

Analysis of the feelings in the reviews of patients versus the evaluation of the ease of use, effectiveness, and satisfaction of prescribed medications

Carlos Agudelo-Santos^{1,*} and Jose Isaac Zablah¹

1 Facultad de Ciencias Médicas, Universidad Nacional Autónoma de Honduras; carlos.agudelo@unah.edu.hn, jose.zablah@unah.edu.hn

** Correspondencia: carlos.agudelo@unah.edu.hn; Tel.: (504) 2239-4292
Available from: <http://dx.doi.org/10.21931/BJ/2024.01.01.24>*

ABSTRACT

The sentimental polarity of patients' medical treatments is decisive for therapeutic adherence, especially in managing chronic diseases. Patients value the effects of medications differently, while health personnel do it from a practical perspective. For this, patient reviews have been taken in unstructured text to the diversity of drugs available on the WebMD site. A numerical assessment accompanies these data on a Likert scale of the variables for "EaseofUse", "Effectiveness," and "Satisfaction". Using an NLP model called RoBERTa; the opinions have been analyzed, finding that neutral opinions are maintained against positive scales of "EaseofUse" and "Effectiveness," but negative opinions regarding "Satisfaction," where the evaluations are divided into the extremes. The analysis has been done statistically using frequencies and diagrams of pairs between feelings and variables of interest.

Keywords: Natural language processing; sentiment analysis; BERT; medication satisfaction

INTRODUCTION

In medical care, patient satisfaction plays a crucial role in assessing the quality of medical services and the effectiveness of drug treatments. Patients' perceptions and experiences about prescribed medications are determining factors in their adherence to treatment and achieving positive health outcomes. Obtaining information in this sense is complex since opinions often need to be captured in databases and remain only in the oral information that feeds back to the doctor in the clinic.¹⁻² Then, when this information is obtained through instruments, unstructured text analysis (ATNE) is an activity that consumes many human resources. In the case of using computers it is a demanding technical activity. In this context, sentiment analysis using natural language processing (NLP) emerges as a tool to understand better and evaluate patient satisfaction with medications.³⁻⁶

The exponential growth of social media, online forums, and product review platforms has generated a massive amount of unstructured data containing patient opinions and experiences about medications. This data represents a rich source of information that can be leveraged using natural language processing techniques to extract valuable insights into patient satisfaction. By understanding their opinions in a broader context, healthcare professionals and researchers can identify trends, spot unreported side effects, and make informed treatment decisions.⁷

Several techniques and models for natural language processing (NLP) have been developed regarding sentiment analysis, increasing accuracy and efficiency as computational capabilities to form and evolve neural networks increase. Pre-trained language models such as BERT, RoBERTa, and GPT-3 have been developed for these purposes. These models are trained on a massive text and code data set, allowing them to learn the relationships between words and phrases. This allows them to identify the nuances of human language that are important for sentiment analysis.⁸⁻¹⁰

Supervised machine learning techniques have been developed, which require training data to be labeled with the polarity of emotions (positive, negative, or neutral). This allows models to learn to identify the language features that are associated with each polarity and assign them a numerical rating that is between zero and one, with the values closest to unity being indicators of more incredible emotion; when comparing polarities, the largest of which is dominant.¹¹⁻¹²

Sentiment analysis models have an accuracy of up to 90% in classifying texts in terms of polarity. At the computational level, it is an activity that requires a lot of CPU processing capacity; because of this, it has been chosen to use a graphics processing unit (GPU). By combining this feature of precision and high performance with the multicore processing available in computers or, failing that, through clusters, large amounts of data can be reduced in less time.¹³

The challenge NLP models face is training data, as sentiment analysis requires large amounts of labeled data to identify the language characteristics associated with each polarity. Collecting this data can be costly and laborious. Additionally, the subjectivity of human language, especially when linked to emotions, makes this task very complex due to the different ways a person can express themselves, making it difficult to assess the polarity of the feelings to be evaluated.¹⁴

This article shows the use of an algorithm with this technology to transform and analyze data of opinions and testimonials of patients from a publicly accessible database and to compare these feelings with the numerical assessment on the Likert¹⁵ scale that has been made of the variables of effectiveness, ease of use, and satisfaction.

