Comparative analysis of reviews of the Russian Agricultural Bank

P A Kuznetsov and A V Plotnikov

1 Perm National Research Polytechnic University, 29 Komsomolsky Av., Perm, 614000, Russia
2 Perm State Agro-Technological University named after Academician D.N. Pryanishnikov, 23, Petropavlovskaia Street, Perm, 614990, Russia

*E-mail: plotnikov-av@mail.ru

Abstract. The paper analyzes reviews of the Russian agricultural Bank (the study's object, one of the largest banks in the Russian Federation, specializing in financing enterprises in the agro-industrial complex). We used the Mann-Whitney U statistical test, The Wilcoxon signed-rank T test nonparametric statistical test, and the Pearson correlation coefficient to compare subsamples. In the study results, we determined that reviews with a higher rating are less understandable and straightforward and contain more complex texts. Conversely, reviews with lower ratings are more specific and easier to understand. Thus, we can reject the null hypothesis that the subgroups of reviews rated "5" and "from 1 to 4" are quantitatively the same.

1. Introduction
The company's online communication with Internet users is often based on two tools: monitoring and official response. Today, in the Russian segment of the Network, users are three-quarters of the country's population. More than 35 million of them regularly visit a particular social network. All virtual world subjects communicate with each other, gather on forums, public pages, discuss brands, companies, and consult about the purchase of goods and services. The final decision to purchase a product is increasingly influenced by the Internet environment or with its participation. Usually, sites for collecting reviews are specialized forums, search aggregators, and social networks.

Any response in the virtual space can significantly affect the company's profitability, bring a start-up to a leading position in the shortest possible time, and make a product or service iconic. Simultaneously, there are risks because a single negative review can shake or even destroy its reputation, which has been formed by marketers for years, which will eventually alienate a significant proportion of potential customers from the company's products. Simple ways to publish and distribute categorical information can quickly have a devastating impact on organizations' online reputation [1]. Reputation has a positive impact on online sales of new organizations' products. Perceived quality and fame are two dimensions of reputation that account for 30% of sales fluctuations [2].

Experts note that every client today is both an audience and a mouthpiece, so it is important to offer them as many reasons as possible to promote the company. The era of Internet communications is the best time for an open, honest business that satisfies consumers. If the client is properly motivated, they will deliver information to a huge number of potential consumers [3].
Virtual space monitoring allows companies to quickly respond to changing audience needs and receive feedback for decision-making. Monitoring with operational measurement of the mood of the company's users and customers; research of competitors' activities; evaluation of the effectiveness of marketing campaigns; participation in discussions and providing feedback [4]. Solving these tasks allows the brand to function effectively in the media space, develop new relevant strategies and ways to manage its reputation [5].

Reputation management is based on measuring the tone of the authors’ reviews [6].

One of the most relevant areas of work with reviews is research in terms of the review's quality, the mood of the review, and the reviewer's characteristics. The research shows that the reviewer's characteristics are good predictors of the usefulness of reviewing [7]. The quality of reviewing and the reviewer’s mood are the opposite. On the contrary, in other research, the information quality and the review's usefulness vary depending on the mood [8].

In the results of a study of another work, the text's readability and the reviewer's characteristics affect the perceived value of reviews [9].

2. Problem statement
We have a review sample of one bank in the dataset, and hypothetically, by default, reviews should have the same characteristics. Applying the Gunning fog-index method, we determine that they have different quantitative values. Moreover, using the graphical method (figure 1), we can visually see the difference between reviews with ratings "1-4" and reviews with ratings "5". Therefore, we need to prove the statistical significance of the differences between the subgroup of reviews rated "1-4" and the second subgroup of reviews rated "5".

3. Purpose of the study
Compare review texts with different ratings using quantitative methods and determine the difference or similarity between them.

4. Research methods
In this paper, we will analyze reviews of the Russian agricultural Bank (the object of research, one of the largest banks in the Russian Federation, specializing in the financing of enterprises in the agricultural sector). We collected 6802 customer reviews (valid N=6802) about banking activities on the site banki.ru and analyzed these reviews' intelligibility on a scale from 1 to 5 (where five is the best rating). We will use the Gunning fog-index (GFI) (the main method) to check the "intelligibility" of the human perception of the reviews left. The GFI method calculates in a modified Textstat package in the Python programming language. We added a correction factor = 0.78, applicable to the Russian text analysis [10]. Compound words are words that consist of more than two syllables. We built several descriptive diagrams. We used the Pandas library (Python) for data analysis and the Matplotlib library for plotting in this work.

\[ GFI = 0.4[0.78\left(\frac{\text{words}}{\text{sentences}}\right) + 100\left(\frac{\text{complex words}}{\text{words}}\right)] \]

The following GFI values are accepted: 70 and higher – no special training is required; up to 70-secondary education; up to 60 – an intellectual level of training; up to 30-scientific level is required for understanding. The lower the GFI, the more complex the text; conversely, the higher the index, the easier the text is perceived.

