

Oleshchenko L.M.

National Technical University of Ukraine
“Igor Sikorsky Kyiv Polytechnic Institute”

Antonenko I.V.

National Technical University of Ukraine
“Igor Sikorsky Kyiv Polytechnic Institute”

SOFTWARE BASED ON A HYBRID NEURAL NETWORK MODEL FOR ANALYZING USER REVIEWS ON E-COMMERCE PLATFORMS

This article presents a hybrid deep learning method and software for sentiment and consistency analysis of customer reviews in e-commerce systems, aimed at improving the accuracy and reliability of text classification and identifying discrepancies between the emotional content of a review and the numerical rating provided by the user. The relevance of this research stems from the continuous growth of online customer interactions and the increasing need for automated tools capable of processing large volumes of textual feedback for effective prioritization and enhanced customer engagement. Existing commercial solutions are often constrained by outdated architectures, limited language support, lack of integration capabilities, and cumbersome interfaces. The proposed approach addresses these limitations through a flexible, scalable, and interpretable hybrid model.

The core architecture integrates Word Embedding (WDE) for semantic representation, Convolutional Neural Networks (CNN) for extracting local textual features, and Long Short-Term Memory (LSTM) networks for capturing sequential dependencies. This combination enables the creation of a rich and multidimensional feature space, facilitating highly accurate sentiment classification across 2-, 3-, and 5-class schemes. The model was trained on the Amazon Product Reviews dataset comprising 3.5 million entries, ensuring diversity and robustness of the learned patterns. Data preprocessing involved balancing class distribution using RandomOverSampler, tokenizing and aligning sequences, and forming the label set based on predefined rating categories. Experimental results demonstrated that the hybrid WDE-CNN-LSTM model outperforms standalone CNN, LSTM, and WDE-LSTM models across precision, recall, accuracy, and F1-score metrics, delivering a consistent 2–5% improvement depending on the classification task. Notably, the hybrid architecture achieves strong validation performance from the very first training epoch, confirming its stability and generalization capability. Additionally, the study introduces a mechanism for evaluating consistency between predicted sentiment polarity and the user’s numerical rating, enabling the detection of ambiguous or contradictory reviews that may require managerial attention. The developed software is suitable for integration into large e-commerce platforms, CRM systems, automated support tools, and customer experience analytics pipelines. Future work will focus on optimizing the model for real-time processing, enhancing cross-domain adaptability, and expanding multilingual support to improve applicability across diverse markets. Overall, the proposed hybrid deep learning method demonstrates high potential for automating sentiment and consistency analysis, increasing the reliability of decision-making systems, and improving the operational efficiency of e-commerce platforms.

Key words: software, machine learning, big data, neural networks, CNN, LSTM, hybrid neural network model, user feedback, e-commerce platforms, Python programming language, TensorFlow.

Formulation of the problem. In modern e-commerce ecosystems, businesses face an exponentially growing volume of customer-generated textual content originating from product reviews, support inquiries, social media responses, and feedback submitted through web interfaces. Manual processing of such information is increasingly impractical, as it requires significant expert time, suffers from human subjectivity, and cannot scale proportionally to customer base

expansion. This situation leads to delayed reaction to critical customer complaints, misprioritized support requests, and a gradual decline in customer satisfaction and brand loyalty. Existing automated sentiment analysis tools offer only partial solutions, as many incorporate outdated deep-learning architectures, exhibit limited multilingual support, or provide insufficient mechanisms for verifying the consistency between textual sentiment and numerical customer ratings. Moreo-

ver, many available systems lack integration APIs, have non-intuitive user interfaces, or demonstrate reduced accuracy when processing heterogeneous review data with substantial class imbalance.

Therefore, the problem addressed in this research consists in developing an advanced hybrid deep learning method capable of performing reliable sentiment classification and evaluating the consistency between customer sentiment and their assigned ratings. The goal is to combine the strengths of multiple neural architectures, specifically WDE, CNN, and LSTM, to build a robust classification model capable of capturing semantic dependencies, local linguistic patterns, and temporal context within textual data. This requires not only designing an effective hybrid neural architecture but also creating a data preprocessing pipeline that addresses text ambiguity, class imbalance, and the need for generalized sentiment–rating alignment. Solving this problem is essential for enabling scalable, accurate, and interpretable sentiment analytics in real-world e-commerce platforms, where timely identification of customer dissatisfaction constitutes a key factor for service optimization and business competitiveness.

Analysis of recent research and publications.

