SENTIMENT ANALYSIS AND MULTIMODAL APPROACH APPLIED TO SOCIAL MEDIA CONTENT IN HOSPITALITY INDUSTRY

Jelena Musanovic
Raffaella Folgieri
Maja Gregoric

Abstract

Purpose – The importance of the "data gold rush" that occurs in real time on various social media platforms is recognized by various tourism stakeholders and researcher. To extract knowledge from textual data, the purpose of this study is to apply text mining techniques to social media data.
Methodology – Descriptive statistical analysis is conducted to quantify the activity of hotel brands on Facebook. The topic modelling technique Latent Dirichlet Allocation (LDA) is used to extract and validate knowledge from text data of 25 Croatian four- and five- star hotel brands that were active on social media in 2019. Sentiment analysis is used to identify personal attitudes expressed through user-generated text that hotel brands promote by posting messages on Facebook pages.
Findings – The LDA analysis of the Croatian hotel posts extracted 6 topics: Wellbeing, Atmosphere, Promotion, Gastronomy, Surrounding and Satisfaction. The results of the sentiment analysis indicated that Facebook page followers are more likely to express positive sentiments reflecting an overall satisfaction with the promoted products, services and staff by hotel brands.
Contribution – It is a unique study that provides an analysis of textual data in Croatian hospitality research. The application of the multimodal approach contributes to a better uses of contents in possible different strategies so that effective indicators can be given to perform an effective communication. This study provides recommendations, challenges, and current insights into applied communication strategies for marketers to increase a greater number of tourists visiting destinations.
Keywords: tourism, hospitality, social media, text analysis, sentiment analysis, Latent Dirichlett Allocation.

INTRODUCTION

Social media has become an integral part of modern life and has caused an explosion in data growth. Tourism stakeholders and researchers have realized the importance of the "data gold rush" created in real time on various social media platforms to engage and delight their customers, optimize business processes, create new and innovative products and services, understand human interactions, and improve conditions for people. To extract knowledge from text data, the purpose of this study is to use a text mining technique to perform topic modelling and sentiment analysis on Croatian hotel brand posts and comments on Facebook.

The negative impact on tourism is reflected in social media and poses a problem for tourism content providers. In fact, managing content to enhance positive sentiment that improves user experience (UX), which is the main goal of any successful advertising

campaign (Brajnik and Gabrielli 2010; Gharibshah and Zhu 2021; Monem 2021) textual information is the main source of cues about preferences, trends and sentiments. Many different algorithms have been used by Artificial Intelligence (AI) to analyse texts. In particular, Natural Language Processing (NLP) approaches are particularly suitable for analysing large amounts of data in real time or offline, as it is a well-developed field of AI. Among NLP approaches, Sentiment Analysis is the most commonly used approach to identify and collect opinions and sentiments in texts posted on social media, blogs, news, forums, etc. The statistics and reports obtained are important to provide indications and forecasts for tourism. These tools are also useful, as shown in our previous work (Folgieri and Bait 2014; Folgieri *et al*. 2018; Mamula *et al*. 2019), to support decisions in the development of tourism campaigns aimed at allocating tourist revenues in periods of the year not usually chosen by visitors, and are essential when unexpected events, such as Covid-19, occur.

The next section is devoted to the theoretical background of the paper, followed by a section on the research methodology, where the process of data collection and screening is presented. The results are then discussed and concluding remarks and further developments have been included in the final conclusion.

