

# Informal evaluation of corporate image based on text mining

*Kostyantyn* Malyshenko<sup>1,\*</sup>, *Vadim* Malyshenko<sup>1</sup>, *Marina* Anashkina<sup>2</sup>, and *Dmitriy* Anashkin<sup>1</sup>

<sup>1</sup>V.I. Vernadsky Crimean Federal University, Republic of Crimea, 295007 Simferopol, Russia

<sup>2</sup>Sevastopol State University, 299053 Sevastopol, Russia

**Abstract.** This article explores the possibilities of using text-mining technologies in order to determine the image of corporations based on data obtained from Twitter social network. The problem of low efficiency of traditional methods of consumer opinion research and the need to develop methods based on unsolicited data has been actualized. Consumer opinion is an indicator of the level of corporate image. Analysis of opinions allows you to develop an effective policy to improve it. The authors have developed a methodology for assessing the corporate image. The article analyzes the work of leading researchers. The features of the use of technologies when working with texts published in Russian have been analyzed. An index of customer (consumer) satisfaction has been developed, which is proposed as a basis for determining the level of corporate image. The obtained results of the study allow to make further adjustments to the corporation's policy in order to improve its image.

## 1 Introduction

In modern conditions of development, corporations play a key role. They are the largest economic entities and perform a number of socio-economic functions, among which are job creation, satisfaction of consumer demand, ensuring capital flow through financial channels, development of various sectors of the economy through investment, etc. The effective functioning of the corporation, its successful development in the changing economic, political and social conditions is the result of a competent policy, which is not possible without an effective analysis of external and internal factors. In the modern customer-oriented economy, the assessment of consumer opinion becomes a key factor in making appropriate management decisions and achieving maximum efficiency of the corporation's activities [1, 2].

A corporation's image is a key factor in its advancement in a competitive market. In this regard, the issue of searching for modern methods of analyzing the consumers' opinions regarding the quality of its work is relevant. Customer satisfaction with individual parameters of goods / works / services sold by the corporation forms the basis for developing specific solutions to eliminate identified problems and build further functioning policies.

---

\* Corresponding author: [docofecon@mail.ru](mailto:docofecon@mail.ru)

The choice of such methods, as a rule, should meet the following requirements

1. Information content, clarity and objectivity of the results.
2. Efficiency of obtaining results.
3. The relevance of the results obtained (must be based on actual data).
4. Possibility of forecasting.
5. Availability of incoming information.
6. Accessibility and ease of use.
7. Minimal financial investment.

The above requirements at the present stage of scientific and technological progress are fully consistent with big data analysis technologies, namely, text mining.

Krylov V.S. in his work "Digital Economy: Intelligent Text Analysis" [3] explores the possibility of using the software modules that allow extracting and visualizing the metadata components of text emotionality. The author focuses on the need to develop the methods of analysis and forecasting necessary for the development of intelligent information systems, their deep learning, neural networks, artificial intelligence, etc.

Chang Ts. [4] in the study "Product sales forecasting using macroeconomic indicators and online reviews: a method combining prospect theory and sentiment analysis" notes that macroeconomic conditions and user reviews have a significant impact on consumer purchasing decisions, and they can potentially be used to make more accurate forecasts of corporate development. The author has developed a sentiment analysis algorithm for converting text online reviews into numerical values.

The need to determine opinions in order to maintain the image of corporations and build an effective policy for their functioning is also considered in the works of El Alau I. [5], Moussa S. [6].

Traditional methods of determining the opinion of consumers today are becoming ineffective, since they are based on historical data. Most often, in order to determine consumer sentiment and satisfaction in Russian practice, statistical and survey methods are used. This, in particular, is reflected in the works of O.V. Takhanova. [7], Amirkhanova L.R. [8]. Pelaez J. The work "Products and services valuation through unsolicited information from social media" [9] raises the problem of the low reliability of the currently existing methods of collecting and evaluating consumer opinions about various goods and services. The author proposes the calculation of a special index based on the processing of unsolicited consumer information using text mining technologies.

In the work of M. Song "Forecasting economic indicators using a consumer sentiment index: Survey-based versus text-based data" [10], the effectiveness of modern methods of processing unsolicited data and their comparison with the traditional method of polling respondents has been investigated. The authors come to the conclusion that methods for analyzing big data have some limitations, but they allow obtaining better information.

