SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES

http://saucis.sakarya.edu.tr/

e-ISSN: 2636-8129
Publisher: Sakarya University

Vol. 8, No. 3, 485-496, 2025
DOI: 10.35377/saucis...1675280

RESEARCH ARTICLE

Bi-directional Encoder Representations from Transformers Based for Sentiment Analysis from Consumer Reviews

Hanife Göker[1]

[1]Gazi University, Health Services Vocational College, Ankara, Türkiye, ror.org/054xkpr46

Corresponding author:
Hanife Göker, Gazi University, Health Services Vocational College, Ankara, Türkiye
gokerhanife@gazi.edu.tr

Article History:
Received: 13.04.2025
Revised: 19.07.2025
Accepted: 17.08.2025
Published Online: 26.09.2025

ABSTRACT

Structured data has a standardized format for easy access, organization, and categorization. However, approximately 95% of data, such as text files or online reviews, is unstructured, and these texts do not have standard rules. Unstructured data analysis, especially when the amount of data to be examined is substantial, requires considerable effort, cost, and time, and classical statistical methods are often insufficient. Transformer models, the latest technological models in natural language processing (NLP), are the strongest candidates to overcome these limits. In this paper, we propose the bi-directional encoder representations from transformers (BERT) model-based solution for sentiment analysis of consumer reviews. The dataset comprises 10975 consumer reviews of technological products from an e-commerce platform and was transformed into a structured dataset using data preprocessing. Then, we compared the performance of the BERT transformer model with deep learning models, specifically convolutional neural networks (CNN), long short-term memory (LSTM), and bidirectional long short-term memory (B-LSTM). Experimental results confirmed that the BERT transformer model achieved a higher kappa of 96.6% and an overall accuracy of 97.78% for multi-classification of consumer reviews. The proposed transformer-based model outperforms the state-of-the-art models, providing a reliable and efficient solution.

Keywords: Transformer model, Deep learning, Natural language processing, Sentiment analysis, E-commerce.

1. Introduction

E-commerce is a business model that involves purchasing products or services and conducting commercial online transactions. This business model offers numerous advantages to both customers and sellers by eliminating the need for face-to-face interaction in a physical environment. Ease of purchase and product variety are the main advantages of e-commerce for customers [1]. It also saves the seller from the expense of renting a store. E-commerce benefits businesses and consumers by introducing various service models into our lives, thereby impacting the entire economy on a macro scale. In 2020, four out of five people using the internet have experienced e-commerce [2]. Globally, e-commerce sales reached $2.98 trillion in 2018, $3.53 trillion in 2019, and $6.98 trillion in 2020 [3]. E-commerce is no longer an alternative to traditional commerce but is increasingly becoming an integral part of daily life, trade, and shopping. Consumers' online evaluations and ratings influence nearly 50% of purchase decisions [4]. Approximately 71% of global online shoppers always read online consumer reviews before purchasing a product [5].

Online consumer reviews are user-generated content that includes information and reviews about a product, company, or service [6]. The experience and satisfaction of those who buy and use a product on e-commerce sites are among the most critical parameters that affect consumers' decisions to purchase a product and product sales [7-8]. The reviews and ratings of other consumers have a significant impact on purchasing decisions [9]. Moreover, consumer evaluations and ratings are more trustworthy than marketing and advertisements [10]. These reviews reduce the consumer's perceived risk associated with online shopping and encourage the purchase decision by providing helpful information about the goods and services being purchased. Companies that want to capture a share of the e-commerce market must understand the rules of the e-commerce ecosystem and develop strategies accordingly [11]. Therefore, e-commerce companies must strive to obtain the best possible reviews and ratings. Because consumer and market preferences vary over time, businesses invest extensively to better assess online consumer reviews. Additionally, technical advancements have increased competition across organizations. E-commerce companies, such as Amazon, invest a significant amount of money in managing sellers and ensuring they act in their best interests [12]. Data analysis generally focuses on structured data. Structured data is stored with specific rules and systems. This data is, therefore, easily accessible, organized, and categorized. However, more than 95% of data generated in electronic media, such as text files, articles, books, online consumer reviews, and blogs, is in an unstructured form [13]. Unstructured data, on the other hand, lacks standard rules, making it challenging to analyze. The language and meaning of the text change depending on its purpose [14]. It is more difficult for a computer to understand and classify unstructured data than structured data;

