

# What Happened in Turkey After Booking.com Limitation: Sentiment Analysis of Tweets via Text Mining



Ekin Akkol, Serkan Alici, Can Aydin, and Cigdem Tarhan

**Abstract** Twitter is one of the most popular applications for sharing feelings and opinions. Sentiment analysis also known as opinion mining is basically used to classify text into three or more categories: positive, negative and neutral sentiments. In this study, sentiment analysis is tested on tweets about [www.booking.com](http://www.booking.com) in Turkey after the court decided to stop the activities of Booking.com. Moreover, after the date that Booking.com stops its services, traffic data of other major Web sites serving in this sector has been obtained and how they are influenced by this activity is also interpreted. As a result of the literature, sentiment analysis on English texts is a highly popular and well-studied topic; however, it has been observed that the study of text mining in Turkish language is limited. The data is obtained on Twitter from starting the date that Booking.com closures in Turkey. The Twitter messages in Turkish were manually obtained from the Internet because of being expensive of old tweet data. The data has been passed through the preprocessing, attribute selection and classification stages. At the end of these processes, the data is analyzed using various text mining algorithms so the success rates achieved are compared and interpreted.

**Keywords** Booking.com · Sentiment analysis · Twitter · Text mining · Machine learning · Turkish tweet

## 1 Introduction

In the twenty-first century, social media platforms offer powerful tools that people can easily share their feelings and opinions about various topics with large crowds. Therefore, the social media usage has become an important part of daily routine in our lives (Gaál et al. 2015; Anstead and O’Loughlin 2015; Chen et al. 2011). Sentiment analysis is known as opinion mining or emotion artificial intelligence in the literature (Yang and Lin 2018; Appel et al. 2018; Öztürk and Ayvaz 2018;

---

E. Akkol · S. Alici · C. Aydin · C. Tarhan (✉)  
Department of Management Information Systems, Dokuz Eylul University, Izmir, Turkey  
e-mail: [cigdem.tarhan@deu.edu.tr](mailto:cigdem.tarhan@deu.edu.tr)

© Springer Nature Switzerland AG 2020

M. Janowicz-Lomott et al. (eds.), *Economic and Financial Challenges for Balkan and Eastern European Countries*, Springer Proceedings in Business and Economics, [https://doi.org/10.1007/978-3-030-39927-6\\_18](https://doi.org/10.1007/978-3-030-39927-6_18)

291

Zheng et al. 2018; Houlihan and Creamer 2017; El-Masri et al. 2017; Geetha et al. 2017; Ma et al. 2017). It is based on the usage of natural language processing, text analysis, computational linguistics and biometrics to systematically identify, extract, quantify and study affective states and subjective information (Antonio et al. 2018; Hameed et al. 2018; Ruan et al. 2018).

Information technology (IT)-based social media data analysis has affected a company’s ability to discover their social media intelligence (Lee 2018). As an example of this type of studies, sentiment analysis is performed to customers’ online or written reviews and survey responses. It has a wide range area for different disciplines, especially for marketing (Abdi et al. 2018; Li 2018; Ruan et al. 2018). Sentiment analysis is classifying the polarity of a given text at the document, sentence or feature/aspect level—whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative or neutral (Daniel et al. 2017; Oscar et al. 2017; Nguyen and Jung 2017).

Almost 70% of adult Internet users use social media, and this percentage is increasing (Pew-Research 2018). Twitter is one of the most popular applications for sharing feelings and opinions (Mondal et al. 2017; Chappel et al. 2017; LaPoe et al. 2017; Sul et al. 2016). According to the ‘Digital in 2017 Global Overview,’ 48 million people use social media actively in Turkey (Digital in 2017 Global Overview 2018). Therefore, social media is an important source of data to analyze people’s feelings about the events that create the country’s agenda. The main reason for this is the intensive use of social media, and there is a great deal of power to spread the news about the agenda directly from social media. Figure 1 shows the social media statis-