Eskisi H. in the study "A text mining application on monthly price developments reports" [11] emphasizes that the use of text mining provides great opportunities for economic research. The paper proposes a methodology for using text mining technologies to measure statistical consistency with the annual consumer price index.

Meanwhile, both of these methods do not meet the requirements put forward for this indicator, which were listed above. So, using only statistical data as a basis for constructing allows you to get results for a period that is significantly remote from the moment of direct receipt of the final result. In this regard, further development of decisions on this basis may have low efficiency, especially in the case when it is necessary to take urgent measures to change the policy of the corporation in order to improve its image.

Thus, the study of the possibilities of using text-mining technologies in order to assess the corporate image of Russian companies is a topical issue. To date, the approaches to data processing by these technologies proposed by the world community are poorly developed.

It is necessary to study the peculiarities of their application in order to analyze the texts left by Russian-speaking users, and also to develop a methodology for using the resulting data in order to assess the corporate image.

The purpose of this work is to determine the possibilities and limitations of the use of text-mining technologies in order to determine the satisfaction of consumers (clients) of corporations and to develop a methodology for assessing the corporate image.

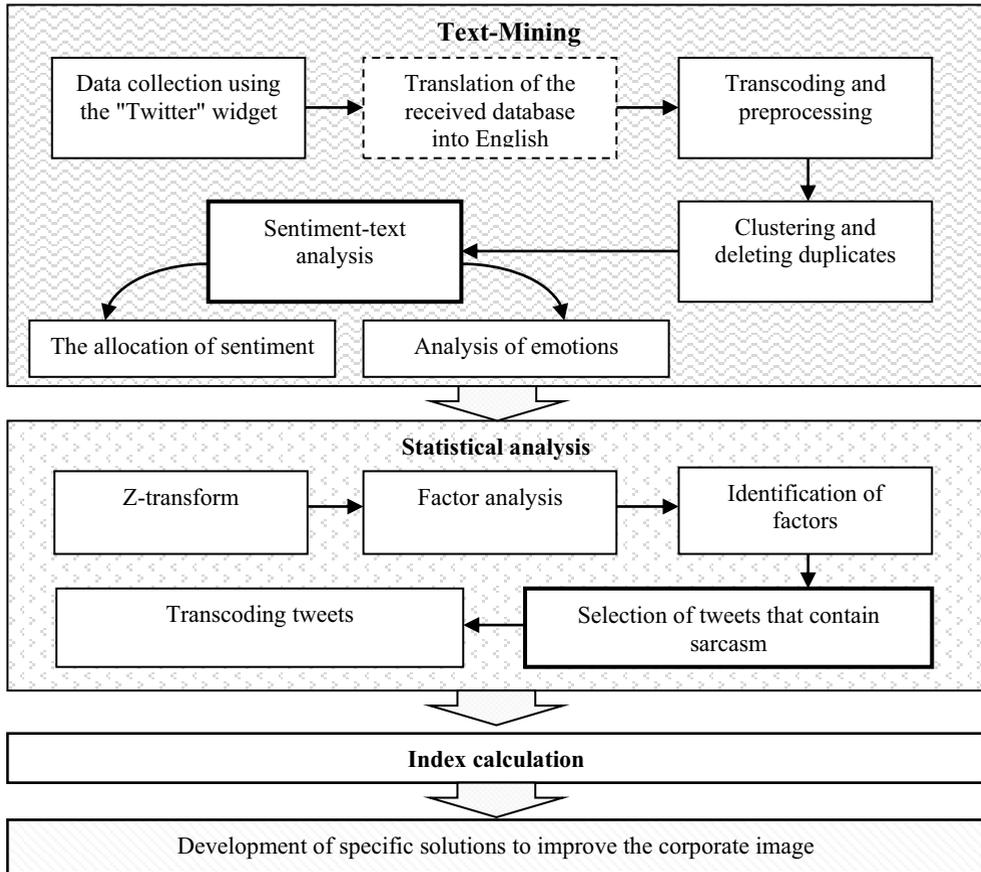
## **2 Materials and Methods**

In this work, we used data on the two largest banks operating in the Russian Federation with a corporate structure. These are banks VTB and Sberbank. We have used the Orange software, which is a set of various big data analysis tools.

It also contains all the necessary set of text mining tools, which allows you to conduct research without hindrance. In particular, the software allows you to automate the data collection process, since it has a special component - "Twitter". Thus, it can be used to extract short user messages on any topic from Twitter social network.

