# THE APPLICATION OF SENTIMENT ANALYSIS IN TOURISM RESEARCH: A BRIEF REVIEW

**[1]Rohani Hashim, [1]Bahiyah Omar, [2]Noria Saeed Abdullah Ba-Anqud & [2]Hosam Al-Samarraie**

*[1]School of Communication, Universiti Sains Malaysia, Malaysia*
*[2]Centre for Instructional Technology & Multimedia, Universiti Sains Malaysia, Malaysia*
*Email: rohani@usm.my; bahiyah@usm.my; nooriasb@gmail.com; myclasy@gmail.com*

## ABSTRACT

Tourism plays a vital role in the development of any country's economy. With the use of Big Data in tourism research, it becomes essential to understand the methodological strengths and limitations in their area. This study reviewed previous studies on the use of sentiment analysis (SA), a computational method to extract people's opinions in social media, in the tourism sector. A total of 50 valid articles focusing on the use of sentiment analysis were identified. The results suggest that the most popular dataset used for analysis was TripAdvisor and the most commonly used sentiment classifications were Machine Learning and Lexicon-based approaches. The major issue for tourism-sentiment analysis research was the use of natural languages which limit the scope of analysis to mainly English texts. We recommend that future research to consider the use of sentiment analysis for evaluating the quality of service and level of satisfaction a visitor may experience when visiting a place.

**Keywords:** sentiment analysis, tourism, opinion mining, tourist experience

## INTRODUCTION

Tourism is an important economic activity in most countries around the world as it creates far-reaching impact on the economy, social and cultural development. The importance of tourism is acknowledged by not only industry practitioners but also the academic community. Research in tourism, in fact, has undergone several stages of development. One important development was the shift from delivery of 'tourism products' to the provision of 'tourism experience' (Kim, Ritchie, & Tung, 2010; Kerdpitak et al., 2019). Many scholars have agreed that visitors' experience of a place is the essence in tourism (e.g., (Binkhorst & Den Dekker, 2009), and, hence, it has received a considerable attention from tourism researchers. The review of previous research on tourism experience revealed different dimensions with regard to various forms of tourist behaviour. For example, Ritchie and Hudson (2009) identified five major categories of tourism experience research, namely (1) conceptualization, (2) experience behaviour, (3) methodologies, (4) types of experience and (5) managerial concerns. They found that most research interest focuses on the experiential effects and types of experience categories, while low attention was given on issues related to use of certain methodologies in analysing visitors' experience of a place.

In this paper, we aim to provide an overview of the current use of sentiment analysis (SA) in analysing tourists experience in the social media era. Sentiment analysis, also known as opinion mining, is a computational process of identifying and categorizing opinions expressed in a piece of written text (Sabatovych, 2019). Sentiment analysis is commonly used to extract

 

sentiments, opinions, and subjectivity in unstructured texts, that is, to identify whether the expressions indicate positive or negative opinions toward the subject (Pang & Lee, 2008). This method has been commonly used by scholars in different contexts, particularly to determine whether individuals' attitude toward a particular topic, product or service is positive, negative, or neutral. Sentiment analysis has become a research tool to understand people's opinion ranging from document level classification to learning the polarity of words and phrases (Kouloumpis, Wilson, & Moore, 2011). Its application in mapping social media users' opinions has been increasingly recognized and become a prevalent culture in the online space. Given the enormous volume of social media data, many researchers have turned to this platform to study people opinions on numerous issues and topics such as politics (Godbole, Srinivasaiah, & Skiena, 2007), education (Munezero, Montero, Mozgovoy, & Sutinen, 2013), and tourism (Alaei, Becken, & Stantic, 2019).

