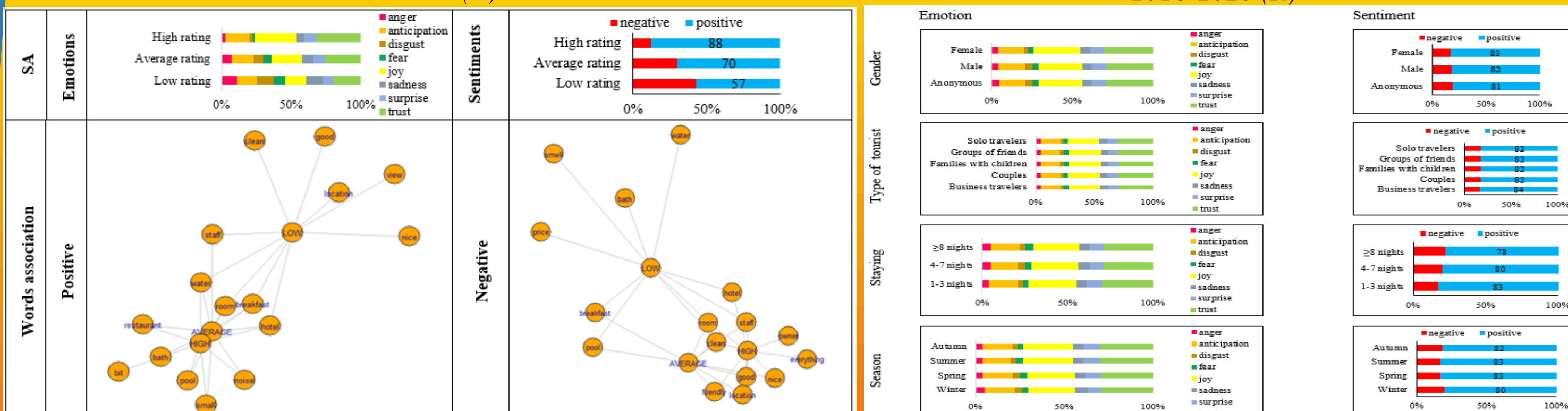


Fig. 3. SA results and words association by years, 2018-2020 (R)

Fig. 4. SA and words association in hotel ratings (R)

Tourist stays ratings equalizes positive and negative emotions and sentiments' values. The low values of ratings in terms of negative emotions and sentiments are more significant than the medium and high values of ratings.

4.2. Rating and SA correlations

Table 1. Regression equation pattern output

| Model | Coefficients ^a | Unstandardized Coefficients | Standardized Coefficients | t | Sig. | Collinearity Statistics | |
|-------|---------------------------|-----------------------------|---------------------------|--------|--------|-------------------------|------------|
| | | | | | | B | Std. Error |
| 1 | (Constant) | 5.626 | 0.065 | 86.845 | 0.000 | 1.000 | 1.000 |
| | Positive | 3.325 | 0.065 | 0.438 | 50.887 | 0.000 | 1.000 |
| | Negative | -0.045 | 0.031 | -0.012 | -1.446 | 0.148 | 1.000 |

a. Dependent Variable: Rating
b. Independent variables: positive & negative sentiments
Source: Extras from SPSS v.17 output

For testing the correlation within SA, regressive analysis has been used to show correlation between Rating and Sentiments. The output pattern adjusted R² explains 19.1% of the variance in the data. The F-test is highly significant (1295.00) so it can assume that the model explains a significant amount of the variance/discrepancy for Rating.

6. ACKNOWLEDGEMENT

The authors have equal contribution on this paper which is part of the PhD Thesis research. The research is financed by the University of Bucharest. Many thanks to VISEGRAD FUND.

7. RESOURCES

1. González-Rodríguez, MR., Martínez-Torres, R., Toral, S., 2016. Post-visit and pre-visit tourist destination image through eWOM sentiment analysis and perceived helpfulness. *International Journal of Contemporary Hospitality Management*, 28(11), pp.2609-2627.
2. Teodorescu, N., Părgaru, I., Stăncioiu, F. A., Matei, E., Botoș, A., 2014. Modelarea cercetării imaginii unei destinații turistice. *Revista Amfiteatral Economic*, 8, pp. 857 – 869.
3. Matei, E., Ilovan, O. R., Sandu, C. B., Dumitrache, L., Istrate, M., Jucu, I.S., Gavrilidis A. A. 2021. Early Covid-19 pandemic impacts on society and environment in Romania. Perception among population with higher education. *Environmental Engineering and Management Journal*, 20(2), pp.319-330.
4. Rodríguez-Díaz, M., Espino-Rodríguez, T.F. 2018. A methodology for a comparative analysis of the lodging offer of tourism destinations based on online customer reviews. *Journal of Destination Marketing & Management*, 8, pp. 147-160.

