Original Scientific Article

Dubai Restaurants: A Sentiment Analysis of Tourist Reviews

Vinaitheerthan Renganathan
Banasthali Vidyapith, India
vinairesearch@yahoo.com

Amitabh Upadhya
American College of Dubai, UAE
upadhyaamitabh@gmail.com

An enormous amount of information is available on innumerable travel websites, social media and blogs, of which a large part is user-generated content. This web content holds great potential to assess visitor sentiment at a destination; as this identifies a need for building automated systems to extract unknown sentiments from these sources. Sentiment analysis, which includes text mining and natural language processing (NLP) techniques, helps in extracting related sentiments from the data thus stored, in unstructured formats. The extracted sentiment would facilitate better tourist decision making and improve customer service and new product development for tourism enterprises. This study presents a sentiment analysis model to extract the hidden sentiments from tourist reviews about restaurants in Dubai that will guide visitors to the city in taking suitable dining decisions. Sentiment analysis is carried out by extracting tourist reviews about restaurants in Dubai using a web scraping method using text mining techniques with the help of the R statistical software package. The resultant data is further analysed by sentiment analysis tools to extract the hidden sentiments, which are categorized under eight heads. The sentiment analysis helped uncover hidden sentiments along with the frequency of each sentiment category. It also helped to find the difference between tourist sentiment scores with respect to different categories of restaurants. The paper provides a sentiment analysis model which can be used in the future to extract the reviews related to other tourism products besides restaurants, such as accommodation, attractions and accessibility.

Keywords: tourist reviews, Dubai restaurants, sentiment analysis, text mining, R statistical package

CC BY-SA https://doi.org/10.26493/2335-4194.14.165-174

Introduction

The internet is now a necessary source of personal and professional information and the brisk-paced evolution of information and communication technologies (ICT) has given rise to Web 2.0 characterized by participatory contribution or user-generated content (UGC), or electronic word of mouth (e-WOM). Businesses hitherto enjoyed a monopoly on the information they possessed; users themselves now determine the information they want to see and to consume

(Breda et.al., 2020). According to the International Telecommunication Union there are approximately three and a half billion people, or 47% of the world population, that use the internet, in turn significantly impacting various sectors of the economy and society including tourism (Buhalis & Law, 2008). Information and communication technology (ICT) and the internet have changed the way individuals and organizations in the tourism sector operate today (Boyer, 2014; Mariani et al., 2014). There are various internet applications such as search engines, social media websites, blogs and review sites that have profound influence on tourist decision making in terms of choice of destination, accommodation and mode of travel (Xiang et al., 2015). There are several online interfaces which enable tourists to share their experiences in the form of text, images and videos. This vast user-generated content is available online in the form of user reviews, comments, feedbacks, messages, posts, and tweets, providing opportunities for better decision making for the stakeholders, especially in the tourism sector, but which cannot be analysed manually and require automated tools like text mining (Hearst, 1992; 2003). Text mining (Renganathan, 2017) involves extracting and analysing information from the unstructured data such as text, opinions, reviews, and comments that is not possible with the traditional statistical tools.

Natural language processing (NLP) (Manning et al., 1999) helps to enable computer systems read and understand natural languages such as English. Sentiment analysis (Jiang et al., 2021; Artemenko et al., 2020; Saad & Aref, 2020) enables understanding of different emotions, attitudes and expressions contained in the textual information using text mining and NLP techniques. Sentiment analysis, or opinion mining, (Pang & Lee, 2008) helps in extracting the hidden sentiments or opinion from the unstructured data using tools such as text mining and natural language processing. This study provides an overview of sentiment analysis, text mining and natural language processing in the tourism sector and builds a sentiment analysis model using a lexicon-based approach (Balasubramanian et al., 2021; Bose et al., 2020). Open source software – R statistical package (see https://www.r-project.org) was used to build the model.