In tourism research, sentiment analysis is often used to determine visitors' attitudes towards the visited places and services by analyzing the big data available online. Social media, professional travel blogs, and websites are the main data sources used by researchers to extract tourists' opinions or thoughts about their visits to certain places (Gräbner, Zanker, Fliedl, & Fuchs, 2012). The importance of sentiment analysis in tourism research is evident as the current trend also shows that people have increasingly relied on 'TripAdvisor' and 'blogs' to decide the must-visit tourist destinations; and in making minor decisions such as choosing the places to stay or eat. According to Kirilenko, Stepchenkova, Kim, and Li (2018), it is not always suitable to use traditional sampling and statistical analysis methods for analyzing tourism data. This is because a diverse set of methods can be used differently to map certain patterns in a large data set. Gao, Hao, and Fu (2015) addressed the potential of studying the role of different sentiment analysis methods and compare their performance when they are used to classify texts in tourism. Nevertheless, sentiment analysis is still facing considerable challenges with regard to its application in the tourism sector (Mkono & Tribe, 2017; Costa et al., 2019). Schuckert, Liu, and Law (2015) argued that online reviews can be placed indifferent ways and generate powerful word-of-mouth online (Schuckert et al., 2015). Many sentiment classification techniques have been developed for English, Japanese, and Chinese languages, but the interest in the sentiment analysis is worldwide. Thus, to provide the necessary support for various natural language applications, researches on automatic sentiment analysis should be conducted in more languages (Zhang, Ye, Zhang, & Li, 2011).

Here, the main aim of this study was to provide a review of past studies that applied sentiment analysis in tourism research. The objectives were to demonstrate a global view about the role of sentiment analysis in tourism research. These include providing a classification of the types of datasets, search queries and sentiment analysis classifications in this field. The reminder of this paper is structured as follows: the second section explains the method used to identify eligible original research articles for analysis; the third section presents the study findings according; and the fourth section provides some recommendations for future research.

## METHODOLOGY

The initial search resulted in 351 articles. We removed 46 articles that were labelled as duplicates. The number of the retrieved articles after removing duplications was 305. After screening the titles and abstracts of the 305 articles, we further excluded 200 articles. The full texts of the remaining 105 articles were checked against the inclusion and exclusion criteria.

This resulted in removing 55 articles, thus leaving us with only 50 articles (see Figure 1). We reviewed 50 journal articles on tourism that used sentiment analysis as their research method. In our searched for the articles, we used different databases such as IEEE, Springer, Scopus, ResearchGate, Science Direct, and SAGE. We used different related keywords, such as "tourism sentiment analysis," OR "sentiment analysis in tourism," OR "text data in tourism," OR "sentiment data on hotel reviews," OR "sentiment analysis of restaurant reviews" to identify the relevant articles.

