ISSN 0869-3617 (Print), ISSN 2072-0459 (Online) http://vovr.elpub.ru

Digital Footprint: Assessing Student Satisfaction with Education Quality

Original article
DOI: 10.31992/0869-3617-2024-33-2-89-108

Mikhail M. Krishtal – Dr. Sci. (Physics-Mathematics), Professor, Rector, ORCID: 0000-0001-7189-0002, office@tltsu.ru

Anna V. Bogdanova – Cand. Sci. (Pedagogical Sciences), Head of Online Education Technologies Department, ORCID: 0000-0002-3553-2272, WoS Reseacher ID: GRO-7042-2022, a.bogdanova@tltsu.ru

Togliatti State University, Togliatti, Russian Federation
Address: 445020, Samara region, Togliatti, Belorusskaya str., 14

Mikhail G. Myagkov – PhD (Pedagogical Sciences), Professor, Leading Researcher, Centre for Cognitive Research and Neuroscience, ORCID: 0000-0002-8419-6404, myagkov@darkwing.uoregon.edu

Yulia K. Alexandrova – Junior Researcher, Centre for Applied Big Data Analysis, ORCID: 0000-0002-6069-779X, jalexandrova@data.tsu.ru

Tomsk State University, Tomsk, Russian Federation
Address: 36, Lenin Ave., Tomsk, 634050

Abstract. The COVID-19 pandemic has changed the way learning is organized around the world. Russian universities have also been faced with the need to quickly transfer all teaching to an online format. The importance of student satisfaction with the education quality in online learning is increasing, since it is an important condition for motivation. The paper shows that based on the analysis of students' messages in the social network, it is possible to observe and analyze the overall dynamics and trends in student community satisfaction with the quality of the learning / the efficiency of universities and conduct a comparative analysis of the identified characteristic data groups with their totality. It is shown that the data on the reaction of students of a particular university may have significant deviations from the totality of data, which reflects the characteristics of the reaction of students of a particular university to events occurring at the same time. This may indicate the internal differences of the university, which form an appropriate response to external events.

To understand the satisfaction of students in the transition to a new implementation format of the learning. The digital traces of students from the social network VKontakte were analyzed using individual Big Data tools on the PolyAnalyst software platform. This made it possible to trace changes in the mood of students and, on the example of a single university, to identify and explain deviations in the attitude of students to the implementation of the learning, as well as to verify the methodology. The methodology developed by authors makes it possible to detect problematic issues in the

Контент доступен под лицензией Creative Commons Attribution 4.0 License.
This work is licensed under a Creative Commons Attribution 4.0 License.
© Krishtal M.M., Bogdanova A.V., Myagkov M.G., Alexandrova Yu.K. , 2024.

university, including the moment of their occurrence, relevance, degree of concern of students. Such content analysis can be used not only to assess students' satisfaction with the quality of the learning, but also to monitor the emergence of any problems that cause concern and strong reactions on the part of the student community, as well as other communities and individual groups.

Keywords: COVID-19 pandemic, Rosdistant, quality of education, online learning, digital footprint, big data, social networks, sentiment analysis

Cite as: Krishtal, M.M., Bogdanova, A.V., Myagkov, M.G., Alexandrova, Yu.K. (2024). Digital Footprint: Assessing Student Satisfaction with Education Quality. *Vysshee obrazovanie v Rossii = Higher Education in Russia.* Vol. 33, no. 2, pp. 89-108, doi: 10.31992/0869-3617-2024-33-2-89-108 (In Russ., abstract in Eng.).

Introduction

With the onset of the COVID-19 pandemic, higher education around the world has faced a number of extraordinary challenges that have not previously had precedents of this magnitude. Face-to-face learning and interaction between students and university has been limited at all levels in a short time. The security measures associated with COVID-19 have made it virtually impossible for many higher education institutions around the world to offer face-to-face teaching. Taking such radical measures is always accompanied by numerous side-effects. This is confirmed by the analytical data published in various countries [1–5].

The insufficient level of digitalization of higher education in almost all countries before COVID-19 and the low pace of development of pedagogical technologies did not allow the transition from face-to-face meetings with the teacher to distance learning (DL) in an online format. The changes that occurred were perceived as destructive events, aggressive and undesirable experiences [3]. Most universities faced the problems of inadequate staff training and lack of understanding of the DL results [4].

Despite the previous trend towards digitalization of almost all social spheres, including higher education, the COVID-19 pandemic acted as a catalyst for it and exacerbated a number of problems [5]. In Russia in 2020, more than 4 million students were affected by the negative changes caused by the COVID-19 pandemic[1]. Therefore, for the higher education system as a whole and for particular universities, the need to assess student satisfaction with the quality of education and the effectiveness of universities in these force majeure conditions has come to the fore.

Social network analysis, which is a continuously updated stream of unstructured data, i.e. big data that carries useful social information, is now being considered as a tool for such tasks, more effective than sociological surveys [6].

During global crises, the production of user-generated content, including posts on social networks, blogs, and forums, increases dramatically [6–11]. A large part of it is "noise", containing no value for analysis. However, some of the available information can be used to detect general trends [7], understand the extent of the crisis, or elucidate characteristic and specific changes in the users' sentiments [9]. A significant flow of information does not allow people to effectively extract, comprehend and organize information, and then make decisions about adequate actions without appropriate IT solutions and services [10]. Therefore, monitoring and analysing social networks as a means of assessing students' satisfaction with the quality of the educational process is of particular interest.

Determination of satisfaction with something can be carried out on the basis of sentiment

[1] Information about the number of students of educational organizations, carrying out educational activities on educational programs of higher education. Ministry of Science and Higher Education of the Russian Federation. URL: https://minobrnauki.gov.ru/opendata/9710062939-svedeniya-o-chislennosti-studentov-obrazovatelnykh-organizatsiy-osushchestvlyayushchikh-obrazovateln. (date of access: 22.12.2022).

analysis of text messages [12]. To do this, linguistic markers are identified that characterize the attitude of the author of the message to the subject of evaluation, and that reflect his emotional coloring and his "sentiment".