The central hypothesis is that reviews that are rated "5" (hypothesis 1) or "4" (hypothesis 2) are likely to differ from other reviews since more positive reviews have a less emotional colour and are an act of politeness rather than a note of protest and indignation. Such reviews are characterized by complexity, many words, and the text's scientific nature. It should be noted that emotions are essential for consumers in online reviews [11].
We used the Mann-Whitney U statistical test, The Wilcoxon signed-rank T test nonparametric statistical test, and the Pearson correlation coefficient to compare subsamples.

5. Research result
The average values of the GFI values: (a:b), where a is a review with a rating; b is the average GFI rounded to an integer: (1:64) (2:60) (3:56) (4:53) (5:27.1). Graphically, the average GFI distribution shows in figure 1, where we can see that reviews with a rating of "5" have a Gunning Fog-index, on average twice as low. Accordingly, reviews with a rating of "5" are not entirely readable.

![Figure 1. Distribution of average GFI by rating review (from 1 to 5).](image)

The statistical Mann-Whitney T test was used to compare pairs of groups of clients who gave different ratings based on banking services (the rating that the client provides after banking services is from 1 to 5 (1 is very bad; 5 is excellent)). The Mann-Whitney U-test calculation allows us to identify differences in the value of a parameter between the samples. It is a nonparametric statistical criterion used to compare two independent variables by the level of a quantified trait [12]. The GFI value was used as the dependent variable.

6. Research questions
So, using the calculation of the Mann-Whitney statistical criterion, two pairs of groups were compared:

- 1st pair: customers who gave ratings "1", "2", "3" and "4" (first subsample) and customers who rated "5" (second subsample);
- 2nd pair: customers who scored "1", "2" and "3" (the first subsample) and customers who scored "4" (the second subsample).

Since we know that the more scientific and incomprehensible the review text is, the lower the GFI (and Vice versa, the simpler and more straightforward the reviews are written, the higher the GFI), we assume that:

- in 1st pair of subsamples: in the subsample of clients who gave ratings "1", "2", "3" and 4 " (first subsample), GFI is higher than in the subsample of customers who rated "5" (second subsample);
• in 2\textsuperscript{nd} pair of subsamples: in the subsample of clients who scored "1", "2" and "3" (the first subsample), the GFI is higher than in the subsample of clients who scored "4" (the second subsample).

6.1. Comparison of ranks

\textbf{Table 1.} Comparison of ranks of subsamples of assessments "1-4" with the subsample of "5".

| Subsamples (Group 1) | N   | Average rank | The sum of the ranks |
|----------------------|-----|--------------|----------------------|
| GFI                  |     |              |                      |
| 1, 2, 3, 4 score rating | 3244 | 2785.56      | 9036351.00           |
| 5 score rating       | 1416 | 1287.98      | 1823779.00           |
| Total                | 4660 |              |                      |

A comparison of the values of the "average rank" corresponding to the subset of customers that have put assessment "1", "2", "3" and 4" (the first subsample) and the subsample of clients who put a score of "5" (the second subsample) shows that in the first subsample of GFI is higher (table 1).

\textbf{Table 2.} Statistical tests for comparison of the subsamples of estimates, "1-4" with the subsample of "5" on the value of the Gunning fog index

| Index                          | Value       |
|--------------------------------|-------------|
| Mann-Whitney U test            | 820543.000  |
| Wilcoxon signed-rank T test    | 1823779.000 |
| Z-Score                        | -34.949     |
| Asymptotic 2-sided significance| .000        |

This difference's statistical significance is confirmed by the value of the asymptotic significance (table 2). Since it is much less than 0.05 (equal to 0.000), we can say that the difference in the GFI value of the studied subsamples is statistically significant.

\textbf{Table 3.} Comparison of the ranks of the reviews subsample with ratings "1-3" with the reviews subsample with rating "4".

| Subsamples (Group 2) | N   | Average rank | The sum of the ranks |
|----------------------|-----|--------------|----------------------|
| GFI                  |     |              |                      |
| 1, 2, 3 score rating | 3134 | 1634.24      | 5121716.50           |
| 4 score rating       | 110  | 1287.94      | 141673.50            |
| Total                | 3244 |              |                      |

A comparison of the values of the "average rank" corresponding to the subset of customers that put the rating "1", "2" and "3" (the first subsample) and the subsample of clients who set a rating of "4" (the second subsample) shows that in the first subsample of GFI is higher (table 3).

\textbf{Table 4.} Statistical tests for comparison of the subsamples of estimates, "1-3" with the subsample of "4" on the value of the Gunning fog index

| Index                          | Value       |
|--------------------------------|-------------|
| Mann-Whitney U test            | 135568.500  |
| Wilcoxon signed-rank T test    | 141673.500  |
| Z                              | -3.812      |
| Asymptotic 2-sided significance| .000        |

This difference's statistical significance is confirmed by the value of the asymptotic significance (table 4). Since it is much less than 0.05 (equal to 0.000), we can say that the difference in the GFI value of the studied subsamples is statistically significant.
Whether there is a pattern: "the higher the score," the lower the GFI value is also confirmed by the correlation coefficient's value between the corresponding indicators.