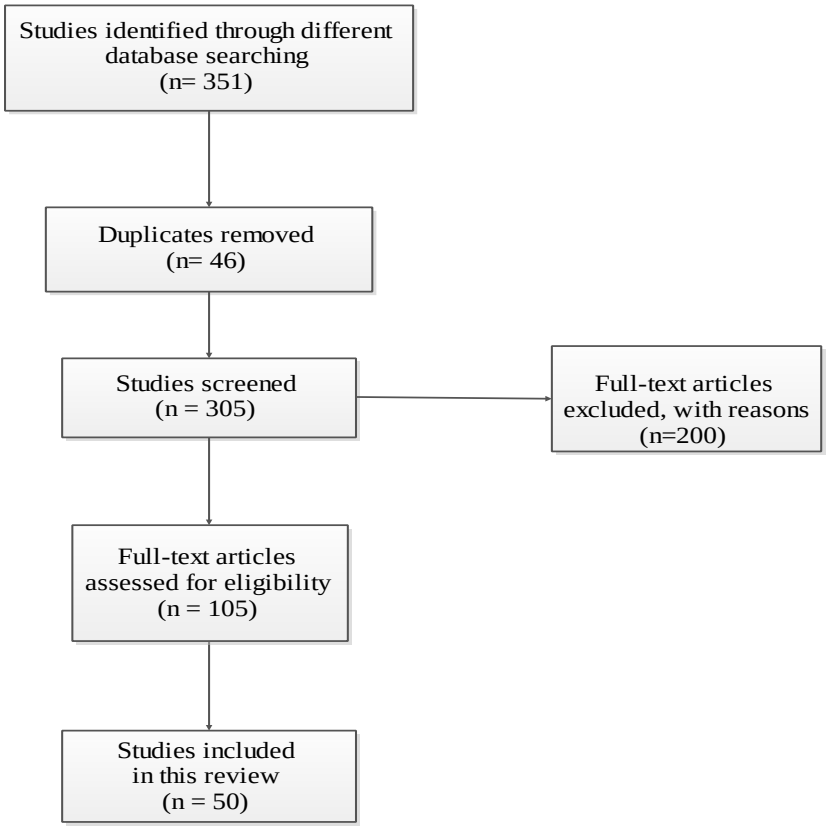


Figure 1: Articles selection phase

## RESULTS

### 3.1 Global use of sentiment analysis in tourism research

We found that 46 percent of researches on tourism-sentiment analysis were conducted in Europe, followed by Asia (21 percent), North America (8 percent), South America (7 percent), Australia (4 percent) and Africa (3 percent). A summary of a global used of sentiment analysis in tourism researches is shown in Figure 2.

