

The Use of Text Mining in New Media Studies: A Systematic Review

Yeni Medya Çalışmalarında Metin Madenciliğinin Kullanımı: Sistematiik Bir İnceleme

Derya ŞAHİN 

Araştırma Makalesi Research Article

Başvuru Received: 10.08.2023 ■ Kabul Accepted: 27.11.2023

ABSTRACT

Text mining is a method that relies on the text as the source and aims to get the data through the text. This method has a strong potential in obtaining real-time findings and effective results in terms of the evaluation of large data sets. It is valuable to understand the people and the whole picture by using such powerful and effective methods obtained from rich sources like new media. In this context, the aim of this study is to define the basic tendencies and the changes that happened in the studies on new media and text mining in the last decade. This study analyses 82 articles published on Google Scholar between 2013 and 2023. The articles were analysed using a systematic analysis method which contains specific criteria in accordance with the aim of this research. The research results indicate that among the text mining methods in articles, Twitter is the most preferred new media platform (40.2%), and in a sectoral context, research primarily focuses on the tourism field (18.3%).

Keywords: Text Mining, New Media, Text Mining in New Media, Text Mining Applications, Future of Text Mining.

ÖZ

Metin madenciliği, bir metni metin kaynağı olarak kabul eden ve metin üzerinden yapılandırılmış veri elde etmeyi amaçlayan bir yöntemdir. Bu yöntem gerçek zamanlı bulgular elde etmek için ve geniş veri setlerinin değerlendirilmesi neticesinde etkin sonuçlar elde edilmesi kapsamında güçlü bir potansiyel sunmaktadır. Özellikle yeni medya gibi zengin kaynaklardan elde edilen verileri anlamak ve kavramak için bu denli etkin ve güçlü yöntemleri kullanmak, insanları anlamaya çalışmak ve bütünü algılamamız adına değerlidir. Bu bağlamda çalışmanın amacı dünyada yeni medya ve metin madenciliği konulu çalışmaların son on yılda nasıl bir gelişim gösterdiğini ve temel eğilimlerini ortaya koymaktır. Araştırmanın amacı ve belirlenen özel kriterler doğrultusunda 2013-2023 yılları arasında yayınlanan ve Google Scholar'da taranan toplam 82 makale sistematiik inceleme metodu kullanılarak analiz edilmiştir. Araştırma sonucunda makalelerde metin madenciliği yöntemi için en çok tercih edilen yeni medya ortamlarının Twitter (%40,2) olduğu ve sektörel bağlamda ise araştırmaların en çok turizm (18,3) alanına odaklanıldığı tespit edilmiştir.

Anahtar Kelimeler: Metin Madenciliği, Yeni Medya, Yeni Medyada Metin Madenciliği, Metin Madenciliği Uygulamaları, Metin Madenciliğinin Geleceği.



Introduction

Information holds indispensable value for societies. It enriches people's lives, solves problems, creates new opportunities, and fosters progress. Information guides the development of societies and enables individuals to build a better future. Therefore, the importance of information for society is significant, and ensuring access to, sharing, and preserving information contributes to the sustainable advancement of communities. It's essential to recognize that information holds immense importance for institutions as well.

Information is an important resource for today's business world. Companies need to have sufficient information about their customers, employees, and their other partners. They obtain structured and unstructured information from various sources such as surveys, tweets, call centre notes, phone records, online customer reviews, emails, social network posts, medical records, and other sources e.g. However, it is not easy to interpret these resources without using convenient text analysis tools. Text analysis process can be conducted manually but it was proven to be ineffective. Moreover, traditional systems often use keywords and cannot provide a complete reading and understanding of the language in emails, tweets, web pages, and text documents. Hence, companies use text analysis programs to analyse large volumes of data. These programs help their users to obtain information from text data in order to take appropriate actions (Çelik, 2020:1344).

With the development of technology, the amount of data recorded on the Internet increases day by day. Among these data, the analysis of unstructured data, such as text files, forum data, and e-mail contents, is a challenging process as traditional query methods are insufficient in the analysis of such data. Therefore, texts should be classified in order to extract valuable information from text stacks. Especially, analyzing a large amount of forum data on the internet requires a lot of effort, cost, and time. In this context, for text mining, automatic detection and statistical evaluation of forum data is important. Nowadays, text mining

and machine learning studies focus on structured data in databases or data warehouses. However, considering the real-life findings, data in text files and on web pages consist of unstructured textual data such as e-mail contents, articles, blogs, and open-ended questionnaire answers (Göker & Tekedere, 2017:292).

In today's rapidly evolving world, the pervasiveness of new media has fundamentally transformed the way we live and behave. As such, it is imperative that we explore and evaluate this complex landscape from multiple perspectives, in order to gain a deeper understanding of its impact on our lives. One promising way to do this is through the use of text mining, which has been shown to be an effective research method (Çalış et al., 2013; He et al., 2013; Hashimi et al., 2015; Gupta et al., 2016; Philer & Zhong, 2016; Kim & Hastak, 2017; Salloum et al., 2017; Al-Daihani & Abrahams, 2018; Nisar & Yeung, 2018; Onan, 2021; Temizhan & Mendeş, 2021) for analyzing and interpreting the vast amounts of textual content available in new media environments.