Research in the field of sentiment analysis and text classification has advanced significantly over the past decade, with deep learning models becoming the dominant methodological foundation. Classical machine-learning techniques have been largely replaced by distributed text representations and neural architectures capable of automatically learning latent linguistic patterns from large datasets. Recent studies highlight the effectiveness of hybrid neural networks that integrate convolutional and recurrent layers, as such models simultaneously capture local syntactic structures and long-range dependencies within text sequences. For instance, CNNs have demonstrated high performance in extracting n-gram features and identifying key textual fragments, while LSTM networks remain the standard for modeling sequential patterns and temporal relationships in natural language input. Contemporary research confirms that combining CNN and LSTM architectures improves classification accuracy, particularly in multi-class and imbalanced datasets, which aligns with findings presented in works such as [1], [2].

In parallel, sentiment analysis research increasingly emphasizes the importance of data consistency and semantic validity when mapping customer reviews to numerical rating scales. Recent studies argue that polarity indicators extracted from textual content can differ markedly from self-reported user

ratings, which complicates the design of reliable sentiment classification models.

Large-scale review datasets, such as the Amazon Product Reviews corpus [3], exhibit strong class imbalance, necessitating sophisticated preprocessing pipelines and resampling strategies to prevent bias in model predictions. Evaluation methodologies also continue to evolve: multi-class classification accuracy is now commonly assessed through precision, recall, and F1-score metrics, as recommended in comparative clustering and classification studies [4]. Despite these advancements, current research still reveals limitations in architectures that rely on single-model approaches, as they often fail to generalize across heterogeneous review domains and lack mechanisms for validating sentiment–rating correspondence. The analysis of existing publications thus indicates a gap in the development of integrated deep-learning approaches that combine multiple neural mechanisms while explicitly addressing the alignment between sentiment polarity and user-assigned ratings. The hybrid method proposed in this study builds upon the trends identified in recent scientific literature, while addressing their limitations by offering a more holistic, resource-efficient, and semantically consistent approach to customer review analysis.

Task statement. Despite the substantial progress achieved in modern deep learning models for text sentiment classification, significant challenges remain in real-world e-commerce environments where customer reviews are highly heterogeneous, linguistically diverse, and often inconsistent with their associated rating scores. Classical single-architecture models such as standalone CNNs or LSTMs struggle to capture simultaneously the local lexical features and long-range semantic dependencies inherent to user-generated content, leading to reduced stability of sentiment predictions and limited reliability when prioritizing customer inquiries. Furthermore, many existing sentiment-analysis systems fail to incorporate mechanisms for verifying the internal consistency between the textual tone of a review and its numerical rating, which is critical for accurate feedback interpretation, customer-support automation, and fraud-detection tasks.

The task of this research is to develop and rigorously evaluate a hybrid deep learning model that integrates word-embedding representations, convolutional feature extraction, and recurrent temporal modeling to jointly assess review sentiment and its consistency with the given rating score. The research aims to determine whether a combined WDE-CNN-LSTM architecture can outperform classical models

in multi-class sentiment classification, ensure higher robustness under class imbalance, and provide more reliable prioritization of customer inquiries in large-scale e-commerce datasets. Additionally, the research seeks to validate the practical applicability of such a model using millions of real Amazon product reviews and to examine its potential for generalization across varying classification granularities (2-class, 3-class, and 5-class tasks), thereby establishing a foundation for further development of intelligent customer-feedback analysis systems.

Outline of the main material of the research.

The research presents the development, implementation, and evaluation of a hybrid deep learning method aimed at improving the accuracy and reliability of sentiment and consistency analysis in e-commerce customer reviews.

The core idea of the research lies in constructing a neural architecture that integrates Word Embedding Encoding (WDE), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) units

to exploit the complementary strengths of each component (Fig. 1). The WDE module captures semantic regularities in textual data, CNN extracts local linguistic patterns, and LSTM models long-range temporal dependencies, collectively forming a unified feature space well-suited for multi-class sentiment classification tasks.

The model architecture includes two parallel computational pathways – convolutional and recurrent – whose outputs are merged and processed by a dense classification layer employing the Softmax activation function, categorical cross-entropy loss, and Adam optimization, which has been shown to be optimal for multi-class text classification problems.