MATERIALS AND METHODS

The data for analyzing patients' feelings regarding prescription drugs has been obtained through the scrapping data science technique, using the API available from the WebMD site. A total of 362,806 opinions have been obtained, which is the totality of them until March 2020. This data is publicly available in the form of a dataset. This includes specific medication information, age, gender, numerical satisfaction rating, ease of use, effectiveness, and text related to patient feedback that reviews contain free-form, unstructured text. All data is written in English.¹⁶⁻¹⁷

The model used to process opinions is called RoBERTa, an acronym for "*A Robustly Optimized BERT Pre-training Approach*". It is a natural language processing (NLP) model that is based on the success and architecture of BERT (*Bidirectional Encoder Representations from Transformers*) but with significant improvements in its training and performance, this model was developed in 2018 by Google AI Languages and has proven to be a valuable tool in various sentiment analysis and word processing tasks.⁸⁻⁹

The usefulness of Roberta for sentiment analysis lies in its ability to understand and represent context, that is, to use the meaning of words in a text accurately. Through a more thorough training process and more data, RoBERTa achieves a deeper understanding of natural language and semantic relationships between words, as its authors have used more than 161GB of initial training data.

When applied to sentiment analysis tasks, RoBERTa can accurately identify emotional nuances and tones in text. You can distinguish between positive, neutral, and negative opinions and understand the intensity of the emotions expressed. Its most excellent utility is to evaluate customer satisfaction from online reviews, for researchers seeking to analyze public perception on specific topics, or for any application that requires the understanding and classifying of sentiment-based opinions in text.

RoBERTa's ability to grasp context and information implicit in language makes it an effective tool for processing complex tasks, such as detecting sarcasm, identifying mixed feelings, and understanding long texts. In addition, its robust training and the availability of pre-trained models in different languages make it a versatile tool for sentiment analysis applications worldwide. For this study, the pre-trained version of Roberta called Twitter-roBERTa-base for Sentiment Analysis updated to 2022 was used; this was trained with more

than 124 million Tweets between January 2018 and December 2021; this has been adjusted for sentiment analysis with the TweetEval reference benchmark.¹⁸⁻²²

The source code of the program used to do sentiment analysis is available in Table 1. NLP was developed using Python v3.10.3 programming language and the Notepad++ v7.9.3 text editor for encoding. Multiple publicly available libraries were used, including Matplotlib v3.7, NLTK v3.8.1, and NumPy v1.18.; Pandas v2.1, SciPy v1.10.1, Seaborn v0.12.2; Torch v2.0.1, TQDM v4.66.1 and Transformers 4.32.1, among others.²³⁻²⁵

```
#####
## nlp_roberta.py
#####
##Libraries import
from transformers import AutoModelForSequenceClassification
from transformers import AutoTokenizer
import pandas as PD
from scipy.special import softmax
import matplotlib.pyplot as plt
import seaborn as sns
from tqdm.notebook import tqdm
from alive_progress import alive_bar;

## Graph style
plt. style.use('ggplot')

## Functions
def polarity_scores_roberta(example):
    encoded_text = tokenizer(example, return_tensors='pt')
    output = model(**encoded_text)
    scores = output[0][0].detach().numpy()
    scores = softmax(scores)
    scores_dict = {
        'roberta_neg' : scores[0],
        'roberta_neu' : scores[1],
        'roberta_pos' : scores[2]}
    return scores_dict

## Input data frame (df) (input file is webmd.csv)
df = pd.read_csv('webmd.csv', delimiter=',', encoding="utf-8-sig")
df = df.reset_index().rename(columns={'index':'Id'})
df['Reviews'] = df['Reviews'].str.slice(start=0, stop=999)

## Write an excel file of data frame used for debugging
with pd.ExcelWriter('data frame.xlsx') as writer:
    df.to_excel(writer)

##Selecting model and variations
# Tasks:
# emoji, emotion, hate, irony, offensive, sentiment
# stance/abortion, stance/atheism, stance/climate, stance/feminist, stance/Hillary

task='sentiment'
MODEL = f'cardiffnlp/Twitter-roberta-base-{task}'

tokenizer = AutoTokenizer.from_pretrained(MODEL)
```

```

## PT (PyTorch)
model = AutoModelForSequenceClassification.from_pretrained(MODEL)
model.save_pretrained(MODEL)

## Array with analysis

res = {}
with alive_bar(len(df)) as bar:
    for i, row in tqdm(df.iterrows(), total=len(pdf)):
        try:
            text = row['Reviews']
            myid = i
            roberta_result = polarity_scores_roberta(text)
            res[myid] = {'**roberta_result'}
            bar()
        except RuntimeError:
            print (f'Error en ID {myid}')

## Transformation with analysis and original dataframe; joining both
results_df = pd.DataFrame(res).T
results_df = results_df.reset_index().rename(columns={'index':'Id'})
results_df = results_df.merge(df, how='left')

## Write excel file with analysis and dataframe
with pd.ExcelWriter('salida.xlsx') as writer:
    results_df.to_excel(writer)

#####
## Use seaborn to make plots comparing Effectiveness, EaseofUse and Satisfaction with NLP results
#####
sns.pairplot(data=results_df,
             vars=['roberta_neg', 'roberta_neu', 'roberta_pos'],
             hue='Effectiveness',
             palette='tab10')
plt.savefig("fig_effectiveness.png", dpi=1200)

sns.pairplot(data=results_df,
             vars=['roberta_neg', 'roberta_neu', 'roberta_pos'],
             hue='EaseofUse',
             palette='tab10')
plt.savefig("fig_easeofuse.png", dpi=1200)

sns.pairplot(data=results_df,
             vars=['roberta_neg', 'roberta_neu', 'roberta_pos'],
             hue='Satisfaction',
             palette='tab10')
plt.savefig("fig_satisfaction.png", dpi=1200)

```

Table 1. Source code of the Python program used to do NLP.

