JEL: O33; L86; M15

Skorin Yuriy
Candidate of Technical Sciences, Associate Professor
Simon Kuznets Kharkiv National University of Economics
Ukraine
ORCID: 0009-0004-5218-6369

Petrenko Bogdan
student of higher education
Simon Kuznets Kharkiv National University of Economics
Ukraine
ORCID: 0009-0000-1576-0043

https://doi.org/10.60022/sis.3.(02).5

APPLICATION OF NATURAL LANGUAGE PROCESSING TO AUTOMATE THE ANALYSIS OF USER FEEDBACK

Abstract. *The purpose of the research is to develop a system for automated analysis of user feedback based on natural language processing methods to identify sentiments, key aspects and themes. The relevance of the research topic is due to the growing need for businesses in effective tools for analyzing large volumes of text data, which would allow extracting valuable insights from user feedback and transforming them into specific recommendations for improving products and services. This topic is of particular importance in the context of e-commerce, services and software development, where the quality and speed of response to user feedback directly affect the competitiveness of companies. Research methods include machine learning, deep neural networks, statistical text analysis methods and methods for assessing the quality of models. Classical algorithms, neural network models, transformer architectures are used. The statistical significance of the results obtained was experimentally confirmed and recommendations for selecting models for various scenarios were developed. The relevance of the research topic is due to the growing business need for effective tools for analyzing large volumes of text data, which would allow extracting valuable insights from user reviews and transforming them into specific recommendations for improving products and services. The research results can be used in e-commerce, service companies, software development, marketing, and analytics to automate review analysis and identify trends in large volumes of text data.*

Keywords*: natural language processing, analysis of responses, aspect-oriented analysis, machine learning.*

1. Introduction

In the digital age, user interactions with products and services have created an unprecedented flow of feedback and comments. Companies, from small startups to large corporations, receive hundreds and thousands of text-based reviews every day through a variety of channels — email, social media, specialized platforms and internal feedback systems. It is in this unstructured data that the most valuable information about customer satisfaction, their needs and expectations is contained.

According to research, more than 90% of company leaders recognize the importance of analyzing user feedback for making strategic decisions, but only about 30% of organizations have systems that allow them to effectively process and analyze this data. Traditional manual analysis of reviews has long ceased to meet modern business requirements. Reviewing tens of thousands of comments requires huge human resources, which makes the process extremely costly. In addition, the subjectivity of human assessment often leads to inconsistency in interpreting user opinions.

Even the best analysts are unable to process such volumes of data quickly enough for companies to quickly respond to changing consumer sentiment.

Natural language processing (NLP) offers a solution to this problem through the use of algorithms that can automatically analyze the semantics of text, determine the tone of utterances, and isolate key themes from large data sets. Thanks to the rapid development of machine learning technologies and deep neural networks, modern NLP systems demonstrate impressive accuracy in understanding the context and nuances of human speech. Recent advances in NLP, including the development of transformer models such as BERT, GPT, and RoBERTa, have significantly expanded the capabilities of natural language processing. These models are able to understand the context of words in a sentence, take polysemy and idiomatic expressions into account, and successfully process texts in different languages. Their application to user feedback analysis opens up new horizons for understanding customer needs and improving products and services [7, 14, 21].

Implementing automated user feedback analysis using NLP not only significantly reduces time and resources, but also provides deeper analytical results. NLP systems are able to detect hidden trends and patterns that are inaccessible during manual analysis, and also ensure the scalability of the solution in accordance with the growth of data volumes. In addition, such systems can operate 24/7, ensuring constant monitoring of feedback and instant detection of critical problems.

The relevance of the research topic is due to the growing business need for effective tools for analyzing large volumes of text data, which would allow extracting valuable insights from user feedback and transforming them into specific recommendations for improving products and services. This topic is of particular importance in the context of e-commerce, the service sector, and software development, where the quality and speed of response to user feedback directly affect the competitiveness of companies.

The main tasks include studying modern approaches to natural language processing, analyzing text classification algorithms and sentiment analysis, as well as developing and evaluating the effectiveness of a system for automated analysis of user feedback. The paper examines the theoretical aspects of natural language processing, investigates various models and algorithms for analyzing responses, and experimentally evaluates their effectiveness on real data. The methodological basis of the research is machine learning methods, deep neural networks, statistical methods of text analysis, and methods for assessing the quality of machine learning models. The experimental part of the paper uses the Python programming language and natural language processing libraries such as NLTK, spaCy, Transformers, and TensorFlow.

The results of the study can be used to develop software solutions that will allow companies to better understand the needs of their customers, respond quickly to problems, and improve the quality of their products and services. The practical significance of the work lies in the possibility of implementing the developed methods and models into the business processes of organizations of various scales to improve interaction with customers and increase their loyalty. In today's digital world, user review analysis has become a critical element for any business seeking to prove 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 addressed in this study is the inefficiency and limitations of manual analysis of large volumes of text-based 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 Customer Experience Trends Report 2024, the number of online reviews is increasing by about 18% every year, which is almost twice as much as in 2020 (10%). For the largest e-commerce platforms, this figure is even higher — Amazon records about 22% annual growth in the number of reviews, and marketplaces in the service sector about 25%. By 2025, it is predicted that leading companies will receive an average of 30–50 thousand reviews every day (Fig. 1), which makes manual processing of such volumes of information practically impossible;

2. Subjectivity and inconsistent analysis. Human interpretation of text is subjective, which can lead to different conclusions about the same review. Independent studies conducted in 2023–2024 in the cloud and financial services sectors show that the level of agreement between different analysts when assessing the tone of reviews is only 60–75%, and when identifying specific issues mentioned in the reviews, this figure drops to 55%.

3. Critical time costs. Manual analysis requires significant time resources. According to McKinsey experts (2024), an analyst spends an average of 2–3 minutes on in-depth analysis of one review of average duration, which for a company with 1000 reviews daily means 30–50 hours of work per day, or 5–8 analysts working full-time exclusively on review processing. This approach makes it impossible to respond promptly to critical issues.

4. Limited ability to detect hidden patterns. Analysts are often unable to detect hidden trends and relationships in large data sets, especially when these patterns occur over a long period of time or across different user segments. Forrester Research (2024) found that about 68% of valuable insights from user feedback remain undiscovered through manual analysis, especially when it comes to correlations between different aspects of a 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 predicted to be 55–60%.