The algorithm shown in Fig. 2 outlines the workflow of a hybrid sentiment-analysis method, beginning with data preprocessing (cleaning, normalization, tokenization, feature extraction), which ensures reliable model performance. The dataset is then split into training, validation, and test subsets to enable learning, hyperparameter tuning, and unbiased eval-

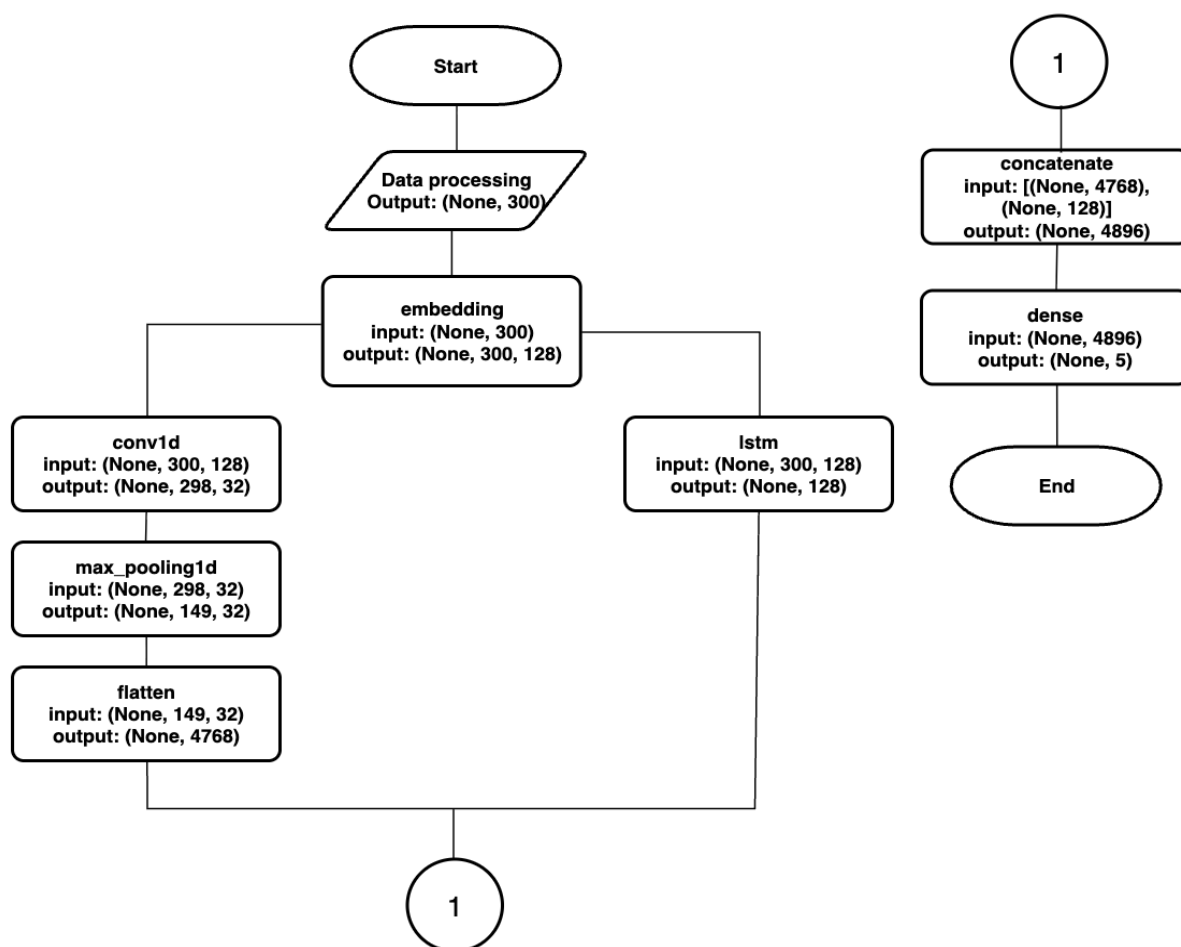


Fig. 1. Algorithm of actions for building hybrid neural network model

uation. After preparation, the model is trained and subsequently tested using standard metrics to assess its generalization capability. This structured pipeline guarantees high classification accuracy and allows objective comparison of different sentiment-analysis models.

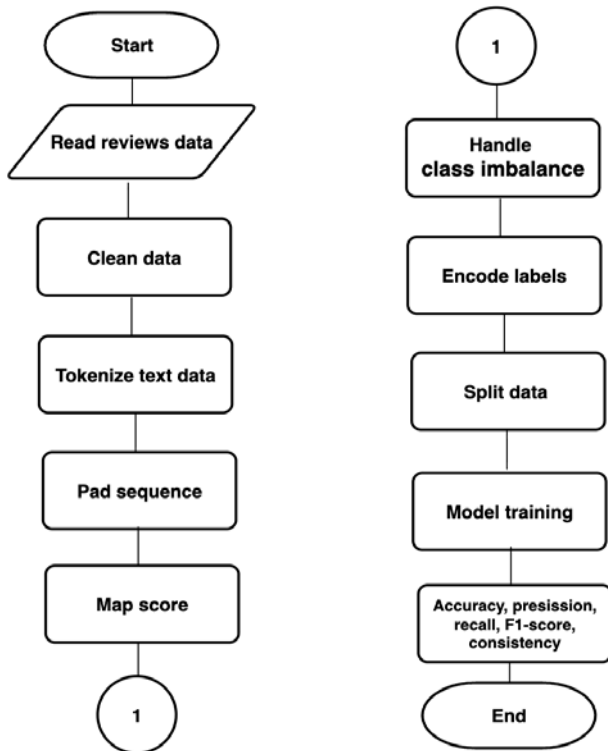


Fig. 2. Workflow of a hybrid sentiment-analysis method

A critical phase of the research involved the construction of a labeled dataset based on the “Amazon Product Reviews” corpus containing 3.5 million customer reviews, each associated with textual content and a rating from one to five stars.

