Metin Madenciliği ve Duygu Analizi Kullanarak Çevrimiçi İncelemelerden Alzheimer İlaçlarına İlişkin Kullanıcı Deneyimlerinin Değerlendirilmesi

***

Evaluating User Experiences of Alzheimer's Drugs from Online Reviews Using Text Mining and Sentiment Analysis

İbrahim BUDAK[1]
Günay KILIÇ[2]
Arzu ORGAN[3]

DOI:10.33461/uybisbbd.1362821

Makale Bilgileri

Makale Türü: Araştırma Makalesi

Geliş Tarihi: 20.09.2023

Kabul Tarihi: 14.12.2023

Öz

Metin madenciliği, yapılandırılmamış metin verilerinden yararlı kalıplar, eğilimler, modeller ve kurallar bulmaya çalışan yeni bir teknolojidir. Metin Madenciliğinde en yaygın kullanılan tekniklerden biri Duygu Analizidir. Duygu analizi, yazarın tutumunu keşfetmek için en yaygın kullanılan sınıflandırma aracıdır. Bir metin aracılığıyla yazarın tutumunun olumlu, olumsuz veya tarafsız olup olmadığını araştırır. İnternet çağında bilginin büyük bir kısmının metin olarak bulunması nedeniyle Duygu analizinin önemi ve kullanım alanları her geçen gün artmaktadır. Sosyal medyada sıklıkla kullanılan duygu analizi, kullanıcıların belirli bir konu veya ürün hakkındaki fikirlerini ortaya çıkarmak için kullanılabilir. Bu çalışmanın amacı, web sitelerindeki ilaç yorumlarını anlamlı bilgilere dönüştürmektir. Bu bilgiler kullanıcılar için karar vermede yardımcı olabilir. Bu çalışmada, 78 kullanıcının, Alzheimer ilaç yorumlarının bulunduğu bir sosyal platformdan elde edilen kişisel veriler değerlendirilmiştir. Özellikle Alzheimer ilaçlarının seçimi diğer ilaçların aksine, hasta ve hasta yakınlarının gözlemlerini birlikte değerlendirmeye imkân vermektedir. Değerlendirmeyi okuyan ve faydalım bulan 3723 kişi yorumun etkisini güçlendirmektedir. Uygulama aşamasında kullanıcı yorumları Duygu analizi ile polarite değerleri hesaplanmış ve geliştirilen formül ile Alzheimer ilaçları sıralanmıştır. Bu sayede tüketicilerin ilaçlara göre memnuniyet düzeyleri belirlenmiştir.

Anahtar Kelimeler: *Metin Madenciliği, Duygu Analizi, Sosyal Medya, İlaç İncelemeleri, Alzheimer İlaçları.*

Article Info

Paper Type: Research Paper

Received: 20.09.2023

Accepted: 14.12.2023

Abstract

Text mining is a new technology that attempts to find useful patterns, trends, patterns and rules from unstructured text data. One of the most commonly used techniques in Text Mining is Sentiment Analysis. Sentiment analysis is the most widely used classification tool to explore an author's attitude. It explores whether the author's attitude is positive, negative or impartial by means of a text. As most of the information in the internet age is found as text, the importance and usage areas of Sentiment analysis are increasing day by day. Sentiment analysis, which is frequently used in social media, can be used to expose users' ideas about a particular topic or product. The aim of this study is to transform drug reviews on websites into meaningful information. This information can help users in decision-making. In this study, personal data obtained from a social platform with Alzheimer's drug reviews of 78 users were evaluated. In particular, the selection of Alzheimer's drugs, unlike other drugs, allows the observations of the patients and relatives of the patient to be evaluated together. The 3723 people who read the review and found it useful strengthens the effect of the comment. In the implementation phase, polarity values of user comments were calculated with Sentiment analysis and Alzheimer's drugs were ranked with the formula developed. In this way, the satisfaction levels of consumers according to the drugs were determined.

Keywords: *Text Mining, Sentiment Analysis, Social Media, Drug Reviews, Alzheimer's Drugs.*

Atıf/ to Cite (APA): Budak I., Kiliç G. & Organ A. (2023). Evaluating User Experiences of Alzheimer's Drugs from Online Reviews Using Text Mining and Sentiment Analysis. International Journal of Management Information Systems and Computer Science, 7(2), 157-167. DOI:10.33461/uybisbbd.1362821

[1] Dr Öğr. Gör., Rektörlük, Kastamonu Üniversitesi, ibudak@kastamonu.edu.tr, Kastamonu, Türkiye.
[2] Öğr. Gör., Rektörlük, Pamukkale Üniversitesi, gkilic@pau.edu.tr, Denizli, Türkiye.
[3] Prof. Dr., İktisadi ve İdari Bilimler Fakültesi, Pamukkale Üniversitesi, aorgan@pau.edu.tr, Denizli, Türkiye.

1. INTRODUCTION

The "text" in text mining is an information-filled writing portion. Any type of written work, including novels, newspapers, blogs, etc., can be used as a text sample. There is a lot of text available now, and it is expanding daily. Text mining uses a variety of methods and techniques, but its main objective is to glean new and valuable knowledge from text documents. Computer programs are used to analyze text, and then people review the findings. Information retrieval, statistics, artificial intelligence, and other disciplines are all combined in text mining (Bilisoly, 2011).