5. Li, J., Xu, L., Tang, L., Wang, S., Li, L., 2018. Big data in tourism research: a literature review. *Tour Management*, 68, pp.301-323.
6. Serrano-Guerrero, J., Olivas, J.A., Romero, F.P, Herrera-Viedma, E. 2015. Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, 311, pp.18-38.
7. Liu, B. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), pp. 1-167.
8. Ainin, S., Feizollah, A., Anuar, N.B., and Abdullah, N.A. 2020. Sentiment analyses of multilingual tweets on halal tourism. *Tourism Management Perspectives*, 34.
9. Zhao, Y. 2012. *R and data mining: examples and case studies*. Academic Press, New York.
10. Khalil, S., Fakir, M. 2017. RCrawler: An R package for parallel web crawling and scraping. *SoftwareX*, 6, pp.98-106.

2. RESEARCH AIM AND OBJECTIVES

The aim of this research is to analyze tourists' satisfaction through sentiment analysis and ratings expressed on booking platforms used by the hotel industry.

Research objectives:

O1: Data mining on main tourism platforms in Romania

O2: Identifying emotions and sentiments, spatial and temporal variables, using the R Software

O3: Creating maps of negative and positive sentiments' classification and rating

3. METHODS AND DATA MANAGEMENT

Sentiment Analysis has been processed through R Software, namely Syuzhet package. Hence the first step was to create database through building a web scraper (data extraction), which extracted targeted data from booking system reviews by means of Rvest package in R Language from RStudio (efficient in data processing and web mining according to Zhao, 2012). Along with RCrawler, RStudio is considered to have a 99.8% scrapping rate (Khalil & Fakir, 2017). Sentiments and emotions have been analyzed using machine learning and automatic annotations. Reviews have been extracted as groups of 10 each, as much as the content of one platform page. All reviews without word content or one meaningless word (such as "and", "by") have been deleted.

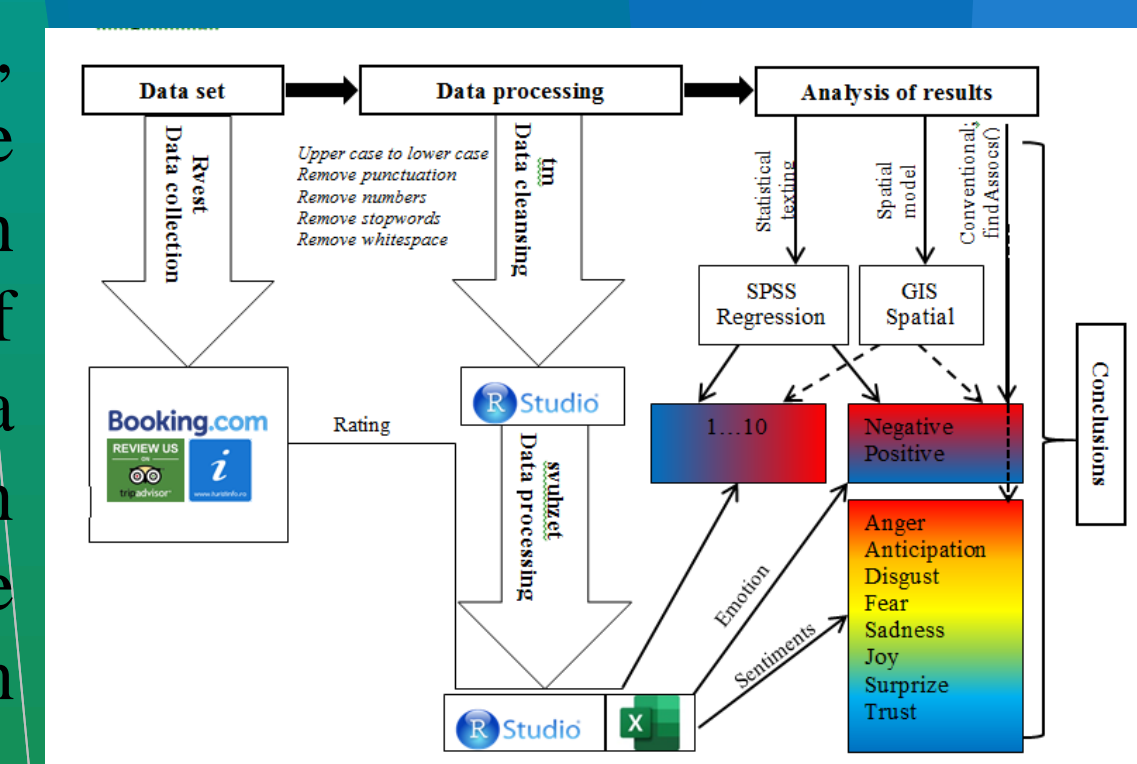


Fig. 1. Scheme of research flow

4.3. Spatial analysis of autocorrelation/clustering

The classification clustering was applied in order to observe any relation with the three topics of analyses. Lower clusters appear in Călimănești Resort, and higher in the eastern part of the area. A second trend of high clustering is in the western part around Târgu Jiu City and its surroundings. Rating clustering is specific in the northern area, where new hotels and boarding houses were built in an area recently declared as resort, nearby the Carpathians Mountains. A low trend of cluster is observed in the previous eastern resorts, characterized by higher clustering of classification. This space is also a subject of low clusters of negative sentiments and to some extent for positive sentiments.