Despite the significant potential of text mining, there is still much we do not know about how it is being used in practice. For instance, we do not have a clear understanding of which new media environments are being studied most frequently, which sectors are most focused on, or which applications are preferred for text mining analysis. To address these gaps in our knowledge, it is essential that we conduct further research into the use of text mining in new media environments. By doing so, we can gain critical insights that will inform future studies and help us to fully realize the potential of this powerful research tool.

At this point, the study aims to reveal how the studies on text mining and new media published in English-language journals scanned in Google Scholar have been developing and trending over the past decade. Moreover, it is to demonstrate how the scientific text mining method reflected upon new media studies in recent years.

Text Mining

Text mining was first discussed in the study of Feldman and Dagan (1995). According to the study, text mining is the process of searching and extracting valuable information from text data. It presents an exciting research area as it attempts to discover information from unstructured texts (Vijayarani et al., 2015:7).

Text mining is defined as the process of structured texts having information obtained from unstructured texts. Texts need to be processed to extract meaningful information. Thus, it consists of some steps like data pre-processing and feature extraction. After these steps, text mining converts unstructured data into a structured format that can be processed by computers (Hotho et al., 2005:22)

Text mining automatically presents individuals with new and previously unknown information from different sources. Its main element is combining extracted information to present new realities or new hypotheses. This also leads to more discoveries. It is also something different from what we are familiar with on the pages about it. While searching, users generally search for something that has already been written or known by someone else. The problem is disregarding all the materials that do not suit the needs to find the information needed. However, in text mining, the objective is to discover information that is unknown and has not been written yet (Gupta & Lehal, 2009:60).

Text mining contains statistical and mathematical methods. It is widely used in different fields such as author recognition, text classification, idea mining, emotional analysis, keyword extraction, and title extraction.

Although it is a sub-branch of data mining, its tools and inputs are quite different. In data mining, data in table format from databases or files is commonly used as input. However, for text mining, input files may not be in tabular format, and they

can be in formats such as HTML web pages, PDFs, or Word documents. In text mining, there is a different approach from a typical search operation performed in a standard database. For instance, in a regular search operation performed with Ctrl+F or SQL commands, the term to be searched is predetermined, and it is queried whether this term exists within a specific source. On the other hand, rather than a search operation, text mining aims to extract the elements that are not known previously either in the text or in the context of it. In other words, text mining is a process aiming to find unknown concepts in texts before their analysis (Atan, 2020:24)

Text mining deals with certain problems. Some of them are as follows:

Information Extraction: Selection of special statements, such as proper names, place names, personal names, time, and event, within unstructured data and methods such as pattern matching, and keyword filtration are within the scope of this field. Information extraction enables us to find the widely used keywords such as verbs, names, locations, etc. in a context like finding the common politician and location names or the most used verbs in a news website. Instead of obtaining a result from information bulks, it differentiates specific matters separate from each other. It also aids to reveal the related topics and categories of information rather than classifying them one by one. This function helps researchers, who follow news about specific people and organisations, to find related news instead of searching with keywords. The ability of an algorithm to extract information such as place names, custom names, dates, events, and events from text stacks can provide significant advantages in cases where there is a large amount of text. For example, models that can constantly scan and filter the events on published news can be developed thanks to algorithms that can parse the name, location, and time of events. Under normal circumstances, it is possible to find such information manually but the words that will be used during the process must

be determined in advance. However, information extraction algorithms can extract such information without needing pre-determined keywords.

Emotional analysis: It is a method that aims to extract emotional expressions in texts. Sentimental polarity is the most common method. In this method, positive and negative comments or articles about a topic are evaluated and divided into two categories. Additionally, emotional analysis also deals with mental states, opinions, and further mixed emotions.

Similarity detection: Text mining technique is also used to detect whether there are any similarities between texts. It is in text mining's scope to determine and understand whether there is a similarity between two texts, or whether a text is derived from another one. Often preferred in the academic field, there are programs developed for this purpose. In this context, programs are not produced only for detecting the similarity between the texts but to prevent the confusion in meaning caused by deceptive programs. As an example, algorithms that correct the text upon entering into search engines and understand what the intended meaning to convey is then provided with the correct text.

Information Retrieval: It is a stage where preliminary information about the corpus (collection) of interest is gathered. For example, in text mining, if the data sources are on the web, this stage involves gathering information such as web page addresses, file dates, user information, file names, directory details, etc., from the web pages or from the file system (Şeker, 2015:32).

The Relationship Between New Media and Text Mining

Today's world is going through a period of rapid and unprecedented change, fueled by globalization and technology. While technology has undoubtedly made human life more convenient, it has also brought about profound and rapid changes within the social framework, leading to radical departures

from old habits (Okmeydan, 2017: 350). The fast-paced transformation in the world of technology has had a significant impact on communication (Postman, 2016:23), diversifying its layered structure and completely altering the source of societal information flow.

