

Evaluation of tourist reviews on TripAdvisor for the protection of the world heritage sites: Text mining approach

Ibrahim Akın Özen

ABSTRACT

Keywords:
World Heritage,
Text mining,
Dictionary-based Sentiment
Analysis,
Cappadocia
Göreme,

Collecting and analyzing online tourist reviews on destinations is important for sustainable tourism. These analyses can give insight into the extent to which natural and cultural assets in the destination are protected. These evaluations should be considered by the authorities as objective and realistic assessments. In this study, 4183 TripAdvisor reviews of foreign tourists visiting "Göreme National Park and Cappadocia Rocky Area", which is listed in the World Heritage Site, were evaluated. The data set consisted of English reviews of foreign tourists visiting the region between the years of 2018 - 2020. Dictionary-based sentiment analysis, one of the text mining methods, was used in the study. According to the analysis results, the positive perceptions of the tourists about the churches, fairy chimneys, valleys and underground cities in the World Heritage Site were found to be significantly high (75%). Negative evaluations were found to be low (33%). In tourist reviews, 63% positive and 10.49% negative evaluations were made about the protection of the region. In addition, the awareness of the tourists about whether the locations they visit are World Heritage Sites was 30.6%. Tourists explained their negative opinions about the locations they visited with the words "extra_payment", "crowded", and "steep". Another finding of the study was that the area is adequately protected. There were, however, some concerns related to protection. The most striking aspects of protection concerns were those in which the words "painted" and "drawn" were used, an important indicator of visitor sensitivity to the protection of frescoes in rock churches. Since the destruction of natural and cultural assets in the region is unacceptable, recommendations are made to take the necessary measures to prevent such damage.

Article History:
Submitted: 14.08.2020
Accepted: 17.02.2021

Doi: <https://doi.org/10.31822/jomat.876175>

1. Introduction

The United Nations Educational, Scientific and Cultural Organization (UNESCO) accepted the Convention on the Protection of the World Cultural and Natural Heritage at the conference held by the member countries on November 16, 1972. According to the convention, it is emphasized that the world's natural and cultural heritage is a source of life and inspiration, not for a single nation but for all humanity, and that it is invaluable and irreplaceable. From this point of view, all countries should protect natural and human-made assets, help them survive and start taking a sustainable perspective. However, due to the increasing threat to natural and cultural heritage over time, the necessity of a continuous legal, administrative and financial functioning has surfaced for the protection of these areas. For this reason, an "Intergovernmental Committee" was established by the contracting countries. The Committee has established the "World Heritage List in Danger"

and the "World Cultural and Natural Heritage Conservation Fund" in order to identify and preserve national heritage and to create an international cooperation and assistance system (Akipek, 2001). Turkey has been a party to the UNESCO World Heritage Convention on March 16, 1983. After that date, first in 1985, Göreme National Park and Cappadocia (Rocky Region), from Turkey, the other two assets (Istanbul Historic Areas and the Great Mosque and Hospital in Sivas) entered the World Heritage List.

According to the World Heritage Convention, if national assets have all or some of the characteristics of cultural and natural heritage, those assets shall be considered as mixed (natural and cultural) heritage assets. Natural and cultural assets are defined as "superior universal value", "cultural and natural value that is extraordinary to the extent that it transcends national borders and has common importance for the present and future generations of all humanity" (Demirçivi, 2017).

Research paper

Ibrahim Akın Özen:

Assist. Prof. Dr., Nevşehir HBV University Department of Tourism Guidance, Turkey, Email: akinozen@gmail.com,
Orcid Id: 0000-0003-1172-5448



This work is licensed under a Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0).

As of 2020, there are 869 cultural, 213 natural and 39 both cultural and natural assets worldwide (UNESCO, 2020). Göreme National Park and Cappadocia (Rocky Regions) Region, also the subject of the study, is an important region that is included in the list of only 39 natural and cultural assets worldwide.

Göreme National Park and Cappadocia Region encompasses in Nevşehir Province, Ürgüp, Avanos, Kaymaklı and Derinkuyu, Karain, Karlık, Yeşilöz and Soğanlı. In these areas, there are Göreme Open Air Museum, Kaymaklı Underground City, Derinkuyu Underground City and Zelve Open Air Museum. In this region there are outstanding rock formations and valleys called fairy chimneys formed by geological erosions.

In Göreme Open Air Museum, there are Girls and Boys Monastery, Tokalı Church, Saint Basil Church, Elmalı Church, Saint Barbara Church, Snake Church, Kitchen-Cellar, Refectory, Dark Church, Saint Catherine Chapel and Çarıklı Church. These settlements belong to the early periods of Christianity (Yılmaz, 2011). The scenes from the Bible and the life of Jesus were depicted on the frescoes in these churches. Due to these characteristics, the Göreme National Park and the Cappadocia region as a whole is considered to be one of the most important tourism centers in our country as a World Heritage Site (WHS) (Tosun, et al., 2020).

The fact that the region is in the WHS has a positive impact on awareness, making it a center of attraction. In this way, the region attracts a large number of visitors to both domestic and foreign tourism.

The tourist density of the region can be threatening natural and cultural assets. Tourism activities in the region should be evaluated and carried out in the context of sustainable tourism principles. For sustainable destination management, on-line communities can be used as a means of protection, evaluation and management of the destination resources. Accordingly, the reviews they share in electronic media on their visiting experiences should be taken into consideration as objective and realistic assessments for evaluating the destination (Aydın, 2016).