The problem of sentiment analysis has become a popular object of research over the past decade. One of the reasons for this is the growing number of text messages that have become available thanks to the development of social networks [12]. Another reason for the increased interest in sentiment analysis is the achievements in the field of natural language processing by means of machine learning and Big Data tools. An overview of various methods, suggested for analysis of opinions and sentiments, presented in the paper [13]. The main conclusion is that such an analysis goes beyond technical issues.

It is important that social networks form communities, for example, university communities, which provide mass communication and interaction of interested users within the framework of certain topics [14]. This facilitates the analysis of information in terms of highlighting messages related to the subject of interest. Communications and information transfer in such communities are much more effective than in other means of mass communication [15; 16]. Posts and comments on social networks convey the dynamics of user opinions with great accuracy and in real time, since they are not subjected to additional censorship and processing, reflecting the current moment during publication [17–21]. That is, social media judgments or opinions provide the latest and most comprehensive information due to the proliferation of social media and the low barrier to posting.

During the pandemic, many researchers turned to the issues of assessing the quality of the educational process by students using the tools of sociology and psychology. In particular, the paper [22] reports on a longitudinal series of 18 interviews with digital accessibility leaders since the official declaration of the COVID-19 pandemic by the World Health Organization in March 2020. The review [23], based on sociological surveys, notes a low level of adoption of online learning technologies among teachers, due to a poor understanding of their applicability and a lack of experience, despite a generally positive attitude towards such technologies. Assessing the quality of education with the tools of sociology has shown that traditional structures and ways of working cannot effectively ensure the creation of hybrid spaces for interaction between students, teachers and university administration, and that there is a need for structural innovations that can create a context of collaboration, interdisciplinarity and new connections within organisations. [24; 25].

Sociological studies in this area also confirm that the assessment of the quality of the educational process is becoming the subject of close attention today. In particular, a study conducted at the Higher School of Economics shows that by September 2021, compared to the same period in 2020, negative assessments and criticism from students regarding quality of education have significantly decreased. At the time of publication of the data from this study, about 70% of the students surveyed claim that the DL allowed them to learn productively and interact with both the teacher and other students [26].

The area of research to which this article is devoted is cross-disciplinary and includes sociology, natural language processing and *Big Data* tools. Growing amounts of data and computing power allow the use of advanced forms of analysis: *Big Data* can become today the dominant tool for analyzing moods and tracking the trend of students' satisfaction with the quality of the educational process in dynamics. The authors of the present research are not aware of any previous attempts of use *Big Data* tools to recognize the dynamics of opinions in messages related to the assessment of students' satisfaction with the quality of the educational process or users' satisfaction with any service of this scale.

The restrictions imposed during the pandemic in Russia on the higher education system can be considered as an ideal case for analysis for the following reasons:

– restrictions regarding the transfer of full-time education to an online format were introduced simultaneously for all universities, regardless of their subordination, and also without the possibility of long-term preparation for them;

– in Russia, digital competencies at the state level are more developed than in the vast majority of countries of the world (the "Electronic Government" project[2], the national program "Digital Economy"[3]), according to the UNO, Russia is among the leaders in the development of digital public services[4];

– in Russia, the digital culture of the population is highly developed, while the vast majority of student youth use the social network VKontakte [27; 28].

All this together makes the situation a model case, i.e. convenient for the purity of research and obtaining unambiguous results when comparing the reactions of particular parts of university communities to this transition on the basis of sentiment analysis.

The purpose of the work is to identify, based on the analysis of the content of social networks, the dynamics and trends in the satisfaction of the student community as a whole and its selected groups (for example, those belonging to a particular university or territory) with the quality of the educational process and the efficiency of universities during the COVID-19 pandemic and the creation methodological basis for the development of a fully automated methodology for monitoring and analyzing social networks based on *Big Data* to assess the attitude of users to large-scale events and transformations, including student satisfaction with the quality of the educational process.

Research Methodology

The initial data of the present research were more than 2 million messages that the authors extracted from 548 communities of higher educational institutions in the Russian Federation (including six communities of Tomsk State University and five communities of Togliatti State University (TltSU)). In Russia, in terms of the number of registered users and published messages, the social network VKontakte is the leader [27; 28], so it was chosen as a research base. The anonymised data were collected using SN lab's social media platform[5], powered by VKontakte's open-source API.

Data processing took place in four stages:

1) Preliminary automatic cleaning of the unloaded array of messages from garbage;

2) Manual deletion of messages that are not related to the research topics (irrelevant messages);

3) Data set markup:

– Manual partitioning;

– Automated partitioning;

4) Visualization of results and conclusions:

– Forming a tag cloud;

– Building temporary dependencies of positive and negative messages.

Thus, the following methods and algorithms were used in the study:

– The Method of Linguistic Markers: search in the text for certain keywords or phrases that are usually associated with the selected topic;

– Sentiment Analysis: natural language processing techniques to analyze tone and emotion expressed in text or characters;

– Machine Learning: machine learning algorithms trained to recognize patterns and

2 System project of the electronic government of the Russian Federation. URL: https://digital.gov.ru/uploaded/files/sistemnyii-proekt-elektronnogo-pravitelstva-rf.pdf?utm_referrer=https%3a%2f%2fyandex.ru%2f free (date of access: 18.06.2023).

3 "Digital Economy of the Russian Federation": Ministry of Digital Development, Communications and Mass Media of the Russian Federation. URL: https://digital.gov.ru/ru/activity/directions/858/?utm_referrer=https%3a%2f%2fyandex.ru%2f free (date of access: 18.06.2023).

4 UNO Survey: E-Government 2018". URL: https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2018-Survey/E-Government%20Survey%202018_Russian.pdf (date of access: 13.06.2023).