As a result of the digitalization of communication through technological infrastructure, we are now experiencing a transformation in the social fabric. In this age, where we are constantly exposed to the powerful effects of the digital world, interpreting and evaluating information found in new media environments has become a crucial matter. Therefore, it is essential to recognize the importance of critical thinking in the age of digital media, in order to navigate the complex web of information flow and make informed decisions.

The collection of large-scale data in new media environments is leading to new approaches in data analysis, which has resulted in significant breakthroughs in understanding different paradigms. A popular approach in recent years is text mining, which allows researchers to interpret massive data sets in new media environments and extract meaningful insights that were previously hidden.

The versatility of text mining is demonstrated in the extensive literature, as it is used across various sectors (Stieglitz & Xuan, 2013; Xiang et al., 2015; Daihani & Abrahams, 2016; Shen et al., 2019; Yost et al., 2021) and topics (Öztürk & Ayvaz, 2018; Tadesse et al., 2019; Chen, et al., 2020; Loureda et al., 2023; Turenne, 2023), providing valuable information that can benefit future studies.

Methodology

Objectives and Methods

The aim of this study is to define the basic tendencies and the changes in the studies on new media in the world and text mining in the last decade. Especially in studies conducted using text mining methods in new media environments, the

question of which applications are predominantly used for research forms the backbone of the study.

Studying published articles provides evidence to identify the different methods and trends used by researchers. As the new media environments have changed and developed significantly, it is possible to observe the change conducted in this field. Particularly in the recent period, it is essential to make a general analysis of new media studies that were prepared with the text mining method. This study aims to conduct a global scale analysis of new media studies that were conducted with text mining method between 2013 and 2023 and to determine which applications are more effective and preferred in which sectors. It also aims to contribute to the literature as something covering recent developments and having a global scale.

The inclusion and exclusion criteria of this study are, as suggested by Pittaway et al. (2004), based on the study title, keywords, and the analysis of summaries. Articles that were not within the scope of this research were found by examining titles, summaries, and keywords. In the end, 82 articles that were in the scope were included.

In the study, new media articles using the text mining method and published between 2013-2023 were systematically analysed. A systematic review can be defined as the scrutiny of a formulated question using systematic and transparent methods to identify, select, critically evaluate, and summarize relevant research while striving to minimize bias. (Needleman, 2002: 6). What distinguishes a systematic review from traditional literature searching is that it is carried out comprehensively and methodically according to a pre-defined protocol to minimize bias. (Hanley and Cutts, 2013: 4).

The search for articles on Google Scholar involved adding specific keywords such as “new media and text mining,” “new media platforms and text mining,” “text mining in new media,” and “text

mining method in new media” to the query. From the search results, 108 articles published between January 2013 and July 2023 were retrieved, and 26 duplicate articles were removed, leaving a total of 82 articles for analysis. As there were a limited number of articles available, no sampling was performed, and the analysis was carried out on the entire population of research.

The limitations of the research include the requirement for articles to be indexed in Google Scholar, the articles to be published in a peer-reviewed journal, the articles to be written in English, and their relevance to the field of new media.

These articles were analysed within the context of the publication date, preferred new media environment, sector, and the software used during the process of text mining. The following question will be answered in this study.

- ▶ In new media environments published between 2013 and 2023, scanned in Google; What's the distribution of studies using text mining method by years?
- ▶ In new media environments published between 2013 and 2023, scanned in Google; Which new media environments were preferred?
- ▶ In new media environments published between 2013 and 2023, scanned in Google; Which sectors were preferred?
- ▶ In new media environments published between 2013 and 2023, scanned in Google; Which programs were preferred?

Findings of the Research

In this section, the data is tabulated and evaluated by examining them in detail. Araştırma kapsamında toplanan verilerin analizinde, SPSS 21.0 programı kullanılmıştır.

Table 1
Distribution of articles by years

Year	n	%
2013	7	8,5
2014	3	3,7
2015	3	3,7
2016	8	9,8
2017	5	6,1
2018	8	9,8
2019	7	8,5
2020	9	11,0
2021	6	7,3
2022	18	22,0
2023	8	9,8
Total	82	100

The data in Table 1 shows the distribution of the articles by year. The number of articles published in English was at its peak in 2022 (22%) Additionally, upon the analysis, no parallel distribution between text mining in new media environments and the studies that were conducted was observed.

Table 2
Distribution of new media platforms

New Media Platforms	n	%
Twitter	33	40,2
Facebook	6	7,3
Web Site	34	41,5
Youtube	2	2,4
Other	7	8,5
Total	82	100

The data in Table 2 shows the preferred new media environments in the published articles. It was found in the articles that the most preferred new media environments for text mining method are Twitter (40.2%) and Website (41.5%). In other new media environments, similar rates were observed (%7,3; %2,4; %8,5).

