Article

Not peer-reviewed version

Application of Machine Learning Methods in Classification of Text Data of Patient Reviews

Irina Kalabikhina *, Vadim Moshkin, Anton Vasilyevich Kolotusha, Maksim Kashin, German Klimenko, Zarina Kazbekova

Posted Date: 26 December 2023

doi: 10.20944/preprints202312.1933.v1

Keywords: machine learning; patient reviews; neural networks; online reviews; review classification; text reviews; quality of medical services; GRU architecture; LSTM; CNN doi:10.20944/preprints202312.1933.v1

Article

Application of Machine Learning Methods in Classification of Text Data of Patient Reviews

Irina Kalabikhina [1,*], Vadim Moshkin [2], Anton Kolotusha [1], Maksim Kashin [2], German Klimenko [1] and Zarina Kazbekova [1]

[1] Department of Population, Faculty of Economics, Lomonosov Moscow State University. Leninskije Gory, GSP-1, Moscow 119991, Russia; tony_kol@mail.ru (A.K.); german89000@mail.ru (G.K.), kazbekova.zarina@bk.ru (Z.K.)
[2] Department of Information Systems, Ulyanovsk State Technical University, Severny Venets str., 32, Ulyanovsk 432027, Russia; v.moshkin@ulstu.ru (V.M.); kashin20031984@mail.ru (M.K.)
* Correspondence: ikalabikhina@yandex.ru

Abstract: The paper aims to develop and test an algorithm for classifying Russian-language text reviews of patients' experiences with medical facilities and physicians, extracted from social media. This is motivated by the limitations of conventional methods of surveying consumers to assess their satisfaction with the quality of services, which are being replaced by automatic processing of text data from social media. This approach enables to get more objective results due to the increased representativeness and independence of the sample of service consumers. The authors have tested machine learning methods using various neural network architectures. A hybrid method was developed to classify text reviews of medical facilities posted by patients on the two most popular physician review websites in Russia. Overall, more than 60,000 reviews were analysed. The main results are as follows: 1) the classification algorithm developed by the authors has a high efficiency, the best result being achieved by the GRU-based architecture (val_accuracy = 0.9271); 2) applying the named entity search method to text messages following their partitioning improved the classification efficiency for each of the classifiers based on artificial neural networks. To further enhance the classification quality, reviews need to be semantically partitioned by target and sentiment and the resulting fragments need to be analysed separately.

Keywords: machine learning; patient reviews; neural networks; online reviews; review classification; text reviews; quality of medical services; GRU architecture; LSTM; CNN

---

1. Introduction

The conventional method of direct questionnaire surveying to assess customer satisfaction is giving way to automatic processing of social media texts allowing to extract semantics. The latter objectively reveals consumer attitudes, as the sample becomes more representative and independent.

To implement this approach, it is essential to develop software classifiers that can group the text data by the following criteria:

- Sentiment;
- Target;
- Causal relationship, etc.

This approach requires that intelligent algorithms be developed for text classification allowing to perform objective analysis of text feedback, considering the genre and speech characteristics of social media texts, with due account of the specificities of the domain.

2. Analysing Medical Service Reviews as a Natural Language Text Classification Task

Online reviews and online ratings make up the so-called electronic word of mouth (eWOM), informal communications targeting consumers, channelled through Internet technologies (online reviews and online opinions) and relating to the use experiences or features of specific products or services or their vendors [1]. The rise of eWOM makes online reviews the most influential source of information for consumer decision-making [2–4].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 December 2023 doi:10.20944/preprints202312.1933.v1

The predominance of positive feedback is an incentive for businesses to raise prices for their goods or services to maximise their profits [5]. This prompts unscrupulous businesses to manipulate the customer reviews and ratings with cheating reviews, rigging the feedback scoring, etc. [6–8]. Some researchers have highlighted a separate threat posed by artificial intelligence, which can generate reviews that are all but indistinguishable from those posted by real users [9,10]. That said, artificial intelligence also serves as a tool to detect fake reviews [9].

eWOM has also become a widespread phenomenon in the healthcare sector: many physician rating websites (PRW) are available now. Physicians themselves are the first to use this opportunity by actively registering on such websites and filling out their profiles. For instance, in Germany, according to 2018 data, more than 52% of physicians had a personal page on online physician rating websites [11].

Online portals provide ratings not only of physicians, but also of larger entities, such as hospitals [12,13]. In many countries, however, a greater number of reviews concern hospitals or overall experiences rather than individual physicians and their actions [14,15].

Creating reviews containing reliable information enhances the efficiency of the healthcare sector by providing the patient, among other things, with trustworthy data and information on the quality of medical services. Despite the obvious benefits of PRWs, they have certain drawbacks, too, such as:

- Poor understanding and knowledge of healthcare on the part of the service consumers, which casts doubt on the accuracy of their assessments of the physician and medical services provided [16,17]. Patients often use indirect indicators unrelated to the quality of medical services as arguments (for example, their interpersonal experience with the physician [18,19]).
- Lack of clear criteria by which to assess a physician / medical service [18].

Researchers have found that online reviews of physicians often do not actually reflect the outcomes of medical services [20–22]. Consequently, reviews and online ratings in the healthcare industry are less useful and effective compared to other industries [23,24]. However, some studies, on the contrary, have revealed a direct correlation between online user ratings and clinical outcomes [25–28].

In general, the healthcare industry shows a high level of concentration of positive reviews and physician ratings [29–35]. However, at the beginning of the COVID-19 pandemic, the share of negative reviews on online forums prevailed [36].

The main factors behind a higher likelihood of a physician receiving a positive review are physician's friendliness and communication behaviour [37]. Shah et al. divide the factors that increase the likelihood of a physician receiving a positive review into factors depending on the medical facility (hospital environment, location, car park availability, medical protocol, etc.) and factors relating to the physician's actions (physician's knowledge, competence, attitude, etc.) [38].

Some researchers have noted that patients mostly rely on scoring alone while rarely using descriptive feedback when assessing physicians [39], which is due to the reduced time cost of completing such feedback [40]. At the same time, consumers note the importance of receiving descriptive feedback as it is more informative than numerical scores [41,42].

Physician's personal characteristics, such as gender, age, specialty, can also influence the patient's assessment and feedback, apart from objective factors [43–46]. For example, according to studies based on German and US physician assessment data, higher evaluations prevail among female physicians [43,44], obstetrician-gynaecologists [47], and younger physicians [47].