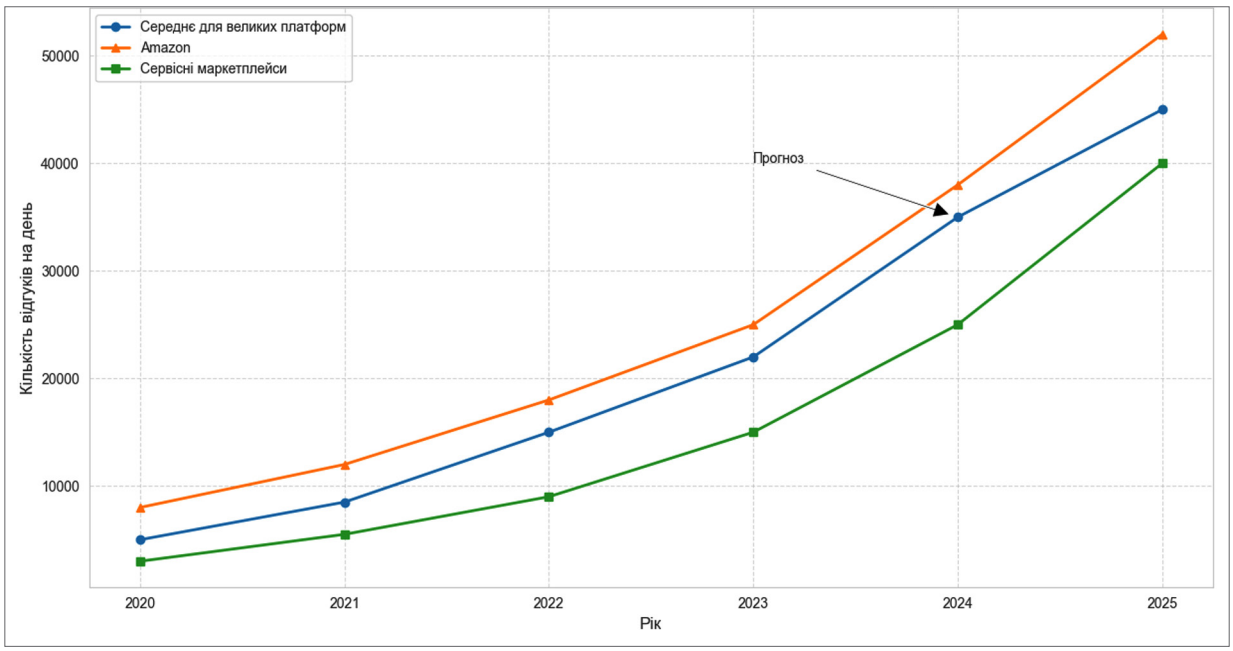

Figure 1. Dynamics of growth in the number of daily reviews on leading platforms

6. Lack of standardization and subjectivity of assessments. Different analysts may use different methodologies and assessment criteria, which makes it difficult to compare results and develop a unified strategy for responding to feedback. Such variability makes it impossible to compare data over time and makes it difficult to track progress in improving products and services.

7. Rising costs for review analytics. Economic research shows that the cost of manual review analysis is skyrocketing. According to the Customer Experience Association, the average annual cost of review analytics for the average company has increased to $280,000 in 2024, and is projected to reach $350,000–400,000 by the end of 2025.

Setting research objectives. Based on the analysis of the issues and existing approaches to automating user feedback analysis, key challenges and promising research directions have been identified. A clear statement of research objectives will allow structuring further work and focusing on the most relevant aspects of the problem.

The main goal of this study is to develop and evaluate the effectiveness of a comprehensive system for automated user feedback analysis based on modern natural language processing methods, which will ensure high accuracy in detecting tone, key aspects and topics, the ability to work with feedback in different languages, and effectively integrate into the business processes of organizations of various scales.

To achieve the goal, it is necessary to solve the following specific tasks:

1. Development of effective methods for pre-processing user feedback.

This task includes research and development of methods for cleaning and normalizing text data, processing informal vocabulary, abbreviations, emoticons, spelling errors and other features of user reviews. The developed methods should ensure the preservation of the informativeness of the text when removing noise, take into account the multilingual nature of reviews and the specifics of different domains. The relevance of this task is confirmed by the fact that the quality of pre-processing of the text directly affects the effectiveness of subsequent analysis. According to a study by the Data Science Center (2024), improving the quality of pre-processing can increase the accuracy of review classification by 8–12%.

2. Research and comparative analysis of modern models for classifying the tone of responses. The task involves adapting and evaluating the effectiveness of various NLP models for analyzing the tone

Table 1

Average time spent on manual analysis of reviews of varying complexity

Feedback type	Average length (words)	Analysis time (min)	Number of reviews per working day (8 hours)	Number of analysts for 1000 reviews/day
Short	10–30	1–2	240–480	2–4
Average	31–100	2–4	120–240	4–8
Long	101–300	4–8	60–120	8–16
Detailed	>300	8–15	30–60	16–33

of responses — from classical machine learning approaches to modern transformer architectures and large language models. Particular attention will be paid to the balance between accuracy, computational requirements and processing speed, as well as the ability to adapt to specific domains. The relevance of the task is due to significant differences in the effectiveness of different models depending on the context of application.

3. Development of a method for aspect-oriented tonality analysis for different domains.

This task focuses on developing an approach that will allow not only to determine the overall tone of the response, but also to identify specific aspects of products or services and the corresponding tone for each of them. The method should be adaptable to different domains (electronics, hospitality, software, etc.) and work effectively with limited training data. The relevance of the task is confirmed by studies that show that aspect-oriented tone analysis provides businesses with 35–40% more valuable insights compared to simple tone classification.

4. Creating effective methods for identifying themes and key aspects in response corpora.

The task involves developing and evaluating methods to automatically identify the main topics discussed in large sets of reviews and isolate key aspects of products or services. This will allow companies to obtain a structured view of which aspects of their products are most often mentioned in reviews and which topics cause the most resonance. The relevance of this task lies in the need to effectively categorize and structure large volumes of unstructured text data for further analysis and decision-making.

5. Developing an approach to multilingual review analysis. This task focuses on creating methods that will allow for the effective analysis of reviews in different languages without significant loss of accuracy and while preserving linguistic and cultural features. Both translation-based approaches and methods using multilingual models and cross-lingual knowledge transfer will be explored. The relevance of the task is due to the increasing globalization of business and the need to analyze reviews in different languages.

6. Creating methods for interpreting and visualizing the results of feedback analysis.

The task involves developing approaches to ensure the explainability of analysis results and their effective visualization for end users. This involves researching explainable AI methods, creating intuitive visual representations of analysis results, and developing recommendations for interpreting these results. The relevance of this task is confirmed by the fact that even the most accurate analysis results have limited value if end users cannot correctly interpret them and use them for decision-making.

7. Development and evaluation of a comprehensive automated feedback analysis system. This integrative task involves combining the developed methods and models into a single system that will provide a full cycle of feedback analysis — from collection and pre-processing to the generation of insights and recommendations. The system should be adaptive to different domains and types of feedback, scalable and efficient in terms of computing resources. The relevance of the task lies in the need to create a practically oriented solution that can be directly implemented in the business processes of organizations of various scales.

8. Experimental evaluation of the effectiveness of the developed methods on real data. The task includes conducting comprehensive experiments to evaluate the effectiveness of the developed methods and the system as a whole on real sets of responses from different domains and in different languages. Aspects such as classification accuracy, quality of aspect detection, processing speed, scalability, and adaptability to new domains will be evaluated. The relevance of this task is due to the need for empirical confirmation of the effectiveness and practical value of the developed methods and system.