Since the studied indicators are quantitative (the scales are metric), and the distribution of indicators is normal, the Pearson parametric correlation coefficient was calculated to identify and assess the tightness of the relationship between the series of comparable quantitative indicators of the questionnaire.

The Pearson correlation coefficient is a parametric method that is used to study the relationship between phenomena statistically. When using the correlation coefficient, the tightness of the relationship between features is conditionally estimated, counting [13]:

- coefficient values equal to 0.3 or less are indicators of weak tightness of communication;
- values greater than 0.4 but less than 0.7 are indicators of moderate tightness,
- values of 0.7 or more are indicators of high tightness of communication.

### Table 5. Pearson Correlation between "Rating" and "GFI".

|       | Rating Pearson Correlation | Rating N | GFI Pearson Correlation | GFI N |
|-------|---------------------------|---------|-------------------------|-------|
| Rating | 1                          | 6813    | -.214*                  | 6803  |
|        | Asymptotic 2-sided significance |  .000    |                         |       |
| GFI    | -.214*                    | 6803    | 1                       | 6803  |
|        | Asymptotic 2-sided significance | .000     |                         |       |

*Correlation is significant at the 0.01 (2-sided).

Thus, the Pearson correlation coefficient between the "Rating" and "GFI" indicators was = -0.214 (the correlation is significant at the level of 0.01), table 5. It means that there is a weak negative relationship between the indicators.

### 7. Findings

Given the direction of the scales of the indicators (for the indicator "rating": the higher the score, the higher the indicator value for the indicator "GFI": the higher the score, the more simple and straightforward opinion), showed correlations should be interpreted in the following way: feedback with a higher rating less intuitive and straightforward (contain more complicated and pseudo-scientific texts), and Vice versa, the reviews with a lower rating more simple and straightforward. Thus, we can reject the null hypothesis, and our two hypotheses turned out to be correct.

### 8. Conclusion

Summing up, it can be noted that GFI is a universal tool for analyzing text information. Thanks to the method, it is possible to determine in advance whether the target audience will perceive this or that review. The disadvantages of GFI are that the method does not consider the analysis of the text's content but considers only its quantitative features. Therefore, if you have extensive text data, then we recommend using word2vec to determine semantic proximity. This study was based on a sample of N=6802, and as N increases, other methods can be applied to analyze the data.

### Acknowledgement

The reported study was funded by RFBR, project number 20-310-70042.

### References

1. Portmann E, Meier A, Cudré-Mauroux P and Pedrycz W 2015 FORA—A fuzzy set-based framework for online reputation management Fuzzy sets and systems 269 90-114
2. Taeuscher K 2019 Reputation and new venture performance in online markets: The moderating role of market crowding J of Business Venturing 34(6) 105944
[3] Rastorguev S V and Tyu Yu S 2019 Digitalization of the Russian economy: trends, personnel, platforms, challenges to the state Public opinion monitoring: economic and social changes 5(153) 136-61
[4] Averin A V and Grigorieva V V 2018 Monitoring and official response as tools for improving the effectiveness of the company's online communication with users of social networks Bulletin of the Moscow University of Finance and law 1 69-79
[5] Belyakov M V 2016 Analysis of news reports on the website of the Russian Foreign Ministry by the method of sentiment analysis (article 2) Bulletin of the Peoples' Friendship University of Russia. Series: Theory of language. Semiotics. Semantics 4 115-24
[6] Giachanou A, Gonzalo J and Crestani F 2019 Propagating sentiment signals for estimating reputation polarity Information Processing & Management 56(6) 102079
[7] Lee P J, Hu Y H and Lu K T 2018 Assessing the helpfulness of online hotel reviews: A classification-based approach Telematics and Informatics 35(2) 436-45 https://doi.org/10.1016/j.tele.2018.01.001
[8] Chua A Y and Banerjee S 2016 Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality Computers in Human Behavior 54 547-54 https://doi.org/10.1016/j.chb.2015.08.057
[9] Fang B, Ye Q, Kucukusta D, and Law R 2016 Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics Tourism Management 52 498-506
[10] Litvinova T, Zagorovskaya O, Litvinova O and Seredin P 2016 Profiling a Set of Personality Traits of a Text’s Author: A Corpus-Based Approach Speech and Computer SPECOM 2016 Lecture Notes in Computer Science 9811 ed A Ronzhin, R Potapova and et al. (Cham: Springer) p 555-62 https://doi.org/10.1007/978-3-319-43958-7_67
[11] Malik M S I and Hussain A 2017 Helpfulness of product reviews as a function of discrete positive and negative emotions Computers in Human Behavior 73 290-302
[12] McKnight P E, and Najab J 2010 Mann-Whitney U Test The Corsini encyclopedia of psychology
[13] Alferes V R and Kenny D A 2009 SPSS programs for the measurement of nonindependence in standard dyadic designs Behaviour Research Methods 41(1) 47-54