Table 3
Distribution of meta-analysis software used in articles

Program	n	%
SPSS	8	9,8
R Studio	12	14,6
RapidMiner	12	14,6
Python	27	32,9
Other	13	15,9
Unspecified	10	12,2
Total	82	100

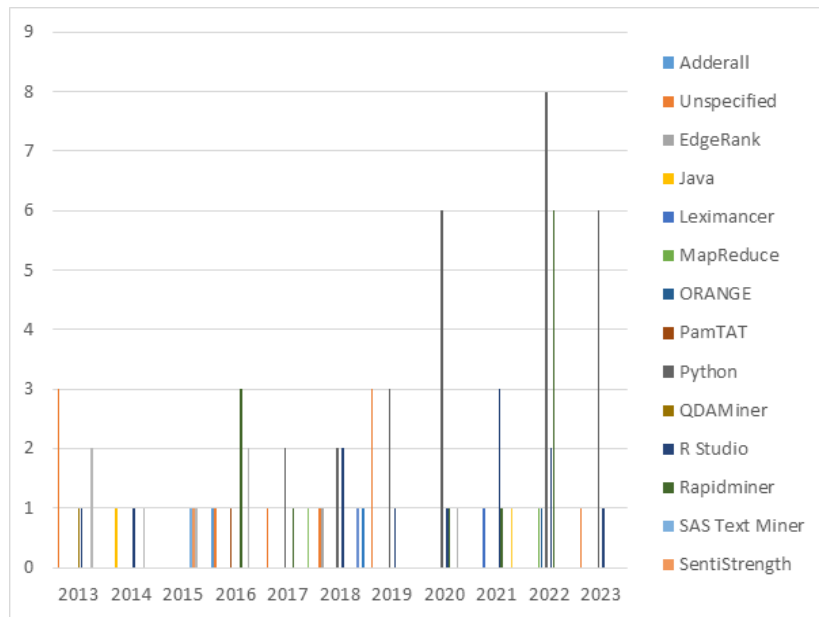
The data in Table 3 shows the preferred software when using the text mining method in the articles. Among the programs, it is observed that Python (32.9%) is the most preferred one in the text mining method. Other programs, respectively, R Studio (14.6%), RapidMiner (14.6%), and SPSS (9.8%) with very close ratios to each other. Furthermore, the programs used in 12.2% of the studies were not specified.

Table 4
Sectoral distribution of research studies

Sector	n	%
Tourism	15	18,3
Health	9	11,0
Communication	11	13,4
Policy	14	17,1
Finance	7	8,5
Marketing	4	4,9
Security	5	6,1
Transportation	7	8,5
Unspecified	10	12,2
Total	82	100

The data in Table 4 shows the distribution of the sectors preferred by the researchers. Upon the analysis, it is observed that the studies were conducted in tourism (18.3%), politics (17.1%), and communication (13.4%) respectively. Moreover, other sectors have similar ratios to each other.

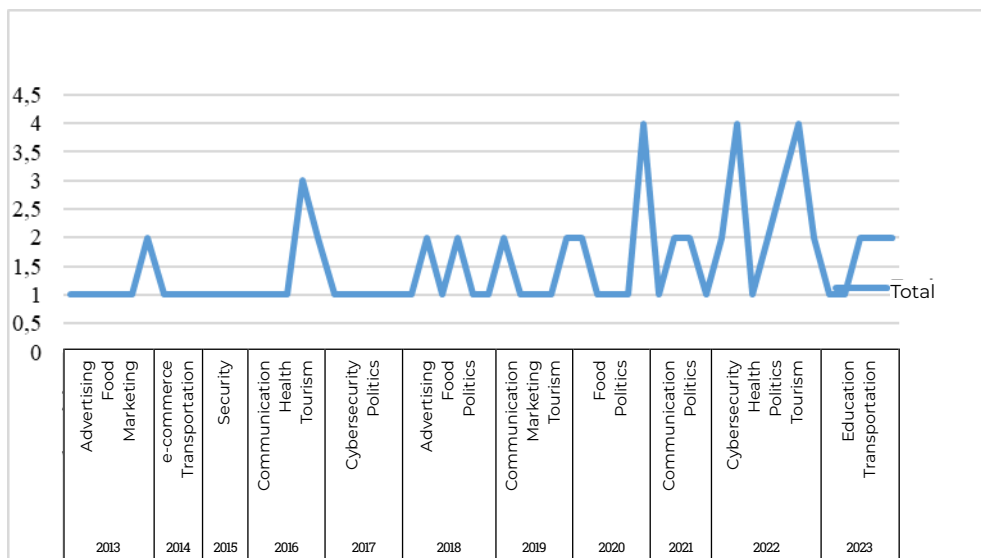
Figure 1
Distribution of software usage by years



The data in Figure 1 shows the preferred programs for text mining method in new media environments by year. According to the data, after 2019, studies use some programs more. Especially, it is found Python, R Studio, and RapidMiner programs were more preferred in the recent ones.

The data in Figure 2 contains the distribution of studies in new media environments by sectors and years. By taking external factors into consideration, in 2019 and the following years, studies on “health” took more place in articles. As a determinative factor, it is necessary to mention “Covid-19” pandemic.