Literature Review

The application of 'text mining' is a tool which is being used in many tourism researches (Thomaz et al., 2016) in the areas of destination branding, destination characteristics, sentiment analysis, tourist online behaviour, tourist purchasing decisions and tourism sector marketing strategies. Tourism can be termed as a product which is intangible, experiential and perishable (Xiang et al., 2015). Similarly, tourism can be defined as a product which exists in the form of information before a tourist makes a purchase decision (Doolin et al., 2002). Therefore, the online medium which acts as a mode of communication providing a platform for the tourism industry in the fields of marketing the tourism product and services (Carson, 2006) also helps in the formation of tourist opinions that have greater influence on their purchasing decisions (Cohen et al., 2014; Litvin et al., 2008).

The online behaviour of tourists can be divided into pre, onsite and post visits to the desired destination. Tourists share their experiences, opinions, comments and suggestions online after the visit, which might be positive, neutral or negative (Kim et al. 2017). Online media, including online review websites, blogs and social networks, enable tourists to share travel experiences in the form of posts, comments, opinions, photos and videos (Xiang & Gretzel, 2010; Law et al., 2017) which in turn become a source of information for future tourists to plan their travels and purchase the tourism products.

Research related to the influence of social media on tourist online behaviour shows that around 46% of tourists shared their travel related experiences on social media and 36% of tourists' choice of destination is influenced by social media posts (Thomaz et al., 2016). Tourist purchasing decisions are influenced by the opinions expressed by fellow tourists who share their experience on tourism products such as destination, accommodation and travel (Godnov & Redek, 2016).

Text mining tools which act as a base for opinion mining enable the study of opinions and sentiments expressed by the tourist (Pang & Lee, 2008; Ye et al., 2009). Opinion mining classifies the text into positive, neutral and negative classes wherein the text classifies

the text into two to thousand different classes (Pang & Lee, 2008).

Information about the destination aids the tourist in understanding the characteristics of the destination (Pang et al., 2011) which they intend to visit in the near future. A vast amount of online information related to a destination is generated by tourists who visited the destination recently. Text mining helps in the study of user-generated content (Choi et al., 2007; Pang et al., 2011; Xue, 2013) and helps the tourism sector as a whole to study the impact of user-generated content in the growth of the particular destination. Text mining models are also used to study destination-specific information from the travelogues (Hao et al., 2010). Text mining models can be used as a decision support system which helps travel and tourist agents to analyse the interesting comments given by the tourist online (Loh et al., 2003). Similarly, it also helps management in the hospitality sector to develop strategies to improve their services and increase occupancy rates by analysing tourist opinion queries from future tourists who are about to visit the destination, expressed in online platforms including the newsgroups postings (Lau et al., 2005; Xiang & Pan, 2011; Qi & Ning, 2017).

The text mining process is divided into the following phases: searching and retrieving the set of documents on a given topic, creating a document corpus, stop word removal, stemming, creating a term document matrix, clustering of documents, finding association between documents and creation of a word cloud (Salton & McGill, 1983; Aggarwal & Zhai, 2012; Vijayarani et al., 2015).

Phase I of the text mining process involves the searching and retrieving of documents which contain comments, opinions and suggestions using an information retrieval process based on the information required by the users (Salton & McGill, 1983).

Phase II of text mining includes pre-processing of documents which involves the removal of stop words present in the documents such as 'and,' 'the' and 'an,' etc. (Vijayarani et al., 2015). Stop words are removed from the documents using methods such as the classic method and Zip's law wherein the former method removes the predefined stop words and the latter method removes the words with high Term Frequency – Inverse Document Frequency (TF – IDF) value (Salton & Buckley, 1988). The TF – IDF of a term is an important measure in the text mining process which is defined as follows:

1. TF – IDF = Frequency $(i) \times N/f(i)$.
2. Term Frequency = Number of times the term appears in the document in comparison with total terms in the document.
3. Inverse document frequency = Total number of documents/number of documents containing the term in consideration.

Phase III of the text mining process includes stemming, which helps to identify the root of each term and where each term is replaced by its root term. For example, 'happiness' or 'happily' is replaced with its root word 'happy.'