A helpful technique for looking at the substance of a text or collection of texts is text mining. The words in the texts form the foundation of several text mining techniques. Conceptually, text mining. Understanding current ideologies and conceptions in relation to events, things, subjects, feelings, and behaviors is aided by these concepts. (Song, 2008).

To evaluate if a text is positive negative, or neutral, the sentiment analysis looks at text elements. Thus, it is possible to identify the eigen values in the text. It is possible to decipher meanings that are intended for the aims but are not stated clearly in the text. Additionally, the text's emotional roles might be examined (Cambria et al., 2017).

As the web forum has become an enormous collection of invaluable opinions and comments, more and more researchers are strongly interested in this issue (Shi et al., 2009). This study, conducted its research with Sentiment analysis and data obtained from a web forum. The drug group known as ‘cholinesterase inhibitors’ used in the treatment of Alzheimer's disease has been discussed. Alzheimer's drugs were taken according to only generics. In this context, 78 users' subjective comments on drugs have formed texts to be analyzed. The number of people reading these texts and finding them useful is 3723. The RapidMiner Program (v9.1) ‘analyze sentiment’ module was used for analysis. The results of the analysis, polarity values (positive, negative, and neutral), and the number of useful comments were evaluated together. In this way, the drug scores were calculated and the drugs were ranked.

Some Alzheimer's patients experience advanced forgetfulness. When drug reviews are examined, user experiences in other diseases are obtained from drug users, while the comments of Alzheimer's drugs also include the comments of patient relatives. In this respect, evaluating Alzheimer's drugs differs from evaluating other drugs. The ranking of Alzheimer's drugs was chosen for reasons such as symptom management, delaying symptom progression, improving quality of life, supporting caretakers and improving patients' daily tasks.

2. LITERATURE REVIEW

Numerous scientific and commercial fields have benefited greatly from sentiment analysis. For instance, user profiling, advertising, and discussion systems, as well as financial and political predictions. Sentiment analysis has grown in popularity as a result of these investigations. Making meaning of ideas or expressions from several media (texts, pictures, videos, audio files, etc.) is exceedingly challenging. The properties of an open and covered, regular and irregular source (linguistic, visual, or auditory) must be thoroughly understood (Dragoni et al., 2018). Some studies conducted with the Sentiment analysis in the literature are as follows:

In the study conducted by Subrahmanian and Reforgiato (2008), adjectives, verbs, and envelope components were used for Sentiment analysis. In the proposed approach, the power of emotions between -1 and +1 was tried to be analyzed. There may be slight differences in the structure of the sentence. But the effects are very important. This study is the first study to explain emotions with adjectives, verbs, and envelope components (Subrahmanian, and Reforgiato, 2008).

In the study conducted by Shi, Sun, and Zhang (2009), Sentiment analysis was done for the forums on the Web. The study focused on different ideas from Chinese Web forums. First of all, web forum headings on the same topic are collected in a cluster. Next, a new classification algorithm called Probability word list was proposed rather than classification algorithms such as SVM (Support Vector Machines) and Naive Bayes. In the study, it has been shown that the proposed algorithm gives better performance (Shi et al., 2009).

Unnamalai (2012) conducted research on the Sentiment analysis of the products on the websites. It is common practice to ask merchants who sell products on the Web to review products and related services from their customers. The number of customer reviews of a product is increasing rapidly because of the increase in e-commerce. In e-commerce, a popular product may have thousands of reviews. In this study, unstructured texts were analyzed by Sentiment analysis (Unnamalai, 2012).

Cai and colleagues (2023) researched sentiment classification of MOOC (a social media) user comments based on machine learning. They created experimental datasets by using artificial labeling on the "Chinese University MOOC" platform as a case study. They analyzed their mistake situations, compared the sentiment categorization of four conventional machine learning models, and offered a model optimization strategy (Cai et al., 2023).

Flynn and colleagues (2021) analyzed comments on health-related social media posts, particularly on the Alzheimer's and dementia subreddits, to better understand the user experience on forums and possibly lead to improvements in their effectiveness. The study analyzed the linguistic characteristics of the comments and identified themes of discussion (Flynn et al., 2021).

Saad et al. (2021) conducted a study to determine the efficacy of drugs under specific conditions from user reviews in health web forums. In this study, a hybrid technique using both learning-based and dictionary-based approaches is proposed to achieve better results, and sentiment analysis techniques in the medical domain using general-purpose sentiment dictionaries such as AFFIN, TextBlob and VADER to annotate reviews. TextBlob has shown promising results with high accuracy when used with certain feature engineering techniques (Saad et al., 2021).

In the study by Ajibade et al. (2022), online drug reviews were analyzed with data mining techniques. The study generally focused on analyzing user satisfaction, side effects and drug efficacy in online drug reviews using sentiment analysis. With the obtained sentiment analysis results, modelling was performed with Emzor and May & Baker. As a result of the study, the accuracy rate of modelling with Emzor was 0.891, while the accuracy rate of modelling with May & Baker was calculated as 0.869 (Ajibade et al., 2022).

In the study conducted by Rea and Parsons (2023), the readability, accessibility, quality, visual design and content of online information on dementia medications were analyzed with a quantitative evaluation. It was observed that the readability of the websites related to dementia medications was poor and the quality of information and content varied among the evaluated websites. As a result, it was found that the evaluated websites did not fully represent the health information available online (Rea & Parsons, 2023).

