BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **24**, No 3
Sofia • 2024 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2024-0028

Sentiment Analysis of Product Reviews Using Transformer Enhanced 1D-CNN and BiLSTM

Muhammad Rizwan Rashid Rana[1], *Asif Nawaz*[1], *Tariq Ali*[1], *Ahmed Saleh Alattas*[2], *Diaa Salama AbdElminaam*[3,4]

[1]*University Institute of Information Technology, PMAS Arid Agriculture University, Rawalpindi, Pakistan*
[2]*Information Science Department, Faculty of Arts & Humanities, King Abdul-Aziz University, Jeddah, Saudi Arabia*
[3]*MEU Research Unit, Middle East University, Amman, Jordan*
[4]*Jadara Research Center, Jadara University, Irbid, Jordan*
E-mails: rizwanrana315@gmail.com asif.nawaz@uaar.edu.pk tariq.ali@uaar.edu.pk aalatas@kau.edu.sa diaa.salama@miuegypt.edu.eg

Abstract: *The rapid growth of Internet-enabled applications, such as social media platforms, e-commerce sites, and blogs, has led to a surge in user-generated content. This vast amount of data has made sentiment analysis increasingly valuable. Modern Aspect-Based Sentiment Analysis* (*ABSA*) *offers a more detailed approach by identifying sentiment trends related to specific aspects within the text. However, the challenge lies in analyzing reviews that are often short, unstructured, and filled with slang and emotive language, making it difficult to gauge customer opinions accurately. To address these issues, we proposed an effective hybrid approach "RoBERTa-1D-CNN-BiLSTM" for ABSA. Initially, the pre*-trained Robustly *Optimized BERT approach* (*RoBERTa*) *and One Dimensional Convolutional Neural Network* (*1D-CNN*) *models are used to extract features at the aspect level from the context of the review, following which classification is performed using Bidirectional Long Short-Term Memory* (*BiLSTM*). *The approach is evaluated on three cross-domain standards datasets, yielding an accuracy of 92.33%. The results of the experiments show that it surpasses the current leading methods in sentiment analysis and product recommendation.*

Keywords: *Sentiment analysis, Deep learning, e-Commerce, Products reviews, Opinion mining.*

1. Introduction

The internet has revolutionized how businesses connect with a global audience, erasing the traditional confines of commerce. In this new landscape, the relevance of brick-and-mortar stores is diminishing, as they are no longer the exclusive channels for product purchases [1]. This shift has been primarily driven by the rise of online marketplaces and e-Commerce platforms, dramatically altering the retail sector. These platforms offer unparalleled convenience, allowing customers to explore and acquire various products from across the globe effortlessly. Reflecting this change, the global e-Commerce market, valued at an impressive $5.4 trillion in 2022, is expected to surge to $7.4 trillion by 2024 [2]. Accompanying this growth is a deluge of online reviews and comments, which have become an integral part of the modern shopping experience.

Customer opinions in the form of reviews are the only true difference when products and services are increasingly commoditized [3]. These reviews are not just opinions; they are powerful influencers shaping others’ purchasing decisions. They offer deep insights into a product’s quality, usability, and overall value. A telling statistic underscores this trend: about 95% of shoppers consult online reviews before buying, and 93% acknowledge that these reviews sway their purchasing choices [4]. This trend has shifted the power balance from businesses to consumers, pushing companies to focus more on customer satisfaction, product quality, and continuous improvement [5]. The daily generation of extensive content on social media platforms necessitates the implementation of automated systems to manage and discern opinions effectively. As a result, many algorithms and methodologies about sentiment analysis have been introduced across a diverse array of literature [6-10].

Sentiment Analysis (SA) is a technique that involves evaluating the opinions and emotions articulated by users across various platforms such as blogs, social media, discussion forums, and websites to glean valuable insights [4]. This method can be a valuable tool for managers, politicians, and decision-makers, aiding them in assessing whether their product, service, or policy meets the community’s approval. The objective of SA is to discern and analyze customer opinions, facilitating well-informed decision-making for buyers and vendors [5]. Opinions conveyed on the internet regarding products are of great importance to both producers and customers. Typically, these reviews are heavily laden with opinions. Customers often inquire about the most well-liked products, superior features, and the reasoning behind positive or negative evaluations. However, the vast amount of data available makes manual analysis unfeasible due to time constraints. As a result, online data processing is mandatory.

Among the various types of sentiment analysis, available, Aspect-Based Sentiment Analysis (ABSA) emerges as one of the most advanced methodologies in contemporary research [6]. ABSA is dedicated to identifying and extracting sentiments about specific attributes or features of a product or service, such as the quality of the customer service, the taste of a food item, or the comfort of a hotel room. This type of analysis requires identifying the relevant aspects in the text and determining the sentiment expressed about each aspect [7]. ABSA can provide more accurate product information by focusing on specific aspects or features. For example, suppose a restaurant owner knows that customers consistently express negative sentiments about the quality of the food. In that case, they can improve the quality of their ingredients or recipes.

The existing methods for ABSA primarily rely on either lexicons or machine learning models. Lexicon and dictionary-based models are domain-specific and may pose challenges in constructing them. In contrast, machine learning models may necessitate structured inputs to train and test the model [4]. Additionally, the reviews available for analysis are typically short, unstructured, and laden with slang words and emotions, making it challenging to extract customers' opinions [5][6]. To address the aforementioned issues, this research aimed to enhance the accuracy of ABSA using the unique hybrid technique "RoBERaT-1D-CNN-BiLSTM".

RoBERTa is a large-scale, pre-trained language model developed by Facebook AI Research (FAIR) [11]. Through extensive exposure to diverse textual material, RoBERTa can learn the statistical patterns and relationships between words and sentences, enabling it to generate high-quality features like aspects, opinions, etc. Neural network models have been developed to improve sentiment analysis in NLP. Because of its memory retention, Recurrent Neural Network (RNN) is effective for extracting long-distance dependent information from review corpora. RNN has limits, while CNNs have troubles when capturing long-term dependencies [12]. In light of these challenges, the BiLSTM was introduced to address these issues. The proposed hybrid model leverages the strengths of RoBERTa, 1D-CNN, and BiLSTM models that can efficiently extract aspects, accurately classify the reviews, and reduce computational overhead, thereby improving the overall accuracy of ABSA. Furthermore, this model has practical applications in analyzing short texts, including tweets, posts, etc., making it a valuable tool for product recommendations.

This research presents a novel approach for ABSA specifically focused on product recommendation. The proposed approach addresses the limitations of baseline methods in this field. The proposed method consists of several stages: preprocessing text data, feature extraction and generation, and product of recommendation. Notably, our proposed model offers the following **research contributions**:

- Investigate and identify the underlying factors that influence the effectiveness of models used for sentiment analysis.
- A hybrid model combining RoBERTa, 1D-CNN, and BiLSTM is designed to leverage the benefits of the Transformer and CNN models.
- The proposed model has been validated using three benchmark datasets. The experiment's findings reveal that the proposed model outperforms recent baseline models in performance, thus validating its effectiveness and validity.

