Communicating Value in Healthcare Marketing from a Social Media Perspective

Sandy Çağlıyor, Petek Tosun, and Nimet Uray

Abstract Sustainable healthcare policies and a developed healthcare industry are vital to countries' competitiveness and productivity. The ongoing transformations in healthcare services and advances in health technologies and analytics make it clear that there is a pressing need for more collaborative and interdisciplinary efforts in the industry. This study aims to explore the effectiveness of online marketing communication for healthcare services in Turkey with regard to the value-driven marketing approach utilized by leading chain hospitals through an examination of two research questions: (1) *Which messages are emphasized in the social media marketing communications of hospitals?* (2) *Which factors increase engagement with healthcare consumers on social media?* To that end, we compiled the Facebook and Twitter posts of three of the largest hospital chains in Turkey for the last 5 years along with the interaction metrics of the posts, ultimately generating a dataset consisting of 9212 posts in total. Using Latent Dirichlet Allocation, we identified four main topics: Posts on holidays and special days/weeks promoting healthy lifestyles, informative posts about the symptoms and treatments of illnesses, posts containing statistics about diseases, and posts including news about the hospital in question. In the following stage, we carried out predictive analysis using three tree-based machine learning algorithms (decision trees, random forests, and gradient boosting trees) to predict total interaction and relative variable importance. Our model performed at an accuracy rate of 70%. The findings of this study indicate that contextual factors such as the number of followers may have more predictive power than content or interactivity factors. Hospitals use social media to improve their brand reputation and increase public awareness about health and critical diseases. The posts about holidays and special days and using links in the posts resulted in the most interaction. Message source was identified as an important factor, so different social media platforms should be treated as separate mediums in the design of marketing communication strategies and the different dynamics

S. Çağlıyor (✉) · P. Tosun · N. Uray
Kadir Has University, Istanbul, Turkey
e-mail: sendi.cagliyor@khas.edu.tr; petek.tosun@khas.edu.tr; nimet.uray@khas.edu.tr

Y. I. Topcu et al. (eds.), *New Perspectives in Operations Research and Management Science*, International Series in Operations Research & Management Science 326,
https://doi.org/10.1007/978-3-030-91851-4_6

of those platforms should be considered instead of posting the same content on various platforms. As such, this research has valuable implications for marketing managers and administrators working in healthcare in terms of the design of their online marketing communication strategies.

Keywords Healthcare marketing · Marketing communication · Social media · Latent Dirichlet allocation · Random forest · Gradient boosting tree · Decision tree

1 Introduction

Social media marketing is becoming increasingly prevalent in business management, particularly in the healthcare sector. As more and more consumers seek out health-related information and advice on social media (Koumpouros et al., 2015), in the last decade hospitals have increasingly turned to social media as a means of communicating with their patients and consumers (Sharma & Gupta, 2019). Through their social media accounts, healthcare institutions promote their brands, share information about well-being, and highlight the advantages of the treatments they provide (Radu et al., 2017). Compared to other mediums of communication, social media marketing provides a number of additional advantages, including the ability to engage with certain communities (Shawky et al., 2019). Moreover, social media makes it easy for consumers to obtain information about hospitals, doctors, and the experiences of other patients (Hackworth & Kunz, 2011), and when consumers have doubts about the trustworthiness of public news sources for specific topics, they tend to rely more on information that is shared on social media (Jang & Baek, 2019).

Advances in digital communications have created a need for observational research that deepens our understanding of the impacts and place of social media in the healthcare sector (Schillinger et al., 2020). The healthcare sector functions as a value network consisting of hospitals, medical laboratories, pharmaceutical companies, medical equipment vendors, and health insurance companies (Sharma & Gupta, 2019). In such an environment, social media facilitates the building of relationships with patients, the transmission of accurate information, and the acquisition of immediate consumer feedback, especially for hospitals (Hackworth & Kunz, 2011). Increased digitalization has driven people to spend more time on social media, which in turn makes it possible for healthcare brands to create online consumer engagement (Haenlein et al., 2020). Social media communication is an effective way for hospitals to act in a socially responsible manner and guide people towards healthier patterns of behavior.

When it comes to information concerning health issues, people tend to think that messages which are conveyed by institutions and include factual information are more believable than those communicated by their peers (van der Meer & Jin, 2020). One of the core responsibilities of hospitals is to provide high-quality healthcare to patients, and that can be taken up in a broader sense to include the dissemination of health-related information that is both correct and proper for the situation at hand. And in light of current developments, social media communications can also provide psychological support to frontline healthcare workers and the community at large (Cheng et al., 2020).

Within that context, this study focuses on the social media communications of hospitals from the perspective of consumer engagement and examines the value that is provided by marketing communication content. Agarwal et al. (2020) have defined value-centered marketing in the healthcare sector as a set of processes that are employed in the delivery and communication of value associated with well-being and treatment. Hospitals interact with their patients and consumers on their social media accounts by sharing content about a wide variety of issues, including disease avoidance and maintaining health. Marketing communication content generates value by associating it with specific outcomes and messages. In line with that framework, the research questions of this study are as follows: (1) *Which messages are emphasized in the social media marketing communications of healthcare providers and/or hospitals*? (2) *Which factors increase engagement with healthcare consumers*? In order to explore those questions, we examined the social media accounts of the three largest hospital chains in Turkey by way of machine learning algorithms. Through the use of web scrapers, we compiled the Facebook and Twitter posts of those hospitals for the last 5 years, and then we employed latent Dirichlet allocation (LDA) for the purposes of topic detection and tree-based machine learning algorithms to carry out our predictive analyses, which will be further discussed in the methodology section. First, however, we provide a summary of the literature on the issue.

The study will continue with the literature review section that includes a synthesis of social media marketing in the healthcare sector and consumer engagement on social media and a brief overview of the Turkish healthcare sector. After the explanation of the research methodology, the findings will be presented. The chapter will conclude with the discussion of findings and conclusion sections.

2 Literature Review

2.1 Social Media Marketing in the Healthcare Sector

In the last decade, hospitals have been increasingly engaging with consumers via their social media accounts, and as a consequence, social media marketing is on the rise in the healthcare sector (Sharma & Gupta, 2019). Making use of an effective marketing strategy that utilizes social media platforms is essential for hospitals to acquire a competitive edge in the healthcare market (Hackworth & Kunz, 2011). As noted by Popović and Smith (Popović 2010), the one-way pattern of communication that predominated in health-related television shows in the 1960s and 1970s, in which inaccessible doctors discussed illnesses, gave way to the medical talk shows of the 2000s that focused on doctor–patient dialogues. Undoubtedly, the underlying reason for the popularity of such two-way interactions on television was steeped in a need for timely, accurate information about health issues. As early as 2011, the US-based *Mayo Clinic* included contact information, videos of doctors explaining issues to patients, discussions raised through the posing of questions, and invitations to consumers to share their comments on Facebook, and all the while *Scripps Health* was communicating daily via Twitter to inform people about health issues (Hackworth & Kunz, 2011).

Shawky et al. (2019) have pointed out that social media has proven to be effective in reaching audiences and developing communications with consumers, and in that way, it has contributed to the communication objectives of organizations. Healthcare service providers use social media to communicate with patients, promote their brands, and attract human resources (Bejtkovsky, 2020). Furthermore, healthcare institutions use social media for advertising, sharing informative content about well-being, and posting messages about the advantages of certain treatments (Radu et al., 2017). Videos, photos, widgets, blogs, and personalized messages are just some of the elements of the rich marketing communication environment that social media platforms provide to marketers and consumers (Popović & Smith, 2010).

Social media marketing makes it possible for physicians and healthcare institutions to satisfy their patients' needs, as indicated by the fact that more than 75% of consumers seek out health-related information on the web (Koumpouros et al., 2015). It has also been found that 62% and 53% of consumers search the internet for health advice and specific products, respectively (Koumpouros et al., 2015). In the past, marketers merely had to increase message frequency to drive up brand awareness, but in today's connected digital world where social media is increasingly used by patients, communities, and other user groups, it is not enough to simply push messages through targeted channels (Popović & Smith, 2010). Hackworth and Kunz (2011) have suggested that healthcare marketers need to share information about their brands, interact with consumers, monitor patient feedback, and regularly keep up with online conversations about their brands. Moreover, it has been found that satisfaction with health-related information obtained online is positively correlated with patient loyalty, and patient loyalty in turn is positively correlated with purchase intentions (Sharma & Gupta, 2019). Given that situation, social media has become a crucial medium for both retaining existing patients and acquiring new ones as well.

