SENTIMENT CLASSIFICATION OF THE LOCAL VISITORS' SOCIAL MEDIA REVIEWS[1]

YERLİ ZİYARETÇİLERİN SOSYAL MEDYA PAYLAŞIMLARININ DUYGU SINIFLANDIRMASI

Doç.Dr. Semra AKTAŞ POLAT

İstanbul Medeniyet Üniversitesi, Turizm Fakültesi, İstanbul/Türkiye

ABSTRACT

The aim of this study is to classify the local visitors' reviews shared on social media after their visits to historical and touristic regions. In this context, sentiment analysis was conducted for social media reviews by using supervised learning approach from machine learning methods. For the analysis, the local visitors' reviews, on Istanbul and Antalya, which are Turkey's two important tourism cities, were imported from TripAdvisor website. The reviews were classified as positive, negative, and neutral. In the study, the data distribution was processed in positive, negative, and neutral classifications with KNIME software, and then it was tested with the learning method of decision trees from the supervised learning models. As a result of the decision trees learning test that is one of the supervised learning models, 75.7% was found as the success rate.

Keywords: Tourism, tourist, touristic experience, data mining, sentiment analysis, visitor reviews.

ÖZ

Bu çalışmanın amacı, tarihi ve turistik bölgelere yaptığı ziyaret sonrası sosyal medya üzerinden değerlendirmelerini paylaşan yerli ziyaretçilerin yorumlarının sınıflandırılmasıdır. Bu çerçevede, makine öğrenmesi yöntemlerinden denetimli öğrenme yaklaşımı kullanılarak sosyal medyada duygu analizi çalışması yapılmıştır. Analiz için Türkiye'nin turizm için önemli görülen iki turizm destinasyonu İstanbul ve Antalya'ya yönelik yerli ziyaretçilerin yorumları TripAdvisor.com'dan alınmıştır. Yorumlar olumlu, olumsuz ve nötr olarak işaretlenerek sınıflandırılmıştır. Çalışmada olumlu, olumsuz ve nötr sınıftaki veri dağılımı KNIME Programı'nda işlenerek denetimli öğrenme modellerinden karar ağaçları öğrenme yöntemiyle sınanmıştır. Denetimli öğrenme modellerinden Karar Ağaçları Öğrenme testi sonucunda başarı oranı %75,7 olarak bulunmuştur.

Anahtar Kelimeler: Turizm, turist, turistik deneyim, veri madenciliği, duygu analizi, ziyaretçi yorumları.

1. INTRODUCTION

Social media is a broad concept defining the different types of content that compose of social networks (Eley & Tilley, 2009, p. 78). According to Mayfield (2008) social media consists of social networks, blogs, wikis, podcasts, forums, content communities, and microblogs. Social media plays an increasingly important role as a source of information for travelers (Xiang & Gretzel, 2010, p. 179).

Knowing what other people think is an important part of information gathering behavior (Pang & Lee, 2008). "Sentiment Analysis (SA) or Opinion Mining (OM) is the computational study of people's opinions, attitudes and emotions toward an entity." and sentiment analysis is an ongoing field of research in the area of

---

[1] This study was presented as a conference paper at International Conference on New Approaches in Social Science and Humanities, October 26-28, 2018, İstanbul, Turkey

text Mining (Medhat, Hassan, & Korsahy, 2014, p. 1093). It is stated that sentiment analysis is also called opinion mining and represents the same field of study (Pang & Lee, 2008; Liu, 2012; Can & Alatas, 2017).

The platforms where millions of people share ideas about different aspects of life every day are rich data sources for sentiment analysis and opinion mining (Pak & Paroubek, 2010). The sentiment analysis is carried out for the following purposes: sentiment and subjectivity classification, vision summarization, vision acquisition, cynicism, and irony etc. (Kaynar, Yıldız, Görmez, & Albayrak, 2016). Sentiment analysis is the definition of feelings, thoughts, and evaluations as positive and negative (Wilson, Wiebe, & Hoffmann, 2005).

