YÖNETİM BİLİŞİM SİSTEMLERİ DERGİSİ
http://dergipark.gov.tr/ybs

Yayın Geliş Tarihi: 02.12.2019 Cilt:6, Sayı:1, Yıl:2020, Sayfa:86-95
Yayına Kabul Tarihi: 31.12.2019 e-ISSN: 2630-550X
Online Yayın Tarihi: 15.05.2020

PREDICTION OF STAR RATINGS OF HOTEL CUSTOMERS USING SENTIMENT ANALYSIS[1]

Dr. Öğr. Üyesi Selçuk KIRAN[1], Prof. Dr. Meltem ÖZTURAN[2]

[1] Business Informatics, Faculty of Business Administration, Marmara University, Istanbul Türkiye

[2] Management Information Systems, School of Applied Disciplines, Bogazici University, Istanbul Türkiye

ABSTRACTS

The aim of this study is three-fold; access some exploratory findings about hotels and their ratings, predict star ratings of customers by sentiment analysis using their eWOM comments, and compare predicted eWOM star ratings of customers with the corresponding online star ratings by statistical analyses. Data of hotels from four cities with different income levels are retrieved using a script written with PHP. The data are cleaned considering various rules to be ready for analyses. Academic demo version of Lexalytics tool is used for sentiment analysis and RStudio for statistical analyses. Results show that, average number of rooms per hotel and their daily average rates increase as the income level of cities increase whereas star ratings of hotels increase as the income level of cities decrease. Analyses show that that for all cities, there is a significant difference between eWOM and online star ratings and also a significant moderate positive relationship between these two star ratings for all cities.

Keywords: sentiment analysis, eWOM, star rating, hotel, income level

DUYGU ANALİZİ İLE OTEL MÜŞTERİLERİNİN YILDIZ DEĞERLENDİRMELERİNİN TAHMİNİ[2]

ÖZET

Bu çalışmanın üç farklı amacı vardır; oteller ve derecelendirmeleri hakkında bazı bilgileri keşfetmek, müşterilerin eWOM yorumları için duygu analizi ile yıldız derecelendirmelerini tahmin etmek ve bu eWOM yıldız derecelendirmelerini istatistiksel analizlerle söz konusu yorumlara karşılık gelen çevrimiçi yıldız derecelendirmeleriyle karşılaştırmaktır. Farklı gelir seviyesine sahip dört ildeki otel verileri PHP ile yazılmış bir program kullanılarak toplanmıştır. Veriler çeşitli kurallar göz önünde bulundurularak temizlenip, analizlere hazır hale getirilmiştir. Lexalytics aracının akademik demo sürümü, duyarlılık analizi, RStudio ise istatistik analizler için kullanılmıştır. Sonuçlar, otel başına ortalama oda sayısının ve günlük ortalama otel ücretlerinin illerin gelir düzeyi arttıkça arttığını, otellerin yıldız puanlarının ise illerin gelir düzeyi azaldıkça arttığını göstermektedir. Analizler tüm iller için eWOM ve çevrimiçi yıldız derecelendirmeleri arasında anlamlı bir fark olduğunu ve ayrıca tüm iller için bu iki yıldız derecelendirme arasında anlamlı ve orta düzeyde bir pozitif ilişki olduğunu göstermektedir.

Anahtar Kelimeler: duygu analizi, eWOM, yıldız derecelendirmesi, otel, gelir düzeyi

---

[1]This study was presented at the 6th International Management Information Systems Conference and the abstract was published in abstract booklet.

[2] Bu çalışma 6. Uluslararası Yönetim Bilişim Sistemleri Konferans'ında bildiri olarak sunulmuş ve çalışmanın özeti özet kitapçığında yayımlanmıştır.

