

Implementation of Text Mining in Socio-Economic Research

Konstantin Malyshenko, Humanitarian and Pedagogical Academy, Crimean Federal University Named After V.I. Vernadsky, Russia*

Vadim Malyshenko, Humanitarian and Pedagogical Academy, Crimean Federal University Named After V.I. Vernadsky, Russia

Marina Anashkina, Humanitarian and Pedagogical Academy, Crimean Federal University Named After V.I. Vernadsky, Russia

Dmitry Anashkin, Crimean Federal University Named After V.I. Vernadsky, Russia

ABSTRACT

This work aims to analyze insights from social networks for identification of population satisfaction with pay level in Russia using the text mining approach. For this, a sentiment analysis framework was developed, which integrates Twitter mining tools and a sentiment index. Sentiments were extracted using Twitter mining and then recoded and substituted into the sentiment formula. The results of sentiment analysis indicate low satisfaction with levels of pay among Russians. Twitter was chosen as the object of research, as one of the most active and independent networks in Russia. It is possible that some of the tweets belong to authors who are not living in Russia at the moment, but their number is not significant and their interest in this issue, in the authors' opinion, only enhances the relevance of the problem under study.

KEYWORDS

Big Data, Content, Data Mining, Mood Index, Sentiment Analysis

INTRODUCTION

In modern conditions, the mood of the population forms the basis of many spheres of society. The influence of this indicator can be traced in the financial and economic spheres.

For example, in their work "Early warning indicators? The effect of consumer and investor sentiments on the restaurant industry," Yost et al. (2020) examine the impact of cyclical fluctuations in the consumer confidence index and the volatility index as early warning indicators of changes in the financial performance of restaurant business entities. Wu et al. (2019) focus on the relationship of consumer spending with changes in the confidence index. The results show that planning consumption growth by considering the mood of consumers and businesses leads to many economic benefits.

The mood (satisfaction) of the population is a set of citizens' opinions regarding the studied object. In this study on population satisfaction, we consider the average assessment of respondents' views obtained by processing messages posted on a social network regarding wage levels (Juhro & Iyke, 2020; Peláez et al., 2020; Yan et al., 2019).

DOI: 10.4018/IJBDCN.341263

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

In world practice, special leading indicators (confidence indices) are actively used for these purposes, based solely on data obtained through a mass survey of citizens. Various economic actors use these indicators to develop effective development policies, but the state apparatus can also use them to assess social tension and economic confidence.

The development of modern technologies opens up a wide range of opportunities for analyzing various spheres of society, most of which have moved partially or entirely to the Internet environment. The massive volume of users generates an increase in information flows, resulting in the possibility of processing and extracting information, which is impossible to obtain during the application of classical methods of text processing and data analysis, becoming of enormous importance. Social networks are a vast information base regarding the population's opinion on various issues today.

OBJECTIVES OF THE STUDY

The effectiveness of the above methods for determining consumer sentiment at the present stage of society's development is declining. Traditional surveys and questionnaires are gradually moving to the Internet. Modern users often try to avoid this form of opinion assessment, which results from the population's distrust due to the growth of Internet fraud. The same applies to traditional contact methods for questionnaires. At the same time, conducting research based on statistical data may not always reflect the actual state of the issue under study due to the influence of numerous factors. Among these are the inability to fully cover subjects' economic activities, their underestimation of the results of their activities, and the growth of the shadow economy. In addition, a significant factor involves the retrospective nature of statistical data. According to some indicators, collecting and processing information takes a long time—up to six months. Thus, a study based on statistical data allows us to obtain a result reflecting consumer sentiment only over past periods. At the same time, to develop effective management decisions in various spheres of society's economic and social life, it is necessary to have a current idea of this indicator.

Notably, in Russian practice, the use of big data processing programs to determine the population's opinion and assess the degree of satisfaction with the economic situation and confidence in state policy in this area has been poorly studied. However, such studies are actively conducted all over the world. Many foreign scientists have studied the possibilities, features, and effectiveness of the application of big data.

Thus, at the moment, no methods exist for using text mining technologies in Russia to determine the population's satisfaction. The possibilities of adapting technologies to work with the Russian text have not been studied, which does not allow for the development of their application in the socio-economic sphere of the country.

The purpose of this work is to assess the population of the Russian Federation's satisfaction with their incomes based on the analysis of social network data using text mining tools. Studying this issue will allow us to get an idea of the features of the use of text mining technologies in the processing of texts from Russian-language sources, as well as to develop a methodology for building on this basis the wage satisfaction index as one of the components of the composite index of economic sentiment.

LITERATURE REVIEW

To solve the problems of analyzing information on the Internet and extracting users' opinions, researchers have begun actively developing big data technologies; in particular, they are investigating and expanding the possibilities of their application for conducting socio-economic research. These technologies comprise a set of various methods, techniques, and tools, of which data mining methods are of the greatest interest for the study of socio-economic phenomena (Alarifi et al., 2020; El Alaoui et al., 2019; Kannan & Kothamasu, 2020; Odendaal et al., 2020; Riahla et al., 2021; Song & Shin, 2019).

Foreign scholars have studied methods of intellectual text analysis. For instance, Abulaish et al. (2019) present a language-independent approach to graph-based tonality analysis, SentiLangN, which uses a symbolic n-gram graph to model text data for processing language-independent unstructured expressions.

Hadi et al. (2019) offer an effective technique for analyzing tonality in the context of big data. First, the collected data is cleaned using the preprocessing method. Then, the authors select optimal functions using a universal algorithm approach. Their research is based on the use of artificial neural networks. Rodrigues and Chiplunkar (2019) propose a hybrid lexicon and naive Bayesian classifier method to conduct sentiment analysis. The suggested method is compared to a naive Bayesian classifier for unigram and bigram features.

Jadon et al. (2019) note that analyzing Twitter data in real time can play an essential role in observing users' thinking and point of view. A cumulative point of view by country can help form a general opinion of various countries' citizens and the criteria for their thinking. The authors propose a model of sentiment analysis to monitor the positive and negative points of view of different countries based on the sentiment analysis approach. Alim and Shukla (2020) offer a big data analysis framework for analyzing human behavior using sampling techniques. The study was conducted based on data from social networks.

Moussa (2019) presents a new metric for tracking consumer emotions concerning various brands. The author analyzed consumer messages from official Twitter accounts. This method of sentiment assessment differs from the traditional one, which is based on text analysis and sentiment extraction. This method is specific because it mainly allows the analysis of information from social networks.

Shehu and Tokat (2020) used the Twitter API to analyze opinions based on 13,000 Turkish tweets. The authors used a machine learning method. They applied preprocessing methods to the received data to remove references, numbers, punctuation marks, and meaningless symbols. The results show that to obtain the most accurate result, both manual text processing methods and methods based on machine learning must be used. Yang et al. (2020) analyze the mood of the text of online reviews of cigarettes to determine their impact on demand. As a result of the research, the authors developed a dictionary of tonality and built a model of sentiment analysis. Alqmase et al. (2021) consider methods of analyzing text in social networks for the presence of fanatical attitudes. Based on Arabic texts, the authors classify the received texts according to the emotional component. Cortis and Davis (2021) investigate public opinion based on artificial intelligence. The authors note that an essential element in such studies is the identification of many aspects of views, moods, and emotions, which contributes to improving the accuracy of the assessment. Radygin et al. (2021) study the application of intellectual analysis to Russian-language texts. The authors offer a set of programs for a narrow subject area (violations of the Federal Procurement Law). The paper notes that text preprocessing (summation and lemmatization) is a mandatory step; ignoring it will decrease the accuracy of the results. Based on data from Russian social networks, Lobantsev et al. (2020) investigate the effectiveness of passive and active data collection methods for further analysis. In the article "Extraction of structural elements of inventions from Russian-language patents," Korobkin et al. (2019) explore the problem of extracting structured data from Russian-language patents in synthesizing new technical solutions. The authors examine existing natural language processing tools in relation to patent processing and propose a new method for extracting predicate–argument constructions, considering the specifics of the patent text, based on shallow sentence parsing and segmentation. Fomin et al. (2019) offer a method for classifying scientific documents, materials, and articles published in Russian. This study allows the use of data mining for Russian-language scientific sources. Kotelnikov et al. (2018), in their work "A Comparative Study of Publicly Available Russian Sentiment Lexicon," also raise the problem of determining moods in the analysis of opinions and note that the analysis of Russian-language texts has its peculiarities. Accounting for unions, in this case, has a significant role. Bobichev et al. (2017) explore the possibility of analyzing sentiments in Ukrainian and Russian news. The authors develop a corpus of Russian and Ukrainian news, dividing it into three categories.