On the other hand, studies have indicated that some doctors and healthcare managers are not comfortable with using social media marketing and still place their trust in conventional marketing channels (Koumpouros et al., 2015). In their examination of the social media marketing activities of healthcare institutions in Poland, Gregor and Gotwald (2013) found that only 18% of healthcare institutions made use of social media, and they traced that low rate of usage led to managerial mistrust of social media marketing (71%), a lack of skilled staff (22%), and managers' assumption that patients would simply not be interested (19%). However, hospitals can provide a major service in terms of improving public health by sharing health-related content and information on their social media accounts, and as such, they not only act as providers of health services but also socially responsible institutions. The corporate social responsibility of hospitals vis-à-vis patients and society has the potential to improve brand advocacy, brand trust, patient-hospital identification, and the positive word-of-mouth intentions of patients (Limbu et al., 2020). Moreover, as Agarwal et al. (2020) have pointed out, patients are increasingly becoming empowered consumers of healthcare services in a value-delivery framework that consists of three dimensions: precision in treatment, the preferences of consumers, and customer-centric processes. In today's highly digitalized world, social media needs to become an embedded part of patient acquisition as well as processes of informing, retaining, and engaging consumers of healthcare.

Consumer and patient engagement on social media make it possible for patients to not only be well-equipped with information but also empowered (Hewitt, 2011). The social media campaigns of hospitals, including those that encourage a healthy diet and exercise as a way to avoid obesity and diabetes, can be taken up as social marketing actions that attempt to influence people's behavior for the benefit of society (Kotler & Zaltman, 1971; Hewitt, 2011). Similarly, Gregor and Gotwald (2013) have noted that healthcare institutions use social media, in particular Facebook, to increase public awareness and prevent diseases. The engagement of hospitals with existing and potential clients on social media has become one element of social marketing that aims to improve public health primarily through the dissemination of information. However, while it is true that the use of social media does offer such benefits, healthcare organizations must also consider legal, medical, and regulatory issues when they utilize social media as a part of integrated marketing communications (Popović & Smith, 2020).

2.2 Consumer Engagement with Social Media in Health Services

Healthcare institutions are increasingly making use of social media to communicate and interact with consumers who have become more conscious about issues such as well-being, critical diseases, and medical diagnostics (Sharma & Gupta, 2019). Since consumers have the option of sharing their opinions in the form of online comments or reviews regarding products and services, social media offers a rich and interactive communication environment. Although websites are the main source of information for health-related topics, consumers also search social media platforms such as YouTube and Facebook to access information about their health-related concerns (Koumpouros et al., 2015; Cangelosi et al., 2019). Studies have shown that Facebook is the most commonly used social media platform for sharing content-based information and raising awareness about social marketing programs that aim to improve public health (Gregor & Gotwald, 2013; Shawky et al., 2019). It has been found that single young adults and people whose employers provide health insurance benefits tend to think that social media is a crucial source of information about healthcare (Cangelosi et al., 2019), and the number of followers that a given page has on social media is often considered to be a primary indicator of how brands interact with consumers (Wang & Jin, 2010).

Consumer comments on social media can be a valuable source of feedback for managers. For example, Chatterjee et al. (2020) have used text-mining to analyze consumer reviews regarding health products and healthcare e-commerce, and they demonstrated that an e-commerce company's service quality and the perceived quality of the physical and online facilities that are integrated into an omnichannel model play a critical role in terms of customer satisfaction in fitness and nutrition services. Moreover, consumers can act as service providers and beneficiaries when they share information with online health communities (Stadtelmann et al., 2019) because their opinions, advice, and comments influence others. A survey conducted about the new patients of a dental clinic revealed that 79% of them were influenced by Internet content in their selection of the clinic, as indicated by the fact that Facebook and Google ads were major sources of information for them (Radu et al., 2022).

Engaging with existing and potential patients on social media platforms contributes to brand equity for healthcare institutions. Open and sincere communication with patients on social media can contribute to the development of brand trust, as perceptions of transparency suggest that the hospital is responsible for its delivery of services (Limbu et al., 2020). Moreover, many consumers find it difficult to ascertain the medical accuracy of the content on websites (Popović & Smith, 2010). When prestigious healthcare institutions share information, answer questions, and engage with people on social media, they are in a better position to build bonds with consumers. In turn, if consumers are satisfied with the richness of information provided by healthcare institutions in online environments, they are more likely to spread positive word-of-mouth (Sharma & Gupta, 2019).

Social media can have both positive and negative impacts on health-related attitudes, behaviors, and norms in society (Schillinger et al., 2020). El-Awaisi et al. (2020) have examined social media posts to explore public perceptions of healthcare sector employees and they pointed out that social media content can generate a wide variety of emotions such as frustration, relief, and gratitude. Patients' levels of involvement in online information about health are significantly correlated with the perceived quality of that content, and people tend to be more selective regarding online content that is posted by healthcare institutions (Sharma & Gupta, 2019). In that context, social media has the potential to influence public health within the scope of a continuum of factors, including disease control, treatment, and the dissemination of disinformation (Schillinger et al., 2020).

2.3 The Healthcare Sector in Turkey

Turkey is a developing country with a population of 81.4 million people and a GDP of $28,423 per capita (OECD, 2021a). For the year 2019, the proportion of health expenditures in relation to GDP was 4.40% and 8.8% in Turkey and the average of OECD countries, respectively (OECD, 2021a). The rates of population growth and health expenditures in Turkey for the 2007–2019 period are shown in Fig. 1.

In parallel with the country's population growth, the number of doctors, nurses, and other healthcare employees has also been increasing in Turkey, the figures for which are shown in Fig. 2. The number of physicians per 100,000 patients has increased from 138 in 2002 to 193 in 2019 (Saglik, 2020). The number of doctors per 1000 people in Turkey was 1.9 in 2017; in comparison, those figures were 0.8 in India, 2.0 in China, 2.4 in Mexico, 3.4 in France, 4.0 in Italy, and 4.3 in Germany (OECD, 2021c).

In 2019, there were 1538 inpatient medical care institutions in Turkey. The distribution of public and private institutions is shown in Fig. 3. The share of private hospitals in the health sector increased from 29% in 2007 to 37% in 2019, and chain hospitals contributed much to that transformation in the healthcare sector.

According to statistics provided by the Turkish Ministry of Health, more than 70% of Turkish citizens have said that they are satisfied with the health services they have received, and the average satisfaction rate was 72.59% for the 2010–2019 period (Saglik, 2020). Although the majority of patients have expressed satisfaction with the health services they have received, the Turkish healthcare sector nonetheless still has great potential to grow and improve.

Within that context, the current study aims to improve our understanding of how hospitals use social media as a marketing communication tool. It utilizes a unique research framework and methodology to reveal the main reasons for social media marketing communication in the Turkish healthcare sector. Moreover, this study will explore online consumer engagement with hospitals and provide

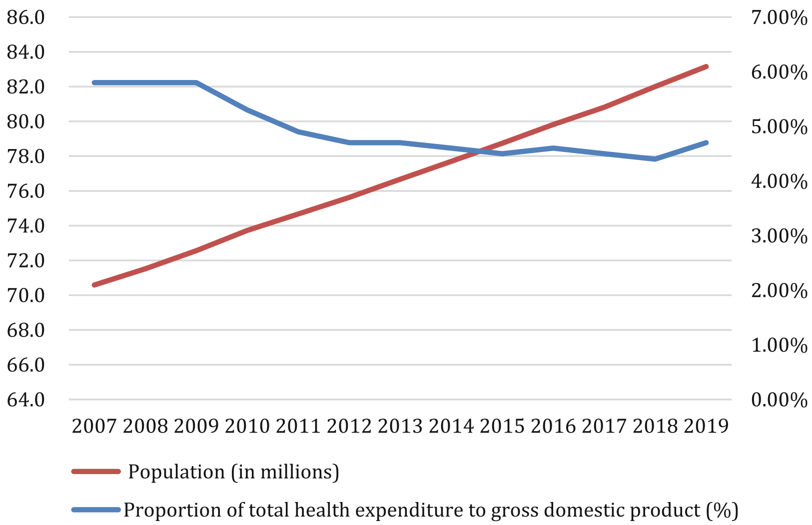

Fig. 1 Population growth and health expenditures in Turkey. (Source: www.saglik.gov.tr)

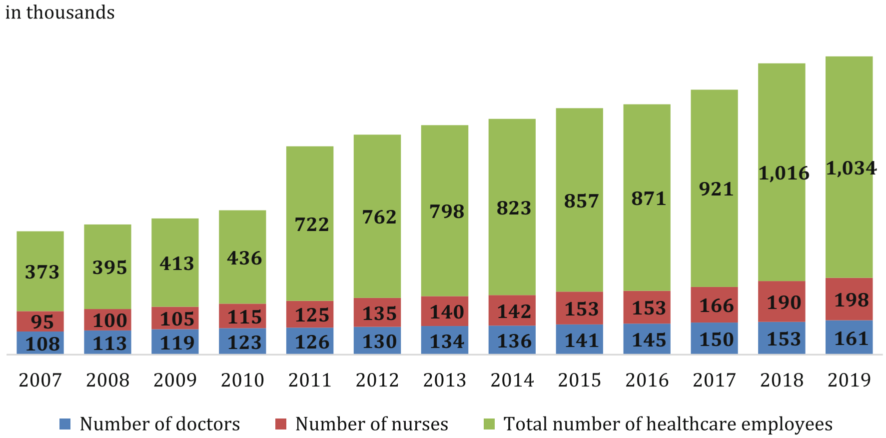

Fig. 2 The number of healthcare sector employees for the years 2007–2019. (Source: www.tuik.gov.tr)