1. THEORETICAL BACKGROUND

In recent years, most researchers have focused on machine learning methods in sentiment analysis of posts on social media (Singh *et al*. 2020). The main feature of topic modelling is that it is a machine learning technique used to determine the unconventional topics discussed in an analysed text (Min *et al*. 2020). In terms of marketing strategy, authors Tirunillai and Tellis (2014) have proposed a unified framework for extracting latent quality dimensions, determining arguments based on unsupervised latent dirichlet allocation, and identifying the main features of these dimensions. This approach allows for the capture of brand attributions, segmentation within a brand and the study of the dynamics of brand positions over time. Godin *et al*. (2013) designed and developed a binary language classifier for tweets based on the Naive Bayes method for text classification and Expectation-Maximization for parameter estimation in Naïve Bayes Models. Lately, the authors applied a Latent Dirichlet Allocation model in the context of tweet hashtag recommendations and trained to group English-language tweets into a group of topics from which main keywords for new tweets can be highlighted. Priyantina and Sarno (2019) used different several review classification techniques in their study. The results of LDA showed that there are five aspects of the hotel: location, meal, service, comfort and cleanness, and concluded that the overall sentiment is influenced by one aspect. Kiatkawsin *et al*. (2020) compared two rival destinations, Hong Kong and Singapore, using latent Dirichlet allocation. In their research, 185,695 from Hong Kong`s and 93,571 from Singapore`s Airbnb were analysed, and a total of twelve topics were identified for Hong Kong, which are be categorized into four different groups, while Singapore's optimal number of themes was only five. The results also identified keywords used in rating the experience, offering deeper insights than the typical numerical ratings. Ye *et al*. (2020) used improved *k*NN model as well as LDA model in order to identify both the competitor set and important attributes in different star-rated hotel market segments. The analysed sample consisted of over 8 million customer

reviews which were collected in 6,409 China hotels. Findings from this research indicate that different market segments define the importance of particular attributes of a hotel according to hotel star ratings.

Although sentiment analysis has emerged in certain research fields since 1990, its importance for entities such as hotel companies has increased since the early 2000s, when researchers began to study the emotions, effects, and sentiments contained in texts (Ma *et al*. 2018). Barbosa *et al*. (2015) conducted research on a reviews and numerical ratings in the hotels which were located in four counutries. Mentioned data were collected from the TripAdvisor and were categorised in two main categories; positive or negative. According to Philander and Zhong (2016), who presented the possibilities of sentiment analysis usage in analysis of data from Twitter platform with the goal of creation of low-cost and real-time measures of hospitality clients attitudes and perceptions. Contribution of this research is in creation of a sentiment index was for each Twitter account associated to an integrated resort property in the Las Vegas. The authors concluded that this kind of approach, from the perspective of hospitality industry allows hospitality companies to keep track on public viewpoints at scale by creating graphical summaries of opinions. Dhaoui, Webster and Tan (2017) used a sample of 850 customer comments from Facebook pages, in order to make a comparison of lexicon-based and machine learning approaches to sentiment analysis. These two approaches were similar in the context of accuracy; but better accuracy has been achieved in classifying positive sentiment than negative sentiment. However, the authors differ significantly in their classification characteristics; the combined approach performed better results in classifying positive sentiment. Singh *et al*. (2020) provide a total overview of the sentiment analysis technique based on recent research and presented machine learning and feature extraction techniques in the context of sentiment analysis over datasets collected on social media. The researchers also addressed issues such as over simplicity in classification, such as the usage of only positive, negative and neutral categories.

Text analysis of hotel social media content provides insights and critical review and reflection on the differences of satisfied and dissatisfied clients in hotel industry. In that sense, information gathered from text mining techniques are of great importance and value in a decision making process.

2. RESEARCH METHODOLOGY

Posts and comments published on hotel brand Facebook sites were captured in 2020 using NCapture for NVivo and Facepager. The collected data ranged from January 2019 to December 2019. To determine the sample size, the following conditions had to be met: first, the first ten pages of the website www.booking.com were scraped, with hotel brands sorted by hotel rating and category; second, only hotel brands that had an official Facebook page were included in the study. Hotel brands that did not meet the conditions, did not post in English or did not post at least three times per month were not included in the study. Finally, the data frame for Croatian hotel brands consists of 2,694 Facebook posts and 5,010 comments.

An important step in natural language processing (NLP) is text pre-processing, where text was lowercased, exclamation marks, web links, tables, extra white spaces, numbers, certain words (e.g. pronouns) and step words were removed, while suffixes and prefixes of words were reduced to facilitate term aggregation. Only English-language posts and comments in English were included in the analysis. Finally, the sample contains 1791 posts and 1430 comments. To better visualize the above described research methodology, a flowchart is shown in Figure 1.