Sentiment analysis is carried out with two different approaches: lexicon based and machine learning. The lexicon-based approach consists of dictionary based and corpus based approaches. The corpus-based approach is divided into statistical and semantic. The machine learning approach is divided into supervised and unsupervised learning. In supervised learning, there are four main classifiers: decision tree classifiers, linear classifiers, rule-based classifiers, and probabilistic classifiers. Linear classifiers are grouped as support vector machines and neural network, and probabilistic classifiers are grouped as naive bayes, bayesian network, and maximum entropy (Medhat, Hassan, & Korsahy, 2014). The stages of sentiment analysis for product reviews are shown in Figure 1.
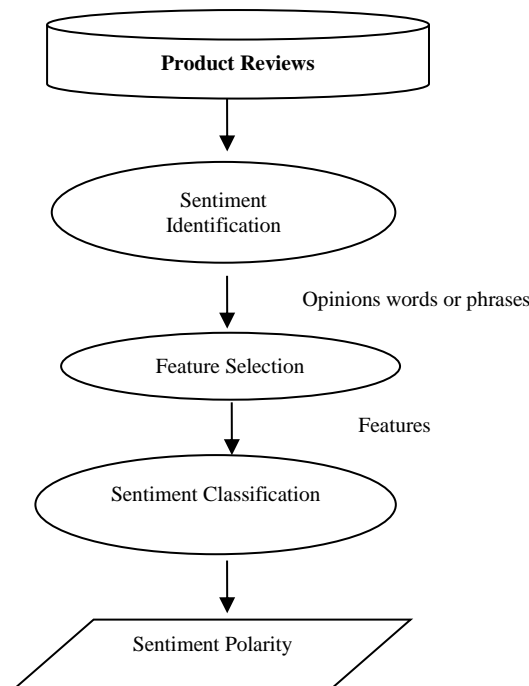

Figure 1. Sentiment Analysis process on product reviews

Source: Medhat, Hassan, & Korsahy, 2014, p. 1094

Sentiment analysis can be applied to various data sources. For examples, IMDB film reviews (Andrew et al., 2011; Kasaba & Yıldıztepe, 2016; Pang, Lee, & Vaithyanathan, 2002) and travel websites such as Twitter (Adak-Kaplan, 2016; Agarwal, Xie, Vovsha, Rambow, & Passonneau, 2011, Çoban, Özyer, & Tümüklü-Özyer, 2015; Go, Bhayani, & Huang, 2009; Kouloumpis, Wilson, & Moore, 2011; Meriç & Diri, 2014, Nizam & Akın, 2014; Onan, 2017) TripAdvisor and Booking.com (Gao, Hao, & Fu, 2015; Oğul & Ercan, 2016).

2. RESEARCH METHODOLOGY

2.1. Purpose of the research

The purpose of this research is to classify the local visitors' reviews shared on social media after their visits to historical and touristic regions. In this context, sentiment analysis was conducted for social media reviews by using supervised learning approach from machine learning methods.

2.2. Sampling and Data Collection

This study has focused on the analysis of local visitors' sentiments. Within the scope of research, the local visitors' reviews shared on TripAdvisor.com were examined after their visits to the historical and touristic

regions of Istanbul and Antalya, in 2018. It is assumed that the visitors those wrote their reviews in Turkish are local visitors.

The best destinations were chosen from TripAdvisor website to visit in Istanbul and Antalya. In this context, 5,000 and more reviewed destinations for Istanbul, and 1,000 and more reviewed destinations for Antalya were included in the study. In selection of the reviews, the gender variable was determined as a prerequisite. The reviews of internet users who do not remark gender on TripAdvisor.com are not included in the study. In addition, the reviews cover the comments shared in 2018. In this context, 625 reviews for Istanbul and 562 reviews for Antalya were included in the analysis. In total 1,187 reviews were taken from TripAdvisor.com.

3. FINDINGS

Table 1 presents the distribution of reviews according to city and gender variables. According to Table 1, 52.6% of the total reviews belong to İstanbul and 47.4% to Antalya. Moreover, 71.1% of the comments were made by males and 28.8% by female visitors.

Table 1. Distribution of Reviews by City and Gender

City	Male		Female		Total	
	n	%	n	%	n	%
İstanbul	437	36.8	188	15.8	625	52.6
Antalya	408	34.3	154	13	562	47.4
Total	845	71.1	342	28.8	1187	100