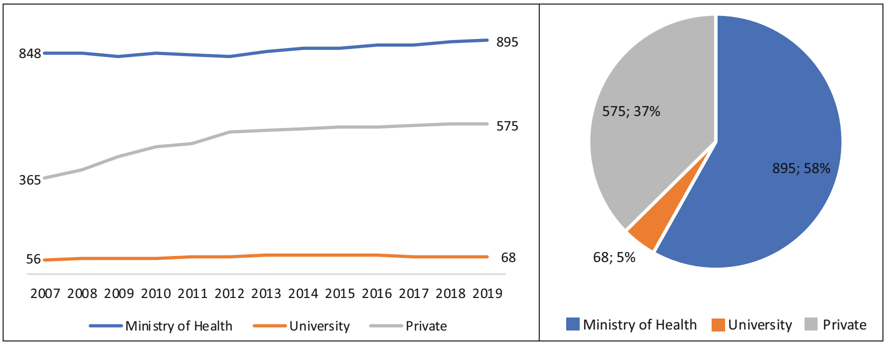

Fig. 3 The total number of inpatient medical institutions in Turkey. (Source: www.tuik.gov.tr)

valuable findings for marketing practitioners. This chapter aims to contribute to the literature by focusing on a relatively less studied topic, social media marketing in the healthcare sector. Furthermore, it bridges a gap between data science and marketing by applying machine learning methodology and interpreting the findings from a social science perspective. The next section will summarize the research methodology.

3 Methodology

This study aimed to explore the effectiveness of online marketing communication for healthcare services in Turkey. In accordance with our aims, this study is based on a descriptive research design. We examined the social media accounts of the three largest hospital chains by way of machine learning algorithms. As illustrated in Fig. 4, the methodology is based on a sequence of successive steps. Although they will be explained in more detail in the following sections, to mention briefly these steps can be summarized as data collection, topic detection, data transformation, and prediction analysis. The first step, in which posts made by Turkey's three largest hospital chains (Acıbadem Healthcare Group, Medical Park Hospital Group, and Medicana Healthcare Group) in the last 5 years were extracted from Facebook and Twitter with the help of two web scrapers along with the interaction metrics of the posts, viz. number of likes, number of shares or retweets, and number of comments, as well as metadata such as video duration, number of video views, and dates.

After cleaning and removing duplicates, our final dataset consisted of 9212 posts, and following preprocessing, topic modeling was applied to the textual data. In the data transformation process, we computed the topics we obtained as new categorical variables, and other metrics such as total interaction, word count, and the presence/lack of a hashtag, link, or question were determined, whereupon the total interaction score was discretized into four classes. By employing three machine learning algorithms—a random forest, gradient boosting tree, and decision tree—we carried out prediction analyses. After the predictive performance of the models' accuracy was evaluated, we examined the importance of the random forest variable with statistical attributes. Topic detection modeling and all predictive analyses were carried out on Knime 4.1.0.

The available contextual factors, content, and interactivity factors were used in the analysis in a manner that is consistent with the literature (Schultz, 2017). In order to represent the contextual factors, we employed a categorical variable indicating the hospitals and a continuous variable indicating the number of followers, since the former provides information about the social media platforms on which posts were made and the latter is considered to be a primary factor that impacts total interaction on brand accounts (Wang & Jin, 2010). In this study, we used the topics generated by LDA as content-related factors and interactivity factors; interactivity factors were coded as three binary variables indicating whether the post contained a hashtag, a link, or a question.

3.1 Data Collection and Formation of the Dataset

Social media and other platforms based on user-generated content have proven to be a great resource for researchers in many disciplines, as there is a plethora of publicly available data about human behavior (Fiesler et al., 2020). Data from social media platforms can be extracted manually (Aydin, 2020) or automatically with a help of a web scraper (Dewi et al., 2019). The process of automatically collecting data from social media and other websites and transforming unstructured data into a structured format is known as web scraping, as well as web harvesting and web data extraction, and the tools used for those purposes are referred to as web crawlers,

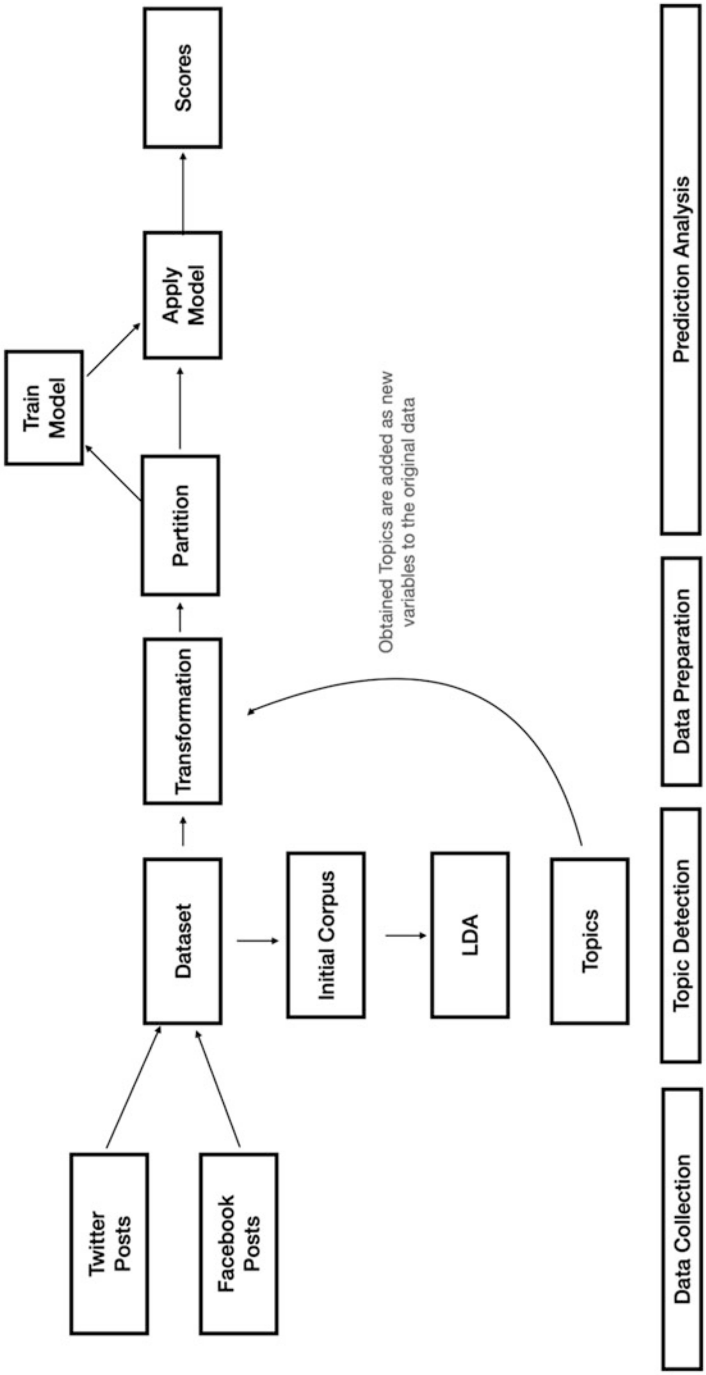

Fig. 4 The workflow

scrapers, robots, bots, and spiders. While those terms are not exactly synonyms, however, they are often used interchangeably since they denote tasks that are quite similar (Algiriyage et al., 2013). While automated scrapers drastically reduce the amount of time needed for data retrieval, they can have certain limitations, as they may require constant adjusting so they can handle the differing formats of websites and they may necessitate a certain amount of technical know-how, especially when websites have complex structures; furthermore, social media platforms may impose restrictions to prevent the collection of data or limit the amount of data that can be gathered by automated means (Batrinca & Treleaven, 2014). Nonetheless, while such limitations prevented us from accessing all the messages posted by hospital chains in Turkey for the time period 2006–2020, we were able to obtain enough data to construct a sufficiently large dataset for our analyses.