Figure 2
Distribution of new media tools across sectors and years



Conclusion

The literature contains various studies prepared by text mining method (Çalış et al., 2013; He et al., 2013; Hashimi et al., 2015; Gupta et al., 2016; Philer & Zhong, 2016; Kim & Hastak, 2017; Salloum et al., 2017; Al-Daihani & Abrahams, 2018; Nisar & Yeung, 2018; Onan, 2021; Temizhan & Mendes, 2021). The focus of these studies is to draw a conclusion from unstructured texts by using new media environments in their research using the method of text mining.

In this study, which aims to discover the uses of text mining methods and to show the tendency of it between 2013 and 2023, 82 English journal articles scanned in Google Scholar are analysed in-depth in terms of publication dates, used new media environments, preferred sectors and programs used for text mining.

This study is based on a systematic review that creates a scope for understanding the role of text mining in new media environments. As stated above in the sections above, the literature review and research sections are based on four basic questions. In order to find answers to the questions, 82 articles were analysed.

This study is based on a systematic review that provides a framework for understanding the role of text mining in new media environments. According to the research results, the text mining method is preferred as an effective research method. When the articles using the text mining method were evaluated accordingly with the first question of this study which is the distribution of the use of text mining, this method has been used more frequently in recent periods (See table 1) This data also shows the effectiveness and preferability of text mining method in new media environments. This finding is consistent with the studies by Al-Mahmoud and Al-Razgan (2015), Salloum and colleagues (2017), and Pavlidou and Tsui (2020).

The answer of the second question, which is “ In new media environments published between

2013 and 2023, scanned in Google; which mining methods are preferred in which environments?”, is determined from the data in Table 2. In the articles, it can be observed that the most preferred new media environments are Twitter (40.2%) and web pages (41.5%). It has been observed that in articles published in the last decade on the topics of new media and text mining, Twitter is preferred more. This finding is in line with the study by Karami and colleagues (2020).

Another important result of the research is related to the topics of the articles. The topics addressed in the articles generally revolve around the application areas of new media and their strategic utilization. Therefore, it has been observed that data on new media is predominantly investigated through text mining methods, particularly in sectors within the service industry.

As a result of this study, it is recommended that researchers provide specific recommendations based on the findings to contribute to future research in the field. In order to expand the scope of research on the role of text mining methods in new media research, it is suggested to focus on specific subfields of new media, such as social media or online news, to gain a more nuanced understanding of the relationship between new media and text mining. This approach could lead to more targeted and insightful findings that could contribute to further advancements in the field.

By utilizing semiautomated techniques to perform time-consuming tasks in systematic reviewing, reviews can be completed more efficiently and effectively. This not only saves time, but also ensures better results by harvesting, filtering, and summarizing more evidence. These techniques have already proven successful in practice. Furthermore, customized searching, screening, and synthesizing can focus on relevant terms, documents, and information fragments, leading to more targeted and insightful (Rea et al., 2009:520). Due to the increase in the popularity of the internet and social media, nowadays, customers are searching and comparing information on the

Internet about products or services before making any purchase decisions. Hence, online reviews and social media data play an important role in the fields of sales, marketing, business, tourism, and airline industries (Kumar et al., 2021:10). Therefore, in new media environments with such vast pools of information, using text mining methods to extract valuable insights and systematically contributing them to the field represents a significant endeavor.

The purpose of this study is to explore the connection between new media and text mining, with the aim of providing guidance for future research. While this research focuses on peer-reviewed journals, it is recommended that future studies also consider seeking the perspectives of experts/practitioners in the field by consulting other publications within the industry. Additionally, Thelen (2021: 8) suggests examining book chapters and articles in commercial publications to gain further insights into the research subject and the perspectives of practitioners.

In conclusion, this study provides insights into the changes related to text mining in new media environments between 2013 and 2023. It is believed that the relationship between new media and text mining will continue to develop each day, and in the coming years, more studies using text mining methods in new media environments will be introduced to the literature.