The data set obtained from the reviews was tailored for the analysis by normalizing, transliterating, and removing the stop words. Moreover, positive and negative dictionary of words was prepared by creating the term document matrix. After this decomposition process, reviews were divided into three categories as positive, negative, and neutral through Excel. Finally, the data recorded in Excel format were analyzed through KNIME. Figure 2 shows the data analysis workflow in KNIME.
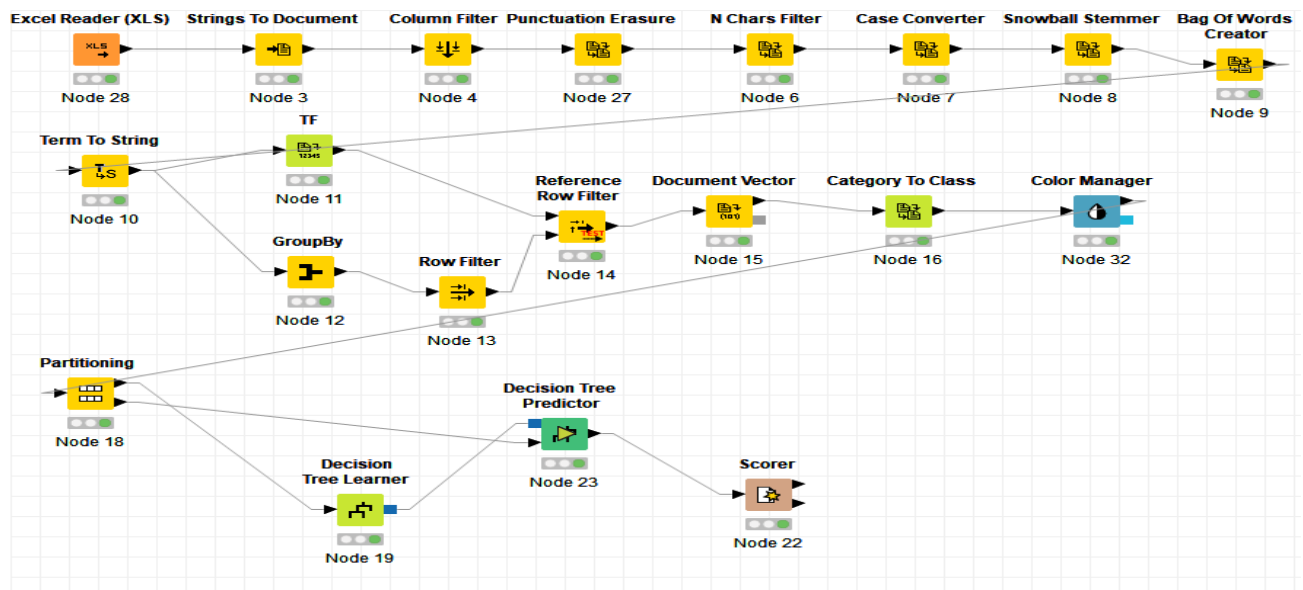

Figure 2. The data analysis workflow in KNIME

The data included in the program with the excel reader command were tested with the process flow shown in Figure 2. After processing, the data were divided two portions as training data (66%) and test data (33%) by the portioning command. And then, the decision tree method, which is one of the supervised learning models, was tested. Decision tree learner and decision tree predictor were used to analyze the sentiment of the data and the results were obtained by scorer command.

Table 2: Sentiment analysis results by decision tree

Prediction	Decision Tree Results				
	True Positives	False Positives	True Negatives	False Negatives	Accuracy
Positives	306	98	0	0	
Negatives	0	0	391	13	75.7%
Neutral	0	0	319	85	

In Table 2, sentiment analysis results are presented according to the decision tree results. As a result of the research, 75.7% was achieved as a success rate.

4. CONCLUSION

Seker (2015, p. 38) emphasizes that techniques for data mining will be fed from a wide range of disciplines such as behavioral sciences, social sciences or business. According to the statement of Gao, Hao, and Fu (2015), sentiment analysis of tourists and local people plays an important role in tourism development. User created content by internet users is considered a credible rumor (Stepchenkova & Zhan, 2013), and it plays an important role in the planning process of travel (O'Connor, 2008).

As a result of study conducted for the reviews of local visitors who shared their evaluations and thoughts on TripAdvisor.com after their visits to the historical and touristic regions in İstanbul and Antalya, the success rate was found 75.7%. Identifying, estimating, and controlling the behavior of individuals in social sciences is an important detail. Consumer experiences are an important element of social and economic issues (Quan & Wang, 2004, p. 297). At this point, knowing the emotions and thoughts of consumers in areas such as tourism marketing, tourism management etc. will be useful for tourism management and marketing of tourism products and services.

The success rate obtained from the sentiment analysis of the reviews shared by the local visitors will be a source of information for the local authorities and destination management/marketing organizations trying to promote the destinations in which they operate. Local visitors' reviews may contribute to the discovery of some issues that local administrators cannot see. Moreover, private sector representatives should also pay attention to the reviews shared on the web by visitors due the fact that the visitors' reviews indicate their intentions to re-visit the same region. In other words, a decrease in the number of visitors leads to a decrease in revenues. In conclusion, sentiment analysis can be used as one of the important tools in tourism development.