In the first part of this study, text mining and its methods were mentioned. In the second part, literature review about text mining in destination management was given. In the third part, the research method was explained. In the findings

and discussion section, the analysis results were evaluated and suggestions were made.

In this study, the reviews of foreign tourists visiting Göreme National Park and Cappadocia Rocky Sites on TripAdvisor were evaluated using text mining methods. The aim of the study is to find answers to following questions in terms of destination management.

1. What are the positive and negative feelings of foreign tourists about natural and cultural assets in the areas they visit?
2. What is the level of awareness of foreign tourists about the fact that the region is in the World Heritage Site?
3. What do foreign tourists think about the protection of natural and cultural assets in the region? Do they have any concerns about this?

2. Literature Review

In the tourism sector, textual content generated by tourists on social media and online platforms is of strategic importance in terms of potential tourists, tourism businesses and destination management (Aydın, 2016; Özen & İlhan, 2020). In the literature, posts with textual content are referred as user-generated content (UGC) or electronic word of mouth (eWOM) (Buzova et al., 2019; Schmunk et al., 2013).

Positive or negative reviews of tourists related to their destination experience are mentioned as unstructured data in text mining. These unstructured data are stored in large data sets on social media and platforms. The extraction of the required information from large unstructured data sets is a very difficult process to do with traditional information systems. Text mining, regarded as the sub-discipline of Data Mining, makes an important contribution to the efforts to extract understandable and interpretable information from unstructured digital texts.

Text mining is the process of transforming unstructured data into numerical values and extracting qualified information. Text mining works with a wide range of disciplines, such as statistics, artificial intelligence and natural language processing, to analyze data from natural language texts (Miner et al., 2012). Text mining uses the methods of text classification, text clustering, topic extraction, sentiment analysis or opinion mining by using the disciplines mentioned above (Li et al., 2019).

Sentiment analysis is a method of text mining that has a vast field of study in a close relationship with the discipline of natural language processing. It allows the sentimental expressions of a text to be extracted (Oğuzlar & Kızılkaya, 2019). Sentiment analysis is categorized as document level, sentence level, and aspect/feature-level (Jo & Oh, 2011).

The objective of the document level sentiment analysis is to identify the positive or negative polarity of the document analyzed. Sentence level sentiment analysis is intended to determine the positive or negative polarity of a sentence. Positive or negative polarity can be defined as a reflection of the positive or negative opinions of the author of a document or sentence.

Two different approaches are used in sentiment analysis research as Machine Learning and Dictionary-Based Sentiment Analysis. Sentiment classification is the focus of the Machine Learning approach using machine learning algorithms (Medhat et al., 2014). In Dictionary-Based Sentiment Analysis, sentiment dictionaries are used. Sentiment Dictionary is a database that keeps the sentiment polarity of the terms (negative / neutral / positive) and sentiment scores numerically. Each term in the content is looked up in the sentiment dictionary while carrying out a sentiment analysis. Polarity and score are added to the total score if the term is available in the dictionary. If the total score is less than zero, it is classified as negative; if it is higher than zero, it is classified as positive; and if it is equal to zero, then the text sentiment is classified as neutral (Hu & Liu, 2004).

Sentiment Analysis is considered as an important opportunity for tourists and suppliers to get to know and understand each other. These opportunities should be evaluated from the perspective of tourist products suppliers (tourism businesses, destination management) and consumers, i.e. tourists.

In the tourism literature, the majority of studies about tourist reviews on electronic platforms are consumer-centered. They are focused mostly on the use of social media by tourists or its impact on tourist behavior. Few studies have examined tourist comments from the viewpoint of tourism providers (Denizci Guillet, Kucukusta, & Liu, 2016). Knowledge on how tourism providers or organizations use social media is limited (Munar, 2012; Uşaklı et al., 2019). It is seen that text mining studies on destination management are genuinely for the whole destination or some

tourism businesses in the destination. As an example; Irawan, Akmalia, & Masrury (2019) analysed TripAdvisor and social media reviews of tourists visiting Indonesia in their study utilizing sentiment analysis and topic extraction techniques. According to the results of the study, it was understood that the "entertainment" aspect was the most important factor in visitor experiences.

In another study, it was inferred from 19.835 online tourists reviews on the website virtualtourist.com utilizing sentiment analysis techniques that the most important negative assessments in Paris were about transportation (Kim et al., 2017).

Some studies using text mining techniques are trend analyzes that have been conducted in terms of destination marketing. In trend analysis; answers are sought to the questions such as what the market thinks about us, how they perceive us, what they value the most and the least, and how they compare us to our competitors. It can be beneficial for destinations to take into account the answers to these questions in order to assess their place in the market. The words used by potential tourists in search engines for a destination are seen as an important data point in trend analysis. In services such as Google Trend and Baidu, vast amounts of search data, i.e. keywords, are organized and stored. Studies use the data sets collected from these services to assess the location and trends of destinations in the tourism market. Web search data are also used to estimate the demand for the tourism market (Li et al., 2018), the number of visitors (Gunter & Önder, 2016), tourist activity (Artola et al., 2015) and demand for hotels (Pan et al., 2012).

The research conducted by Dinis et al. (2019) can be given as an example in determining the trends towards the destinations. In the research, the interests of Spanish, British and German tourists towards Portuguese touristic products were analyzed. The analysis was carried out through categorizing by geographical locations based on the keywords used in the Google search engine. From this point of view, marketing strategies can be developed by segmenting the market. Determining the interests of potential tourists for a destination can provide a strategic competitive advantage.