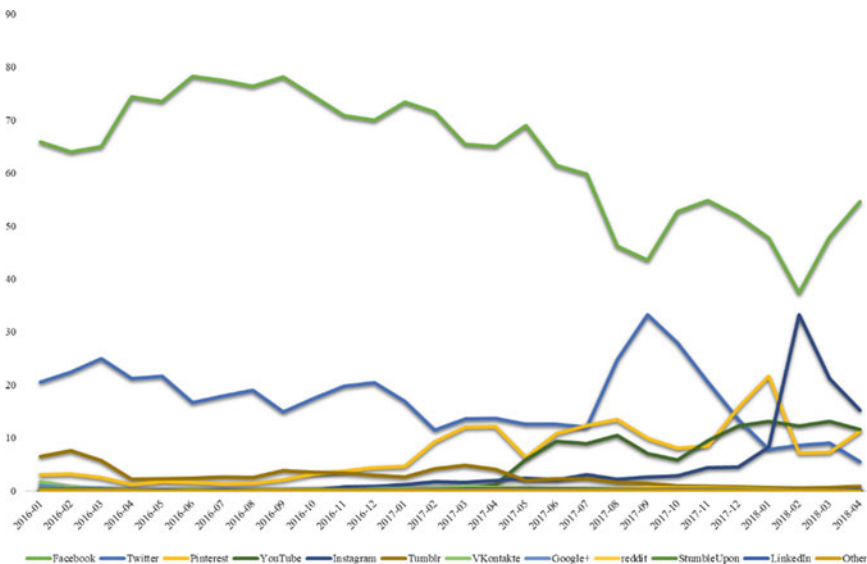


Fig. 1 Social media statistics of Turkey from January 2016 to April 2018 (StatCounter—Global-Stats 2018)

tics of Turkey from January 2016 to April 2018. According to the statistics, Twitter is the second social media platform used in Turkey with 17.67% ratio.

With the increasing popularity of the Internet, tourism facilities have become more digital with increased interconnections between customers, suppliers and firms (Zhou et al. 2014; Philips et al. 2016; Podnar and Javernik 2012; Filieri and McLeay 2013). Therefore, in order to understand why customers choose a product or a service, social media data analysis plays an important role in competitive advantage (Brooks et al. 2014). At this point, Booking.com is an example of IT-based tourism facilities. Booking.com firm was established in 1996 in Amsterdam and has grown from a small Dutch start-up to one of the largest travel e-commerce companies in the world. Now, the firm employs more than 15,000 employees in 198 offices in 70 countries worldwide. The Booking.com Web site and mobile apps are available in over 40 languages, offer 1,742,015 properties and cover 130,452 destinations in 227 countries and territories worldwide. Each day, more than 1,550,000 room nights are reserved on its platform (Booking 2018). Moreover, the firm has long-term relationship with more than 560,000 hotels worldwide, 40,000,000+ guest reviews, 750,000 rooms booked per day, # 1 most visited travel site by traffic, 100+ million visits a month and access to over 180 countries (DPO 2018).

In 2017, an Istanbul court ordered the suspension of the activities of the Web site ([www.booking.com](http://www.booking.com)) in Turkey on March 29, citing accusations of unfair competition, following a lawsuit filed by the Association of Turkish Travel Agencies (TURSAB). At the end of the lawsuit, it has been concluded to limit [www.booking.com](http://www.booking.com)'s services for the hotel search and booking in Turkey since 2017. The Web site, which had around 13,000 hotel members from Turkey, halted selling rooms in Turkey to Turkish users on March 30, one day after the court decided to block the Web site in the country. The Web site can still be used from foreign countries to make reservations for Turkish hotels and from Turkey to make reservations abroad (Figs. 2 and 3). According to a sector player, Turkey's city hotels take around 35% of their reservations via Web