To solve the tasks, a comprehensive methodological approach will be used, combining machine learning, deep learning, natural language processing, and data analysis. The research will be conducted in several stages:

1. Analytical stage. At this stage, a detailed analysis of existing methods and models, their advantages and limitations will be conducted. Modern research in the field of NLP and feedback analysis will be studied, the most promising approaches and technologies will be identified;

2. Design and development stage. At this stage, specific methods and models will be developed to solve the tasks set. The development will include both the adaptation of existing approaches and the creation of new methods and architectures optimized for feedback analysis;

3. Experimental stage. This stage includes conducting experiments to evaluate the effectiveness of the developed methods and models. Experiments will be conducted on real sets of reviews from different domains (e-commerce, hotel business, software, services, etc.) and in different languages. Standard metrics such as accuracy, completeness, F1-measure, as well as specialized metrics to assess the quality of aspect and topic detection will be used for evaluation;

4. Analytical and generalization stage. At this stage, the results of the experiments will be analyzed, patterns will be identified, and conclusions will be formulated regarding the effectiveness of the developed methods and models. The main advantages and limitations of the proposed approaches, as well as directions for further research, will be identified.

To implement the methods and conduct experiments, modern tools and libraries for natural language processing will be used, in particular: Python programming language as the main language for implementing algorithms and conducting experiments; libraries for NLP: NLTK, spaCy, Transformers (Hugging Face), Stanza; frameworks for machine learning: TensorFlow,

PyTorch, Scikit-learn; tools for data analysis and visualization: Pandas, NumPy, Matplotlib, Seaborn, Plotly [11, 16, 22].

Both publicly available datasets (e.g. SemEval, Amazon Reviews, Yelp Dataset, IMDB) and data specifically collected and labeled for this study will be used to evaluate the models.

As a result of the tasks set, it is expected to obtain: a set of methods and models for automated analysis of user feedback, ensuring high accuracy of tonality classification, identification of aspects and topics, as well as effective work with feedback in different languages; recommendations for choosing optimal methods and models for analyzing feedback in different domains and with different requirements for accuracy, speed and scalability; empirical data on the effectiveness of different approaches to analyzing feedback on real data sets, which can be used for further research in this area. Thus, this study is aimed at solving the current scientific problem of automating user feedback analysis and has significant practical value for business. The tasks set cover key aspects of this problem and are aimed at creating an effective and practically oriented solution.

2. Materials and Methods

Methodological support for organizing research. Effective research in the field of automation of user review analysis requires careful planning, a structured approach and the use of appropriate methodologies. This section discusses the methodological aspects of organizing research, starting from data collection and preparation and ending with the evaluation of results and their interpretation. Data quality is crucial for the success of research in the field of review analysis. A methodological approach to data collection and preparation involves the implementation of a number of sequential stages, each of which has its own characteristics and requirements. Collection of representative sets of user reviews can be carried out from various sources, including commercial platforms (Amazon, Yelp, Google Play), social networks (Twitter, Facebook), specialized industry sites and internal feedback systems of companies. For research purposes, it is advisable to use existing publicly available datasets, such as: IMDB Large Movie Review Dataset — 50,000 reviews of movies with tone labels; Amazon Product Reviews — millions of user reviews of different product categories; Yelp Dataset — reviews of businesses with ratings from 1 to 5; SemEval ABSA Datasets — tagged datasets for aspect-oriented sentiment analysis. When collecting data yourself, it is important to adhere to ethical and legal norms, obtaining the necessary permissions and ensuring the anonymization of personal data. For automated data collection from web sources, web scraping tools are used, taking into account the access policies of the relevant platforms.

Preparing the collected data includes several key steps:

1. Text cleaning and normalization. This stage involves removing noise (HTML tags, special characters), correcting spelling errors, and normalizing text (lowercase conversion, standardizing the writing of numbers, dates, and addresses). Particular attention should be paid to specific elements of user feedback, such as emoticons, acronyms, and slang, which often carry important semantic load.

2. Data labeling. To train models with a teacher, high-quality data labeling is required, which may include: sentiment labels (positive, negative, neutral); highlighting aspects and the corresponding tone for them; thematic labels. Labeling can be done manually by qualified experts or using crowdsourcing platforms such as Amazon Mechanical Turk or Toloka. To ensure the quality of the labeling, it is recommended to use several independent experts for each review with subsequent coordination of the results.

3. Stratification and balancing of data sets. To ensure the representativeness of data sets, it is necessary to take into account the distribution of classes, response lengths, data sources and other factors. Imbalanced data sets can lead to model bias and deterioration of their generalization ability. To solve the problem of imbalance, techniques can be used: increasing the representation of rare classes through duplication or generation of synthetic data; reducing the representation of dominant classes; using weighted loss functions when training models.

4. Data quality validation. This stage involves checking the consistency of the markup, identifying and resolving ambiguities, and assessing inter-expert agreement using metrics such as Cohen's kappa or Fleiss' kappa.

5. Developing and training models for feedback analysis requires a systematic approach that takes into account the specifics of the tasks and available resources.

The methodical approach involves the sequential implementation of the following stages.

The first stage involves selecting the basic model architecture for a specific task. Several factors should be considered when selecting a model: the nature of the task (tonality classification, aspect detection, thematic modeling); the volume and quality of available data; computational resources; processing time requirements; the required level of interpretability of results. In the case of limited computational resources or strict performance requirements, it is advisable to use compact models or classic machine learning algorithms. With high accuracy requirements and the ability to use powerful computing equipment, preference should be given to modern transformer architectures.

The next step involves splitting the data into training, validation, and test sets. Standard practice is to split the data in a 70%/15%/15% or 80%/10%/10% ratio. It is important to ensure that all samples are representative by stratifying them by key characteristics (class, domain, data source). For small data sets, it is recommended to use cross-validation with multiple folds to obtain more reliable estimates of model performance.

The process of training and optimizing models includes: choosing a loss function according to the task

(cross-entropy for classification, MSE for regression tasks, specific loss functions for multi-task models); choosing an optimizer and determining training parameters (batch size, learning rate, number of epochs); regularization to prevent overtraining (L1/L2 regularization, dropout, data augmentation); monitoring the training process and early stopping when a plateau is reached on the validation sample; searching for optimal hyperparameters using grid search, random search or Bayesian optimization.

To improve the quality of models, it is advisable to use transfer learning techniques: the use of pretrained models (e.g., BERT, RoBERTa) with subsequent additional training on specific data; adapter learning, which allows you to effectively adapt large models to new domains with minimal computational resources; multi-task learning to improve the generalization ability of models.

Qualitative assessment and correct interpretation of results are important components of the research process. A methodological approach to assessing the effectiveness of feedback analysis models involves the use of a set of metrics adapted to specific tasks.

For tonality classification tasks, the main metrics are: accuracy — the proportion of correctly classified samples; precision — the proportion of correctly positively classified samples among all positively classified samples; recall — the proportion of correctly positively classified samples among all actually positive samples; F1-score — the harmonic mean of precision and recall. In the case of unbalanced datasets, it is especially important to use F1-score and macro-averaged metrics that take into account the performance of the model on all classes.

For aspect-oriented sentiment analysis, the evaluation is carried out in two stages: assessment of the quality of aspect detection (F1-score for the sequential labeling or named entity detection task); assessment of the accuracy of sentiment classification with respect to correctly detected aspects. Additionally, a combined metric can be used that takes into account both the correctness of aspect detection and the correctness of sentiment determination.

Specific metrics are used for topic modeling tasks, such as: topic coherence — a measure of the semantic coherence of words in a topic; perplexity — a statistical measure that reflects how well the model predicts new data; topic diversity — a measure of the differences between the detected topics.