The use of the data of this social network meets the requirements of the research - the need to obtain information in real time, or with a minimum gap in the time window between the collection and the direct receipt of the final analysis result. Tweets posted by users are selected according to a given topic (tag). Thus, filtering tweets by corporation name allows us to extract consumer feedback about it, presented in the form of user posts on the social network.

With the help of the software component, we collected data on the two above corporations using the tags "VTB" and "Sberbank" for November 2020. Their volume was 505 tweets. The study was carried out in several stages, shown in Figure 1.



**Fig. 1.** Procedure for evaluating the corporate image.

Let's look at these stages in more detail.

### 3 Results and Discussion

As mentioned, the first step in research is to collect the necessary data. As a result, we obtained two databases of tweets with the tags "VTB" and "Sberbank", containing the texts published by users regarding these corporations. The software we use has a number of functions that allow automating the process of adding new tweets to the existing database. Thus, it is possible to obtain data in real time.

The main problem of using text mining technologies for analyzing texts left in Russian is the lack of necessary dictionaries in existing software for their further processing. In other words, the software does not recognize the meaning of such texts and, accordingly, does not allow calculating sentiment and extracting emotions from it. To solve this problem in this study, at the next stage, it is necessary to translate the obtained bases into English, since it is the language of the software developer.

At the next stage, the databases are loaded in English and prepared for further processing. For this, the software also has a special component. It performs conversion and preprocessing of the text - extra characters, punctuation marks, etc. are removed.

Next, you need to clear the text from duplicate tweets that are published by robots. This is necessary to obtain a more accurate result. Such tweets can affect the overall assessment

of consumers' opinion on the functioning of the corporation, since it increases the share of certain emotions inherent in such a text through its numerous repetitions.

After that, sentiment and emotions are directly extracted according to the algorithms set by the program. We used the Vader algorithm for sentiment extraction, since it gives the most adequate assessment of the analyzed text. The Plutchik algorithm was used to highlight the emotions. It allows you to highlight all the groups of emotions necessary to define sarcasm in the text. The results obtained as a result of processing by text-mining technology require further statistical processing in order to clarify the belonging of individual tweets to one of three groups: positive, negative, neutral. This must be done in order to obtain the most accurate assessment of customer opinion, since the software does not recognize sarcasm.

For these purposes, we used SPSS statistical data processing software. The first step is to standardize the scores (Z-transform). This procedure is necessary because the obtained variables have different ranges and meanings in orders. They must be brought to a common denominator. Further, on this basis, we carried out a factor analysis, as a result of which the main factors in the selected groups of tweets corresponding to sarcasm were identified. As a result, it was revealed that sarcasm is a ratio of disgust, humor, sadness.

The next step is to define the criteria boundaries of these factors and identify the tweets containing sarcasm. Thus, tweets that are defined as positive and neutral, and contain sarcasm, should be categorized as negative ones. For this, we have carried out the appropriate conversion.

The resulting estimates are necessary to build an index of customer satisfaction. The higher the index, the higher the corporate image, and vice versa. We have formed three groups of tweets: positive, negative and neutral. To calculate the index, it is proposed to use the following formula developed by the authors:

$$Ind = (pos - neg)/(tot - neu) \tag{1}$$

In this formula, *pos* is the number of positive tweets, *neg* is the number of negative tweets, *tot* is the total number of tweets, *neu* is the number of neutral tweets.

The results of the frequency analysis are presented in Tables 1 and 2.

**Table 1.** Frequency analysis based on opinions about Sberbank.

If compound is greater than zero (positive tweet) - 1 if less than zero (negative tweet) - 2 if zero (neutral tweet) - 0					
		Frequency	Percentage	Percentage of valid	Accumulated percent
Valid	.00	159	28.7	31.5	31.5
	1.00	218	39.4	43.2	74.7
	2.00	128	23.1	25.3	100.0
	Total	505	91.2	100.0	
Missing	Systemic	49	8.8		
Total		554	100.0		

**Table 2.** Frequency analysis based on opinions about VTB Bank

If compound is greater than zero (positive tweet) - 1 if less than zero (negative tweet) - 2 if zero (neutral tweet) - 0					
		Frequency	Percentage	Percentage of valid	Accumulated percent
Valid	.00	148	29.4	29.4	29.4
	1.00	203	40.3	40.3	69.6
	2.00	153	30.4	30.4	100.0
	Total	504	100.0	100.0	