INTRODUCTION

Online star ratings of customers, ranging from 1 to 5, not only helps where the hotel stands out amongst its competitors, but also helps in the decision-making process of consumers for their travel arrangements. The more stars the hotel has, the more sales it will likely get. Besides star ratings, electronic word of mouth (eWOM), in other words online reviews and comments about the hotel, is likewise very important for consumers. The key issue here is to be able to read all of these reviews and comments before making a decision. On the other hand, it should be considered that the most important issue to managing a hotel's online reputation is not only being rated on its own web site but also being available online on different review sites like TripAdvisor. In this study, sentiment analysis approach is used to quantify English eWOM data for hotels similar to star rating. Hence, it aims to give consumers the chance to focus only on the hotels with the eWOM star rating they prefer for detailed customer opinions, rather than going over the whole reviews and comments. An academic demo version tool is used for sentiment analysis approach at this stage. Moreover, eWOM star ratings of customers are compared with their online star ratings to find out how similar the ratings are. Also, some exploratory findings about hotels and their ratings are presented.

The paper is structured as follows; the rest of the introduction section provides a background and a literature survey on sentiment analysis and its use in tourism industry; next two sections describe the methodology and findings of the study; the last section consists of discussion of the findings, limitations of the study and directions for future research.

Sentiment analysis is defined as a computational study to detect and extract subjective information, such as opinion and emotions, expressed in text using a series of methods, tools and techniques (Liu, 2009). Usually, sentiment analysis is about opinion polarity, where there is positive, neutral, or negative sentiment towards something (Dave et al., 2003). There are two major techniques for sentiment analysis; machine learning based techniques and lexicon based techniques. In a machine learning based technique, two sets of text data are needed: training and a test set. A training set is used by an automatic classifier to learn the differentiating characteristics of the text data, and a test set is used to check how well the classifier performs (Vohra, & Teraiya, 2013). The lexicon based technique involves calculating orientation for a text data from the semantic orientation of words in the text whose sentiment values are determined prior to their use (Turney 2002). Each of two techniques has advantages and disadvantages that have been discussed in various studies (Neviarouskaya et al., 2015, Taboada et al., 2011, and Hailong et al., 2014). Therefore, many researchers have used hybrid technique which uses the combination of both machine learning technique and lexicon based technique (Prabowo, & Thelwall, 2009, Malandrakis et al., 2013 and Sommar, & Wielondek, 2015). The basic goal of this combination is to yield the best and optimum results using the effective feature set of both machine learning and lexicon based techniques, and to overcome the deficiencies and limitations of both techniques (Ahmad et al, 2017). Almost in all production and service sectors, sentiment analysis is being used to present exciting opportunities for marketers to generate market intelligence on consumer attitudes and brand opinions (Rambocas, & Pacheco, 2018). Tourism is one of these sectors and some of the related studies can be seen in Table 1.

Table 1. Use of Sentiment Analysis Techniques in Tourism

REFERENCE	CONTENT OF THE STUDY
Alaei et al., 2019	Different sentiment analysis approaches applied in tourism are reviewed and assessed in terms of the datasets used and performances on key evaluation metrics.
Thelwall, 2019	The main sentiment analysis approaches with a focus on practical descriptions of how the methods work and how they can be applied in tourism sector are reviewed.
Fu et al., 2019	The design effects on predictive accuracy using a sentiment analysis experiment for Chinese travel news are examined.
Kirilenko et al., 2018	The suitability of different types of automated sentiment analysis for applications typical in tourism, hospitality, and marketing studies by comparing their performance to that of human raters is evaluated.
Roy et al., 2018	Sentiment analysis is done in a summarized opinion so that interested tourists do not go through all the reviews, rather they go through summarized documents with the overall sentiment about target place.
Gitto, & Mancuso, 2017	Sentiment analysis is used for measuring the level of customer satisfaction of airport passengers to provide a feedback to airport managers.
Lak, & Turetken, 2014	Sentiment analysis results with star ratings in three different domains (technology, tourism, health) are compared to explore the promise of this analysis.
Garcia et al., 2012	Sentiment analysis of user reviews in Spanish for the accommodation, and food and beverage sectors are done by using lexical databases.