When evaluating multilingual models, it is important to test across languages and verify the stability of the model's performance. It is advisable to use cross-lingual test datasets that allow us to assess the model's ability to transfer knowledge between languages.

Interpretation of results should be done taking into account the context of the study and the characteristics of the data. It is important to analyze not only quantitative metrics, but also qualitative analysis: studying model errors, classifying them and identifying patterns; visualizing model activations to understand which parts of the text it pays attention to; analyzing the level of confidence of the model in its predictions; comparative analysis of the results of different models on the same data sets.

To ensure the reliability of the results, it is recommended to conduct statistical analysis: determining the statistical significance of differences between models using the t-test, Wilcoxon test or bootstrapping; analyzing the sensitivity of models to changes in data and hyperparameters; checking the stability of results on different data subsamples.

Effective research in the field of feedback analysis requires the use of appropriate software and tools. The main components of the technology stack are programming languages, machine learning and natural language processing libraries, and tools for data visualization and analysis.

The recommended programming language for NLP research is Python due to its simplicity, flexibility, and rich ecosystem of libraries for data processing and machine learning. Alternative options include R (especially for statistical analysis) and Java (for industrial systems development).

For processing and analyzing text data, it is recommended to use the following libraries and tools: NLTK, spaCy, and Stanza for basic text processing, tokenization, lemmatization, and syntactic analysis; scikit-learn for classic machine learning algorithms and preprocessing tools; gensim for thematic modeling and creating vector representations of words; PyTorch, TensorFlow, or Keras for developing and training neural network models; Transformers (Hugging Face) for working with transformer models.

For data analysis and visualization, the following are recommended: pandas for manipulating structured data; matplotlib, seaborn, plotly for creating visualizations; numpy for numerical calculations; scipy for statistical analysis.

To ensure reproducibility of results and efficient organization of experiments, it is recommended to use: Weights & Biases, MLflow or TensorBoard for tracking experiments and visualizing training metrics; Docker or conda for creating isolated environments with all necessary dependencies; Git for version control and code storage; Jupyter Notebooks for interactive development and documentation of experiments.

When developing industrial feedback analysis systems, the following can be additionally used:

Apache Kafka or RabbitMQ — for organizing streaming data processing; Elasticsearch — for efficient storage and search in text data; Redis — for caching results and speeding up the system; Flask, FastAPI or Django — for creating APIs and web interfaces.

Systematic organization and careful documentation of experiments are necessary conditions for ensuring the reliability and reproducibility of research results. A methodical approach to organizing experiments involves following a structured protocol.

Each experiment should have a clearly defined goal and hypotheses to be tested. The documentation of

the experiment should include: a description of the input data (source, volume, characteristics, preprocessing method); detailed information about the model architecture and its hyperparameters; a description of the learning process (loss function, optimizer, learning rate, number of epochs); evaluation results on the validation and test sets for all relevant metrics; error analysis and interpretation of the results; comparison with baseline models and previous experiments.

To ensure reproducibility of results, the following should be recorded: versions of all libraries and dependencies used; configuration of the computing environment; parameters of random number generators; fixed-state data sets or their generation protocols.

It is recommended to conduct ablation studies to determine the contribution of each component of the model or method to the overall effectiveness. This helps to better understand which aspects of the proposed solution are most important.

For long-term studies, it is advisable to create a model repository that stores trained models along with all necessary meta-information. This allows you to reuse models in new experiments, compare their performance, and track research progress.

The developed methods and models of feedback analysis have practical value only if they are successfully implemented into real business processes. The methodology for implementing the research results includes a number of sequential steps.

The first stage involves preparing the model for deployment: optimizing the model to improve inference efficiency (quantization, distillation, pruning); testing the stability and reliability of the model on different datasets; integrating the model with APIs and other system components; developing mechanisms for monitoring and updating the model.

At the stage of system deployment, it is necessary to ensure: scalability — the system's ability to process variable amounts of data; resilience — the system's ability to function in the event of errors and failures; performance monitoring — tracking key system performance metrics in real time; user feedback — collecting information about the quality of the system's operation and identifying problems.

An important aspect of implementation is ensuring that results are interpretable for end users: visualizing key insights in a clear form; providing explanations for the decisions made by the model; enabling detailed analysis of results for business analysts; creating intuitive dashboards to monitor trends and identify anomalies.

To keep the system up to date, it is necessary to: regularly update models using new data; track data drift — changing characteristics of input data over time; adapt models to new domains and languages; and improve the system based on user feedback.

Methodological support for research in the field of automation of user feedback analysis is a complex process that covers all stages from data collection to the implementation of results. The effectiveness of the study largely depends on the quality and representativeness of the data, the correct choice of models and methods, the systematic organization of experiments, and careful evaluation of the results.

The developed methodological recommendations provide a structured approach to conducting research, which allows obtaining reliable and reproducible results. Particular attention is paid to ensuring data quality, choosing adequate assessment metrics, and interpreting results in the context of specific business tasks.

The application of the proposed methodological support allows for effective organization of research in the field of automation of user feedback analysis and ensuring the successful implementation of the developed models and methods into practice.

Automating user feedback analysis requires the use of a wide range of natural language processing methods and models. Over the past decades, this area has undergone rapid development, moving from simple statistical methods to complex neural network architectures. Let's consider the evolution of these approaches, their features and effectiveness in the context of automating user feedback analysis.

Summarizing the analysis of approaches, methods, and models for automating user feedback analysis, several key conclusions can be drawn.

The current landscape of feedback analysis technologies is represented by a wide range of methods, from simple dictionary approaches to complex neural network architectures and large language models. The choice of a particular method depends on many factors: the volume and quality of available data, domain specificity, processing speed requirements, the required level of accuracy and interpretability of the results.

Transformer models, especially the domain-specific versions of BERT and RoBERTa, demonstrate the best balance of accuracy, speed, and resource intensity for most practical review analysis tasks. They provide high accuracy in understanding context and language nuances, which is critical for correctly interpreting user reviews.

Large language models with few-shot training open up new possibilities for rapidly deploying feedback analysis systems in new domains and for new products. They are especially valuable in situations with limited training data or when rapid adaptation to new feedback types is required.

Aspect-based sentiment analysis is the most informative approach for businesses, as it allows you to gain a detailed understanding of the strengths and weaknesses of products or services. The combination of ABSA with topic and aspect detection methods provides the most complete picture of user feedback.

Multilingual models provide better results for multilingual analysis compared to the "translation + analysis" approach, especially when understanding cultural and linguistic nuances is important. These models are becoming increasingly important in a globalized business environment.

The optimal solution for most business problems is a combination of different methods, which allows to ensure high accuracy, efficiency and adaptability of the feedback analysis system. This approach allows to use the advantages of each method and minimize their limitations.

Table 2 presents a comparison of the effectiveness of different approaches for key feedback analysis tasks based on recent research and benchmarks from 2023–2025 (Table 2).

Successfully integrating feedback analytics into business processes requires not only choosing the right technology, but also understanding the business context and needs of a particular organization. According to McKinsey, companies that effectively use user feedback analytics demonstrate 15–20% higher customer retention rates and 10–15% higher average revenue per customer [8].