Cite as: H. Göker, "Bi-directional Encoder Representations from Transformers Based for Sentiment Analysis from Consumer Reviews" *Sakarya University Journal of Computer and Information Sciences,* vol. 8, no. 3, pp. 485-496, 2025. doi: 10.35377/saucis...1675280

therefore, classical statistical methods are insufficient for analyzing this type of data. Especially if the number of texts to be classified is large, this process requires much more effort, time, and cost [15]. NLP, a machine learning technology, enables computers to understand, analyze, and interpret human language, with a wide range of applications, including chatbots, text categorization, and language translation. NLP is used to develop chatbots that can interact with customers in natural language, analyze customer feedback to improve products and services, and translate product descriptions into multiple languages to reach a wider audience. It is an alternative solution to analyze this unstructured data. Deep learning and traditional machine learning algorithms are widely used for NLP tasks, especially sentiment analysis.

Sentiment analysis based on NLP techniques has significant applications in e-commerce. NLP is utilized for sentiment analysis of consumer reviews to enhance the effectiveness of recommendation systems. These sentiment analysis-based e-commerce applications have focused on investigating the connection between e-commerce logistical services and customer satisfaction [16], social media texts to analyze consumer behavior in sustainable fashion marketing [17], evaluation of airline passenger reviews [18], and modeling of political views from data on the web and predicting [19]. Aftab et al. (2021) present a solution for predicting consumers' sentiment analysis regarding social media marketing in the online industry, as well as whether consumers will purchase products. The methodology employed the recurrent neural networks (RNN), Random Forest (RF), Naive Bayes, and Bayes Net models for sentiment analysis on social media. The RNN model achieved an accuracy of 97.15% and a precision of 94.25% [20]. Jabbar et al. (2019) analyzed both positive and negative reviews to understand people's sentiments about e-commerce products using a support vector machine (SVM) algorithm. The algorithm achieved an F1-score of 93.54% [21]. Dey et al. (2020) compared the sentiment analysis of consumer reviews of Amazon products. They used Naïve Bayes and Linear SVM machine learning approaches. Features were extracted using the TF-IDF method. The results showed that Linear SVM had an accuracy of 84% in classifying Amazon product feedback, while the Naïve Bayes algorithm achieved an accuracy of 82.87% [22]. Pratama et al. (2022) utilized the SVM algorithm to categorize reviews of cosmetic products. Features were extracted using the TF-IDF method after labeling the data as positive and negative. The SVM algorithm classified with 80.06% accuracy [23]. Although NLP techniques are an alternative solution to analyze unstructured data on e-commerce platforms, separating texts into distinct groups is an ideal problem for NLP. The design of neural architecture for data representation is a crucial intermediate step. Effective text classification has always depended on this deep language representation. BERT transformers have been proposed in recent years and have become an effective representation model successfully used in developing state-of-the-art models for numerous NLP applications, such as sentiment analysis [24]. The BERT transformers outperform numerous task-specific architectures and perform best on token-level and sentence-level tasks.