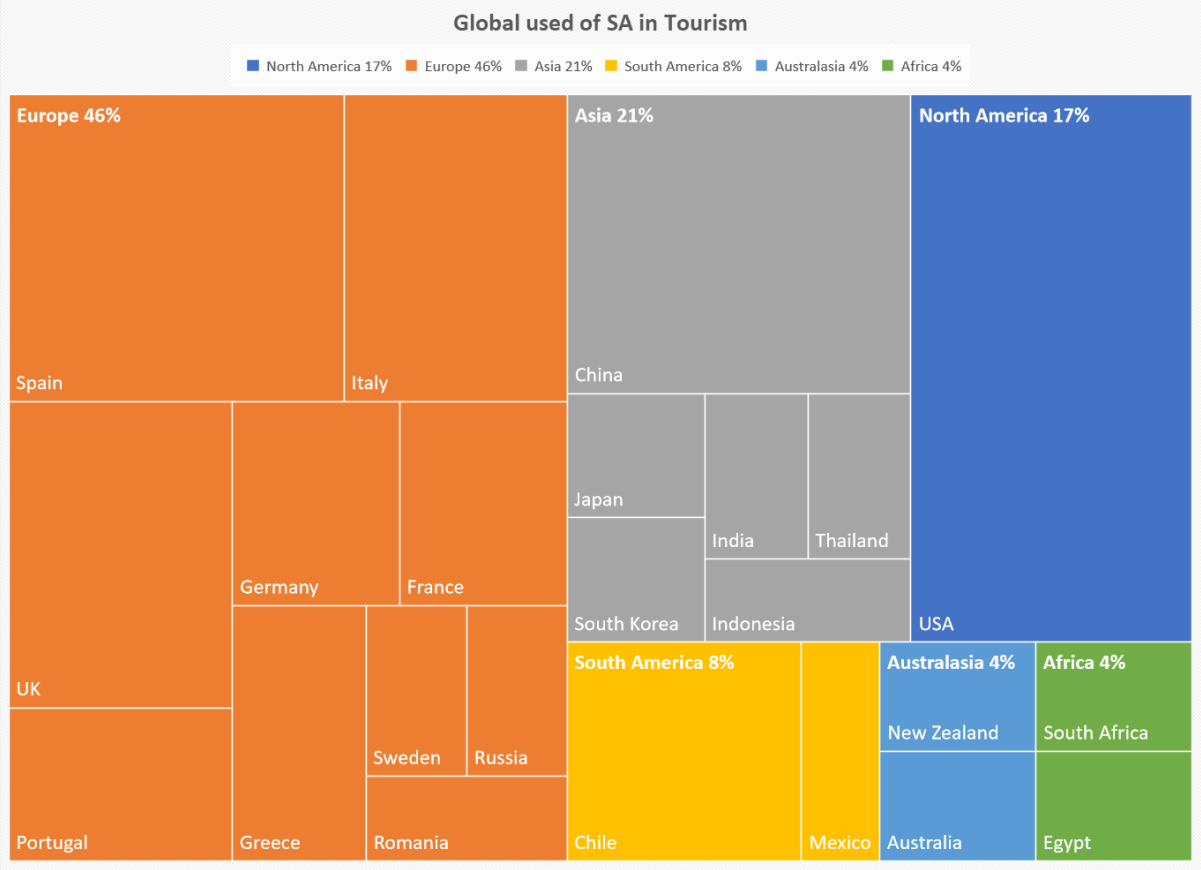


Figure 2. Global use of sentiment analysis (SA) in tourism research

### 3.2. Datasets for tourism research

The results showed different datasets for sentiment analysis purposes were used in tourism studies (see Figure 3). Our findings suggest that previous studies obtained their data from: official and professional traveling websites (e.g., TripAdvisor, Expedia, Ctrip, and Booking), and social network posts (e.g., Twitter or travel blogs). We categorized the dataset usage in the tourism industry into six main dimensions: hotel booking (Shi & Li, 2011; Xiang, Schwartz, Gerdes Jr, & Uysal, 2015), travel blog (Rahmani, Gnoth, & Mather, 2018; Volo, 2010), travel website (García-Pablos, Cuadros, & Linaza, 2015; Gonzalez-Rodriguez, Martínez-Torres, & Toral, 2014), TripAdvisor (Hwang, Lai, Chang, & Jiang, 2015), Twitter (Meehan, Lunney, Curran, & McCaughey, 2013) and others (Kasper & Vela, 2011). The results showed that TripAdvisor was commonly used by researchers to extract visitors' sentiments on their experiences at tourist destinations. The details of the six main parts are as follows:

1. TripAdvisor is a professional travel website used by people from different countries (Bjørkelund, Burnett, & Nørvåg, 2012; Bucur, 2015; Putri & Kusumaningrum, 2017; Rossetti, Stella, & Zanker, 2016), and it alone accounts for 455 million unique visitors per month. This platform provides over 570 million reviews and opinions covering the world's largest selection of travel listings on aspects related to accommodations, airlines, attractions, and restaurants. Additionally, TripAdvisor allows visitors to compare between more than 200 hotel booking sites.
2. Twitter is a social networking service in which users post and interact with messages known as "tweets". Tweets are originally restricted to 140 characters, but since November 2017, this limit was doubled to users in all languages except Chinese, Japanese, and Korean. Many

previous studies (e.g., Claster, Dinh, and Cooper (2010); Shimada, Inoue, Maeda, and Endo (2011); Barbagallo, Bruni, Francalanci, and Giacomazzi (2012), and Serna, Gerrikagoitia, and Bernabé (2016) reported that Twitter is the greatest dataset source for Application Programming Interface (API) that can be effectively used for different sentiment analysis purposes. The users of this platform usually use keywords to retrieve information about hotels, accommodations, places, city attractions, restaurants, food, and weather conditions.

3. Travel website is a collection of related network web resources, such as web pages, multimedia content, travel services, and social community that are typically identified with a common domain name, and published on at least one web server. Based on our review (e.g., Tan and Wu (2011); García-Pablos et al. (2015); Xiang, Du, Ma, and Fan (2017), most travel websites are used to collect data related to tourism and sentiment analysis. Several categories are included, such as Yelp, Ctrip, Google Place, and Open Rice (food and restaurant website).
4. Hotel booking consists of all professional websites or apps that are used to manage hotel and decide bookings such as booking.com (Schmunk, Höpken, Fuchs, & Lexhagen, 2013), expedia.com, orbitz.com (Duverger, 2013), and hotels.com (Duan, Cao, Yu, & Levy, 2013).
5. Travel blog is a free online travel diary for travelers across the world. It provides an easy access to the users to various reviews about certain experiences from internet cafes and computers worldwide. In our categories it included all information, experience, and recommendations that are written by travelers about food, favorite place, and accommodation. From our review travelblog is the least dataset used by the researchers for collecting data.
6. Others, here, included all the apps and websites which are not often used as a dataset for sentiment analysis such as Facebook, Instagram, and YouTube (Oliveira & Panyik, 2015).