The developed code was executed on a computational instance deployed in the Akamai computational cloud; this had the CentOS Stream 9 operating system of 64bit, with 32 CPU cores, 192GB of RAM, 3840GB of

storage, and connected to the internet at a speed of 12Gbps. The dataset processing required seven and a half hours to complete NLP by delivering the file with the relevant scores and graphs.²⁶⁻²⁷

According to the emotion identified with RoBERTa, the polarity scores have been compared with the numerical value in the range of 1 to 5, where the higher value represents the maximum possible according to the experience of satisfaction, ease of use, and effectiveness perceived by the patient. These values were captured following the Likert scale; histograms represent their frequencies. Our objective is to determine if the sentiment analysis using NLP is by numerical assessments. Pairwise diagrams have been used for data analysis, as a fundamental technique in applied computer science in high-performance computing and medical data research, to compare numerical values of effectiveness, ease of use, and satisfaction regarding the sentiment scores resulting from NLP with RoBERTa.

RESULTS

With the numerical data of the perception of ease of use, effectiveness, and satisfaction, a distribution of frequencies was grouped according to the Likert scale developed. It highlights that most of the responses obtained a rating of 5 in perceptions. More specifically, in the case of the variable "Effectiveness," most patients rated 4 and 5, which indicates that the perception of effectiveness is primarily positive. Values 1 and 2 have a lower frequency, suggesting that a minority of respondents have a negative or neutral perception regarding this variable.

Regarding "EaseofUse," most opinions lean towards values 4 and 5, indicating a high level. However, it is essential to note that values 1 and 2 also have a significant portion of correct answers, suggesting that many patients believe that ease of use could be improved or is unsatisfactory.

In the variable "Satisfaction," the answers are concentrated in the values 4 and 5 of the scale. Values 1 and 2 also have a significant number of responses, which suggests that a considerable proportion of opinions express dissatisfaction, with the most significant number of them in this variable compared to the previously described variables. Graphically, all variables and their frequencies are implicit in Figure 1.

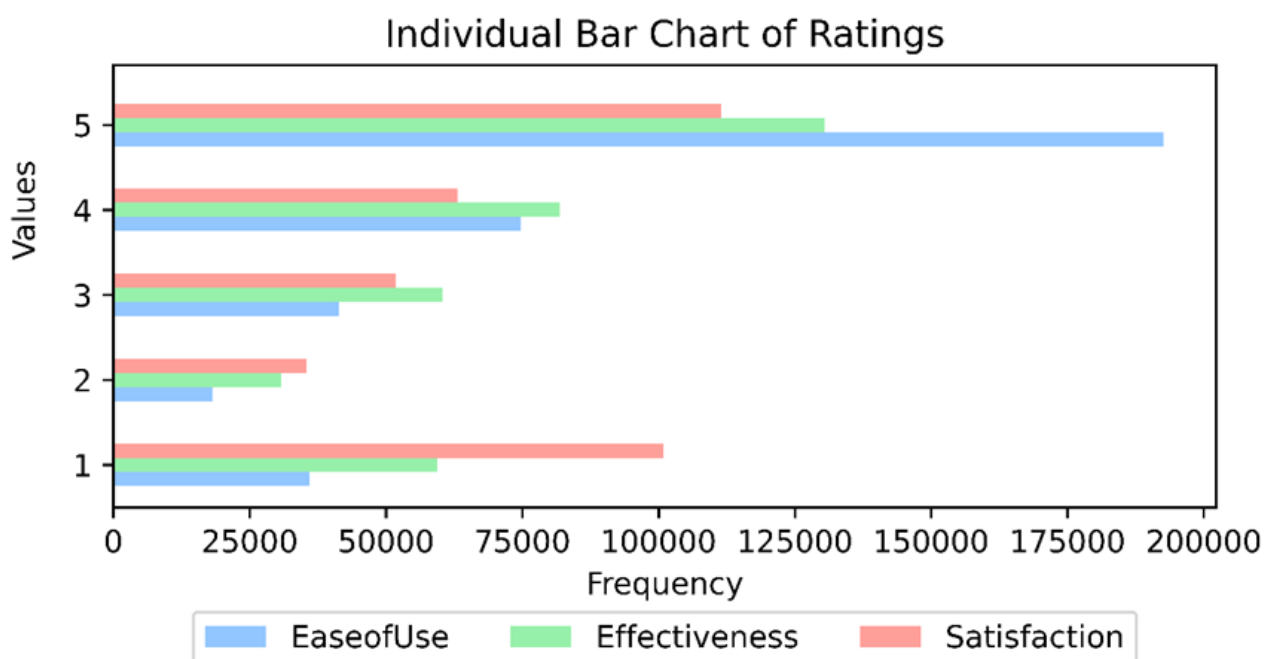


Figure 1. The frequency of the study variables was grouped according to the Likert scale.