In the paper, we proposed a BERT transformer model-based solution for sentiment analysis of online consumer reviews on an e-commerce platform. Unlike many existing studies that focus on binary sentiment classification, this paper emphasizes a multi-class sentiment analysis approach using a large-scale, real-world dataset obtained from an e-commerce environment. Moreover, we comparatively evaluate BERT against deep learning models (CNN, LSTM, and B-LSTM) under the same experimental setup, which provides a broader perspective on model performance and robustness. This approach highlights the practical advantages of BERT in real-world e-commerce applications and contributes to the novelty of the study by evaluating its relative performance against alternative models and distinguishing it from sentiment analysis research. The following is a summary of this paper's contributions:

i) The proposed transformer-based sentiment analysis model received raw inputs without the use of traditional feature extraction techniques, such as bag-of-words, n-grams, and TF-IDF.

ii) The performances of transformer and deep learning models were compared for multi-labelled sentiment analysis from consumer reviews on e-commerce platforms. To the best of our knowledge, this paper represents a rare attempt at multi-label sentiment analysis, comparing the transformer model and deep learning algorithms.

iii) The BERT transformer-based model 10975 was evaluated on consumer reviews. It can work efficiently on large datasets and process texts more effectively by taking long-distance relationships into account. These are important for sentiment analysis because e-commerce reviews are often large datasets, and customer sentiments are often expressed through words and phrases that appear in different parts of the texts.

iv) Experimental results confirmed that the proposed transformer-based model outperforms deep learning models. Moreover, the proposed transformer model outperforms other state-of-the-art models in the relevant literature.

2. Materials and Methods

2.1. The Proposed Transformer-Based Sentiment Analysis Model

The proposed transformer-based model includes the following stages: i) In the initial stage (data acquisition), the dataset was collected from 10975 consumer reviews on technological products, taken from the electronics category on Amazon, the most well-known e-commerce site. ii) In the second stage, data preprocessing was carried out, which included data transformation, stemming, stop word removal, and tokenization. iii) Subsequently, the dataset was split into training and test datasets to assess the classification performance via the holdout method. iv) Lastly, based on the model evaluation criteria, the performances of the BERT transformer and deep learning models were compared. Figure 1 presents the stages of the proposed transformer-based sentiment analysis model.