Bae and Lee (2012) used Sentiment analysis to analyze the masses on Twitter. Three million tweets from the popular users were analyzed. In the study, Twitter messages are a reflection of the sensitivity of the platform's most well-liked users. Sentiment analysis was employed as a reliable indicator or gauge of popularity. First, there is a differentiation between the favorable and unfavorable popular user masses. Second, it was discovered that popular users emotionally touch the audience through their tweets. Thirdly, using the data from the first two discoveries, a positive-negative measure for this effect was created. The positive-negative sensitivity change in the time series of viewers has been linked to popular users' real-world sensitivity opinions, according to a Granger causality analysis. (Bae and Lee, 2012). On the other hand, in the literature, using the RapidMiner program with Sentiment analysis, Arabic tweets (Duwairi et al., 2014), English tweets (Tripathi et al., 2015), film reviews (Alsaqer and Sasi, 2017), hotel evaluations (Markopoulos et al., 2015), and complaint detections (Tayel et al., 2013) were also conducted.

4. MATERIAL AND METHOD

In this part of the study, text mining and Sentiment analysis are mentioned.

3.1. Text Mining

Text mining, a part of data mining with textual data, is a new technology. By text mining, meaningful information can be extracted from unstructured text data. It is imperative to use computer techniques to quickly extract useful information from a large number of textual documents. Text mining's aim is to find helpful information (models, patterns, etc.) from various unstructured text data, such as (text files, emails, etc.). Text mining is used efficiently and systematically to get information from texts. In health, business, and education text mining is successfully used by researchers to analyze extensive text data. Witten, Don, Dewsnip, and Tablan have used text mining in digital library documents to find metadata and mark documents (He et al., 2013).

Text mining is a different type of data mining that aims to find different patterns from large databases (Hearst, 2003). The most of data stored are texts, so it is believed that text mining's commercial potential is greater than data mining. A recent study has shown that the ratio of text documents in firms to total information is 80 percent. Text mining is more complex than data mining, as it deals with more complex and natural data. Text mining is an area where multiple disciplines, such as information retrieval, text analysis, clustering analysis, data mining, etc., are used (Tan, 1999).

3.2. Sentiment Analysis

Sentiment analysis is about analyzing people's opinions from a written text. It is one of the most popular research areas in natural language processing. At the same time, it has been working extensively in the fields of text mining, data mining, and web mining. Sentiment analysis is frequently encountered in management sciences and social sciences, except computer science, because of its importance as a whole for business and society. Sentiment analysis is encountered in different aspects of social media such as Twitter, social networks, blogs, forums, comments, etc. (Liu, 2012).

Sentiment analysis helps to reveal positive or negative opinions. It conducts sentence-level analysis with tasks such as answering multiple perspective questions and summarizing, extraction of idea-oriented information, and text analysis. For example, if a system about asking questions and wondering about answers contains people's opinions, the sentiment analysis needs to be able to fully identify the expressions of positive and negative emotions (Wilson, 2005).

The polarity of the given text is first calculated at the sentence level. The words in a sentence are encoded in the order of the sentence. With the help of the association lexicon, it is defined as +1 for positive terms, -1 for negative, and 0 for neutral words which are not defined in the lexicon. The word values are weighted according to negators, amplifiers de-amplifiers, and contrasting words. In this way, the polarity of words is better measured by emphasizing or decreasing. The aim is to calculate the polarity more accurately by using different words. The polarity value of the text is calculated using the equations (1), and (2) given below (Balbi et al., 2018).

The polarity value of a sentence is calculated by equality (1).

$$p_{sij} = \frac{\sum_{k=1}^{p_j} r^*_{w_{ijk}}}{\sqrt{p_j}} \tag{1}$$

p_j means the word count of a sentence, and $r^*_{w_{ijk}}$ means each weighted term score.

After calculating the polarity of each sentence, the document score of review (r_{di}) will be calculated by its sentence polarities. A down-weighted zeros average is used, where neutral sentiment has minor weight.

$$r_{di} = \frac{\sum_{j=1}^{q_i} r_{sij}}{\hat{q}_i + \sqrt{\log(2-\hat{q}_i)}} \tag{2}$$

$\hat{q}$ means the number of sentences with negative or positive semantic orientation. The aim of giving less impact to neutral sentences is that they have less emotional impact. After calculating all r_{di} of documents they will normalized and brought into a [0,1] range, where the maximum negativity is represented by 0 and the maximum positivity is represented by 1.

4. APPLICATION

The application of the study was evaluated by following the flow chart in Figure 1. The evaluation was carried out using the open source Rapidminer application.