The characteristics of the patients have an impact on the distribution of scores. For example, according to a study by Emmert and Meier based on the online physician review data from Germany, older patients tend to score physicians higher than younger patients [43]. However, according to their estimates, doctors' assessments/scores don't depend on the respondent's gender [43]. Having an insurance policy has a significant influence on the feedback sentiment [43,46]. Individual studies have demonstrated that negative feedback is prevalent among patients from rural areas [48]. Some studies have focused on the characteristics of online service users, noting that there are certain characteristics of people that are indicative of the PRW use frequency [49,50]. Depending on this, users having different key characteristics will differ significantly in their ratings of importance of online physician reviews [51].

 doi:10.20944/preprints202312.1933.v1

A number of studies have used both rating scores and commentary texts as data [52]. In particular, the study [52] identified the factors influencing more positive ratings that would be related to both the physician's characteristics and other factors beyond the physician's control.

A number of studies use arrays of physician review texts as the data basis [53–57]. Researchers have found that physician assessment services can complement the information provided by more conventional patient experience surveys while contributing to a better understanding by patients of the quality of medical services provided by a physician or health care facility [58–60].

Social media analysis can be viewed as a multi-stage activity involving collection, comprehension, and presentation:

1. In the capture phase, relevant social media content will be extracted from various sources. Data collection can be done by an individual or third-party providers [61].
2. At the second phase, relevant data will be selected for predictive modelling of sentiment analysis.
3. At the third phase, important key findings of the analysis will be visualised [62].

Supervised or unsupervised methods can be used to effectively analyse the sentiments based on social media data. [63] gives an overview of these methods. The main approaches to classify the polarity of analysed texts are based on word, sentence, or paragraphs.

In [64], various text mining techniques for detecting different text patterns in a social network were studied. The text mining using classification based on various machine learning and ontology algorithms was considered, along with a hybrid approach. The experiments described in the above paper showed that there is no single algorithm that would perform best across all data types.

In [65], different classifier types were analysed for text classification and their respective pros and cons. Six different algorithms were considered:

- Bayesian classifier;
- decision tree;
- k-nearest neighbour algorithm (k-NN);
- support vector machine (SVM);
- artificial neural network based on multilayer perceptron;
- Rocchio algorithm.

A limited performance is a common drawback of all the above algorithms. Some of these are easy to implement, but their performance is poor. Some others perform well while requiring some extra time for training and parameter setting.

Lee et al. [66] classified trending topics on Twitter [now X] by using two approaches to topic classification, a well-known Bag-of-Words method for text classification and a network classification. They identified 18 classes and assigned trending topics to the respective categories. Ultimately, the network classifier significantly outperformed the text classifier.

[67] discusses methods addressing the challenges of short text classification based on streaming data in social networks.

[68] compared six feature weighting methods in the Thai document categorisation system. They found that the using SVM score thresholding with ltc yielded the best results for the Thai document categorisation.

[69] proposed a multidimensional text document classification framework. The paper reported that classifying text documents based on a multidimensional category model using multidimensional and hierarchical classifications was superior to the flat classification.

[70] compared four mechanisms to support divergent thinking using associative information obtained from Wikipedia. The authors used the Word Article Matrix (WAM) to compute the association function. This is a useful and effective method for supporting divergent thinking.

[71] proposed a new method to fine-tune a model trained on some known documents with richer contextual information. The authors used WAM to classify text and track keywords from social media to comprehend social events. WAM with a cosine similarity is an effective method of text classification.

As is clear from the review of the current state of automatic processing of unstructured social media data, there is no single approach at this time to achieve effective classification of text resources. The classification results will depend on the domain, representativeness of the training sample, and other factors. Therefore, it is important to develop and apply such intelligent methods for analysing reviews of medical services provided.

3. Classification Models for Text Reviews of the Quality of Medical Services in Social Media

In this study, we have developed a hybrid method of classification of text reviews from social media. This resulted in the classification of text reviews based on:

- text sentiment: positive or negative;
- target: a review of a medical facility or an individual physician.

Initially, the task was to break the multitude of reviews down into four classes. To solve this, we started with testing machine learning methods using various neural network architectures.

Mathematically, a neuron is a weighted adder whose single output is defined by its inputs and weight matrix as follows:

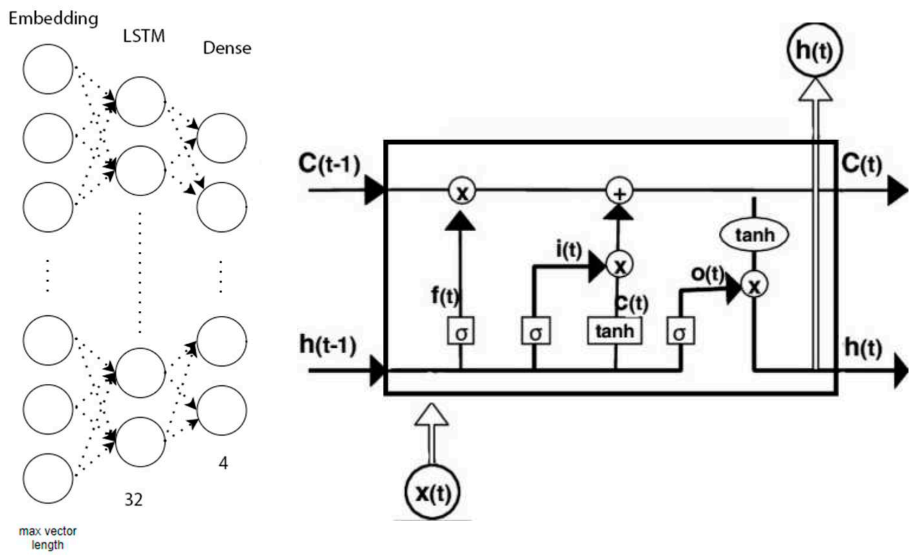
$$y = f(u), \qquad \text{where } u = \sum_{i=1}^{n} w_i x_i + w_0 x_0$$

where x_i and w_i are the neuron input signals and the input weights, respectively; the function u is the induced local field, and $f(u)$ is the transfer function. The signals at the neuron inputs are assumed to lie in the interval [0,1][0,1]. The additional input x_0 and its corresponding weight w_0 are used for neuron initialisation. Initialisation here means the shifting of the neuron's activation function along the horizontal axis.

There are a large number of algorithms available for context-sensitive and context-insensitive text classification. In this study, we propose three neural network architectures that have shown the best performance in non-binary text classification tasks. We compared the effectiveness of the proposed algorithms with the results of text classification using the models applied in our previous studies, which achieved good results in binary classification (BERT and SVM) [72,73].

3.1. LSTM Network

Figure 1 shows the general architecture of the LSTM network.