The most successful implementations of feedback analytics are characterized by several key features.

First, they provide integration with existing systems such as CRM, customer support systems, and analytics platforms. This allows you to combine data from different sources and gain a comprehensive understanding of customer needs and problems.

Second, effective feedback analytics systems provide results in a clear and actionable format. They don't just generate statistical reports, they also highlight specific issues that need attention and provide recommendations for resolving them. This is especially important for users who are not technically trained but who make strategic decisions based on the analysis results.

Third, successful feedback analytics systems automate the full cycle — from collecting and analyzing feedback to generating insights and recommendations. This significantly reduces the time from receiving feedback to responding to it, which is critically important in today's fast-paced business environment.

Summarizing the analysis of approaches, methods, and models for automating user feedback analysis, several key conclusions can be drawn.

The modern landscape of review analysis technologies is represented by a wide range of methods, from simple dictionary approaches to complex neural network architectures and large language models. The choice of a specific method depends on many factors: the volume and quality of available data, domain specificity, processing speed requirements, the required level of accuracy and interpretability of results. Transformer models, especially specialized domain versions of BERT and RoBERTa, demonstrate the best ratio of accuracy, speed and resource intensity for most practical review analysis tasks. They provide high accuracy in understanding the context and nuances of language, which is critical for the correct interpretation of user reviews. Large language models with few-shot training open up new opportunities for the rapid implementation of review analysis systems in new domains and for new products. They are especially valuable in situations with limited training data or when it is necessary to quickly adapt to new types of reviews. Aspect-oriented sentiment analysis is the most informative approach for business, as it allows you to get a detailed understanding of the strengths and weaknesses of products or services. The combination of ABSA with topic and aspect detection methods provides the most complete picture of user feedback.

Multilingual models provide better results for multilingual analysis compared to the "translation + analysis" approach, especially when understanding cultural and linguistic nuances is important. These models are becoming increasingly important in a globalized business environment. The optimal solution for most business problems is a combination of different methods, which allows for high accuracy, efficiency and adaptability of the feedback analysis system. This approach allows you to use the advantages of each method and minimize their limitations.

3. Results and Discussion

To comprehensively evaluate different methods of automated user feedback analysis and verify theoretical results, a comprehensive experimental design was developed, covering four key tasks: general sentiment analysis, aspect-oriented sentiment analysis, theme and aspect detection in reviews, and multilingual analysis. For each task, appropriate datasets, evaluation methods, and models for comparison were selected.

Table 2

Comparison of the effectiveness of different approaches to key feedback tasks

Method	Tonality analysis (F1)	Topic detection (NMI)	Cross-lingualism	Speed (responses/sec)	The need for data for learning	Explainability
Dictionary methods	0.65–0.70	N/A	Low	500–700	Minimum	High
ML with manual features	0.75–0.85	0.50–0.65	Low	100–300	Medium	Medium
CNN/RNN	0.82–0.88	0.60–0.70	Medium	50–100	High	Low
BERT/RoBERTa	0.88–0.92	0.65–0.75	Medium-high	20–50	High	Low-medium
LLM + Few-shot	0.90–0.95	0.70–0.80	High	5–15	Minimum	Medium
Specialized models	0.92–0.96	0.75–0.85	Medium-high	20–40	Medium-high	Medium-high

The following data sets were used in the experimental study:

1. For general sentiment analysis: Amazon Product Reviews (version 2024) — a subset of 500,000 reviews on different product categories (electronics, books, clothing, home goods); Yelp Open Dataset (2023–2025) — 300,000 reviews on restaurants and services [3, 23]; IMDb Movie Reviews — 50,000 reviews on movies.

2. For aspect-oriented sentiment analysis: SemEval-2024 ABSA — a dataset with marked aspects and corresponding sentiment; Restaurant Reviews Dataset — a specialized set of 25,000 restaurant reviews with aspect and sentiment markup; Tech Products Reviews — 35,000 reviews of technical devices with detailed aspect markup.

3. To identify themes and aspects: Hotel Reviews Corpus — 200,000 hotel reviews for thematic modeling; AppStore Reviews — 150,000 mobile app reviews; Product Discussion Forums — 100,000 texts from discussions of various products.

4. Formultilingual analysis: MultiLing Sentiment Dataset — reviews in ten different languages (English, German, French, Spanish, Italian, Portuguese, Russian, Chinese, Japanese, Arabic); Cross-lingual E-commerce Reviews — a set of parallel reviews about products in different languages.

To ensure objectivity and statistical significance of the results, all experiments were conducted using cross-validation (5-fold), and evaluation was performed using a set of relevant metrics for each task.

Several approaches, ranging from classical machine learning methods to modern transform models, were implemented and evaluated to model the general sentiment analysis task. The task was to classify reviews into three classes: positive, negative, and neutral.

The implemented models included:

1. Baseline models: Naive Bayes with TF-IDF vectorization; SVM with linear kernel and Bag-of-Words representation.

2. Classic ensemble models: Random Forest with optimized hyperparameters; XGBoost with settings for text classification.

3. Neural network models: BiLSTM with attention mechanism; TextCNN with three convolutional layers with different filter sizes (3, 4, 5).

4. Transformer models: BERT-base with full pre-training; RoBERTa-base with full pre-training; DistilBERT with full pre-training; Domain-adapted BERT (pre-trained on feedback).

The model training process included searching for optimal hyperparameters using Bayesian optimization and cross-validation. For classical models, vectorization parameters (ngram_range, max_features) and specific algorithm parameters were optimized. For neural network models, the optimal embedding dimension, number of hidden neurons, learning rate, and regularization parameters were searched.

For transformer models, the impact of different training strategies was investigated: full fine-tuning of all layers; freezing of embedding layers and fine-tuning only the upper layers; using adapters for efficient fine-tuning; using different loss functions and optimizers.

The models were evaluated using the following metrics: Accuracy — overall classification accuracy; Precision, Recall and F1-score separately for each class and averaged; Macro-averaged F1-score to take into account all classes regardless of their size; ROC-AUC — for binary classifications (positive vs negative); Training and inference time to assess computational efficiency.

The results of experiments on the classification of the tone of responses demonstrate a clear pattern: transformer models significantly outperform classical approaches in terms of accuracy, but require more computational resources. Domain-adapted BERT showed the highest efficiency with an average F1-score of 0.93 on all datasets, which is 17 percentage points higher