5 Portal for working with social network data. URL: https://lk.opendata.universitet/ (date of access: 22.12.2022).

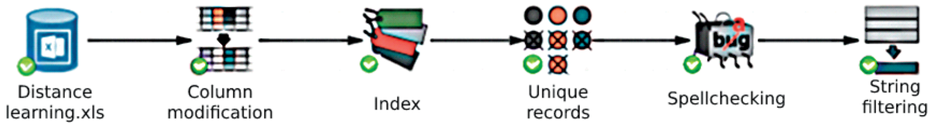

Fig. 1. PolyAnalyst data filtering script (screenshot of the program editor's workspace)

characteristics of texts using labeled training datasets.

1. Preliminary automatic cleaning of the uploaded array of messages from garbage

The data were automatically processed using the *PolyAnalyst* software platform (PolyAnalyst)[6] before manual labeling (Fig. 1).

The number of messages for further investigation was reduced from 2 million to less than 500 thousand. The work of the *PolyAnalyst* software according to the scenario presented took less than 1 hour. The posts that did not contain text and posts published by community administrators were excluded, as they did not carry any semantic load within the framework of the research. Duplicate posts have been removed and spelling errors in the text have been corrected. Irrelevant messages that are not related to the issues studied are excluded. They were screened out by the presence of marker words "buy", "sell", "rent", "rent", "seek", "get acquainted", their synonyms and cognate words.

2. Manual cleaning of messages that do not recur to research topics and manual markup

Manual marking was carried out by 25 markers according to the compiled message marking instructions. Posts and comments classified as an opinion or a value judgment (attitude) were accepted for sorting, and the rest were marked as "garbage". All posts that did not cover the topics chosen by the authors were also marked as "garbage" during the manual markup phase for convenience.

To analyze the emotional coloring of the messages, the authors used the classification according to two emotions of Ekman [29]: "happiness" and "sadness", which are antonyms of each other, and therefore it is easy to distinguish them. The tone of the message was determined in accordance with the rules that allow to answer the question "How does the author of the message relate to the subject of the message?":

– positive – indicates the state of happiness, joy, satisfaction, etc.;

– neutral – information, a statement of fact without attitude, which does not reflect the emotionally expressive coloring;

– negative – indicates a state of sadness, discontent, regret, etc.

The authors used lingual markers to recognize emotions in text messages, examples of which are shown in Table 1.

Stage I of manual marking (Marking algorithm A):

1) take one file for work, open it on a PC;

2) using filters in Excel, look for marker words and words with the same root in messages, mark the tone of the message and verify the markers;

3) after markup using filters, sort the remaining messages and mark them by tone;

4) in the process of markup, mark messages that are not related to the topic of research or do not carry a semantic load as "garbage";

5) give an opinion on the quality of linguistic markers: which of them allow you to uniquely identify the message as positive or negative;

6) at the end of the work, save the changes in the file, close the file.

Based on the results of Algorithm A, the marker lists were adjusted (Table 2).

[6] PolyAnalyst – software product for deep data analysis – Megaputer Intelligence. URL: https://www.megaputer.com/ru/polyanalyst/ (date of access: 22.12.2022).

Table 1

Recommended linguistic markers

The tone of emotion in the message	Examples of linguistic markers	Comments
Negative	Rage, hate, anger, hysterical, hurtful, annoying, outrageous, envious, scared, suspicious, anxious, unruly, apprehensive, fearful, trick, ashamed, guilty, bitter, melancholy, laziness, dissatisfied, upset, hopeless, arrogant, incomplete, uncomfortable, awkward, insecurity, fatigue, depressed, indifference, bad, nightmare, horror, disgusting, vile, strange.	Words used to convey the negative emotions that the speaker (writer) experiences
	Harm, apathy, weird, dead end, never	Words carrying a negative "charge", often used in a story about something unpleasant for someone who speaks (writes)
	Why? What for?	By themselves, these questions are not always unambiguous markers of negativity in messages, but they require attention, because, asking them, a person is trying to find out the cause of something, more often – something that does not suit him. Although they also occur in the context of theoretical questions, such as "Why is everything so beautiful?!", but much less often
	You	Our dissatisfaction is more often reported through You-statements, accusing, forcing to defend yourself, and automatically making the party to which we are addressing an "enemy". A person uses this speech paradox when embarrassed by his emotions and feelings. By saying "you" instead of "I" and thereby alienating one's own needs, feelings, words and thoughts, one attributes them to other people [30].
Positive	Happy, good, delight, enthusiastic, interest, care, hope, excellent, amazing, solemn, cheerful, surprised, class, cool, inspiration, tenderness, warmth, trust, love, native, proud, admiration, appreciation, high, sympathetic, sincere, friendship, charming, kind, well done, dear, thank you, super, great.	Words used to convey the positive emotions that the speaker (writer) experiences.
	Me, I, we.	"We" is used in speech to join. In a positive context, this means a desire to demonstrate an equal position in communication [30]. "I", "me" are used when they are not ashamed of their emotions, feelings; usually it is characteristic of the positive mood of the speaker (writer).

Table 2

Recommended "unambiguous" linguistic markers

The tone of emotion in the message	Enumeration of recommended "unambiguous" linguistic markers
Negative	Infuriating, hateful, angry, hurtful, outrageous, scared, frightened, anxious, humiliating, fearful, ashamed, longing, harm, upset, desperate, arrogant, uncomfortable, awkward, dead end, tired, depressed, bad, nightmare, horror, tin, trash, disgusting, vile, strange.
Positive	Happy, good, delight, enthusiastic, interesting, excellent, cool, inspiration, high, love, proud, sincere, friendship, kind, well done, thank you, great.