After carrying out a preliminary examination of the social media usage of the hospital chains in question, we decided to focus our efforts on postings made via Facebook and Twitter. While some of the hospitals also had Instagram and LinkedIn accounts, there were either fewer messages posted on those accounts or they paralleled postings that were made on Facebook. We used two web scraping tools, Octoparse and Outwit Hub, for the extraction of data from the official Twitter and Facebook pages of the Acıbadem Healthcare Group (@acibademsaglik), Medical Park Hospital Groups (@MedicalParkHG), and Medicana Healthcare Group (@medicanasaglik). Octoparse is a machine learning-based visual tool with a point-click interface that can run extractions on a cloud with multiple servers or a local device (Octoparse, 2021), and Outwit Hub is also a web scraping tool but it works by breaking down the elements on a webpage into their different constituents (Outwithub, 2021). In order to extract the Twitter posts of the accounts, we created yearly filters through the use of advanced search functions for each account and scraped them individually with Octoparse. Because Twitter sets limitations for the collection of past data, we utilized a yearly filter to create a more homogeneous dataset. For each account, we extracted text, timestamps, numbers of likes, numbers of comments, numbers of retweets, and the image links of tweets, and timestamps were converted into dates. For Facebook, once we had obtained the URLs of the posts, their contents were scraped with the use of Outwit Hub. As an extension to the "like" button, Facebook also offers a reaction feature (in the form of a heart, sad face, angry face, etc.), which we also took into account. For those posts we collected text, numbers of reactions, numbers of likes, numbers of shares, dates, and comments. When posts included videos, we were also able to obtain the number of views and duration of the videos, but instead of an exact date or time stamp, the scraper only harvested the year of the post. After the results were compiled and duplicates were removed, the overall data set consisted of 9212 posts, of which only 742 contained visual but not textual content. In total, we were able to obtain 6147 Facebook posts and 3065 Twitter posts for the three hospitals. Tables indicating the distribution and descriptive statistics of the scraped posts can be found below. Table 1 presents a summary of the descriptive statistics for the number of likes, retweets/shares, and comments for our whole dataset, and Table 2 shows the composition of the posts in terms of the social media platform used and the hospital that posted them, along with the number of followers for each one. Lastly, Table 3 illustrates the distribution of the posts with regard to the number of posts by year, which hospital made them, and the social media platform that was used.

3.2 *Topic Detection*

In order to determine what kinds of messages were emphasized in the social media marketing communications of the hospitals in question, we employed LDA Topic Detection modeling for the text corpora we obtained—in other words, the textual content of the posts. LDA is a tri-level hierarchical Bayesian model that is frequently

Table 1 Summary statistics for the social media metrics in this study

	Min.	Max.	Mean	Median	Std. Deviation
Likes	0	18, 823	174.32	0	16.51
Shares/retweets	0	1400	16.27	2	58.49
Comments	0	746	2.26	21	574.50

Table 2 Number of followers and number of posts

	# of Followers: Twitter	# of Followers: Facebook	# of Posts: Twitter	# of Posts: Facebook
Acıbadem	28, 263	908,489	1544	2370
Medical Park	14, 483	219,548	854	1965
Medicana	8884	220,635	676	1805

Table 3 Distribution of posts (by hospital per year)

	Hospital	Facebook	Twitter	Total	
2016	Acıbadem	504	407	911	2212
	Medical Park	665	292	957	
	Medicana	220	124	344	
2017	Acıbadem	567	435	1002	2051
	Medical Park	559	187	746	
	Medicana	189	114	303	
2018	Acıbadem	536	344	880	2002
	Medical Park	174	134	308	
	Medicana	646	168	814	
2019	Acıbadem	305	122	427	1353
	Medical Park	236	126	362	
	Medicana	424	140	564	
2020	Acıbadem	473	236	708	1594
	Medical Park	329	106	435	
	Medicana	321	130	451	
Total		6148	3065		9212

used in topic modeling because of its ability to identify and describe latent thematic structures in compiled text documents (Blei et al., 2003). The input for LDA is a textual corpus in which each document is considered as a collection of topics and each word in the document corresponds to one of those topics (Lash & Zhao, 2016). LDA clusters each topic with a set of words that best describes the group in which each item is modeled as a finite mixture over an underlying set of topics.

The initial step in topic modeling is data cleaning and preprocessing. In our study, we started that process by removing links from the corpus and then eliminating terms that did not contain content. We followed the usual preprocessing steps such as stop word filtering, punctuation erasing, case conversion, number filtering, RegEX filtering, stemming, and lemmatization through the use of Knime's related nodes. The ZemberekNLP (Natural Language Processing) Library was used for stemming and POS tagging. The data cleaning process can be difficult for user-generated comments because of typos, spelling mistakes, and the usage of "social media jargon," but in our case the cleaning and preprocessing was fairly straightforward since the corpus was obtained from the posts of official accounts, which tend to be more standard in terms of grammar and spelling. All the same, it wasn't enough to just preprocess the nodes, as some manual interventions were also needed, which we carried out with a dictionary filter node. We also found 147 reply posts in our dataset, which we excluded.

In the LDA model, the number of topics (K) and the values for parameters such as document-topic density (α) and topic word density (β) are not defined beforehand. Therefore, as suggested in the literature, we tested for different K values and different combinations of α and β so that we could choose the best model in terms of the LDA evaluation, and the outcomes were evaluated on the basis of their significant differences and interpretability (Maier et al., 2018). Among all the combinations, the most representative model was obtained when $K = 4$, $\alpha = 0.03$, and $\beta = 0.01$.

3.3 Prediction Analysis

Due to their ability to process high-dimensional and complex data and reveal hidden patterns (Wuest et al., 2016), machine learning algorithms are increasingly being used in prediction models. In studies that utilize machine learning algorithms, several different algorithms are usually used instead of a single algorithm and the best-performing model is chosen. Since the performance of models may vary as a result of the structures of the problem, variables, and the dataset, there is no across the board formula for determining which model should be selected (Delen & Sharda, 2009; Hur et al., 2016). Tree-based models have proven to be a potent solution for machine learning problems in various fields (Asadi et al., 2014). In this study, three tree-based machine learning algorithms—a decision tree (DT), random forest (RF), and gradient boosting tree (GBT)—were employed for the predictive analysis. The partitioning ratio of the training set to test ratio was fixed at 70:30 for all the algorithms and stratified sampling was adopted. In that way, a stratification variable target value was selected.

The first algorithm used in the model was a DT, which is a supervised machine learning algorithm that consists of a root node, branches, and leaf nodes, and while each node represents an attribute, each branch represents a decision (Naik & Samant, 2016). The ability of such algorithms to handle both continuous and categorical data, as well as their insensitivity to outliers and missing values (Breiman Breiman et al., 1984) and their utility in providing insights about variable importance, make DTs ideal for such cases (Tso & Yau, 2007). Despite some disadvantages such as overfitting problem, repetition problem, or fragmentation problem, decision tree is still a highly preferred algorithm due to its speed and simplicity (Patel & Rana, 2014). There are different quality measures for decision trees. In this study, we chose the Gini Index, also known as Gini Impurity, as the quality measure (Kingsford & Salzberg, 2008). The minimum number of records per node was set at two and a reduced error pruning option was selected.

The second algorithm we used was a random forest, which is an ensemble model consisting of simple decision trees that can generate a response to predictor values. Random forest algorithms primarily use bagging, by means of which each model runs individually, and in our study the result was calculated by the mean of each individual output. Unlike decision trees, random forests do not learn from a single tree but from a number of different simple trees, and that reduces the risk of learning from noise (Breiman, 2001). In this way, it allows to obtain more accurate results especially for noisy data, and avoids the problem of overfitting, but it may be slower and more difficult to interpret because of higher number of trees (Reis et al., 2018). In this study, information gain ratio was chosen as a split criterion for the tree option, and we did not interfere with the tree depth and minimum node size. As for the forest option, 100 models were selected.

The third algorithm, a gradient boosting tree, is also an ensemble model, but unlike bagging it uses boosting to achieve a strong learner (Zhang & Haghani, 2015). Like bagging algorithms, boosting trees also consist of simple trees, but unlike bagging, in which each model runs individually, boosting trees compute a sequence of trees, and each tree learns from the preceding one. Like random forest, GBDT can better prevent the possibility of overfitting, and also has superior robustness compared to decision tree because its performance is less likely to be affected by outliers and irrelevant features (Cui et al., 2018). In terms of tree options, the tree depth was set to 4, and in terms of boosting options, the number of the model was set to 100 and the learning rate was set to 0.1.