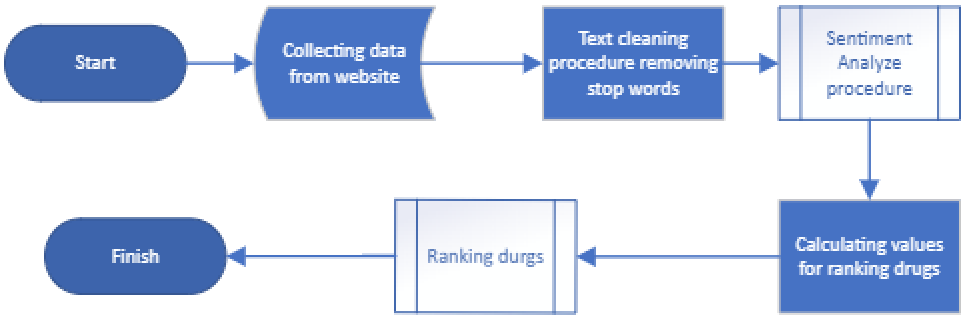

Figure 1. Flowchart for the Evaluation of Alzheimer's Drugs

Figure 1 shows the input-output flowchart of the algorithm for the evaluation of Alzheimer's drugs. In this direction, the data were taken from the website and cleaned from unnecessary content (preposition, conjunction, emoji, etc.). Then, drug reviews were analyzed with the sentiment module. Finally, the drugs were ranked with the module outputs.

4.1. RapidMiner

RapidMiner, stands as the pre-eminent open source framework for the extraction of information. This framework can be accessed as an autonomous tool for comprehensive data analysis, or as an engine for data mining, seamlessly integrated into its own suite of products. RapidMiner operates on a client/server model, with the server component offered through cloud infrastructures as software as a service. The capabilities of RapidMiner encompass an array of data mining and machine learning procedures, which include data loading and transformation, data preprocessing and visualization, as well as predictive analytics and statistical modeling. Furthermore, RapidMiner is constructed using the Java programming language. To facilitate the design and execution of analytical workflows, RapidMiner equips users with a graphical user interface. These workflows, known as “Processes” within the RapidMiner framework, consist of a series of interconnected “Operators”. Each Operator undertakes a specific task within the process, with the output of one Operator serving as the input for the next. Alternatively, the engine of RapidMiner can be invoked from other software applications or utilized as an application programming interface. Moreover, individual functions can be accessed through the command line interface. To supplement its existing repertoire, RapidMiner offers additional learning schemes, models, and algorithms from Python, Weka, and R scripts, which can be accessed through extensions (RapidMiner, 2023).

RapidMiner also offers the following features: Open source and autonomy from operating systems. Fascinating possibilities for plotting in high dimensions, combined with the concept of a multi-layered data view, contribute to the efficiency of data processing. It covers the functionalities of the WEKA data mining tool. It accesses information from databases such as Excel, Access, Oracle, IBM Db2, Microsoft SQL, Sybase and so on. It helps to establish operator chains for complex tasks (Gupta and Malhotra, 2015).

RapidMiner version 10.1, equipped with the Analyze sentiment feature, which includes the Aylien text analysis and Text processing extensions, was employed for the execution of the study. This software encompasses a variety of operators that are applicable to diverse types of data, including textual documents. It encompasses two distinct perspectives, namely "design" and "result". In the design perspective, users have the capability to visually construct a model by means of dragging and dropping operators onto the workspace and connecting them. Moreover, it provides users with the ability to execute their own code utilizing a specific extension. On the other hand, the result perspective offers advanced approaches to examine outcomes through the utilization of graphs and statistical tables containing pertinent information. Furthermore, RapidMiner is among the collection of big data analysis tools that are accessible (Alsaqer and Sasi, 2017). Predefined operations within the module are readily available in the application's operators. Consequently, the application employed in the study can be analyzed without necessitating the composition of code, in contrast to coding environments such as Python and R. The sentiment analysis module that was employed in the study is presented in Figure 2.

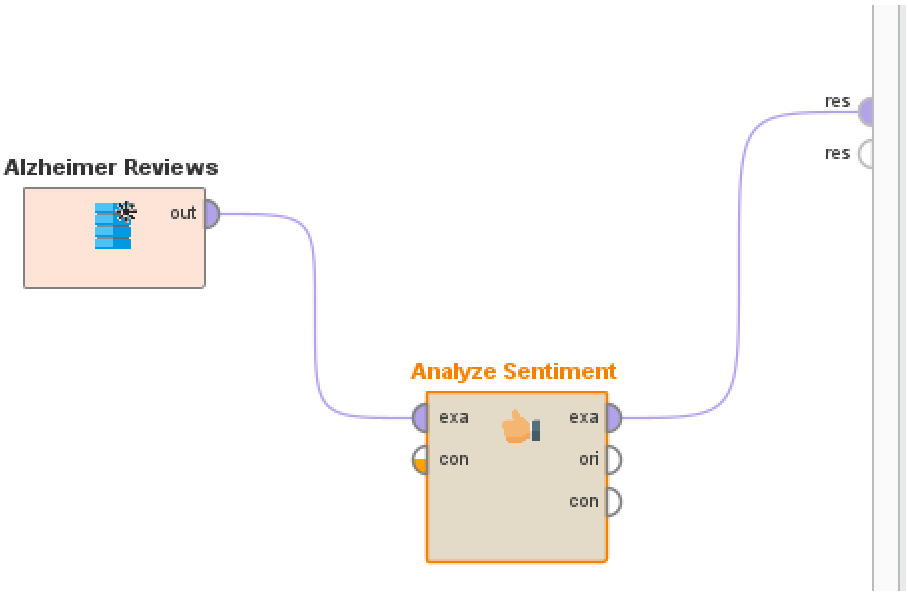

Figure 2. Example of Analyze Sentiment Module

Figure 2 shows the "Analyse sentiment" module for the evaluation of Alzheimer comments. Comments are uploaded as input to the module and analyzed by a word-based algorithm according to the weights of predetermined words. The results are defined as polarity (positive, negative and neutral).