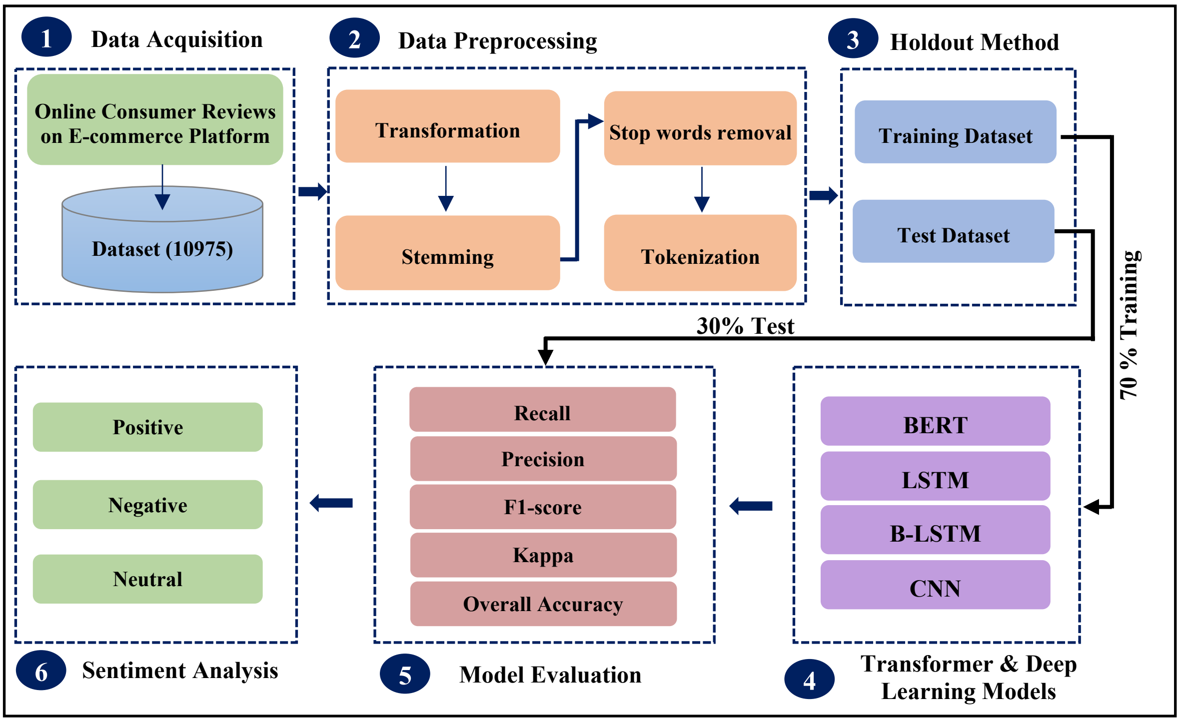

Figure 1. The Stages of the Proposed Transformer-Based Sentiment Analysis Model

2.2. Data Acquisition

Data acquisition is the process of collecting the initial data related to the problem under consideration, defining the obtained data, and determining whether the study meets the needs and the quality of the data. The dataset was collected from 10975 consumer reviews of technological products. Consumer reviews were sourced from the electronics category on Amazon, the most well-known e-commerce site. Consumer reviews were categorized into three categories, namely “positive”, “negative”, and “neutral”. The characteristics of the consumer reviews dataset are presented in Table 1.

Table 1. The Characteristics of the Consumer Reviews Dataset

Label	Number of consumer reviews
Negative	4048
Neutral	2751
Positive	4176
Total	10975

The received data consisted of text files. Therefore, using data pre-processing, this unstructured data needs to be transformed into structured data for the purpose of sensitivity analysis.

2.3. Data Preprocessing

Data preprocessing includes “transformation”, “stemming”, “stop words removal”, and “tokenization”. On the internet, documents are typically stored in various formats, such as HTML and XML. In the transformation step, the texts are cleared of HTML and XML tags. Mentions (@), hashtags (#), and links in the data obtained through social networks are deleted. In the tokenization step, the text is separated into words. Punctuation marks, symbols, and numbers are extracted from the text [25]. In addition, all words are converted to lowercase letters. In the step of “stop and redundant words removal”, prepositions (like, but, for, etc.), pronouns (you, we, they, etc.), and conjunctions (and, or, with, etc.) are removed. Although these terms are frequently used in the text, they do not have specific meanings. Stop words make up 20-30% of the total number of words in a given text document [26]. While looking at the frequencies of the words, in the step of removing the stop words, the roots of the words are found, so that the words from the same root with different suffixes are perceived as the same word. “Stemming” matches similar words in the text document and can reduce indexing size by up to 40-50%.

In the data preprocessing stage of the proposed transformer-based sentiment analysis model, data transformation, removal of stop words and redundant words, tokenization, and stemming were carried out. Text files containing online consumer reviews were pulled from the e-commerce platform. The consumer reviews contained characters in different languages. Therefore, to ensure standardization, the other characters in the texts were converted to English characters during the data transformation step. Then, the stop and redundant words that were present in the text but had no meaning in the classification process were removed from the text (and, or, because, but, with, both, moreover, this, that, these, those, etc.). In the tokenization step, terms such as punctuation, numbers, and symbols were removed. In addition, user names and URL addresses from online consumer reviews were removed. The texts were converted to lowercase. Finally, in the stemming step, the roots of the words were determined.

2.4. Holdout Method

The consumer reviews dataset was split into training and test sets to evaluate the classification performance using the holdout method. In the process, a certain ratio of the dataset is allocated as the test set, and the remaining data is used as the training set. The holdout method is widely preferred for random division of the training and test datasets. In this paper, 70% of the dataset was used as training data (n = 7683), and 30% was used as test data (n = 3292). The distribution of the consumer reviews dataset is given in Table 2:

Table 2. Distribution of the Consumer Reviews Dataset

Dataset	Label	Number of consumer reviews	Total
Training set	Negative	2834	7683 (70% of the dataset)
	Neutral	1926	
	Positive	2923	
Test set	Negative	1214	3292 (30% of the dataset)
	Neutral	825	
	Positive	1253	