Several metrics can be used to measure the predictive performance of a machine learning algorithm. In this study, the average percentage hit rate (APHR), which is the ratio of true positives to the total number of samples, was used to evaluate performance.

$$\mathrm{APHR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{1}$$

Although machine learning algorithms often have higher predictive power than linear models, they may fall short with regard to explaining the effect of independent variables on dependent variables. Nevertheless, even if they do not fully describe the impact of independent variables on dependent variables where the direction is unclear, the contribution of an individual variable to the predictive power of the model—in other words, the relative variable importance—can be calculated by way of additional scruties, such as sensitivity analysis (Sharda & Delen, 2006). On the other hand, unlike other popular machine learning algorithms like support vector machines and neural networks, tree-based methods can establish a relationship between dependent and independent variables without requiring additional analysis (Zhang & Haghani, 2015). In Knime, a Random Forest Learner node includes an output showing how often a single feature is used at different levels of the tree. Frequent selection as the best split indicates that the variable has a strong informative effect, and hence the importance of the variables in the model can be compared relatively in terms of the contributions they make to the predictive power of the model by dividing the split number by the number of candidates at its level and adding the values.

$$\text{Variable importance score} = \Sigma \frac{\text{Splits}_{\text{Level}_\text{i}}}{\text{Candidates}_{\text{Level}_\text{i}}} \tag{2}$$

3.4 Variables

Online consumer engagement and interaction are considered important metrics as they not only indicate the response of a given customer to a specific issue, but also provide clues about how much the consumer wants to interact with the brand as well as the customer's feelings about and attitude towards the brand (Demmers et al., 2020). There is an extensive body of literature dedicated to improving our understandings of the antecedents and consequences of customer engagement on social media networks. In order to capture customer engagement or interaction on social media platforms, most of these studies use behavioral metrics such as the number of likes, comments, and shares generated in response to posts made by brands. Those metrics can be evaluated separately (Fiok et al., 2020) or combined into a single metric (Liu et al., 2019). Alternatively, they can also be transformed into an engagement rate by dividing the total number of interactions by the number of followers and generating a percentage.

As regards target value, we tested both customer engagement and total interactions. As is frequently done in machine learning studies, we approached the subject from a classification point of view in which the target value is discretized into classes. Since a wide range and the distribution of values complicates discretization, we first needed to compute the normalization. Even though we carried out several trials of different normalization and binning techniques, we were unable to effectively discretize customer engagement. We tried using classes with different ranges, but in every attempt, all the algorithms yielded low-quality results, so we decided to go with total interaction as the target value. By computing the log of values, a wide range can be compressed to a narrow range. After several trials with different binning techniques and class numbers, we found that 3 and 4 were the most favorable number of classes. Posts that received a total number of likes, shares, or comments in the range of 0–10 were designated as class C, those with 11–100 were designated as class B, and those that received total interaction higher than 101 were designated as class A for the analysis with 3 classes. For the analysis with 4 classes, the range was set as follows: 0–10, 11–100, 101–1000, >1000. Since machine learning algorithms can handle outliers, we did not intervene in them. Furthermore, at this point, we should clarify that the correlation between the number of likes and the total interaction score does not affect the performance of the model, in which there are no assumptions associated with them, unlike in the case of linear regression models. However, it is to be expected that the number of followers will be the best split in most of the sub-sampling.

As stated earlier, the available contextual, content, and interactivity factors constituted the dependent variables in the prediction analysis. The number of followers, the hospital's name, and the social media platforms utilized were taken up as contextual factors, and the topics generated by the LDA topic modeling were used as content-related variables. Details about the results yielded by the LDA will be presented in full in the findings section, but it will be useful to discuss them briefly here to give an idea about the classes of the topics. We identified four topics in total: greetings on special days, healthy ways of living, information about serious diseases, and messages promoting videos or events in which doctors provide information about particular medical conditions. For posts that did not contain textual data, we appended a dummy variable. The majority of the non-textual posts consisted of photographs of symposiums or events organized by hospitals, and the remainder were health-related posts in which text was superimposed over the photo, and captions were not used.

The impacts of interactivity factors such as links, questions, polls, hashtags about customer engagement, and total interaction have been investigated at length. Although there are variations in terms of the metrics used, it has been found that platforms, the industry in which the brand operates, and the use of interactive elements enhance customer interaction and engagement (Schultz, 2017). For example, Kujur and Singh (2016) have demonstrated that while including links in posts has a negative impact on likes, it has a positive effect on increasing share numbers. In this study, we coded interactivity factors as three binary variables indicating whether posts contain a hashtag, a link, or a question.

In terms of variables, this study has two drawbacks. First of all, we could only identify the year that videos were posted but not the date or time, so we were unable to use variables related to time or post frequency. For dates and times, Facebook uses the elements: "<abbr title="" data-time="" data-shorten="" class="" >. We modified the Xpath element of the scraper, but it still did not work. Another type of data that could not be extracted was related to video content on Twitter. Although the scraper successfully detected photo content, it could not detect videos on Twitter,

Table 4 Key variables of the study

Variable	Type	Coding
Total interaction	Categorical	4 Classes
Social media platform	Categorical	2 Classes
Content	Categorical	4 Classes and a dummy
Number of followers	Continuous	Number
Hashtag	1/0 Binary	1/0 Binary
Question	1/0 Binary	1/0 Binary
Link	1/0 Binary	1/0 Binary
Video duration	Categorical	3 Classes (<4 min.; ≤4–8 min.; ≥8 min.
Number of views	Continuous	Number

Table 5 Topics and top words

Topic_0	Topic_1	Topic_2	Topic_3
Holiday	Water	Health	Doctor
Day	Health	Cancer	Professor
Week	Drink	Treatment	Specialist
Happy	Eat	Control	Question
World	Caution	Symptom	Information
Health	Protect	Heart	Broadcast

and as a consequence vividness factors could not be included in the analysis. We did not, however, experience such difficulties with Facebook video data, and since we had enough Facebook posts (6147), we were able to do a second analysis by filtering the Twitter data and adding the vividness variables (Table 4).

4 Findings

4.1 LDA Topic Detection

After we evaluated various models with a different number of topics (K), document-topic density (α) and topic-word density (β) values four topics emerged as a result of the LDA topic modeling. This section presents our findings as well as the average and maximum total engagement ratios per topic and some notable examples, and the next section offers an interpretation of those findings. All the posts that are presented as examples in this section (with the exception of one) were originally in Turkish and have hence been translated. Table 5 shows the top representative words with the highest term weights for each topic.

All of the terms in this table were translated from Turkish by the authors.

The first topic, Topic_0, mainly includes posts that were observances of national or religious holidays, commemorations of special global days and days/weeks dedicated to the awareness of certain causes, or hospital-related issues such as obituaries and messages posted following natural disasters. 19% of the total posts (1745) fell within this category. The engagement average for this group was found to be 0.13%, and the average interaction rate was 280. The post with the highest engagement rate had a score of 8.91% and the highest total interaction scored 19,555 (that was the highest score not only in this group but also in the data set as a whole). It was dated October 27, 2018 and posted on Facebook by Medical Park. In terms of visual content, this post included a photo of Atatürk in front of the Turkish National Flag.

> Happy 95th anniversary of our Republic! #29October #Republic Day

Medical Park also posted a message that had one of the lowest scores in this group. The message was posted on Twitter on December 30, 2019:

> Do you remember phonebooks? On special days you would go through the pages of those phonebooks and call people one by one. Call your loved ones, share your good wishes and increase your happiness in the new year. ▯ #newyear #health #healthforeveryone #medicalpark

The other posts that had high engagement scores included commemorations of the 10th of November (the anniversary of the death of Atatürk, who was the founder of the Republic of Turkey), the 18th of March (marking the victory of Turkish forces in the Battle of Gallipoli, also commemorated as Martyrs' Day), and the 25th anniversary of the founding of Acıbadem Hospital.

The second topic, Topic_1, included posts aimed at inspiring people to adopt healthier lifestyles, dealing with issues like quitting smoking, eating healthier, sleeping better, and exercising, as well as messages informing people about mild medical conditions like insulin resistance, gastroesophageal reflux disease (GERD), cramps, fungal infections, and allergies. The percentage ratio of the messages in this group was 27% with a total of 2533 messages, and the average engagement rate was 0.10% with an average total interaction of 245. A message posted on Facebook by Acıbadem on July 2, 2020 received the highest interaction rate in this group with a score of 9115.

> Cinnamon, cloves, basil, cumin, cardamom, mint, coriander, rosemary, garlic, ginger, and turmeric are all health-friendly spices we often use in our kitchens. They help prevent inflammation and also have detox properties.

The message that scored the highest engagement rate was a tweet posted by Medical Park on the 1st of February in 2020. The engagement rate for the post was 3.91%.

> If you want to stay healthy during the winter months and prevent weight gain:
>
> Eat more fruits and vegetables!
> Make sure you get enough protein.
> Drink plenty of fluids.
> Increase your physical activity!
> Eat less often.
> #healthforeveryone #medicalpark

Examinations of other messages in this group that had high engagement rates revealed that they were about healthy lifestyles and included recommendations about improving nutrition, recipes for healthy meals, and nutrition facts. Below are some examples of such posts that did not receive responses:

> Remember to take precautions against the harmful effects of the sun when the temperature is higher than the seasonal norm. #healthyweekend (Twitter, Medical Park, June 25, 2016)

> What should you do to get relief from cramps? (Twitter, Medicana, June 28, 2016)

The posts in the third group, Topic_2, were mostly messages aimed at raising awareness, and they provided information about the symptoms and treatment of serious diseases as well as statistical facts. The illnesses they covered include cancer, Alzheimer's, COPD, osteolysis, bipolar disorder, autism, obesity, and cardiovascular diseases. Containing a total of 3031 posts, this group accounted for 33% of the overall data with an average engagement rate of 0.08% and an average total interaction score of 178. The post with the highest interaction score (8695) was made by Acıbadem on Facebook on June 1, 2018:

> Scoliosis can be treated through early detection and in the course of treatment patients will retain their mobility. For detailed information, visit the website omurgasagligimerkezi.com.