The organization of this paper is as follows: Section 2 reviews previous methods in ABSA. Section 3 outlines the problem formulation. A comprehensive overview of the RoBERTa-BiLSTM model structure is provided in Section 4. Section 5 demonstrates the efficacy of our approach through experimental validation, comparing our model's performance against other methods. Finally, Section 6 concludes with a discussion of conclusions and potential directions for future research

2. Literature review

ABSA is a burgeoning area of research that aims to identify and analyze the sentiment expressed in the text [6-10]. This literature review examines the present methodologies utilized in ABSA and the difficulties and opportunities researchers face in this area.

The authors in [13] have introduced a recommendation system that incorporates Sentiment Analysis (SA) and a Hybrid Recommendation Model (HRM) to tackle cold-start and data sparsity issues and enhance accuracy. HRM generates an initial list of recommendations, which SA then refines. The proposed system outperforms traditional models across a variety of evaluation metrics. Meanwhile, another study in [14] proposed a Knowledge-aware Dependency Graph Network (KDGN) that combines domain knowledge, dependency labels, and syntax path. The findings show that KDGN outperforms previous methods, highlighting the significance of domain knowledge of ASBA. Another research presented a model in [15] for ABSA that differentiates the polarity of each aspect's sentiment. The model uses contextual location data, position weight functions, and bidirectional GRU layers.

Garcia and Berton [16] introduced a hybrid methodology for sentiment analysis, integrating lexicon-based and learning-based approaches, and applied this combined technique to Twitter data. Similarly, Gupta and Joshi [17] proposed a distinctive hybrid deep-learning architecture aimed at categorizing sentiments. Deep convolutional networks have exhibited significant success in extracting local features through recurrent structures, which generally produce outstanding results in the sequential analysis of extensive texts. Moreover, Kim and Jeong [18] highlighted the growing interest in sentiment analysis among major corporations, a trend fueled by the escalating utilization of social networking sites and e-commerce platforms. This study evaluates the sentiment analysis efficacy of multiple classifiers. Results showed that Random Forest yielded highly accurate results compared to other algorithms.

Nandal, Tanwar and Pruthi [19] conducted a study on ABSA challenge related to bipolar words. The research focused on how the context can affect the polarity of words and, in turn, affect the overall rating of products and specific aspects. Their findings were impressive. Parthi, Raparthi and Gopalachari [20] proposed a solution to the cold start issue in ABSA by developing an automatic method to calculate sentiments for dynamic aspects in customer-generated reviews obtained from various sources using web scraping. The authors also improved the system’s accuracy by introducing new stop words.

Shams, Khoshavi and Baraani-Dastjerdi [21] proposed an approach for ABSA that centers on unsupervised learning. This approach addresses time, and cost complexity issues, and consists of three stages, each with detailed operations. In the first stage, a preliminary lexicon for polarity and a group of aspect words are chosen to extract pertinent details from the dataset. Subsequently, these resources are inputted into an expectation-maximization algorithm that computes the likelihood of a term according to its emotion and aspect. In the final step, the document is divided into components to analyze their polarity. Londhe and Rao [22] proposed a pioneering strategy for ABSA by leveraging a deep learning classifier called LSTM-RNN. The amalgamation of LSTM-RNN showcased exceptional precision in scrutinizing and forecasting the sentiment of various aspects. Additionally, the proposed technique was proficient in extracting multiple aspects from lengthy reviews comprising many sentences.

Mutinda, Mwangi and Okeyo [23] presented a classification model combining sentiment lexicons, N-grams, BERT, and CNN. It uses these components to vectorize selected words from the input text and employs CNN for sentiment prediction. The model was evaluated on three public review datasets and outperformed existing models with an impressive 88.73% F-measure score in binary sentiment classification. Huang et al. [24] devised a technique that automatically hones in on critical emotional cues within sentences. This approach further explores the interaction between aspect words and context, emphasizing the nuanced interplay of sentiments within a text. Li et al. [25] focused on the attention mechanism and features that posed challenges to CNN. They introduced a novel module, TNet, characterized by word vectors generated through the synergy of CNN and RNN. This innovative approach aimed to optimize the attention mechanism and mitigate the limitations of CNN. Further, Li et al. [26] amalgamated LSTM and CNN in a parallel configuration and introduced an innovative sentiment-filling technique. This novel approach aimed to cultivate a methodology better tailored for sentiment analysis, enhancing the capability of the model to discern and interpret sentiments more effectively.

Furthermore, Table 1 presents a comprehensive summary of additional approaches addressing the sentiment analysis.

Table 1. Comprehensive summary

Reference	Methodology	Dataset	Result accuracy	Limitation	Evaluation strategy
[27]	BiLSTM with an attention mechanism	Amazon	86%	Required complex training	Single run
[28]	Dynamic memory networks	Semeval 2014 and Semeval 2016	84%	Loss of sequential and contextual information	Cross-validation with ten folds
[29]	CNN-RNN	Sentiment140	81%	Dictionary-based and domain-specific	Single run
[30]	CNN with LSTM	Tweets	89%	Cannot balance word sentiment	Average accuracy from three runs
[31]	Deep intelligent contextual embedding	Airline dataset	95%	Required a more dependable recurrent model	Single run
[32]	VGGNet-16 and LSTM	Weibos dataset	88%	Cannot handle other domain words	Cross-validation with five folds
[33]	LSTM	Yelp 2014 and Yelp 2015	63.9%	Unable to normalize polarity intensity of words	Single run
[34]	AHRM	Flicker8k dataset, Getty dataset	87%	Features are lost during extraction	Cross-validation with five folds
[35]	Lightweight multilayer interactive attention network	Laptop and restaurant reviews	85.88%	High memory consumption.	Single run
[36]	LSTM-RNN	The Stanford sentiment treebank (SST)	86.62%	Features are redundant	Average accuracy from five runs

In brief, ABSA has emerged as a highly debated and researched topic in the domain of sentiment analysis, garnering the attention of numerous scholars [6-13]. Significant progress has been made in the practical implementation of ABSA, particularly in the realm of product reviews. However, the accuracy of ABSA tends to diminish when dealing with short, unstructured, and multiple aspects present in the reviews. Given this backdrop, the present study proposes the utilization of RoBERTa-1D-CNN-BiLSTM for aspect-level semantic sentiment analysis of product reviews.

3. Problem formulation

The objective is to classify each aspect of a product mentioned in a review based on sentiment polarity (recommend, not recommend). A Review (R) is defined as a sequence of n words, denoted as R = $\{R_1 + R_2 + R_3 + \cdots + R_n\}$. Aspects ($A$) are specific features or attributes of the product mentioned within the review. There are m aspects in the review, denoted as A= $\{A_1 + A_2 + A_3 + \cdots + A_m\}$ where each aspect A_i, i = 1, …, m, is a sub-sequence of the review focused on a particular feature or attribute. The goal is to automatically identify these aspects and classify the sentiment associated with each one into either "recommend" or "not recommend". The aspects are extracted as relevant sub-sequences of words that represent specific features or attributes. For example, the review "Its quality is good, but the price is high". contains two aspects: [A_1: quality] and [A_2: price], which are supposed to produce [quality]: recommend and [price]: not recommend outputs. Another example, "The screen size is excellent and the cost is reasonable, but the quality is poor", consists of three aspects: [A_1: screen], [A_2: cost], and [A_3: quality], which is supposed to produce [screen]: recommend, [cost]: recommend and [quality]: not recommend output. The model aims to perform this identification and classification process automatically. The aggregation of aspect sentiments to determine the overall review sentiment could be done using methods such as majority voting or weighted voting.