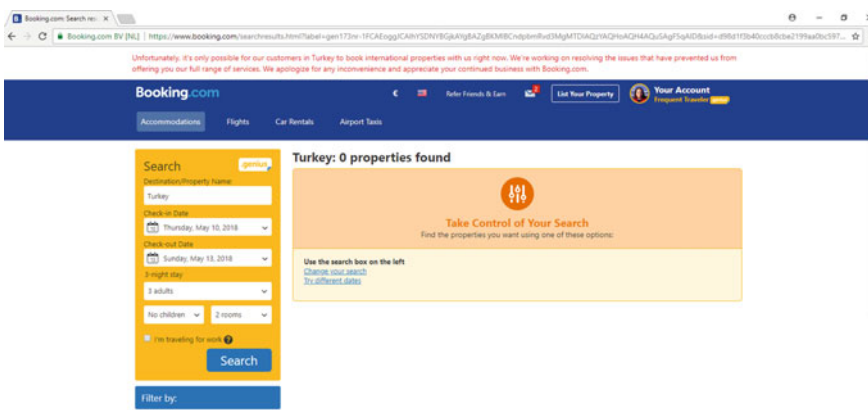


Fig. 2 Booking.com Web site homepage in Turkey

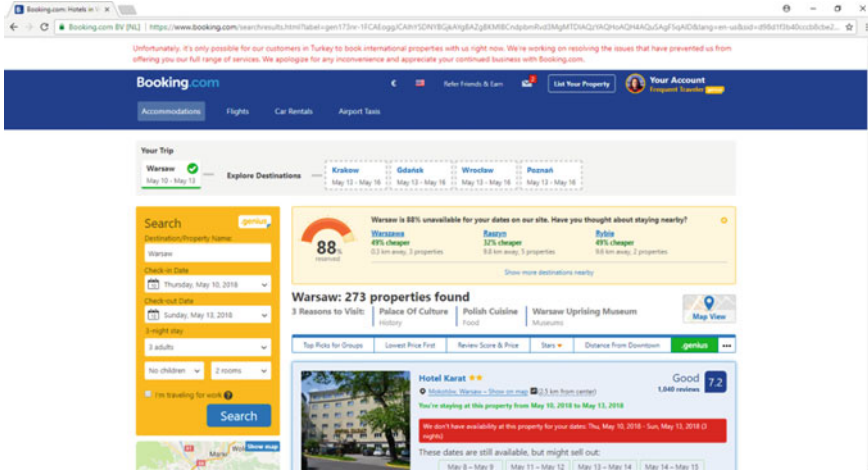


Fig. 3 Booking.com Web site search example for Warsaw

sites, with Booking.com taking a large share of this total (Independent 2018; Hotel Management 2018; Hurriyet 2018).

In this study, sentiment analysis is tested on Turkish tweets about [www.booking.com](http://www.booking.com) in Turkey after the court decided to stop the activities of Booking.com. Moreover, after the date that Booking.com stops its services, traffic data of other major Web sites serving in this sector has been obtained and how they are influenced by this activity is also interpreted. The data is obtained on Twitter from starting the date that Booking.com closures in Turkey. The Twitter messages in Turkish were manually obtained from the Internet because of being expensive of old tweet data. The data has been passed through the preprocessing, feature selection and classification stages. At the end of these processes, the data is analyzed using various text mining algorithms so the success rates achieved are compared and interpreted.

## 2 Web Traffic of Online Reservation in Turkey After Booking.com Limitation

There are two main aims in this study. The first one is to determine the emotional analysis of the customers after Booking.com in Turkey; the other one is to analyze how other companies in the same sector are affected by this activity. According to Turkish media reports, up to 30% increase in sales of other companies has been observed after stopping the activities of Booking.com (Posta 2018). In order to be able to determine how other companies in the sector are affected, traffic data of the Web pages of the other companies with a high market share in Turkey was obtained from [www.semrush.com](http://www.semrush.com) Web site.

The numbers of ‘unique visitors’ of the pages are taken into account and interpreted. The number of unique visitors is the number of entries on the same rope over a given period of time, counted as a single entry. The data obtained before and after stopping the activities of Booking.com was analyzed. The company has ceased its activities in Turkey on March 29, 2017. Table 1 shows the traffic analytics of unique visitors between April 2016 and October 2017 (k).

Traffic data shows (Fig. 4) that the density of the Web pages of other companies has started to increase visibly from April 2017. Although this is thought to be related to the opening of the summer season and many other factors, it is observed that there is an increase in the number of unique visitors compared to the data in the summer of 2017, which appears in the graphs. Therefore, stopping the activities of Booking.com has led users to intensively refer to the sites of other holiday agencies, which has a favorable effect on other agencies. Figure 4, shown, represents the other firms’ Web traffic starting from April 2016 to October 2017. According to the statistics, the firms called ‘ETS Tur, Jolly Tur, Neredekal.com, Tatilbudur.com, tatilsepeti.com, Tatil.com, trivago and Am Tur’ have increased their unique visitor numbers.