Figure 1: **Flowchart of the applied methodology**
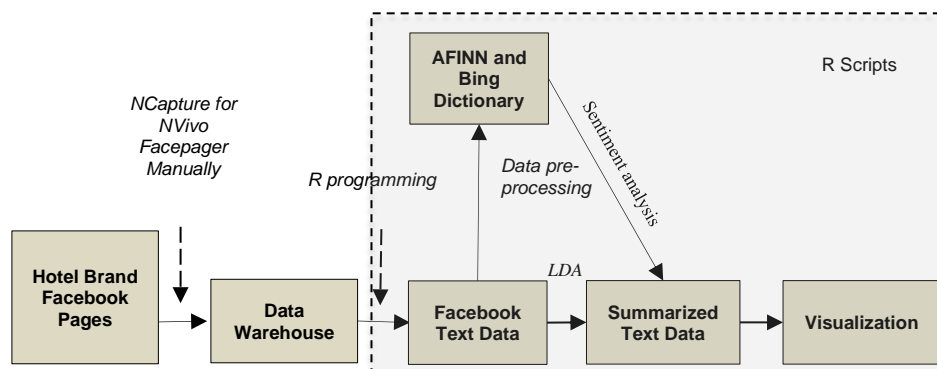
Source: Authors

Descriptive statistical analysis is performed to quantify the activity of hotel brands on Facebook. Latent Dirichlet Allocation (LDA) technique for topic modelling and sentiment analysis were applied to the text data. To perform LDA and define the optimal number of topics, the authors evaluated the performance of different selections of K as the total number of topics. As suggested by Kiatkawsin *et al*. (2020), the results of each method were normalised on a scale from 0 to 1 before the LDA models ran the results. It is worth noting that every possible number of topics between 2 and 50 was identified and scored through a grid search. The measure proposed by Arun *et al*. (2010) and Cao *et al*. (2009) aims to minimize the proposed criteria, while Griffiths and Steyvers (2004) and Deveaud *et al*. (2014) aim to maximize the log-likelihood of the word topic probability in the documents. Customers' feelings toward the hotel offer that is promoted by posting messages on Facebook pages are identified and compared using sentiment analysis. Sentiment analysis was conducted using the AFINN and Bing dictionaries. Each word in the non-binary AFINN lexicon is assigned a numerical value between -5 and 5, with negative values reflecting negative emotions and positive values reflecting positive sentiment. The higher the number, the more positive the sentiment and vice versa. The Bing dictionary divides words into positive and negative groups using a binary technique.

3. FINDINGS

*Descriptive statistical analysis*

According to the analysis, in 2019 a total of 2,694 posts and 5,010 comments of Croatian four- and five-star hotel brand were recorded. In total 489,580 page followers are recorded on Facebook pages regarding Croatian hotel brands. Croatian hotel brands mostly posted in the afternoon (2 p.m. – 6 p.m.) and in the evening (6 p.m. – 9 p.m.) and at least in the morning (6 a.m. – 9 a.m.) and midday (9 a.m. – 12 p.m.). The Facebook page fans most actively comment on in the noon (12 p.m. – 2 p.m.) and afternoon (2 p.m. – 6 p.m.). The descriptive analysis also showed that the Croatian hotel brands were posting the most from July to December.

The Index of Interaction Rate is used to determine certain text qualities of hotel companies' social media postings that contribute to customer involvement. It is calculated using consumer reactions, comments, shares, and the number of page followers. Croatian hotel companies have a 109 engagement rate, which represents the average number of interactions per 100 page fans. The interaction rate is low, and the campaign cannot be deemed a success.

Latent Dirichlet Allocation

The number of topics extracted using the LDA model for Croatian hotel brands is shown in Table 1. As indicated by Sutherland *et al.* (2020), keywords were determined by ranking the words with the highest beta value within the topic. Since LDA is a mixture model, the same words may appear in multiple topics. However, since LDA does not explicitly state the topic names, authors must interpret the overarching topics of the keywords in each topic and specify names. The validation of the topics was accomplished by analyzing the Facebook posts of the hotel brands with the highest proportion of extracted keywords in the relevant topic to confirm that the topic labels appropriately reflect the topic content (Kiatkawsin et al. 2020).