4. Material and methods

This section offers an in-depth overview of the model proposed in this study. The proposed methodology aims to refine the aspect term extraction process and pinpoint the sentiments in online reviews concerning various aspects. The approach encompasses multiple phases, including preprocessing of data, i.e., removal of HTML tags and Punctuations, Tokenization, etc. Features are then extracted using RoBERTa and 1D-CNN. Subsequently, we predict the product recommendation from the reviews using a classification method employing the Bidirectional Long Short-Term Memory (BiLSTM) model. Fig. 1 illustrates the process of proposed model.

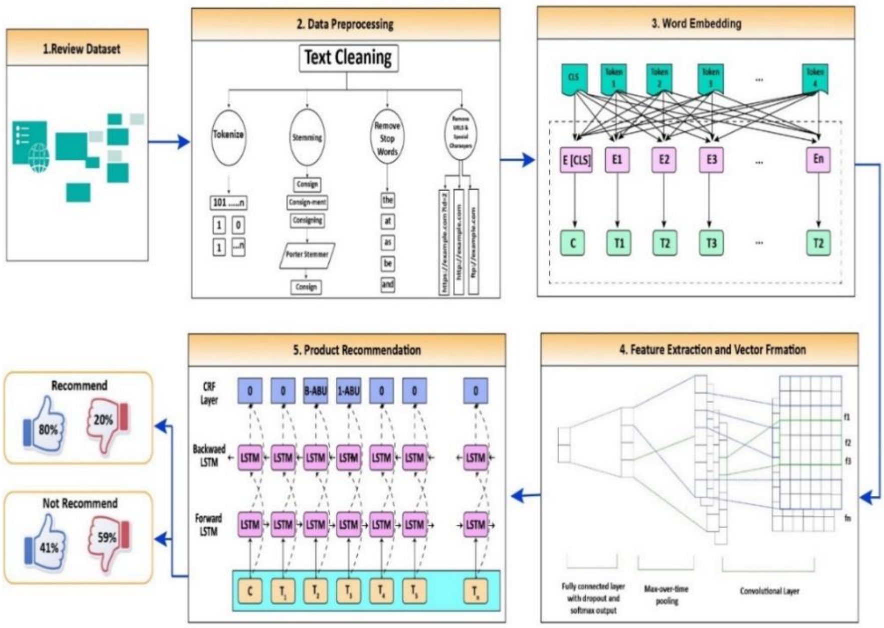


Fig. 1. Proposed methodology for sentiment classification and product recommendation

4.1. Dataset description

Data is the most important prerequisite for effectively conducting sentiment analysis. Insufficient data can significantly hamper the accuracy of the analysis. However, in recent years, sentiment analysis research has gained significant momentum due to the profusion of reviews on e-commerce sites, resulting in the availability of numerous review datasets. Three multidomain review benchmark datasets are used for the experiments, as shown in Table 2. The reason for using benchmark datasets for sentiment analysis experiments is that they provide a standard way of evaluating the performance of different sentiment analysis models. A benchmark dataset is a well-known dataset that the research community has used for developing and testing sentiment analysis models.

Table 2. Dataset breakdown

No	Dataset	Domain	Reviews about
1	Hu and Liu Dataset [37]	Products	MP3 Players, Canon, Apex AD 2600, Nikon and Nokia 6610
2	SemEval Dataset [38]	Product and hotels	Laptop and Restaurants
3	Bo Pang and Lillian Lee [39]	Movies	Kindle and Fire TV Stick

4.2. Data preprocessing

The primary aim of the preprocessing phase is to cleanse and standardize the data, thus enhancing the accuracy of the ABSA model [4]. Preprocessing techniques include tokenization, lemmatization, and stemming and eliminating stop words, URLs, and HTML tags. The process of segmenting the text into separate words or phrases is known as tokenization. Lemmatization and stemming are two techniques employed to condense the vocabulary by transforming inflected word forms back to their root forms, thereby reducing vocabulary size. Eliminating stop words, URLs, and HTML tags is crucial for removing phrases and elements that do not add value to the text's meaning. The process of data preprocessing is described in detail in Algorithm 1.

Algorithm 1. Data preprocessing for sentiment analysis

Step 1. *Input:* DS_i : A dataset of reviews
Step 2. *Output:* W^{all} : All preprocessed words
Step 3. Initialize $W^{all} \longleftarrow \Phi$
Step 4. For each DS ϵ DS_i do
Step 5. St ⟵ tokenize (DS_i) // Tokenize the reviews
Step 6. St′ ⟵ lowercase (St) // Convert all words to lowercase
Step 8. For each S_i ϵ St′ do
Step 9. St′′ ⟵ normalize (S_i)
Step 10. St′′′ ⟵ stem (St′′)
Step 11. St′′′′′ ⟵ transliteration (St′′′′)
$W^{all} \longleftarrow W^{al}$ U { St′′′′}
Step 12. **End For**
End For
Step 13. Return W^{all}

4.3. Word embedding

The integration of word embeddings in sentiment analysis is imperative, especially preceding feature extraction with models like 1D-CNNs. Word embeddings transform words into dense, low-dimensional vectors, encapsulating semantic nuances and inter-word relationships based on contextual usage, thereby enriching the model's semantic understanding. These embeddings facilitate dimensionality reduction, converting sparse, high-dimensional text data into a more condensed, manageable form. It is particularly critical in sentiment analysis, where discerning the subtle shades of meaning is pivotal for accurate classification. Furthermore, embeddings facilitate dimensionality reduction, converting sparse, high-dimensional text data into a more condensed, manageable form, optimizing learning efficiency and enhancing model performance. They also adeptly preserve the context in which words appear, a feature exploited by 1D-CNNs through their local feature detectors to identify and extract pertinent features indicative of sentiment. Moreover, the ability of embeddings to handle various forms of a word and reduce noise ensures a more robust and cleaner representation of the text, ultimately leading to more accurate and nuanced sentiment analysis through improved feature extraction.

BERT is a pioneering approach that employs a deeply bidirectional self-attention mechanism pre-trained on a vast corpus of data [40]. It includes the BookCorpus, which comprises over 11000 unreleased books from 16 distinct genres and 2.5 billion words of text passages from English Wikipedia. RoBERTa (Robustly optimized BERT approach) is an improved version of BERT that shares many similar configurations [41]. RoBERTa achieves its superior performance through several adjustments to BERT, including the use of more extensive training data, dynamic masking patterns, longer sequence training, and the replacement of the next sentence prediction. RoBERTa undergoes training on substantially larger datasets, utilizing an increased batch size and longer sequence length, spanning an extended duration. This system's training involves using four sets of data: CC-News, BookCorpus + English Wikipedia, OpenWebText, and Stories. RoBERTa enhances BERT's performance by increasing the dataset size and optimizing hyperparameters.

To generate word embeddings from the review texts, this research utilized the RoBERTa model. The process began with tokenization, where each word in the review texts was tokenized, and each token was assigned a unique input identifier corresponding to its respective index in the RoBERTa vocabulary. Following tokenization, an attention mask was allocated to each token to indicate its relevance within the context of the input sequence, prioritizing important tokens while disregarding those of lesser significance. These input identifiers and attention masks were then utilized by the RoBERTa model, which comprises 12 layers with 768 hidden states each. By utilizing a self-attentive approach, the RoBERTa model effectively processes the input sequence, enabling it to capture contextual information across various levels of abstraction.