### 3 Sentiment Analysis of Turkish Tweets via Text Mining

In this study, the data set has been generated from Twitter messages about Booking.com after stopping its activities in Turkey. The data was obtained manually, filtering from Twitter in Turkish. The reason for the manual acquisition of the data is the need for tweet in the past and in large quantities, which leads to huge costs. In the tweets obtained, the parts that are not used in the study such as user name and liking are separated and only a data set consisting of messages is created. Messages in the generated data set are grouped into three categories: positive, negative and neutral. After text preprocessing—cleaning step, the data set of the study is composed of 2000 tweets. The results show that 382 of 2000 tweets are positive, 1274 of them are negative and 344 of them are neutral opinions.

Firstly, the typing mistakes in the data set and the correction of the marking mistakes have been corrected. In addition to this step, the abbreviation is made when the tweet is thrown, and the letter repetition made to emphasize the written word is also corrected. RapidMiner software was used for further preprocessing, machine learning, classification and analysis phases. With the RapidMiner, all letters in the data set are converted to lowercase, clearing of unnecessary characters such as @ and #, clearing of punctuation marks, clearing of words with more or less than a certain number of characters, clearing of stall words that do not make sense in working according to the generated stall word dictionary, identification of roots, disintegration of data. The Snowball library, which was developed for the Turkish language, was used to determine the roots of the words. N-gram model was used for feature selection process. Then, the term frequency weighting method was applied to determine how many times a term has passed in the data set. The supervised learning technique of the machine learning method is used in the study.

**Table 1** Traffic analytics: unique visitor table—April 2016–October 2017 (k)

	ETS Tur	Jolly Tur	Neredekal.com	Tatilbudur.com	tatilsepeti.com	Tatil.com	trivago.com	anitur.com.tr
April 2016	100	50	40	70	70	70	50	25
May 2016	90	40	35	50	60	60	40	20
June 2016	105	60	40	55	70	70	50	35
July 2016	100	40	38	55	60	60	55	20
August 2016	98	45	36	60	70	70	55	30
September 2016	90	30	30	50	50	50	40	20
October 2016	80	40	35	60	50	50	45	24
November 2016	90	80	180	110	90	90	200	40
December 2016	230	75	150	115	100	100	220	40
January 2017	220	110	140	120	130	130	200	70
February 2017	260	140	145	120	200	150	260	90
March 2017	300	150	140	115	130	130	240	60
April 2017	250	110	170	125	250	180	350	75
May 2017	450	180	190	280	350	300	600	120
June 2017	700	270	250	290	350	350	650	110
July 2017	650	250	260	310	330	300	700	40
August 2017	675	260	300	280	300	280	600	30
September 2017	575	240	250	170	150	150	280	20
October 2017	300	120	180	100	140	140	260	20

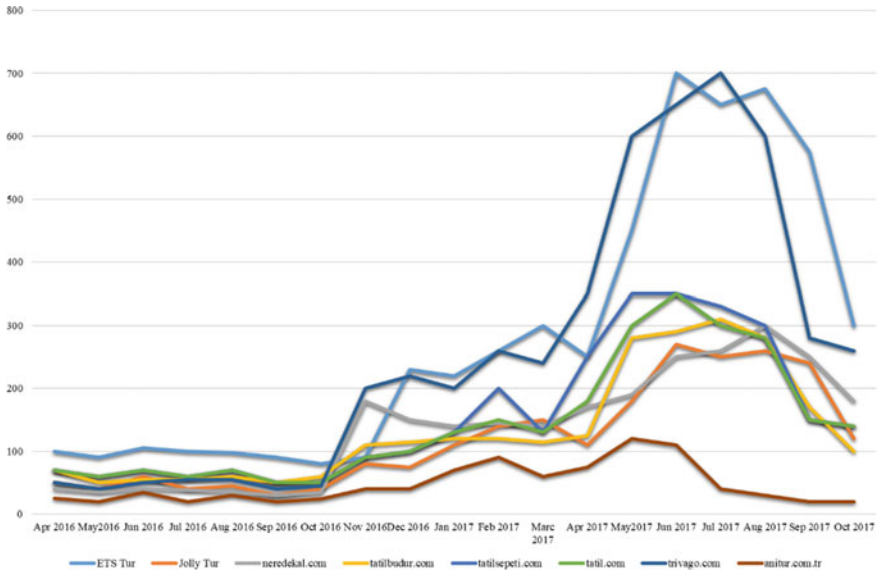


Fig. 4 Traffic analytics: unique visitor chart—April 2016–October 2017 (k)

After all these steps have been completed, experimental results have been obtained by Naive Bayes (kernel), gradient boosted tree, Naive Bayes, k-nearest neighbors (KNNs), sequential minimal optimization (SMO), random forest, decision tree (J48) algorithms. For each algorithm, training data set was selected as shuffled sampling at 0.7, 0.5 and 0.3 ratios and analyzed separately. The success rates obtained as a result of the analyses made are compared. Figure 5 shows the accuracy rate.