Al-Obeidat et al. (2019) address the problem of high data variability on consumer satisfaction with certain products due to their short validity period. The research focuses on studying existing sentiment analysis methods and developing a new approach considering the specific characteristics of the chosen industry. The authors present a new sentiment analysis structure with a scaling technique that uses data mining strategies to obtain, identify, and analyze fast fashion social networks to determine customer satisfaction.

Arora et al. (2019) offer a mechanism for measuring the index of influencers in popular social networks, including Facebook, Twitter, and Instagram. Srivastava et al. (2019) present a hybrid approach using naive Bayesian analysis to study Twitter datasets. Mathews and Abraham (2019) focus on determining the polarity of words entered by various users in their reviews. Takhanova (2018) explores the phenomenon of trust as a determining factor contributing to the strengthening of social capital. The authors conducted a review of research on the classification of types of trust by domestic and foreign authors. They note that the existing methods of assessing interpersonal trust are based on sociological survey methods. The paper presents an assessment of the measurement of interpersonal trust using the method of economic calculation. Amirkhanova and Bikmetov (2017) undertake an interdisciplinary theoretical analysis of the content and conditions of trust.

Thus, various authors have deeply researched the use of text mining tools; in Russia, however, this direction is just beginning to develop. At the same time, the study and development of methods and approaches for building a trust index based on unsolicited data extracted from social networks and their adaptation to work with Russian text is an urgent and vital research direction.

Podoprigora (2016) examines the transformation of the sources and decisive factors of the dynamics of the economy and society at the present stage of development of post-industrial society. The work reveals the genesis, interrelation, and influence of such socio-psychological phenomena as trust and the paradigm of moral behavior on the nature and effectiveness of economic processes in the conditions of the dominance of the “creative industry.”

RESEARCH DESIGN

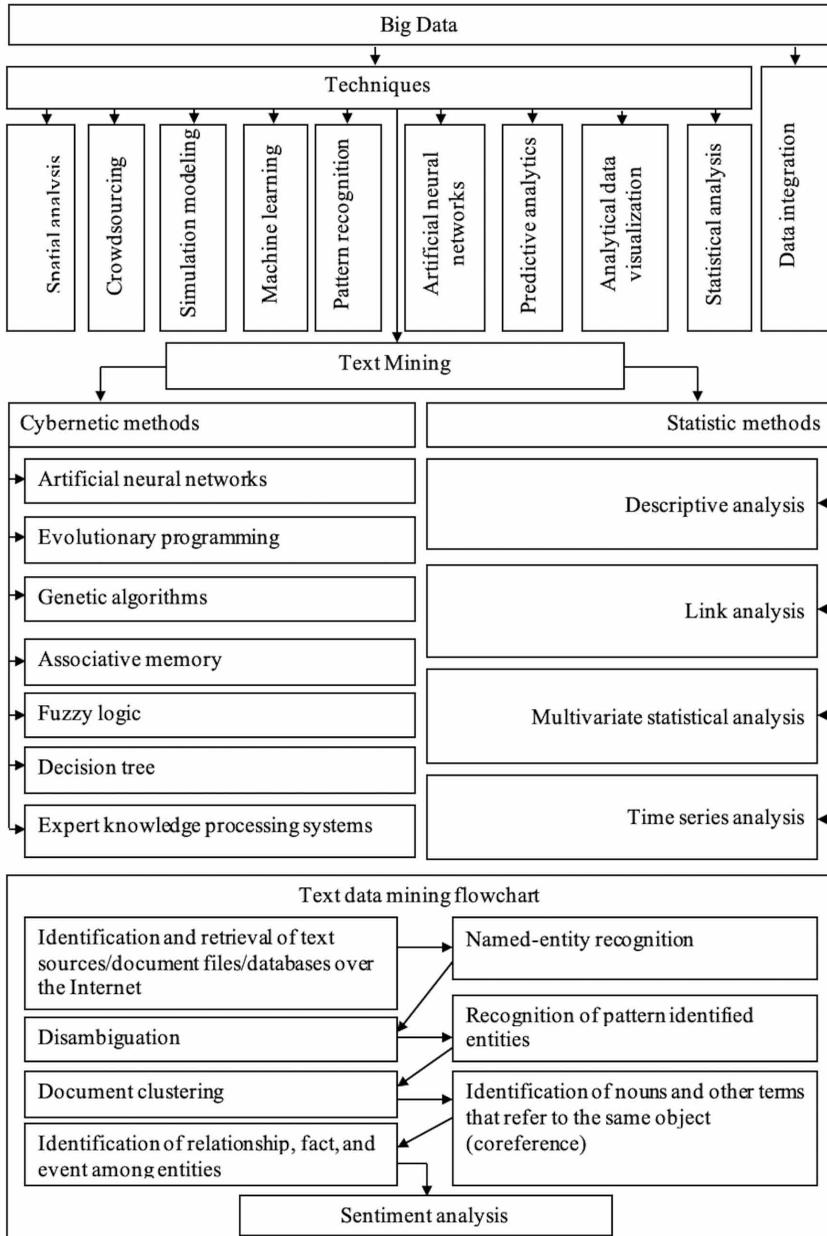
The term “big data” today has many different definitions. Initially, big data was understood as data of a massive volume, the analysis and management of which require significant material and technical investments.

With the development of technologies that can process large amounts of data, the term gradually began to be applied directly to them. Today, in most foreign studies, this term is understood as the object of research, while in Russian practice, technologies are increasingly being understood. In this paper, big data refers to a set of methods, technologies, and tools for processing large amounts of data to extract structured information that is not extracted by applying traditional approaches and methods of analysis. These methods’ main advantages are the ability to process vast volumes of various data, including constantly increasing and changing data, and to work with unstructured and poorly structured data. The classification of big data techniques is presented in Figure 1. As seen in the figure, we have allocated a group of text mining methods in a separate block, which formed the basis of this study. Text mining is a methodology for extracting previously unknown information from information flows and data arrays that can be used practically.

In Figure 1, we divide text mining into cybernetic and statistical methods. The first group is a combination of approaches from computer mathematics and artificial intelligence, which allows for a deep analysis of the text and the identification of hidden relationships.

To assess the Russian population’s satisfaction with wage levels, we used the Orange program, a data visualization and analysis tool. This open-source program has the necessary tools for conducting this research. The program tools used in this study belong to the cybernetic group. We used a sentiment analysis tool based on artificial neural networks.

Figure 1. Classification of big data techniques
 (Note: Chart developed by the authors)



POPULATION AND SAMPLING TECHNIQUE

We can now proceed to the study of the degree of satisfaction of the population of the Russian Federation with the current wage levels. Using the Twitter widget of the Orange program, we formed a database of messages from the social network Twitter for three months (July 1–October 1, 2020), selected by the tags “salary in Russia” and “salary in Russia.” We chose these tags because wages are the primary source of income for most citizens and determine their socio-economic position

in society. The size of citizens' salaries affects their confidence in the near future; therefore, the assessment of salary satisfaction can be used to construct an indicator of the economic mood of the population. Based on this indicator, in the future, we can forecast such economic indicators as the volume of consumption of goods and services, supply and demand for certain types of goods, and lending volumes. Including references to Russia ("RF" and "Russia") in the body of the tag allowed us to limit the search results in a certain way, selecting only those tweets in which Russian-speaking users touched on the topic of salaries in Russia. Thus, tweets concerning wages in other countries where Russian-speaking users may reside were excluded from the sample.

The total number of tweets during this period was 711. This period is not tied to any particular event and covers the content produced preceding the start of direct processing by TextMining technologies.

Twitter users post short text messages daily, based on which it is possible to determine their attitude to a particular research subject. This social network is the most popular among the networks in Russia; each active user posts an average of 47 messages per month. For comparison, in the next most popular platforms, Instagram and VKontakte, active users posted, on average, 6.2 and 17 messages per month in 2020, respectively. Another reason for choosing this social network for research is that users can freely publish messages reflecting their opinions. Located in the so-called "legal vacuum" of Russian territory, this company usually does not respond to the regulator's requirements to remove certain content. Although this is a negative factor for the country, it opens up research opportunities. Thus, the data obtained from Twitter directly reflects users' opinions. This ultimately ensures greater reliability of the results obtained. Unlike other social networks (such as Instagram, where users prefer to post images), Twitter is dominated by text content, which is essential for our analysis. Social networks such as Facebook, Telegram, and VKontakte are most likely to publish news, entertainment, and informative posts. Communication in these social networks occurs mainly in groups or conversations that can be closed to most users or directly by leaving comments under a specific news item. At the same time, Twitter provides open communication among users. Thus, this social network covers individual users and not their interest groups. Obtaining data from the social network Twitter is also possible due to particular components in the program that automate data collection, eliminating the influence of the human factor and ensuring the study's objectivity.