References

- Al-Daihani, S.M. & Abrahams, A. (2018). Analysis of academic libraries' Facebook posts: Text and data analytics. *J. Acad. Librariansh.* 44, 216–225, doi:10.1016/j.acalib.2018.02.004
- Al-Mahmoud, H. & Al-Razgan, M. (2015). Arabic Text Mining: A Systematic Review of the Published Literature 2002-2014, *IEEE*
- Atan, S. (2020). Metin madenciliği: İmkanlar, yöntemler ve kısıtlar. *Mehmet Akif Ersoy University Journal of Social Sciences Institute*, 31, 220-239, <https://orcid.org/0000-0003-3170-0969>
- Chen, W.K., Riantama, D., & Chen, L.S. (2020). Using a Text Mining Approach to Hear Voices of Customers from Social Media toward the Fast-Food Restaurant Industry, *sustainability*, 13(1), 1-17, <https://doi.org/10.3390/su13010268>
- Çalış, K., Gazdağı, O. & Yıldız, O. (2013). Reklam içerikli epostaların metin madenciliği yöntemleri ile otomatik tespiti. *Bilişim Teknolojileri Dergisi*, 6(1), 1-7.
- Çelik, S. (2020). Metin madenciliği ile Shakespeare külliyyatının incelenmesi. *MANAS Sosyal Araştırmalar Dergisi*, 9(3), 1344-1357.
- Daihani, S.M., & Abrahams, A. (2016). A Text Mining Analysis of Academic Libraries' Tweets, *The Journal of Academic Librarianship*, 42(2), 135-143, <https://doi.org/10.1016/j.acalib.2015.12.014>
- Feldman, R. & Dagan, I. (1995). Kdt - Knowledge discovery in texts. In *Proc. of The First Int. Conf. on Knowledge Discovery (KDD)*, 112–117.
- Göker, H. & Tekedere, H. (2017). FATİH projesine yönelik görüşlerin Metin madenciliği yöntemleri ile otomatik değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 10(3), 291-299, DOI:10.17671/gazibtd.331041
- Gupta, V. & Lehal, G.S. (2009). A survey of text mining techniques and applications. *Journal of Emerging Technologies in Web Intelligence*, 1(1), 60-76.
- Gupta, B., Sharma, S. & Chennamaneni, A. (2016). Twitter sentiment analysis: An examination of cybersecurity attitudes and behavior. In *Proceedings of the 2016 Pre-ICIS SIGDSA/IFIP WG8.3 Symposium: Innovations in Data Analytics*, Dublin, Ireland, 11 December, 17.
- Hanley, T. & Cutts, L. (2013). What is a systematic review. *Counselling Psychology Review*, 28(4), 3-6.

- Hashimi, H., Hafez, A. & Mathkour, H. (2015). Selection criteria for text mining approaches. *Computers in Human Behavior*.
- He, W., Zha, S. & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *Int. J. Inf. Manag.*, 33, 464–472, doi:10.1016/j.ijinfomgt.2013.01.001
- Hotho, A., Nurnberger, A. & Paaß, G. (2005). A brief survey of text mining. *LDV Forum - GLDV Journal for Computational Linguistics and Language Technology*, 20(1), 19-62.
- Karami, A., Lundy, M., Webb, F. & Dwivedi, Y.K. (2020). Twitter and Research: A Systematic Literature Review Through Text Mining, in *IEEE Access*, vol. 8, pp. 67698-67717, 2020, doi:10.1109/ACCESS.2020.2983656.
- Kılınç, D., Borandağ, E., Yücalar, F., Tunalı, V., Şimşek, M. & Özçift, A. (2016). KNN algoritması ve R dili ile metin madenciliği kullanılarak bilimsel makale tasnifi. *Marmara Fen Bilimleri Dergisi*, 28(3), 89-94.
- Kim, J. & Hastak, M. (2018). Social network analysis: Characteristics of online social networks after a disaster. *Int. J. Inf. Manag.*, 38, 86–96, doi:10.1016/j.ijinfomgt.2017.08.003
- Koyuncugil, A.S. (2006). Bulanık veri madenciliği ve sermaye piyasalarına uygulanması. Ankara Üniversitesi Fen Bilimleri Enstitüsü, Yayınlanmış doktora tezi.
- Kumar, S., Kar, A.K. & Ilavarasan, P.V. (2021). Applications of text mining in services management: A systematic literature review, *Information Management Data Insights*, 1(1), 1-11, https://doi.org/10.1016/j.ijime.2021.100008
- Loureda, M.C., Baquerio, A.H., & Tamés-Muñoz, E. (2023). A text mining analysis of human flourishing onTwitter, *Scientific Reports*, 13.
- Needleman, I. G. (2002). A guide to systematic reviews. *Journal of Clinical Periodontology*, 29(s3), 6-9. https://doi.org/https://doi.org/10.1034/j.1600-051X.29.s3.15.x
- Nisar, T.M. & Yeung, M. (2018). Twitter as a tool for forecasting stock market movements: A short-window event study. *J. Financ. Data Sci.*, 4, 101–119, doi:10.1016/j.jfds.2017.11.002
- Okmeydan, S. B. (2017). Yeni İletişim Teknolojilerini Sorgulamak: Etik, Güvenlik ve Mahremiyetin Kesiştiği Nokta. *Gümüşhane Üniversitesi İletişim Fakültesi Elektronik Dergisi*, 5 (1), 347-371.
- Onan, A. (2021). COVID-19 ile ilgili sosyal medya gönderilerinin metin madenciliği yöntemlerine dayalı olarak zaman-mekansal analizi. *Avrupa Bilim ve Teknoloji Dergisi*, 26, 138-143.
- Öztürk, N., & Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis, *Telematics and Informatics*, 35, 136-147.
- Pavlidou, I. & Tsui, E. (2020). Crowdsourcing: a systematic review of the literature using text mining, *Industrial Management & Data Systems*, 120(11), 2041-2065, DOI 10.1108/IMDS-08-2020-0474
- Philer, K. & Zhong, Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *Int. J. Hosp. Manag.*, 55, 16–24, doi:10.1016/j.ijhm.2016.02.001
- Pittaway, L., Robertson, M., Munir, K., Denyer, D., & Neely, A. (2004). Networking and innovation: A systematic review of the evidence. *International Journal of Management Reviews*, 5(3-4), 137- 168.
- Pilavcılar, İ.F. (2007). Metin madenciliği ile metin sınıflandırma. Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü. Yayınlanmış Published Yüksek lisans tezi.