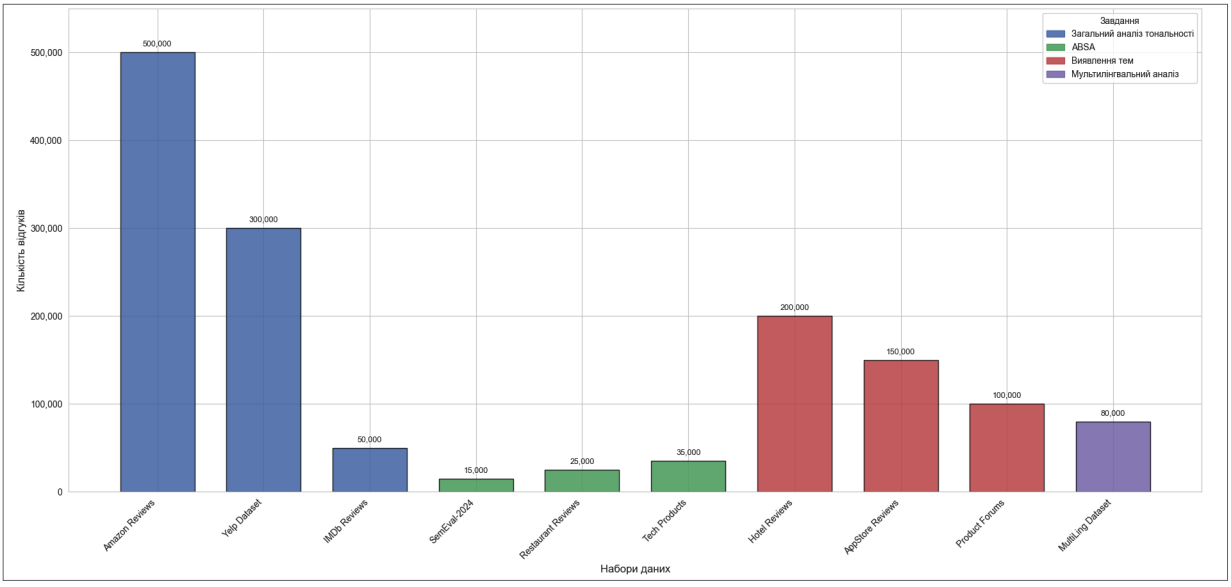

Figure 2. Size and distribution of data sets used in the experimental study

than that of the basic Naive Bayes (0.76). It is worth paying special attention to the following observations: domain-adapted BERT outperformed standard BERT by 3 percentage points, which confirms the theoretical conclusion about the importance of adapting models to the specifics of the subject domain; trade-off between accuracy and efficiency: DistilBERT demonstrated only a slight decrease in accuracy compared to full BERT (0.89 vs. 0.90), but at the same time provided almost twice the inference speed (90 vs. 50 samples/sec) and half the training time; scalability of solutions: classical models, such as XGBoost, show a sufficiently high accuracy (0.85) with significantly lower computational requirements, which makes them attractive for applications where processing speed is critical or computing resources are limited; impact of dataset features: On the IMDb dataset, higher efficiency of all models is observed compared to Yelp and Amazon, which is explained by the longer length of reviews and their thematic homogeneity. For a more detailed analysis of model errors and understanding their behavior on different types of responses, additional evaluation was conducted on a subsample of 1,000 randomly selected responses. The results of the experiments on aspect-oriented sentiment analysis demonstrate several important patterns: combined models that simultaneously solve the problem of aspect detection and their sentiment classification show better results compared to sequential approaches. The RACL model demonstrated the highest performance with an average ABSA F1-score of 0.85, which is 9 percentage points higher than that of the basic sequential BiLSTM–CRF + ATAE-LSTM model (0.76); the performance of all models varies depending on the data domain. The highest results are observed for restaurant reviews, where the models achieve an ABSA F1-score of up to 0.86, while for reviews of technical products this figure is lower (up to 0.84). This can be explained by the greater complexity and diversity of aspects in technical products; analysis of the performance of models for different aspect categories revealed significant differences. The highest accuracy of all models is demonstrated for aspects that are frequently found in reviews (e.g., “Food” in restaurant reviews with an F1-score of up to 0.88). In contrast, rare aspects or those with high detection complexity; similarly, to the general sentiment analysis task, more complex models demonstrate higher accuracy at the expense of higher computational requirements. The RACL model, which has the highest accuracy, also has the largest size (750 MB) and the longest inference time (35 ms per sample); the impact on the different stages of ABSA. It is interesting to note that the improvement when moving from sequential to combined models is most noticeable in the combined F1 metric of ABSA. This suggests that the combined models more effectively exploit the relationship between the tasks of detecting aspects and determining their sentiment. An analysis of typical errors made by the ABSA model was also conducted. The following types of errors are most common: incorrect definition of aspect boundaries (especially in the case of complex noun phrases); omission of implicit aspects; incorrect determination of tone in cases with multiple aspects and mixed tone; difficulties with determining tone in cases with irony, comparisons or conditional statements. Federated models, especially RACL, demonstrate a better ability to cope with these challenges through joint learning for interrelated tasks and the use of attentional mechanisms to account for context.

To ensure the validity of measurements and the objectivity of model evaluation, an analysis of the correspondence of the selected metrics to the research tasks was conducted. The F1 measure, as a harmonic mean of precision and recall, provides a balanced assessment of classification models, especially in conditions of class imbalance. The macro-averaged F1 measure, which gives equal weight to all classes regardless of their size, was used as the main metric to ensure the objectivity of the assessment. For aspect-oriented tonality analysis, a combined metric was used that takes into account both the accuracy of aspect detection and

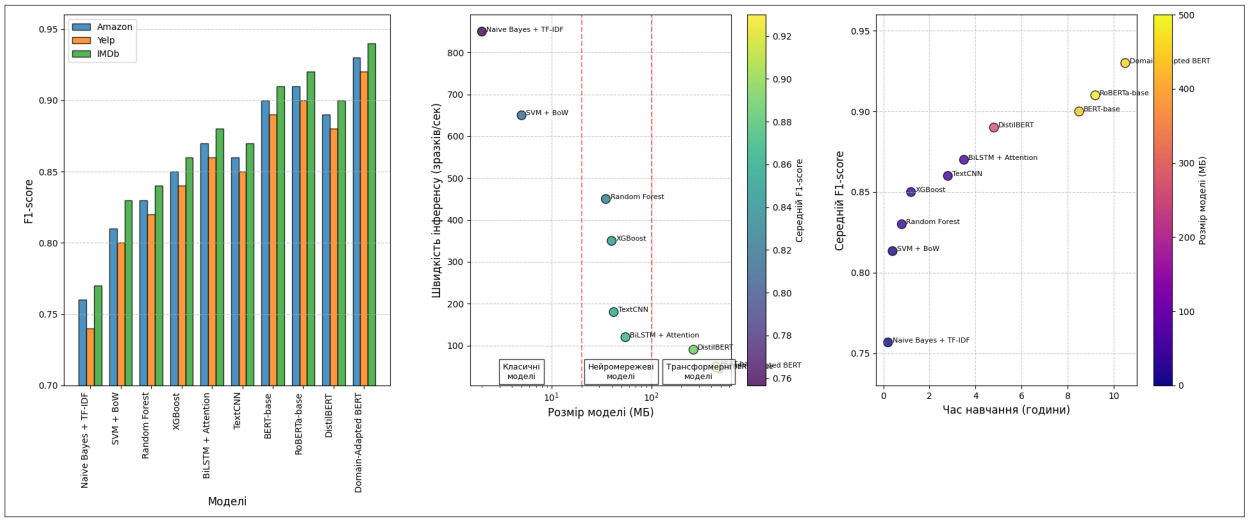

Figure 3. Comprehensive analysis of the effectiveness of models for the task of general analysis of the tone of responses