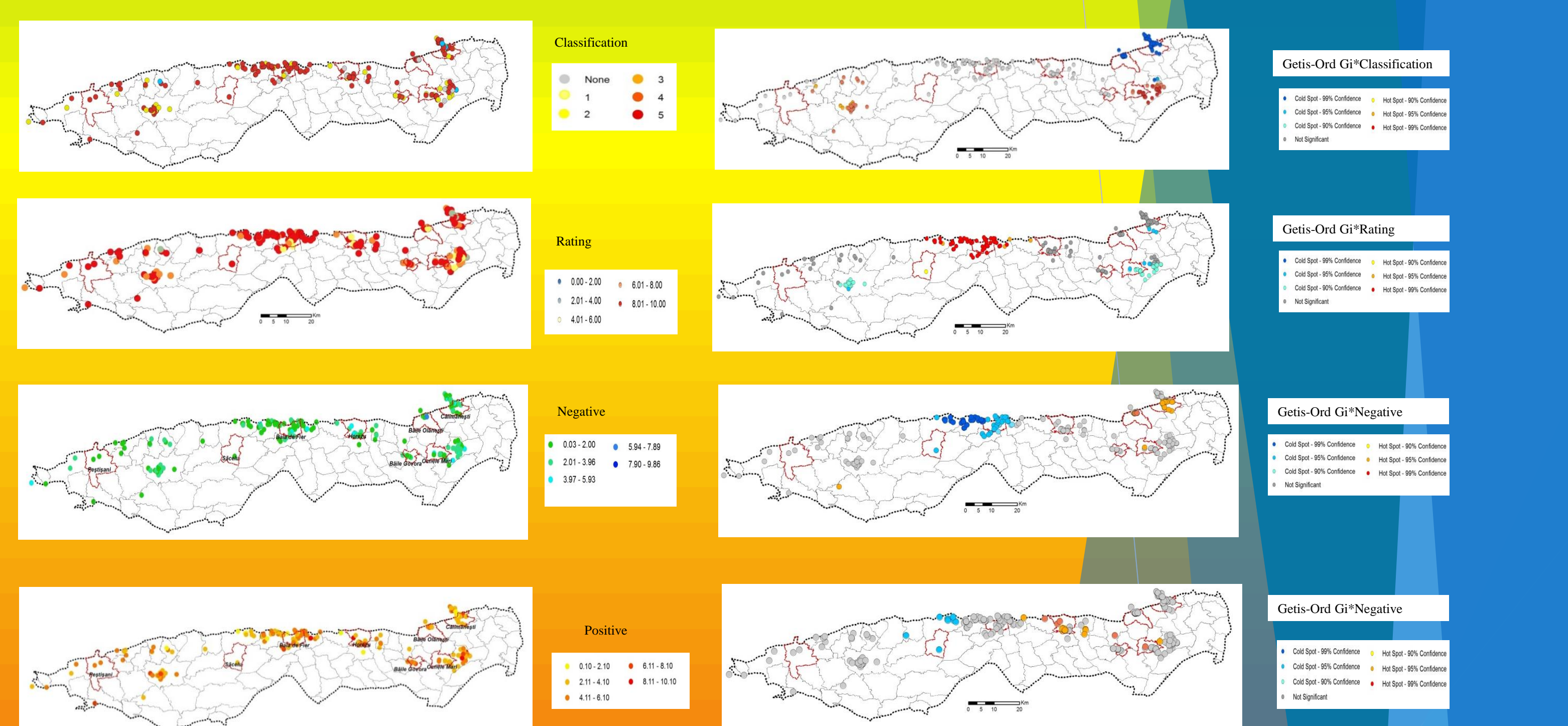


Fig. 6. Classification, rating, negative and positive sentiments maps - clustering by Getis-Ord Gi* (GIS)

Table 2. Statistics of hot spot and cold spot hotels in the studied area

| | Confidence | Classification | Rating | Positive | Negative |
|-----------|------------|----------------|--------|----------|----------|
| Hot Spot | 99% | 21 | 59 | 0 | 0 |
| | 95% | 41 | 4 | 5 | 3 |
| | 90% | 8 | 1 | 13 | 34 |
| Cold spot | 99% | 98 | 48 | 0 | 28 |
| | 95% | 1 | 7 | 8 | 27 |
| | 90% | 0 | 0 | 1 | 6 |
| Total | | 159 | 121 | 31 | 98 |

5. CONCLUSIONS

SA has recently received special attention in the research field. The study reveals that SA is a very useful tool to process large amount of data and analyze them qualitatively. The outputs show that most positive reviews are associated with short stays, 5 stars tourist accommodations or unclassified rent rooms in familiar, friendly and private atmosphere.

The research found that positive reviews are associated with the national Turistinfo.ro platform, but also highlights positive reviews of foreigners.

This study shows that destination attracts by accommodation structures located nearby the Carpathians in a beautiful landscape.

The study contributed to the existing knowledge about SA and tourism in Romania.