Let's calculate the indices according to the above formula:

$$Ind (Sberbank) = (218 - 128) / (505 - 159) = 0.26 \quad (2)$$

$$Ind (VTB) = (203 - 153)/(504 - 148) = 0,14 \quad (3)$$

We will interpret the index values according to the following table based on the Likert scale:

**Table 3.** Customer satisfaction scale.

Index value	Satisfaction level	Interpretation of the value	The level of the image of the Corporation
$1 \geq Ind > 0.5$	high	High level - the respondents' mood regarding the corporation is positive, the policy is effective, no revision is required	high
$0.5 \geq Ind > 0$	average	Average level - the general mood of the respondents towards the corporation is positive, the policy is effective, it needs to be continued	average
$0 (\pm 0.01)$	neutral	Neutral level - respondents' mood regarding the corporation is identified as both positive and negative, the policy is ineffective, it is necessary to take certain actions to improve the corporation's image	neutral
$0 > Ind \geq -0.5$	low	Low level - the general mood of respondents towards the corporation is negative, a revision of the existing policy is necessary	low
$-0.5 > Ind \geq -1$	crisis	Crisis level - respondents' mood regarding the corporation is negative, a complete change of the existing policy is required	crisis

Thus, the values of the indices obtained in this study for both corporations refer to the average level. That is, the customers are satisfied with the quality of services provided by the banks, dissatisfaction is generally insignificant. This is in line with the average corporate image score. The policy of their functioning should be continued, or it is possible to carry out a detailed analysis of customer dissatisfaction in order to eliminate the shortcomings and achieve a high level of image.

## Conclusion

This article explores the possibilities of using text-mining technologies in order to determine the image of corporations based on data obtained from Twitter social network . We have developed a methodology for assessing the corporate image. The methodology was verified on the basis of data from VTB and Sberbank operating in Russia and having a corporate structure.

We have identified certain limitations regarding the possibility of using text-mining technologies when analyzing texts published in Russian. Existing software can analyze text mainly in the developer's language (English). At the same time, the training of the program is possible, but this requires the development of a Russian dictionary and its further introduction into the software.

To date, the application of the methodology proposed in the work requires the addition of an obligatory step - the translation of the resulting database into English. The results obtained do not have material distortions and allow further calculation of the index.

The proposed index can be interpreted according to the scale. Its results can be used as the basis for the development of specific management decisions aimed at changing the corporate image. This indicator has a number of advantages, in particular, it allows you to monitor the current state of the company's image, since it is based on current data. This allows you to timely respond to negative trends and take measures to prevent a fall in the image of corporations and a decrease in demand for their services.

## Acknowledgment

The reported study was funded by RFBR and EISR, project number 20-011-31388 "Development of a methodology for using text-mining technologies to determine mood indices based on social networks".

## References

1. K.A. Malyshenko, V.A. Malyshenko, M.V. Anashkina, Proceedings of the 32nd International Business Information Management Association Conference, IBIMA 2018 - Vision 2020: Sustainable Economic Development and Application of Innovation Management from Regional expansion to Global Growth, 1030, (2018)
2. K.A. Malyshenko, V.A. Fastynova, D.A. Mardar, M.A. Shostak, M.V. Anashkina, V.A. Malyshenko, *Revista Inclusiones*, **7**, 739 (2020)
3. V.S. Krylov, Scientific notes of the Crimean engineering and pedagogical University, **3(65)**, 120 (2019)
4. C.Zhang, Y.-X. Tian, Z.-P. Fan, Y. Liu, L.-W. Fan. *Soft Computing*, **24(9)**, 6213 (2020)
5. I. El Alaoui, Y. Gahi, R. Messoussi, Y. Chaabi, A. Todoskoff, A. Kobi, *Journal of Big Data*, **5**, 1 (2018)
6. S. Moussa, *Journal of Retailing and Consumer Services*, **49**, 86 (2019)
7. O.V. Tahanova, J.D. Tsibikdorzhieva, G.E. Alekseeva, A.R. Tsyrenov, *Internet-journal "Science studies"*, 9, 6. (2017), <https://naukovedenie.ru/>
8. L.R. Amirkhanova, E.Yu. Bikmetov, *Eurasian economic journal*, **1(104)**, 349 (2017)
9. J.I. Peláez, E.A.Martínez, L.G.Vargas, *Soft Computing*, **24(3)**, 1775 (2020)
10. M. Song, K-S. Shin, *Journal of forecasting*, **38(6)**, 504 (2019)
11. H.B. Eskici, N.A. Kocak, *Central bank review*, **18(2)**, 51 (2018)