RoBERTa implements dynamic masking in each epoch by replicating the training dataset ten times for each training instance. This allows each sequence to be masked in ten different ways over the course of forty epochs. RoBERTa employs not just the representation of the [CLS] token but also integrates all of the final hidden representations to enhance its performance. The [CLS] token, which stands for "classification token", is used to indicate the beginning of a sequence of tokens that needs to be classified. Mathematically, the output of the RoBERTa model can be represented as $T_i = \mathbb{R}_s$, where s denotes the hidden size of RoBERTa. The proposed model takes the input {[CLS], $\mathrm{tok}_1, \mathrm{tok}_2, \ldots, \mathrm{tok}_n$} and {[CLS], $a_1, a_2, \ldots, a_n$} and provide the result A $\in \mathbb{R}_{(m+1)\times s}$ and O $\in \mathbb{R}_{(n+1)\times s}$, where [CLS] is represented by [A] and [O].

The feature extraction process employed a 1D-CNN to improve the outcomes yielded by the RoBERTa model. By incorporating the 1D-CNN into the RoBERTa framework, utilizing both the contextual information gathered by RoBERTa and the long-range connections among tokens became feasible. This combined approach facilitates the production of more precise feature extraction. The process involves generating contextual embeddings for each token in the input sequence through RoBERTa. These embeddings are then fed into the 1D-CNN, which captures long-range dependencies and enhances feature extraction. Subsequently, the outputs from RoBERTa and the 1D-CNN are concatenated by aligning the representations along the hidden dimensions to form a unified feature vector. These concatenated features are used for downstream tasks, such as classification or regression, leveraging the rich contextual and long-range information captured by the model.

4.4. Feature vector generation

Different methods exist for feature extraction but a one-Dimensional Convolutional Neural Network (1D-CNN) is one the best algorithm among existing methods, which is commonly used for processing sequential data, such as text data [42]. In a

1D-CNN, the input data is typically represented as a sequence of values or features arranged in a one-dimensional array. The network applies filters or kernels, small sliding windows, to the input data. These filters move across the input, performing convolution operations at each position. The filters learn to extract relevant patterns or features from the input data by detecting local dependencies or correlations between neighboring elements. The convolutional layer in a 1D-CNN is responsible for learning the filters, and the output of this layer is often passed through an activation function to introduce non-linearity. Multiple convolutional layers can be stacked to learn increasingly complex features.

Moreover, pooling layers are usually used to reduce their spatial dimensions to effectively downsize the convolutional layer output and capture the most prominent features. The final layers of a 1D-CNN typically consist of fully connected layers, which learn to map the extracted features to the desired output, such as classification labels or regression values. Activation functions and regularization techniques can follow these fully connected layers to improve the model's performance.

The two-dimensional matrix is $F \in \mathbb{R}_{(n+1)\times H}$, R can be represented as

(1) $$R = r_0 \oplus r_1 \oplus \ r_2 \oplus \dots \oplus r_n,$$

where $\oplus$ stands for the concatenation operation.

A filter, represented by $w \in \mathbb{R}_{1\times H}$, is utilized during each convolutional operation to process a window consisting of l words, creating a new feature. Fig. 2 illustrates the configuration of a 1D-CNN featuring a single filter. For example, a feature c_i can be utilized from a window of features $r_{i:i+l-1}$ by the equation

(2) $$c_i = f\,(w \cdot r_{i:i+l-1} + b),$$

where the bias term is denoted as b and a non-linear function is represented by f. The feature map of the review/sentence is

(3) $$c = [c_0, c_1, \dots ,\ c_{n-l+1}],$$

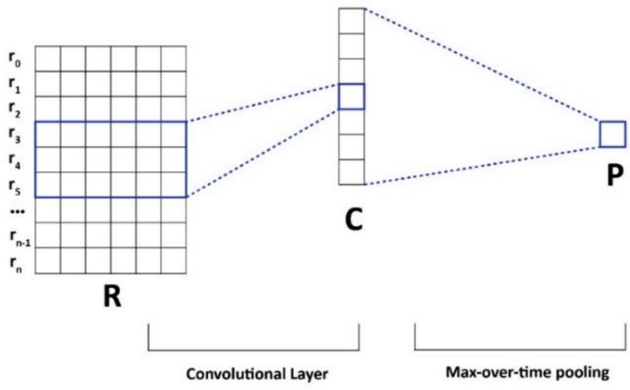


Fig. 2. 1D-CNN framework

The 1D-CNN model resembles an n-gram as it generates the representation of features by capturing information from the local filter region. This characteristic empowers the 1D-CNN to detect phrase features in textual data efficiently. Lastly, the size of these features is reduced through max-over-time pooling. It involves selecting the maximum value for each feature dimension across all time steps in the sequence. By capturing the most salient features, max-over-time pooling creates a fixed-length representation of the sequence that summarizes its important information. This pooling operation is commonly used in text mining and sentiment analysis tasks, where it helps identify key features and their positions within the sequence. By applying a 1D-CNN with H filters to text features, the output is $p \in \mathbb{R}_H$.

4.5. Product recommendation

This section wraps up the process by implementing a BiLSTM model for final classification once feature vector generation is completed. BiLSTM is chosen for its proficiency in handling sequential data in both forward and backward directions, taking advantage of the sharing of internal weights across sequences [10]. In contrast to conventional LSTM models, which may face certain limitations in this aspect. The mathematical representation of LSTM is as follows:

(4) $$h_t = f(m_1.e_i + m_2.s_{i-1} + b_i).$$

That equation is formulated in the following manner: e_i denotes the word embedding, whereas the weight metrics are indicated by m_1 and m_2, respectively. The bias term is denoted by b_i and the non-linear function is represented by $f(x)$. s_{i-1} is the hidden state vector at the previous time step, which helps in maintaining the continuity of information through the sequence. The unidirectional nature of sentence processing can pose difficulties for conventional LSTM models handling post-word information. However, BiLSTM models have the ability to operate bidirectionally, thus overcoming this limitation. Additionally, BiLSTM models excel in capturing vanishing errors and long-term dependencies by utilizing logical gates.

Each LSTM layer executes both forward and backward computations within the neural networks in this context. It implies that the output layer may encounter challenges with the dimensionality of the vector representation. Two hidden states h_t^{forward} and h_t^{backword} from these LSTM units are combined into a final hidden state h_t^{bilstm}. It is shown in the next equation,

(5) $$h_t^{\text{bilstm}} = h_t^{\text{forward}} + h_t^{\text{backword}}.$$

Within the sphere of binary recommendation systems, the principal aim is to ascertain the suitability of recommending a specific item to a user. The framework operates on pairs of user-product vectors, meticulously constructed by leveraging the rankings associated with various products. The essence of this recommendation approach hinges on the assignment of binary labels. In this context, a label of ‘1’ indicates an item being recommended, while alternately, a label of ‘0’ signifies a non-recommendation status.

Achieving this classification necessitates the implementation of a specific threshold value, represented as g. This value is applied to the ratings of the training examples, each of which is a compilation of a user vector, an item vector, and a rating. The classification hinges on a comparison between the product rating r_{ji} and the threshold g. If r_{ji} is greater than or equal to g, the item is recommended and labeled as 1. On the other hand, if it does not meet the threshold, it is classified as 0, indicating it should not be recommended.