Many tools have been developed and used in recent years for analyzing sentiments of online texts. HubSpot's ServiceHub, Quick Search, Repustate, Lexalytics, Critical Mention, Brandwatch, Social Mention and Sentiment Analyzer are well known tools. Among them HubSpot's ServiceHub, Lexalytics (academic demo version), Social Mention and Sentiment Analyzer are free where the others are commercial. For this study, Lexalytics is chosen as the tool since its academic demo version is free, it has a very large dictionary of words together with their qualified sentiment scores and it uses hybrid sentiment analysis technique.

METHOD

This study has three main goals; accessing some exploratory findings about hotels and their ratings by descriptive statistics, predicting star ratings of customers by sentiment analysis using their English eWOM comments and comparing predicted eWOM star ratings of customers with the corresponding online star ratings.

Data Collection and Preparation

Source of data is chosen to be the tourism review web site, TripAdvisor. The other related sites are found unsuitable since there are some drawbacks for them like difficulty in fetching hotel data, non-compact hotel data, prohibited in some countries, no support for JSON that facilitates data collection.

For determining the appropriate hotels to retrieve data from TripAdvisor, purposive sampling is considered where firstly the world regions are categorized according to their income economies namely; low, lower-middle, upper-middle and high (WB, 2019). Afterwards, one city is chosen from each region that its country belongs to, considering the city’s tourism attraction and the number of hotels of the city existing in TripAdvisor as; NewYork City from high, İstanbul from upper-middle, Hanoi from lower-middle and Kathmandu from low. General information about these cities are given in Findings section.

In order to scrape the necessary data that include hotel's basic characteristics, its eWOM comments and the corresponding star ratings, and to store them in MySQL database, a PHP script that resolves HTML and JSON codes of TripAdvisor is written. The data collection process has been realized between July 18, 2019 and Jul 25, 2019 for the predetermined four cities. During this time period, 7,243 hotels' basic characteristics and their accumulated customer online star ratings are retrieved. Also, to apply sentiment and statistical analyses, customers' individual eWOM comments and the corresponding online star ratings are collected. During this collection, due to the limitations of the academic demo version of the tool Lexalytics, for the hotels that have more than 20 eWOM comments and corresponding online star ratings, only the first 20 recent data for each of those hotels are gathered which add up to 37,767 eWOM comments and 37,767 corresponding online star ratings. Once the data are stored in the database, a data cleaning process is applied following the below rules sequentially;

- All the data that are retrieved problematically due to network errors are removed from the database together with their all related data

- All the hotels which stay as link in TripAdvisor but don't have any information are removed from the database together with their related eWOM comments and corresponding online star ratings

- eWOM data written in a language different than English are removed from the database

- Hotels that remain with less than 5 English reviews are removed from the database together with their related eWOM comments and corresponding online star ratings.

- In order to analyze close number of reviews for each city, number of hotels is reduced, by taking only the first 300 hotels data from each city

At the end of the data cleaning process, 1,200 hotels with 19,327 eWOM comments and corresponding star ratings are made ready for sentiment and statistical analyses.

Analyses

Sentiment analysis

In order to feed the data to the tool Lexalytics for sentiment analyses, the cleaned data in MySQL database are exported to a csv file since the demo version of the tool only permits its Excel API to calculate sentiment scores. Lexalytics assigns a sentiment score to each phrase on a scale of -1 (very negative) to +1 (very positive). But as explained by Lexalytics (2019), there is no set scale for sentiment scores as they are based on the sentiment-bearing phrases contained within an individual document. Since the aim of this study is to compare analyzed eWOM star ratings with the corresponding 1-5 online star ratings, the sentiment scores are recoded to 1-5 after making normalization based on the minimum and maximum sentiment score values scale as given in Table 2.