(**a**) (**b**)

Figure 1. LSTM-network: general architecture (a), LSTM-layer (b).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 December 2023 doi:10.20944/preprints202312.1933.v1

The proposed LSTM network architecture consists of the following layers:

- Embedding, the neural network input layer consisting of neurons:

$$Emb = \{Size(D), Size(S_{vec}), L_{Sec}\}$$

where *Size(D)* — dictionary size in text data;

$Size(S_{vec})$ – size of the vector space in which the words will be inserted;

$Size(S_{vec}) = 32$;

L_{Sec} – length of input sequences, equal to the maximum size of the vector generated during word pre-processing.

- LSTM Layer — recurrent layer of the neural network; includes 32 blocks.
- Dense Layer — output layer consisting of four neurons. Each neuron is responsible for an output class. The activation function is "Softmax".

3.2. A Recurrent Neural Network

Figure 2 shows the general architecture of a recurrent neural network.

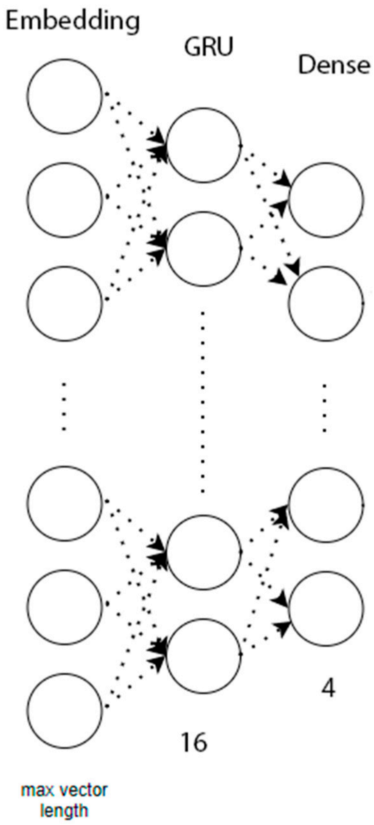


Figure 2. General architecture of recurrent neural network.

The proposed recurrent neural network architecture consists of the following layers:

- Embedding — input layer of the neural network.
- GRU — recurrent layer of the neural network; includes 16 blocks.
- Dense — output layer consisting of four neurons. The activation function is "Softmax".

3.3. A Convolutional Neural Network

Figure 3 shows the general architecture of a convolutional neural network (CNN).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 December 2023 doi:10.20944/preprints202312.1933.v1

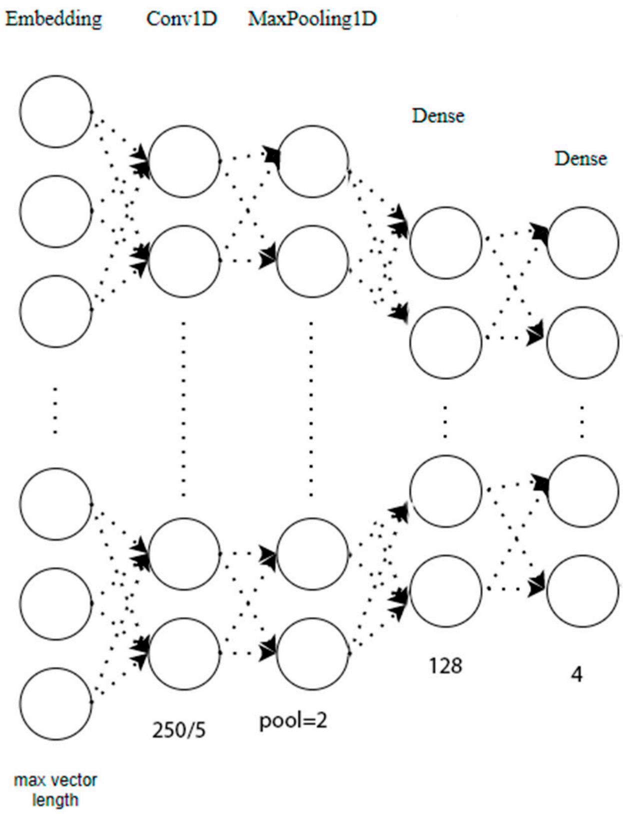


Figure 3. General architecture of a convolutional neural network.

The proposed convolutional neural network architecture consists of the following layers:

- Embedding — input layer of the neural network.
- Conv1D — convolutional layer required for deep learning. This layer improves the accuracy of text message classification by 5-7%. The activation function is "ReLU".
- MaxPooling1D — layer which performs dimensionality reduction of generated feature maps. The maximum pooling is equal to 2.
- Dense — first output layer consisting of 128 neurons. The activation function is "ReLU".
- Dense — final output layer consisting of four neurons. The activation function is "Softmax".

3.4. Using Linguistic Algorithms

We observed that some text reviews contained elements from different classes. To account for this, we added two new review classes: mixed positive and mixed negative.

To improve the classification quality, we used hybridisation of the most effective machine learning methods and linguistic methods that account for the speech and grammatical features of the text language.

Figure 4 shows the general algorithm of the hybrid method.

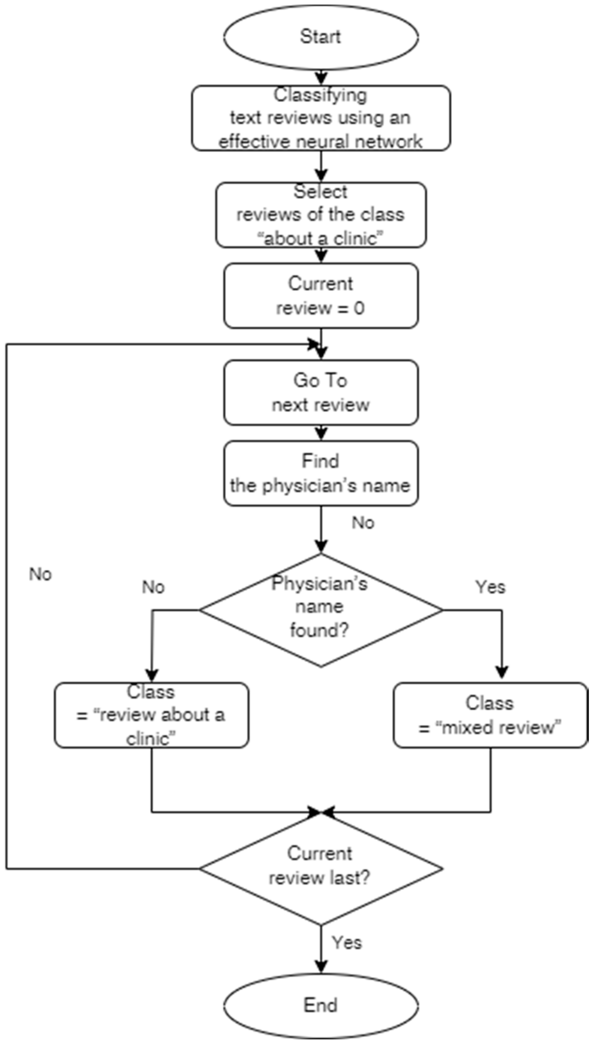


Figure 4. General flowchart of the hybrid classification algorithm.