2.5. Transformer and Deep Learning Models

Deep learning models, including LSTM, B-LSTM, and CNN, have made significant contributions to learning sequential data and have been successfully applied to NLP tasks, such as text classification, sentiment analysis, and language modeling. However, these models have limitations in modeling long dependencies and reducing processing times. Transformer models, presented as a solution to these problems, can more effectively model the structural relationships of language with their attention-based structure. The transformer model is a deep learning architecture based on the self-attention mechanism, providing high accuracy and efficiency in NLP tasks that require sequential data processing. The model can process all parts of the data simultaneously, allowing it to be trained more effectively and quickly. In this study, the performances of the BERT transformer and deep learning models were compared based on the model evaluation criteria. The hyper-parameters used in training these models were applied under the same training conditions in all models and are presented in Table 3:

Table 3. The hyper-parameters of the classification models

Hyper-Parameters	LSTM	B-LSTM	CNN	BERT
MaxEpochs	30	30	30	30
Optimizer	'adam'	'adam'	'adam'	'adam'
MiniBatchSize	128	128	128	128
InitialLearnRate	1.0000e-03	1.0000e-03	1.0000e-03	1.0000e-03
SequenceLength	'longest'	'longest'	'longest'	'longest'
L2Regularization	1.0000e-04	1.0000e-04	1.0000e-04	1.0000e-04
LearnRateDropFactor	0.1000	0.1000	0.1000	0.1000
LearnRateDropPeriod	10	10	10	10
ValidationFrequency	50	50	50	50
Shuffle	'every-epoch'	'every-epoch'	'every-epoch'	'every-epoch'

Table 3 presents the hyper-parameters for the LSTM, B-LSTM, CNN, and BERT models. The LSTM, B-LSTM, CNN, and BERT models were configured to use a maximum of 30 training epochs, with the 'adam' optimizer and a mini-batch size of 128. The initial learning rate was set to 0.001, and L2 regularization was applied with a coefficient of 0.0001 to prevent overfitting. A learning rate drop factor of 0.1 was employed, with a drop period of every 10 epochs. Validation was performed every 50 iterations. The sequence length was set to 'longest' across all models to ensure uniformity in input size, and the training data were shuffled at the end of each epoch. These standardized hyper-parameter settings enabled the evaluation of model performance based solely on model architecture. The theoretical background of the BERT transformer model algorithm, as well as LSTM, B-LSTM, and CNN deep learning algorithms, is presented in the following sections.

2.5.1. BERT Transformer Model Algorithm

The BERT model utilizes a multilayer, bi-directional transformer encoder and has an attention mechanism that captures contextual connections between words within a text. Bi-directional language models acquire more general language representations. One sentence or two sentences can be represented by the BERT input representation in a single token sequence [24]. The transformer encoder takes the entire sequence of words simultaneously, unlike linear directional models that read text input in left-to-right or right-to-left order. So, it is considered bi-directional, but it would be more accurate to describe it as non-directional. This enables the model to understand the context of a word, based on its entire surroundings [27]. BERT is trained on a vast dataset of text and code, making it highly successful at performing a variety of NLP tasks, including text understanding, question answering, language translation, and summarization. Figure 2 shows the BERT transformer model.