4.2. Drug Comments Sentiment Analysis

In this section, four generic groups of Alzheimer's drugs were evaluated by Sentiment analysis. Generics groups are coded as Drug1, Drug2, Drug3 and Drug4. The drug reviews were accessed on 15.05.2023 by Drugs (Drug, 2023). The Drugs site was referred to as Global 'the most popular, most comprehensive and up-to-date drug information source online' in the Healthcare Global (Healtcareglobal, 2023). The site is ranked fourth among the most popular websites in the field of health. It is also mentioned that Drugs, an independent portal on the same site, provides information and data on 24,000 prescription medicines, including both professional health workers and consumers. In addition, on Similarweb, which gives information about web traffic, the Drugs website ranks 2060 globally and ranks 586 in the United States. Also, in the Health> Pharmacy category Drugs website is the second most popular website in the USA (Similarweb, 2023).

Text analysis was performed with the “analyze sentiment” operator using the RapidMiner package program. An example comment of Drug3 from Drugs is given in Table I. There are 160 different drug reviews at Drugs. These drug reviews may be found helpful by others. These 160 comments have a total of 1183 different useful opinions. The total number of useful finds (uf) for 160 drugs allows us to evaluate the opinions of 11183 people, although there are 160 comments.

Table 1. Alzheimer's Disease Drug Review

N	Drug Name	Review	uf
140	Drug4	"Besides nausea this has improved my mothers life 500 percent. Overall improvement wonderful."	16

In Table 1, the sample drug comment for Drug4 and the number of useful findings (uf) ‘16’ is given. The number of useful finds shows how many users find this comment useful. Drug names (Dn) are numerically encoded (mc) and the results of the analysis of the data and also the measure of the severity of the polarity (p+ and p-), drug codes (mc), and the number of useful finds (uf) are also shown in Table 2.

Table 2. Sentiment Analysis Results

N	Dn	mc	polarity	p-	p+	uf	N	Dn	mc	polarity	p-	p+	uf	N	Dn	mc	polarity	p-	p+	Uf
1	Drug2	2	negative	3,5	0	228	**55**	Drug2	2	positive	0,3	2,4	102	**109**	Drug3	3	positive	0	0,4	62
2	Drug2	2	positive	2,8	5	384	**56**	Drug2	2	negative	5,8	5,6	66	**110**	Drug3	3	positive	1,1	0	22
3	Drug2	2	negative	0,5	0	313	**57**	Drug2	2	positive	0,7	1,1	93	**111**	Drug3	3	positive	0	0,8	7
4	Drug2	2	positive	0	1,3	285	**58**	Drug2	2	positive	0,8	1	101	**112**	Drug3	3	positive	0	0,4	12
5	Drug2	2	negative	1	0,7	127	**59**	Drug2	2	positive	0,3	1,8	81	**113**	Drug3	3	positive	0	0	0
6	Drug2	2	negative	2,8	0,5	119	**60**	Drug2	2	negative	0,9	0,5	158	**114**	Drug4	4	positive	2,1	3,9	79
7	Drug2	2	positive	0,2	1,5	212	**61**	Drug2	2	positive	0	1,7	137	**115**	Drug4	4	positive	3,1	2,9	71
8	Drug2	2	positive	1,1	2,3	146	**62**	Drug2	2	positive	0,6	1,5	52	**116**	Drug4	4	positive	0,3	0,7	51
9	Drug2	2	negative	3,4	3,4	196	**63**	Drug2	2	negative	3,6	0	80	**117**	Drug4	4	positive	2,1	2,5	37
10	Drug2	2	positive	0,5	1,9	160	**64**	Drug2	2	negative	2,3	1,8	41	**118**	Drug4	4	positive	1,2	0,3	41
11	Drug2	2	positive	0,9	2,2	71	**65**	Drug2	2	positive	0,6	1,5	75	**119**	Drug4	4	positive	1,9	0	19
12	Drug2	2	positive	0,3	3	66	**66**	Drug2	2	positive	0	2,7	51	**120**	Drug4	4	positive	0,9	0	35
13	Drug2	2	positive	0,2	0,5	185	**67**	Drug2	2	positive	2,9	3,4	55	**121**	Drug4	4	positive	0,4	0,2	23
14	Drug2	2	positive	1,1	2,1	160	**68**	Drug2	2	positive	1,5	2	42	**122**	Drug4	4	positive	0	0,4	35
15	Drug2	2	positive	0,3	2,4	102	**69**	Drug2	2	negative	2,4	0	58	**123**	Drug4	4	positive	0	3,2	12
16	Drug2	2	positive	0,7	1,1	93	**70**	Drug2	2	negative	0,6	0,5	94	**124**	Drug4	4	positive	1,6	0,8	20
17	Drug2	2	positive	0,8	1	101	**71**	Drug2	2	positive	0,6	0,9	41	**125**	Drug4	4	positive	0,3	0,6	14
18	Drug2	2	positive	0,3	1,8	81	**72**	Drug2	2	negative	1,6	0,6	45	**126**	Drug4	4	positive	3,7	3,9	13
19	Drug2	2	negative	0,9	0,5	158	**73**	Drug2	2	negative	2	1,1	62	**127**	Drug4	4	positive	3,2	3	15
20	Drug2	2	positive	0	1,7	137	**74**	Drug2	2	negative	1,8	0	32	**128**	Drug4	4	positive	0,8	1,5	10
21	Drug2	2	positive	0,6	1,5	52	**75**	Drug2	2	negative	2,3	0,8	5	**129**	Drug4	4	positive	4,6	2,1	5
22	Drug2	2	positive	0	2,7	51	**76**	Drug2	2	negative	3,6	0,5	30	**130**	Drug4	4	positive	0,9	0,3	8
23	Drug2	2	negative	0,6	0,5	94	**77**	Drug2	2	negative	2,2	1,2	26	**131**	Drug4	4	positive	0,3	1,6	2
24	Drug2	2	positive	0,6	0,9	41	**78**	Drug2	2	negative	2,6	0,9	30	**132**	Drug4	4	positive	2,6	0	14
25	Drug2	2	negative	2,3	0,8	5	**79**	Drug2	2	negative	5,8	1,5	3	**133**	Drug4	4	positive	0,6	1,1	1
26	Drug2	2	negative	5,8	1,5	3	**80**	Drug2	2	negative	0,7	0,4	29	**134**	Drug4	4	positive	2,9	1,7	0
27	Drug2	2	positive	0,5	0,9	161	**81**	Drug2	2	negative	3,4	0,5	3	**135**	Drug4	4	positive	1,1	1	102
28	Drug2	2	positive	0,4	0,7	147	**82**	Drug2	2	positive	0	1,3	23	**136**	Drug4	4	positive	0,5	0,5	63
29	Drug2	2	positive	0	0,4	50	**83**	Drug2	2	negative	1,3	0,5	16	**137**	Drug4	4	positive	1	0	29
30	Drug2	2	positive	0,4	0,7	73	**84**	Drug2	2	negative	1,8	1,4	17	**138**	Drug4	4	positive	0	0	46
31	Drug2	2	negative	1	0	74	**85**	Drug2	2	positive	0,3	3,6	1	**139**	Drug4	4	positive	1,5	0	16
32	Drug2	2	neutral	0	0	55	**86**	Drug2	2	negative	1,9	0	3	**140**	Drug4	4	positive	0	1,7	16
33	Drug2	2	positive	0	1,6	3	**87**	Drug2	2	negative	1,3	0,8	1	**141**	Drug4	4	positive	0,7	0,3	8
34	Drug2	2	positive	0,9	1,4	232	**88**	Drug2	2	positive	0,4	2	0	**142**	Drug4	4	positive	0	1,5	1