Figure 1: **Appropriate K-value for the optimal number of topics on Croatian hotel brand posts**
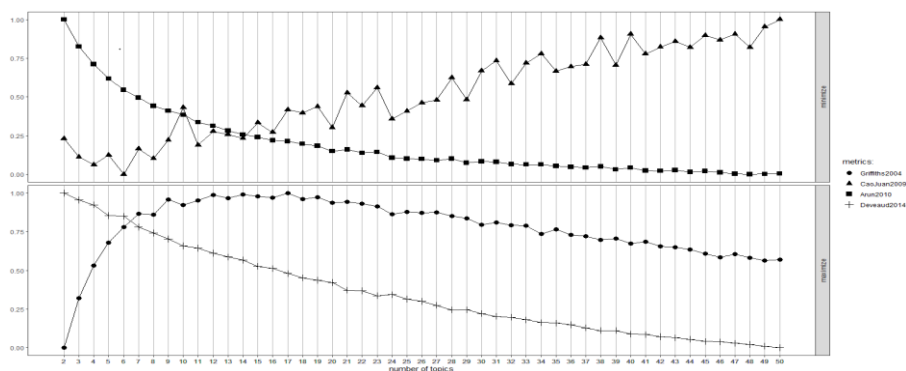
Source: Authors

The metrics od "Deveaud2014" and "CaoJuan2009" indicate that the optimal number of topics in the sample of Croatian hotel brand posts is between 2 and 6 topics. Therefore, the authors define K = 6 as the total number of topics and perform 1000 iterations of the Gibbs sampling procedure to derive the subsequent probabilistic distribution via topic assignment. After evaluating the extracted keywords per topic, the authors decided to define 6 topics as the keywords perfectly matched the content of the posts: *wellbeing, atmosphere, satisfaction, advertising, gastronomy and surrounding.*

Table 1: **Extracted latent topics with its related keywords and posts – Croatian hotel brand posts**

Topic 1 – Wellbeing
Keywords: enjoy, view, relax, pool, time, room, spa, day, see, suit

Think it couldn't get more exclusive? Think again! Reserve one of our Executive suites, for the intimate retreat overlooking Town of X. Stunning views, private jacuzzi, and spacious terrace will be the main features responsible for your most exquisite vacation experience yet.

Relax. Enjoy. Repeat. Only at X.

Topic 2 – Atmosphere
Keywords: day, bar, feel, sunset, sun, weekend, even, good, wine, night

Stunning X sunsets will create a special place in your ❤

Could you imagine pampering yourself in here? Pick a spot at our X bar to enjoy one of the books you've been longing to read or simply enjoy the quiet company of art embellishing the walls.

Topic 3 – Satisfaction
Keywords: hotel, love, best, palac, happi, guest, world, meet, year, celebr

Take a moment in time to enjoy in this view. And that's all you need for happiness. Have a happy summer at the happy island of X

Surrounded by history, forged in luxury! Experience the best of the happy island X.

Topic 4 – Advertising
Keywords: book, summer, stay, holiday, offer, experi, check, next, plan, sea

Do you have reservation for this summer? You still have time, we are waiting for you ❤ *just check our offers and book on X.*

Spread your wings and fly above our luxury resort! Check the views, the sea, the history, everything there invites you at X.

Topic 5 – Gastronomy
Keywords: beauti, restaur, best, welcom, adriat, sure, tast, dinner, food, blue

Looking for a shaded place during hot summer afternoon   Take a break   Have a lunch at X.
Dine and shine! The X restaurant is the best spot for a quick, light lunch. After be sure to hit the pool for a refreshing swim!

Topic 6 – Surrounding
Keywords: perfect, vacat, resort, place, photo, visit, beach, island, amaz, away

It's a perfect day to explore the island of X with e-bike, mountain bike or segway. X offers you a unique experience and active vacation...

Hidden in a magnificent forest on the X peninsula, the X invites you to experience the best holiday yet!

Source: Authors

Sentiment analysis

When analysing the frequency of Croatian hotel brands, Facebook followers mentioned the following negative words most often: *hard* and *limited* (3 times), *bad*, *bloody*, *cold*, *hell*, *jealous*, *sad* and *unfortunately* (2 times). The most positive words in the followers' comments are: *beautiful* (84 times), *best* (60 times), *love* (57 times), *great* (55 times), *nice* (40 times), *happy* (37 times), *amazing* (28 times), *well* (25 times), *perfect* (17 times), etc.