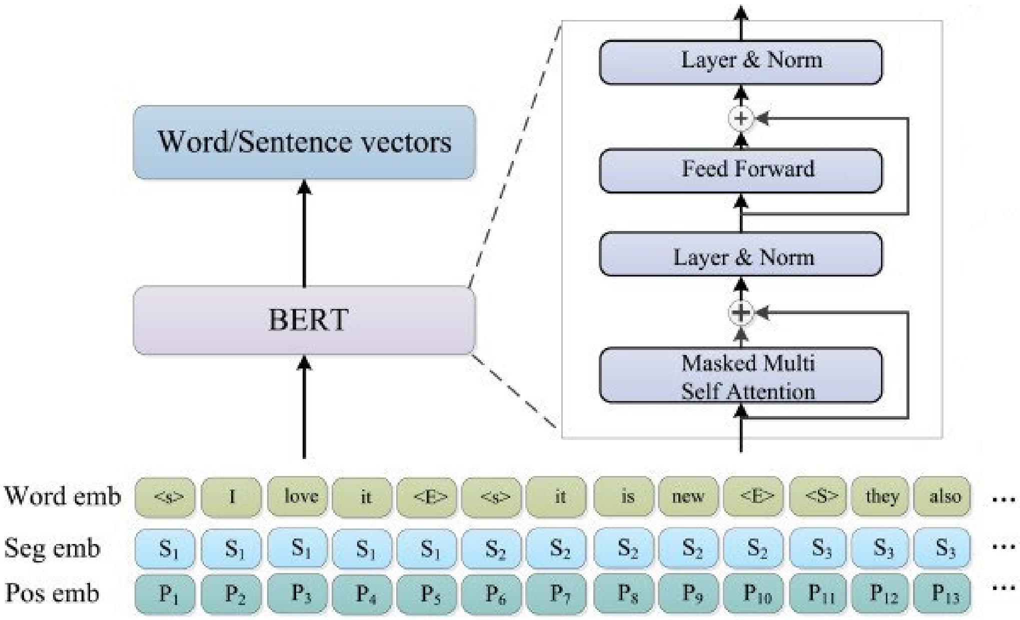

Figure 2. The BERT Transformer Model [28]

The main characteristic of BERT is that it is a bi-directional encoder. This technique is used to create representations of each word, taking into account both the left and right context of the text. This allows BERT to better understand the overall meaning of the text. Additionally, it can achieve robust and generalizable performance by training on vast datasets.

2.5.2. Long Short-Term Memory (LSTM) Deep Learning Algorithm

LSTM networks are a specialized form of RNN that enhances performance on both current and future inputs by leveraging past information. The network consists of hidden states and feedback loops. The network can work on sequences and store past information in confidence thanks to the loop structure [29]. The LSTM cell not only processes the current entries but also processes the historical data it keeps in its memory. The LSTM cell contains gate units that enable it to remember past states and determine whether to use these states [30]. Figure 3 shows the LSTM block:

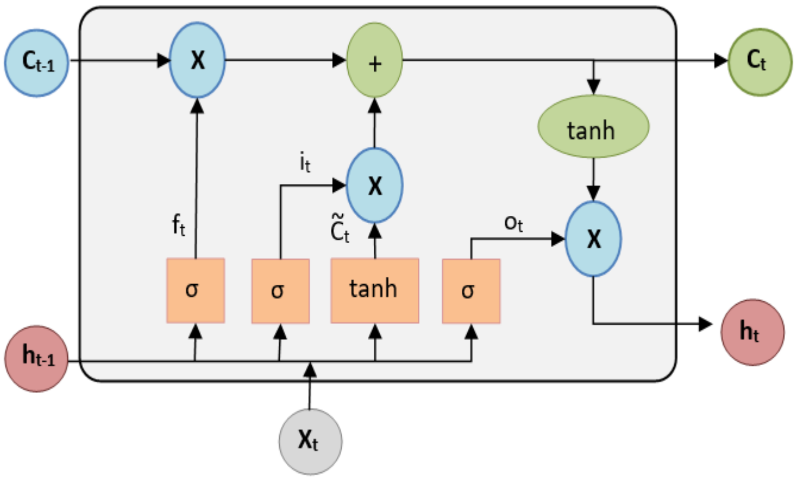

Figure 3. The LSTM block [30]

The symbols for input, forget, and output gates are represented by it, ft, and ot. W represents the recurrent connection, which is between the current hidden layer and the previously hidden layer. $\tilde{C}$ represents a candidate hidden unit, and U represents the weight matrix. C represents the internal memory of the unit. This internal memory is the combination of the input gate, multiplied by the newly calculated hidden unit, and the previous memory, multiplied by the forget gate.