In tourism literature, Text mining studies on WHS are limited. Antonio Correia, & Ribeiro (2020) analyzed and compared 8.638 online travel reviews from TripAdvisor for Salamanca (Spain) and

Coimbra (Portugal), listed on the WHS, in their research. They used text mining, sentiment analysis and market basket models. In the research, keywords were defined in order to be used by the destination management and organizations for promoting these cities. In the promotion of these regions, it was emphasized that the keywords for history and world heritage such as "architectural style", "names of kings" and "place names" should be brought to the fore. The endeavors of WHS to reposition them in tourism market using the approach of data science and smart tourism applications are understood.

In another study about WHS, using text mining methods, it was aimed to determine the destination image of the Macau tourism region of China (Liu et al., 2020). Macau is an important tourist destination with its historic streets, Portuguese and Chinese style religious buildings, which was added to the UNESCO World Heritage list in 2005. In the study, 51,191 visitor reviews on websites such as Ctrip and TripAdvisor were collected and analyzed using text mining techniques. It was observed that the image of Macau changed over the years 2014-2018. It was determined, according to the results, that foreign and domestic tourists perceived the image of Macau's destination differently. It also showed that there are differences between the image of the destination projected by the government and the image of the destination perceived by the tourist. While the image of Macau was described by tourists as the "City of Culture" and "World Tourism and Entertainment Center", the official institutions were defined as "Entertainment and Gambling Center".

In order to analyze and track the changing destination image, destination management can use online data. It is possible to design destinations according to changing tourist demands.

In the tourism sector, textual content generated by tourists on social media and online platforms is of strategic importance in terms of potential tourists, tourism businesses and destination management (Aydın, 2016; Özen & İlhan, 2020). In the literature, posts with textual content are referred as user-generated content (UGC) or electronic word of mouth (eWOM) (Buzova et al., 2019; Schmunk et al., 2013).

Positive or negative reviews of tourists related to their destination experience are mentioned as unstructured data in text mining. These unstructured data are stored in large data sets on

social media and platforms. The extraction of the required information from large unstructured data sets is a very difficult process to do with traditional information systems. Text mining, regarded as the sub-discipline of Data Mining, makes an important contribution to the efforts to extract understandable and interpretable information from unstructured digital texts.

Text mining is the process of transforming unstructured data into numerical values and extracting qualified information. Text mining works with a wide range of disciplines, such as statistics, artificial intelligence and natural language processing, to analyze data from natural language texts (Miner et al., 2012). Text mining uses the methods of text classification, text clustering, topic extraction, sentiment analysis or opinion mining by using the disciplines mentioned above (Li et al., 2019).

Sentiment analysis is a method of text mining that has a vast field of study in a close relationship with the discipline of natural language processing. It allows the sentimental expressions of a text to be extracted (Oğuzlar & Kızılkaya, 2019).

Sentiment analysis is categorized as document level, sentence level, and aspect/feature-level. (Jo & Oh, 2011)

The objective of the document level sentiment analysis is to identify the positive or negative polarity of the document analyzed. Sentence level sentiment analysis is intended to determine the positive or negative polarity of a sentence. Positive or negative polarity can be defined as a reflection of the positive or negative opinions of the author of a document or sentence.

Two different approaches are used in sentiment analysis research as Machine Learning and Dictionary-Based Sentiment Analysis. Sentiment classification is the focus of the Machine Learning approach using machine learning algorithms (Medhat et al., 2014). In Dictionary-Based Sentiment Analysis, sentiment dictionaries are used. Sentiment Dictionary is a database that keeps the sentiment polarity of the terms (negative / neutral / positive) and sentiment scores numerically. Each term in the content is looked up in the sentiment dictionary while carrying out a sentiment analysis. Polarity and score are added to the total score if the term is available in the dictionary. If the total score is less than zero, it is classified as negative; if it is higher than zero, it is classified as positive; and if it is equal to zero, then

the text sentiment is classified as neutral (Hu & Liu, 2004).

Sentiment Analysis is considered as an important opportunity for tourists and suppliers to get to know and understand each other. These opportunities should be evaluated from the perspective of tourist products suppliers (tourism businesses, destination management) and consumers, i.e. tourists.

In the tourism literature, the majority of studies about tourist reviews on electronic platforms are consumer-centered. They are focused mostly on the use of social media by tourists or its impact on tourist behavior. Few studies have examined tourist comments from the viewpoint of tourism providers (Denizci Guillet, Kucukusta, & Liu, 2016). Knowledge on how tourism providers or organizations use social media is limited (Munar, 2012; Uşaklı et al., 2019). It is seen that text mining studies on destination management are genuinely for the whole destination or some tourism businesses in the destination. As an example; Irawan, Akmalia, & Masrury (2019) analysed TripAdvisor and social media reviews of tourists visiting Indonesia in their study utilizing sentiment analysis and topic extraction techniques. According to the results of the study, it was understood that the "entertainment" aspect was the most important factor in visitor experiences.

In another study, it was inferred from 19.835 online tourists reviews on the website *virtualtourist.com* utilizing sentiment analysis techniques that the most important negative assessments in Paris were about transportation (Kim et al., 2017).