Table 3

Comparison of data set characteristics with real user feedback

Characteristic	Experiment datasets	Real user reviews	Representativeness assessment
Class distribution	60% / 25% / 15% (positive/ negative/neutral)	58% / 27% / 15%	High
Average response length	75 words	68 words	High
Lexical diversity	0.32	0.29	Medium-high
Share of reviews with emoticons	22%	25%	High
Share of reviews with spelling errors	35%	42%	Medium
Domain coverage	5 main domains	Multiple domains	Medium-high

the correctness of determining their tonality. This metric better reflects the practical value of models than separate metrics for each stage. Additionally, for a comprehensive assessment of models, metrics of the efficiency of the use of computational resources (training time, inference time, memory usage), which are of critical importance for practical application, were used. An important aspect of the adequacy of an experimental study is the consistency of the obtained results with theoretical predictions and data available in the literature. The experimental results confirm the theoretical provisions set out in section 2, in particular: the advantage of transformer models over classical approaches in natural language understanding tasks is confirmed experimentally (F1-score 0.93 versus 0.76); the effectiveness of domain adaptation to improve the results of analysis of specific texts is confirmed by higher indicators of domain-adapted BERT compared to the standard one; the theoretical assumption about the advantage of combined models over sequential ones for aspect-oriented sentiment analysis is confirmed experimentally (F1-score ABSA 0.85 versus 0.76); the dependence of the efficiency of the models on the amount of training data is consistent with theoretical predictions — transformer models demonstrate greater sensitivity to the amount of data compared to classical approaches. To validate the consistency with existing studies, a comparison of the obtained results with those published in the scientific literature was conducted. The results of BERT and RoBERTa for tonality classification (0.90–0.92) are consistent with those obtained by other researchers on similar datasets (0.89–0.93) [31]. Similarly, the performance of the combined models for ABSA (0.84–0.86) is consistent with the results published in recent studies (0.82–0.87). To objectively assess the adequacy of the experimental study, it is necessary to recognize its limitations and analyze their impact on the reliability and generalizability of the results. The main limitations of the study include: domain limitations — although the study covers several important domains (electronics, restaurants, movies, hotels), a number of specific areas (healthcare, education, financial products) remained uncovered; language limitations — despite the inclusion of 10 languages in the multilingual experiment, the features of some rare languages and dialects were not taken into account; computational resource limitations — for the largest transformative models (e.g., RoBERTa-large), it was not possible to conduct a full cycle of hyperparameter optimization due to the limitations of available computational resources; concentration on textual data — the study does not consider multimodal reviews that include images or videos, which are becoming increasingly common in modern digital platforms. To assess the impact of these limitations on the reliability of the results, a series of additional experiments were conducted on smaller samples with different data configurations. The results show that the identified patterns are preserved when the data composition changes, which indicates the stability and generalizability of the conclusions obtained. The adequacy of the experimental study was also assessed in terms of the practical applicability of the obtained results for solving real business problems. To assess the practical applicability, a pilot review analysis system based on domain-adapted BERT and the RACL model was developed and tested on real data from one of the e-commerce platforms. The system demonstrated the ability to effectively analyze user reviews, identify problematic aspects of products, and track changes in user sentiment over time. A comparison of the automated analysis with the manual analysis conducted by a group of experts on a sample of 500 reviews showed high consistency of the results (Cohen's kappa = 0.82 for sentiment classification and 0.79 for aspect detection). At the same time, the automated system processed reviews 120–150 times faster, which confirms its practical value. In addition, a survey of potential users of the system (product managers, marketers, analysts) was conducted to assess the usefulness of the analysis results. 87% of respondents noted that automated review analysis provides valuable insights for making decisions about product improvement. Based on a comprehensive analysis of statistical reliability, data representativeness, metric validity, and compliance with theoretical predictions, it can be concluded that the conducted experimental study is highly adequate for solving the tasks set. The main factors confirming the adequacy of the study are: the use of a comprehensive set of statistical methods to ensure the reliability of the results; the use of representative data sets that reflect the diversity of

real user feedback; the selection of appropriate evaluation metrics that take into account all aspects of the effectiveness of the models; the consistency of the experimental results with theoretical predictions and existing research; successful pilot implementation, which confirms the practical applicability of the developed models. The identified limitations of the study do not reduce its overall adequacy, but identify areas for further research and improvements. Thus, the conducted experimental study is adequate for assessing the effectiveness of various methods of automating user feedback analysis and developing practical recommendations for their implementation.

Conclusions. The conducted research on the application of natural language processing to automate the analysis of user feedback allowed us to obtain a comprehensive understanding of the effectiveness, advantages and limitations of various NLP methods and models in solving this problem. As a result of the research, all the goals and tasks were achieved, which allowed us 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 volumes; traditional manual analysis faces fundamental limitations associated with time costs, subjectivity of assessments and the inability to effectively process large data sets; these limitations are especially acute in the context of business globalization, when companies receive feedback in different languages and from different sources. Theoretical research on 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 own advantages and limitations, and the choice of a particular method depends on the specifics of the task, available computational resources, and requirements for accuracy and speed of analysis. Based on experimental research, it was found that transformer models, in particular domain-adapted BERT, provide the highest accuracy for analyzing the tone of responses (F1-score up to 0.93), significantly outperforming classical methods such as Naive Bayes (F1-score 0.76) and XGBoost (F1-score 0.85). At the same time, compact versions of transformers, such as DistilBERT, demonstrate only a slight decrease in accuracy (0.89 vs. 0.90) at twice the inference speed, which makes them attractive for practical application. In the field of aspect-oriented tone analysis, combined models that simultaneously solve the task of detecting aspects and classifying their tone have shown a significant advantage over sequential approaches. The RACL model demonstrated the highest performance with an average ABSA F1-score of 0.85, which is 9 percentage points higher than the baseline sequential BiLSTM–CRF + ATAE-LSTM model (0.76). It was found that the performance of all models significantly depends on the domain specificity and the quality of data preprocessing. Domain adaptation and additional training on specific data corpora significantly increase the accuracy of the analysis, especially for specialized industries with their own terminology. The study confirmed that multilingual models, such as XLM-RoBERTa, provide effective analysis of reviews in different languages without significant loss of accuracy compared to models trained for specific languages. This is of particular importance for global companies operating in different markets and receiving reviews in many languages. Statistical validation of the results using cross-validation, bootstrapping, and t-tests confirmed the reliability and statistical significance of the results obtained. The 95% confidence intervals for the F1-metric of the transformer models are within ±1.5%, which indicates high stability of the results. The developed pilot system of automated feedback analysis demonstrated high consistency with manual analysis by experts (Cohen's kappa = 0.82 for tone classification and 0.79 for aspect detection) at a significantly higher processing speed (120–150 times), which confirms the practical applicability of the proposed solutions. A survey of potential users of the system revealed that the most valuable functions of automated feedback analysis are the identification of critical problems (95% of respondents) and the identification of key aspects of products (92% of respondents), which emphasizes the practical significance of the research results. Despite the high efficiency, the developed methods have some limitations related to the difficulties of analyzing irony and sarcasm, the difficulties of processing informal vocabulary, and the dependence on the quality of training data. These aspects determine the directions for further research. The cost-effectiveness of implementing automated feedback analysis systems is confirmed by a significant reduction in the time and human resources required to process feedback, as well as an increase in the quality of the insights obtained, which allows companies to make more informed decisions on the development of products and services. The practical significance of the results obtained lies in the possibility of using them to create effective automated user feedback analysis systems that can be integrated into the business processes of companies of various scales and industries. Such systems allow not only to reduce the cost of analyzing feedback, but also to obtain deeper and more objective insights into the perception of products and services by users, to identify hidden trends and patterns, and to respond promptly to critical problems. Further research can be aimed at developing methods for analyzing multimodal feedback, which include not only text, but also images and videos, improving irony and sarcasm recognition algorithms, and creating more effective models for low-resource languages. Another promising direction is the development of methods for generating automatic recommendations for improving products based on the analysis of user feedback. Thus, the research makes a significant contribution to the development of methods for automating the analysis of user feedback based on natural language processing and creates a basis for further improvement of such systems.