Phase IV of the text mining process involves preparation of a term document matrix wherein the rows present the terms and columns represent the document. For example, if the word 'Dubai' appears 17 times in a traveller's blog article on different dates and there were 50 dates of blog articles that were considered for the text mining analysis, and out of the 50 documents, 48 contain the term Dubai then:

$$\text{TF} - \text{IDF (Dubai)} = 15 \times 50/48 = 15.625.$$

The web scraping technique enables the extraction of the content from the webpages embedded in HyperText Markup Language (HTML) tags and store it in text format (Prameswari et al., 2017).

Natural language processing (Pang & Lee, 2008) helps in understanding the interaction between the computer systems and the human language such as English. Natural language processing techniques involve studying syntactic (grammar), morphological (different forms of words), semantic (meaning) and pragmatic (context) aspects within a given text. There are different approaches, such as statistical, rule based, linguistic or mixed, used in the field of NLP.

Natural language processing tools are used in the tourism sector (Pekar & Ou, 2008; Özen and Ilhan, 2020) to obtain tourist evaluations of services and products offered by hotels and restaurants from the reviews available in the online medium.

Sentiment analysis helps in understanding sentiments from the user-generated content (UGC) (Gräbner et al., 2012; Schmunk et al., 2013; Calheiros et al., 2017; Chen et al., 2020) in the form of opinions, views and comments available in various online platforms such as social media, groups and blogs, using text mining and natural language processing techniques.

Sentiment analysis is generally carried out using machine learning (ML) (Duong & Nguyen-Thi, 2021; Yi & Liu, 2020), deep learning (Li et al., 2020), lexicon-based and hybrid (combination of two methods) approaches. Each method has its own advantages and disadvantages (Divaka et.al., 2016).

The machine learning-based approach (Nehe et al., 2020) uses the train and test datasets to classify the text into positive and negative sentiments. It includes classifiers such as support vector machines (SVM) and Naive Bayes classifiers (Yusof et al., 2015; Alaei et al., 2019).

Deep learning methods (Karas & Schuller, 2021), which are similar to machine learning methods, are also used for sentiment analysis. Deep learning is more powerful in terms of classification accuracy (Zhang et al., 2018). It includes convolutional neural network (CNN), reinforcement learning, and long short term memory (LSTM) models for classification purposes.

The following are the advantages of machine learning (ML) and deep learning (DL) methods (Yi & Liu, 2020):

1. They are faster to implement.
2. They can handle large volumes of data sets.
3. Training accuracy increases with the increase of dataset size.

The following are the disadvantages of ML and DL methods:

1. They require the users to provide labels for the training data set in supervised learning models.
2. The model built on one domain may not be suitable for another domain.

The lexicon-based approach (Faheem et al., 2020; Yu et al., 2019) uses language dictionaries to classify the text into positive or negative sentiments. Following are the advantages of this method:

1. It attaches sentiment to each word.
2. It does not need any training dataset.
3. Easy to implement (Alessia et al., 2015).

The following are the disadvantages of the lexicon-based approach:

1. It is language specific.
2. If any sarcasm is present, it might not capture that.

This paper uses the lexicon-based approach in carrying out the sentiment analysis as the tourist reviews are collected in the English language. The National Research Council Canada (NRC) Word-Emotion Association Lexicon is used (Mohammad & Turney, 2013) in this paper.

Tourist reviews are available online at social media websites like Twitter and Facebook and also on popular sites like Tripadivisor.com, Expedia.com and Booking.com. An interesting and noteworthy example of sentiment analysis was carried out by Valdivia et al. (2017), uncovering users' sentiment about three well-known monuments in Spain: Alhambra, Mezquita Córdoba, and Sagrada Familia, with the help of user ratings available at tripadvisor.com. Also, Philander and Zhong (2016) captured tourist sentiments through their tweets on Las Vegas resorts.

Analysis of variance (ANOVA) is a statistical model which is used to find out whether the groups or categories in the study differ with respect to the outcome variable (Sun et al., 2020). A post hoc comparison test is used to test which groups differ among themselves (Chen & Scovino, 2020).

Methodology

The study aims to find the hidden sentiments within the tourist reviews on restaurant service and find whether any difference among restaurants exists based on the sentiment scores. The study also addresses the following research question:

RQ1 *Are there any significant differences among restaurant categories (Indian, Chinese, Italian, Middle East and Café Food restaurants) in terms of sentiment score?*

Based on the above research question the following null hypothesis and alternative hypothesis are formed which will be tested using the analysis of variance method.