Table 2. Conversion of Sentiment Score to 1-5 Scale

SENTIMENT SCORE	1-5 SCALE
$\geq$ - 1.672 and < - 0.956	1
$\geq$ - 0.956 and $\leq$ - 0.240	2
$\geq$ - 0.240 and < 0.476	3
$\geq$ 0.476 and < 1.192	4

≥ 1.192 and ≤ 1.908	5

Statistical analyses

Using RStudio statistical tool, through descriptive statistics, some characteristics of hotels, their analyzed eWOM and online star ratings are explored and tabulated in Findings section. Lastly, using the same tool, the analyzed eWOM and corresponding online star ratings are compared using paired sample t-test, which is a statistical procedure used to test if the means of two metric variables are different in the same sample within a significance level and also using Pearson correlation test that indicates the extent to which two variables are linearly related. The results of these analyses are also given in Findings section.

FINDINGS

General information about the cities considered in this study are given in Table 3. New York is seen as the city that has the highest tourist attraction where Kathmandu comes out to be lowest. These numbers show parallelism with income levels of the cities.

Table 3. General Information about Cities

	NEW YORK CITY (USA)	İSTANBUL (TURKEY)	HANOI (VIETNAM)	KATHMANDU (NEPAL)
Income Level (WB, 2019)	High	Upper-Middle	Lower-Middle	Low
Population in 2018	8,601,186	14,750,771	7,781,631	1,329,732
Area (km^2)	784	5,461	3,329	51
Number of Hotels Hosted in TripAdvisor in 2019	688	3,790	1,923	842
Number of Tourist Arrivals in 2018 [1,000]	13,500	12,120	5,740	969

Descriptive statistics for the characteristics of hotels in four cities are given in Table 4. The results show that;

- Only New York City (high income level) has hotels with more than 600 rooms (7.0%) whereas Kathmandu (low income level) has the highest percent of hotels with room numbers less than 50 (83.0%).

- Average number of rooms per hotel in cities are seen to be positively related to the income levels of the cities; higher the income level of a city, higher the average number of rooms per hotel in that city.

- Hotels in New York City (high income level) have the highest average daily rate per hotel ($2,108) whereas hotels in Kathmandu (low income level) have the lowest average daily rate per hotel ($247).

- Average daily rates of hotels are seen to be positively related to the income levels of the cities; higher the income level of a city, higher the average daily rate of hotels in that city.

- Average accumulated online star ratings is smallest (4.04) in New York City (high income level) whereas it is highest (4.60) in Hanoi (lower-middle income level).

- Average accumulated online star rating values of hotels in cities are mostly negatively related to the income levels of cities; higher the income level of a city, lower the average online star rating of that city.

- Average analyzed online star ratings is smallest (3.94) in New York City (high income level) whereas it is highest (4.42) in Kathmandu (low income level).

- Average analyzed online star rating values of hotels in cities are negatively related to the income levels of cities; higher the income level of a city, lower the average online star rating of that city.

- In overall, average of analyzed online star ratings for all cities (4.19) is very close to accumulated online star ratings for all cities (4.14).

- Average number of eWOM comments and online star ratings of hotels is seen to be highest in New York City (1,942); this may be due to the consideration of only English eWOM comments in this study where New York City is the only city among the others that has English as its native language.

- Although it is aimed to analyze a total of 6,000 eWOM comments from each city (300 hotels x 20 eWOM comments per hotel), the aimed number is not reached since the average eWOM comments are less than 20 especially for low and lower-middle income level cities.

- Average eWOM star ratings is smallest (3.43) in New York City (high income level) whereas it is highest (3.58) in Kathmandu (low income level).

- Average eWOM star rating values of hotels in cities are negatively related to the income levels of cities; higher the income level of a city, lower the average eWOM star rating of that city.

- The percentage of the eWOM star rating 1 is very low in total (0.1%) whereas for online star ratings it is higher (4.0%); this is not consistent for the ratings from 2 to 4 where percentages of eWOM star ratings are higher than the ones for online star ratings.

- In total, almost half of the online star ratings is 5 (51.5%), whereas almost half of the eWOM star ratings per city is 4 (52.8%); this may be due to various reasons which are discussed in detail for the results of paired sample t-test given in Table 5.