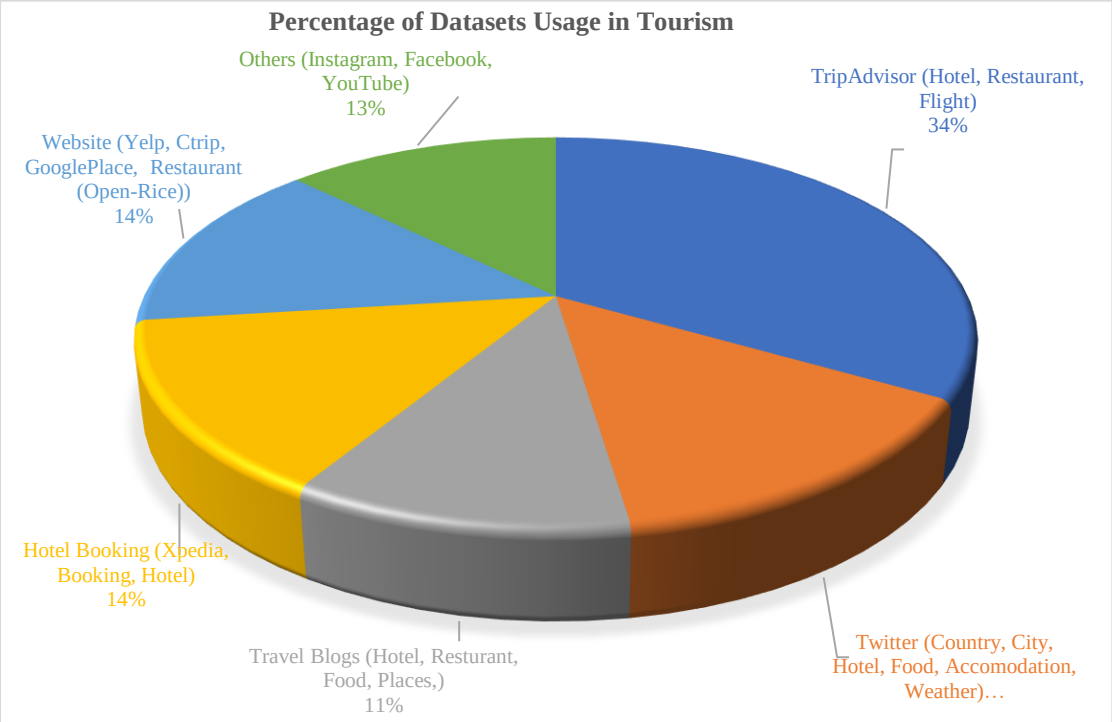


Figure 3. Datasets usag

### 3.3 Search queries in tourism

The search queries used in most previous studies on tourism were mostly related to hotels, restaurants, accommodations (e.g., villa, resort, apartment, cottage, and cabin), food, holiday, locations, rooms, attractions, old town, beach, scenery, relaxation, nature, mountain, architecture, city walk, ambience, and culture. A small number of researchers focused on restaurants reviews, such as Ganu, Elhadad, and Marian (2009) who identified text classification-strategies as the means to allow users to get better recommendations on specific aspects of restaurants search such as food or ambience, and soft clustering-based approaches that group users based on their reviewing styles and interest similarities.

### 3.4 Sentiment analysis classifications

Our findings showed three types of sentiment analysis classifications: i) Machine Learning approach, ii) Lexicon-based approach, and iii) Hybrid approach, which were commonly used in tourism research. Descriptions of the three approaches are as follows:

i) Machine Learning refers to the use of scientific algorithms and statistical models in order to effectively perform a specific task without using explicit instructions. It is divided into supervised learning, unsupervised learning, and semi-unsupervised learning. Support vector machine and Naïve Bayes are the standard classifiers used in most tourism-based studies.
ii) Lexicon-based approach consists of dictionary based and corpus based which is also divided into statistical and semantic. Based on our review, dictionaries for Lexicon-based approach can be created manually as realized by Shimada et al. (2011) who produced query list, namely the union of basic queries and related words. In addition, Hwang et al. (2015) proposed an approach to automatically identifying customer reviews on hotel management by using SentiWordNet. These approaches were found to offer an in-depth understanding of visitors' opinions of a place.
iii) Hybrid approach is another classification method that combine machine learning and lexicon-based methods.