To pinpoint the exact threshold value, represented as g_{ji}, the system employs a technique known as linear mapping. It is represented mathematically as

(6) $$g_{ji} = W_g \cdot \begin{matrix} u_i \\ v_j \end{matrix} + b_g.$$

Here, we deal with two specific vectors: u_i representing the user, and v_i representing the item. The process involves combining these two vectors to fabricate a new column vector, denoted as W_g. This new vector W_g embodies the parameters essential for linear mapping and is part of the space $W_g \in R^{1\times 2n}$. Additionally, the equation features a bias term b_j that resides in R, influencing the determination of the threshold.

To ascertain the likelihood of recommending the product v_j to the user u_i the system adopts the logistic sigmoid function. This function is pivotal as it maps any real-valued number to a range between 0 and 1, making it an ideal candidate for estimating recommendation scores. The representation of this function is

(7) $$f(g_{ij}) = \frac{1}{1+e^{-g_{ij}}}.$$

Through this detailed process, the binary recommendation system effectively determines whether a product aligns with a user's preferences and should be recommended based on the calculated scores and the established threshold.

5. Results

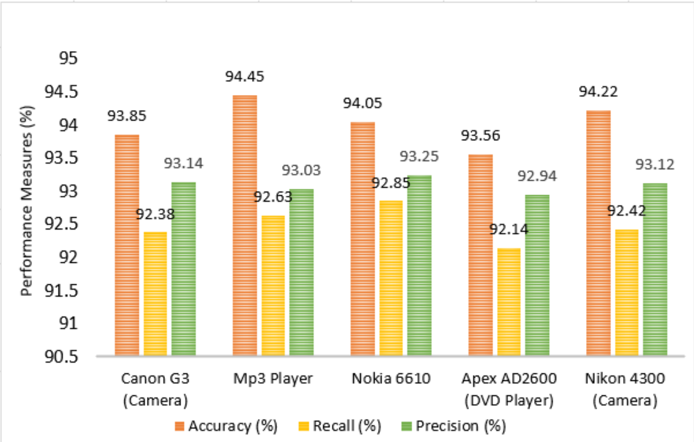
The k-Fold cross-validation method [40] is utilized to assess the effectiveness of sentiment classification. The hyperparameters listed in Table 3 were selected based on preliminary experiments conducted using k-Fold cross-validation. Initially, a range of possible values for each hyperparameter was evaluated using GridSearchCV. The k-Fold cross-validation method, with *K* set to 5, was utilized to assess the effectiveness of sentiment classification. The dataset was divided into 5 equally sized bins, one of which was used for testing while the remaining four were used for training. This process was repeated 5 times, each time with a different test bin, and the results were averaged to provide a comprehensive performance evaluation. During the hyperparameter tuning phase, GridSearchCV was employed to systematically explore various combinations of hyperparameters and identify the optimal values. The final hyperparameters reported in Table 3 are the ones that yielded the best performance across the folds during this tuning process.

Table 3. Details of hyperparameters

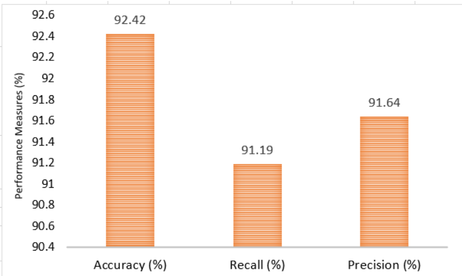
Hyperparameter	Values
Hidden units	128
Dropout rate	0.5
Learning rate	1×10^{-4}
Batch size	32
Sequence length	200
Epochs	20

The code is available at the following repository: **Sentiment Classification Repository.**

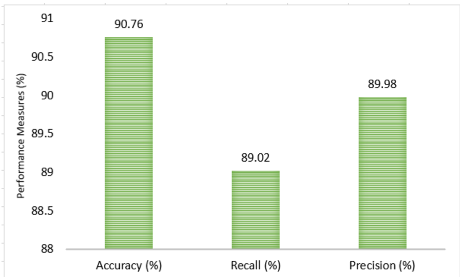
The experimental results indicate that the proposed model has achieved exceptionally high levels on Hu and Liu Dataset (DS-1) featuring products like cameras and MP3 players, showing performance with accuracy ranging from 92.56% up to 93.45%, recall from 91.14% up to 91.85%, and precision from 91.94% up to 92.25%. The SemEval Dataset (DS-II) encompasses a broader spectrum of products and hotels, showing slightly lower performance with 92.21% accuracy, 91.19% recall, and 91.64% precision, indicating that models may perform better on more homogenous data. The Bo Pang and Lillian Lee Dataset (DS-III) with movie reviews has the lowest metrics (90.76% accuracy, 89.02% recall, and 89.98% precision), suggesting the subjective nature of movie reviews introduces complexity, making accurate classification more challenging. Graphically, the results are shown in Fig. 3. Overall, datasets with more specific and objective content, such as individual product reviews, appear to yield higher performance metrics, while more varied and subjective data can result in a slight decrease in model effectiveness.



(a) DS-1



(b) DS-2

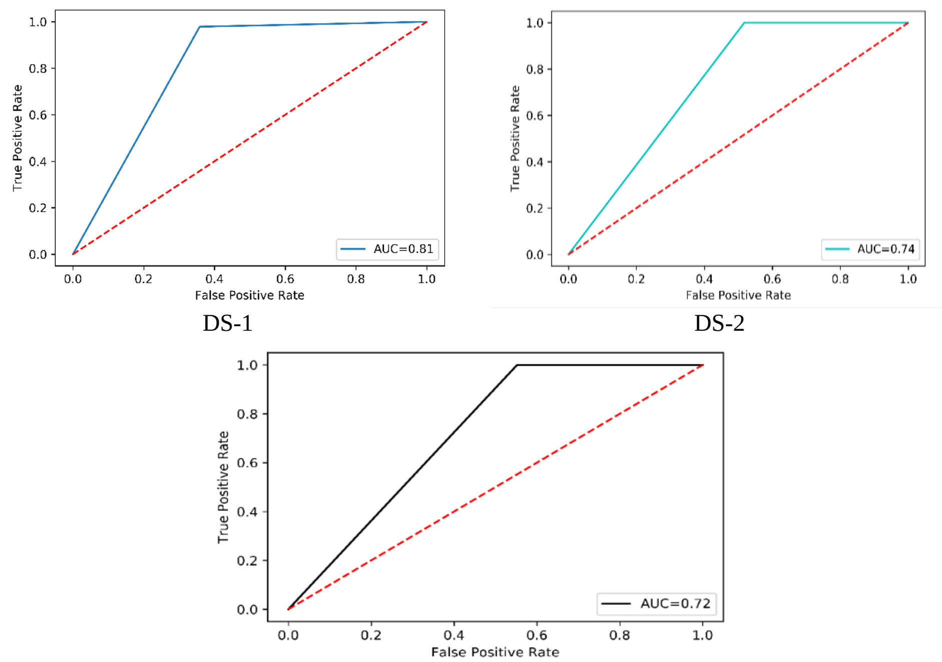


(c) DS-3

Fig. 3. Experimental results of proposed approach on DS-1, DS-2 and DS-3

The Receiver Operating Characteristic (ROC) has been employed to measure the performance of the proposed framework, specifically in terms of True Positive