Some studies using text mining techniques are trend analyzes that have been conducted in terms of destination marketing. In trend analysis; answers are sought to the questions such as what the market thinks about us, how they perceive us, what they value the most and the least, and how they compare us to our competitors. It can be beneficial for destinations to take into account the answers to these questions in order to assess their place in the market. The words used by potential tourists in search engines for a destination are seen as an important data point in trend analysis. In services such as Google Trend and Baidu, vast amounts of search data, i.e. keywords, are organized and stored. Studies use the data sets collected from these services to assess the location and trends of destinations in the tourism market.

Web search data are also used to estimate the demand for the tourism market (Li et al., 2018), the number of visitors (Gunter & Önder, 2016), tourist activity (Artola et al., 2015) and demand for hotels (Pan et al., 2012).

The research conducted by Dinis et al. (2019) can be given as an example in determining the trends towards the destinations. In the research, the interests of Spanish, British and German tourists towards Portuguese touristic products were analyzed. The analysis was carried out through categorizing by geographical locations based on the keywords used in the Google search engine. From this point of view, marketing strategies can be developed by segmenting the market. Determining the interests of potential tourists for a destination can provide a strategic competitive advantage.

In tourism literature, Text mining studies on WHS are limited. Antonio Correia, & Ribeiro (2020) analyzed and compared 8.638 online travel reviews from TripAdvisor for Salamanca (Spain) and Coimbra (Portugal), listed on the WHS, in their research. They used text mining, sentiment analysis and market basket models. In the research, keywords were defined in order to be used by the destination management and organizations for promoting these cities. In the promotion of these regions, it was emphasized that the keywords for history and world heritage such as "architectural style", "names of kings" and "place names" should be brought to the fore. The endeavors of WHS to reposition them in tourism market using the approach of data science and smart tourism applications are understood.

In another study about WHS, using text mining methods, it was aimed to determine the destination image of the Macau tourism region of China (Liu et al., 2020). Macau is an important tourist destination with its historic streets, Portuguese and Chinese style religious buildings, which was added to the UNESCO World Heritage list in 2005. In the study, 51,191 visitor reviews on websites such as Ctrip and TripAdvisor were collected and analyzed using text mining techniques. It was observed that the image of Macau changed over the years 2014-2018. It was determined, according to the results, that foreign and domestic tourists perceived the image of Macau's destination differently. It also showed that there are differences between the image of the destination projected by the government and the image of the destination perceived by the tourist. While the image of Macau was described by tourists as the "City of Culture" and "World Tourism and

Entertainment Center", the official institutions were defined as "Entertainment and Gambling Center".

In order to analyze and track the changing destination image, destination management can use online data. It is possible to design destinations according to changing tourist demands.

3. Method

The aim of this study is to evaluate TripAdvisor reviews of foreign tourists visiting Göreme National Park and Cappadocia Rocky Sites. What were the views of foreign tourists regarding the protection of natural and cultural assets in the WHS region revealed within the scope of this aim? In addition, the awareness of tourists has emerged about the fact that the place they visited is a World Heritage Site.

The application process was carried out in 5 steps. These steps are, respectively; (1) data collection, (2) model building, (3) text pre-processing analysis (4) dictionary-based sentiment analysis (5) findings and information representation. The sample of this research consists of English texts, gathered from the comments of foreign tourists about Göreme National Park and Cappadocia Region.

Data Collection

Post comments on TripAdvisor about the areas in Göreme National Park and the Cappadocia region that attract the attention of foreign tourists were investigated during the data collection process. According to the results of the research, it was observed that the locations most frequently reviewed on TripAdvisor are Göreme Open Air Museum, Zelve Open Air Museum (Paşabağ and fairy chimneys) and Kaymaklı Underground City.

TripAdvisor reviews of these regions for the time frame of 2018-2020 were collected using web scraping technique. Descriptive statistics of the collected data are presented in Table 1.

Table 1. Data Set Descriptive Statistics

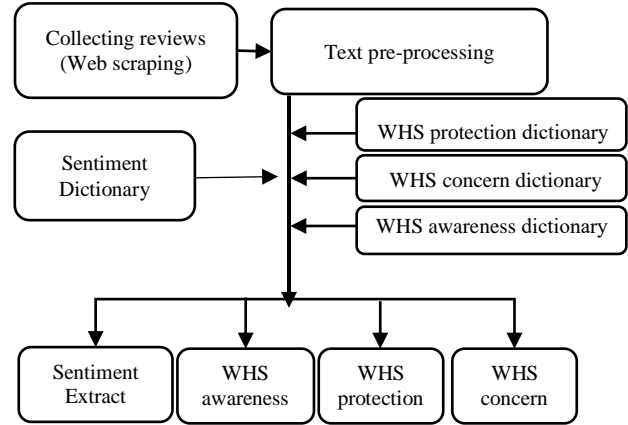
| WHS and Tourist Attraction Centers | Frequency | Total % |
|--|-------------|-------------|
| Göreme Open Air Museum | 2884 | 68.9% |
| Zelve Open Air Museum (Paşabağ and Fairy chimneys) | 1000 | 23.9% |
| Kaymaklı Underground City | 299 | 7.1% |
| Total | 4183 | 100% |
| Total number of sentences: 12.182 | | |
| Total word count: 173.052 | | |
| Words per sentence: 14.2 | | |
| Words per paragraph: 41.4 | | |

Source: Author

Collected data were transferred to the Wordstat software as the analysis software.