For the variable "EaseofUse," all its significant scales (4 and 5) are grouped with the lowest scores of negative polarities; The other values 1,2, and 3, are concentrated with the highest values of negative polarity. The neutral polarity scores between 0 and 0.5 concentrate the most significant number of opinions with a scale 5 of "EaseofUse"; to a lesser extent, all the neutral score values have a homogeneous distribution of the other scale values. The values are grouped as the neutral score approaches zero. The positive polarity scores have a

lower frequency in scale values, concentrating all when the score approaches zero. Figure 2 has a pairwise diagram comparing polarities and numerical scale.

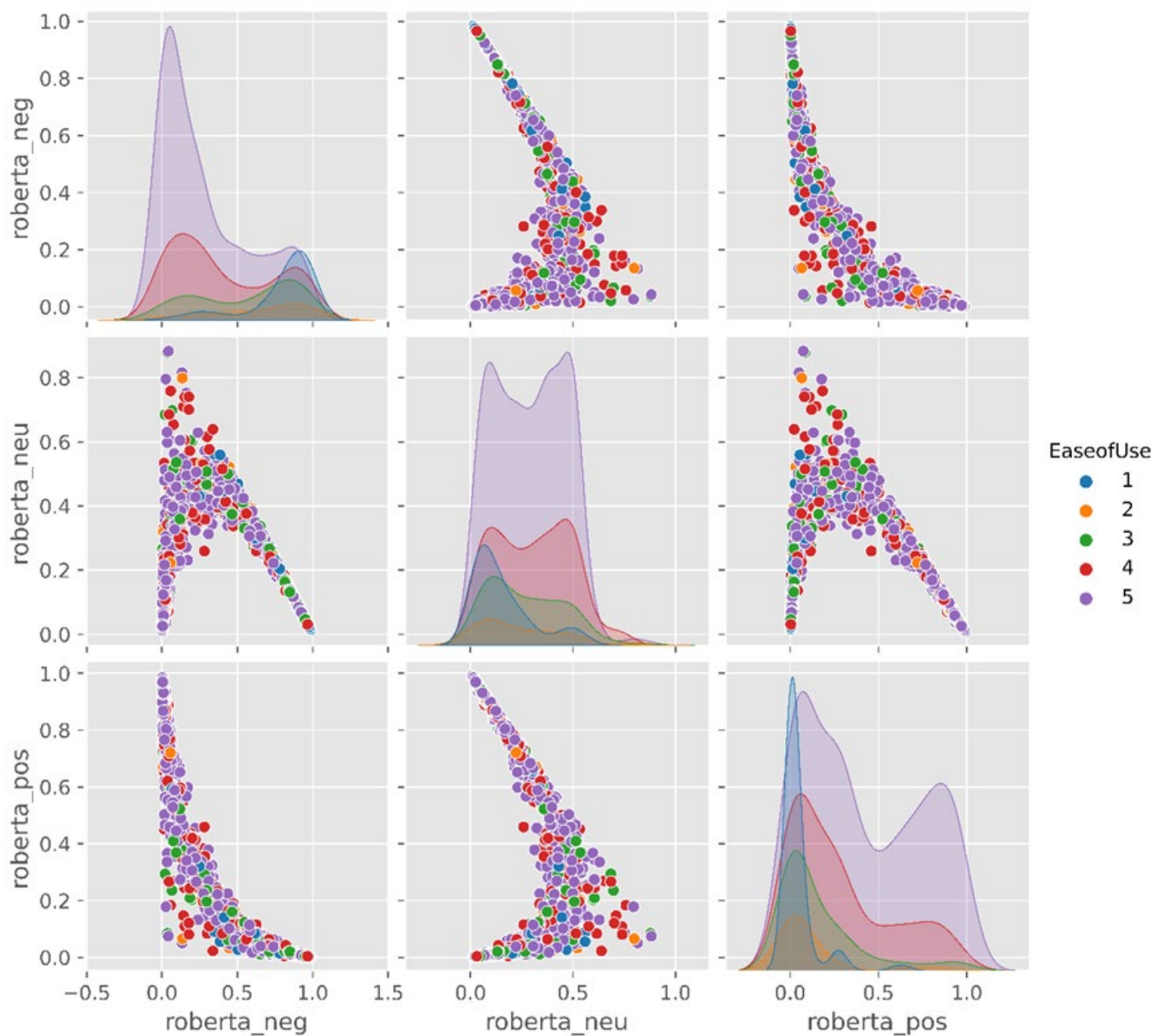


Figure 2. "Ease of Use" pairwise comparison between numerical scale and NLP polarity.