- Postman, N. (2016). Teknopoli Kültürün Teknolojiye Teslim Oluşu, Sentez Yayıncılık İstanbul.
- Rea, B., Okazaki, N., Procter, R. & Thomas, J. (2009). Supporting Systematic Reviews Using Text Mining, *Social Science Computer Review*, 27(4), 509-523.
- Salloum, S.A., Mhamdi, C., Al-Emran, M. & Shaalan, K. (2017). Analysis and classification of Arabic newspapers' Facebook pages using text mining Techniques. *Int. J. Inf. Technol. Lang. Stud*, 1, 8-17.
- Shen, C., Chen, M., & Wang, C. (2019). Analyzing the trend of O2O commerce by bilingual text mining on social media, *Computers in Human Behavior*, 101, 474-483, <https://doi.org/10.1016/j.chb.2018.09.031>
- Stieglitz, S., & Xuan, L.D. (2013). Social media and political communication: a social media analytics framework, *Soc. Netw. Anal. Min.*, 3, 1277-1291.
- Şeker, S.E. (2015). Metin madenciliği (text mining). *YBS Ansiklopedi*, 2(3), 30-32.
- Tadesse, M.M., Lin, H., Xu, B., & Yang, L. (2019). Detection of Depression-Related Posts in Reddit Social Media Forum, *IEEE Access*, 7, 44883 – 44893, [10.1109/ACCESS.2019.2909180](https://doi.org/10.1109/ACCESS.2019.2909180)
- Temizhan, E. & Mendeş, M. (2021). COVID-19 pandemisi ile ilgili Twitter mesajlarının metin madenciliği tekniği ile değerlendirilmesi. *Türkiye Klinikleri Biyoistatistik Dergisi*, 13(2), 185-200, DOI: 10.5336/biostatic.2020-79992
- Thelen, P. D. (2021). The status of public relations research addressing Latin America: A content analysis of published articles from 1980 to 2020. *Public Relations Review*, 47(4), 102079. <https://doi.org/https://doi.org/10.1016/j.pubrev.2021.102079>
- Turenne, N. (2022). Net activism and whistleblowing on YouTube: a text mining analysis, *Multimedia Tools and Applications*, 82, 9201-9221, <https://doi.org/10.1007/s11042-022-13777-0>
- Xiang, Z., Schwartz, Z., Gerdes, J.H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction?, *International Journal of Hospitality Management*, 44, 120-130, <https://doi.org/10.1016/j.ijhm.2014.10.013>
- Vijayarani, S., Ilamathi, J. & Nithya (2015). Preprocessing techniques for text mining - an overview. *International Journal of Computer Science & Communication Networks*, 5(1), 7-16.
- Yost, E., Zhang, T., & Qi, R. (2021). The power of engagement: Understanding active social media engagement and the impact on sales in the hospitality industry, *Journal of Hospitality and Tourism Management*, 46, 83-95, <https://doi.org/10.1016/j.jhtm.2020.10.008>

Genişletilmiş Özet

Teknolojinin gelişmesiyle birlikte internet ortamında kaydedilen verilerin miktarı gün be gün artmaktadır. Bu bilgiler arasında yer alan metin dosyaları, forum verileri ve e-posta içerikleri gibi yapısal olmayan verilerin analizinde klasik sorgu yöntemleri yetersiz kalmaktadır. Bu nedenle, değerli bilginin metin yığınlarından çıkarılabilmesi için metinlerin sınıflandırılması gerekmektedir. Özellikle büyük miktardaki forum verilerinin internet ortamında analiz edilmesi, çok fazla emek, maliyet ve zaman gerektiren bir süreçtir. Bu bağlamda, forum verilerinin otomatik olarak tespit edilmesi ve istatistiksel olarak değerlendirilmesi metin madenciliği yönteminin önemini ortaya koymaktadır.

İlk olarak Feldman ve Dagan (1995) tarafından gerçekleştirilen çalışmada ele alınan metin madenciliği, esas olarak yapısal olmayan metinlerden bilgi içeren yapısal metinleri üretme süreci olarak tanımlanmıştır. Anlamlı bilgilerin elde

edilmesi için metinlerin işlenmesi gerekmektedir, bu da veri ön işleme ve özellik çıkartımı gibi adımları içerir. Bu aşamalardan sonra yapısal olmayan veriler metin madenciliği tarafından kullanılıp bilgisayarlar tarafından işlenebilen yapısal bir biçime dönüştürülmektedir (Hotho ve ark., 2005:22). Metin madenciliği yöntemlerinin özünde istatistiksel ve matematiksel yöntemler yer almaktadır. Metin madenciliği, yazar tanıma, metin sınıflandırma, fikir madenciliği, duygu analizi, anahtar kelime çıkartma, başlık çıkartma gibi farklı alanlarda sıklıkla kullanılmaktadır (Kılınç ve ark., 2016:90).