In this group, the highest engagement score was 2.33% for a post made on Twitter by Medical Park on January 20, 2018:

> Make sure you keep up with your regular gynecological check-ups so your doctor can make an early diagnosis of cervical cancer. Don't be late for life—catch cancer early!
> #cervicalcancer #cancer #cervix #gynecology #women'shealth #health #smeartest #healthcare #medicalparkTwitter

Thirty-eight messages in this group did not receive any responses. Two such postings are as follows:

> Growth to abnormal heights, infertility, reduced visual acuity, weakness... These are all symptoms that may indicate the presence of pituitary gland tumors. (Facebook, Acıbadem, February 20, 2017)

> Narrowing of the coronary artery can occur suddenly, and if the flow of blood is not supplemented from other arteries, it can lead to a heart attack. When this occurs together with a disturbance in the rhythm of the heartbeat, the heart may not be able to supply enough blood or it may even stop completely, leading to death if an intervention is not carried out. (Facebook, Medicana, September 22, 2020)

The last topic includes hospital-related posts. The majority of the posts in this group consisted of messages promoting videos, events, or live broadcasts in which doctors discuss specific subjects. Additionally, there were posts providing information about the opening of new hospitals, health centers, and clinics, as well as other news about the hospitals. A total of 159 of the messages in this group received responses. When we calculated the average for total interactions, we excluded posts that did not get responses and we did not carry out predictive analyses of those posts.

In this group, the post with the highest total interaction score (9600) and highest customer engagement ratio (4.37%) was a video posted by Medical Park on Facebook in 2017. It was about someone who survived cancer as a child and

Table 6 Summary of the descriptive statistics of the topic groups

Topics	# of Posts	%	Mean Engagement (%)	Mean Interaction	Max. Engagement (%)	Max. Interaction	# without responses
0	1745	21	0.13	280	8.91	19, 555	17
1	2533	30	0.10	245	3.91	9155	14
2	3031	36	0.08	178	2.33	8695	35
3	1014	12	0.06	134	4.37	9600	58

Table 7 Accuracy of prediction analysis: 3 classes

Algorithm	Accuracy
DT	70.136%
RF	70.84%
GBT	70.732%

later in life volunteered to organize musical events for the children being treated at the pediatric oncology clinic of a Medical Park hospital. The caption of the video read as follows:

> Together it is possible. We believe in the importance of psychological support in the struggle against cancer. In the fight against despair and giving up, we say #TogetherItIsPossible.

Two other messages that also had high interaction scores are presented below. The second post was made in English.

> Medical Park and Liv Hospital are going public! (Facebook, Medical Park, 2018. Total Interaction Score: 7,900)

> At 14:00 on Natural TV, Dr. Ulas Sozener will talk about Kidney Transplantation. (Facebook, Medicana, 2018. Total Interaction Score: 8,900)

Below are two posts from this group that did not receive responses:

> The 19th episode of the TV series "Eşkıya Dünyaya Hükümdar Olmaz" was shot at our hospital. #eskiyadunyayahukumdarolmaz (Twitter, Medical Park, February 4, 2016)

> Prof. Dr. Vildan Çerçi explained the McKenzie technique, which provides treatment for waist and neck hernias without surgery. (Twitter, Medicana, September 8, 2016)

A summary of the descriptive statistics of the topic groups is provided in Table 6.

4.2 Prediction Analysis

In the first experiment, after excluding the reply posts we tested both the Twitter and Facebook posts along with independent variables such as the social media platform used, the number of followers, the hospital, questions, hashtags, links, and topics. As can be seen in Tables 7 and 8, for both experiments where the target value was discretized into three and four classes, all the algorithms yielded very close results. For three classes we obtained an approximate performance accuracy of 70% and for four classes it was 65%.

Table 8 Accuracy of prediction analysis: 4 classes

Algorithm	Accuracy
DT	65.378%
RF	65.016%
GBT	65.197%

Table 9 Accuracy of prediction analysis: Facebook

Algorithm	Accuracy
DT	69.075%
RF	68.588%
GBT	69.643%

Table 10 Average attribute statistics for 4 classes

Row ID	#splits (level 0)	#splits (level 1)	#splits (level 2)	#candidates (level 0)	#candidates (level 1)	#candidates (level 2)
Number of followers	19	39	65	22	57	108
Platform	28	21	17	28	52	111
No text	17	32	25	23	59	118
Hospital	7	26	59	29	43	110
Links	16	21	17	34	64	110
Topic 3	7	11	31	25	51	103
Topic 1	4	17	32	39	57	106

To make it possible to observe the impacts of vividness, we conducted another analysis of the Facebook posts for which we included variables such as the number of views and whether the post included a video or photo, and if it did include a video, the number of people who viewed it and the duration of the video (Table 9).

In order to observe the relative importance of the variables, namely the relative contributions to the models' predictive power, we used statistic attributes, dividing the number of splits to their candidates for each level and adding them together. As a means of avoiding the randomness factor, we repeated three individual analyses and took the average values. When the Facebook and Twitter posts were analyzed together, we found that the number of followers, the social media platform used, non-textual posts, and the hospital in question made the largest contributions to the predictive power of the models, while links, Topic 0, and Topic 1 did so to a lesser extent. The variables Topic 2, Topic 3, questions, and hashtags were found to be the variables that contributed the least. For Facebook, while the number of followers, views and lack of text made the most contributions, the hospital, links, video duration, and Topic 3 contributed moderately to the model. We found that questions, hashtags, Topic 0, Topic 1, and Topic 2 were the variables that contributed the least (Tables 10, 11 and 12).

Table 11 Average attribute statistics for 3 classes

Row ID	#splits (level 0)	#splits (level 1)	#splits (level 2)	#candidates (level 0)	#candidates (level 1)	#candidates (level 2)
Number of followers	19	41	66	22	57	107
Platform	28	19	18	28	52	109
No text	17	32	28	23	59	117
Hospital	9	28	60	29	43	108
Links	16	15	12	34	64	108
Topic 3	6	12	27	25	51	103
Topic 1	4	16	33	39	57	106

Table 12 Average attribute statistics for facebook

Row ID	#splits (level 0)	#splits (level 1)	#splits (level 2)	#candidates (level 0)	#candidates (level 1)	#candidates (level 2)
Number of followers	20	38	47	20	50	76
Views	10	20	35	17	38	85
No text	19	25	10	22	51	67
Hospital	5	20	36	17	40	73
Links	16	20	22	27	48	80
Video duration	8	13	30	26	35	84
Topic 3	3	18	26	14	48	108

5 Discussion

In order to improve our understanding of how hospitals use social media as a marketing communication tool, this study first sought to shed light on what kinds of posts they make on those platforms. For that reason, we conducted LDA topic modeling analysis, which is frequently used in text classification and is known for its ability to generate satisfactory outcomes. Our topic modeling analysis revealed that hospitals use social media for four main reasons: to send greetings for holidays and other special days, to promote healthy lifestyles, to disseminate information related to diseases, and to make news announcements about the hospitals themselves. In terms of post frequency, it was found that hospitals mostly use social media to circulate information about illnesses and promote healthy habits, followed by postings that include observations of holidays and special days/weeks and hospital-related news such as announcements about upcoming talks, conferences, and seminars. Although our study is unique in terms of its research framework and methodology, these findings are nonetheless consistent with other studies that have examined the content of hospitals' social media messages. For example, research by İlgün and Uğurluoğlu (2019) indicated that hospitals primarily use social media to create public awareness about health issues, commemorate special days/weeks, and make announcements about upcoming training sessions, congresses, conferences, and seminars, and also to promote hospital services and their departments (İlgün & Uğurluoğlu, 2019). In another study, Huang and Dunbar (2013) categorized the social media messages posted by hospitals as "hospital news," "event announcements," "patient stories," "holiday salutations," "public service announcements," and "polls" (Huang & Dunbar, 2013).

In terms of total interaction and customer engagement, we found that greetings for holidays and special days/weeks resulted in the most interaction, followed by postings that target the promotion of healthy lifestyles, the dissemination of information about diseases, and hospital-related announcements. However, with regard to the contributions they made to the predictive power of the model, they were ranked as hospital-related announcements, the promotion of healthy habits, and observations of holidays and special days/weeks. The fact that their relative importance is low does not mean that those variables are insignificant; rather, when predicting interaction levels, contextual factors such as the number of followers were found to matter more than interactivity or content-related factors. In short, a larger number of followers means more visibility. Since sources were identified as an important factor, Twitter and Facebook should be treated as separate mediums in the design of marketing communication strategies, and the different dynamics of those platforms should be carefully considered instead of simply allowing the same material to be posted on various platforms.