As for the "Effectiveness" variable, minor (1) and intermediate (3) scales are grouped with the highest negative polarity scores; the other values 2,4, and 5 are concentrated with the lowest negative polarity score values. The neutral polarity scores are close to zero and concentrate on the "Effectiveness" scale 1, while the other values are distributed along the score between 0 and 0.5 neutral polarity. The positive polarity scores have a lower frequency in scale values, the most significant being the values 1, 3, and 5 of the "Effectiveness" scale, which is associated with the zero positive polarity score. Figure 3 has a pairwise diagram comparing polarities and numerical scale.

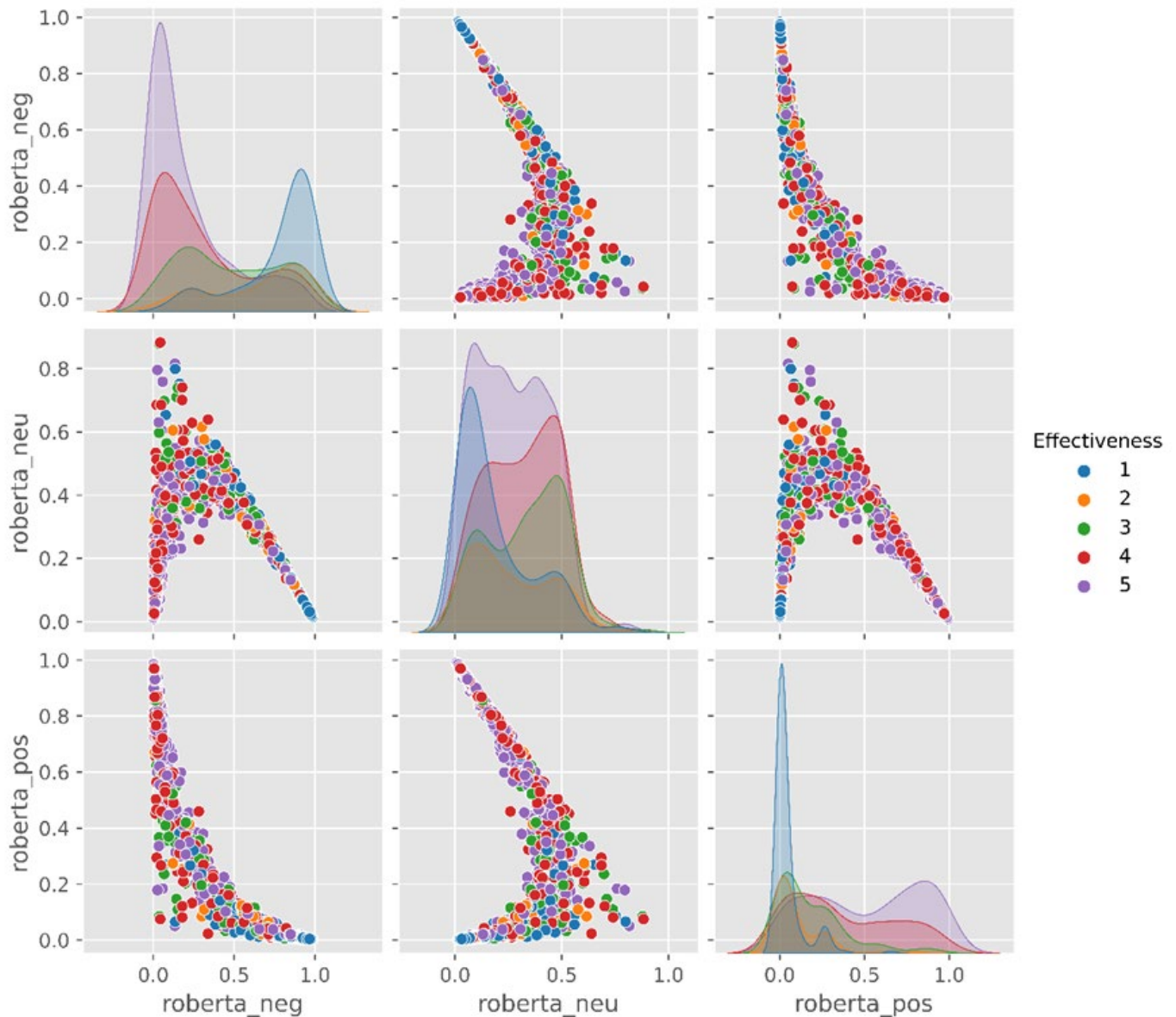


Figure 3. "Effectiveness" pairwise comparison between numerical scale and NLP polarity.