These three classifiers were commonly used in tourism research (Serna et al., 2016; Xiang et al., 2017), mainly due to their abilities to provide a better prediction accuracy of opinions (Broß, 2013; Kirilenko, Stepchenkova, Kim, & Li, 2017).

### 3.5 Issues in tourism-sentiment analysis research

In general, sentiment analysis research is concerned about analyzing texts using natural language other than English. This is because English has the richest set of sentiment analysis resources such as SentiWordNet (Bucur, 2015; Hwang et al., 2015), while in reality many different languages are used to express views and opinions on social media. Hence, there is a strong demand for sentiment analysis for other languages besides English.

Our review of previous works revealed different issues with regard to the use of sentiment analysis methods in tourism research. For example, Kasper and Vela (2011) addressed the needs for providing accurate dictionary tools for collecting users' comments written in German language. Similarly, other previous studies found it difficult to specify the type of service when filtering visitors' comments (Zhang et al., 2011). This, as a result, led many previous studies to propose different solutions to overcome these issues. For example, Hassan, Bakr, and Ziedan

(2018) presented a novel framework for Arabic concept-level sentiment analysis using SenticNet. Similarly, ElSahar and El-Beltagy (2015) presented an approach for detecting subjective slang terms with the aim of building Arabic slang lexicons without depending on any language-dependent part of speech taggers or parsers. The proposed technique showed that it was able to deal with the evolving nature of the language used in social media. Molina-González, Martínez-Cámara, Martín-Valdivia, and Urena-López (2014) described a new corpus for sentiment analysis composed by hotel reviews written in Spanish. They used the corpus COAH to carry out a set of experiments for unsupervised polarity detection using different lexicons. Moreover, Shimada et al. (2011) proposed a tourism information analysis system for a local city using the manually produced query list, namely the union of basic queries and related words in Japanese language. Our review of previous studies showed that various forms of sentiment analysis models have been developed in natural languages (ElSahar & El-Beltagy, 2014; García, Gaines, & Linaza, 2012; Molina-González et al., 2014; Shimada et al., 2011; Zhang et al., 2011). This is mainly because natural languages allow the use of linguistic tools to extract textual information for sentiment analysis. This means that the current use of sentiment analysis of natural language texts is growing, and can potentially offer a better performance result in tourism research (Calheiros, 2015).

## CONCLUSION

This study concludes that using sentiment analysis in tourism research can open new and interesting avenues for the development of tourism programs, especially in developing countries. The review results showed that using sentiment analysis on social media sites can help decision makers understand the sophisticated nature of tourists' opinions. In addition, future development can be expected regarding the application of certain classification methods. This include the language used for processing and filtering visitors' comments in the tourism sector (Alaei et al., 2019). This review found a common issue in the use of natural language processing in tourism research, particularly in the use lack of linguistic resources other than English (Molina-González et al., 2014). The lack of resources to carry out accurate sentiment classification was also found to influence the prediction performance of visitors' opinions (Ravishankar, 2017). Since, sentiment analysis requires deep analysis in natural language processing, it is important to design methods that can perform precise sentiment classifications in natural language settings. Overall, this review on tourism-sentiment analysis researches concludes that using sentiment analysis approach will help integrating a great potential for generating future insights at scales not seen before practically in tourism. It is recommended that future research to consider the use of sentiment analysis for evaluating the quality of service and level of satisfaction a visitor may experience when visiting a place.

## REFERENCES

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. *Journal of Travel Research, 58*(2), 175-191.

Barbagallo, D., Bruni, L., Francalanci, C., & Giacomazzi, P. (2012). *An empirical study on the relationship between twitter sentiment and influence in the tourism domain*: na.

Binkhorst, E., & Den Dekker, T. (2009). Agenda for co-creation tourism experience research. *Journal of Hospitality Marketing & Management, 18*(2-3), 311-327.

Bjørkelund, E., Burnett, T. H., & Nørvåg, K. (2012). *A study of opinion mining and visualization of hotel reviews.* Paper presented at the Proceedings of the 14th

International Conference on Information Integration and Web-based Applications & Services.