Table 4. Descriptive Statistic for Hotels (300 Hotels from Each City)

	NEW YORK CITY (USA)	İSTANBUL (TURKEY)	HANOI (VIETNAM)	KATHMANDU (NEPAL)	ALL FOUR CITIES
Size Classification [Number of Rooms / Hotel]					
< 50	64 - 21.3%	192 - 64.0%	219 - 73.0%	249 - 83.0 %	724 - 60.3 %
50 - 150	94 - 31.3%	65 - 21.7%	51 - 17.0%	30 - 10.0 %	240 - 20.0 %
151 - 299	79 - 26.3%	27 - 9.0%	9 - 3.0%	7 - 2.3 %	122 - 10.2 %
300 - 600	34 - 11.3%	12 - 4.0%	6 - 2.0%	1 - 0.3 %	53 - 4.4 %
> 600	21 - 7.0%	0 - 0.0%	0 - 0.0%	0 - 0.0 %	21 - 1.8 %
Missing	8 - 2.8%	4 – 1.0%	15 - 5.0%	13 - 4.3 %	40 - 3.3 %
Average Number of Rooms / Hotel	207	67	43	32	87
Price Classification [Average Daily Room Rate ($)]					
< 125	0 - 0.0%	0 - 0.0%	39 - 13.0%	78 - 26.0%	117 - 9.8%
>= 125 and < 250	0 - 0.0%	14 - 4.7%	109 - 36.3%	87 - 29.0%	210 - 17.5%

>= 250 and < 375	1 - 0.3%	45 - 15.0 %	77 - 25.7%	37 - 12.3%	160 - 13.3%
>= 375	263 - 87.7%	241 - 80.3%	57 - 19.0%	37 - 12.3%	598 - 49.8%
Missing	36 - 12.0%	0 - 0.0%	18 - 6.0%	61 - 20.3%	115 - 9.6%
Average Daily Room Rate / Hotel	2,108	665	307	247	831
Accumulated Online Star Ratings Distribution					
1	29,173 - 5.0%	3,106 - 2.8%	1,925 - 1.6%	1,227 - 2.0 %	35,431 - 4.0%
2	32,068 - 5.5%	2,926 - 2.7%	2,129 - 1.7%	1,660 - 2.7 %	38,783 - 4.4%
3	78,801 - 13.5%	8,414 - 7.6%	6,266 - 5.1%	4,700 - 7.8 %	98,181 - 11.2%
4	187,985 - 32.3%	27,752 - 25.1%	22,642 - 18.5%	14,254 - 23.6%	252,633 - 28.8%
5	254,505 - 43.7%	68,171 - 61.8%	89,427 - 73.1%	38,664 - 63.9%	450,767 - 51.5%
Average	4.04	4.40	4.60	4.45	4.19
Number of eWOM Comments and Online Star Ratings					
Total	582,532	110,369	122,389	60,505	875,795
Average / Hotel	1,942	368	408	202	730
Maximum / Hotel	24,331	3,915	4,207	2,074	24,331
Minimum / Hotel	19	21	12	5	5
Analyzed Online Star Ratings Distribution					
1	537 - 9.3%	481 - 8.2%	273 - 6.8%	136 - 3.7 %	1427 - 7.4%
2	379 - 6.6%	277 - 4.7%	179 - 4.5%	103 - 2.8 %	938 - 4.9%
3	722 - 12.6%	556 - 9.4%	309 - 7.7%	278 - 7.6 %	1865 – 9.6%
4	1,370 - 23.8%	1,494 - 25.3%	782 - 19.5%	704 – 19.2%	4,350 - 22.5%
5	2,747 - 47.7%	3,091 – 52.4%	2,465 - 61.5%	2,444 - 66.7%	10,747 - 55.6%
Average	3.94	4.09	4.24	4.42	4.14
Number of Analyzed eWOM Comments					
Total	5,755	5,899	4,008	3,665	19,327
Average / Hotel	19.2	19.7	13.4	12.2	16.1
Maximum / Hotel	20	20	20	20	20
Minimum / Hotel	16	17	8	5	5
Analyzed eWOM Star Ratings Distribution					
1	12 - 0.2%	11 - 0.2%	4 - 0.1%	1 - 0.0%	28 - 0.1%
2	349 - 6.1%	363 - 6.2%	159 - 4.0%	93 - 2.5%	964 - 5.0%
3	2,620 - 45.5%	2,473 - 41.9%	1,516 - 37.8%	1,380 - 37.7%	7,989 - 41.3%
4	2,730 - 47.4%	3,008 - 51.0%	2,294 - 57.2%	2,168 - 59.2%	10,200 - 52.8%
5	44 - 0.8%	44 - 0.7%	35 - 0.9 %	23 - 0.6%	146 - 0.8%
Average	3.43	3.46	3.55	3.58	3.49