As for the variable "Satisfaction," its minor scale (1) is grouped with the highest scores of negative polarities; The other values 2,3,4, and 5, are concentrated with the lowest values of this score. When neutral polarity scores approach zero, they concentrate the "Satisfaction" scale 1, while the other values are distributed along the score in the range between 0 and 0.5 neutral polarity. The positive polarity scores have a lower frequency in scale values, the most significant being the value 1 of the "Satisfaction" scale associated with the zero positive polarity score. Figure 4 has a pairwise diagram comparing polarities and numerical scale.

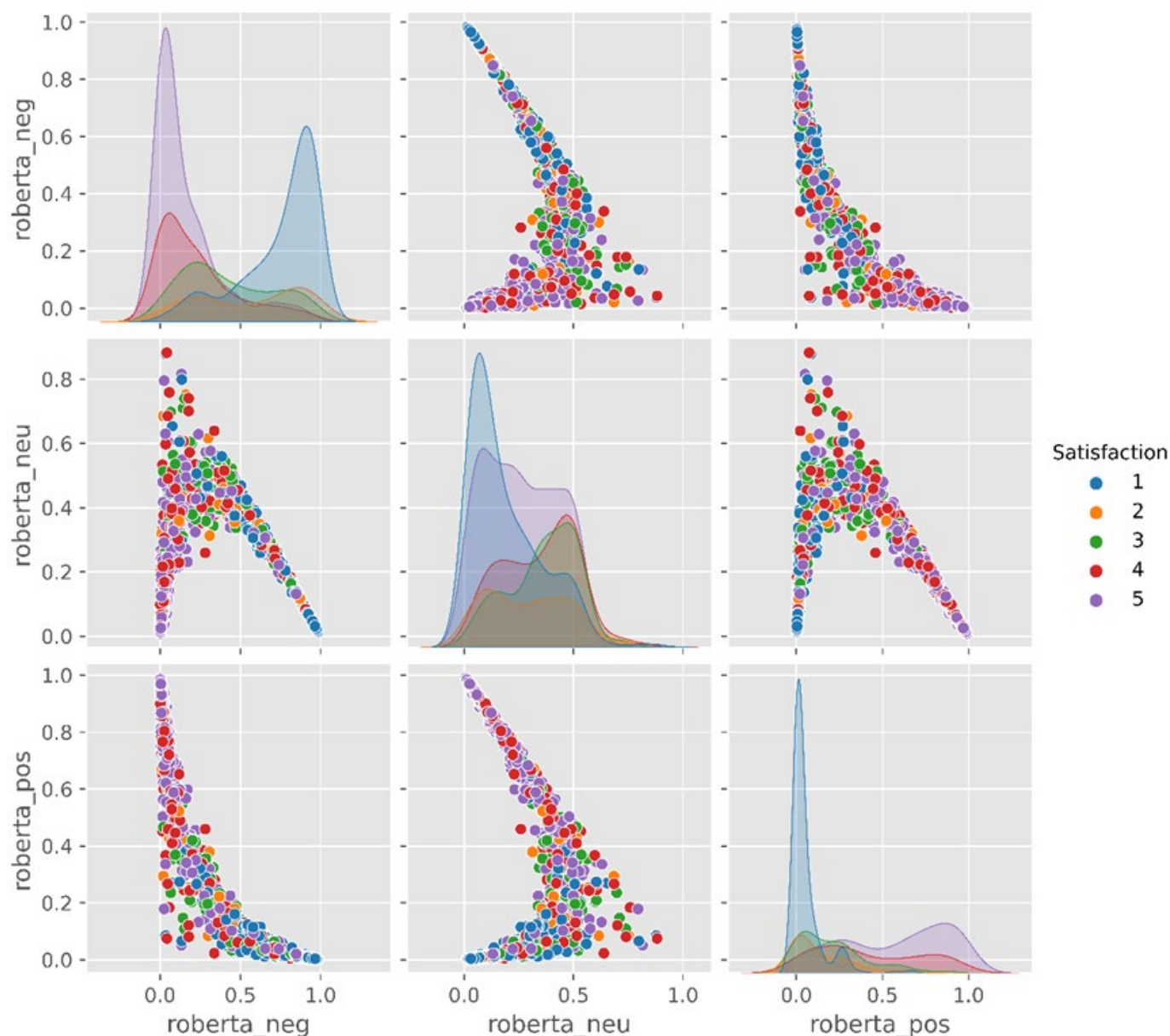


Figure 4. "Satisfaction" pairwise comparison between numerical scale and NLP polarity.