## 4 Conclusion

Turkish tweet sentiment problem is a challenging problem with the fact that the Turkish expressions are short and contain different interpretations in terms of semantics. At the classification stage, the attributes were tested in three different ways for the N-gram model: 2, 3 and 4 grams. The best result is observed from 3-gram model. The success rates of the study were affected from the out of balance of the data set. The existing negative sentences in the data set cause increase in the success rates predicted by negative cues. In this study, machine learning-based approaches were used for sentiment analysis. The highest success rate with 79.29% was obtained with sequential minimal optimization (SMO) algorithm. It was also seen that the highest success rate was always obtained when 0.7 training sets were selected. It is obviously determining that the success rates of the Naive Bayes (kernel) algorithm are almost the same as the SMO algorithm. Among the algorithms used in the study, the KNN algorithm is the most unsuccessful algorithm because of giving unsuccessful results.

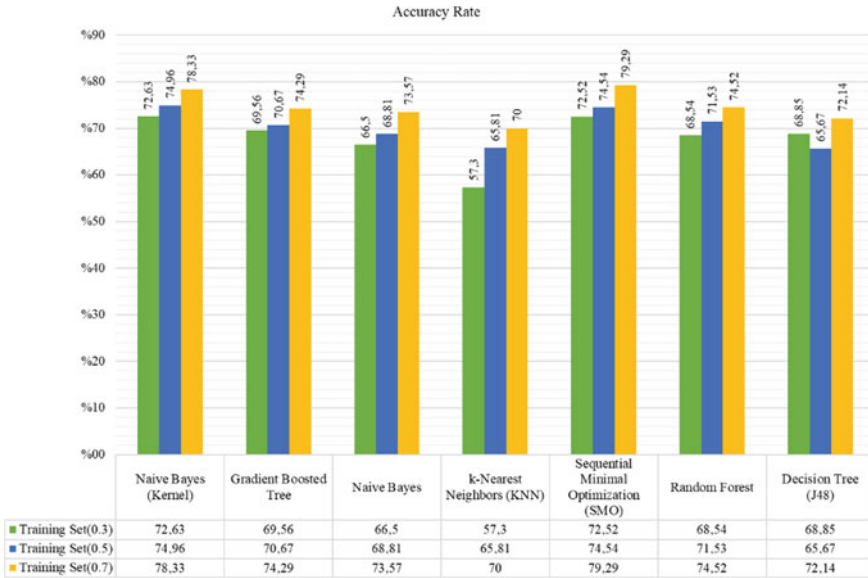


Fig. 5 Accuracy rate

In addition to the accuracy rate, the kappa statistical results obtained are also compared. The kappa statistic ranges from +1 to -1, but gives the relationship between the observed compliance and the chance-based compliance among the classes. When kappa statistic is equal to individual, full harmony is mentioned, while if it is greater than zero, harmony observed is greater or equal to harmony depending on chance. If the kappa statistic is less than zero, it is understood that the classification is not reliable (Aha and Kibler 1991; Nizam and Akin 2014). Table 2 shows the kappa statistical results of the algorithms used in the study. The kappa statistic of sequential minimal optimization, which is the most successful algorithm, was measured as 0.583. This value is an indicator that the classification is reliable.

Table 2 Kappa statistical values

	0.3	0.5	0.7
Naive Bayes (kernel)	0.478	0.520	0.577
Gradient boosted tree	0.343	0.356	0.391
Naive Bayes	0.353	0.409	0.487
k-nearest neighbors (KNNs)	0.301	0.389	0.451
Sequential minimal optimization (SMO)	0.455	0.491	0.583
Random forest	0.285	0.372	0.416
Decision tree (J48)	0.339	0.317	0.407



It is known that the training data used in the classification run and the attributes extracted from the data set are direct effects. For this reason, in the future studies, it is considered to construct a data set consisting of samples with much better discriminability. It is also contemplated to apply additional semantic and mathematical methods in order to increase the size of the data by increasing the number of messages, thereby increasing the ability of the classifiers to generalize, and extracting attributes with better distinguishing characteristics at the feature extraction stage.