Stage II of manual labeling (Labeling algorithm B):

1) take one file for work, open it on a PC;
2) using filters in Excel, look for marker words and words with the same root in messages, mark the tone of the message according to the approved classification of unambiguous identification without analyzing their texts (see Table 2);
3) using filters in Excel, look for the remaining marker words and words with the same root in messages, mark the tone of the message (see Table 1);
4) after markup using filters, sort the remaining messages and mark them by tone;
5) in the process of markup, mark messages that are not related to the topic of research or do not carry a semantic load as "garbage";
6) at the end of the work, save the changes in the file, close the file.

The markup files were shared with other developers for markup cross-checking. Checking critical errors in the markup did not reveal. After all the files were marked up, they were combined into one, in which each message was marked by tone and in relation to one of the research topics.

Since none of the existing tools makes it possible to clearly classify the text as an opinion or attitude, it is impossible to fully automate this part of the work. When manually marking up the text, the mood of each comment was assessed by the markers, taking into account not only individual words, but also context, style and wording to get a complete picture of the mood. Combinations of punctuation marks that form the so-called "emoticons" were taken into account. Such an integrated approach allowed us to more accurately determine the mood.

The set of linguistic markers used for recognizing emotions in text messages was formed by authors in the interests of the study in order to recognize the emotional coloring of text messages. It builds on the research in language psychology and emotional linguistics listed in the introduction and contains the most commonly used words that help determine the tone of the message. The set is not exhaustive and can be supplemented with other linguistic markers, depending on the specific task and the context in which the analysis of emotions is performed, in order to achieve maximum accuracy and efficiency in recognizing the emotional coloring of text messages.

3. Automated partitioning

To test the possibility of using fully automatic markup, messages after automatic cleaning of "garbage" in parallel with manual markup were marked up with *PolyAnalyst* software tools that allow identifying the tone of statements (Fig. 2). Entity Extraction node is means of formalizing unstructured text fields in databases, a text field is represented as a set of Boolean features based on the presence and/or frequency of words, phrases or their synonyms. Sentiment Analysis node is performs sentiment analysis text by identifying the object of evaluation, determining the evaluation itself, sentiment, and the degree of its severity; the algorithm is the intellectual property of Megaputer.

This was done in order to check whether this tool gives sufficiently accurate results.

Fig. 2. Sentiment Analysis script in PolyAnalyst (screenshot of the program editor's workspace)

4. Visualization of results and conclusions

4.1. Forming a tag cloud

In the *PolyAnalyst* software package, the following nodes have been added to the script: "Term Relationship" and "Keyword Extraction" to create a hierarchical catalog of *PolyAnalyst* software products and identify relationships between concepts in the form of a tag cloud (Fig. 3).

4.2. Building Temporal Dependencies on Positive and Negative Messages

To visualize fluctuations in the number of posts of negative and positive , which well illustrate the change of moods in moments, but do not work well for analysis in long periods, polynomial curves were used to approximate the dependencies built on points obtained with different averaging periods using the "sliding window" method. A linear trend was used to identify general trends.

For the category "Student satisfaction with the educational process and interaction with the university", the authors compared the number of messages per user in the TltSU communities and all research communities. To do this, the authors used the ratio of the number of messages to the number of users registered in the communities, estimating the relative number of negative K_{NEG} and positive K_{POS} messages per community user:

$$K_{NEG} = \frac{\text{Number of Negative Messages}}{\text{Number of User Communities}},$$

$$K_{POS} = \frac{\text{Number of Positive Messages}}{\text{Number of User Communities}}. \quad (1)$$

For the calculation, the values of the number of users as of the research period were taken: for TltSU communities – 32,400 people, for all communities – 2,455,040 people. Aggregated values for the number of messages are taken in increments of 14 days, since an average of 1 to 4% of the messages of the entire array were generated during this period.

Results and discussion

The use of *Big Data* tools made it possible to carry out automated markup of messages, which increased the speed of data processing and reduced the likelihood of errors. Their limi-

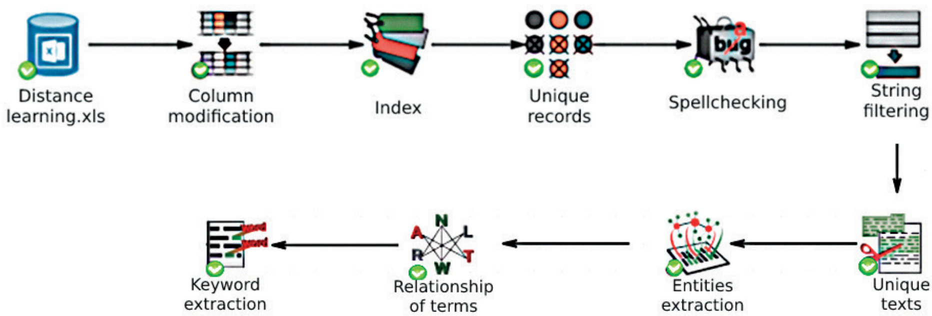

Fig. 3. Script for analyzing the sentiment of messages in the *PolyAnalyst* (screenshot of the program editor's workspace)

tations were taken into account and the stage of manual markup was included in the study, since *Big Data* tools cannot always take into account the context and meaning of messages, which affects the results. To determine the accuracy and level of reliability of the results, authors compared the quantitative results of both markups of messages by emotional coloring (manual and obtained using a script on the *PolyAnalyst* software) without dividing into topics, obtaining the following discrepancies:

– the number of messages marked as positive according to the results of marking on the *PolyAnalyst* is 6% higher;

– the number of messages marked as negative according to the results of marking on the *PolyAnalyst* is less by 5.8%.

Thus, the discrepancy between the results of automated and manual partitioning is no more than $\sqrt{6^2+5,8^2}=8,35\%$.