We carried out the study in several stages (Figure 2), and we can consider the presented methodology in more detail. In the first stage (Figure 2, block 1), we selected tweets according to the specified parameters on the topic "wages." The selection was carried out using the Twitter widget. With this widget, one can create a database of tweets by content, author, or both parameters at once.

If further data expansion is needed, the widget provides for the accumulation of results.

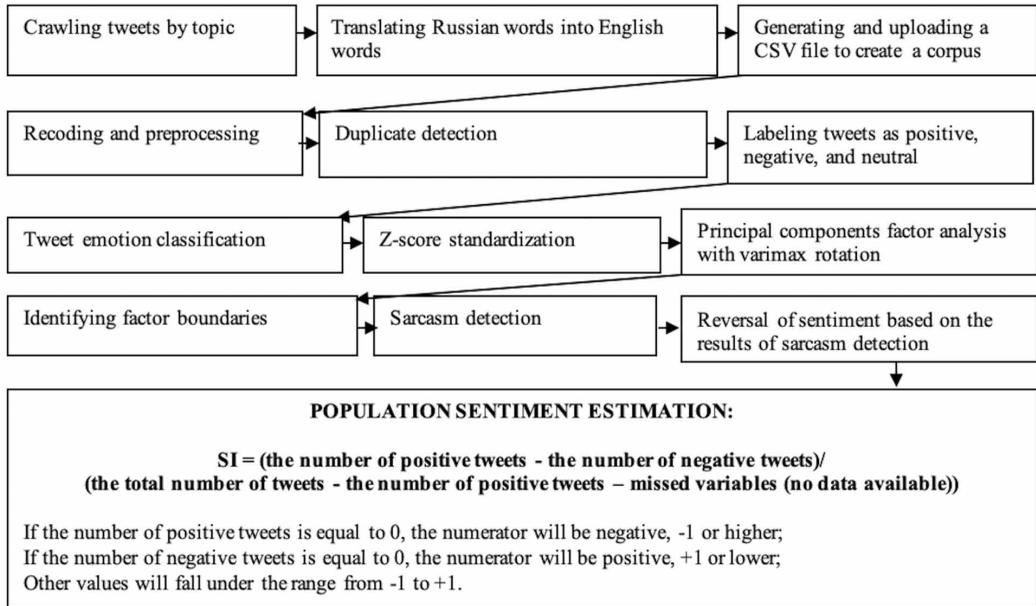
DATA COLLECTION INSTRUMENT

To activate the widget, users must create an account and get the key and secret code required to complete the form. To create a database, the parameters must be defined. The program requires users to provide a list of keywords and select search filters (by author and/or content), language, and the upper limit of tweets (the maximum amount of information). It also includes the additional parameter "allow retweets."

Query execution allows users to collect information on the specified keywords, content, and authors. The "collect results" box must be checked to update the database.

To continue the process, the generated database must be translated into English (Figure 2, block 2). It does not include a Russian dictionary of sentiment, and our program does not have one either. This is a disadvantage since the Russian language differs significantly from English and may contain various proverbs, stable expressions, and slang. Accordingly, the idiom of an entire expression can change from positive to negative. This is especially true of sarcastic (and, to a lesser extent, ironic)

Figure 2. Flowchart of sentiment estimation process
 (Note: Flowchart developed by the authors)



expressions. To solve this problem (to some extent), we used a more accurate classification by Plutchik (2001), which we discuss in more detail in stage 7.

Initially, we aimed to form and use a training sample to evaluate Russian-language tweets, with their subsequent classification depending on the emotional load. However, this turned out to be impossible for the same reason: Orange cannot “read” tweets in Russian.

The third stage involves purely technical actions, namely, converting the tweet database (in English) into a file type readable by the program (.csv) to form the body of the text (Figure 2, block 3). After that, we uploaded it to the Ogapdes, indicating the type and role of variables.

The fourth step entails recoding and preprocessing (Figure 2, block 4). This step involves clearing the text and converting emoticons into strings (words), as shown in Table 1.

We conducted the recoding directly using the Orange tools. Figure 3 shows the workflow and the results of this utility.

Next, we carried out a standard preprocessing procedure, including various actions:

- Tokenization of the corpus (splitting the text into small blocks or tokens),
- Text filtering
- Normalization (definition of boundaries, lemmatization)
- Creating an n-gram
- Marking tokens with part-of-speech tags

The listed steps are applied in a specific sequence but can be enabled or disabled.

In the fifth stage (Figure 2, block 5), we removed repetition in tweets—that is, messages regularly published by robots. This operation can also be performed with SPSS tools, but Orange has a convenient tool specially created for this. The program provides clustering, allowing users to find text similarities in a document.

Table 1. Emoticons and related emotional content of the tweet

| Emoticon | Replacement string | Emoticon | Replacement string |
|----------|-----------------------------|----------------------|---|
| :) or =) | smiley_face |]:-< | devil |
| :D | laughing | :-[| embarrassed; shy_face |
| :(| sad_face | @->- or @ }->-- | rose |
| :) | winking_face | <:-(>:-(>:-[| sorrowful_face frowning_face mad_face |
| :- | neutral_face | :-x | dirty_language |
| :-\ | confused_face; annoyed_face | 8-) | cool |
| :-* | kissing_face | :-! | nauseated_face; disgust |
| :P | tongue_sticking_out | :-< | pensive_face |
| :-o | face_with_open_mouth | :-t | angry_face |
| =-o | astonished_face | :3 | face_with_melt_emotion |
| O:-) | angel; saint; halo | | |

To determine the similarity, the vertical line must be moved in the open visualization window. The distance of the line installation corresponds to the degree of similarity of the messages: The further it is exposed, the more coincidences they include. This threshold can also be set manually in the control area (Figure 4).

When determining the parameters, the type of communication must be determined (Figure 5, block 2). The program provides the following types of communication:

- single
- average
- full
- weighted
- according to Ward

Figure 3. Flowchart of text recoding and preprocessing

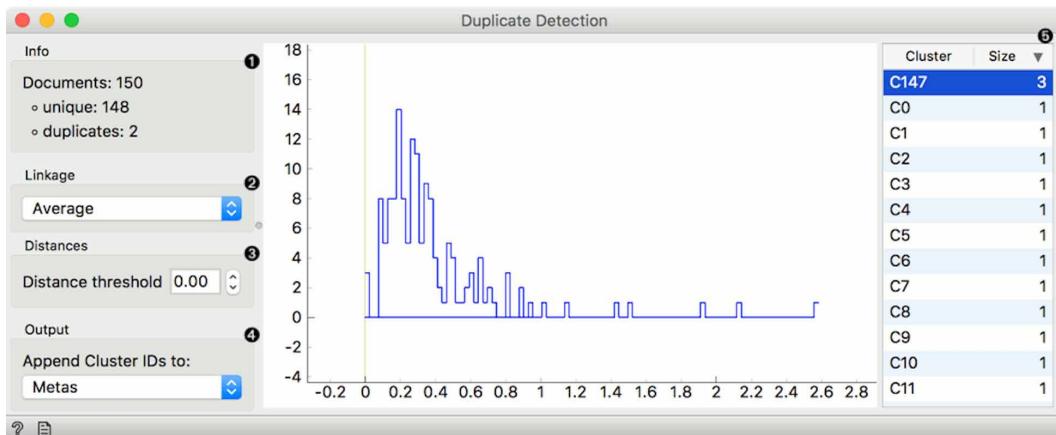
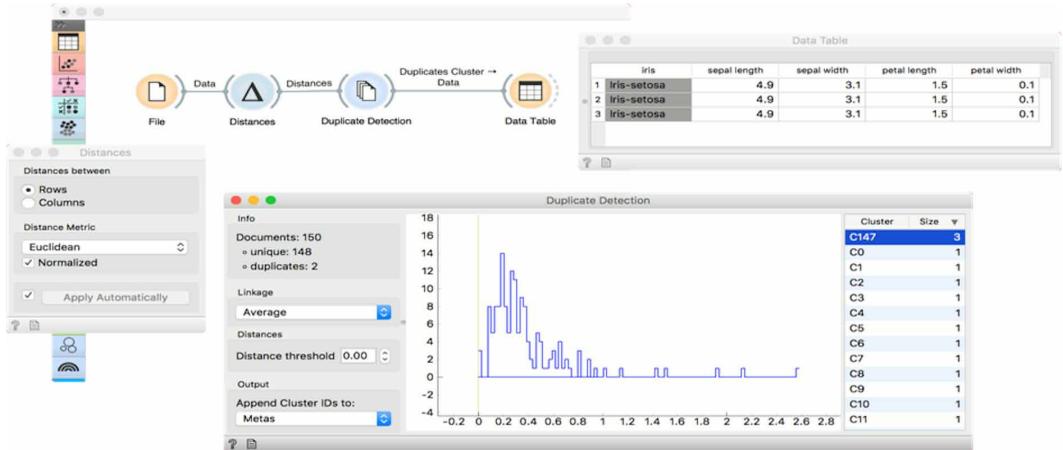


Figure 4. Duplicate detection in orange



By default, the program sorts clusters by size. To view them, users must click on the cluster of interest. The window displays its contents (Figure 5).