Comprehensive preprocessing was carried out to prepare the data for neural training: irrelevant attributes were removed, review title and body were merged into a single text sequence, tokenization and

sequence padding were performed, and class imbalance was mitigated using RandomOverSampler.

The resulting dataset included 871,492 training instances and 217,874 validation instances, enabling robust performance evaluation across two-class, three-class, and five-class categorization settings. The research further introduces a consistency-checking mechanism that compares the sentiment polarity of a review – computed using TextBlob – against its numerical rating. This dual-labeling procedure allows the system to identify mismatches between textual sentiment and the assigned rating, enhancing the interpretive value of the processed customer feedback while enriching the dataset with additional semantic metadata.

The hybrid model was implemented in Python using the TensorFlow framework and trained on the preprocessed dataset (Table 1).

Consider the representation of the hybrid method in the form of formulas (1 – 8).

1. Input data is a set of user feedback data, we represent them in the form of the formula:

$$D = \{(x_i; y_i)\}, \quad (1)$$

where x_i is sequence of actions; y_i is future action (category).

2. Vector representation layer:

$$X = D * WDE, \quad (2)$$

where D is input sequence; WDE is vector representation matrix.

3. CNN model branch starts with the Conv1D layer, which applies 32 convolutional filters with kernel size 3 to the embedded sequences:

$$F = \text{Conv1D}(W, X), \quad (3)$$

where W are parameters of the convolutional layer, i.e. the weight matrix for 32 filters with kernel size 3, which are learned during training.

4. The convolution operation is accompanied by a ReLU activation function to introduce nonlinearity,

Table 1

Parameters of the hybrid neural network model

Layer (type)	Output Shape	Param #	Connected to
Input_1	(None, 300)	0	-
Embedding_1	(None, 300, 128)	1,280,000	input_layer_1[0][0]
Conv1D_1	(None, 298, 32)	12,320	embedding_1[0][0]
MaxPooling1d_1	(None, 149, 32)	0	conv1d_1[0][0]
Layer (type)	Output Shape	Param #	Connected to
Flatten_1	(None, 4768)	0	max_pooling1d_1[0][0]
LSTM_1	(None, 128)	131,584	embedding_1[0][0]
Concatenate_1	(None, 4896)	0	flatten_1[0][0], lstm_1[0][0]
Dense_1	(None, 3)	14,691	concatenate_1[0][0]

which allows the model to capture more complex patterns in the data:

$$R = \text{ReLU}(X * W_c + b_c), \quad (4)$$

where W_c is the convolution filter; b_c is the offset applied during the convolution operation.

5. After the convolution layer, a MaxPooling1D layer with a window size of 2 is used to reduce the dimensionality of the feature maps. The pooling operation selects the maximum value from the regions of the feature map, thereby preserving the most expressive features while simultaneously reducing the computational complexity:

$$M = \text{MaxPooling}(R, 2). \quad (5)$$

6. The LSTM branch starts with an LSTM layer, which consists of 128 units. This layer processes embedded sequences and captures temporal dependencies in the text. The output of the LSTM is the final hidden state h_T at the last time step T , while the state of the cell c_T is also updated during the calculations:

$$h_T, c_T = \text{LSTM}(X, h_{T-1}, c_{T-1}), \quad (6)$$

where T is the final time step; h_{T-1} is the output of the LSTM at a specific time step $T-1$; c_{T-1} is the internal “long-term memory” of the LSTM, which stores key information throughout the sequence at step $T-1$.

7. Next, the Concatenate layer is used, which combines the outputs of both branches – CNN and LSTM. This layer combines the local features extracted by the CNN with the temporal dependencies captured by the LSTM, forming a complex representation of the features. The resulting concatenated vector contains information from both branches:

$$C = \text{Concatenate}(M, h_t), \quad (7)$$

where M is the output of the CNN branch; h_t is the final hidden state of the LSTM branch.