2.5.3. Bi-directional LSTM (B-LSTM) Deep Learning Algorithm

The LSTM deep network only allows forward recall of information. This ensures that the data is processed only in one direction. To overcome this problem, a B-LSTM network has been proposed. The B-LSTM deep network can learn the knowledge of the future backward, whereas the LSTM deep network is limited to understanding the information of a one-way sequence [31]. Figure 4 shows the B-LSTM block:

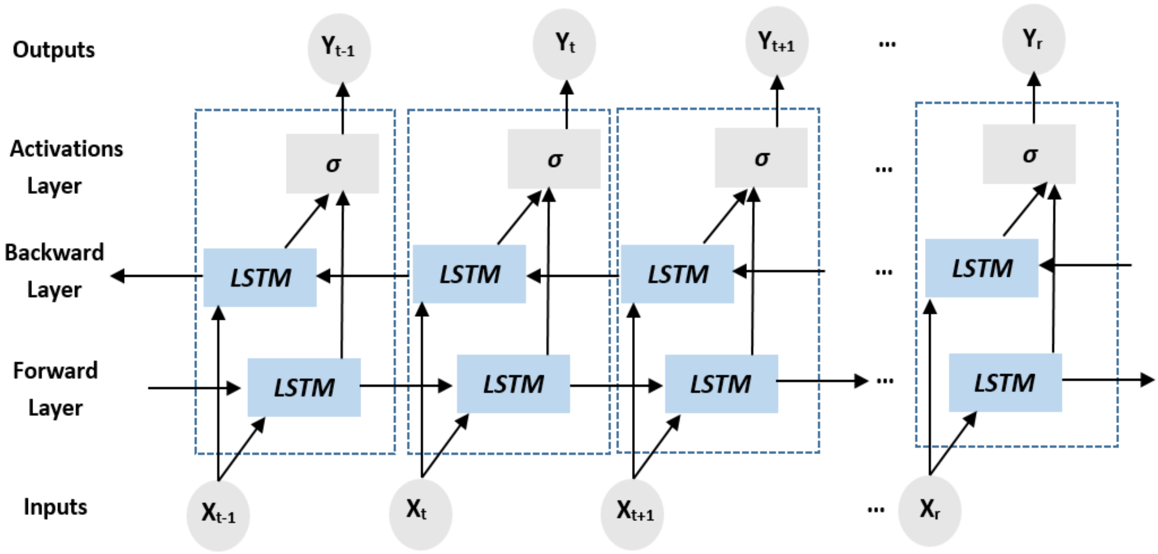

Figure 4. The B-LSTM block [31]

The B-LSTM deep learning network integrates the data gathered from operations carried out in both directions and assesses forward and backward computations concurrently, as two-way networks are more efficient than one-way networks and provide an advantage for data processing [31-32].

2.5.4. Convolutional Neural Network (CNN) Deep Learning Algorithm

The CNN algorithm has been successfully applied for NLP tasks in recent years, although it is a model traditionally used in image processing. The algorithm is a powerful technique to automatically separate text documents into predefined categories based on their content [33]. This approach delivers accurate and effective classification performance by harnessing the CNN algorithm's capacity to discern intricate relationships within text data. However, it requires large amounts of data, is complex to train, and computationally costly. Figure 5 shows the CNN block for text classification.