Graph 2: **Automatic classification of sentiment polarity of Croatian hotel brand comments**
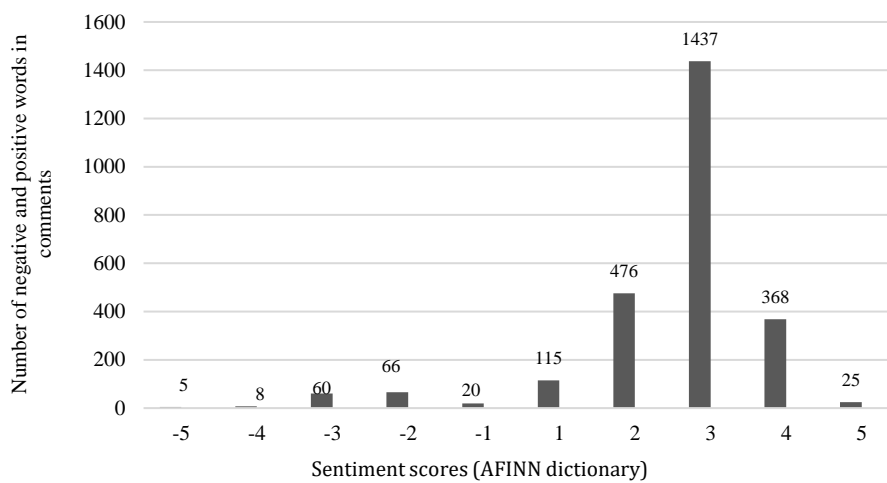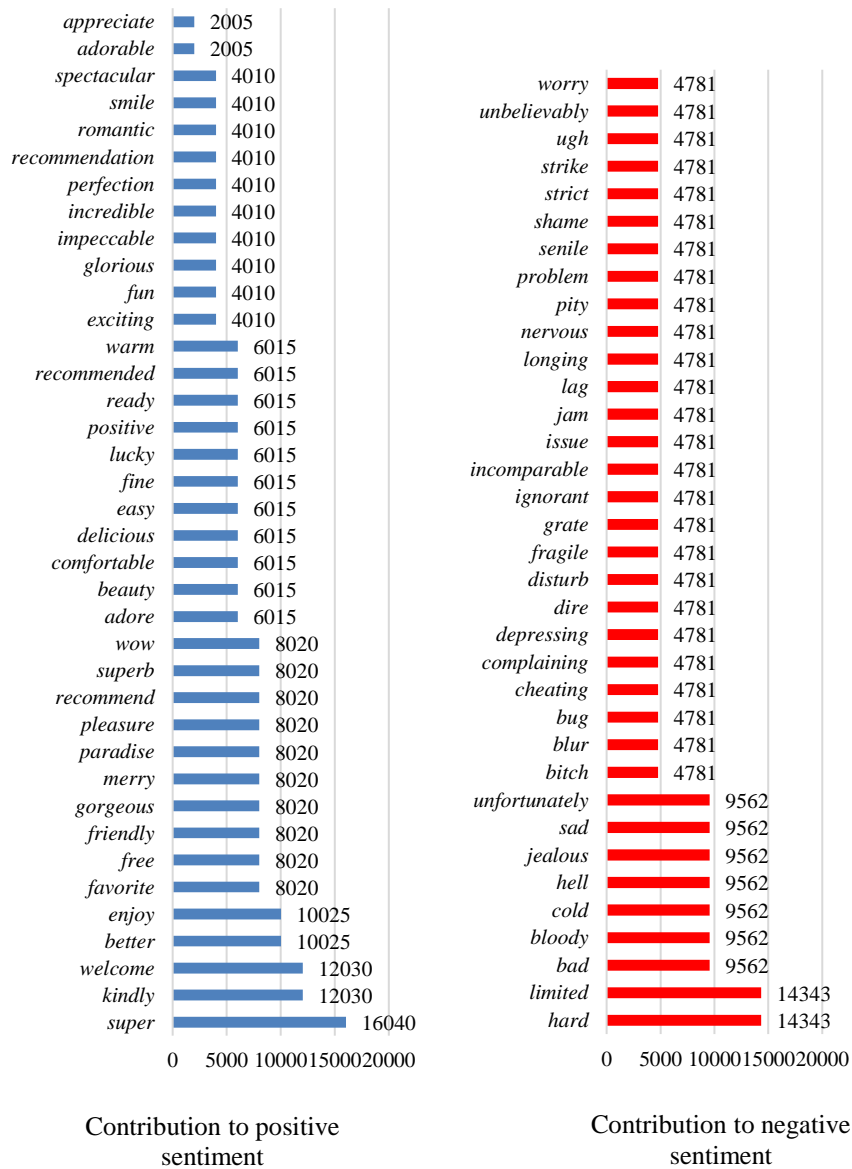

Source: Authors

The results of the automatic sentiment polarity classification show that the sample of Croatian hotel brand comments contained 5 words that were rated -5. The most positive words received sentiment score 2 (476 words), score 3 (1437 words) and score 4 (368 words).