Orange allows users to make the case smaller to make it easier to visualize. We selected only about seven hundred tweets for our study, so we did not need to use this function.

After identifying and deleting duplicate tweets, we proceeded to the definition of sentiment. In the sixth stage, we used Orange to determine sentiment, using the Vader algorithm, compound index (Figure 2, block 6).

In the text analysis process, we determined the tonality of each tweet. The program offers three algorithms of sentiment analysis: LiuHu, Vader, and Multi-lingual sentiment. Each of these mood models is based on vocabulary (Liu, 2020).

We recoded the sentiment with Vader (based on a simple scale: negative values—negative, positive values—positive, zero—neutral) using SPSS tools.

In the next stage (Figure 2, block 7), we identified sarcasm based on Plutchik’s assessments using Orange tools for eight emotions (anger, disgust, fear, humor, sadness, surprise, trust, and expectation). The Tweet Profiler in Ogap receives mood information from the server for each given tweet (or document). The widget sends the data to the server, where the model calculates the probabilities of emotions and estimates. The widget supports three classifications of emotions, namely Ekman, Plutchik, and profile of mood state (POMS).

Plutchik (2001) first proposed his cone-shaped model (3D) or wheel model (2D) in 1980 to describe the relationships among emotions. He suggested eight basic bipolar emotions: joy vs. sadness, anger vs. fear, trust vs. disgust, and surprise vs. expectation. In addition, his circumplex model combines the idea of a circle of emotions and a color circle. Like colors, basic emotions can be expressed with different intensities and can combine to form different emotions.

Plutchik (2001) proposed this theory to explain the basic protection mechanisms of the human psyche. He suggested eight elements of psychological protection based on the eight basic emotions.

The study’s next step involved normalizing the data by calculating standard z-scores for each variable (Figure 2, block 8). Following this operation, we carried out a principal components analysis using the SPSS software. Standardization is crucial if variables ranging over different orders of magnitude lead to a common denominator and if single statistical methods are used. In this case, we first had to standardize the variables and then compute the mean scores: Z-score(Anger), Z-score(Fear), Z-score(Trust), Z-score(Anticipation), Z-score(Surprise), Z-score(Sadness), Z-score(Joy), and Z-score(Disgust). The z-scoring process can be expressed as:

Figure 5. Exploring the content of the cluster

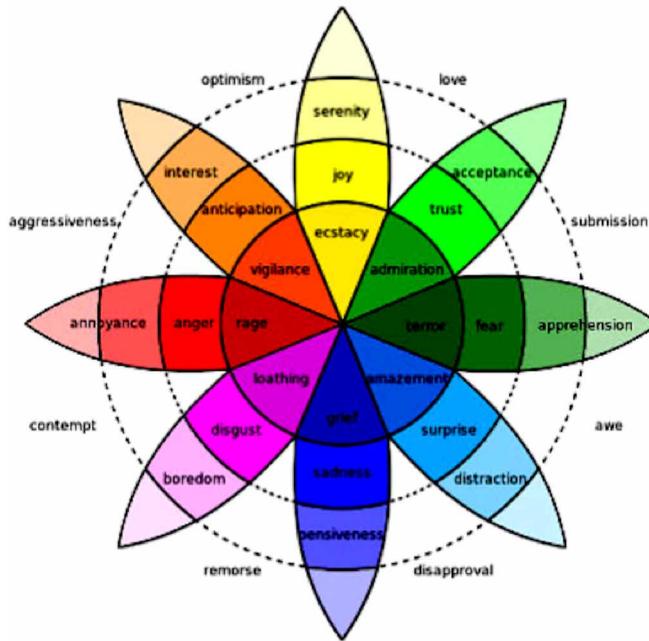
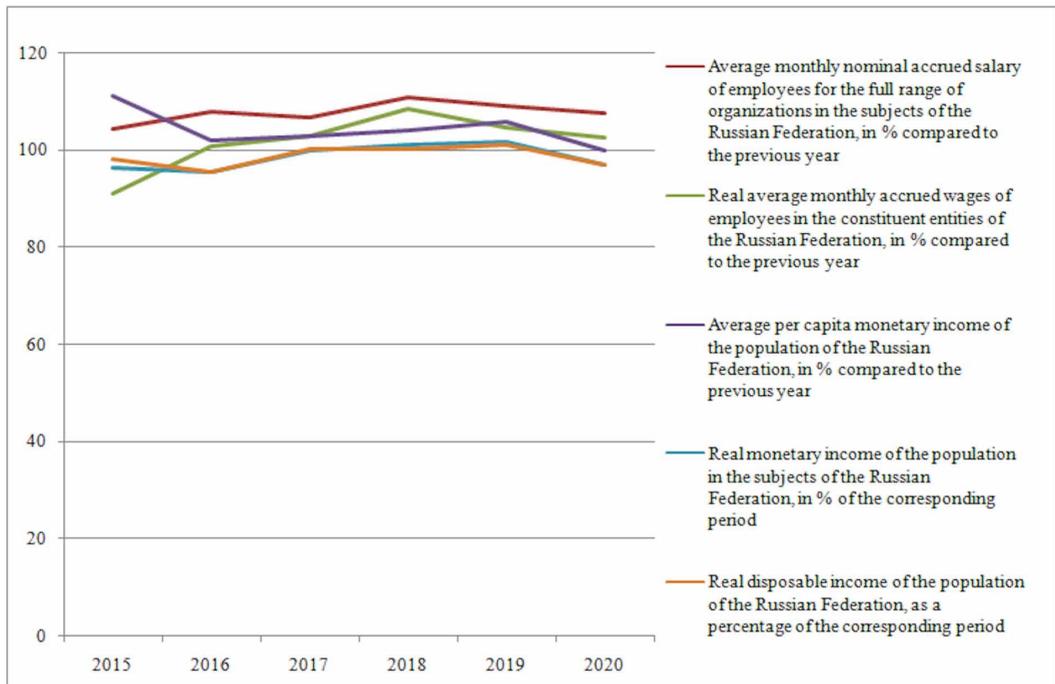


Figure 6. Plutchik's wheel of emotions
 (Source: Plutchik, 2001)



$$z = (x - m)/S \tag{1}$$

where m is the mean value of the variable, and S is the standard deviation.

In the ninth stage, we determined the factors (emotions) included in these components (Figure 2, block 9), identifying the factors responsible for sarcasm in messages (rotation method, Varimax, SPSS). This stage aims to discover which factors may be responsible for the manifestation of sarcasm. To do this, we performed a factor analysis using SPSS tools.

We calculated Pearson correlation coefficients between the variables under consideration. The correlation matrix becomes the basis for subsequent calculations. To construct it, the eigenvalues and eigenvectors must be determined using the estimated values of the diagonal elements of the matrix—the relative variances of simple factors.

When selecting factors by the number of eigenvalues, these values are sorted in descending order, while the corresponding eigenvectors form these factors. They can be taken as correlation coefficients between variables and factors.

To solve the task set before us, it is necessary to apply the method of determining the main components. The calculation steps presented above do not allow us to obtain a specific solution to the problem of determining factors. The search for an unambiguous solution based on the geometric representation of the problem is called the problem of rotation of factors. Many methods exist, but the most popular is the Varimax method, which determines which factor loads of a rotated matrix can be considered as a result of the factor analysis procedure. The factors are interpreted based on the values obtained.

If the interpretation of factors is possible, factor values are assigned to individual observations as the final step of factor analysis. As a result, each observation of the values of a large number of variables is translated into the values of a few factors.

As a result of the above procedures, we obtained Table 2.

Thus, we identified three components: 1—the component included factors responsible for negativity (anger, fear, trust, expectation, surprise); 2—the component included only one factor (humor, joy), mainly responsible for positivity; 3—the component included sadness and disgust, having a neutral value.

DATA ANALYSIS

Next, we determined the boundaries of the selected factors included in sarcasm (disgust, humor, sadness) and their criteria boundaries (based on the chosen quartiles; Figure 2, block 10).