35	Drug2	2	negative	3,5	0	229	**89**	Drug4	4	positive	2,1	3,9	79	**143**	Drug4	4	positive	0	0,1	1
36	Drug2	2	positive	2,8	5	384	**90**	Drug4	4	negative	3,1	2,9	71	**144**	Drug2	2	positive	0,9	0,1	40
37	Drug2	2	negative	2,2	0,6	97	**91**	Drug4	4	positive	0,3	0,7	51	**145**	Drug2	2	positive	0,5	0,3	72
38	Drug2	2	positive	1	2,2	122	**92**	Drug4	4	positive	2,1	2,5	37	**146**	Drug1	1	positive	1,1	1,1	21
39	Drug2	2	negative	0,5	0	313	**93**	Drug4	4	negative	1,2	0,3	41	**147**	Drug1	1	positive	0,9	1,7	11
40	Drug2	2	positive	0	1,3	285	**94**	Drug4	4	negative	0,9	0	35	**148**	Drug1	1	positive	0	2,1	10
41	Drug2	2	negative	1	0,7	127	**95**	Drug4	4	negative	0,4	0,2	23	**149**	Drug1	1	positive	1,1	1,1	3
42	Drug2	2	negative	2,8	0,5	119	**96**	Drug4	4	positive	0	0,4	35	**150**	Drug1	1	positive	0,7	0	21
43	Drug2	2	positive	0,8	2,1	134	**97**	Drug4	4	negative	1,1	1	102	**151**	Drug1	1	positive	0	0,9	4
44	Drug2	2	positive	0,2	1,5	212	**98**	Drug4	4	negative	0,5	0,5	63	**152**	Drug1	1	positive	1,1	1,1	21
45	Drug2	2	positive	1,1	2,3	146	**99**	Drug4	4	negative	1	0	29	**153**	Drug1	1	positive	0,9	1,7	11
46	Drug2	2	positive	1,8	3,6	93	**100**	Drug4	4	neutral	0	0	46	**154**	Drug1	1	positive	0	2,1	10
47	Drug2	2	negative	3,4	3,4	196	**101**	Drug3	3	positive	0	1,9	26	**155**	Drug1	1	positive	1,1	1,1	3
48	Drug2	2	positive	0,5	1,9	160	**102**	Drug3	3	positive	0,3	2,9	20	**156**	Drug1	1	positive	0,7	0	21
49	Drug2	2	positive	0,3	2,5	52	**103**	Drug3	3	positive	0	1,4	31	**157**	Drug1	1	positive	0	0,9	4
50	Drug2	2	positive	0,9	2,2	71	**104**	Drug3	3	positive	0	2,3	26	**158**	Drug1	1	positive	1	0	0
51	Drug2	2	positive	0,3	3	66	**105**	Drug3	3	positive	0,9	2,3	12	**159**	Drug3	3	positive	0	0,4	62
52	Drug2	2	positive	0,2	0,5	185	**106**	Drug3	3	positive	0,8	0,8	9	**160**	Drug3	3	positive	1,1	0	22
53	Drug2	2	positive	1,1	2,1	160	**107**	Drug3	3	positive	0,3	1,2	5							
54	Drug2	2	negative	2,9	0,4	41	**108**	Drug3	3	positive	0,2	0,5	6							