In addition, the authors draw word contribution on the sentiments of the comments using the Bing dictionary, which clustered the words into "positive" and "negative" sentiments.

Graph 3: **Words contribution to positive and negative sentiment using Bing dictionary - Croatian hotel brand comments**

Contribution to positive sentiment

Contribution to negative sentiment

Source: Authors

Graph 3 shows the contribution of words to the positive and negative sentiment of the Croatian hotel brand comments. The words *super*, *kindly*, *welcome*, *better* and *enjoy* had the highest contribution to the positive sentiment ranging from 16,040 to 10,025. The words *appreciate*, *adorable*, *spectacular*, *smile*, *romantic*, etc. had the lowest

contribution to sentiment frequencies, which ranged from 2.005 to 4.010. A total of 35 words contributed to negative sentiment with frequencies ranging from 9,562 to 14,343. These are the words: *hard*, *limited*, *bad*, *bloody*, *cold*, *hell*, *jealous*, *sad*, *unfortunately*, etc.

The application of sentiment analysis made it possible to empirically confirm that the reputation of Croatian hotel brands is positive, which made it possible to build trust and facilitate the growth of the relationship with followers.

CONCLUSION

This paper aims to analyse social media content in the hospitality industry with the application of a multimodal approach. A brief overview of the applied methodology in tourism research is provided in the theoretical part of the paper. In a more deeper analysis of the result, it can be concluded that Croatian hotel brand posts are hardly based on relationship building, publicity or message strategies, brand building, market research and sales promotion as presented by, Loo (2020) and Ashley and Tuten (2015). They also emphasize their comparative and competitive advantages in tourism at the micro and macro levels. Core resources and attractors such as physiography (landscape and climate of the destination), culture, history, activities, special events of the destination, supporting resources and services are very popular thematic posts of hotel brands. Based on the interaction rate index, it is clear that followers behave more like passive users (lurkers) and tend to read posts and comments, view posted photos or videos, and engage less with the hotel brand and other page followers.

The findings offer meaningful implications for interactive marketing practitioners, online consultants, social media website operators and academics. This study makes a theoretical and empirical contribution to the field of tourism and hospitality on social media, as the multimodal approach is not often used. The information gained from the analysis can increase brand awareness and the growth of electronic word-of-mouth by understanding consumer behaviour and encouraging and facilitating customer engagement. To increase engagement rates and brand awareness, hotel brands and marketers should focus on growing their Facebook followers as they have the power to influence other customers and increase their continued behaviour. Therefore, it is very important to increase the number of active users who successfully interact and share the offered and owned content.

Although social media has become an important part of the marketing mix, little is understood about how to manage the creative aspects of social brand content to maximise psychological engagement. Further research which includes other factors that influence the user experience could provide further insight into a comprehensive strategy that takes into account all possible variables involved in visitor engagement through Internet media. In further research, comparing the effects of content shared on the Internet during the Covid-19 emergency and during unaffected periods should provide a picture of the substantive impact of messages on potential visitors. Such an analysis could provide important information on how to contrast the tourism impacts associated with unexpected events.

REFERENCES

Arun, R., Suresh, V., Madhavan, C.V. and Murthy, M.N. (2010), "On finding the natural number of topics with Latent Dirichlet Allocation: Some observations", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6118 LNAI (part 1), pp. 391–402. https://doi.org/10.1007/978-3-642-13657-3_43

Ashley, C. and Tuten, T. (2015), "Creative Strategies in Social Media Marketing: An Exploratory Study of Branded Social Content and Consumer Engagement", *Psychology and Marketing*, Vol. 32, No. 1, pp. 15–27. doi: https://doi.org/10.1002/mar.20761

Barbosa, R.R.L., Sánchez-Alonso, S. and Sicilia-Urban, M.A. (2015), "Evaluating hotels rating prediction based on sentiment analysis services", *Aslib Journal of Information Management*, Vol. 67, No. 4, pp. 392–407. doi: https://doi.org/10.1108/AJIM-01-2015-0004.