The impacts of informative content have been studied for a wide variety of industries, and the outcomes have been found to vary (De Vries et al., 2012). In this study, contextual factors were found to be more critical than informative factors in terms of interaction, but it should be noted here that informative content in the healthcare industry differs from that of other industries. In the field of healthcare, informative content can refer to information about illnesses or medical conditions. Even if a follower of a page or someone else who reads the post is interested in the content, he or she may not want to "like" or respond to it because they might be concerned that their friends will see their reaction. The fact that a lack of text in posts was found to be significant may be indicative of two things. Non-textual postings are usually photos of events that are held at or organized by hospitals, but they also include some health-related posts in which text is superimposed over the photo without a caption,

As regards interactivity factors such as hashtags, questions, and links, we observed that hashtags and questions make less of a contribution than links or other factors. Scholars who have investigated the impact of links have arrived at differing conclusions. For example, Sabate et al. (2014) found no significant relationship between the number of likes and links, whereas Kujur and Singh (2016) found a negative relationship and speculated that links may have a negative impact on total interaction because they direct users to another website. Keeping both views in mind, we checked some of the posts that contained YouTube links to see if one or the other also held true for the healthcare industry. For instance, the Tweets mentioned in the results section that had an interaction score of 0 were viewed by 6721 people and received 42 likes on YouTube. All the same, further research is needed to arrive at concrete conclusions regarding this issue. Lastly, among all the variables, questions were found to have the least impact. However, it would be misleading to compare that finding with the results of studies examining other industries, as the vast majority of the posts that contained a question included answers as well, and as such the questions seemed to be rhetorical in nature. It can thus be concluded that in the healthcare sector, adding questions to posts may not increase interactivity, but it does increase noticeability.

As a final point, we obtained a performance accuracy of around 70% for all the predictive models in the study. Those results could be improved, however, by means of hyperparameter optimization or the application of other models such as neural networks, support vector machines, or fusion models, in addition to the inclusion of other variables.

6 Conclusion

In this study, we attempted to fill in several gaps in the literature. First of all, few studies have sought to determine how social media is used in the health sector, especially by hospitals, and what kinds of outcomes are produced. Through the development of a comprehensive model, we sought to build upon the existing knowledge on this subject. Secondly, there are two poles in social media studies in the literature. While the first group approaches the subject from a data science perspective and focuses on developing larger, more advanced datasets but ultimately fails to interpret the results from a marketing or social sciences point of view, studies taking up the subject from a social sciences standpoint try to explain human behavior but work with a limited amount of data and models that do not represent the complex nature of social media. Given that situation, we set ourselves the goal of creating a bridge between those two approaches, and through the use of NLP techniques and machine learning algorithms, we also interpreted the messages of posts from a marketing communication perspective.

Nonetheless, this study has some limitations that should be mentioned. First of all, the fact that we were unable to take into account the days and times of posts may have affected our results; as the literature suggests, they can have a significant impact on total interaction or customer engagement. Moreover, in this study, textual data consist of captions where the texts on the photographs were not included in the analysis. An evaluation of the pictures and videos that were included in posts not just in terms of their mere presence but also their visuality and content could also improve the performance of the models. Another noteworthy limitation is the fact that we only looked at two social media platforms, Facebook and Twitter, whereas YouTube and Instagram are also extensively used by hospitals as marketing communication tools. Further research that includes metrics from YouTube and/or Instagram could create a more inclusive understanding of how hospitals use social media and which factors affect interaction levels.

Moreover, future studies could include more hospitals to enlarge the sample size and examine interaction metrics separately with regard to comments and shares. Also, the hospitals in this study have international social media accounts, too. Comparing the hospitals' local and international accounts from a global marketing perspective could potentially contribute to the literature on medical tourism, which is a major sector in Turkey and as such could also tell an important story.

Last but not the least, the dynamics of the health sector in Turkey differ from those of many other service sectors because of the nature of the services they provide. In that regard, the meanings and impacts of interactions might differ in other industries. For example, a recent study has demonstrated that having a high number of likes or followers is not necessarily a good thing and that low customer engagement can negatively impact perceived account credibility (De Vries, 2019). As a consequence of those findings, we are left with a pressing question: Does that also hold true for the health sector, where trust is a key building block? If so, how would that affect the decision-making processes of consumers and their sense of loyalty?

Social media has become indispensable for almost every industry as a means of interacting with consumers. However, there is much to learn about how it can be used effectively. Companies need to think carefully when determining their social media strategies because they need to be compatible with their target audiences as well as their marketing and communication tactics and objectives.

References

Agarwal, R., Dugas, M., Gao, G., & Kannan, P. K. (2020). Emerging technologies and analytics for a new era of value-centered marketing in healthcare. *Journal of the Academy of Marketing Science, 48*, 9–23. https://doi.org/10.1007/s11747-019-00692-4

Algiriyage, N., Jayasena, S., Dias, G., Perera, A., & Dayananda, K. (2013). Identification and characterization of crawlers through analysis of web logs. In *2013 IEEE 8th international conference on industrial and information systems, ICIIS 2013 - conference proceedings*, 150–155. https://doi.org/10.1109/ICIInfS.2013.6731972

Asadi, N., Lin, J., & De Vries, A. P. (2014). Runtime optimizations for tree-based machine learning models. *IEEE Transactions on Knowledge and Data Engineering, 26*(9), 2281–2292. https://doi.org/10.1109/TKDE.2013.73

Aydin, G. (2020). Social media engagement and organic post effectiveness: A roadmap for increasing the effectiveness of social media use in hospitality industry. *Journal of Hospitality Marketing and Management, 29*(1), 1–21. https://doi.org/10.1080/19368623.2019.1588824

Batrinca, B., & Treleaven, P. C. (2014). Social media analytics: A survey of techniques, tools and platforms. *AI Soc, 30*(1), 89–116. https://doi.org/10.1007/s00146-014-0549-4

Bejtkovsky, J. (2020). Social media platforms as HR marketing tool in selected healthcare service providers. *Marketing and Management of Innovations, 1*, 294–302. https://doi.org/10.21272/mmi.2020.1-25

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation David. *Journal of Machine Learning Research, 3*(3), 993–1022. https://doi.org/10.1016/B978-0-12-411519-4.00006-9

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classificaiton and regression trees*. Chapman & Hall/CRC.

Cangelosi, J., Kim, D., Griffin, K., & Ranelli, E. (2019). The role of social media and social networking as marketing delivery systems for preventive health care information. *Atlantic Marketing Journal, 8*(1), Article 1. https://digitalcommons.kennesaw.edu/amj/vol8/iss1/1

Chatterjee, S., Goyal, D., Prakash, A., & Sharma, J. (2020). Exploring healthcare/health-product eCommerce satisfaction: A text mining and machine learning application. *Journal of Business Research, 131*, 815–825. https://doi.org/10.1016/j.jbusres.2020.10.043

Cheng, P., Xia, G., Pang, P., Wu, B., Jiang, W., Li, Y. T., Wang, M., Ling, Q., Chang, X., Wang, J., Dai, X., Lin, X., & Bi, X. (2020). COVID-19 epidemic peer support and crisis intervention via social media. *Community Mental Health Journal, 56*, 786–792. https://doi.org/10.1007/s10597-020-00624-5

Cui, H., Huang, D., Fang, Y., Liu, L., & Huang, C. (2018). Webshell detection based on random forest-gradient boosting decision tree algorithm. In *Proceedings – 2018 IEEE 3rd international conference on data science in cyberspace, DSC 2018*, 153–160. https://doi.org/10.1109/DSC.2018.00030

De Vries, E. L. (2019). When more likes is not better: the consequences of high and low likes-to-followers ratios for perceived account credibility and social media marketing effectiveness. *Marketing Letters, 30*(3), 275–291.

De Vries, L., Gensler, S., & Leeflang, P. S. H. (2012). Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *Journal of Interactive Marketing, 26*(2), 83–91. https://doi.org/10.1016/j.intmar.2012.01.003

Delen, D., & Sharda, R. (2009). Predicting the financial success of hollywood. *Endüstri Mühendisliði Dergisi, 21*(1), 30–37.