A set of methods of pre-processing, validation and detection of named entities representing the physicians' names in the clinic was the linguistic component of the hybrid method we developed.

4. Software Implementation of a Text Classification System

The Natasha library in Python was used as a linguistic analysis module for texts in the natural Russian language. The library performs basic tasks of natural Russian language processing: token / sentence segmentation, morphological analysis and parsing, lemmatisation, extraction, normalisation, and named entities detection.

For this project, the library functionality was mainly used for searching and retrieving named entities.

The following libraries were also used to implement the processes of initialisation, neural network training, and evaluation of classification performance:

- Tensorflow, an open-source machine learning software library developed by Google for solving neural network construction and training problems.
- Keras, a deep-learning library that is a high-level API written in Python and capable of running on top of TensorFlow.
- Numpy, a Python library for working with multidimensional arrays.
- Pandas, a Python library that provides special data structures and operations for manipulating numerical tables and time series.

Google's Colab service was used to train the models, which provides significant computing power for data mining techniques.

5. Experimental Results of Text Review Classification

5.1. Using Dataset

A number of experiments were conducted on classifying Russian-language text reviews of medical services provided by clinics or individual physicians with a view to evaluating the effectiveness of the proposed approaches.

Feedback texts from the aggregators prodoctorov.ru and infodoctor.ru were used as input data.

All extracted data to be analysed had the following list of variables:

- city — city where the review was posted;
- text — feedback text;
- author_name — name of the feedback author;
- date — feedback date;
- day — feedback day;
- month — feedback month;
- year — feedback year;
- doctor_or_clinic — a binary variable (the review is of a physician OR a clinic);
- spec — medical specialty (for feedback on physicians);
- gender — feedback author's gender;
- id — feedback identification number.

The experiments were designed to impose a **90-word limit** on the feedback length.

5.2. Experimental Results on Classifying Text Reviews by Sentiment

In the first experiment, a database of 5,037 reviews from the website prodoctorov.ru with initial markups by the sentiment and target was built to test the sentiment analysis algorithms.

The RuBERT language model was used as a text data vectorisation algorithm. The Transformer model was used for binary classification of text into positive / negative categories. The training and test samples were split in an **80:20 ratio**.

The results of the classifier on the test sample were as follows: Precision = 0.9857, Recall = 0.8909, F1-score = 0.9359.

The classifier performance quality metrics confirm the feasibility of using this binary text sentiment classifier architecture for alternative sources of medical feedback data.

The LSTM classifier, the architecture of which is described in Section 3.1, was also tested on this sample. The reviews were supposed to be classified into four classes as follows:

- Positive review of a physician;
- Positive review of a clinic;
- Negative review of a physician;
- Negative review of a clinic.

The split between the training and test samples was also 80/20. Figure 5 gives the classification results.

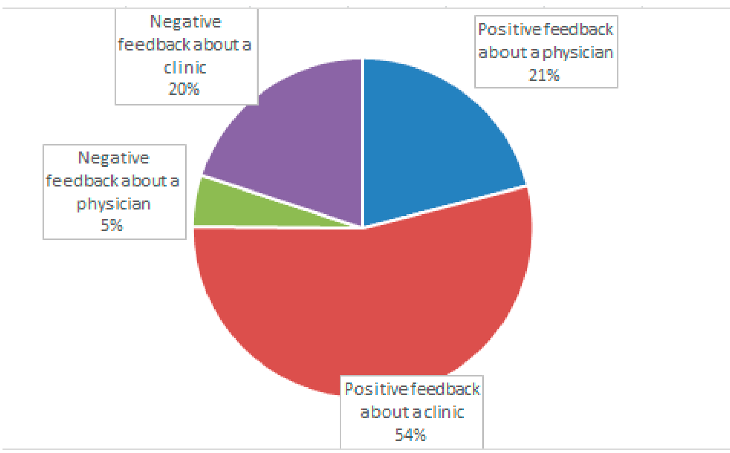


Figure 5. LSTM-network-based classification results for feedback posted on the website prodoctorov.ru.

5.3. A Text Feedback Classification Experiment Using Various Machine Learning Models

We used data from the online aggregator infodoctor.ru to classify Russian-language text feedback into four (later, six) classes using the machine learning models described in Section 3. This aggregator has a comparative advantage over other large Russian websites (prodoctorov.ru, docdoc.ru) in that it groups reviews by their ratings on the scale from one star to five stars with a breakdown by Russian cities, which greatly simplifies the data collection process.

We collected samples from Moscow, St. Petersburg and 14 other Russian cities with a million-plus population, for which we could obtain minimally representative samples by city (with a minimum 1,000 observations per city). The samples spanned the time period from July 2012 to August 2023.

We retrieved a total of 58,246 reviews. These reviews either contained only positive experiences with a physician or a clinic or had mixed sentiments and targets. Table 1 summarises some selected feedback examples.

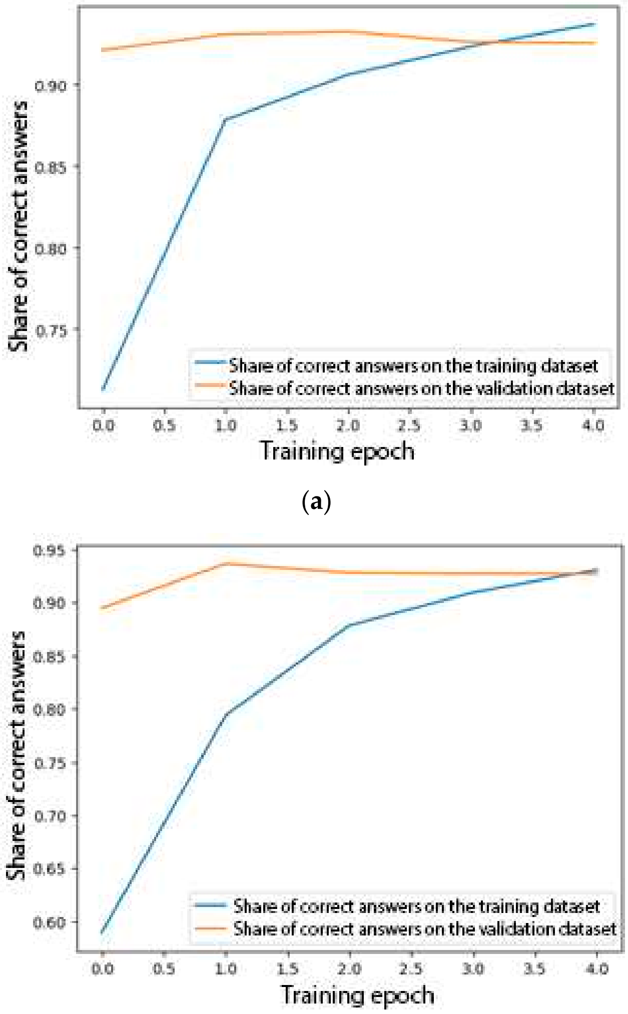
Table 1. Feedback examples, as posted on the website infodoctor.ru.