## References

- Abdi, A., Shamsuddin, S. M., & Aliguliyev, R. M. (2018). QMOS: Query-based multi-documents opinion-oriented summarization. *Information Processing and Management*, *54*, 318–338.
- Aha, D., & Kibler, D. (1991). Instance-based learning algorithms. *Machine Learning*, *6*(1).
- Anstead, N., & O’Loughlin, B. (2015). Social media analysis and public opinion: The 2010 UK general election. *Journal of Computer-Mediated Communication*, *20*, 204–220.
- Antonio, N., Almedia, A., Nunes, L., Batista, F., & Ribeiro, R. (2018). Hotel online reviews: Different languages, different opinions. *Information Technology & Tourism*, *18*(1–4), 157–185. <https://doi.org/10.1007/s40558-018-0107-x>.
- Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2018). Successes and challenges in developing a hybrid approach to sentiment analysis. *Applied Intelligence*, *48*(5), 1176–1188. <https://doi.org/10.1007/s10489-017-0966-4>.
- Booking. (2018). Retrieved January 18, 2018, from <http://www.booking.com>.
- Brooks, G., Heffner, A., & Henderson, D. (2014). A SWOT analysis of competitive knowledge from social media for a small startup business. *Review of Business Information Systems*, *8*(1), 23–34.
- Chappel, P., Tse, M., Zhang, M., & Moore, S. (2017). Using GPS geo-tagged social media data and geodemographics to investigate social differences: A Twitter pilot study. *Sociological Research Online*, *22*(3), 38–56. <https://doi.org/10.1177/1360780417724065>.
- Chen, Y., Fay, S., & Wang, Q. (2011). The role of marketing in social media: How online consumer reviews evolve. *Journal of Interactive Marketing*, *25*, 85–94.
- Daniel, M., Neves, R. F., & Horta, N. (2017). Company event popularity for financial markets using Twitter and sentiment analysis. *Expert Systems with Applications*, *71*, 111–124. <https://doi.org/10.1016/j.eswa.2016.11.022>.
- Digital in 2017 Global Overview. (2018). Retrieved January 18, 2018, from <https://wearesocial.com/special-reports/digital-in-2017-global-overview>.
- DPO. (2018). Retrieved January 18, 2018, from <http://blog.directpay.online/booking-com>.
- El-Masri, M., Altrabsheh, N., & Mansour, H. (2017). Successes and challenges of Arabic sentiment analysis research: A literature review. *Social Network Analysis Mining*, *7*, 54. <https://doi.org/10.1007/s13278-017-0474-x>.
- Filieri, R., & McLeay, F. (2013). E-WOM and accommodation an analysis of the factors that influence travelers’ adoption of information from online reviews. *Journal of Travel Research*, *53*(1), 44–57.
- Gaál, Z., Szabó, L., Obermayer-Kovács, N., & Csepregi, A. (2015). Exploring the role of social media in knowledge sharing. *The Electronic Journal of Knowledge Management*, *13*(3), 185–197.
- Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels—An empirical analysis. *Tourism Management*, *61*, 43–54.
- Hameed, M., Tahir, F., & Shahzad, M. A. (2018). Empirical comparison of sentiment analysis techniques for social media. *International Journal of Advanced and Applied Sciences*, *5*(4), 115–123.