DISCUSSION

In the variable "EaseofUse," when the values 1, 2, and 3 coincide with a negative polarity, if the values are 4 and 5, they are associated with a neutral polarity. If the patient does not perceive the prescribed drugs that meet the criteria of convenience for him, the polarity is harmful and is directly associated with such perception. However, the drug is easy to use. In that case, it is not reflected in patients' opinions because, at present, the trend in medicines follows this pattern of behavior, so the consumer could be associating it with something that he takes for granted.

In the variable "Effectiveness," the numerical values of the scale do not coincide with the polarity of the opinions, except when it is ineffective (1). When the "Effectiveness" ranges from 2 to 5, patient opinions tend to be of neutral polarity. This condition occurs when the patient lacks the relevant information about the medications prescribed. Although the central pathology may disappear, side effects or adverse effects will likely occur, which the patient did not expect; this is reflected by the highest number of neutral opinions across the "Effectiveness" scale.

In the variable "Satisfaction," the numerical values of the scale are according to the polarity of opinions obtained because it is very likely that the knowledge of the health specialist does not coincide with the

perceptions perceived by patients since they may have expectations much higher than the results obtained. This indicates a very different vision of the patient from that of the health personnel since the neutral and negative polarities on the drugs accumulate in this variable of "Satisfaction."

The behavior observed in the results is similar to that observed in similar works published in the medical sciences and other areas of knowledge. It can be noted that numerical and scalar values do not represent what can be obtained from open-ended responses made up of unstructured text. The aspect of feelings can change the way research data is interpreted, which is consistent with the data we have obtained.²⁸⁻²⁹

CONCLUSIONS

As for "EaseofUse," the ratings tend to be polar-neutral due to the international trend that drugs must follow a pattern that facilitates therapeutic adherence so the consumer could associate this with any prescription. In the variable "Effectiveness," the opinions are concentrated in the neutral polarity, which does not coincide with the frequencies of the responses. Since the numerical trend is concentrated in high effectiveness, this may occur because the drug was beneficial to alleviate the central pathology but could have side or adverse effects. The variable "Satisfaction" concluded that the drugs that originated the patients' opinions did not meet their expectations since the comments had a negative polarity towards them.

These results were obtained using a natural language processing (NLP) model known as Roberta, which has effectively found new data on the relationship between drugs and patient opinion. This opens new possibilities for healthcare workers to fine-tune their interactions with patients. These tools can boost the development of precision and personalized medicine.

With the advent of immersive technologies, ubiquitous computing, and more excellent connectivity through multiple hyper-converged technologies, BigData will be generated, requiring automated deep learning tools to manage it and make decisions in the medical field. The exploration and adoption of these technologies will represent a critical condition in future doctor-patient interactions. The use of NLP is promising in the medical field, and with multiple applications across the healthcare spectrum, this knowledge can be integrated into autonomous care teams operated by artificial intelligence.

One of the limiting conditions of these technologies is the requirement of high capacities of computational resources to perform the reduction, processing, and analysis of data. In the dataset used, which is of minimum size, several hours of processing were required to obtain a dimension that describes the behavior of the data. This should be considered in future studies that follow the development of this line of research.

Supplementary Materials: They have not been provided.

Author Contributions: This research received no external funding. C. A and JZ conceptualized the idea and conducted the study and statistical analyses. They have read, reviewed, and edited the manuscript.

Funding: Not applicable.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: WebMD <https://www.webmd.com/drugs/2/index>

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

1. Naidu A. Factors affecting patient satisfaction and healthcare quality. *Int J Health Care Qual Assur*, 2009, 22(4):366–81. DOI:10.1108/09526860910964834