#	Feedback Text	Feedback Data	Sentiment Class	Target Class
1	"The doctor was really rude, she had no manners with the patients, she didn't care about your poor health, all she wanted was to get home early. I never want to see that doctor again. She's rubbish, I wouldn't recommend her to anyone."	Ekaterina, April 13, 2023, Moscow	Negative	physician
2	"I had to get an MRI scan of my abdomen. They kept me waiting. They gave me the scan results straight away; I'll show them to my doctor. It was easy for me to get to the clinic. Their manners were not very good. I won't be going back there."	Kamil, April 17, 2023, Moscow	Negative	clinic
3	"All those good reviews are written by their staff marketers, they try to stop the bad ones, they don't let any real complaints get through. The clinic is very pricey, they just want to make money, no one cares about your health there."	Anonymous, April 10, 2023, Moscow	Negative	clinic
4	"What they do in this clinic is rip you off because they make you do checkups and tests that you don't need. I found out when I was going through all this stuff, and then I wondered why I had to do it all."	Arina, March 2, 2023, Moscow	Negative	clinic
5	"Rubbish doctor. My problem is really bad skin dryness and rashes because of that. ######## just said, "you just moisturise it" and that was it. She didn't tell me how to moisturise my skin or what to use for moisturiser. I had to push her asking for advice on care and what to do next. She didn't give me anything except some cream, and that only after I asked her".	Anonymous, May 11, 2023, Moscow, Russia	Negative	physician
6	"My husband had a bad tooth under the crown, the dentist said he had to redo his whole jaw and put all new crowns again, like he had to sort everything out to fit the new crowns after the tooth was fixed. In the end we trusted the dentist and redid my husband's whole jaw. The bridge didn't last a month, it kept coming out. In the end we had to do it all over again with another dentist at another clinic. He was awful, he only rips you off. I don't recommend this dentist to anyone."	Tatyana, April 13, 2023, Moscow	Negative	physician

7	" In 2020, I was going to a doctor at the clinic #######.ru for 3 months for the pain in my left breast. He gave me some cream and told me to go on a diet, but I was getting worse. I went to see another doctor; it turned out it was breast cancer. Nearly killed me…"	Maya, March 27, 2023, Moscow	Negative	physician
8	"####### nearly left my child with one leg. A healthy 10-month-old child had to have two surgeries after what this “doctor” had given him. It’s over now, but the “nice” memory of this woman will stay with me forever."	Elizaveta, March 16, 2023, Moscow	Negative	physician

Note: the above translation is made from the reviews in their original form.

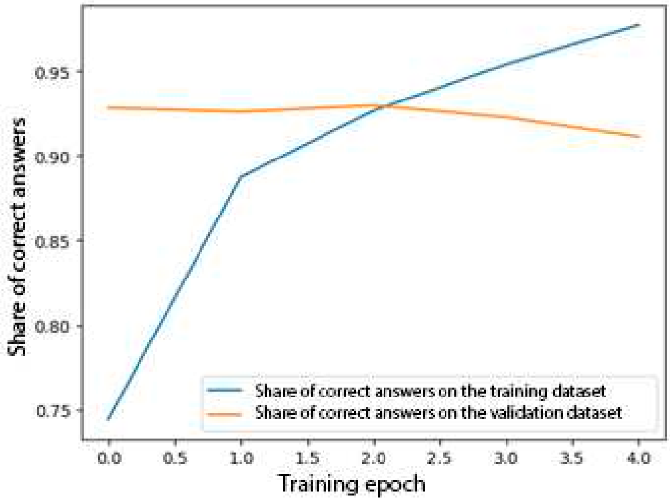
All the algorithms had an 80:20 split between the training and test samples. Figure 6 gives the graphs showing the classification results on the training and test datasets for LSTM, GRU, and CNN architectures.



(a)

(b)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 December 2023 doi:10.20944/preprints202312.1933.v1



(**c**)

Figure 6. Classification results on the training and test datasets for LSTM-network (a), GRU-network (b), and CNN-network (c).

Table 2 compares the performance indicators of text feedback classification using the above approaches.

Table 2. Performance indicators of text feedback classification

	LSTM	GRU	CNN	SVM	BERT
Accuracy	0.9369	0.9309	0.9772	0.8441	0.8942
Val_accuracy	0.9253	0.9271	0.9112	0.8289	0.8711
Loss	0.1859	0.2039	0.0785	0.3769	0.1729
Val_loss	0.2248	0.2253	0.3101	0.3867	0.2266

where:

- Accuracy — training accuracy;
- Val_accuracy — validation accuracy;
- Loss — training loss;
- Val_loss — validation loss.

To compare the proposed models with other methods, we performed experiments using the support vector machine (SVM) and RuBERT algorithms on the same dataset. As shown in Table 2, these algorithms had slightly lower performances than our models.

As mentioned earlier, one of the difficulties with the text reviews analysed was that they could contain elements of different classes within the same review. For instance, a short text message could include both a review of a physician and a review of a clinic.

Hence, we introduced two additional classes, mixed positive and mixed negative, while applying the linguistic method of named entity recognition (as described in Section 3.4) to enhance the quality of classification of each feedback class.