The median is usually used to determine the boundaries of the class. The median is the value in the center of all values when ordered by magnitude. Based on this definition, we can characterize the concept of percentiles, which allows us to determine the values below which 10% of the values lie and, accordingly, above which the remaining 90% lie. The percentage is set at 10%. As a result of the calculation, the following results were obtained (see Table 3).

We determined the boundaries of the selected factors included in sarcasm (disgust, humor, sadness) and calculated their criterion boundaries (based on the chosen quartiles).

Based on the data presented in Table 4, disgust (disgust) had only positive values (above 0,0180118); joy (humor, joy) had positive values (above 0.3680564); sadness (sadness) had positive values (above 0.3202377).

Based on the determined factors, we identified tweets containing sarcasm (Figure 2, block). The boundaries of sarcasm can be expressed as follows:

Table 2. Rotated component matrix

| Variable | Component | | | Factor |
|-----------------------|--------------|--------------|---------------|--------------|
| | 1 | 2 | 3 | |
| Z-score(Anger) | 0.728 | 0.201 | 0.063 | Anger |
| Z-score(Fear) | 0.687 | 0.253 | 0.056 | Fear |
| Z-score(Trust) | 0.609 | 0.214 | 0.030 | Trust |
| Z-score(Anticipation) | 0.578 | 0.102 | -0.257 | Anticipation |
| Z-score(Surprise) | 0.559 | -0.309 | -0.225 | Surprise |
| Z-score(Sadness) | -0.146 | -0.916 | -0.080 | Sadness |
| Z-score(Joy) | -0.668 | 0.703 | -0.183 | Joy |
| Z-score(Disgust) | 0.041 | -0.022 | 0.953 | Disgust |

Extraction method: principal components analysis
 a. 3 components extracted

$$if \left\{ \begin{array}{l} ZSadness \geq 0.018 \\ ZJoy \geq 0.35 \\ ZDisgust \geq 0.320 \end{array} \right\} \rightarrow \text{the polarity of a tweet is flipped to negative} \quad (2)$$

Finally, if we detected that the tweet classified as positive or neutral in step 6 contained sarcasm, we reversed the tweet's sentiment to a negative one (Figure 2, block 12). The results are depicted in Table 4.

Knowing the accurate proportion of positive, negative, and neutral sentiments in tweets, one can compute the sentiment index (SI) as follows:

$$SI = \frac{\text{Number of positive tweets} - \text{number of negative tweets}}{\text{Total number of tweets} - \text{number of neutral tweets} - \text{missed variables (no data available)}} \quad (3)$$

The SI scores are interpreted on a scale from -1 to +1 (Likert, 1932), as shown in Table 5. We can compute the sentiment index using the above formula:

$$SI = (198 - 372) / (711 - 141 - 0) = -174 / 570 = -0.305$$

The resulting SI value is relatively low, which suggests that the population of the Russian Federation is unhappy with the pay levels. Pay level dissatisfaction primarily affects consumption volumes and consumer preferences. Namely, citizens with an insufficient amount of funds at their disposal are forced to live under purchase limits and carefully decide what they need most. They may also apply for a bank loan if the situation demands it. Hence, dissatisfaction with pay level is likely to correlate with the lending volume.

Table 3. Descriptive statistics for all variables under investigation

| | | Zscore (anger) | Zscore (disgust) | Zscore (fear) | Zscore (joy) | Zscore (sadness) | Zscore (surprise) | Zscore (trust) | Zscore (anticipation) |
|----------------------------|------------|-------------------------|-------------------------|-------------------------|--------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| N | Valid | 711 | 711 | 711 | 711 | 711 | 711 | 711 | 711 |
| | Missed | 35 | 35 | 35 | 35 | 35 | 35 | 35 | 35 |
| Mean | | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 |
| Standard error of mean | | 0.03750293 | 0.03750293 | 0.03750293 | 0.03750293 | 0.03750293 | 0.03750293 | 0.03750293 | 0.03750293 |
| Median | | -0.4005288 ^a | -0.2147819 ^a | -0.3777157 ^a | -0.3397351 ^a | -0.5070130 ^a | -0.4478429 ^a | -0.4230535 ^a | -0.3688541 ^a |
| Mode | | -0.40053 | -0.22328 | -0.37772 | 10.45808 | -0.90176 | -0.72965 | -0.42308 | -0.37346 |
| Standard deviation | | 10.00000000 | 10.00000000 | 10.00000000 | 10.00000000 | 10.00000000 | 10.00000000 | 10.00000000 | 10.00000000 |
| Variance | | 10.000 | 10.000 | 10.000 | 10.000 | 10.000 | 10.000 | 10.000 | 10.000 |
| Skewness | | 40.214 | 60.102 | 40.886 | 0.264 | 0.969 | 10.906 | 30.009 | 40.550 |
| Standard error of Skewness | | 0.092 | 0.092 | 0.092 | 0.092 | 0.092 | 0.092 | 0.092 | 0.092 |
| Kurtosis | | 250.976 | 360.854 | 310.581 | -10.605 | -0.544 | 30.358 | 90.206 | 250.656 |
| Standard error of Kurtosis | | 0.183 | 0.183 | 0.183 | 0.183 | 0.183 | 0.183 | 0.183 | 0.183 |
| Range | | 100.85007 | 70.42256 | 90.81346 | 20.61604 | 30.24404 | 50.08106 | 50.51407 | 80.71325 |
| Minimum | | -0.40053 | -0.22328 | -0.37772 | -10.15731 | -0.90230 | -0.73014 | -0.42308 | -0.37362 |
| Maximum | | 100.44954 | 70.19929 | 90.43575 | 10.45873 | 20.34174 | 40.35092 | 50.09099 | 80.33963 |
| Total | | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| Percentile rank | 10 | -0.4005293 ^b | -0.2232370 ^b | -0.3777157 ^b | -10.1283619 ^b | -0.8825617 ^b | -0.7236484 ^b | -0.4230781 ^b | -0.3734659 ^b |
| | 20 | -0.4005293 | -0.2227230 | -0.3777157 | -10.0347067 | -0.8276422 | -0.6994014 | -0.4230769 | -0.3731985 |
| | 25 | -0.4005293 | -0.2222081 | -0.3777157 | -9.9795950 | -0.7899203 | -0.6769337 | -0.4230761 | -0.3729440 |
| | 30 | -0.4005292 | -0.2216742 | -0.3777157 | -8.8941578 | -0.7395791 | -0.6529173 | -0.4230746 | -0.3726562 |
| | 40 | -0.4005291 | -0.2190923 | -0.3777157 | -7.077671 | -0.6493635 | -0.5770395 | -0.4230688 | -0.3715492 |
| | 50 | -0.4005288 | -0.2147819 | -0.3777157 | -0.3397351 | -0.5070130 | -0.4478429 | -0.4230535 | -0.3688541 |
| | 60 | -0.4005281 | -0.2057600 | -0.3777156 | 0.3680564 | -0.2180959 | -0.2591978 | -0.4230040 | -0.3586090 |
| | 70 | -0.4005232 | -0.1821212 | -0.3777146 | 0.9044909 | 0.3202377 | 0.0676693 | -0.4222736 | -0.2917433 |
| | 75 | -0.2106141 | -0.1603951 | -0.1118695 | 10.0958493 | 0.6888830 | 0.2956563 | -0.1771986 | -0.2049837 |
| | 80 | 0.1343573 | -0.1222146 | 0.1629894 | 10.2379296 | 10.1347431 | 0.5951613 | 0.0945848 | 0.0022895 |
| 90 | 10.0190353 | 0.0180118 | 0.7774799 | 10.4147154 | 10.7548875 | 10.4664905 | 10.0753285 | 0.8472849 | |

a. calculated for grouped data
 b. calculated for grouped data

When the overwhelming majority of the population is dissatisfied with pay levels, society may face growing tension and mistrust, followed by an economic crisis. Low satisfaction with pay level has a demotivating effect on the manufacturing industry. This can cause the performance of multiple business entities to decline. By doing nothing, this issue will inevitably worsen, and the situation may even scale up. The proposed sentiment analysis model for measuring population satisfaction can ease the monitoring process, does not require significant costs, and helps promptly address challenges.

Table 4. Results of compound sentiment reversal based on the results of sarcasm detection

| | | Frequency | Percent |
|-------|-----------------|-----------|---------|
| Valid | Neutral (0.00) | 141 | 19.8 |
| | Positive (1.00) | 198 | 27.8 |
| | Negative (2.00) | 372 | 52.3 |
| | Total | 711 | 100.0 |

DISCUSSION OF FINDINGS

To verify the results obtained, we will consider the leading indicators of the standard of living of the population of the Russian Federation.