Model

A model was created for processing the data set transferred to Wordstat software. Application model is shown in Figure 1.



Source: Author

Figure 1. Research model

According to the model in Figure 1; collected comments were subjected to text pre-processing. The main purpose of pre-processing steps is to reduce the word count without disturbing the general meaning. The purpose of text pre-processing in text mining is to reduce the number of words in the model. Otherwise, the high-size word count model becomes impractical (Akbıyık, 2019).

This stage consists of three sub-stages. (a) Word separation; it performs the process of separating the given text into word parts through desired properties. (b) Semantic rooting; words are separated from their suffixes and reduce them to a common root. (c) Word exclusion: Words which are in the texts but do not add meaning to the sentence are removed.

After this stage, dictionary-based sentiment analysis was applied. Pre-built and tested dictionaries can be used in dictionary-based sentiment analysis. Sentiment dictionaries are techniques frequently used in sentiment extraction research.

In dictionary-based sentiment extraction, predetermined positive and negative words are compared with the words in the texts used in the model and classified as positive and negative.

In this model; the sentiment dictionary, created by Kiritchenko, Zhu, and Mohammad (2014) after evaluating 1.6 million Twitter messages, was used. By means of this dictionary, the feelings of foreign tourists towards Göreme Open Air Museum, Zelve

Open Air Museum (Paşabağ and fairy chimneys) and Kaymaklı Underground City were extracted. In addition, WHS awareness, WHS protection, and WHS concern dictionary were used in the model. These dictionaries were created by the researcher according to the term frequencies after the texts were pre-processed. Thanks to these dictionaries, the awareness of foreign tourists towards WHS, their positive opinions about the protection of these areas and their concerns about their protection were identified.

4. Findings

In this part of the research, a dictionary-based sentiment analysis was applied on the comments about the churches, fairy chimneys, valleys and underground city in WHS visited by tourists. The obtained results are presented in Table 2.

Table 2. WHS Sentiment Analysis Results

| Sentiment | Frequency | No. Cases | % Cases |
|---------------------------|-------------|-------------|---------------|
| Positive Sentiment | 5519 | 2970 | 75.40% |
| Churches | 217 | 199 | 51.5% |
| Fairy Chimneys | 59 | 58 | 14% |
| Valleys | 44 | 41 | 10.4% |
| Underground City | 38 | 38 | 8.1% |
| Negative Sentiment | 1801 | 1409 | 33.68% |
| Churches | 82 | 79 | 19.48% |
| Fairy Chimneys | 14 | 14 | 3.33% |
| Underground City | 10 | 10 | 2.13% |
| Valleys | 5 | 5 | 1.19% |

Source: Author

According to Table 2, positive sentiments were detected in 74.4% of 2970 cases and negative sentiments in 33.68% of 1409 cases in all tourist reviews. In WHS, positive sentiments were seen in Churches 51%, fairy chimneys 14%, valleys 10.4%, and underground city 8.1%. In addition, negative sentiments were seen in Churches 19%, fairy chimneys 3.3%, underground city 2.13% and valleys 1.19%.

As a result of the sentiment analysis, positive and negative words associated with WHS are presented in Table 3.

Table 3. WHS Positive and Negative words

| Word | Frequency | No. Cases | % Cases |
|---|-----------|-----------|---------|
| Positive Words | | | |
| Worth, Beaut*, Well, Great, Good, Best, Natural, Lov*, Beauty, Spot | 3096 | 2902 | 69.38% |
| Negative Words | | | |
| Extra_Payment, Crowded, Claustrophob*, Steep, Hard | 433 | 418 | 10.0% |

Source: Author

According to Table 3, positive words were seen in 2902 cases with a rate of 69% in all tourist reviews. Negative words were seen in 418 cases with a rate of 10%. The positive sentiments of the tourists were

determined by the words "Worth", "Beaut*" "Well", "Great", "Good", "Best", "Natural", "Lov*", "Beauty", "Spot". The negative sentiments of the tourists were determined by the words "Extra_Payment", "Crowded", "Claustrophob*", "Steep", "Hard".

On the other hand, according to the dictionary-based sentiment analysis results; the protection, concerns and awareness of tourists for the natural and cultural assets in WHS were identified. Analysis results are shown in Table 4.

Table 4. Protection, concern and awareness analysis results in WHS.

| About World Heritage Site | Frequency | No. Cases | % Cases |
|--|------------|------------|---------------|
| PROTECTED | 262 | 254 | 63.34% |
| Well_Preserved. Preserved. Protected. Restored. Best_Preserved | | | |
| CONCERN | 122 | 122 | 10.49% |
| Vandalism. Worry. Concerned. Painted. drawn. destroyed. destroying | | | |
| AWARENESS | 123 | 123 | 30.68% |
| Unesco*, Heritage, World Heritage Site | | | |

Source: Author

According to Table 4, it was observed that there were 254 cases at a rate of 63% for WHS protection and 122 cases at 10.49% for concern about WHS. There were 123 cases at a rate of 30.6% of the tourists' awareness of whether the locations they visited are in WHS.

Tourists have expressed their opinions about the protection of WHS using the words "Well Preserved", "Preserved", "Protected", "Restored" and "Best Preserved". On the other hand, tourists stated their concern for WHS using the words "Vandalism", "Worry", "Concerned", "painted", "drawn", "destroyed", and "destroying". Tourists also expressed their awareness about the places they visited are included in WHS by using the terms "Unesco*", "Heritage" "World_Heritage_Site".