Results of paired sample t-test for testing if the means of eWOM star ratings and of online star ratings are different are given in Table 5. The results show that for all cities, there is a significant difference in the eWOM (M = 3.49, SD = 0.61) and online (M = 4.14, SD = 1.22) star ratings; t(9,326) = - 91.48, p < 2.2e-16 where the mean of differences is - 0.65. As it can be seen from Table 5, this significance difference is valid for each city also and eWOM star ratings are lower than online star ratings for all cities. This difference may be due to;

- the attitudes of tourists as giving higher star ratings but at the same time criticizing the points that they are not happy with,

- some drawbacks of the algorithm of Lexalytics for sentiment analysis,

Table 5. Results of Paired Sample t-Test for Comparison of eWOM & Online Star Ratings

	eWOM STAR RATINGS AVERAGE	**ONLINE STAR RATINGS AVERAGE**	**t**	**df**	**p-VALUE**	**MEAN DIFFERENCES**
ALL	3.49	4.14	-91.48	19,326	<2.2e-16	-0.65
NEW YORK	3.42	3.94	-37.13	5,754	<2.2e-16	-0.52

ISTANBUL	3.46	4.09	-49.32	5,898	<2.2e-16	-0.63
HANOI	3.55	4.24	-45.24	4,007	<2.2e-16	-0.70
KATHMANDU	3.58	4.42	-58.88	3,664	<2.2e-16	-0.85

Results of Pearson correlation test analyzing if there is a relationship between the eWOM and online star ratings are given in Table 6. Results show that there is a significant (p < 2.2e-16) significant moderate positive relationship between these two star ratings for each city hence for all cities together.

Table 6. Results of Pearson Correlation for Comparison of eWOM & Online Star Ratings

	t	**df**	**p-VALUE**	**PEARSON CORRELATION COEFFICIENT (r) BETWEEN eWOM and ONLINE STAR RATINGS**
ALL	102.590	19,325	<2.2e-16	0.5937861
NEW YORK	57.334	5,753	<2.2e-16	0.6030076
ISTANBUL	60.093	5,897	<2.2e-16	0.6162792
HANOI	45.606	4,006	<2.2e-16	0.5846044
KATHMANDU	35.371	3,663	<2.2e-16	0.5045715

DISCUSSION AND CONCLUSIONS

In this study, based on the data of hotels and their eWOM and online ratings collected from four cities from different income levels are analyzed using sentiment and statistical analyses.

Descriptive results show that average number of rooms per hotel and daily average rates of hotels increase as the income level of cities increase whereas star ratings of hotels increase as the income level of cities decrease. Statistical analyses show that that for all cities, there is a significant difference between eWOM and online star ratings where eWOM star ratings are lower than online star ratings. This difference may be due to the attitudes of tourists as giving higher star ratings but at the same time criticizing the points that they are not happy with or due to some drawbacks of the algorithm of Lexalytics for sentiment analysis. In addition, results show that there is a significant moderate positive relationship between these two star ratings for each city hence for all cities together.

As a managerial implication, based on these results, it can be recommended to hotel managers to pay more attention to eWOM comments since though customers give high star ratings, they actually have some complaints about the hotel.