We have selected five main indicators based on which it is possible to assess the change in the population's standard of living, namely the values of the increase in the average monthly nominal accrued wages, the real average monthly accrued wages, the average per capita monetary income of the population, real monetary income, and real disposable monetary income. We obtained the data from an official source (Federal State Statistics Service of the Russian Federation).

For convenience, we present the selected values in the form of graphs. As seen in Figure 7, during 2015–2020, indicators' growth dynamics were ambiguous. The indicators show the greatest importance in 2015. However, our goal at this stage of the study was to verify the adequacy of the obtained values of the satisfaction of the Russian population with wages. Therefore, it is worth paying attention to their value in 2020 since we analyzed the period from July 1, 2020 to October 1, 2020.

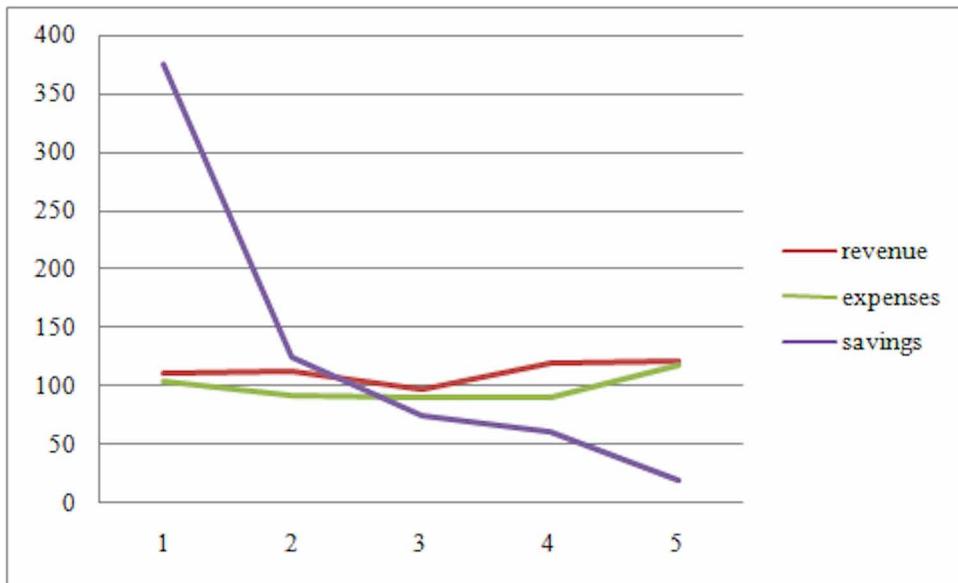
As seen in the chart, all indicators show a decrease in dynamics. The population's real monetary incomes and real disposable monetary incomes had a negative growth rate. At the same time, the average per capita monetary income had zero growth, and the values of the average monthly nominal and real average monthly wages increased slightly—by 1 and 2%, respectively.

Next, we considered the rates of monetary incomes and expenditures of the population of the Russian Federation (Figure 8). The graph shows that since 2015, the population's savings volume has significantly decreased. While in 2015, the growth rate was 375% compared to the previous year, in 2019, its value was 18%. We also see an increase in household spending while maintaining the level of income. Thus, we can conclude that the existing level of income (the main component of which is

Table 5. Understanding scores on the population satisfaction with pay level

| Sentiment Index (SI) | Satisfaction Level | Description |
|----------------------|--------------------|--|
| $1 \geq SI > 0.5$ | High | The overall sentiment towards the item under investigation is very positive, and the majority of the population is satisfied with the current state pay policy |
| $0.5 \geq SI > 0$ | Moderate | The overall sentiment towards the item under investigation is positive, and the current state pay policy can continue |
| $0 (\pm 0.01)$ | Neutral | The overall sentiment towards the item under investigation is neither positive nor negative, which necessitates policy actions to prevent negative sentiments |
| $0 > SI \geq -0.5$ | Low | The overall sentiment towards the item under investigation is negative, and the current state pay policy must be revised |
| $-0.5 > SI \geq -1$ | Critical | The overall sentiment towards the item under investigation is very negative, such that the current state pay policy must be abandoned |

Figure 7. Dynamics of changes in the main indicators of the standard of living of the population of the Russian Federation



wages) is insufficient to maintain the standard of living familiar to citizens, which may explain the decline in the level of savings in relation to 2019.

Next, we considered the values of the indices calculated from Russian statistics. The consumer confidence index, reflecting the aggregate consumer expectations of the population, decreased in 2020 and amounted to -26% (Figure 9). This indicator is low, and as can be seen from the graph, according to the dynamics, this level indicates a decline in the trend relative to previous periods. Notably, the consumer confidence index is one of the most sensitive indicators of assessing the market situation. Thus, the detected dynamics directly indicate the actual situation. However, we should not forget that official statistics tend to “smooth the corners” during crisis events.

The index of changes in personal financial situation also decreased and amounted to -20% (Figure 10). As seen in the chart, both indexes have negative values. They reached a positive mark only in 2007. Thus, we can conclude that the data obtained from applying the methodology proposed in this work are adequate.

Using text mining technologies to assess the satisfaction of the population avoids the need to conduct surveys or questionnaires, which are necessary when using traditional models to calculate such indicators in Russia. Obtaining the essential information is possible in real time, which is also a benefit since it allows researchers to omit parts of the study, such as preparing and conducting the survey and preparing the results for further processing and analysis. In addition, existing programs allow researchers to automate the calculation of indicators. To do this, it is sufficient to save the data processing algorithm and set the appropriate marks to ensure that the existing database is replenished with new data.

A conceptual analysis of existing case studies has shown that in Russia, the determination of moods is most often based on traditional methods that lose effectiveness in the modern world. At the same time, the study of the possibility of conducting such studies using text-mining technologies based on unsolicited data is poorly developed. The main hindrance to developing this scientific direction is the lack of adapted methods for analyzing Russian text. This is primarily due to the specifics of existing programs. They are mainly presented in English and have a well-developed English dictionary that allows researchers to distinguish sentiment and the main groups of emotions in the analysis process.

Figure 8. Dynamics of monetary incomes and expenditures of the population of the Russian Federation, in % to the corresponding period



Such programs, as a rule, allow for their training because they use artificial intelligence. In other words, such programs can be trained by adding specific dictionaries, algorithms, and databases.

RECOMMENDATION

The problem of the impossibility of processing Russian text has several solutions. One of them, presented in this paper, consists of adding a mandatory step to the text analysis process to translate the resulting database into English with maximum preservation of meaning. Thus, we provide the possibility of text processing and obtaining adequate results. The second solution is to develop a Russian dictionary and add it to the work program for further training. This method is preferable because it will allow researchers to obtain the most reliable results that consider all the features of the language. However, it is time-consuming and requires a long time interval for implementation. Such work should be carried out jointly with philologists, taking modern colloquial speech into