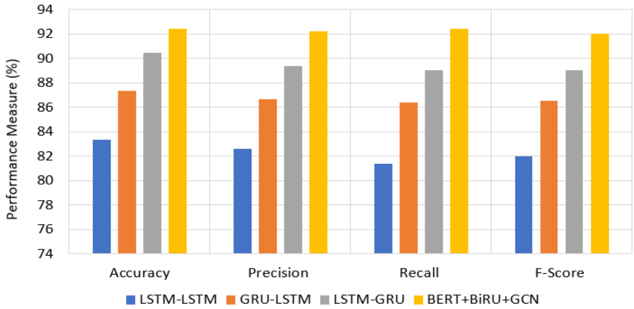
Rate (TPR) and False Positive Rate (FPR). The ROC curve is a graphical representation of the true positive rate against the false positive rate for a diagnostic test's different possible cut points. Fig. 4 presents the ROC curves for each dataset, illustrating the corresponding TPR and FPR values. Each ROC curve is analyzed based on the area it covers; in this scenario, all ROC curves span more than 80% of the area. Covering such a substantial area indicates high precision in the model's predictions. This extensive coverage substantiates the efficacy of the proposed framework, verifying its capability to accurately distinguish between positive and negative classes and, thus, deeming it a reliable model for the task at hand.



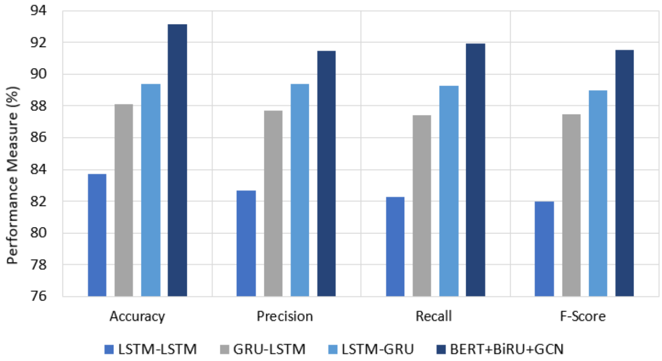
DS-3

Fig. 4. The ROC curve on multidomain datasets

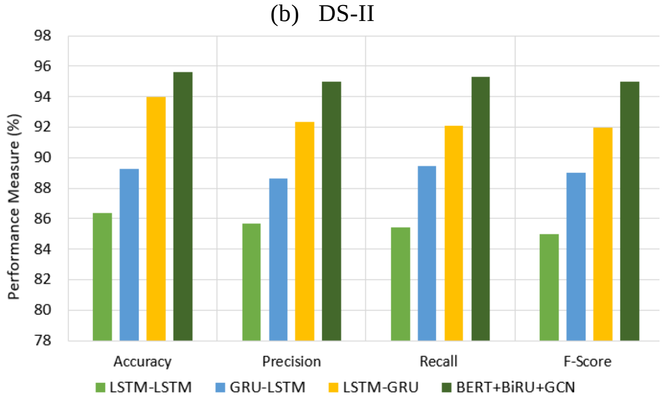
The initial analysis compares the performance of the proposed model with three unique combinations of standard deep learning architectures, each consisting of stages for feature extraction and sentiment classification. These configurations include LSTM-LSTM, GRU-LSTM, and LSTM-GRU. We evaluated each model's performance using k-fold cross-validation; the results are presented in Fig. 5. These results indicate that while the LSTM-GRU layers have led to reasonably high-performance metrics, their combination does not outperform the proposed model, RoBERTa+1D-CNN+BiLSTM, which achieved the best results among these models.



(a) DS-1



(b) DS-II



(c) DS-III

Fig. 5. Comparative analysis of proposed model with LSTM-LSTM, GRU-LSTM and LSTM-GRU

The second comparative analysis, the proposed model performance, was compared with baseline sentiment analysis approaches. The goal is to evaluate each model's effectiveness and identify each approach's strengths and weaknesses. The eight baseline approaches of deep learning and machine learning are:

1. LMIAN [35] employed the BERT model in conjunction with Point-wise Convolution Transformation (PCT) to establish correlations between aspect words and their respective contexts for enhanced sentiment analysis.

2. BERT-ADA [43] introduces a domain-adaptive approach that makes use of a task-specific corpus to further train BERT-BASE for ABSA, resulting in a remarkable performance on the Laptop dataset.

3. MGAN [44] uses a Bi-directional LSTM network and incorporates a multi-granularity attention mechanism to capture the intricate relationship between aspects and contexts.

4. BERT+ARM [5] captures the aspects and opinions using ARM and classifies them using BERT.

5. SVM + Decision tree [45] Supervised machine learning models such as Support Vector and Decision tree used for sentiment analysis.

6. ABSA-PER [46]: Designed to estimate the polarity of customer reviews based on ABSA using aspect co-occurrence calculation and sentiment classification.

7. ABALGA [47]: ALGA (Adaptive Lexicon learning using a Genetic Algorithm) has been specifically developed to facilitate learning lexicons at the tweet level.

8. POS-Lexicon [4]: ABSA was proposed using a hybrid lexicon-based approach (POS, NGD, and GA).

These existing approaches are experimentally tested using the same training and evaluation procedures as the proposed method. The comparison of the proposed model with baseline models is shown in Table 4.

Table 4. Comparisons with the baseline models

Models	Product reviews (Hu and Liu)		Movie reviews		Products reviews (SemEval)	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
LMIAN	91.09	89.64	90.56	88.45	89.02	88.52
BERT-ADA	84.35	80.12	80.24	78.65	83.56	81.56
MGAN	81.25	78.95	76.64	75.02	81.01	79.95
BERT-ARM	89.45	87.65	84.64	82.12	88.62	86.45
SVM+ Decision Tree	87.62	85.25	86.12	84.85	86.72	84.96
ABSA-PER	89.82	88.32	87.45	85.96	89.12	87.12
ABALGA	90.56	88.98	88.12	87.13	87.45	85.19
POS-Lexicon	89.12	87.42	86.82	85.23	88.45	86.59
RoBERTa-BiLSTM	94.03	93.44	90.76	88.50	92.21	91.41

When it comes to comparing different deep learning models, the RoBERTa-BiLSTM model stands out as achieving the highest accuracy and F1 scores across all three datasets. It has outperformed several other models, including BERT-BASE, BERT-ADA, MGAN, and BERT-ARM. Although ABSA-PER, ABALGA, and POS-Lexicon have demonstrated competitive results, RoBERTa-BiLSTM has still managed to outperform them. While SVM+ Decision Tree has also performed well, it has fallen slightly behind the RoBERTa-BiLSTM model. The proposed technique, RoBERTa-BiLSTM, has shown superior performance, indicating its effectiveness in sentiment analysis tasks across different datasets.

The outcomes of the preceding experiments substantiate the evident advantages of the model introduced in this paper. These advantages stem from the potent feature extraction and nonlinear fitting capabilities intrinsic to deep learning, significantly enhancing the model's predictive performance. Additionally, the model showcased in this research exhibits pronounced superiority compared to other deep learning models. It can be attributed to the incorporation of RoBERTa and the unique timing characteristics intrinsic to the BiLSTM model. BiLSTM, being an amalgamation of forward and backward LSTM, is adept at capturing contextual relationships, thereby enabling improved predictive performance. Moreover, the model employs 1D-CNN for dimensionality reduction of features, enhancing its ability to extract features for sentiment analysis effectively.