As it is true for all studies, this study also has some limitations such as: those coming from the sentiment analysis tool, Lexalytics; analyzing only English eWOM comments; retrieving data only from TripAdvisor; using inefficient network environment during data collection.

Associated with the findings of the study, future studies can be recommended as: using different tools for sentiment analysis to test the accuracy of prediction of eWOM star ratings; repeating this study with eWOM comments of various languages; retrieving data from various review sites.

REFERENCES

Ahmad, M., Aftab, S., Ali, I., & Hameed, N. (2017). Hybrid tools and techniques for sentiment analysis: a review. International Journal of Multidisciplinary Sciences and Engineering, 8(4):28-33.

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

Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. Proceedings of the 12th International Conference on World Wide Web, 519-528.

Fu, Y., Hao, J. X., Li, X., & Hsu, C. H. (2019). Predictive accuracy of sentiment analytics for tourism: A metalearning perspective on chinese travel news. Journal of Travel Research, 58(4): 666-679.

García, A., Gaines, S., & Linaza, M. T. (2012). A lexicon based sentiment analysis retrieval system for tourism domain. e-Review of Tourism Research (eRTR), 10(22), 35-38.

Gitto, S., & Mancuso, P. (2017). Improving airport services using sentiment analysis of the websites. Tourism Management Perspectives, 22, 132-136.

Hailong, Z., Wenyan, G., & Bo, J. (2014). Machine learning and lexicon based methods for sentiment classification: A survey. 2014 11th Web Information System and Application Conference: 262-265.

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.

Lak, P., & Turetken, O. (2014). Star ratings versus sentiment analysis--a comparison of explicit and implicit measures of opinions. 2014 47th Hawaii International Conference on System Sciences, 796-805

Lak, P., & Turetken, O. (2014). Star ratings versus sentiment analysis--a comparison of explicit and implicit measures of opinions. In 2014 47th Hawaii International Conference on System Sciences, 796-805

Lexalytics (2019). Freqeuntly asked questions. http://dev.lexalytics.com/wiki/pmwiki.php?n=Main.FAQ, (29 July 2019)

Liu, B. (2009). Handbook chapter: Sentiment analysis and subjectivity. Handbook of Natural Language Processing. Marcel Dekker, Inc. New York, NY, USA.

Malandrakis, N., Kazemzadeh, A., Potamianos, A., & Narayanan, S. (2013, June). SAIL: A hybrid approach to sentiment analysis. Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), 438-442.

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2015). Attitude sensing in text based on a compositional linguistic approach. Computational Intelligence, 31(2): 256-300.

Park, E., Kang, J., Choi, D., & Han, J. (2018). Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews. Current Issues in Tourism, 1-7.

Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. Journal of Informetrics, 3(2): 143-157.

Rambocas, M., & Pacheco, B. G. (2018). Online sentiment analysis in marketing research: a review. Journal of Research in Interactive Marketing, 12(2): 146-163.

Roy, A., Guria, S., Halder, S., Banerjee, S., & Mandal, S. (2018). Summarizing Opinions with Sentiment Analysis from Multiple Reviews on Travel Destinations. International Journal of Synthetic Emotions (IJSE), 9(2): 111-120.

Sommar, F., & Wielondek, M. (2015). Combining Lexicon-and Learning-based Approaches for Improved Performance and Convenience in Sentiment Classification.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational Linguistics, 37(2): 267-307.

Thelwall, M. (2019). Sentiment Analysis for Tourism. Marianna Sigala, Roya Rahimi, Mike Thelwall (Der.) Big Data and Innovation in Tourism, Travel, and Hospitality içinde 87-104, Singapore: Springer.

Vohra, S. M., & Teraiya, J. B. (2013). A comparative study of sentiment analysis techniques. Journal of Information, Knowledge and Research in Computer Engineering (JIKRCE), 2(2): 313-317.

WB (2019). Country Classification: World Bank Country and Lending Groups. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups, (24 July 2019)