Figure 9. Dynamics of the consumer confidence index

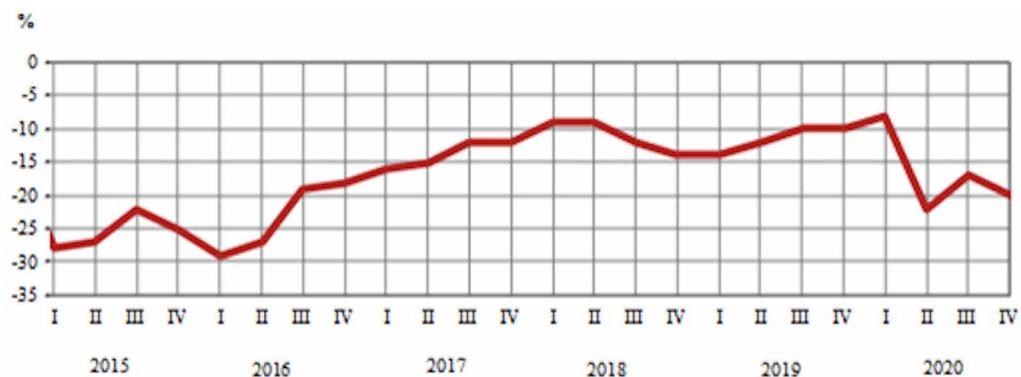
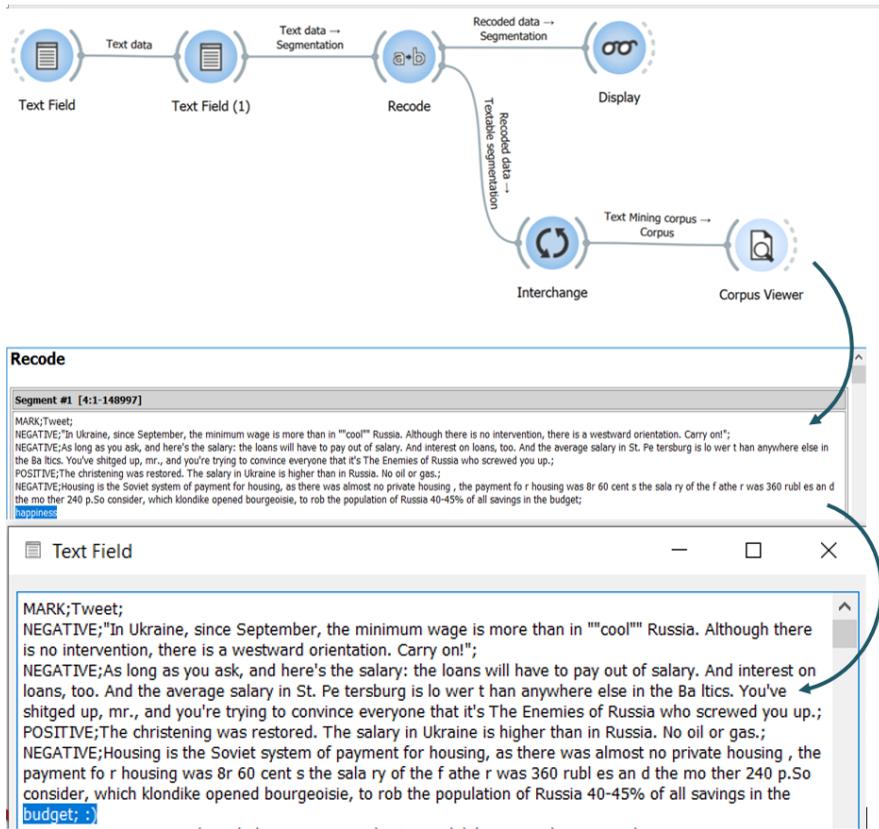


Figure 10. Dynamics of the index of changes in personal financial situation



account. This requirement is mandatory since the texts published on social networks primarily use spoken language.

Despite their apparent advantages, text mining methods are poorly developed in the Russian Federation today. Even with the existing limitations, their use has several significant advantages described earlier in this paper. They allow researchers to conduct a general analysis of moods and to study it relative to a separate subject. Thus, focusing on a particular problem in detail, taking into account the needs of society, is ensured. Studying the moods of the population with the help of these methods will allow obtaining timely, relevant, and maximally complete results, as opposed to traditional survey methods, ensuring the development of appropriate and effective methods of solving problems.

The methodology presented in this paper fills the existing gap regarding the use of text mining methods to study the moods (satisfaction) of the Russian population. It examines the main features of the application of existing programs and the direction of their further development. In particular, we determined that to analyze Russian-language text, it is necessary to identify sarcasm since existing algorithms cannot recognize it. At the same time, the presence of sarcasm in the text completely changes its meaning, affecting the final result. We carried out identification according to the data extracted by the program based on Plutchik's estimates. As a result, we revealed that sarcasm includes emotions such as disgust, humor, and sadness. They are crucial to highlighting messages. As we showed, the program algorithm identified messages containing sarcasm as positive or neutral, which is an incorrect assessment since sarcasm carries a negative message. Accordingly, recoding these messages into negative ones is necessary to obtain reliable data. On this basis, it is possible to build a population satisfaction index.

Thus, in this paper, we have proposed several additional actions that ensure reliable data when analyzing the Russian text to determine the population's mood and satisfaction regarding specific socio-economic issues.

CONCLUSION

In sum, it is worth emphasizing once again the importance of studying and developing approaches to assessing the opinion of the population on various issues in socio-economic life. As the research analysis has shown, scientific study in this area has been sufficient outside of Russia, while interest remains extremely low in Russia. Few Russian researchers have conducted studies in this field, and most of these are theoretical.

In this study, we adapted the methodology of analyzing text-based Russian-language content of social networks based on text mining technologies to assess population satisfaction. We also proposed a formula for calculating the satisfaction index. The methodology considers the peculiarities of working with unsolicited data published in Russian, allowing us to obtain a final indicator based on the interpretation of the values and making it possible to assess the Russian population's overall level of satisfaction.

The results of this work allow us to advance in the study of the application of existing social network content analysis programs based on text mining technologies to Russian-language messages. This work aimed to develop applications of text mining technologies to determine the mood and satisfaction of the population of the Russian Federation. The results allow us to identify several features of the analysis of texts left by Russian-speaking users in social networks concerning the subject under study.

In particular, using text mining technologies when working with Russian texts has its specifics since existing programs focus on working with English. The program used in this study lacks the dictionaries necessary for processing texts posted by Russian-speaking users. In this regard, it is impossible to isolate sentiment from them. To do this, the text must be translated into English. As a result of this operation, the meaning of messages can be distorted since users may use phrases, expressions, and idioms that are difficult to translate. However, this distortion is generally insignificant. The text's general meaning is preserved, which makes it possible to further extract sentiment and evaluate emotions based on the translated text. In the future, developing a Russian dictionary and training a program based on it is necessary to solve the existing problem.

The data obtained will also allow us to build a composite index of economic confidence. Such an index makes it possible to adjust and forecast macroeconomic indicators, such as supply and demand, lending volumes, spending volumes, population welfare, and economic growth, which can help adjust and build a strategy for the functioning and development of various economic entities. Using the proposed methodology ensures the calculation of the indicator in real time. Based on the obtained indicator, it is possible to develop timely measures to eliminate the identified problems.

In subsequent works, the authors will construct a composite index of economic sentiment, one component of which is the value of the population's satisfaction with wage levels. To do this, it is necessary to clarify the ranges of values on the scale interpreting the results obtained since the division presented in this paper is subjective and based on the Likert scale.

COMPETING INTERESTS

The authors of this paper declare there are no competing interests.

FUNDING

This research was supported by the Russian Science Foundation grant No. 23-28-00946, <https://rscf.ru/en/project/23-28-00946/>.

REFERENCES

- Abulaish, M., Rahimi, M. B., Ebrahemi, H., & Sah, A. K. (2019). SentiLangN: A Language-Neutral Graph-Based Approach for Sentiment Analysis in Microblogging Data. In *Proceedings of WI '19: IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 461–465). ACM. doi:10.1145/3350546.3352568
- Al-Obeidat, F., Hani, A. B., Benkhelifa, E., Adedugbe, O., & Majdalawieh, M. (2019). A Sentiment Analysis Approach of Data Volatility for Consumer Satisfaction in the Fashion Industry. In *Proceedings of the Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)* (pp. 262–267). IEEE. doi:10.1109/SNAMS.2019.8931876
- Alarifi, A., Tolba, A., Al-Makhadmeh, Z., & Said, W. (2020). A big data approach to sentiment analysis using greedy feature selection with cat swarm optimization-based long short-term memory neural networks. *The Journal of Supercomputing*, 76(6), 4414–4429. doi:10.1007/s11227-018-2398-2
- Alim, A., & Shukla, D. (2020). A big data parameter estimation approach to develop big social data analytics framework for sentiment analysis. *Lecture Notes in Networks and Systems*, 100, 755–764. doi:10.1007/978-981-15-2071-6_63
- Alqmase, M., Al-Muhtaseb, H., & Rabaan, H. (2021). The formalism of sports fanaticism for the analysis of moods in the Arabic text. *Analysis and Mining of Social Networks*, 11. doi:10.1007/s13278-021-00757-9
- Amirkhanova, L. R., & Bikmetov, E. Yu. (2017). The interdisciplinary conceptualization and the methods of economic evaluation of trust between the subjects of market relations. [in Russian]. *Eurasian Law Journal*, 1(104), 349–352.
- Arora, A., Bansal, S., Kandpal, C., Aswani, R., & Dwivedi, Y. (2019). Measuring social media influencer insights from Facebook, Twitter and Instagram. *Journal of Accounting, Auditing & Finance*, 34(3), 385–410. doi:10.1177/0148558X17719206
- Bobichev, V., Kanishcheva, O., & Cherednichenko, O. (2017). Sentiment analysis in the Ukrainian and Russian news. In *Proceedings of the IEEE 1st Ukrainian Conference on Electrical and Computer Engineering, UKRCON 2017 – Materials* (pp. 1050–1055). IEEE. doi:10.1109/UKRCON.2017.8100410
- Cortis, K., & Davis, B. (2021). Over a decade of public opinion research: A systematic review. *Artificial Intelligence Review*, 54(7), 4873–4965. doi:10.1007/s10462-021-10030-2 PMID:34188346
- El Alaoui, I., Gahi, Y., Messoussi, R., Chaabi, Y., Todoskoff, A., & Kobi, A. (2019). Correction to: A novel adaptable approach for sentiment analysis on big social data. *Journal of Big Data*, 6(1), 76. doi:10.1186/s40537-019-0239-7
- Fomin, V., Osochkin, A., & Zhuk, Yu. (2019). Frequency and morphological patterns of recognition and thematic classification of essays and full-text scientific publications. In *Proceedings of the CEUR Workshop* (p. 2401). CEUR.
- Hadi, K. A., Lasri, R., & El Abderrahmani, A. (2019). An efficient approach for sentiment analysis in a big data environment. *International Journal of Engineering and Advanced Technology*, 8(4), 263–266.
- Jadon, P., Bhatia, D., & Mishra, D. K. (2019). A BigData approach for sentiment analysis of Twitter data using Naive Bayes and SVM Algorithm. In *Proceedings of the Sixteenth International Conference on Wireless and Optical Communication Networks (WOCN)* (pp. 1–6). IEEE. doi:10.1109/WOCN45266.2019.8995109
- Juhro, S. M., & Iyke, B. N. (2020). Consumer confidence and consumption expenditure in Indonesia. *Economic Modelling*, 89, 367–377. doi:10.1016/j.econmod.2019.11.001
- Kannan, E., & Kothamasu, L. A. (2020). A pattern based approach for sentiment analysis using ternary classification on Twitter data. *International Journal on Emerging Technologies*, 11(2), 811–816.
- Korobkin, D. M., Vasiliev, S. S., Fomenkov, S. A., & Lobeyko, V. I. (2019). Extraction of structural elements of inventions from Russian-language patents. In *Proceedings of the Multi-conference on Informatics and Information Systems, MCCSIS 2019 and Proceedings of International Conferences Big Data Analytics, Data Mining and Computational Intelligence 2019 and Theory and Practice in Modern Computing 2019* (pp. 159–166). IEEE. doi:10.33965/tpmc2019_201907L020