This approach improved the classification performance for all the three artificial neural network architectures. Figure 7 illustrates the classification results obtained using the hybrid algorithms.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 December 2023 doi:10.20944/preprints202312.1933.v1

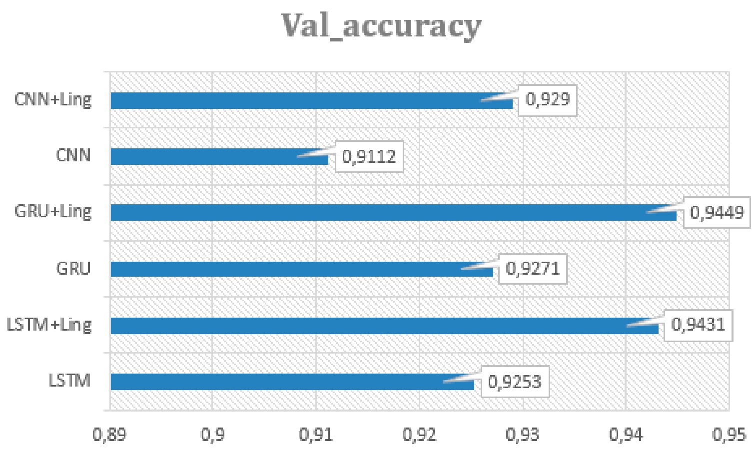


Figure 7. Classification results obtained using hybrid algorithms

We applied the linguistic method to the reviews that were classified as "clinic review" by the neural network at the first stage, regardless of their sentiment. The named entity recognition method improved the classification performance when it was used after the partitioning of text messages. However, some reviews were still misclassified even after applying the linguistic method. Those reviews were long text messages that could belong to different classes semantically. The reasons for the misclassification were as follows:

- Some reviews were of both a clinic and a physician without mentioning the latter's name. This prevented the named entity recognition tool from assigning the reviews to the mixed class. This problem could be solved by parsing the sentences further with identifying a semantically significant object unspecified by a full name.
- Some reviews expressed contrasting opinions about the clinic, related to different aspects of its operation. The opinions often differed on the organisational support versus the level of medical services provided by the clinic.

For example, the review "This doctor's always absent sick, she's always late for appointments, she's always chewing. Cynical and unresponsive" expresses the patient's dissatisfaction with the quality of organisation of medical appointments, which is a negative review of organisational support. The review "I had meniscus surgery. He fooled me. He didn't remove anything, except damaging my blood vessel. I ended up with a year of treatment at the Rheumatology Institute. ####### tried to hide his unprofessional attitude to the client. If you care about your health, do not go to ########" conveys the patient's indignation at the quality of treatment, which belongs to the medical service class. A finer classification of clinic reviews will enhance the meta-level classification quality.

6. Conclusions

In this paper, we propose a hybrid method for classifying Russian-language text reviews of medical facilities extracted from social media.

The method consists of two steps: first, we use one of the artificial neural network architectures (LSTM, CNN, or GRU) to classify the reviews into four main classes based on their sentiment and target (positive or negative, physician or clinic); second, we apply a linguistic approach to extract named entities from the reviews.

We evaluated the performance of our method on a dataset of more than 60,000 Russian-language text reviews of medical services provided by clinics or physicians from the websites prodoctorov.ru and infodoctor.ru. The main results of our experiments are as follows:

- The neural network classifiers achieve high accuracy in classifying the Russian-language reviews from social media by sentiment (positive or negative) and target (clinic or physician) using various architectures of the LSTM, CNN, or GRU networks, with the GRU-based architecture being the best (val_accuracy=0.9271).

- The named entity recognition method improves the classification performance for each of the neural network classifiers when applied to the segmented text reviews.
- To further improve the classification accuracy, semantic segmentation of the reviews by target and sentiment is required, as well as a separate analysis of the resulting fragments.

As future work, we intend to develop a text classification algorithm that can distinguish between reviews of medical services and reviews of organisational support for clinics. This would enhance the classification quality at the meta-level, as it is more important for managers to separate reviews of the medical services and diagnosis from those of organisational support factors (such as waiting time, cleanliness, politeness, location, speed of getting results, etc.).

Moreover, in a broader context, refined classification of social media users' statements on review platforms or social networks would enable creating a system of standardised management responses to changes in demographic and social behaviours and attitudes towards socially significant services and programmes [73].

Author Contributions: Conceptualization, I.K.; Formal analysis, V.M., Z.K., G.K, Software, M.K.; A.K. All authors have read and agreed to the published version of the manuscript.

Funding: The study was supported by the Russian Science Foundation, project No. 23-71-01101 "Development of models and methods for improving the performance of data warehouses through predictive analysis of temporal diagnostic information" and within the framework of the research project "Population reproduction in socio-economic development" with the support of the Faculty of Economics of Lomonosov Moscow State University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. Tourism management, 29(3), 458-468
2. Ismagilova, E., Dwivedi, Y. K., Slade, E., & Williams, M. D. (2017). Electronic word of mouth (eWOM) in the marketing context: A state of the art analysis and future directions.
3. Cantallops, A. S., & Salvi, F. (2014). New consumer behavior: A review of research on eWOM and hotels. International Journal of Hospitality Management, 36, 41-51.
4. Mulgund, P., Sharman, R., Anand, P., Shekhar, S., & Karadi, P. (2020). Data quality issues with physician-rating websites: systematic review. Journal of Medical Internet Research, 22(9), e15916
5. Ghimire, B., Shanaev, S., & Lin, Z. (2022). Effects of official versus online review ratings. Annals of Tourism Research, 92, 103247.
6. Xu, Y., & Xu, X. (2023). Rating deviation and manipulated reviews on the Internet—A multi-method study. Information & Management, 103829.
7. Hu, N., Bose, I., Koh, N. S., & Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. Decision support systems, 52(3), 674-684.
8. Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. Management Science, 62(12), 3412-3427.
9. Namatherdhala, B., Mazher, N., & Sriram, G. K. (2022). Artificial Intelligence in Product Management: Systematic review. International Research Journal of Modernization in Engineering Technology and Science, 4(7)
10. Jabeur, S. B., Ballouk, H., Arfi, W. B., & Sahut, J. M. (2023). Artificial intelligence applications in fake review detection: Bibliometric analysis and future avenues for research. Journal of Business Research, 158, 113631
11. Emmert, M., & McLennan, S. (2021). One decade of online patient feedback: longitudinal analysis of data from a German physician rating website. Journal of Medical Internet Research, 23(7), e24229
12. Kleefstra, S. M., Zandbelt, L. C., Borghans, I., de Haes, H. J., & Kool, R. B. (2016). Investigating the potential contribution of patient rating sites to hospital supervision: exploratory results from an interview study in the Netherlands. Journal of Medical Internet Research, 18(7), e201