We conducted an ablation study to evaluate the contribution of each component within our model, which primarily includes RoBERTa, 1D-CNN, and BiLSTM. This study aimed to discern the individual impact and relative importance of these elements. In this study, unlike in our original model proposal, we integrated additional components to observe their effects. The results, as detailed in Table 5, demonstrate a clear positive relationship between the complexity of the model, measured by the number of components and its overall performance. A standout finding is a marked improvement in performance across three datasets when 1D-CNN is used in conjunction with RoBERTa and BiLSTM. This improvement was especially pronounced in the product dataset. These results underscore the effectiveness of our proposed model, particularly its design, which adeptly focuses on aspect-related words.

The different components in our model are combined by first using RoBERTa to generate contextual embeddings, followed by processing these embeddings through either 1D-CNN or BiLSTM for pattern recognition. For two-component combinations, RoBERTa’s embeddings are either fed into 1D-CNN to capture local dependencies or into BiLSTM to capture long-term dependencies, enhancing performance by focusing on specific aspects of the data. These configurations effectively leverage the strengths of each component to improve overall model accuracy.

Table 5. Ablation study

Components						Performance metrics (Accuracy)		
RoBERTa	1D-CNN	BiLSTM	BERT	CNN	BiGRU	Product reviews	Movie reviews	Product reviews 2
			✓	✓	✓	85.36	86.32	84.63
			✓	✓		78.63	77.45	79.25
✓		✓		✓		86.89	87.12	86.34
	✓	✓	✓			89.36	90.05	90.56
	✓		✓		✓	90.45	88.36	87.56
✓	✓	✓				94.03	90.76	92.21

6. Conclusion

Online product reviews and comments have unprecedentedly impacted consumer behavior and sales. These reviews help improve product and service quality and offer valuable insights for potential buyers seeking recommendations. This study introduces a transformer-based deep learning approach to conduct Aspect-Based Sentiment Analysis (ABSA) on user reviews. The proposed approach leverages

RoBERTa, BiLSTM, and 1D-CNN to create a significantly enhanced deep-learning model. This model has been rigorously evaluated across three benchmark datasets. The results demonstrate that our model surpasses existing state-of-the-art baselines in multidomain benchmarks. Future research might explore more efficient decision-making techniques to refine further the accuracy of our model. Additionally, applying our approach to other domains, such as event identification, could open new pathways for research and development, broadening the scope of its applicability.

References

1. Kim, Y. J., H. S. Kim. The Impact of Hotel Customer Experience on Customer Satisfaction through Online Reviews. – Sustainability, Vol. **14**, 2022, No 2, p. 848.
2. Ketipov, R., V. Angelova, L. Doukovska, R. Schnalle. Predicting User Behavior in e-Commerce Using Machine Learning. – Cybernetics and Information Technologies, Vol. **23**, 2023, No 3, pp. 89-101.
3. Solairaj, A., G. Sugitha, G. Kavitha. Enhanced Elman Spike Neural Network Based Sentiment Analysis of Online Product Recommendation. – Applied Soft Computing, Vol. **132**, 2023, No 1, pp. 109789.
4. Lee, G. T., C. O. Kim, M. Song. Semisupervised Sentiment Analysis Method for Online Text Reviews – Journal of Information Science, Vol. **47**, 2021, No 3, pp. 387-403,
5. Tan, K. L., C. P. Lee, K. M. Lim. A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. – Appl. Sci., Vol. **13**, 2023, No 7, pp. 45-50.
6. Rana, M. R. R., S. Ur Rehman, A. Nawaz, T. Ali, M. Ahmed. A Conceptual Model for Decision Support Systems Using Aspect Based Sentiment Analysis. – Proc. of Romanian Academy Series A-Mathematics Physics Technical Sciences Information Science, Vol. **22**, 2021, No 4, pp. 381-390.
7. Rana, M. R. R., A. Nawaz, J. Iqbal. A Survey on Sentiment Classification Algorithms, Challenges and Applications. – Acta Universitatis Sapientiae, Informatica, Vol. **10**, 2018, No 1, pp. 58-72,
8. Mehbodniya, A., M. V. Rao, L. G. David, K. G. Joe Nigel, P. Vennam. Online Product Sentiment Analysis Using Random Evolutionary Whale Optimization Algorithm and Deep Belief Network. – Pattern Recognition Letters, Vol. **159**, 2022, No 1, pp. 1-8,
9. Khan, A. Improved Multi-Lingual Sentiment Analysis and Recognition Using Deep Learning. – Journal of Information Science, 2023.
10. Rana, M. R. R., A. Nawaz, T. Ali, A. M. El-Sherbeeny, W. Ali. A BiLSTM-CF and BiGRU-Based Deep Sentiment Analysis Model to Explore Customer Reviews for Effective Recommendations. – Engineering, Technology & Applied Science Research, Vol. **13**, 2023, No 5, pp. 11739-11746.
11. Gu, T., H. Zhao, Z. He, M. Li, D. Ying. Integrating External Knowledge into Aspect-Based Sentiment Analysis Using Graph Neural Network. – Knowledge-Based Systems, Vol. **259**, 2023, No 1, 110025.
12. Trillo, J. R., E. Herrera-Viedma, J. A. Morente-Molinera, F. J. Cabrerizo. A Large Scale Group Decision Making System Based on Sentiment Analysis Cluster. – Information Fusion, Vol. **91**, 2023, pp. 633-643.
13. Karn, A. L., R. K. Karna, B. R. Kondamudi, G. Bagale, D. A. Pustokhin, S. Sengan. Customer Centric Hybrid Recommendation System for e-Commerce Applications by Integrating Hybrid Sentiment Analysis. – Electronic Commerce Research, Vol. **23**, 2022, No 1, pp. 279-314.
14. Wu, H., C. Huang, S. Deng. Improving Aspect-Based Sentiment Analysis with Knowledge-Aware Dependency Graph Network. – Information Fusion, Vol. **92**, 2023, pp. 289-299.
15. Huang, B., R. Guo, Y. Zhu, Z. Fang, G. Zeng, J. Liu, Y. Wang, H. Fujita, Z. Shi. Aspect-Level Sentiment Analysis with Aspect-Specific Context Position Information. – Knowledge-Based Systems, Vol. **243**, 2022, 108473.