REFERENCES

Adak-Kaplan, B. (2016). Twitter üzerindeki Türkçe mesajlarda veri madenciliğiyle duygu analizi. *Yayınlanmamış Yüksek Lisans Tezi*. İstanbul: Beykent Üniversitesi Fen Bilimleri Enstitüsü.

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. In, *Proceedings of the Workshop on Languages in Social Media* (pp. 30-38). Association for Computational Linguistics.

Andrew, L. M., Raymond, E. D., Peter, T. P., Dan, H., Andrew, Y. N., & Christopher, P. (2011). Learning word vectors for Sentiment Analysis. In, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, (pp. 142–150), Portland, Oregon, June 19-24, 2011.

Can, U., & Alatas, B. (2017). Duygu Analizi ve Fikir Madenciliği algoritmalarının incelenmesi. *Int. J. Pure Appl. Sci. 3(1),* 75-111.

Çoban, Ö., Özyer, B., & Tümüklü-Özyer, G. (2015). Sentiment Analysis for Turkish Twitter feeds. *23nd Signal Processing and Communications Applications Conference (SIU).*

Eley, B., & Tilley, S. (2009). *Online Marketing Inside Out.* Austraila, Sitepoint Pty. Ltd.

Gao S., Hao, J., & Fu, Y. (2015). The application and comparison of web services for Sentiment Analysis in tourism. In *Service Systems and Service Management (ICSSSM), 2015 12th International Conference on* (pp. 1-6). IEEE.

Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report*, Stanford, 1-12.

Kasaba, E., & Yıldıztepe, E. (2016). Destek vektör makinesi yöntemi ile bir duygu çözümlemesi. *Akademik Bilişim*.

Kaynar, O., Yıldız, M., Görmez, Y., & Albayrak, A. (2016). Makine öğrenmesi yöntemleri ile Duygu Analizi. *International Artificial Intelligence and Data Processing Symposium (IDAP'16)*, September 17-18, Malatya/TURKEY, pp. 234-241.

Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the omg!. *Fifth International AAAI Conference on Weblogs and Social Media, 11,* 538-541.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Mayfield, A. (2008). *What is social media? iCrossing*. http://www. icrossing. co. uk/fileadmin/uploads/eBooks/What_is_Social_Media_ Crossing_ebook. pdf.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment Analysis algorithms and applications: A survey. *Ain Shams Engineering Journal, 5,* 1093-1113. doi.org/10.1016/j.asej.2014.04.011

Meriç, M., & Diri, B. (2014). Twitter üzerinde duygu analizi. *IEEE 22. Sinyal İşleme ve İletişim Uygulamaları Kurultayı,* s. 23-25.

Nizam, H., & Akın, S. S. (2014). Sosyal medyada makine öğrenmesi ile duygu analizinde dengeli ve dengesiz veri setlerinin performanslarının karşılaştırılması. *XIX. Türkiye'de İnternet Konferansı.*

O'Connor, P. (2008). User-Generated content and travel: A case study on Tripadvisor.com. In O'Connor P., Höpken W., Gretzel U. (eds), *Information and Communication Technologies in Tourism*, Springer, Vienna, pp. 47-58. doi.org/10.1007/978-3-211-77280-5_5

Oğul, B. B., & Ercan, G. (2016). Türkçe otel yorumlarından Duygu Analizi. *24th Signal Processing and Communication Application Conference (SIU),* 16-19 May 2016, Zonguldak, Turkey. doi.org/10.1109/SIU.2016.7495786

Onan, A. (2017). Sentiment Analysis on twitter messages based on Machine Learning Methods. *Yönetim Bilişim Sistemleri Dergisi, 3(2).*

Quan, S., & Wang, N. (2004). Towards a structural model of the tourist experience: An illustration from food experiences in tourism. *Tourism Management 25,* 297–305. doi:10.1016/S0261-5177(03)00130-4

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for Sentiment Analysis and opinion mining. In *LREc, 10(2010),* 1320-1326.

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

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. *Proceedings of the ACL-02 conference on Empirical methods in Natural Language Processing,* Philadelphia, July 2002, pp. 79-86. Association for Computational Linguistics.

Seker, S. E. (2015). Sosyal ağlarda veri madenciliği (Data mining on social networks). *YBS Ansiklopedi, 2(2),* 30-39.

Stepchenkova, S., & Zhan, F. (2013). Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. *Tourism Management, 36,* 590-601. doi.org/10.1016/j.tourman.2012.08.006

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 347-354). Association for Computational Linguistics.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management, 31(2),* 179-188. doi.org/10.1016/j.tourman.2009.02.016