H0 *There is no significant difference among restaurant categories in terms of sentiment scores.*

H1 *There is significant difference among restaurant categories in terms of sentiment scores.*

The study involves extracting tourist reviews from tourist review websites. The extracted reviews are then parsed and converted into documents. The documents are further analysed to find the hidden sentiments in the reviews and sentiment scores are computed from the analysis. The study also focuses on finding whether any significant difference between restaurants exists in terms of sentiment scores. The study includes tools such as web scraping, text mining and sentiment analysis as the three methods are related and form the basis for the other method. Web scraping is required to extract the text from online websites, text mining tools are required to parse the texts and sentiment analysis tools are required to extract the sentiment from the text.

The reviews of tourists about restaurants in Dubai are extracted using a web scraping technique from www.tripadvisor.com for a period of three months. The sample included tourist reviews data from webpages related to different types of restaurants (Indian, Chinese, Café, Italian and Middle Eastern food-type restaurants).

The extracted text data was stored in a text file for further processing. The content of the text files was then fed into r environment using the 'readline' function.

The resultant text was converted into vector. The text data was preprocessed by removing the stop words, numbers, and punctuation using the tm_map function of the tm package. Here, a single sentence in the text document is treated as one single document for the analysis purpose. The model produced a word cloud output which is a graphical representation of terms present in the reviews, with the font of the words showing the frequency of occurrence. The resultant data is further analysed by sentiment analysis tools to extract the hidden sentiments, which are categorized under eight headings.

To carry out sentiment analysis, the following built-in packages were installed through RStudio environment: TM – text mining package, NLP – natural language processing package, Syuzhet – sentiment analysis package, and ggplot2 – graphical package (see https://www.rstudio.com).

Sentiments which are categorized into *positive, negative, anger, anticipation, disgust, fear, trust, sadness,* and *surprise* headings were extracted using the 'nrc' dictionary present in the Syuzhet package. The 'nrc' function created the sentiment score matrix which is used to find the difference in sentiments across different types of restaurants.

The analysis of variance model is used to test the difference between the sentiment scores among the restaurants and a post hoc comparison test – Tukey's HSD test – is applied to see which restaurants differ among them.

Results and Discussion

The sentiment analysis model provided a word cloud output which is given in Figure 1. The size of each word in the word cloud indicates the importance or frequency of each word appearing in the reviews given by the customers. Four out of five word clouds expressed positive sentiments whereas the Chinese food restaurants word cloud includes negative words like 'overpriced' and 'terrible.' The word clouds which are obtained above are in line with similar studies conducted on customer reviews with respect to restaurant quality (Kamerer, 2014; Gadidov & Priestley, 2018).

The sentiment analysis model provided the following sentiment score matrix (Table 1) with respect to sentiment type and restaurant food type, and the scores in each category are represented as percentages. The positive sentiments come out top in all the food categories, ranging from 32 to 34.54%, and negative sentiment ranged from 1.26% to 4.86%. Café food-type restaurant customers expressed positive sentiments (34.564%) compared to other types of customers. Chinese restaurant customers expressed the highest percentage of negative sentiments compared to other types of customers (2.21%). Previous researches also

Café Food Restaurant

Indian Food Restaurant

Middle Eastern Food Restaurant

Italian Food Restaurant

Chinese Food Restaurant

Figure 1 Word Clouds for Different Restaurant Types

obtained similar sentiment scores for the eight sentiment categories (Samuel et al., 2020; Ray et al., 2020).

The sentiment scores are further analysed using the analysis of variance method which is given in Table 2. The ANOVA method was used to test the hypothesis H1 that the sentiment scores differ with respect to restaurant category such as Indian, Chinese, Italian and Middle Eastern. The ANOVA method was also used to check whether there is any interaction effect in terms of type of sentiment and category of restaurants.