8. Finally, the Dense Output Layer takes the feature connections and partitions them into target classes. The number of units in this layer depends on the classification problem and can be 5, 3, or 2, and uses the Softmax activation function to obtain a probability distribution over the possible labels. The output is a vector representing the predicted probabilities for each class:

$$y = \text{Softmax}(C * W_d + b_d), \quad (8)$$

where W_d is the weight matrix; b_d is the dense layer offset (Dence).

Experimental evaluation shows that the model converges rapidly, achieving high validation accuracy already during the first epoch and demonstrating stable performance without overfitting after five epochs (Fig. 3).

A comparative analysis of four architectures – CNN, LSTM, WDE-LSTM, and WDE-CNN-LSTM – was conducted across different rating granularities using Precision, Recall, F1-Score, and Accuracy metrics.

The results indicate that the hybrid WDE-CNN-LSTM model consistently produces superior performance across all metrics. In the two-class setting, although CNN performs weaker, the hybrid model maintains high accuracy. For three-class classification, performance improves across all architectures, with the hybrid model achieving the strongest results. In the five-class scenario, where the task becomes more complex, all models exhibit some degradation; the hybrid model again outperforms the benchmarks, achieving accuracy around 91%, with CNN showing the steepest decline. These findings confirm that integrating multiple deep learning mechanisms significantly improves classification robustness, especially for tasks requiring nuanced sentiment discrimination and stable performance under multi-class conditions. The research results demonstrate that the proposed hybrid architecture effectively addresses limitations of single-architecture models, offers strong generalization for large and diverse real-world datasets, and provides businesses with an enhanced toolset for automated sentiment analysis and review consistency evaluation (Table 2).

Software includes a consistency-checking module that evaluates whether the predicted sentiment aligns with the numerical rating provided by users. Parallel request processing and a retraining mechanism ensure scalability and adaptability to evolving linguistic patterns. The implementation leverages GPU acceleration, message queues, and a modular cloud architecture based on AWS services (Lambda, ECS, S3, SQS/SNS, SageMaker), enabling real-time processing and large-scale data handling. The practical system consists of a client interface, a multi-module backend (authentication, data intake, preprocessing, sentiment analysis, statistics, monitoring), and a cloud-oriented deployment architecture (Fig. 4).

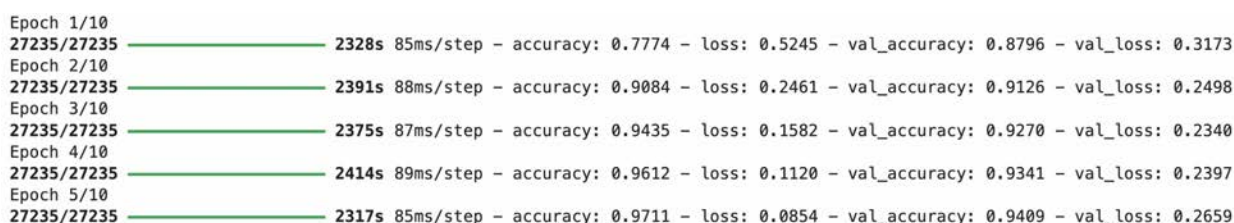


Fig. 3. Result of the hybrid neural network training process

Table 2

Comparison of neural network accuracy indicators on the test dataset

Rating	Model	Precision	Recall	F1 -Score	Accuracy
2	CNN	94.22	94.19	94.19	87.19
	LSTM	94.13	94.11	94.12	94.11
	WDE-LSTM	94.21	94.18	94.18	94.18
	WDE-CNN-LSTM	94.52	94.49	94.49	94.00
3	CNN	93.95	93.35	93.93	93.94
	LSTM	94.08	94.04	94.04	94.04
	WDE-LSTM	94.04	94.04	94.04	94.04
	WDE-CNN-LSTM	94.26	94.26	94.26	94.26
5	CNN	89.33	89.33	89.32	89.35
	LSTM	89.51	89.53	89.50	89.53
	WDE-LSTM	89.50	89.48	89.48	89.55
	WDE-CNN-LSTM	91.21	91.21	91.20	91.21