Because a full substantive review of each tagged message would take a critically long time, authors randomly reviewed the messages labeled in the *PolyAnalyst* as positively and neutrally colored (400 messages) to find possible causes of the discrepancies. The main problem that the authors found when selectively comparing the results of manual marking and automatic sentiment detection using *PolyAnalyst* software was that the algorithms embedded in it in some cases cannot effectively capture sarcasm, irony and passive aggression in comments to posts. For example, if one of the users wrote a post with negative content, and another left a comment to it "Fine!", then *PolyAnalyst* marks such a comment as positive, while a living person will see this and labels the message as negative or neutrally colored depending on the context. Therefore, further work was carried out by the authors based on the file obtained by manual markup. When highlighting keywords and forming a tag cloud from posts that are not marked as positive (Fig. 4), it became clear that posts and comments that have a negative and neutral connotation, students try to leave anonymously, write not about themselves, but about all students abstractly, probably fearing reactions and responses from the university administration or, possibly, fellow students. Manifestations of stress in students attracted attention. It was revealed that about 2.5% of the total number of messages contain significant mentions for the analysis in the category "Stress and psychological health of students" – all of them are negative tonality. This suggests that the dissatisfaction of students with the interaction or organization of the educational process affects their psychological state enough to share this information, to give an outlet to emotions in the form of a message. Indirectly, this is confirmed by the frequent use of markers of neurosis and ignorance, like "in general", "nobody", etc.

Changes in the number of posts on the main topics of the category "Stress and psychological health of students" (Fig. 5) can be

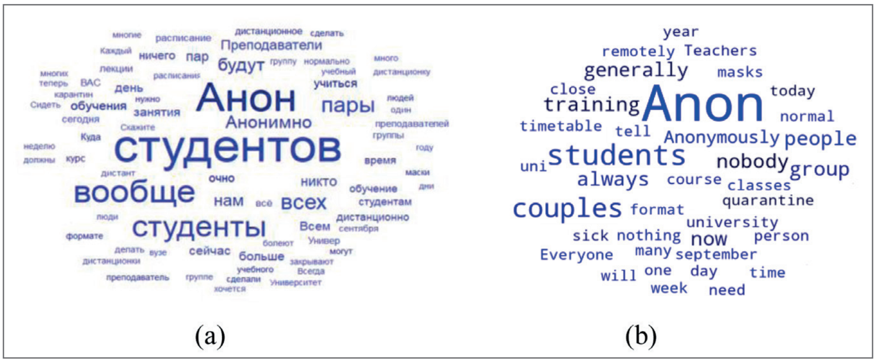

Fig. 4. Keyword cloud in posts related to the category "Stress and Psychological Health of Students": a – result obtained using PolyAnalyst; b – translation of the keyword cloud obtained using PolyAnalyst.

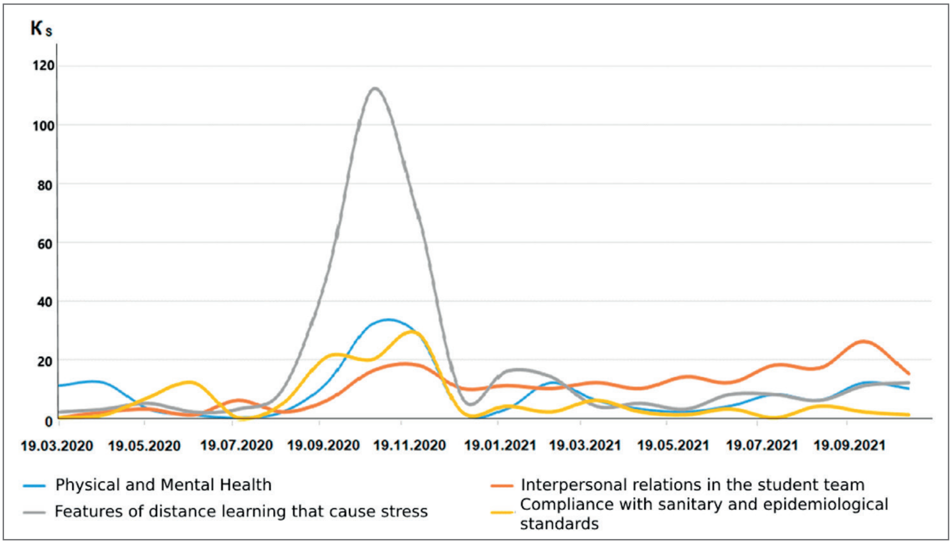

Fig. 5. Dynamics of messages containing information about the manifestations of stress in students

seen that, although universities managed to relieve the main tension regarding pre-school education, over time in 2021, more and more students show signs of health impairment, including psychological well-being in the course of forced isolation.

The peak of activity on most of the subcategories of this topic falls on September-November 2020, with the exception of the topic of relationships in the student collective. This topic is not the most relevant and is evenly distributed throughout the time periods of 2020-2021 with a slight predominance in the spring and summer of 2021. The importance of DL issues and students' concerns for their physical and psychological health in the second semester is sharply reduced. This is probably because educational institutions managed to restructure to work in a pandemic, as indicated by other results of the study.

Since the "Online Higher Education System" has been implemented at TltSU since 2014 [31], this gave us the opportunity to verify the methodology. The Rosdistant project developed within its framework in 2015 (transformed into the NewGenUniv project since 2022) has become an effective platform for analyzing and using *Big Data*. Figures 6 presents the results obtained for the entire set of communities and communities of TltSU in the category "Student satisfaction with the educational process and interaction with the university", including linear trends.

The main trends are as follows:

– in general, the attitude of TltSU students in comparison with the entire set of communities to the topics under study is at a lower level of interest, which indicates a less emotional reaction to the events taking place;

– relatively the number of negative responses for TltSU communities is significantly lower than for all communities, and at first by about 29%, and by the end of the study period by about 38% (when comparing linear trends), despite the fact that there is a clear tendency to reduce the number of negative reviews for all communities;

– the relative number of positive messages for TltSU communities is growing much more confidently than for all communities, that is, TltSU students by the end of the period under review are increasingly willing to speak about the organization of distance and hybrid learning.