5. Conclusion and Discussion

In this study, the views of foreign tourists visiting Göreme National Park and Cappadocia Rocky Sites in WHS were evaluated using text mining methods. According to the results of the dictionary-based sentiment analysis, the positive perceptions of the tourists about the churches, fairy chimneys, valleys and the underground city were found to be significantly high (75%). Negative perceptions of tourists were found at a low rate (33%). This situation indicates that the region is considered positively as a tourism attraction center. The words "extra payment", "crowded", "steep" stand out for the negative perceptions of tourists. Tourists found the museum entrance fees as high and disapproved the additional fees to the entrance fee in some churches. As Doğan and Karakuş (2014) pointed

out, the determination of the visitors' expectations for museums and churches serves as a precursor to meeting these expectations.

It is recommended to eliminate this negative result in terms of destination management. Moreover, it is understood that the steep slopes in the visiting areas are difficult to walk for the tourists. It is understood that there are not enough arrangements in the visiting areas. These findings support the results of the research conducted by Yılmaz (2011). Visiting areas should be arranged in such a way that especially elderly and disabled tourists can move freely without any fear (Topsakal, 2018).

In addition, tourists are negatively affected by the crowdedness of the visiting areas. As stated by Karakuş and Kalay (2017), crowding in a destination can be an undesirable situation for tourists. Destination managers and local authorities need to develop a strategy to solve the problem of excessive tourism (De Luca et al., 2020).

According to the other results of the study, it is understood that WHS is sufficiently protected. However, there are some concerns about protection. The most striking words in concerns about protection are the words "painted", "drawn". It is unacceptable to destroy the natural and cultural assets of WHS. Conservation of World Heritage Sites is an important issue (Yılmaz et al., 2020). There are some similar findings from previous research, showing the concerns of local citizens on conservation and protection of cultural and natural assets in the Göreme National Park Area (Çeşmeci and Tekeli, 2018). Necessary rules should be established to prevent these damages. Preserving the world heritage sites and transferring them to future generations is an extremely important issue for destination management. The way of transferring World Heritage Sites to future generations by allowing tourists to visit is possible with sustainable tourism (Su & Wall, 2011).

Tourist reviews used in this study consist of texts written entirely voluntarily by tourists visiting WHS. The objective and reliable data obtained from tourist reviews have enabled the exploration of tourist perceptions about WHS using text mining methods.

The original aspect of this study is the use of dictionary-based sentiment analysis for World Heritage Sites. The sentiment dictionary used in the study enables people to understand their positive or negative feelings from texts written in

social media platforms. Whereas the words "old", "cut", "complex", "ancient" express negative opinions in the dictionary, they describe the characteristics of touristic areas in WHS in this study. Therefore, the mentioned words were removed from the dictionary or evaluated as positive.

In dictionary-based sentiment analysis studies, dictionaries to be used should be created in a domain-dependent manner. In future research, it is suggested to develop dictionaries to understand the feelings of tourists in World Heritage Sites.

In this study, the views of foreign tourists visiting Göreme National Park and Cappadocia Rocky Sites in WHS were evaluated using text mining methods. According to the results of the dictionary-based sentiment analysis, the positive perceptions of the tourists about the churches, fairy chimneys, valleys and the underground city were found to be significantly high (75%). Negative perceptions of tourists were found at a low rate (33%). This situation indicates that the region is considered positively as a tourism attraction center. The words "extra payment", "crowded", "steep" stand out for the negative perceptions of tourists. Tourists found the museum entrance fees as high and disapproved the additional fees to the entrance fee in some churches. As Doğan and Karakuş (2014) pointed out, the determination of the visitors' expectations for museums and churches serves as a precursor to meeting these expectations.

It is recommended to eliminate this negative result in terms of destination management. Moreover, it is understood that the steep slopes in the visiting areas are difficult to walk for the tourists. It is understood that there are not enough arrangements in the visiting areas. These findings support the results of the research conducted by Yılmaz (2011). Visiting areas should be arranged in such a way that especially elderly and disabled tourists can move freely without any fear (Topsakal, 2018).

In addition, tourists are negatively affected by the crowdedness of the visiting areas. As stated by Karakuş and Kalay (2017), crowding in a destination can be an undesirable situation for tourists. Destination managers and local authorities need to develop a strategy to solve the problem of excessive tourism (De Luca et al., 2020).

According to the other results of the study, it is understood that WHS is sufficiently protected. However, there are some concerns about protection. The most striking words in concerns

about protection are the words "painted", "drawn". It is unacceptable to destroy the natural and cultural assets of WHS. Conservation of World Heritage Sites is an important issue (Yılmaz et al., 2020). There are some similar findings from previous research, showing the concerns of local citizens on conservation and protection of cultural and natural assets in the Göreme National Park Area (Çeşmeci and Tekeli, 2018). Necessary rules should be established to prevent these damages. Preserving the world heritage sites and transferring them to future generations is an extremely important issue for destination management. The way of transferring World Heritage Sites to future generations by allowing tourists to visit is possible with sustainable tourism (Su & Wall, 2011).

Tourist reviews used in this study consist of texts written entirely voluntarily by tourists visiting WHS. The objective and reliable data obtained from tourist reviews have enabled the exploration of tourist perceptions about WHS using text mining methods.