13. Bardach, N. S., Asteria-Peñaloza, R., Boscardin, W. J., & Dudley, R. A. (2013). The relationship between commercial website ratings and traditional hospital performance measures in the USA. BMJ quality & safety, 22(3), 194-202
14. Van de Belt, T. H., Engelen, L. J., Berben, S. A., Teerenstra, S., Samsom, M., & Schoonhoven, L. (2013). Internet and social media for health-related information and communication in health care: preferences of the Dutch general population. Journal of medical Internet research, 15(10), e220.
15. Hao, H., Zhang, K., Wang, W., & Gao, G. (2017). A tale of two countries: International comparison of online doctor reviews between China and the United States. International journal of medical informatics, 99, 37-44.
16. Bidmon, S., Elshiewy, O., Terlutter, R., & Boztug, Y. (2020). What patients value in physicians: analyzing drivers of patient satisfaction using physician-rating website data. Journal of medical Internet research, 22(2), e13830
17. Ellimoottil, C., Leichtle, S. W., Wright, C. J., Fakhro, A., Arrington, A. K., Chirichella, T. J., & Ward, W. H. (2013). Online physician reviews: the good, the bad and the ugly. Bulletin of the American College of Surgeons, 98(9), 34-39
18. Bidmon, S., Terlutter, R., & Röttl, J. (2014). What explains usage of mobile physician-rating apps? Results from a web-based questionnaire. Journal of medical Internet research, 16(6), e3122
19. Lieber, R. (2012). The web is awash in reviews, but not for doctors. Here's why. New York Times, (March 9)
20. Daskivich, T. J., Houman, J., Fuller, G., Black, J. T., Kim, H. L., & Spiegel, B. (2018). Online physician ratings fail to predict actual performance on measures of quality, value, and peer review. Journal of the American Medical Informatics Association, 25(4), 401-407
21. Gray, B. M., Vandergrift, J. L., Gao, G. G., McCullough, J. S., & Lipner, R. S. (2015). Website ratings of physicians and their quality of care. JAMA internal medicine, 175(2), 291-293
22. Skrzypecki, J., & Przybek, J. (2018, May). Physician review portals do not favor highly cited US ophthalmologists. In Seminars in Ophthalmology (Vol. 33, No. 4, pp. 547-551). Taylor & Francis
23. Widmer, R. J., Maurer, M. J., Nayar, V. R., Aase, L. A., Wald, J. T., Kotsenas, A. L., ... & Pruthi, S. (2018, April). Online physician reviews do not reflect patient satisfaction survey responses. In Mayo Clinic Proceedings (Vol. 93, No. 4, pp. 453-457). Elsevier
24. Saifee, D. H., Bardhan, I., & Zheng, Z. (2017). Do Online Reviews of Physicians Reflect Healthcare Outcomes?. In Smart Health: International Conference, ICSH 2017, Hong Kong, China, June 26-27, 2017, Proceedings (pp. 161-168). Springer International Publishing
25. Trehan, S. K., Nguyen, J. T., Marx, R., Cross, M. B., Pan, T. J., Daluiski, A., & Lyman, S. (2018). Online patient ratings are not correlated with total knee replacement surgeon–specific outcomes. HSS Journal, 14(2), 177-180
26. Doyle, C., Lennox, L., & Bell, D. (2013). A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ open, 3(1), e001570
27. Okike, K., Uhr, N. R., Shin, S. Y., Xie, K. C., Kim, C. Y., Funahashi, T. T., & Kanter, M. H. (2019). A comparison of online physician ratings and internal patient-submitted ratings from a large healthcare system. Journal of General Internal Medicine, 34, 2575-2579
28. Rotman, L. E., Alford, E. N., Shank, C. D., Dalgo, C., & Stetler, W. R. (2019). Is there an association between physician review websites and press ganey survey results in a neurosurgical outpatient clinic?. World Neurosurgery, 132, 891-899
29. Lantzy, S., & Anderson, D. (2020). Can consumers use online reviews to avoid unsuitable doctors? Evidence from RateMDs. com and the Federation of State Medical Boards. Decision Sciences, 51(4), 962-984
30. Gilbert, K., Hawkins, C. M., Hughes, D. R., Patel, K., Gogia, N., Sekhar, A., & Duszak Jr, R. (2015). Physician rating websites: do radiologists have an online presence?. Journal of the American College of Radiology, 12(8), 867-871
31. Okike, K., Peter-Bibb, T. K., Xie, K. C., & Okike, O. N. (2016). Association between physician online rating and quality of care. Journal of medical Internet research, 18(12), e324
32. Imbergamo, C., Brzezinski, A., Patankar, A., Weintraub, M., Mazzaferro, N., & Kayiaros, S. (2021). Negative online ratings of joint replacement surgeons: An analysis of 6,402 r
33. Mostaghimi, A., Crotty, B. H., & Landon, B. E. (2010). The availability and nature of physician information on the internet. Journal of general internal medicine, 25, 1152-1156

34. Lagu, T., Hannon, N. S., Rothberg, M. B., & Lindenauer, P. K. (2010). Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. Journal of general internal medicine, 25, 942-946
35. López, A., Detz, A., Ratanawongsa, N., & Sarkar, U. (2012). What patients say about their doctors online: a qualitative content analysis. Journal of general internal medicine, 27, 685-692
36. Shah, A. M., Yan, X., Qayyum, A., Naqvi, R. A., & Shah, S. J. (2021). Mining topic and sentiment dynamics in physician rating websites during the early wave of the COVID-19 pandemic: Machine learning approach. International Journal of Medical Informatics, 149, 104434.
37. Bidmon, S., Elshiewy, O., Terlutter, R., & Boztug, Y. (2020). What patients value in physicians: analyzing drivers of patient satisfaction using physician-rating website data. Journal of medical Internet research, 22(2), e13830.
38. Shah, A. M., Yan, X., Tariq, S., & Ali, M. (2021). What patients like or dislike in physicians: Analyzing drivers of patient satisfaction and dissatisfaction using a digital topic modeling approach. Information Processing & Management, 58(3), 102516
39. Lagu, T., Metayer, K., Moran, M., Ortiz, L., Priya, A., Goff, S. L., & Lindenauer, P. K. (2017). Website characteristics and physician reviews on commercial physician-rating websites. Jama, 317(7), 766-768
40. Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. Management science, 54(3), 477-491
41. Pavlou, P. A., & Dimoka, A. (2006). The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. Information systems research, 17(4), 392-414
42. Terlutter, R., Bidmon, S., & Röttl, J. (2014). Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. Journal of medical Internet research, 16(3), e97
43. Emmert, M., & Meier, F. (2013). An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. Journal of medical Internet research, 15(8), e2655
44. Nwachukwu, B. U., Adjei, J., Trehan, S. K., Chang, B., Amoo-Achampong, K., Nguyen, J. T., ... & Ranawat, A. S. (2016). Rating a sports medicine surgeon's "quality" in the modern era: an analysis of popular physician online rating websites. HSS Journal®, 12(3), 272-277
45. Obele, C. C., Duszak Jr, R., Hawkins, C. M., & Rosenkrantz, A. B. (2017). What patients think about their interventional radiologists: assessment using a leading physician ratings website. Journal of the American College of Radiology, 14(5), 609-614
46. Emmert, M., Meier, F., Pisch, F., & Sander, U. (2013). Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. Journal of medical Internet research, 15(8), e2702.
47. Gao, G. G., McCullough, J. S., Agarwal, R., & Jha, A. K. (2012). A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. Journal of medical Internet research, 14(1), e38
48. Rahim, A. I. A., Ibrahim, M. I., Musa, K. I., Chua, S. L., & Yaacob, N. M. (2021, October). Patient satisfaction and hospital quality of care evaluation in malaysia using servqual and facebook. In Healthcare (Vol. 9, No. 10, p. 1369). MDPI
49. Galizzi, M. M., Miraldo, M., Stavropoulou, C., Desai, M., Jayatunga, W., Joshi, M., & Parikh, S. (2012). Who is more likely to use doctor-rating websites, and why? A cross-sectional study in London. BMJ open, 2(6), e001493
50. Hanauer, D. A., Zheng, K., Singer, D. C., Gebremariam, A., & Davis, M. M. (2014). Public awareness, perception, and use of online physician rating sites. Jama, 311(7), 734-735
51. McLennan, S., Strech, D., Meyer, A., & Kahrass, H. (2017). Public awareness and use of German physician ratings websites: Cross-sectional survey of four North German cities. Journal of medical Internet research, 19(11), e387