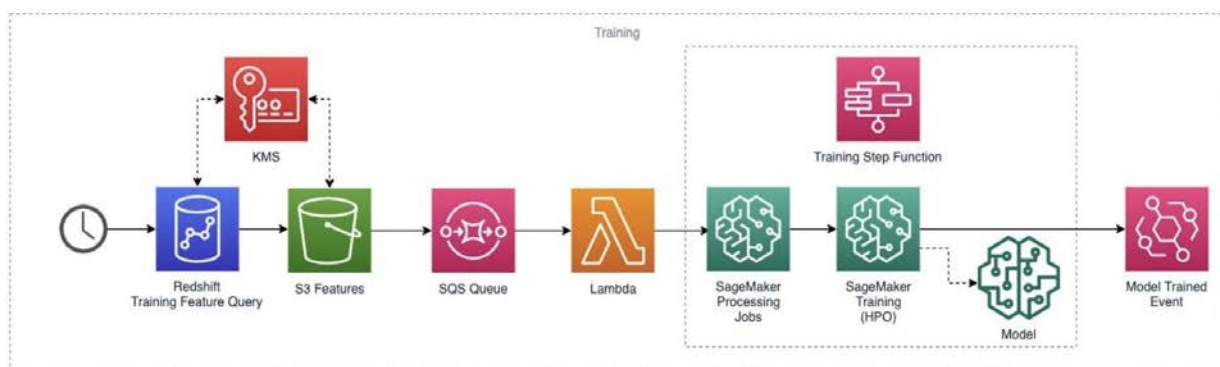


Fig. 4. Diagram of system components in the “AWS” annotation. Model training

The software architecture consists of several main components. The client part is built to interact with the server. It provides a graphical interface for interacting with system functions, such as viewing analytics, processing feedback, and feedback information. The server part is a REST API. API Gateway performs the function the only one incoming points for all external requests to server parts system (Fig. 5).

The product list view function includes viewing product information, its average sentiment score and average user score, the total number of product reviews, a list with detailed information about each. The page has a function to add images to the review. A webhook that is triggered when a new review arrives on the platform allows you to integrate the API with own solutions and receive notifications (Fig. 6). The architecture of the created software provides for regular retraining of the model several times a month. The speed of training is not a key factor and does not affect the choice of the optimal classification method. frequent launch of training processes, this feature may become critical.

Review dashboard (Fig. 7) was added to quickly view the most important reviews, as they are divided

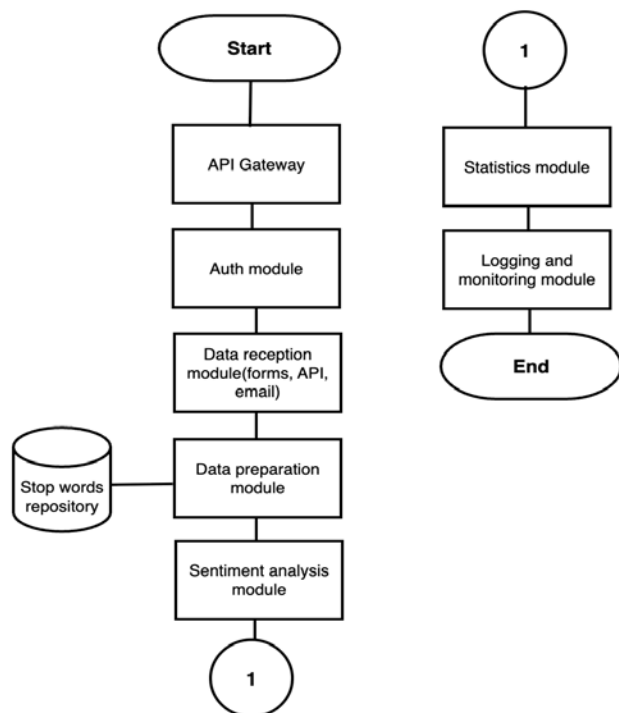


Fig. 5. Structure of software modules

into important, medium-urgent and non-urgent based on sentiment analysis. Each column contains information about the review text, images added by the customer, date, sentiment, review metadata such as email and communication channel.

The experimental results showed that the average training time of the improved model is the longest – 87ms, but not much longer than the LSTM training – 74ms, while the CNN – the fastest 34ms. It is also worth noting that the developed model was trained in fewer epochs and showed better results.

Even the duration of 87 seconds meets the performance requirements for use in the “AWS Lambda” environment, which allows us to consider it acceptable at the current stage of development. In this work,

priority is given to the accuracy of the model, so the training time is not a decisive factor.

Conclusions. The research presents a hybrid deep learning architecture that successfully integrates WDE, CNN, and LSTM components to improve the accuracy and robustness of sentiment classification and consistency analysis in e-commerce customer reviews. Experimental results demonstrate that the proposed model surpasses traditional CNN and LSTM architectures across multiple evaluation metrics and classification granularities, achieving high effectiveness in two-class, three-class, and five-class settings.