The original aspect of this study is the use of dictionary-based sentiment analysis for World Heritage Sites. The sentiment dictionary used in the study enables people to understand their positive or negative feelings from texts written in social media platforms. Whereas the words "old", "cut", "complex", "ancient" express negative opinions in the dictionary, they describe the characteristics of touristic areas in WHS in this study. Therefore, the mentioned words were removed from the dictionary or evaluated as positive.

In dictionary-based sentiment analysis studies, dictionaries to be used should be created in a domain-dependent manner. In future research, it is suggested to develop dictionaries to understand the feelings of tourists in World Heritage Sites.

6. References

- Akbıyık, A. (2019). *Sosyal Bilimlerde Metin Madenciliği*. Sakarya Yayıncılık.
- Akipek, S. (2001). Dünya kültürel ve doğal mirasının korunmasına dair sözleşmenin değerlendirilmesi. *Ankara Üniversitesi Hukuk Fakültesi Dergisi*, 50(4), 13–39.
- Antonio, N., Correia, M. B., & Ribeiro, F. P. (2020). Exploring User-Generated Content for Improving Destination Knowledge: The Case of Two World Heritage Cities. *Sustainability*, 12(22), 9654. <https://doi.org/10.3390/su12229654>
- Artola, C., Pinto, F., & Pedraza, P. De. (2015). Can internet searches forecast tourism inflows? *International Journal of Manpower*, 36(1), 103–116. <https://doi.org/10.1108/IJM-12-2014-0259>
- Aydın, B. (2016). Restaurant Image in Social Media: The Case of Tripadvisor. *Journal of Multidisciplinary Academic Tourism*, 1(1), 13–30. <https://doi.org/10.31822/jomat.287998>
- Buzova, D., Sanz-Blas, S., & Cervera-Taulet, A. (2019). Does culture affect sentiments expressed in cruise tours' eWOM? *The Service Industries Journal*, 39(2), 154–173. <https://doi.org/10.1080/02642069.2018.1476497>
- Çeşmeci, N., & Tekeli, E. K. (2018). Göreme Tarihi Milli Parkı'nın Sürdürülebilir Turizm Açısından Değerlendirilmesi. *Turan Stratejik Araştırmalar Merkezi Dergisi*, 10(40), 148. <https://doi.org/http://dx.doi.org/10.15189/1308-8041>
- De Luca, G., Shirvani Dastgerdi, A., Francini, C., & Liberatore, G. (2020). Sustainable Cultural Heritage Planning and Management of Overtourism in Art Cities: Lessons from Atlas World Heritage. *Sustainability*, 12(9), 3929. <https://doi.org/10.3390/su12093929>
- Demirçivi, B. M. (2017). Göreme Millî Parkı ve Kapadokya Kayalık Bölgeleri'ne İlişkin UNESCO Raporu Değerlendirmeleri ve Öneriler. *Turizm Akademik Dergisi*, 4(2), 91–106. <https://dergipark.org.tr/en/pub/touraj/issue/33509/356565>
- Denizci Guillet, B., Kucukusta, D., & Liu, L. (2016). An Examination of Social Media Marketing in China: How do the Top 133 Hotel Brands Perform on the Top Four Chinese Social Media Sites? *Journal of Travel & Tourism Marketing*, 33(6), 783–805. <https://doi.org/10.1080/10548408.2015.1064337>
- Dinis, M. G. F., da Costa, C. M. M., & Pacheco, O. M. da R. (2019). Composite indicators for measuring the online search interest by a tourist destination. In *Big Data and Innovation in Tourism, Travel, and Hospitality: Managerial Approaches, Techniques, and Applications* (pp. 1–19). Springer Singapore. https://doi.org/10.1007/978-981-13-6339-9_1
- Doğan, N. Ö., & Karakuş, Y. (2014). Evaluating the service quality in the tourism industry using QFD-AHP integrated method: an application on Göreme Open Air Museum. In *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* (Vol. 19, Issue 3). SDU İktisadi ve İdari Bilimler Fakültesi.
- Gunter, U., & Önder, I. (2016). Forecasting city arrivals with Google Analytics. *Annals of Tourism Research*, 61, 199–212. <https://doi.org/10.1016/j.annals.2016.10.007>

- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '04*, 168. <https://doi.org/10.1145/1014052.1014073>
- Irawan, H., Akmalia, G., & Masrury, R. A. (2019). Mining tourist's perception toward Indonesia tourism destination using sentiment analysis and topic modelling. *ACM International Conference Proceeding Series*, 7–12. <https://doi.org/10.1145/3361821.3361829>
- Jo, Y., & Oh, A. H. (2011). Aspect and sentiment unification model for online review analysis. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining - WSDM '11*, 815. <https://doi.org/10.1145/1935826.1935932>
- Karakuş, Y., & Kalay, N. (2017). A Study on The Concept and Causes of Destination Rejection. *International Journal of Management Economics and Business*, 13(3), 1–16. <https://doi.org/http://dx.doi.org/10.17130/ijmehb.2017331320>
- Kim, K., Park, O. joun, 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, 362–369. <https://doi.org/10.1016/j.techfore.2017.01.001>
- Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment Analysis of Short Informal Texts. *Journal of Artificial Intelligence Research*, 50, 723–762. <https://doi.org/10.1613/jair.4272>
- Li, Q., Li, S., Zhang, S., Hu, J., & Hu, J. (2019). A review of text corpus-based tourism big data mining. *Applied Sciences (Switzerland)*, 9(16). <https://doi.org/10.3390/app9163300>
- Liu, M. T., Liu, Y., Mo, Z., & Ng, K. L. (2020). Using text mining to track changes in travel destination image: the case of Macau. *Asia Pacific Journal of Marketing and Logistics*, 33(2), 379–395. <https://doi.org/10.1108/APJML-08-2019-0477>
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- Miner, G., Elder IV, J., Fast, A., Hill, T., Nisbet, R., & Delen, D. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Academic Press.
- Munar, A. M. (2012). Social Media Strategies and Destination Management. *Scandinavian Journal of Hospitality and Tourism*, 12(2), 101–120. <https://doi.org/10.1080/15022250.2012.679047>
- Oğuzlar, A., & Kızılkaya, Y. M. (2019). *Metin Madenciliğinde Duygu Analizi*. Dora Yayıncılık.
- Özen, İ. A., & İlhan, İ. (2020). Opinion Mining in Tourism: A Study on "Cappadocia Home Cooking" Restaurant. In E. Çeltek (Ed.), *Handbook of Research on Smart Technology Applications in the Tourism Industry* (pp. 43–64). IGI Global.
- Pan, B., Wu, D. C., & Song, H. (2012). Forecasting hotel room demand using search engine data. *Journal of Hospitality and Tourism Technology*, 3(3), 196–210. <https://doi.org/10.1108/17579881211264486>
- 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 International Publishing. https://doi.org/10.1007/978-3-319-03973-2_19
- Su, M. M., & Wall, G. (2011). Chinese Research on World Heritage Tourism. *Asia Pacific Journal of Tourism Research*, 16(1), 75–88. <https://doi.org/10.1080/10941665.2011.539392>
- Topsakal, Y. (2018). Akıllı Turizm Kapsamında Engelli Dostu Mobil Hizmetler: Türkiye 4.0 İçin Öneriler. *Journal of Tourism Intelligence and Smartness*, 1(1), 1–13. <https://dergipark.org.tr/en/pub/jtis/issue/39024/445128>
- Tosun, C., Dedeoğlu, B. B., Çalışkan, C., & Karakuş, Y. (2020). Role of place image in support for tourism development: The mediating role of multi-dimensional impacts. *International Journal of Tourism Research*, jtr.2405. <https://doi.org/10.1002/jtr.2405>
- UNESCO. (2020). *UNESCO World Heritage Centre - World Heritage List Statistics*. UNESCO. <http://whc.unesco.org/en/list/stat>
- Uşaklı, A., Koç, B., & Sönmez, S. (2019). Social Media Usage Among Top European DMOs. In *In Tourist Destination Management* (pp. 1–14). Springer, Cham. https://doi.org/10.1007/978-3-030-16981-7_1
- Yılmaz, İ. (2011). Müze ziyaretçilerinin hizmet kalitesi algılamaları: Göreme Açık Hava Müzesi örneği. *Anatolia: Turizm Araştırmaları Dergisi*, 22(2), 183–193. <https://dergipark.org.tr/en/pub/atad/issue/16802/174525>
- Yılmaz, İ., Tabak, G., & Samur, A. N. (2020). Vandalism problem in tourism and solution proposals: The case of Nevşehir. *Journal of Multidisciplinary Academic Tourism*, 5(2), 95–103. <https://doi.org/10.31822/jomat.706294>

INFO PAGE**Evaluation of tourist reviews on TripAdvisor for the protection of the world heritage sites: Text mining approach****Abstract**

Collecting and analyzing online tourist reviews on destinations is important for sustainable tourism. These analyses can give insight into the extent to which natural and cultural assets in the destination are protected. These evaluations should be considered by the authorities as objective and realistic assessments. In this study, 4183 TripAdvisor reviews of foreign tourists visiting "Göreme National Park and Cappadocia Rocky Area", which is listed in the World Heritage Site, were evaluated. The data set consisted of English reviews of foreign tourists visiting the region between the years of 2018 - 2020. Dictionary-based sentiment analysis, one of the text mining methods, was used in the study. According to the analysis results, the positive perceptions of the tourists about the churches, fairy chimneys, valleys and underground cities in the World Heritage Site were found to be significantly high (75%). Negative evaluations were found to be low (33%). In tourist reviews, 63% positive and 10.49% negative evaluations were made about the protection of the region. In addition, the awareness of the tourists about whether the locations they visit are World Heritage Sites was 30.6%. Tourists explained their negative opinions about the locations they visited with the words "extra_payment", "crowded", and "steep". Another finding of the study was that the area is adequately protected. There were, however, some concerns related to protection. The most striking aspects of protection concerns were those in which the words "painted" and "drawn" were used, an important indicator of visitor sensitivity to the protection of frescoes in rock churches. Since the destruction of natural and cultural assets in the region is unacceptable, recommendations are made to take the necessary measures to prevent such damage.

Keywords: World Heritage, Text mining, Dictionary-based Sentiment Analysis, Cappadocia, Göreme

Authors

| Full Name | Author contribution roles | Contribution rate |
|---------------------------|--|-------------------|
| <i>İbrahim Akın Özen:</i> | <i>Conceptualization, Methodology, Software, Validation, Formal analysis, Data Curation, Writing - Original Draft,</i> | <i>100%</i> |

Author statement: Author(s) declare(s) that All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

This paper does not required ethics committee report

Justification: This research was conducted before January 1, 2020. For this reason, it is exempt from "ULAKBİM TRDizin" criterion.