From Table 2, the ANOVA model indicates that the p-values of category of restaurants, sentiment type and interaction effect are less than 0.05. Hence we conclude that there is a difference within the sentiment types and type of restaurant category. There is also an interaction between type of sentiment and category of the restaurant (Qamar & Alassaf, 2020). Since the p-value for category of restaurant is less than 0.05, we will reject the null hypothesis (H0) but accept the alternative hypothesis (H1) that the category of restaurant differs with respect to sentiment score.

Since there is significant difference among the restaurant type in terms of sentiment scores, a multiple comparison test, Tukey's HSD test, is used to check which type of restaurants differ among them. The Tukey's HSD results on sentiment scores with respect to restaurant types are provided in Table 3. The p-values marked with (*) are statistically significant at 5% level of significance (Qamar & Alassaf, 2020).

From the above table we can infer that Chinese food-type restaurants differ with respect to Café food, Indian food, and Middle Eastern food restaurants. Similarly, Italian food restaurants differ with respect to Café food, and Indian food and Middle Eastern food differ with respect to Indian food as the p values are less than 0.05 ($p < 0.05$).

Table 1 Sentiment Scores with Respect to Restaurant Food Type Expressed in Percentage

Restaurant	Positive	Joy	Trust	Anticip.	Surprise	Negative	Sadness	Anger	Fear	Disgust
Café food	34.54	20.10	18.89	11.38	7.10	2.82	2.10	1.13	0.81	1.13
Chinese	32.51	18.75	17.99	11.17	6.06	4.86	2.46	2.21	2.27	1.70
Indian food	32.00	22.97	20.69	13.03	6.97	1.26	1.49	0.69	0.91	0.00
Italian	33.35	22.71	21.74	10.19	5.52	2.47	1.36	1.30	0.45	0.91
Middle Eastern	32.41	21.09	20.34	11.55	5.22	3.43	2.38	1.86	1.19	0.52
Overall	33.00	20.97	19.88	11.31	6.08	3.15	1.99	1.52	1.17	0.94

Table 2 Analysis of Variance

Variables	DF	Sum Sq	Mean Sq	*F* Value	Pr(>*F*)
Category of restaurant	4	1027	257	11.591	$4.18e^{-9}$
Sentiment type	9	66600	7400	334.124	$<2e^{-16}$
Category of restaurant × sentiment type	36	1345	37	1.687	0.008
Residuals	670	14839	22		

Table 3 The Multiple Comparison Test – Tukey's HSD Results

Type of restaurants	Upper value	Lower value	*p* adj.	Difference
Chinese–Café	2.300000	0.813638	3.786362	0.000254*
Indian food–Café	−0.968330	−2.544860	0.608191	0.447034
Italian–Café	2.013333	0.526972	3.499695	0.002127*
Middle Eastern–Café	0.686667	−0.799690	2.173028	0.713661
Indian food–Chinese	−3.268330	−4.844860	−1.691810	$2e^{-7}$*
Italian–Chinese	−0.286670	−1.773030	1.199695	0.984536
Middle Eastern–Chinese	−1.613330	−3.099690	−0.126970	0.025659*
Italian–Indian	2.981667	1.405142	4.558191	0.000003*
Middle Eastern–Indian	1.655000	0.078475	3.231525	0.034152*
Middle Eastern–Italian	−1.326670	−2.813030	0.159695	0.105744

Notes * Statistically significant at 5% level of significance.

Conclusion

This study provided an overview of sentiment analysis, text mining tools and natural language processing techniques in the tourism sector. The paper provided a base for analysing the sentiments of customers' perceptions about restaurant service. It also highlighted the difference between restaurants categories in terms of sentiment scores using an analysis of variance model. The developed sentiment analysis model for Dubai restaurants can also be extended to extract reviews related to tourism products such as accommodation, attractions and accessibility, with credible efficiency proving to be of greater utility to the tourism sector. The study has some limitations. It has used only limited data and only one sentiment analysis model, which is based on the lexicon approach. Hence it could not compare the accuracy of the proposed model with other models such as machine learning and deep learning models. Future research can focus on building a hybrid model which includes both lexi-

cons and the machine learning based model and so the accuracy of the predicted sentiments can be measured.