The model shows particularly strong performance when applied to large-scale real-world datasets such as Amazon Product Reviews, confirming its suitability

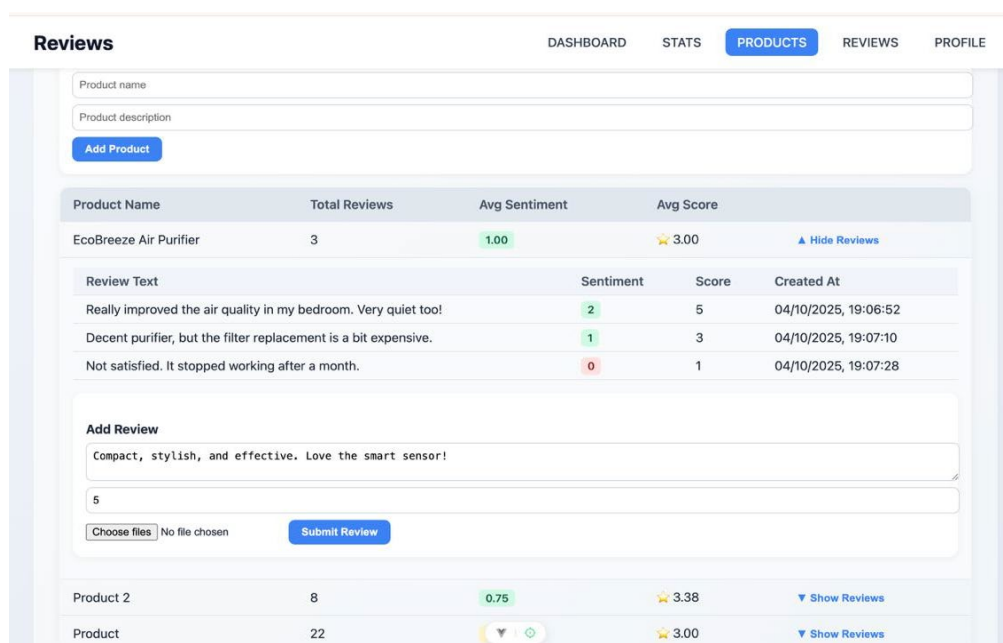


Fig. 6. Feedback creation and viewing screen of developed software

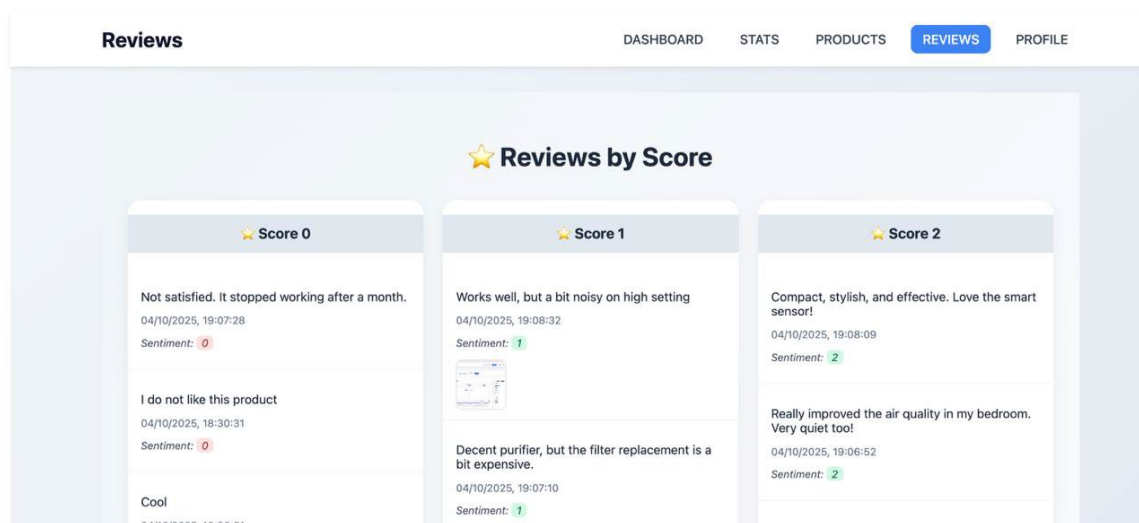


Fig. 7. Quick review screen of developed software

bility for industrial e-commerce applications. The research also contributes a sentiment–rating consistency methodology, enabling deeper interpretability of user feedback by identifying discrepancies between textual sentiment and assigned ratings.

The findings validate the feasibility of using hybrid neural architectures for processing large volumes of customer feedback with improved accuracy and reliability. While the proposed model demonstrates strong performance, further research is needed to assess its

applicability to other types of textual data and to optimize the system for real-time deployment.

Future work may include expanding the model to support additional languages, enhancing computational efficiency, and integrating advanced embedding techniques or transformer-based architectures. Nonetheless, the results of this study highlight the potential of hybrid deep learning approaches to significantly enhance automated customer feedback analysis and support data-driven decision-making in modern e-commerce systems.

