NATURAL LANGUAGE PROCESSING TO ANALYZE USER FEEDBACK TO IDENTIFY TRENDS IN LARGE VOLUMES OF TEXTUAL BUSINESS DATA

Bohdan PETRENKO, master student
ORCID ID: 0009-0000-1576-0043
Semen Kuznets Kharkiv National University of Economics,
Kharkiv, Ukraine

In today's digital world, user feedback analysis has become a critical element for any business looking to improve the quality of its products and services. However, the scale and unstructured nature of this data pose significant challenges for traditional analysis. The main problem solved in this study is the inefficiency and limitations of manual analysis of large volumes of text user reviews. This problem has several important aspects:

1. Exponential growth in data volumes. Modern companies receive thousands of new reviews every day through various channels. According to the data, the number of online reviews is increasing by about 18% each year, which is almost double compared to 2020 (10%).

2. Subjectivity and inconsistency of analysis. Human interpretation of a text is subjective, which can lead to different conclusions about the same response. Independent studies show that the level of consistency between different analysts in assessing the tone of reviews is only 60-75%, and when identifying specific problems mentioned in reviews, this figure drops to 55% [1].

3. Critical time costs. Manual analysis requires significant time resources. According to experts, an analyst spends an average of 2-3 minutes on a thorough analysis of one medium-length review, which for a company with 1000 daily reviews means 30-50 hours of work daily, or 5-8 analysts.

4. Limited opportunities to detect hidden patterns. Analysts are often unable to identify hidden trends and relationships in large data sets, especially when these patterns manifest themselves over a long period of time or in different user segments. The study found that most valuable insights from user reviews go undetected in manual analysis, especially when it comes to correlations between different aspects of the product and user reactions.

5. The increasing complexity of multilingual analysis. The globalization of business has led to the need to analyze reviews in many languages. By 2025, this figure is projected at the level of 55-60%.

6. Lack of standardization and subjectivity of assessments. Different analysts may use different methods and evaluation criteria, making it difficult to compare results and form a unified strategy for responding to feedback.

7. Growing costs for review analytics. Economic studies show a rapid increase in the cost of manual review analysis. It is estimated that the average annual review analysis cost for the average company has risen to $280,000 in 2024, and is projected to be between $350,000 and $400,000 by the end of 2025.

The purpose of the study is to develop a system for automated analysis of user feedback based on natural language processing techniques to identify sentiment, key aspects and topics. In today's digital landscape, the amount of text data generated by users of products and services is growing rapidly. In particular, user reviews have become an integral element of the modern business ecosystem. The ability to effectively analyze these reviews has become a critical success factor for businesses looking to remain competitive. Traditional methods of analyzing reviews based on manual processing face many limitations. Analysts can only efficiently process a limited number of texts, and the subjectivity of human interpretation leads to inconsistencies in results. From a business perspective, user reviews provide invaluable information about: the strengths and weaknesses of a product or service; expectations and needs of customers; comparison with competitors; ideas for improvement and innovation; problems that need to be solved immediately. However, extracting this information from a large array of unstructured text data remains a challenging task. This is where natural language processing (NLP) technologies come in, an interdisciplinary field that integrates computer science, artificial intelligence, and linguistics to enable interaction between computers and human language. The use of NLP to analyze user feedback is a complex task that includes a number of subtasks: pre-processing of text; tonality detection; identification of key aspects and topics; classification of reviews into different categories; identification of emotional coloring; summarizing and generating reports. The tasks require the use of various models and algorithms – from classical statistical methods to modern deep learning approaches using transformers and large language models (LLMs). To solve the tasks, a comprehensive methodological approach will be used, which combines the methods of machine learning, deep learning, natural language processing and data analysis. The study will be carried out in several stages:

1. Analytical stage. At this stage, a detailed analysis of existing methods and models, their advantages and limitations will be carried out;

2. Design and development stage. At this stage, specific methods and models will be developed to solve the tasks;

3. Experimental stage. This stage includes conducting experiments to evaluate the effectiveness of the methods and models developed. Standard metrics such as accuracy, completeness, F1-measure will be used for

evaluation, as well as specialized metrics to assess the quality of the detection of aspects and topics;

4. Analytical-generalizing stage. At this stage, the results of experiments will be analyzed, patterns will be identified and conclusions will be formulated regarding the effectiveness of the developed methods and models.

A study conducted on the use of natural language processing to automate the analysis of user feedback provided a comprehensive view of the effectiveness, advantages, and limitations of various NLP methods and models in solving this problem. As a result of the study, all the goals and tasks were achieved, which made it possible to formulate the following conclusions: the analysis of the current state of the problem confirmed the critical importance of automating the analysis of user feedback in the context of the constant growth of text data; traditional manual analysis faces fundamental limitations related to time costs, subjectivity of assessments and the impossibility of efficiently processing large amounts of data; These limitations are especially acute in the context of business globalization, when companies receive feedback in different languages and from different sources. A theoretical study of natural language processing methods has demonstrated the evolution of approaches to text analysis - from simple dictionary methods to complex transformer architectures. Each of these approaches has its advantages and limitations, and the choice of a specific method depends on the specifics of the task, available computing resources, and requirements for accuracy and speed of analysis.

Thus, the conducted research makes a significant contribution to the development of methods for automating the analysis of user reviews based on natural language processing and creates the basis for further improvement of such systems.

References

1. Stanford Sentiment Treebank. URL: https://nlp.stanford.edu/sentiment/.
2. State of AI Report. URL : https://www.stateof.ai/.
3.TensorFlow Documentation. URL: https://www.tensorflow.org/api_docs.