References

Aggarwal, C. C., & Zhai, C. (2012). *Mining text data.* Springer Science & Business Media.

Alessia, D., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications, 125*(3), 26–33.

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

Artemenko, O., Pasichnyk, V., Kunanets, N., & Shunevych, K. (2020). Using sentiment text analysis of user reviews in social media for e-tourism mobile recommender systems. In *COLINS 2020: Computational linguistics and intelligent systems* (pp. 259–271). http://ceur-ws.org/Vol-2604/

Balasubramanian, S., Kaitheri, S., Nanath, K., Sreejith, S., & Paris, C. M. (2021). Examining post-COVID-19 Tourist concerns using sentiment analysis and topic modeling. In W. Wörndl, K. Chulmo, & J. S. Stienmetz (Eds.), *Information and Communication Technologies in Tourism 2021: Proceedings of the ENTER 2021 E-Tourism Conference, January 19–22, 2021* (pp. 564–569). Springer.

Bose, R., Dey, R. K., Roy, S., & Sarddar, D. (2020). Sentiment analysis on online product reviews. In C. Marolla, *Information and communication technology for sustainable development* (pp. 559–569). CRC Press.

Breda, Z., Costa, R., Dinis, G., & Martins, A. A. (2020). eWOW of guests regarding their hotel experience: Sentiment analysis of TripAdvisor reviews. In C. M. Q. Ramos, C. R. de Almeida, & P. O. Fernandes (Eds.), *Handbook of research on social media applications for the tourism and hospitality sector* (pp. 295–308). IGI Global.

Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet – The state of eTourism research. *Tourism Management, 29*(4), 609–623.

Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment classification of consumer-generated online reviews using topic modeling. *Journal of Hospitality Marketing & Management, 26*(7), 675–693.

Carson, D. (2006). Developing regional tourism using information communications technology. In S. Marshall, W. Taylor, & X. H. Yu, *Encyclopedia of developing regional communities with information and communication technology* (pp. 176–181). Idea Group Reference.

Chen, M. M., & Scovino, A. I. R. (2020). Which photo themes evoke higher intention to visit Switzerland? In J. Neidhardt & W. Wörndl (Eds.), *Information and communication technologies in tourism 2020* (pp. 53–64). Springer.

Chen, W., Xu, Z., Zheng, X., Yu, Q., & Luo, Y. (2020). Research on sentiment classification of online travel review text. *Applied Sciences, 10*(15), 5275. https://doi.org/10.3390/app10155275

Choi, S., Lehto, X. Y., & Morrison, A. M. (2007). Destination image representation on the web: Content analysis of Macau travel related websites. *Tourism Management, 28*(1), 118–129.

Cohen, S. A., Prayag, G., & Moital, M. (2014). Consumer behaviour in tourism: Concepts, influences and opportunities. *Current Issues in Tourism, 17*(10), 872–909.

Doolin, B., Burgess, L., & Cooper, J. (2002). Evaluating the use of the Web for tourism marketing: A case study from New Zealand. *Tourism management, 23*(5), 557–561.

Duong, H. T., & Nguyen-Thi, T. A. (2021). A review: Preprocessing techniques and data augmentation for sentiment analysis. *Computational Social Networks, 8*(1), 1–16. https://doi.org/10.1186/s40649-020-00080-x

Faheem, A., Awan, T. M., Hussain, S. J., Aisha, K., & Mahwish, P. (2020). Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. *Palgrave Communications, 7*(1), 1–9.

Gadidov, B., & Priestley, J. L. (2018). Does Yelp matter? Analyzing (and guide to using) ratings for a quick serve restaurant chain. In S. Srinivasan (Ed.), *Guide to big data applications* (pp. 503–522). Springer.

Godnov, U., & Redek, T. (2016). Application of text mining in tourism: Case of Croatia. *Annals of Tourism Research, 58*(C), 162–166.

Gräbner, D., Zanker, M., Fliedl, G., & Fuchs, M. (2012). Classification of customer reviews based on sentiment analysis. In M. Fuchs, F. Ricci, & L. Cantoni (Eds.), *Information and communication technologies in tourism 2012: Proceedings of the International Conference in Helsingborg, Sweden, January 25–27, 2012* (pp. 460–470). Springer.