Demmers, J., Weltevreden, J. W. J., & van Dolen, W. M. (2020). Consumer engagement with brand posts on social media in consecutive stages of the customer journey. *International Journal of Electronic Commerce, 24*(1), 53–77. https://doi.org/10.1080/10864415.2019.1683701

Dewi, L. C., Meiliana, & Chandra, A. (2019). Social media web scraping using social media developers API and regex. *Proc Comput Sci, 157*, 444–449. https://doi.org/10.1016/j.procs.2019.08.237

El-Awaisi, A., O'Carroll, V., Koraysh, S., Koummich, S., & Huber, M. (2020). Perceptions of who is in the healthcare team? A content analysis of social media posts during COVID-19 pandemic. *Journal of Interprofessional Care, 34*(5), 622–632. https://doi.org/10.1080/13561820.2020.1819779

Fiesler, C., Beard, N., & Keegan, B. C. (2020). No robots, spiders, or scrapers: legal and ethical regulation of data collection methods in social media terms of service. In *Proceedings of the international AAAI conference on web and social media*, *2020* (ICWSM), 187–196. Retrieved from www.aaai.org

Fiok, K., Karwowski, W., Gutierrez, E., & Ahram, T. (2020). Predicting the volume of response to tweets posted by a single twitter account. *Symmetry, 12*(6), 1–15. https://doi.org/10.3390/sym12061054

Gregor, B., & Gotwald, B. (2013). Potential of social media and mobile marketing in marketing communication management in health care institutions. *International Journal of Contemporary Management, 12*(3), 38–49.

Hackworth, B. A., & Kunz, M. B. (2011). Health care and social media: Building relationships via social networks. *Academy of Health Care Management Journal, 7*(2), 1–14.

Haenlein, M., Anadol, E., Farnsworth, T., Hugo, H., Hunichen, J., & Welte, D. (2020). Navigating the new era of influencer marketing: How to be successful on Instagram, TikTok, & co. *California Management Review, 63*(1), 5–25.

Hewitt, A. M. (2011). Aligning social media, social networking and social marketing: engagement as the key to healthcare marketing. *Proceedings of Northeast Business and Economics Association, 4*(2), 211–214.

Huang, E., & Dunbar, C. L. (2013). Connecting to patients via social media: A hype or a reality? *Journal of Medical Marketing, 13*(1), 14–23. https://doi.org/10.1177/1745790413477647

Hur, M., Kang, P., & Cho, S. (2016). Box-office forecasting based on sentiments of movie reviews and independent subspace method. *Information Sciences, 372*, 608–624. https://doi.org/10.1016/j.ins.2016.08.027

İlgün, G., & Uğurluoğlu, Ö. (2019). How Turkish private hospitals use social media: A qualitative study. *Journal of Social Service Research, 45*(1), 34–43. https://doi.org/10.1080/01488376.2018.1479339

Jang, K., & Baek, Y. M. (2019). When information from public health officials is untrustworthy: the use of online news, interpersonal networks, and social media during the MERS outbreak in South Korea. *Health Communication, 34*(9), 991–998. https://doi.org/10.1080/10410236.2018.1449552

Kingsford, C., & Salzberg, S. L. (2008). What are decision trees? *Nature Biotechnology, 26*(9), 1011–1012. https://doi.org/10.1038/nbt0908-1011

Kotler, P., & Zaltman, G. (1971). Social marketing: An approach to planned social change. *Journal of Marketing, 35*, 3–12.

Koumpouros, Y., Touliasa, T. L., & Koumpouros, N. (2015). The importance of patient engagement and the use of social media marketing in healthcare. *Technology and Health Care, 23*, 495–507. https://doi.org/10.3233/THC-150918

Kujur, F., & Singh, S. (2016). Social networking sites as a multimedia tool for brand popularity - an exploratory study. *Indian Journal of Science and Technology, 9*, 45. https://doi.org/10.17485/ijst/2016/v9i45/98151

Lash, M. T., & Zhao, K. (2016). Early predictions of movie success: The who, what, and when of profitability. *Journal of Management Information Systems, 33*(3), 874–903. https://doi.org/10.1080/07421222.2016.1243969

Limbu, Y. B., Pham, L., & Mann, M. (2020). Corporate social responsibility and hospital brand advocacy: Mediating role of trust and patient-hospital identification and moderating role of hospital type. *International Journal of Pharmaceutical and Healthcare Marketing, 14*(1), 159–174. https://doi.org/10.1108/IJPHM-04-2019-0029

Liu, X., Shin, H., & Burns, A. C. (2019). Examining the impact of luxury brand's social media marketing on customer engagement: Using big data analytics and natural language processing. *Journal of Business Research, 125*, 815–826. https://doi.org/10.1016/j.jbusres.2019.04.042

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., … Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures, 12*(2–3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

Meer, T. G. L. A., & Jin, Y. (2020). Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. *Health Communication, 35*(5), 560–575. https://doi.org/10.1080/10410236.2019.1573295

Naik, A., & Samant, L. (2016). Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime. *Proc Comput Sci, 85*(Cms), 662–668. https://doi.org/10.1016/j.procs.2016.05.251

Octoparse. (2021). Web Scraping Tool & Free Web Crawlers | Octoparse. Retrieved January 20, 2021, from https://www.octoparse.com/

OECD. (2021a). Gross domestic product (GDP) (indicator). Retrieved January 3, 2021, from https://doi.org/10.1787/dc2f7aec-en

OECD. (2021b). Health expenditure. Retrieved January 24, 2021, from https://www.oecd.org/health/health-expenditure.htm#:~:text=Health%20spending%20as%20a%20share,second%20and%20third%20highest%20spenders

OECD. (2021c). Health workforce. Retrieved January 24, 2021, from http://www.oecd.org/els/health-systems/workforce.htm

Patel, B. R., & Rana, K. K. (2014). A survey on decision tree algorithm for classification. *IJEDR, 2*(1), 1–5.

Popović, K., & Smith, D. C. (2010). Tweeting@doctorwelby: Practical examples of social media in healthcare. *Journal of Communication in Healthcare, 3*(2), 138–151. https://doi.org/10.1179/175380710X12813488575925

Radu, G., Solomon, M., Gheorghe, C. M., Hostiuc, M., Bulescu, I. A., & Purcarea, V. L. (2017). The adaptation of health care marketing to the digital era. *Journal of Medicine and Life, 10*(1), 44–46.

Reis, I., Baron, D., & Shahaf, S. (2018). Probabilistic random forest: A machine learning algorithm for noisy data sets. *The Astronomical Journal, 157*(1), 16. https://doi.org/10.3847/1538-3881/aaf101

Sabate, F., Berbegal-Mirabent, J., Cañabate, A., & Lebherz, P. R. (2014). Factors influencing popularity of branded content in Facebook fan pages. *European management journal, 32*(6), 1001–1011.

Saglik (2020). Health statistics yearbook. Retrieved January 2, 2021, from https://www.saglik.gov.tr/TR,62400/saglik-istatistikleri-yilligi-2018-yayinlanmistir.html

Schillinger, D., Chittamuru, D., & Ramírez, A. S. (2020). From "infodemics" to health promotion: a novel framework for the role of social media in public health. *American Journal of Public Health, 110*(9), 1393–1396. https://doi.org/10.2105/AJPH.2020.305746

Schultz, C. D. (2017). Proposing to your fans: Which brand post characteristics drive consumer engagement activities on social media brand pages? *Electronic Commerce Research and Applications, 26*, 23–34. https://doi.org/10.1016/j.elerap.2017.09.005

Sharda, R., & Delen, D. (2006). Predicting box-office success of motion pictures with neural networks. *Expert Systems with Applications, 30*(2), 243–254. https://doi.org/10.1016/j.eswa.2005.07.018

Sharma, V., & Gupta, M. (2019). Correlation between attributes concerning social media marketing and private healthcare sector. *MANTHAN: Journal of Commerce and Management, 6*(1), 1–15. https://doi.org/10.17492/manthan.v6i1.182678

Shawky, S., Kubacki, K., Dietrich, T., & Weaven, S. (2019). Using social media to create engagement: a social marketing review. *Journal of Social Marketing, 9*(2), 204–224. https://doi.org/10.1108/JSOCM-05-2018-0046

Stadtelmann, M., Woratschek, H., & Diederich, C. (2019). Actor engagement in online health communities: A netnographic analysis of engagement practices and roles. *International Journal of Pharmaceutical and Healthcare Marketing, 13*(4), 500–515. https://doi.org/10.1108/IJPHM-06-2018-0033

Tso, G. K. F., & Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy, 32*(9), 1761–1768. https://doi.org/10.1016/j.energy.2006.11.010

Wang, R., & Jin, Y. (2010). An empirical study on the relationship between the followers' number and influence of microblogging. *Proc Int Conf E-Bus E-Gov ICEE, 2010*(2007), 2014–2017. https://doi.org/10.1109/ICEE.2010.509

Wuest, T., Weimer, D., Irgens, C., & Thoben, K. D. (2016). Machine learning in manufacturing: Advantages, challenges, and applications. *Prod Manuf Res, 4*(1), 23–45. https://doi.org/10.1080/21693277.2016.1192517

Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. *Transport Res Part C: Emerg Technol, 58*, 308–324. https://doi.org/10.1016/j.trc.2015.02.019