2. Ferrand YB, Siemens J, Weathers D, Fredendall LD, Choi Y, Pirrallo RG, et al. Patient satisfaction with healthcare services A critical review. *Qual Manag J*, **2016**, 23(4):6–22. DOI: 10.1080/10686967.2016.11918486
3. King G, Lam P, Roberts ME. Computer-assisted keyword and document set discovery from unstructured text. *Am J Pol Sci*, **2017**, 61(4):971–88. DOI:10.1111/ajps.12291
4. Borodkin A, Lisin E, Strielkowski W. Data algorithms for processing and analysis of unstructured text documents. *Appl Math Sci*, **2014**, 8:1213–22. DOI:10.12988/ams.2014.4125
5. Locke S, Bashall A, Al-Adely S, Moore J, Wilson A, Kitchen GB. Natural language processing in medicine: A review. *Tren Anaesth Crit Care*, **2021**, 38:4–9. DOI:10.1016/j.tacc.2021.02.007
6. Harrison CJ, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction to natural language processing. *BMC Med Res Methodol*, **2021**, 21(1). DOI:10.1186/s12874-021-01347-1
7. Pandita R. Internet a change agent: An overview of internet penetration and growth across the world. *Int J Inf Dissem Technol*, **2017**, 7(2):83. DOI:10.5958/2249-5576.2017.00001.2
8. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pretraining of deep bidirectional Transformers for language understanding. *arXiv [cs.CL]*, **2018**. DOI: 10.48550/arXiv.1810.04805
9. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv [cs.CL]*, **2019**. DOI: 10.48550/arXiv.1907.11692
10. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language Models are Few-Shot Learners. *arXiv [cs.CL]*, **2020**. DOI: 10.48550/arXiv.2005.14165
11. Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: A survey. *Wiley Interdiscip Rev Data Min Knowl Discov*, **2018**, 8(4). DOI:10.1002/widm.1253
12. Feldman R. Techniques and applications for sentiment analysis. *Commun ACM*, **2013**, 56(4):82–9. DOI:10.1145/2436256.2436274
13. Liu Z, Li G, Cheng J. Hardware acceleration of fully quantized BERT for efficient natural language processing. In: *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE; 2021.
14. Leeson W, Resnick A, Alexander D, Rovers J. Natural language processing (NLP) in qualitative public health research: A proof of concept study. *Int J Qual Methods*, **2019**, 18:160940691988702. DOI:10.1177/1609406919887021
15. Albaum G. The Likert scale revisited. *J Mark Res Soc*, **1997**, 39(2):1–21. DOI:10.1177/147078539703900202
16. Krotov V, Johnson L, Silva L, Legality and ethics of web scraping. *Commun Assoc Inf Syst*, **2020**, 47:539–63. DOI:10.17705/1cais.04724
17. WebMD's A to Z drug database. Available online: <https://www.webmd.com/drugs/2/index> (accessed on 6 September 2023)
18. Loureiro D, Barbieri F, Neves L, Anke LE, Camacho-Collados J. TimeLMs: Diachronic language models from Twitter. *arXiv [cs.CL]*, **2022**. DOI:10.48550/ARXIV.2202.03829
19. Barbieri F, Anke LE, Camacho-Collados J. XLM-T: Multilingual language models in Twitter for sentiment analysis and beyond. *arXiv [cs.CL]*, **2021**. DOI: 10.48550/ARXIV.2104.12250
20. Barbieri F, Camacho-Collados J, Neves L, Espinosa-Anke L. TweetEval: Unified benchmark and comparative evaluation for tweet classification. *arXiv [cs.CL]*, **2020**. DOI: 10.48550/arXiv.2010.12421
21. Mohammad S, Bravo-Marquez F, Salameh M, Kiritchenko S. SemEval-2018 Task 1: Affect in Tweets. In: *Proceedings of The 12th International Workshop on Semantic Evaluation*. Stroudsburg, PA, USA: Association for Computational Linguistics; 2018.
22. Rosenthal S, Farra N, Nakov P. SemEval-2017 Task 4: Sentiment Analysis in Twitter. In: *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Stroudsburg, PA, USA: Association for Computational Linguistics; 2017.
23. Python. Available online: <https://www.python.org/> (accessed on 6 September 2023)
24. Notepad++. Available online: <https://notepad-plus-plus.org/> (accessed on 6 September 2023)
25. Pip 23.2.1. Available online: <https://pypi.org/project/pip/> (accessed on 6 September 2023)
26. Linode. Available online: <https://www.linode.com/> (accessed on 6 September 2023)
27. CentOS stream 9. Available online: <https://centos.org/stream9/> (accessed on 6 September 2023)
28. Rajput A. Natural language processing, sentiment analysis, and clinical analytics. In: *Innovation in Health Informatics*. Elsevier; 2020. p. 79–97.
29. Aattouchi I, Elmendili S, Elmendili F. Sentiment analysis of health care: Review. *E3S Web Conf*, **2021**, 319:01064. DOI:10.1051/e3sconf/202131901064

Received: October 9th 2023/ **Accepted:** January 15th 2024 / **Published:** 15 March 2024

Citation: Agudelo-Santos C. , Isaac Zablah J. Analysis of the feelings in the reviews of patients versus the evaluation of the ease of use, effectiveness, and satisfaction of prescribed medications. Bionatura Journal 2024; 1 (1) 1. <http://dx.doi.org/10.21931/BJ/2024.01.01.24>

Additional information Correspondence should be addressed to carlos.agudelo@unah.edu.hn

Peer review information. Bionatura thanks anonymous reviewer(s) for their contribution to the peer review of this work using <https://reviewerlocator.webofscience.com/>

All articles published by Bionatura Journal are made freely and permanently accessible online immediately upon publication, without subscription charges or registration barriers.

Publisher's Note: Bionatura stays neutral concerning jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2024 by the authors. They were submitted for possible open-access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).