Table 2 shows the textual contradictions (positive, negative and neutral) of the polarity values for the 160 drug reviews. The Table also shows the numerical values generated from textual expressions. Numerical transformation was performed as follows: p+ and p- were used to quantify the polarity of the drug.

4.3. Ranking Drugs

The user values for the drugs to be calculated and used in the sequence are expressed as V_j. The number n represents the number of comments, and m represents the count of drug types in the data set. With the numerical values obtained by the Sentiment analysis, V_j is calculated as follows. For the drug coded as j; the total calculated values of the drug TW_j, in the data set are divided by the total number of drugs (TC_j) for each drug.

This calculation is shown by Equation (3).

$$V_j = \frac{TW_j}{TC_j} \tag{3}$$

TW_j value is calculated with Equation (4) using p^+, p^-, uf, and mc_i.

$$TW_j = \sum_{i=1}^{n} \begin{cases} uf_i * (p_i^+ * + * p_i^- * -1), & mc_i = j \\ 0 & , mc_i \neq j \end{cases},$$

$$i=1, 2, \ldots, n, j=1, 2, \ldots, m \tag{4}$$

TC_j value is calculated with Equation (5) using mc_i.

$$TC_j = \sum_{i=1}^{n} \begin{cases} 1, & mc_i = j \\ 0, & mc_i \neq j \end{cases}, i=1, 2, \ldots, n, j=1,2,\ldots,m \tag{5}$$

Table 3. Drug Rankings

Drug Name	TW_j	TC_j	V_j	Ranking
Drug1	38	13	2,923	3
Drug2	3342	90	37,133	1
Drug3	238	15	15,866	2
Drug4	78,92	42	1,879	4

Table 3 shows that when the drugs are ranked according to the calculated V_j, the drugs are from the highest to the lowest as follows “Drug2”, “Drug3”, “Drug1” and “Drug4”.

According to Table 3, it is seen that Drug2 is the most used drug according to the number of comments and the number of user comments found useful. It can be inferred from this that this drug is given to patients more by healthcare professionals. At the same time, it means that the drug that is found to be the most useful drug with comments is given most frequently by doctors. This can be said that healthcare professionals choose the right drug in patient treatment. Table 3 also shows that according to user experience, Drug2 is the most useful drug in Alzheimer's disease. This table reflects the opinions of users or relatives of users as opposed to the opinions of health professionals. When using medication for any disease, the advice of professional healthcare professionals should be followed. However, the results of this table are useful for healthcare professionals to have an idea of the users' experiences in advance.

5. CONCLUSION

Nowadays, it is very important in terms of both time and cost to produce meaningful results from many texts, such as business reports, comments on social media, and printed documents. Text mining is used in many areas, such as health, communication, astronomy, and production.

In this study, reviews of Alzheimer's drugs were evaluated with Sentiment analysis, and the drugs were ranked. Although the number of comments made for each drug was different, a total of 160 reviews and 11183 helpful comments were analyzed in the study including the number of useful comments in the analysis means adding more people to the data set. The polarity (positive, negative, and neutral) values of the reviews were calculated with the Sentiment analysis. Total values were obtained by multiplying the number of useful comments by these values. Average values were obtained by dividing the total values by the number of drugs, and the drugs were ranked.

Considering that most of the information is stored as a text document, it increases the importance of text mining. In many other studies, as in this study, text mining can be used to determine whether textual information is meaningful in a short time. It may also help to decide the decision-making units according to the positive, negative, and neutral status of the text. The examined texts can be separated and/or classified according to the polarity value.

Sentiment analysis can be used to rank other drugs such as Alzheimer's drugs. The study may provide guidance for patients and their relatives who will be using the drug for the first time or who are not satisfied with the medication used. This study can also be found useful in terms of evaluating patient opinions from pharmaceutical companies. Many products can be evaluated similarly, such as the evaluation of other drugs by this method.

Since the comments received in the study on Alzheimer's drugs did not include demographic data, no evaluation was made. In future studies, patient data to be collected from a different website or hospitals and analyses on demographic factors can be made and compared with this study.

Medication reviews can be combined with the opinions of healthcare professionals to make more accurate decisions. In order to make similar drug evaluations, healthcare professionals can further expand their drug evaluations by adding categorical data according to patient age, gender or presence of other diseases. In this way, data and text mining techniques can be used to make more accurate decisions.