Broß, J. (2013). Aspect-oriented sentiment analysis of customer reviews using distant supervision techniques.

Bucur, C. (2015). Using opinion mining techniques in tourism. *Procedia economics and finance, 23*, 1666-1673.

Calheiros, A. C. d. S. (2015). *Sentiment analysis in hospitality using text mining: the case of a Portuguese eco-hotel.*

Claster, W. B., Dinh, H., & Cooper, M. (2010). *Naïve Bayes and unsupervised artificial neural nets for Cancun tourism social media data analysis.* Paper presented at the Nature and Biologically Inspired Computing (NaBIC), Second World Congress on 2010.

Costa, T., Umbelino, J., Calisto, L., Afonso, V. & Nunes, S**.** (2019). Residents Perceptions of Tourism Impact, A Tool for governance: A study on residents in the regional tourism area of Lisbon. *International Journal of Business Tourism and Applied Sciences.* 7(1). 10-21.

Duan, W., Cao, Q., Yu, Y., & Levy, S. (2013). *Mining online user-generated content: using sentiment analysis technique to study hotel service quality*. Paper presented at the System Sciences (HICSS), 46th Hawaii International Conference on 2013.

Duverger, P. (2013). Curvilinear effects of user-generated content on hotels' market share: a dynamic panel-data analysis. *Journal of Travel Research, 52*(4), 465-478.

ElSahar, H., & El-Beltagy, S. R. (2014). *A fully automated approach for arabic slang lexicon extraction from microblogs.* Paper presented at the International conference on intelligent text processing and computational linguistics.

ElSahar, H., & El-Beltagy, S. R. (2015). *Building large arabic multi-domain resources for sentiment analysis.* Paper presented at the International Conference on Intelligent Text Processing and Computational Linguistics.

Ganu, G., Elhadad, N., & Marian, A. (2009). *Beyond the stars: improving rating predictions using review text content.* Paper presented at the WebDB.

Gao, S., Hao, J., & Fu, Y. (2015). *The application and comparison of web services for sentiment analysis in tourism.* Paper presented at the 2015 12th International Conference on Service Systems and Service Management (ICSSSM).

García-Pablos, A., Cuadros, M., & Linaza, M. T. (2015). OpeNER: open tools to perform natural language processing on accommodation reviews. In *Information and Communication Technologies in Tourism 2015* (pp. 125-137): Springer.

García, A., Gaines, S., & Linaza, M. T. (2012). A lexicon based sentiment analysis retrieval system for tourism domain. *Expert Syst Appl Int J, 39*(10), 9166-9180.

Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. *Icwsm, 7*(21), 219-222.

Gonzalez-Rodriguez, M., Martínez-Torres, M. R., & Toral, S. (2014). *Monitoring travel-related information on social media through sentiment analysis.* Paper presented at the Proceedings of the 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing.

Gräbner, D., Zanker, M., Fliedl, G., & Fuchs, M. (2012). *Classification of customer reviews based on sentiment analysis*: Citeseer.

Hassan, H. G., Bakr, H. M. A., & Ziedan, B. E. (2018). A Framework for Arabic Concept-Level Sentiment Analysis using SenticNet. *International Journal of Electrical and Computer Engineering, 8*(5), 4015.

Hwang, S.-Y., Lai, C.-Y., Chang, S., & Jiang, J.-J. (2015). The identification of noteworthy hotel reviews for hotel management. *Pacific Asia Journal of the Association for Information Systems, 6*(4).

Kasper, W., & Vela, M. (2011). *Sentiment analysis for hotel reviews.* Paper presented at the Computational linguistics-applications conference.

Kerdpitak, C., Hotrawaisaya, C., Khaengkhan, C. (2019). Assisting Tourism Supply Chain Performance in Thailand through Big Data Analytics: Moderating Role of IT Capability. *Internatioanl Journal of Supply Chain Management*. 8(6), 189-197.

Kim, J.-H., Ritchie, J., & Tung, V. W. S. (2010). The effect of memorable experience on behavioral intentions in tourism: A structural equation modeling approach. *Tourism Analysis, 15*(6), 637-648.

Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. (2017). Automated sentiment analysis in tourism: Comparison of approaches. *Journal of Travel Research*, 0047287517729757.

Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. (2018). Automated sentiment analysis in tourism: Comparison of approaches. *Journal of Travel Research, 57*(8), 1012-1025.

Kouloumpis, E., Wilson, T., & Moore, J. (2011). *Twitter sentiment analysis: The good the bad and the omg!* Paper presented at the Fifth International AAAI conference on weblogs and social media.

Meehan, K., Lunney, T., Curran, K., & McCaughey, A. (2013). *Context-aware intelligent recommendation system for tourism.* Paper presented at the Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on.

Mkono, M., & Tribe, J. (2017). Beyond reviewing: Uncovering the multiple roles of tourism social media users. *Journal of Travel Research, 56*(3), 287-298.

Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Urena-López, L. A. (2014). *Cross-domain sentiment analysis using Spanish opinionated words.* Paper presented at the International Conference on Applications of Natural Language to Data Bases/Information Systems.

Munezero, M., Montero, C. S., Mozgovoy, M., & Sutinen, E. (2013). *Exploiting sentiment analysis to track emotions in students' learning diaries.* Paper presented at the Proceedings of the 13th Koli Calling International Conference on Computing Education Research.

Oliveira, E., & Panyik, E. (2015). Content, context and co-creation: Digital challenges in destination branding with references to Portugal as a tourist destination. *Journal of Vacation Marketing, 21*(1), 53-74.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval, 2*(1–2), 1-135.

Putri, I., & Kusumaningrum, R. (2017). *Latent Dirichlet Allocation (LDA) for Sentiment Analysis Toward Tourism Review in Indonesia.* Paper presented at the Journal of Physics: Conference Series.

Rahmani, K., Gnoth, J., & Mather, D. (2018). A Psycholinguistic View of Tourists' Emotional Experiences. *Journal of Travel Research*, 0047287517753072.

Ravishankar, N. (2017). *Grammar Rule Based Sentiment Analysis Techniques for Tamil Tweets Classification.* BS ABDUR RAHMAN UNIVERSITY,

Ritchie, J. B., & Hudson, S. (2009). Understanding and meeting the challenges of consumer/tourist experience research. *International Journal of Tourism Research, 11*(2), 111-126.

Rossetti, M., Stella, F., & Zanker, M. (2016). Analyzing user reviews in tourism with topic models. *Information Technology & Tourism, 16*(1), 5-21.

Sabatovych, I. (2019). Use of Sentiment Analysis for Predicting Public Opinion on Referendum: A Feasibility Study. *The Reference Librarian*, 1-10.

Schmunk, S., Höpken, W., Fuchs, M., & Lexhagen, M. (2013). Sentiment analysis: Extracting decision-relevant knowledge from UGC. In *Information and Communication Technologies in Tourism 2014* (pp. 253-265): Springer.

Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing, 32*(5), 608-621.

Serna, A., Gerrikagoitia, J. K., & Bernabé, U. (2016). Discovery and classification of the underlying emotions in the user generated content (UGC). In *Information and communication technologies in tourism 2016* (pp. 225-237): Springer.

Shi, H.-X., & Li, X.-J. (2011). *A sentiment analysis model for hotel reviews based on supervised learning.* Paper presented at the Machine Learning and Cybernetics (ICMLC), 2011 International Conference on.

Shimada, K., Inoue, S., Maeda, H., & Endo, T. (2011). *Analyzing tourism information on twitter for a local city.* Paper presented at the 2011 First ACIS International Symposium on Software and Network Engineering.

Tan, S., & Wu, Q. (2011). A random walk algorithm for automatic construction of domain-oriented sentiment lexicon. *Expert Systems with Applications, 38*(10), 12094-12100.

Volo, S. (2010). Bloggers' reported tourist experiences: Their utility as a tourism data source and their effect on prospective tourists. *Journal of Vacation Marketing, 16*(4), 297-311.

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management, 58*, 51-65.

Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management, 44*, 120-130.

Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications, 38*(6), 7674-7682.