Bu bilgiler ışığında çalışmanın amacı, Dünyada yeni medya ve metin madenciliği konulu çalışmaların son on yılda nasıl bir gelişim gösterdiğini ve temel eğilimlerini ortaya koymaktır. Özellikle yeni medya ortamlarında metin madenciliği yöntemi kullanılarak yapılan çalışmalarda daha çok hangi uygulamalar üzerinden araştırmaların gerçekleştirildiği sorusu da çalışmanın temel iskeletini oluşturmaktadır. Böylesine güçlü bir değişimin yaşandığı yeni medya çalışmalarının metin madenciliği yöntemi ile yapılan çalışmalarda hangi uygulamanın daha etkili olduğunu ve hangi sektörlerde daha çok tercih edildiğini belirlemek amacıyla 2013-2023 yıllarının Dünya genelinde bir incelemesinin ele alınması amaçlanmaktadır. Çalışmanın hem güncel bir zaman aralığını kapsamaması hem de Dünya genelinde bir araştırma olması açısından literatüre katkı sağlaması hedeflenmektedir.

Çalışmada 2013-2023 yılları arasında Google Scholar'da yayınlamış yeni medya ortamlarında metin madenciliği yöntemi kullanılarak yazılmış makalelerin sistematik bir analizi yapılmış ve bu makaleler; yayın yılı, tercih edilen yeni medya ortamı, sektör ve metin madenciliği analizinde kullanılan programlar kapsamında incelenmiştir.

Araştırma sonuçları, metin madenciliği yönteminin yeni medya ortamlarında etkin bir araştırma yöntemi olarak tercih edildiği göstermektedir. Çalışmanın ilk sorusu olan yıllara göre metin

madenciliği yöntemini kullanan makaleler değerlendirildiğinde son dönemlerde bu yöntemin daha sık bir şekilde kullanıldığı (Bknz: Tablo 1) tespit edilmiştir. Bu veri de yıllar içerisinde metin madenciliği yönteminin yeni medya ortamı için daha etkin ve tercih edilebilir bir yöntem olduğunu göstermektedir.

Makalelerde metin madenciliği yöntemi için en çok tercih edilen yeni medya ortamlarının Twitter (%40,2) ve Web sitesi (%41,5) olduğu tespit edilmiştir. Özellikle Twitter ve bireylerin rahatlıkla yorumlarını yapabildikleri web siteleri yazınsal metin özellikleri taşıması bakımından araştırmalar için çok büyük bir veri sunmaktadır. Benzer şekilde makalelerde kullanılan programlar incelendiğinde araştırmaların özellikle üç programda yoğunlaştığı (Bknz: Tablo 3) görülmektedir.

Son olarak sektörel bağlamda değerlendirildiğinde ise araştırmaların en çok turizm, iletişim ve politika sektörlerine (Bknz:Tablo 4) yöneldiği tespit edilmiştir. Ülkemizdeki genel paradigmalarda düşünüldüğünde bu üç konunun daha fazla çalışmalarda yer alması beklenen bir sonuç olarak karşımıza çıkmaktadır. Bununla birlikte araştırma kapsamında elde edilen veriler ışığında 2019 ve sonrasında "Sağlık" konulu çalışmaların daha çok makalelerde yer aldığı göz önüne alındığında, bu durumun en önemli belirleyicisi olarak "Covid 19" salgınına belirtmek önemlidir.

Sonuç olarak bu çalışmanın, 2013-2023 yılları arasında yeni medya ortamlarında metin madenciliğiyle bağlantılı meydana gelen değişikliklere dair literatüre fikir vermesi umulmaktadır. Yeni medya ve metin madenciliği ilişkisinin her geçen gün daha da gelişeceği ve gelecek yıllarda yeni medya ortamlarında metin madenciliği yöntemini kullanacak başka çalışmaların da literatüre kazandırılacağı düşünülmektedir.

Yazar Bilgileri

Author details

(Sorumlu Yazar **Corresponding Author**) Dr. Öğr. Üyesi,
İstanbul Aydın Üniversitesi Anadolu Bil Meslek Yüksekokulu,
deryasahin2@aydin.edu.tr

Destekleyen Kurum/Kuruluşlar

Supporting-Sponsor Institutions or Organizations:

Herhangi bir kurum/kuruluştan destek alınmamıştır. None

Çıkar Çatışması

Conflict of Interest

Herhangi bir çıkar çatışması bulunmamaktadır. None

Kaynak Göstermek İçin

To Cite This Article

Şahin, D. (2023). The use of text mining in new media studies:
a systematic review. *Yeni Medya*, (15), 35-47, <https://doi.org/10.55609/yenimedya.1340590>