Brajnik, G. and Gabrielli, S. (2010), "A review of online advertising effects on the user experience", *International Journal of Human-Computer Interaction*, Vol. 26, No. 10, pp. 971–997. https://doi.org/10.1080/10447318.2010.502100

Cao, J., Xia, T., Li, J., Zhang, Y. and Tang, S. (2009), "A density-based method for adaptive LDA model selection", *Neurocomputing*, Vol. 72, No. 7–9, pp. 1775–1781. https://doi.org/10.1016/j.neucom.2008.06.011

Deveaud, R., SanJuan, E. and Bellot, P. (2014), "Accurate and effective Latent Concept Modeling for ad hoc information retrieval", *Document Numerique*, Vol. 17, No. 1, pp. 61–84. https://doi.org/10.3166/dn.17.1.61-84

Dhaoui, C., Webster, C.M. and Tan, L.P. (2017), "Social media sentiment analysis: lexicon versus machine learning", *Journal of Consumer Marketing*, Vol. 34, No. 6, pp. 480–488. https://doi.org/10.1108/JCM-03-2017-2141

Folgieri, R. and Bait, M. (2014), "The new profile of the virtual tourist-traveller: communicative perspectives and technological challenges", in *Tourism and Hospitality Industry 2014. Trends in tourism and Hospitality Industry*, Opatija: University of Rijeka, Faculty of Tourism and Hospitality Management.

Folgieri, R., Baldigara, T. and Mamula, M. (2018), "Sentiment analysis and artificial neural networks-based econometric models for tourism demand forecasting", in *Biennial International Congress.Tourism and Hospitality Industry*. Opatija: University of Rijeka, Faculty of Tourism and Hospitality Management, pp. 88–97. doi: https://doi.org/10.20867/tosee.04.10

Gao, S., Janowicz, K. and Couclelis, H. (2017), "Extracting urban functional regions from points of interest and human activities on location-based social networks", *Transactions in GIS*, Vol. 21, No. 3, pp. 446–467. doi: https://doi.org/10.1111/tgis.12289

Gharibshah, Z. and Zhu, X. (2021), "User Response Prediction in Online Advertising", *ACM Comput. Surv*, Vol 37, No. 4, pp. 1–49. doi: https://doi.org/10.1145/3446662

Godin, F., Slavkovikj, V., De Neve, W., Schrauwen, B. and Van de Walle, R. (2013), "Using Topic Models for Twitter Hashtag Recommendation", in *WWW '13 Companion: Proceedings of the 22nd International Conference on World Wide Web*, pp. 593–596. https://doi.org/10.1145/2487788.2488002.

Griffiths, T.L. and Steyvers, M. (2004), "Finding scientific topics", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 101 (suppl. 1), pp. 5228–5235. https://doi.org/10.1073/pnas.0307752101

Grönroos, C. (2008), "Service logic revisited: Who creates value? And who co-creates?", *European Business Review*, Vol. 20, No. 4, pp. 298–314. doi: 10.1108/09555340810886585

Guo, Y., Barnes, S. J. and Jia, Q. (2017), "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation", *Tourism Management*, Vol. 59, pp. 467–483. https://doi.org/10.1016/j.tourman.2016.09.009

Kiatkawsin, K., Sutherland, I. and Kim, J. Y. (2020), "A comparative automated text analysis of airbnb reviews in Hong Kong and Singapore using latent dirichlet allocation", *Sustainability (Switzerland)*, Vol. 12, No. 16, 6673. https://doi.org/10.3390/su12166673

Lies, J. and Fuß, S. (2019), "Factors influencing the success of social media postings: A quantitative analysis of over 60 million Facebook postings | HSTalks", *Journal of Digital & Social Media Marketing*, Vol. 7, No. 2, pp. 137–147.