- Kotelnikov, E., Peskischeva, T., Kotelnikova, A., & Razova, E. (2018). A comparative study of publicly available Russian sentiment lexicon. In D. Ustalov, A. Filchenkov, L. Pivovarova, & J. Žižka, (Eds.), *Proceedings of the Conference on Artificial Intelligence and Natural Language (AINL 2018), Communications in Computer and Information Science*, (Vol. 930, pp. 139–151). Springer. doi:10.1007/978-3-030-01204-5_14
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives de Psychologie*, 140, 1–55.
- Liu, X. (2020). Analyzing the impact of user-generated content on B2B firms' stock performance: Big data analysis with machine learning methods. *Industrial Marketing Management*, 86, 30–39. doi:10.1016/j.indmarman.2019.02.021
- Lobantsev, A., Loginova, V., Burlakova, Yu., Andreev, N., Matveeva, V., Filimonova, I., Dobrenko, N., & Gusarova, N. (2020). Public resources for identifying mentions of medical data in Russian social networks. *Communications in Computer and Information Science*, 1242, 479–489. doi:10.1007/978-3-030-65218-0_35
- Mathews, D. M., & Abraham, S. (2019). Twitter data sentiment analysis on a Malayalam dataset using rule-based approach. *Advances in Intelligent Systems and Computing*, 906, 407–415. doi:10.1007/978-981-13-6001-5_33
- Moussa, S. (2019). An emoji-based metric for monitoring consumers' emotions toward brands on social media. *Journal of Retailing and Consumer Services*, 49, 86–101. doi:10.1016/j.jretconser.2019.03.012
- Odendaal, H., Reid, M., & Kirsten, J. F. (2020). Media-based sentiment indices as an alternative measure of consumer confidence. *The South African Journal of Economics*, 88(4), 409–434. doi:10.1111/saje.12261
- Peláez, J. I., Martínez, E. A., & Vargas, L. G. (2020). Products and services valuation through unsolicited information from social media. *Soft Computing*, 24(3), 1775–1788. doi:10.1007/s00500-019-04005-3
- Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4), 344–350. doi:10.1511/2001.28.344
- Podoprigora, A. V. (2016). Trust as a key resource for the socio-economic development of the information society. *Vestnik Chelyabinskogo gosudarstvennogo universiteta (The Chelyabinsk State University Bulletin), Series [in Russian]. Economics*, 11(393), 9–18.
- Radygin, V. Yu., Kupriyanov, D. Yu., Bessonov, R. A., Ivanov, M. N., & Oslyakova, I. V. (2021). *Application of technologies of intellectual analysis of text in Russian for solving problems of primary financial monitoring*. In Proceedings of the International Conference on Brain-Inspired Cognitive Architectures for Artificial Intelligence: Eleventh Annual Meeting of the BICA Society, BICA*AI 2020, Rio Grande do Norte. <https://doi.org/10.1016/j.process.2021.06.078>
- Riahla, M. A., Goumiri, S., Tamine, K., & Hamadouche, M. H. (2021). A new cooperation model for dynamic networks. [IJBDNCN]. *International Journal of Business Data Communications and Networking*, 17(2), 76–90. doi:10.4018/IJBDNCN.286704
- Rodrigues, A. P., & Chiplunkar, N. N. (2019). A new big data approach for topic classification and sentiment analysis of Twitter data. *Evolutionary Intelligence*, 1, 1–11. doi:10.1007/s12065-019-00236-3
- Shehu, H. A., & Tokat, S. (2020). A hybrid approach for the sentiment analysis of Turkish Twitter data. *Lecture Notes on Data Engineering and Communications Technologies*, 43, 182–190. doi:10.1007/978-3-030-36178-5_15
- Song, M., & Shin, K.-S. (2019). Forecasting economic indicators using a consumer sentiment index: Survey-based versus text-based data. *Journal of Forecasting*, 38(6), 504–518. doi:10.1002/for.2584
- Srivastava, A., Singh, V., & Drall, G. S. (2019). Sentiment analysis of Twitter data: A hybrid approach. *International Journal of Healthcare Information Systems and Informatics*, 14(2), 1–16. doi:10.4018/IJHISI.2019040101
- Takhanova, O. V. (2018). Interpersonal trust as a factor of social and economic development of society. *Bulletin of Udmurt University, Series [in Russian]. Economics and Law*, 28(1), 49–55.
- Wu, W., Liu, W., Suardi, S., & Chang, Y. (2019). Tiered information disclosure: An empirical analysis of the advance peek into the Michigan Index of Consumer Sentiment. *Financial Review*, 54(3), 541–582. doi:10.1111/fire.12184

- Yan, Y., Dongmei, Z., & Weining, Y. (2019). Role of platform, supplier and medium in online trust formation. [IJBDCN]. *International Journal of Business Data Communications and Networking*, 15(1), 58–70. doi:10.4018/IJBDCN.2019010104
- Yang, C., Zhang, H., Huang, J., & Wan, J. (2020). Text sentiment analysis of online cigarette reviews. *Acta Tabacaria Sinica*, 26(2), 92–100. doi:10.16472/j.chinatobacco.2019.259
- Yost, E., Ridderstaat, J., & Kizildag, M. (2020). Early warning indicators? The effect of consumer and investor sentiments on the restaurant industry. *International Journal of Hospitality Management*, 89, 102575. doi:10.1016/j.ijhm.2020.102575

Konstantin Malysenko is a PhD, Candidate of Economic Sciences, an PhD, Assistant Professor at the Department of Economics and Finance of the Crimean Federal University named after V.I. Vernadsky, Humanitarian and Pedagogical Academy (branch), Institute of Economics and Management in Yalta, Russian Federation. He is an expert in economics, effective development policies, social tension and economic confidence etc.

Vadim Malysenko is a PhD, Candidate of Economic Sciences, an PhD, Assistant Professor at the Department of Economics and Finance of the Crimean Federal University named after V.I. Vernadsky, Humanitarian and Pedagogical Academy (branch), Institute of Economics and Management in Yalta, Russian Federation. He is specialized in development of modern technologies, Internet environment, processing and extraction of information etc.

Marina Anashkina is an Assistant at the Department of Economics and Finance of the Crimean Federal University named after V.I. Vernadsky, Humanitarian and Pedagogical Academy (branch), Institute of Economics and Management in Yalta, Russian Federation. She is interested in Big Data technologies, socio-economic research and so on.

Dmitry Anashkin is a Student at the Department of Mathematics, Theory and Methods of Teaching of Mathematics of the Crimean Federal University named after V.I. Vernadsky in Yalta, Russian Federation. His research interests include big data analysis, social networks, consumer emotions, etc.