REFERENCES

Ajibade, S. S. M., Zaidi, A., Tapales, C. P., Ngo-Hoang, D. L., Ayaz, M., Dayupay, J. P., ... & Adediran, A. O. (2022, December). Data Mining Analysis of Online Drug Reviews. *In 2022 IEEE 10th Conference on Systems, Process & Control (ICSPC) (pp. 247-251).* IEEE.

Alsaqer, A. F., & Sasi, S. (2017, July). Movie Review Summarization and Sentiment Analysis Using Rapidminer. *In 2017 International Conference on Networks & Advances in Computational Technologies (NetACT) (329-335).* IEEE.

Bae, Y., & Lee, H. (2012). Sentiment Analysis of Twitter Audiences: Measuring the Positive or Negative Influence of Popular Twitterers. *Journal of the American Society for Information Science and technology*, 63(12), 2521-2535.

Balbi, S., Misuraca, M., & Scepi, G. (2018). Combining Different Evaluation Systems on Social Media for Measuring User Satisfaction. *Information Processing & Management*, 54(4), 674-685.

Bilisoly, R. (2011). *Practical Text Mining with Perl*. John Wiley & Sons.

Cai, T., Zhu, Y., & Liu, Y. (2023, April). Research on Sentiment Classification of MOOC User Comments Based on Machine Learning. *In 2023 8th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA) (152-156).* IEEE.

Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (Eds.). (2017). *A Practical Guide to Sentiment Analysis (Vol. 5).* Cham: Springer International Publishing.

Dragoni, M., Poria, S., & Cambria, E. (2018). OntoSenticNet: A Commonsense Ontology for Sentiment Analysis. *IEEE Intelligent Systems*, 33(3), 77-85.

Drug, www.drugs.com, (15.05.2023).

Duwairi, R. M., Marji, R., Sha'ban, N., & Rushaidat, S. (2014, April). Sentiment Analysis in Arabic Tweets. *In 2014 5th International Conference on Information and Communication Systems (ICICS) (1-6).* IEEE.

Flynn, O. A., Murugadass, A., & Xiao, L. (2021). Attracting Attention in Online Health Forums: Studies of r/Alzheimers and r/dementia. In Diversity, Divergence, Dialogue: *16th International Conference, iConference 2021, Beijing, China, March 17–31, 2021, Proceedings, Part I 16 (377-395).* Springer International Publishing.

Gupta, G. ve Malhotra, S. (2015). Text Document Tokenization for Word Frequency Count Using RapidMiner (Taking Resume as an Example). *International Journal of Computer Applications*, 975, 8887, 24-26.

He, W., Zha, S., & Li, L. (2013). Social Media Competitive Analysis and Text Mining: A case study in the Pizza İndustry. *International Journal of Information Management*, 33(3), 464-472.

Healthcare Global, www.healthcareglobal.com/top-10/top-10-healthcare-websites, (15.05.2023)

Hearst, M. (2003). *What is Text Mining*. SIMS, UC Berkeley.

Liu, B. (2022). *Sentiment Analysis and Opinion Mining*. Springer Nature.

Markopoulos, G., Mikros, G., Iliadi, A., & Liontos, M. (2015). Sentiment Analysis of Hotel Reviews in Greek: A Comparison of Unigram Features. *In Cultural Tourism in a Digital Era: First International Conference IACuDiT, Athens, 2014 (373-383).* Springer International Publishing.

Rapidminer, https://rapidminer.com/, Erişim Tarihi: 09.12.2023.

Rea, R., & Parsons, C. (2023). Readability, Accessibility, Quality, Visual Design and Content of Online Information on Dementia Medication: A Quantitative Evaluation. *International Journal of Pharmacy Practice.* 31(S1), i5-i6.

Saad, E., Din, S., Jamil, R., Rustam, F., Mehmood, A., Ashraf, I., & Choi, G. S. (2021). Determining the Efficiency of Drugs under Special Conditions from Users' Reviews on Healthcare Web Forums. *IEEE Access*, 9, 85721-85737.

Similarweb, https://www.similarweb.com/, (15.05.2023).

Shi, L., Sun, B., Kong, L., & Zhang, Y. (2009, October). Web forum Sentiment Analysis Based on Topics. *In 2009 Ninth IEEE International Conference on Computer and Information Technology (2, 148-153)*. IEEE.

Song, M., & Brook Wu, Y. F. (Eds.). (2008). *Handbook of Research on Text and Web Mining Technologies*. IGI Global.

Subrahmanian, V. S., & Reforgiato, D. (2008). AVA: Adjective-Verb-Adverb Combinations for Sentiment Analysis. *IEEE Intelligent Systems*, 23(4), 43-50.

Tan, A. H. (1999, April). Text Mining: The State of the Art and the Challenges. *In Proceedings of the Pakdd 1999 Workshop on Knowledge Disocovery from Advanced Databases*, 8, 65-70.

Tayel, S., Reif, M., & Dengel, A. (2013). Rule-based Complaint Detection Using Rapidminer. *In Conference: RCOMM*, 141-149.

Tripathi, P., Vishwakarma, S. K., & Lala, A. (2015, December). Sentiment Analysis of English Tweets using Rapidminer. *In 2015 International Conference On Computational Intelligence and Communication Networks (CICN) (668–672)*. IEEE.

Unnamalai, K. (2012). Sentiment Analysis of Products Using Web. *Procedia Engineering*, 38, 2257-2262.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. *In Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 347-354