Hao, Q., Cai, R., Wang, C., Xiao, R., Yang, J. M., Pang, Y., & Zhang, L. (2010). Equip tourists with knowledge mined from travelogues. In *Proceedings of the 19th International Conference on World Wide Web* (pp. 401–410). ACM.

Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In *COLINS 1992: The 14th International Conference on Computational Linguistics* (Vol. 2, pp. 539–545). Association for Computational Linguistics.

Hearst, M. (2003). *What is text mining.* https://people.ischool.berkeley.edu/ hearst/text-mining.html

Jiang, Q., Chan, C. S., Eichelberger, S., Ma, H., & Pikkemaat, B. (2021). Sentiment analysis of online destination image of Hong Kong held by mainland Chinese tourists. *Current Issues in Tourism, 24*(6), 1–22.

Kamerer, D. (2014). Understanding the Yelp review filter: An exploratory study. *First Monday, 19*(9). https://doi.org/10.5210/fm.v19i9.5436

Karas, V., & Schuller, B. W. (2021). Deep learning for sentiment analysis: An overview and perspectives. In F. Pinarbasi & M. Nurdan Taskiran, *Natural language processing for global and local business* (pp. 97–132). IGI Global.

Kim, K., Park, O. J., Yun, S., & Yun, H. (2017). What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management. *Technological Forecasting and Social Change, 123*(C), 362–369.

Lau, K. N., Lee, K. H., & Ho, Y. (2005). Text mining for the hotel industry. *Cornell Hotel and Restaurant Administration Quarterly, 46*(3), 344–362.

Law, R., Fong, L. H. N., Koo, C., & Ye, B. H. (2017). Social media in hospitality and tourism. *International Journal of Contemporary Hospitality Management, 29*(2), 646–647.

Li, W., Jin, B., & Quan, Y. (2020). Review of research on text sentiment analysis based on deep learning. *Open Access Library Journal, 7*(3), 1–8.

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management, 29*(3), 458–468.

Loh, S., Lorenzi, F., Saldaña, R., & Licthnow, D. (2003). A tourism recommender system based on collaboration and text analysis. *Information Technology & Tourism, 6*(3), 157–165.

Manning, C. D., Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing.* MIT Press.

Mariani, M., Baggio, R., Buhalis, D., & Longhi, C. (Eds.). (2014). *Tourism management, marketing, and development: The importance of networks and ICTS* (Vol. 1). Springer.

Mohammad, S. M., & Turney, P. D. (2013). *NRC emotion lexicon.* National Research Council Canada.

Nehe, M. P. B., & Nawathe, A. N. (2020). *Aspect based sentiment classification using machine learning for online reviews* (EasyChair Preprint 3051). https://easychair.org/publications/preprint/xnVW

Özen, I. A., & Ilhan, I. (2020). Opinion mining in tourism. In E. Çeltek (Ed.), *Handbook of research on smart technology applications in the tourism industry* (pp. 43–64). IGI Global.

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

Pang, Y., Hao, Q., Yuan, Y., Hu, T., Cai, R., & Zhang, L. (2011). Summarizing tourist destinations by mining user-generated travelogues and photos. *Computer Vision and Image Understanding, 115*(3), 352–363.

Pekar, V., & Ou, S. (2008). Discovery of subjective evaluations of product features in hotel reviews. *Journal of Vacation Marketing, 14*(2), 145–155.

Philander, K., & Zhong, Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management, 55,* 16–24.

Prameswari, P., Surjandari, I., & Laoh, E. (2017). Opinion mining from online reviews in Bali tourist area. In R. Drezewski, G. Chakraborty, S. Nazir, L. Septem Riza, U. Raba'ah Hashim, A. Prasetyo Wibawa, Y. Wihardi, A. Pranolo, E. Junaeti, S.-J. Horng, H. S. Lim, L. Hernandez (Eds.), *3rd International Conference on Science in Information Technology* (pp. 226–230). IEEE.