It should be noted the peaks of negatively colored messages on the communities of TltSU in May 2020 and in May 2021, which clearly ex-

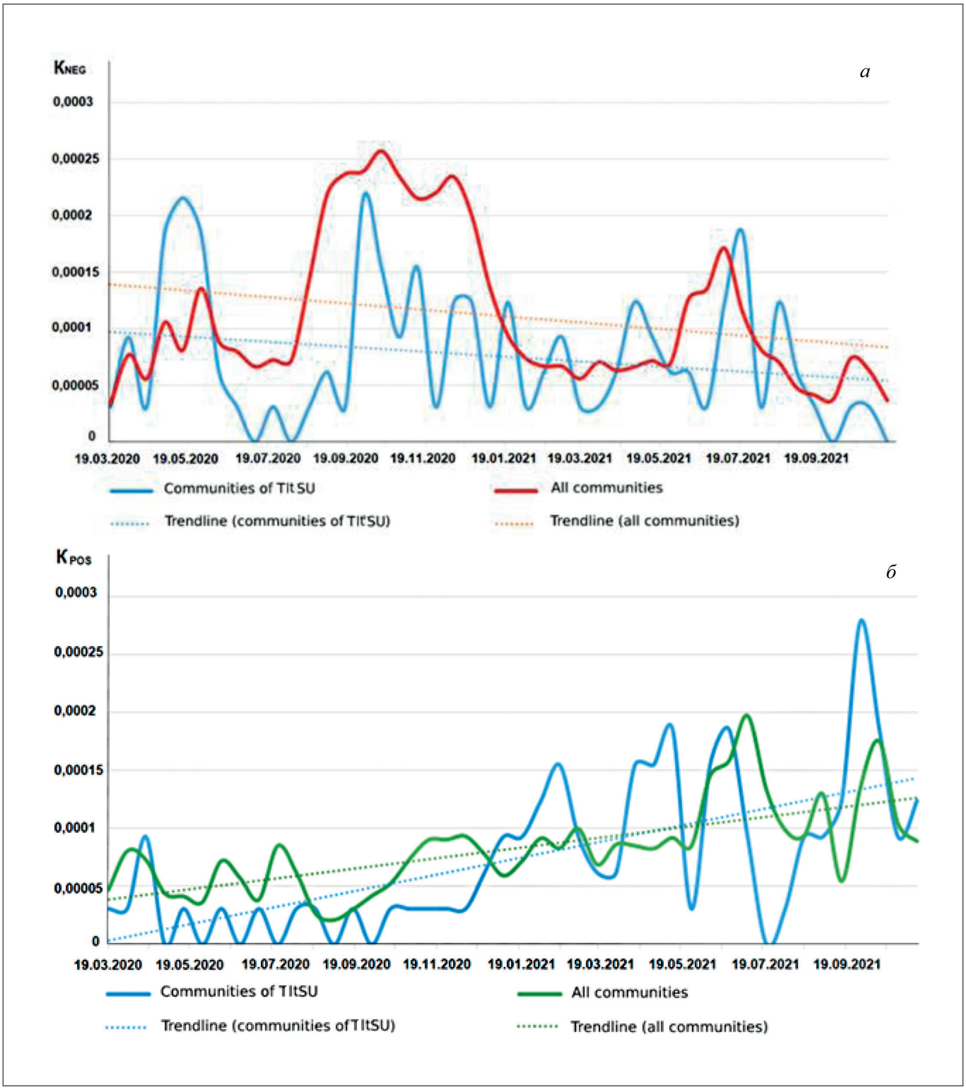

Fig. 6. Number of negative (a) and positive (б) messages per user

ceed the relative number of negative messages for all communities as a whole. In authors opinion, they are associated with wide information in the LMS of TltSU and through other channels of full-time students about conducting remote defense of the final qualifying works. The figures show, that these peaks are not long-term, but are limited to the step of the curve, which indicates effective explanatory work and the willingness of all internal services to respond quickly to requests from students.

Findings

1. Comparison of the results of manual and automated message markup by tone (emotional coloring) showed the high efficiency of *PolyAnalyst* in solving this problem. If the problem of automated tone detection in sarcastic, ironic and passive-aggressive texts can be solved in the future, this tool could be used as a replacement for manual markup, which would significantly reduce the labour and time required for such markup tasks.

2. A methodology has been developed for identifying and comparing characteristic groups of data (for example, those related to a particular university or territory) with their totality, which makes it possible to determine the specifics and causes of deviations and evaluate their relationships with controlled factors.

3. It is shown that based on the analysis of students' messages in the social network (using the social network VKontakte as an example), one can observe and analyze the general dynamics and trends in student community satisfaction with the quality of the educational process / the efficiency of universities and conduct a comparative analysis of the identified characteristic data groups with their totality.

4. The results of the analysis show that the data on the reaction of students of a particular university may have significant deviations from the entire set of data, which reflects the characteristics of the reaction of students of a particular university to events occurring at the same time. This may indicate the internal differences of the university, which form an appropriate response to external events.

5. The evidence demonstrate that full-time students at TltSU were significantly more prepared for the transition to online learning and perceived it significantly more positively, which can be linked to the success of the 'Rosdistant' project [32; 33], as a result of which the university structures, teachers, and the educational process support system and the students themselves were prepared to work in an online format, which was widely used by TltSU in normal times before the pandemic as an element of a hybrid learning format, and during the pandemic became the only possible learning format.

Conclusion

This study showed that after solving the main organizational issues of transferring learning to a distance format, it is necessary to pay attention to the cognitive-psychological side of the transition. The lack of social contact that promotes teacher-student interaction, which is inherent in face-to-face teaching, has led students to talk a lot about feelings of anxiety, loss of confidence in new conditions. The data obtained underscore the importance of a student-centered approach to organizing all forms of interaction at a university, which becomes especially critical in the forced circumstances of such a pandemic. In addition to the educational process, it is important to transform extracurricular and social activities into a hybrid form, providing students with the effect of presence, full interaction with teachers and fellow students.