- Hotel Management. (2018). Retrieved January 18, 2018, from <https://www.hotelmanagement.net/development/why-turkey-s-hotels-are-uniting-behind-booking-com>.
- Houlihan, P., & Creamer, G. G. (2017). Can sentiment analysis and options volume anticipate future returns? *Computational Economics*, *50*, 669–685. <https://doi.org/10.1007/s10614-017-9694-4>.
- Hurriyet. (2018). Retrieved January 18, 2018, from <http://www.hurriyetdailynews.com/open-branch-in-turkey-to-avoid-ban-association-tells-bookingcom-111663>.
- Independent. (2018). Retrieved January 18, 2018, from <https://www.independent.co.uk/travel/news-and-advice/bookingcom-turkey-priceline-hotels-court-istanbul-ankara-ban-competition-authority-a7658251.html>.
- LaPoe, V. L., Olson, C. C., & Eckert, S. (2017). LinkedIn is my office; Facebook my living room, Twitter the neighborhood bar. *Journal of Communication Inquiry*, *41*(3), 185–206. <https://doi.org/10.1177/0196859917707741>.
- Lee, I. (2018). Social media analytics for enterprises: Typology, methods, and processes. *Business Horizons*, *61*, 199–210. <https://doi.org/10.1016/j.bushor.2017.11.002>.
- Li, L. (2018). Sentiment-enhanced learning model for online language learning system. *Electronic Commerce Research*, *18*, 23. <https://doi.org/10.1007/s10660-017-9284-5>.
- Ma, B., Yuan, H., & Wu, Y. (2017). Exploring performance of clustering methods on document sentiment analysis. *Journal of Information Science*, *43*(1), 54–74.
- Mondal, M., Messias, J., Ghosh, S., Gummadi, K. P., & Kate, A. (2017). Managing longitudinal exposure of socially shared data on the Twitter social media. *International Journal of Advances in Engineering Sciences and Applied Mathematics*, *9*, 238. <https://doi.org/10.1007/s12572-017-0196-3>.
- Nguyen, H. L., & Jung, J. E. (2017). Statistical approach for figurative sentiment analysis on social networking services: A case study on Twitter. *Multimedia Tools and Applications*, *76*, 8901–8914. <https://doi.org/10.1007/s11042-016-3525-9>.
- Nizam, H., & Akın, S. S. (2014). Sosyal medyada makine öğrenmesi ile duygu analizinde dengeli ve dengelessiz veri setlerinin performanslarının karşılaştırılması. In *XIX. Türkiye’de İnternet Konferansı*.
- Oscar, N., Fox, P. A., Croucher, R., Wernick, R., Keune, J., & Hooker, K. (2017). Machine learning, sentiment analysis, and tweets: An examination of Alzheimer’s disease stigma on Twitter. *The Journals of Gerontology: Series B*, *72*(5), 742–751. <https://doi.org/10.1093/geronb/gbx014>.
- Öztürk, N., & Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telematics and Informatics*, *35*, 13–147.
- Pew-Research. (2018). Retrieved January 18, 2018, from <http://www.pewinternet.org/fact-sheet/social-media/>.
- Philips, P., Barnes, S., Zigan, K., & Schegg, R. (2016). Understanding the impact of online reviews on hotel performance: An empirical analysis. *Journal of Travel Research*, *56*(2), 235–249. <https://doi.org/10.1177/0047287516636481>.
- Podnar, K., & Javernik, P. (2012). The effect of word of mouth on consumers’ attitudes toward products and their purchase probability. *Journal of Promotion Management*, *18*(2), 145–168.
- Posta. (2018). Retrieved January 18, 2018, from <http://www.posta.com.tr/booking-com-yasagi-rakiplerine-yaradi-haberi-1292278>.
- Ruan, Y., Durressi, A., & Alfantouk, L. (2018). Using Twitter trust network for stock market analysis. *Knowledge-Based Systems*, *145*, 207–218.
- StatCounter—GlobalStats. (2018). Retrieved January 18, 2018, from <http://gs.statcounter.com/social-media-stats/all/turkey/#monthly-201601-201804-bar>.
- Sul, K. S., Dennis, A. R., & Yuan, L. I. (2016). Trading on Twitter: Using social media sentiment to predict stock returns. *Decision Sciences*, *48*(3), 454–488.
- Yang, H. L., & Lin, Q. F. (2018). Opinion mining for multiple types of emotion-embedded products/services through evolutionary strategy. *Expert Systems with Applications*, *99*, 44–55.
- Zheng, L., Wang, H., & Gao, S. (2018). Sentimental feature selection for sentiment analysis of Chinese online reviews. *International Journal of Machine Learning and Cybernetics*, *9*, 75–84.

Zhou, L., Ye, S., Pearce, P., & Wu, M. (2014). Refreshing hotel satisfaction studies by reconfiguring customer review data. *International Journal of Hospitality Management*, 38, 1–10.