Loo, P.T. (2020), '"Exploring airline Companies' engagement with their passengers through social network: An investigation from their Facebook pages", *Tourism Management Perspectives*, Vol. 34, 100657. https://doi.org/10.1016/j.tmp.2020.100657

Ma, E., Cheng, M. and Hsiao, A. (2018), "Sentiment analysis – a review and agenda for future research in hospitality contexts", *International Journal of Contemporary Hospitality Management*, Vol. 30, No. 11, pp. 3287–3308. https://doi.org/10.1108/IJCHM-10-2017-0704

Mamula, M., Folgieri, R. and Duvnjak, K. (2019), "Some considerations about artificial neural networks in hotel industry: state of the art and future developments", in *ToSEE - Tourism in Sourthern and Eastern Europe*. Opatija: University of Rijeka, Faculty of Tourism and Hospitality Management, pp. 431–440. doi: 10.20867/tosee.05.8

Min, K.B., Song, S.H. and Min, J.Y. (2020), "Topic modeling of social networking service data on occupational accidents in Korea: Latent dirichlet allocation analysis", *Journal of Medical Internet Research*, Vol. 22, No. 8, e19222. https://doi.org/10.2196/19222

Ministry of Tourism and Sport of the Republic of Croatia (2020), *21 million tourists visit Croatia in 2019, a 5% increase*, viewed 16 May 2021, https://mint.gov.hr/news-11455/21-million-tourists-visit-croatia-in-2019-a-5-increase/20800

Monem, N.H.A. (2021), "The effectiveness of Advertising Personalization: a consumers' psychology perspective", *Journal of Design Sciences and Applied Arts*, Vol. 2, No. 1, pp. 114–121. https://doi.org/10.21608/jdsaa.2021.31121.1061

Pearce, P. L. and Wu, M. Y. (2018), "Entertaining International Tourists: An Empirical Study of an Iconic Site in China", *Journal of Hospitality and Tourism Research*, Vol. 42, No. 5, pp. 772–792. https://doi.org/10.1177/1096348015598202

Philander, K. and Zhong, Y.Y. (2016), "Twitter sentiment analysis: Capturing sentiment from integrated resort tweets", *International Journal of Hospitality Management*, Vol. 55, pp. 16–24. https://doi.org/10.1016/j.ijhm.2016.02.001

Priyantina, R.A. and Sarno, R. (2019), "Sentiment Analysis of Hotel Reviews Using Latent Dirichlet Allocation, Semantic Similarity and LSTM", *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 4, pp. 142–155. https://doi.org/10.22266/ijies2019.0831.14

Singh, N. K., Tomar, D. S. and Sangaiah, A.K. (2020), "Sentiment analysis: a review and comparative analysis over social media", *Journal of Ambient Intelligence and Humanized Computing*, Vol. 11, No. 1, pp. 97–117. https://doi.org/10.1007/s12652-018-0862-8

Sutherland, I., Sim, Y., Lee, S. K., Byun, J. and Kiatkawsin, K. (2020), "Topic Modeling of Online Accommodation Reviews via Latent Dirichlet Allocation", *Sustainability*, Vol. 12, No. 5, 1821. https://doi.org/10.3390/su12051821

Tirunillai, S. and Tellis, G.J. (2014), "Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation", *Journal of Marketing Research*, Vol. 51, No. 4, pp. 463–479. https://doi.org/10.1509/jmr.12.0106

Tsai, C.F., Chen, K., Hu, Y.H. and Chen, W.K. (2020),"Improving text summarization of online hotel reviews with review helpfulness and sentiment", *Tourism Management*, 80, 104122. https://doi.org/10.1016/j.tourman.2020.104122

Xiang, Z., Du, Q., Ma, Y. and Fan, W. (2017), "A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism", *Tourism Management*, Vol. 58, pp. 51–65. https://doi.org/10.1016/j.tourman.2016.10.001

Ye, F., Xia, Q., Zhang, M., Zhan, Y. and Li, Y. (2020), "Harvesting Online Reviews to Identify the Competitor Set in a Service Business: Evidence From the Hotel Industry", *Journal of Service Research*. https://doi.org/10.1177/1094670520975143

Jelena Musanovic, PhD, Postdoctoral Researcher
University or Rijeka, Faculty of Tourism and Hospitality Management, University of Rijeka
Department of Quantitative Economics
Primorska 46, 51410 Opatija, Croatia
+385 51 294 698
jelenak@fthm.hr

Raffaella Folgieri, PhD, Associate Professor
Università degli Studi di Milano
Department of Philosophy
Via Festa del Perdono 7, 20122 – Milano, Italia
+39250312739
raffaella.folgieri@unimi.it

Maja Gregoric, PhD, Assistant Professor
University or Rijeka, Faculty of Tourism and Hospitality Management, University of Rijeka
Department of Quantitative Economics
Primorska 46, 51410 Opatija, Croatia
+385 51 294 698
majam@fthm.hr