Although at the stage of development and verification of the methodology the authors used elements of manual text processing, it can be assumed that the described methodology of social network analysis can be further automated using *PolyAnalyst* software, which is a logical development of the research topic. To do this, it is necessary to develop a methodology for fine-tuning nodes by editing the ontologies / dictionaries used in the analysis and compiling special queries. It is also necessary to solve the problem of automated sentiment is defined in sarcastic, ironic, and passive-aggressive texts. So far, there is very little research [34] on the definition of irony and sarcasm in the Russian language.

With this information, pain points can be managed in time and predictions can be made. That is why the next step in the development of this topic is the automation of the described methodology. User-generated data is an important and easily accessible source of public opinion that can successfully replace the sociological surveys familiar to us. The methodology refinement phase is used to work with data generated in the past. Once automated, analysing them in real time will provide a powerful tool for measuring tensions through indices and metrics, using this information to track tensions spikes and anticipatory management.

The study presents a conceptual framework for highlighting key aspects of student satisfaction assessment through a digital footprint on social media based on sentiment analysis. Through this, universities can gain a deeper understanding of the educational experience, identify areas for improvement and target interventions.

References

1. Bakker, A., Cai, J. & Zenger, L. (2021). Future themes of mathematics education research: an international survey before and during the pandemic. Educational Studies in Mathematics. No. 107, pp. 1-24, doi: 10.1007 / s10649-021-10049- w
2. Turnbull, D., Chugh, R. & Luck, J. (2021). Transitioning to E-Learning during the COVID-19 pandemic: How have Higher Education Institutions responded to the challenge? *Education and Information Technologies*. No. 26, pp. 6401-6419, doi: 10.1007/s10639-021-10633-w
3. Rapanta, C., Botturi, L., Goodyear, P. et al. (2021). Balancing Technology, Pedagogy and the New Normal: Post-pandemic Challenges for Higher Education. *Postdigital Science and Education*. No. 3, pp. 715-742, doi: 10.1007/s42438-021-00249-1
4. El-Sayad, G., Md Saad, N.H. & Thurasamy, R. (2021). How higher education students in Egypt perceived online learning engagement and satisfaction during the COVID-19 pandemic. *Journal of Computers in Education*. No. 8, pp. 527-550, doi: 10.1007/s40692-021-00191-y
5. Osina, D.M., Tolstopyatenko, G.P., Malinovsky, A.A. (2021). Digitalization of higher legal education in Russia in the era of COVID-19. In: Ashmarina, S., Mantoulenko, V., Vochozka, M. (eds.). *Engineering Economics: Decisions and Solutions from Eurasian Perspective*. Vol. 139, pp. 392-398, doi: 10.1007/978-3-030-53277-2_47
6. Shchekotin, E., Goiko, V., Myagkov, M., Dunaeva, D. (2021). Assessment of quality of life in regions of Russia based on social media data. *Journal of Eurasian Studies*. Vol. 12, no. 2, pp. 182-198. DOI: 10.1177/18793665211034185
7. Zielinski, A., Bügel, U. (2012). Multilingual analysis of Twitter news in support of mass emergency events. *Proc. of the Ninth International Conference on Information Systems for Crisis Response and Management (ISCRAM 2012)*, Vancouver, Canada, doi: 10.24406/publica-fhg-377702
8. Shchekotin, E., Myagkov, M., Goiko, V., Kashpur, V. (2021). Digital methods of analysis of subjective quality of life: Case of Russian regions. *Administratie si Management Public*. No. 36, pp. 25-48, doi: 10.24818/amp/2021.36-02
9. Yin, J., Lampert, A., Cameron, M.A., Robinson, B., Power, R. (2012). Using social media to enhance emergency situation awareness. *IEEE Intell Syst*. Vol. 27, no. 6, pp. 52-59, doi:10.1109/MIS.2012.6
10. Orlova, V., Goiko, V., Alexandrova, Yu., Petrov, E. (2020). Analysis of approaches to study identification in social media. *E3S Web of Conferences*. Vol. 210: 16034, doi: 10.1051/e3sconf/202021016034
11. Verma, S., Vieweg, S., Corvey, W., Palen, L., Martin, J., Palmer, M., Schram, A., & Anderson, K. (2021). Natural Language Processing to the Rescue? Extracting “Situational Awareness” Tweets During Mass Emergency. *Proc. of the International AAAI Conference on Web and Social Media*. Vol. 5, no. 1, pp. 385-392, doi: 10.1609/icwsm.v5i1.14119
12. Liu, B. (2010). Sentiment analysis and subjectivity. In Indurkhya, N., Damerau, F.J. *Handbook of Natural Language Processing, Chapman & Hall/CRC Machine Learning & Pattern Recognition Series*. Florida: Taylor & Francis Group, Boca Raton, pp. 627-666. Available at: https://www.routledge.com/Handbook-of-Natural-Language-Processing/Indurkhya-Damerau/p/book/9781420085921 (accessed 22.12.2022).
13. Pang, B., Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*. Vol. 2, no. (1-2), pp. 1-135, doi: 10.1561/1500000011
14. Treude, C., Barzilay, O., Storey, M.A. (2011). How do programmers ask and answer questions on the web? (nier track). *Proc. of the 33rd International Conference on Software Engineering, ICSE*. Vol. 11, pp. 804-807, doi: 10.1007/s11390-016-1672-0