5. References

[1] Acheampong, F. A., Wenyu C., Nunoo-Mensah H. Text-based emotion detection: Advances, challenges, and opportunities. Engineering Reports. 2020. Vol. 2, No. 7.

[2] Amazon Comprehend Documentation [Електронний ресурс] // Amazon Web Services. 2024. URL: https://docs.aws.amazon.com/comprehend/.

[3] Amazon Reviews Dataset [Електронний ресурс] // Registry of Open Data on AWS. 2023. URL: https://registry.opendata.aws/amazon-reviews/.

[4] Bird, S., Klein, E., Loper, E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, 2019. 504 p.

[5] Chen, Z., Qian, T. Transfer Capsule Network for Aspect Level Sentiment Classification. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019. P. 547–556.

[6] Conneau, A. Unsupervised Cross-lingual Representation Learning at Scale. Proc. ACL 2020, pp. 440–451.

[7] Devlin, J., Chang, M.W., Lee, K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT 2019. 2019. P. 171–186.

[8] Global AI Survey: AI proves its worth, but few scale impact [Електронний ресурс] // McKinsey & Company. 2023. URL: https://www.mckinsey.com/capabilities/quantumblack/our-insights/global-ai-survey-ai-proves-its-worth-but-few-scale-impact.

[9] Goldberg, Y. Neural Network Methods for Natural Language Processing. Synthesis Lectures on Human Language Technologies. 2017. Vol. 10, No. 1. P. 1–309.

[10] Google Cloud Natural Language API [Електронний ресурс] // Google Cloud. 2024. URL: https://cloud.google.com/natural-language.

[11] Hugging, Face Transformers Documentation [Електронний ресурс] // Hugging Face. 2024. URL: https://huggingface.co/docs/transformers/.

[12] Jurafsky, D., Martin, J.H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 3rd ed. draft. Pearson, 2023. [Електронний ресурс]. URL: https://web.stanford.edu/~jurafsky/slp3/.

[13] Kim, Y. Convolutional Neural Networks for Sentence Classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014. P. 746–751.

[14] Liu, Y., Ott M., Goyal N. et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach [Електронний ресурс] // arXiv preprint. 2019. URL: https://arxiv.org/abs/1907.11692.

[15] Mikolov, T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space. Proceedings of the International Conference on Learning Representations (ICLR 2013). 2013. P. 16–19.

[16] NLTK Documentation [Електронний ресурс] // Natural Language Toolkit. 2024. URL: https://www.nltk.org/.

[17] NLP Progress: Sentiment Analysis [Електронний ресурс] // NLP Progress. 2023. URL: http://nlpprogress.com/english/sentiment_analysis.html.

[18] Pontiki, M., Galanis D., Papageorgiou H. et al. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. Proceedings of the 10th International Workshop on Semantic Evaluation. 2016. P. 19–30.

[19] Poria S., Cambria E., Gelbukh A. Aspect Extraction for Opinion Mining with a Deep Convolutional Neural Network. Knowledge-Based Systems. 2016. Vol. 108. P. 42–49.

[20] PyTorch Documentation [Електронний ресурс] // PyTorch. 2024. URL: https://pytorch.org/docs/stable/.

[21] Sanh, V., Debut L., Chaumond J., Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter [Електронний ресурс] // arXiv preprint. 2019. URL: https://arxiv.org/abs/1910.01108.

[22] SemEval-2022 Tasks [Електронний ресурс] // International Workshop on Semantic Evaluation. 2022. URL: https://semeval.github.io/SemEval2022/tasks.html.

[23] Yelp Open Dataset [Електронний ресурс] // Yelp. 2023. URL: https://www.yelp.com/dataset.

Received: 23/08/2025
Accepted: 16/11/2025
Published: 15/12/2025

Скорін Юрій
кандидат технічних наук, доцент
Харківський національний економічний університет імені Семена Кузнеця,
Україна
ORCID: 0009-0004-5218-6369

Петренко Богдан
Студент вищого навчального закладу
Харківський національний економічний університет імені Семена Кузнеця,
Україна
ORCID: 0009-0000-1576-0043

https://doi.org/10.60022/sis.3.(02).5

ЗАСТОСУВАННЯ ПРИРОДНОЇ МОВНОЇ ОБРОБКИ ДЛЯ АВТОМАТИЗАЦІЇ АНАЛІЗУ ВІДГУКІВ КОРИСТУВАЧІВ

Анотація. *Метою представленої статті є дослідження застосування штучного інтелекту в автоматизованих системах управління фінансовими ризиками для підвищення точності, оперативності та ефективності прийняття управлінських рішень у фінансовій сфері. Дослідження полягає у комплексному вивченні теоретичних і методичних основ застосування штучного інтелекту в системах управління, аналізу сучасних підходів до класифікації та оцінки фінансових ризиків із використанням алгоритмів машинного навчання, формуванні архітектури системи підтримки прийняття рішень на основі моделей прогнозування, а також оцінці ефективності побудованих моделей за результатами симуляційного моделювання. У межах проведеного дослідження розроблено модель прогнозування кредитного ризику клієнтів банку, яка дозволяє здійснювати оцінку платоспроможності на основі історичних даних та сучасних методів машинного навчання. Методом дослідження є моделювання за допомогою інструментів машинного навчання, зокрема нейронних мереж і методів ансамблевого навчання (Random Forest), а також аналізування даних із використанням платформ для візуалізації результатів і оцінки ефективності моделей. Особливу увагу приділено підготовці даних, вибору релевантних ознак, оцінці точності моделей та побудові інтерпретованих візуалізацій, таких як SHAP-графіки, ROC-криві тощо. Результатом проведеного дослідження стало створення ефективної моделі прогнозування кредитного ризику, яка демонструє достатньо високий рівень точності класифікації та здатність до адаптації при зміні вхідних умов. Практичне значення дослідження полягає у можливості впровадження розробленої моделі до існуючих автоматизованих систем управління фінансовими ризиками банківських установ, що дозволить зменшити рівень кредитних втрат, підвищити фінансову стабільність і забезпечити більш точне управління ризиками. Такий підхід сприяє розвитку інтелектуальних фінансових систем, підвищенню рівня автоматизації управлінських рішень та зміцненню конкурентоспроможності фінансових установ у сучасних ринкових умовах.*

Ключові слова*: штучний інтелект, фінансові ризики, машинне навчання, кредитний ризик, нейронні мережі, автоматизована система управління, прогнозування, класифікація.*