Bibliography:

1. Hasib K. M., Azam S., Karim A., Marouf A. A., Shamrat F. M. J. M., Montaha S., Yeo K. C., Jonkman M., Alhaji R., Rokne J. G.. MCNN-LSTM: Combining CNN and LSTM to Classify Multi-Class Text in Imbalanced News Data. *IEEE Access*, 11, 2023. 93048-93063. Article 93048. <https://doi.org/10.1109/ACCESS.2023.3309697> (access date: 15.11.2025).
2. Margarita Rodríguez-Ibáñez, Antonio Casáñez-Ventura, Félix Castejón-Mateos, Pedro-Manuel Cuenca-Jiménez. A review on sentiment analysis from social media platforms. *Expert Systems with Applications*, Vol. 223, 2023, 119862. <https://doi.org/10.1016/j.eswa.2023.119862> (access date: 15.11.2025).
3. Amazon Product Reviews (Dataset), 2021. URL: <https://www.kaggle.com/datasets/arhamrumi/amazon-product-reviews> (access date: 15.11.2025).
4. Amigó E., Gonzalo J., Artiles J. et al. A comparison of extrinsic clustering evaluation metrics based on formal constraints. *Information Retrieval*, 12, 461–486 (2009). <https://doi.org/10.1007/s10791-008-9066-8> (access date: 15.11.2025).

Олещенко Л.М., Антоненко І.В. ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ НА ОСНОВІ ГІБРИДНОЇ МОДЕЛІ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ АНАЛІЗУ ВІДГУКІВ КОРИСТУВАЧІВ ПЛАТФОРМ ЕЛЕКТРОННОЇ КОМЕРЦІЇ

У статті представлено гібридний метод глибинного навчання та програмне забезпечення для аналізу тональності та узгодженості клієнтських відгуків у сфері електронної комерції, спрямований на підвищення точності класифікації текстових даних та виявлення невідповідностей між емоційним змістом відгуку та оцінкою користувача. Актуальність дослідження зумовлена різким зростанням обсягів клієнтських звернень у цифрових сервісах та потребою бізнесу в автоматизованих інструментах обробки текстових повідомлень для оперативного реагування та підвищення лояльності користувачів. Запропонована модель усуває обмеження традиційних рішень, які часто мають складний інтерфейс, підтримують лише обмежену кількість мов та інтегрують застарілі архітектури без можливості адаптації до власних застосунків. Гібридна архітектура поєднує можливості Word Embedding (WDE) для семантичного узагальнення тексту, згорткових нейронних мереж CNN для виділення локальних ознак та рекурентних мереж LSTM для оброблення часових залежностей. Така комбінація дає змогу сформувати комплексний простір ознак та здійснювати класифікацію за 2, 3 або 5 класами з високою точністю. Модель навчається на великому наборі Amazon Product Reviews, який містить 3,5 млн. відгуків, що забезпечує достатню різноманітність прикладів для коректного узагальнення та перевірки. Використання RandomOverSampler дозволило усунути проблему дисбалансу класів, а комбінована мережева архітектура забезпечила стабільність навчання та високу валідаційну точність уже на перших епохах.

Отримані результати показали, що гібридна модель перевершує класичні CNN, LSTM та WDE-LSTM за точністю, повнотою та F1-метрикою, зокрема демонструючи від 2 до 5% приросту продуктивності залежно від кількості класів. Найкращих показників модель досягає для 3-класової та 5-класової класифікації, де вона демонструє найвищу узгодженість між прогнозом та реальними оцінками. Окремо реалізовано механізм оцінювання узгодженості між емоційною тональністю тексту та числовою оцінкою користувача, що дозволяє автоматично виявляти суперечливі та потенційно проблемні відгуки. Розроблене програмне рішення забезпечує масштабованість, можливість інтеграції через API та потенційну підтримку мультимовної обробки даних. Запропонований підхід може застосовуватися у великих торговельних платформах, CRM-системах, чат-ботах та сервісах аналітики клієнтського досвіду. У майбутніх дослідженнях планується оптимізація продуктивності моделі для роботи в реальному часі, дослідження можливості перенавчання на інших наборах даних та розширення підтримки багатомовного аналізу текстів. Представлений метод демонструє значний потенціал для автоматизації робочих процесів у сфері електронної комерції та підвищення точності інтелектуальних систем обробки текстових звернень.

Ключові слова: програмне забезпечення, машинне навчання, великі дані, нейронні мережі, CNN, LSTM, гібридна модель нейронної мережі, відгуки користувачів, платформи електронної комерції, мова програмування Python, TensorFlow.

Дата надходження статті: 26.11.2025

Дата прийняття статті: 12.12.2025

Опубліковано: 30.12.2025