Qamar, A. M., & Alassaf, M. (2020). Improving sentiment analysis of Arabic tweets by One-Way ANOVA. *Journal of King Saud University – Computer and Information Sciences.* https://doi.org/10.1016/j.jksuci.2020.10.023

Qi, S., & Ning, C. (2017). 'Thank you for your stay,' and then what? Macau hotels' responses to consumer online reviews. In R. Schegg & B. Stangl (Eds.), *Information and communication technologies in tourism 2017* (pp. 559–569). Springer International Publishing.

Ray, A., Bala, P. K., & Jain, R. (2020). Utilizing emotion scores for improving classifier performance for predicting customer's intended ratings from social media posts. *Benchmarking: An International Journal, 28*(2). https://doi.org/10.1108/BIJ-01-2020-0004

Renganathan, V. (2017). Text mining in biomedical domain with emphasis on document clustering. *Healthcare Informatics Research, 23*(3), 141–146.

Saad, S., & Aref, M. (2020). A survey on sentiment analysis in tourism. *International Journal of Intelligent Computing and Information Sciences.* http://doi.org/10.21608/IJICIS.2020.22851.1014

Salton, G., & Buckley, C (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management, 24*(5), 513–523.

Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval.* McGraw-Hill.

Samuel, J., Rahman, M. M., Ali, G. M. N., Samuel, Y., Pelaez, A., Chong, P. H. J., & Yakubov, M. (2020). Feeling positive about reopening? New normal scenarios from COVID-19 US reopen sentiment analytics. *IEEE Access, 8,* 142173–142190.

Schmunk, S., Höpken, W., Fuchs, M., & Lexhagen, M. (2013). Sentiment analysis: Extracting decision-relevant knowledge from UGC. In Z. Xiang & I. Tussyadiah (Eds.), *Information and communication technologies in tourism 2014* (pp. 253–265). Springer.

Standing, C., Tang-Taye, J.-P., & Boyer, M. (2014). The impact of the Internet in travel and tourism: A research review 2001–2010. *Journal of travel & tourism marketing, 31*(1), 82–113.

Sun, S., Law, R., & Zhang, M. (2020). An updated review of tourism-related experimental design articles. *Asia Pacific Journal of Tourism Research, 25*(7), 710–720.

Thomaz, G. M., Biz, A. A., Bettoni, E. M., Mendes-Filho, L., & Buhalis, D. (2016). Content mining framework in social media: A FIFA world cup 2014 case analysis. *Information & Management, 54*(6), 786. https://doi.org/10.1016/j.im.2016.11.005

Valdivia, A., Luzón, M. V., & Herrera, F. (2017). Sentiment analysis in tripadvisor. *IEEE Intelligent Systems, 32*(4), 72–77.

Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining: An overview. *International Journal of Computer Science & Communication Networks, 5*(1), 7–16.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management, 31*(2), 179–188.

Xiang, Z., & Pan, B. (2011). Travel queries on cities in the United States: Implications for search engine marketing for tourist destinations. *Tourism Management, 32*(1), 88–97.

Xiang, Z., Magnini, V. P., & Fesenmaier, D. R. (2015). Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet. *Journal of Retailing and Consumer Services, 22*(C), 244–249.

Xue, P. E. N. G. (2013). *A study on the tourism destination image of Japan in the Chinese market* (Working Paper Series No. 9). The International Centre for the Study of East Asian Development.

Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications, 36*(3), 6527–6535.

Yi, S., & Liu, X. (2020). Machine learning based customer sentiment analysis for recommending shoppers, shops based on customers' review. *Complex & Intelligent Systems, 6*(3), 621–634.

Yu, C., Zhu, X., Feng, B., Cai, L., & An, L. (2019). Sentiment analysis of Japanese tourism online reviews. *Journal of Data and Information Science, 4*(1), 89–113.

Yusof, N. N., Mohamed, A., & Abdul-Rahman, S. (2015). Reviewing classification approaches in sentiment analysis. In M. W. Berry, A. Mohamed, & B. W. Yap (Eds.), *Soft computing in data science: First International Conference* (pp. 43–53). Springer.

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wires: Data Mining and Knowledge Discovery, 8*(4), e1253.