52. Lin, Y., Hong, Y. A., Henson, B. S., Stevenson, R. D., Hong, S., Lyu, T., & Liang, C. (2020). Assessing patient experience and healthcare quality of dental care using patient online reviews in the United States: mixed methods study. Journal of Medical Internet Research, 22(7), e18652
53. Emmert, M., Meier, F., Heider, A. K., Dürr, C., & Sander, U. (2014). What do patients say about their physicians? An analysis of 3000 narrative comments posted on a German physician rating website. Health policy, 118(1), 66-73
54. Greaves, F., Ramirez-Cano, D., Millett, C., Darzi, A., & Donaldson, L. (2013). Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. BMJ quality & safety, 22(3), 251-255.
55. Hao, H., & Zhang, K. (2016). The voice of Chinese health consumers: a text mining approach to web-based physician reviews. Journal of medical Internet research, 18(5), e108.
56. Shah, A. M., Yan, X., Shah, S. A. A., & Mamirkulova, G. (2020). Mining patient opinion to evaluate the service quality in healthcare: a deep-learning approach. Journal of Ambient Intelligence and Humanized Computing, 11, 2925-2942
57. Wallace, B. C., Paul, M. J., Sarkar, U., Trikalinos, T. A., & Dredze, M. (2014). A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. Journal of the American Medical Informatics Association, 21(6), 1098-1103
58. Ranard, B. L., Werner, R. M., Antanavicius, T., Schwartz, H. A., Smith, R. J., Meisel, Z. F., ... & Merchant, R. M. (2016). Yelp reviews of hospital care can supplement and inform traditional surveys of the patient experience of care. Health Affairs, 35(4), 697-705.
59. Hao, H. (2015). The development of online doctor reviews in China: an analysis of the largest online doctor review website in China. Journal of medical Internet research, 17(6), e134
60. Jiang, S., & Street, R. L. (2017). Pathway linking internet health information seeking to better health: a moderated mediation study. Health Communication, 32(8), 1024-1031
61. Hotho, A. Nürnberger and G. Paaß, "A Brief Survey of Text Mining", LDV Forum - GLDV Journal for Computational Linguistics and Language Technology, vol. 20, pp.19-62, 2005.
62. V. Păvăloaia, E. Teodor, D. Fotache and M. Danileț, M. "Opinion Mining on Social Media Data: Sentiment Analysis of User Preferences", Sustainability, 11, 4459, 2019.
63. D. Bespalov, B. Bing, Q. Yanjun and A. Shokoufandeh, "Sentiment classification based on supervised latent n-gram analysis", Proceedings of the 20th ACM international conference on Information and knowledge management (CIKM '11). Association for Computing Machinery, New York, NY, USA, 375–382, 2011.
64. Irfan, R., King, C.K., Grages, D., Ewen, S., Khan, S.U., Madani, S.A., Kolodziej, J., Wang, L., Chen, D., Rayes, A., Tziritas, N., Xu, C.Z., Zomaya, A.Y., Alzahrani, A.S. & Li, H., A Survey on Text Mining in 196 Phat Jotikabukkana, et al. Social Networks, Cambridge Journal, The Knowledge Engineering Review, 30(2), pp. 157-170, 2015.
65. Patel, P. & Mistry, K., A Review: Text Classification on Social Media Data, IOSR Journal of Computer Engineering, 17(1), pp. 80-84, 2015.
66. Lee, K., Palsetia, D., Narayanan, R., Patwary, Md.M.A., Agrawal, A. & Choudhary, A.S, Twitter Trending Topic Classification, in Proceeding of the 2011 IEEE 11 th International Conference on Data Mining Workshops, ICDW'11, pp. 251-258, 2011.
67. Kateb, F. & Kalita, J., Classifying Short Text in Social Media: Twitter as Case Study, International Journal of Computer Applications, 111(9), pp. 1-12, 2015.
68. Chirawichitichai, N., Sanguansat, P. & Meesad, P., A Comparative Study on Feature Weight in Thai Document Categorization Framework, 10th International Conference on Innovative Internet Community Services (I2CS), IICS, pp. 257-266, 2010.
69. Theeramunkong, T. & Lertnattee, V., Multi-Dimension Text Classification, SIIT, Thammasat University, 2005.http://www.aclweb.org /anthology/C02-1155 (25 October 2023).
70. Viriyayudhakorn, K., Kunifuji, S. & Ogawa, M., A Comparison of Four Association Engines in Divergent Thinking Support Systems on Wikipedia, Knowledge, Information, and Creativity Support Systems, KICSS2010, Springer, pp. 226-237, 2011.
71. Sornlertlamvanich, V., Pacharawongsakda, E. & Charoenporn, T., Understanding Social Movement by Tracking the Keyword in Social Media, in MAPLEX2015, Yamagata, Japan, February 2015.

72. Konstantinov A., Moshkin V., Yarushkina N. Approach to the Use of Language Models BERT and Word2vec in Sentiment Analysis of Social Network Texts. In: Dolinina O. et al. (eds) Recent Research in Control Engineering and Decision Making. ICIT 2020. Studies in Systems, Decision and Control, vol 337. Springer, Cham. pp 462-473 https://doi.org/10.1007/978-3-030-65283-8_38
73. Kalabikhina, I., Zubova, E., Loukachevitch, N., Kolotusha, A., Kazbekova, Z., Banin, E., Klimenko, G., Identifying Reproductive Behavior Arguments in Social Media Content Users' Opinions through Natural Language Processing Techniques, Population and Economics, 7(2), pp. 40-59, 2023.