16. G a r c i a, K., L. B e r t o n. Topic Detection and Sentiment Analysis in Twitter Content Related to COVID-19 from Brazil and the USA. – Applied Soft Computing, Vol. **101**, 2021, 107057.
17. G u p t a, I., N. J o s h i. A Review on Negation Role in Twitter Sentiment Analysis. – International Journal of Healthcare Information Systems and Informatics, Vol. **16**, 2021, No 4, pp. 1-19.
18. K i m, H., Y. S. J e o n g. Sentiment Classification Using Convolutional Neural Networks. – Applied Sciences, Vol. **9**, 2019, No 11, 2347.
19. N a n d a l, N., R. T a n w a r, J. P r u t h i. Machine Learning Based Aspect Level Sentiment Analysis for Amazon Products. – Spatial Information Research, Vol. **28**, 2020, No 5, pp. 601-607.
20. P r a t h i, J. K., P. K. R a p a r t h i, M. V. G o p a l a c h a r i. Real-Time Aspect-Based Sentiment Analysis on Consumer Reviews. – In: Proc. of 3rd Data Engineering and Communication Technology: ICDECT-2K19, 2020, pp. 801-810.
21. S h a m s, M., N. K h o s h a v i, A. B a r a a n i-D a s t j e r d i. LISA: Language-Independent Method for Aspect-Based Sentiment Analysis. – IEEE Access, Vol. **8**, 2020, No 1, pp. 31034-31044.
22. L o n d h e, A., P. P. R a o. Aspect Based Sentiment Analysis – An Incremental Model Learning Approach Using LSTM-RNN. – In: Proc. of 5th International Conference in Advances in Computing and Data Sciences, 2021, pp. 23-24.
23. M u t i n d a, J., W. M w a n g i, G. O k e y o. Sentiment Analysis of Text Reviews Using Lexicon-Enhanced Bert Embedding (LeBERT) Model with Convolutional Neural Network. – Applied Sciences, Vol. **13**, 2023, No 3, 1445.
24. H u a n g, B., R. G u o, Y. Z h u, Z. F a n g, G. Z e n g, J. L i u, Y. W a n g, H. F u j i t a, Z. S h i. Aspect-Level Sentiment Analysis with Aspect-Specific Context Position Information. – Knowledge-Based Systems, Vol. **243**, 2022, No 1, 108473.
25. L i, X., L. B i n g, W. L a m, B. S h i. Transformation Networks for Target-Oriented Sentiment Classification. – In: Proc. of 56th Annual Meeting of the Association for Computational Linguistics ACL, 2018, pp. 946-956.
26. L i, W., L. Z h u, Y. S h i, K. G u o, E. C a m b r i a. User Reviews: Sentiment Analysis Using Lexicon Integrated Two-Channel CNN-LSTM, Family Models. – Applied Soft Computing, Vol. **94**, 2020, 106435.
27. Y u a n, Z., S. W u, F. W u, J. L i u, Y. H u a n g. Domain Attention Model for Multidomain Sentiment Classification. – Knowledge-Based Systems, Vol. **155**, 2018, No 155, pp. 1-10,
28. Z h a n g, Z., L. W a n g, Y. Z o u, C. G a n. The Optimally Designed Dynamic Memory Networks for Targeted Sentiment Classification. – Neurocomputing, Vol. **309**, 2018, pp. 36-45.
29. B a s i r i, M. E., S. N e m a t i, M. A b d a r, E. C a m b r i a, U. R. A c h a r y a. ABCDM: An Attention-Based Bidirectional CNN-RNN Deep Model for Sentiment Analysis. – Future Generation Computer Systems, Vol. **115**, 2021, No 1, pp. 279-294.
30. O n a n, A. Sentiment Analysis on Product Reviews Based on Weighted Word Embeddings and Deep Neural Networks. – Concurrency and Computation: Practice and Experience, Vol. **33**, 2021, No 23, e5909.
31. N a s e e m, U., I. R a z z a k, K. M u s i a l, M. I m r a n. Transformer Based Deep Intelligent Contextual Embedding for Twitter Sentiment Analysis. – Future Generation Computer Systems, Vol. **113**, 2020, No 1, pp. 58-69.
32. W a n g, L., W. G u o, X. Y a o, Y. Z h a n g, J. Y a n g. Multimodal Event-Aware Network for Sentiment Analysis in Tourism. – IEEE MultiMedia, Vol. **28**, 2021, No 2, pp. 49-58.
33. R a o, G., W. H u a n g, Z. F e n g, Q. C o n g. LSTM with Sentence Representations for Document-Level Sentiment Classification. – Neurocomputing, Vol. **308**, 2018, No 1, pp. 49-57.
34. X u, J., Z. L i, F. H u a n g, C. L i, P. S. Y u. Social Image Sentiment Analysis by Exploiting Multimodal Content and Heterogeneous Relations. – IEEE Transactions on Industrial Informatics, Vol. **17**, 2021, No 4, pp. 2974-2982.
35. Z h e n g, W., S. Z h a n g, C. Y a n g, P. H u. Lightweight Multilayer Interactive Attention Network for Aspect-Based Sentiment Analysis. – Connection Science, Vol. **35**, 2023, No 1, 2189119.
36. W u, D., M. C h i. Long Short-Term Memory with Quadratic Connections in Recursive Neural Networks for Representing Compositional Semantics. – IEEE Access, Vol. **5**, 2017, pp. 16077-16083.
37. H u, M., B. L i u. Mining and Summarizing Customer Reviews. – In: Proc. of 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004, pp. 168-177.

38. L i n, Y., C. W a n g, H. S o n g, Y. L i. Multi-Head Self-Attention Transformation Networks for Aspect-Based Sentiment Analysis. – IEEE Access, Vol. **9**, 2021, No 1, pp. 8762-8770.
39. P a n g, B., L. L e e. Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales. – arXiv preprint cs/0506075, 2005.
40. O l i a e e, A. H., S. D a s, J. L i u, M. A. R a h m a n. Using Bidirectional Encoder Representations from Transformers (BERT) to Classify Traffic Crash Severity Types. – Natural Language Processing Journal, Vol. **3**, 2023, No 1, 100007.
41. L i a o, W., B. Z e n g, X. Y i n, P. W e i. An Improved Aspect-Category Sentiment Analysis Model for Text Sentiment Analysis Based on RoBERTa. – Applied Intelligence, Vol. **51**, 2020, No 1, pp. 3522-3533.
42. O z c a n l i, A. K., M. B a y s a l. Islanding Detection in Microgrid Using Deep Learning Based on 1D CNN and CNN-LSTM Networks. – Sustainable Energy, Grids and Networks, Vol. **32**, 2022, No 1, 100839.
43. R i e t z l e r, A., S. S t a b i n g e r, P. O p i t z. Adapt or Get Left Behind: Domain Adaptation through Bert Language Model Finetuning for Aspect-Target Sentiment Classification. – In: Proc. of 12th Language Resources and Evaluation Conference, 2020, pp. 4933-4941.
44. F a n, F., Y. F e n g, D. Z h a o. Multi-Grained Attention Network for Aspect Level Sentiment Classification. – In: Proc. of 2018 Conference on Empirical Methods in Natural Language Processing, 2023, pp. 3433-3442.
45. R a h m a n, H., J. T a r i q, M. A. M a s o o d, F. S u b a h i. Multi-Tier Sentiment Analysis of Social Media Text Using Supervised Machine Learning. – Computers, Materials & Continua, Vol. **74**, 2023, pp. 5527-5543.
46. B a n j a r, A., Z. A h m e d, A. D a u d, R. A. A b b a s i, H. D a w o o d. Aspect-Based Sentiment Analysis for Polarity Estimation of Customer Reviews on Twitter. – Computers, Materials & Continua, Vol. **67**, 2021, pp. 2203-2225.
47. M o w l a e i, M. E., M. S a n i e e A b a d e h, H. K e s h a v a r z. Aspect-Based Sentiment Analysis Using Adaptive Aspect-Based Lexicons. – Expert Systems with Applications, Vol. **148**, 2020, 113234.

Received: 16.04.2024; Second Version: 13.08, Third Version – 15.08; Accepted: 15.08.2024