15. Mezouar, M.E., Zhang, F., Zou, Y. (2018). Are tweets useful in the bug fixing process? An empirical study on firefox and chrome. *Empirical Software Engineering*. Vol. 23, no. 3, pp. 1704-1742, doi: 0.1007/s10664-017-9559-4
16. Aniche, M., Treude, C., Steinmacher, I., Wiese, I., Pinto, G., Storey, M.A., Gerosa, M.A. (2018). How modern news aggregators help development communities shape and share knowledge. *Proc. of the 40th International Conference on Software Engineering, ICSE ' 18*. Pp. 499-510, doi: 10.1145/3180155.3180180
17. Sharma, U., Datta, R.K., Pabreja, K. (2020). Sentiment Analysis and Prediction of Election Results 2018. In: Shukla, R., Agrawal, J., Sharma, S., Chaudhari, N., Shukla, K. (eds). *Social Networking and Computational Intelligence. Lecture Notes in Networks and Systems*. Vol. 100, pp. 727-739, doi: 10.1007/978-981-15-2071-6_61
18. Georgiadou, E., Angelopoulos, S., Drake, H. (2020). Big data analytics and international negotiations: Sentiment analysis of brexit negotiating outcomes. *International Journal of Information Management*. Vol. 51, no. 2: 102048, doi: 10.1016/j.ijinfomgt.2019.102048
19. Si, X.-M., Li, C. (2018). Bounded confidence opinion dynamics in virtual networks and real networks. *Journal of Computers*. Vol. 29, no. 3, pp. 220-228, doi: 10.1016/j.ijinfomgt.2019.102048
20. Kahani, N., Bagherzadeh, M., Dingel, J., Cordy, J.R. (2016). The problems with eclipse modeling tools: A topic analysis of eclipse forums. *Proc. of the ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems, MODELS ' 16*, pp. 227-237, doi: 10.1145/2976767.2976773
21. Squire, M. (2015). Should we move to stack overflow? Measuring the utility of social media for developer support. *Proc. of the 37th International Conference on Software Engineering, ICSE' 15*. Vol. 2, pp. 219-228, doi: 10.1109/ ICSE.2015.150
22. Lazar, J. (2021). Managing digital accessibility at universities during the COVID-19 pandemic. *Universal Access in the Information Society*. Vol. 21, pp. 749-765, doi: 10.1007/s10209-021-00792-5
23. Ligthart, A., Catal, C. & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. *Artificial Intelligence Review*. Vol. 54, no. 2, pp. 1-51, doi: 10.1007/s10462-021-09973-3
24. Li, K.C., Wong, B.Tm. (2021). The Opportunities and Challenges of Social Media in Higher Education: A Literature Review. *SN Computer Science*. No. 2: 455, doi: 10.1007/s42979-021-00857-5
25. Purcell, W.M., Lumbreras, J. (2021). Higher education and the COVID-19 pandemic: navigating disruption using the sustainable development goals. *Discover Sustainability*. Vol. 2: 6, doi: 10.1007/ s43621-021-00013-2
26. Abramova, M.O., Barannikov, K.A., Gruzdev, I.A. et al. (2021). Quality of Education in Russian Universities: What We Understood in the Pandemic: Analytical Report / sci. ed. by E.A. Sukhanova, I.D. Froumin. Tomsk: Tomsk State University Press, 46 p. Available at: https://publications.hse.ru/pubs/share/direct/506461902.pdf (accessed: 18.06.2023). (In Russ.).
27. The most popular social networks in Russia. *Kommersant* Newspaper. No. 174/П of 27.09.2021, p. 10. Available at: https://www.kommersant.ru/doc/5006356 (accessed: 22.12.2022). (In Russ.).
28. Demchenko, N., Makarov, I. *Instagram Bypassed "VKontakte" in the Number of Active Authors in Russia*. Available at: https://www.rbc.ru/technology_and_media/30/11/2020/5fc4aa2c9a7947f4bfbce6db# (accessed: 22.12.2022). (In Russ.).
29. Magdin, M., Balogh, Z., Reichel, J. et al. (2021). Automatic detection and classification of emotional states in virtual reality and standard environments (LCD): comparing valence and arousal of induced emotions. *Virtual Reality*. No. 25, pp. 1029-1041, doi: 10.1007/s10055-021-00506-5

30. Sigaeva, E. *Speech Paradoxes: We, You or is it Me?* Available at: https://www.b17.ru/article/5993/ (accessed: 22.12.2022). (In Russ.).
31. Patent No. 2769644 the Russian Federation, МПК G06Q 50/20 (2012.01), СПК G06Q 50/20 (2021.08). *Online Higher Education System*: No. 2021117849, applied 19.06.2021: published: 04.04.2022 / Krishtal, M.M. , Boyur, R.V., Baboshina, E.S., Kutuzov, A.I., Sokolova, T.A., Drozdova, M.A., Repina, E.A., Denisova, O.P., Bogdanova, A.V., Khamidullova, L.R., Gasanova, R.M.; Patent Holder: Togliatti State University. 44 p.
32. Krishtal, M.M., Boyur, R.V., Baboshina, E.S., Kutuzov, A.I., Sokolova, T.A., Denisova, O.P. (2021). Experience of building an online higher education system (Rosdistant project). Prerequisites for the emergence of the project. P. 1. *Standarty i Kachestvo = Standards and Quality*. No. 11, pp. 52-56, doi: 10.35400/0038-9692-2021-11-52-56 (In Russ., abstract in Eng.).
33. Krishtal, M.M., Boyur, R.V., Baboshina, E.S., Kutuzov, A.I., Sokolova, T.A., Denisova, O.P. (2021). Experience of building an online higher education system (Rosdistant project). The main institutional changes and effects. P. 2. *Standarty i Kachestvo = Standards and Quality*. No. 12, pp. 82-87, doi: 10.35400/0038-9692-2021-12-82-87 (In Russ., abstract in Eng.).
34. Zefirova, T., Lukachevich, N. (2019). Expression of irony and sarcasm on Twitter. *Proc. of the Third Workshop "Computational linguistics and language science"*. Vol. 4, pp. 45-49, doi: 10.29007/tpzw

The paper was submitted 05.12.2023
Accepted for publication 22.01.2024