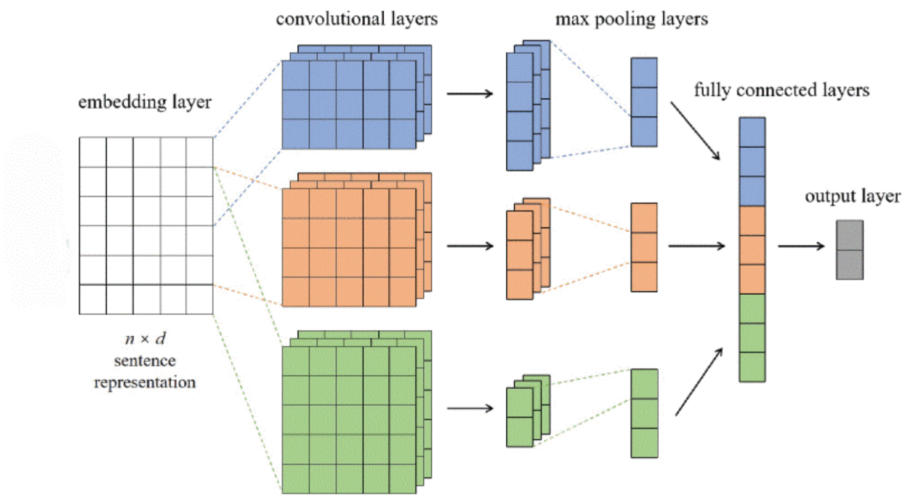

Figure 5. The CNN block for text classification [34]

The CNN algorithm consists of different layers, depicted in Figure 5. The embedding layer initializes random or pre-trained word embedding weights after receiving integer-encoded texts as input. The convolution layer takes as input texts that are converted into sequences of word embedding using convolution filters and generates feature vectors by analyzing word embedding for each text. The convolution layer's variable-length feature vectors are fed into the pooling layer, which converts them into fixed-length vectors. Fixed-length feature vectors are sent into the softmax layer after first passing through a fully connected layer. The layer learns non-linear feature combinations, and the values for each label are the output of this fully connected layer [33,34].

2.6. Model Evaluation

The model's performance is related to the number of correctly labeled data and the number of incorrectly labeled data. The results of the model are expressed using the confusion matrix parameters, namely true positive (TP), false positive (FP), false negative (FN), and true negative (TN). The model evaluation criteria are calculated using the parameters of the confusion matrix, and are recall (RCL), precision (PRC), F1-score (F1), overall accuracy (ACC), and kappa statistic (kappa).

$$RCL = TP / (TP+FN) \tag{1}$$

$$PRC = TP / (TP + FP) \tag{2}$$

$$F1 = 2TP / (2TP + FP + FN) \tag{3}$$

$$ACC = \text{Correctly classified samples} / \text{Total number of samples} \tag{4}$$

Kappa is a frequently used statistical method that measures agreement between two observers. The closeness of these values to one confirms that the model does not achieve random performance. The mathematical equation of kappa is given below [35]:

$$Pr_a = (TP + TN) / (TP + FN + FP + TN)$$

$$P\,c_1 = [(TP+FN) / (TP+FP+FN+TN)] \times [(TP+FP) / (TP+FP+FN+TN)]$$

$$P\,c_2 = [(FP+TN) / (TP+FP+FN+TN)] \times [(FN+TN) / (TP+FP+FN+TN)]$$

$$Pr_e = P\,c_1 + P\,c_2$$

$$Kappa = (Pr_a - Pr_e) / (1- Pr_e) \tag{5}$$

3. Experimental Results and Discussion

In this study, the performances of transformer and deep learning models were compared for sentiment analysis of online consumer reviews on e-commerce sites. Transformers and deep learning models were applied to 10975 online consumer reviews collected from Amazon, one of the most popular e-commerce platforms. Figure 6 shows the word clouds of the dataset containing online consumer reviews. Word clouds, which represent the frequency of words within a text, are a good way to quickly understand the general topic or theme of a text. They can also be used to identify keywords and trends in text.

Figure 6. The word clouds for the consumer reviews

Transformer and deep learning models take consumer reviews as input data and perform a classification task to determine whether they are "positive", "negative", or "neutral". Confusion matrix parameters were calculated to evaluate the models. The confusion matrix parameters indicate how many samples with the correct label and how many with the incorrect label are classified by the algorithm. Figure 7 presents the confusion matrix of the transformer and deep learning models.

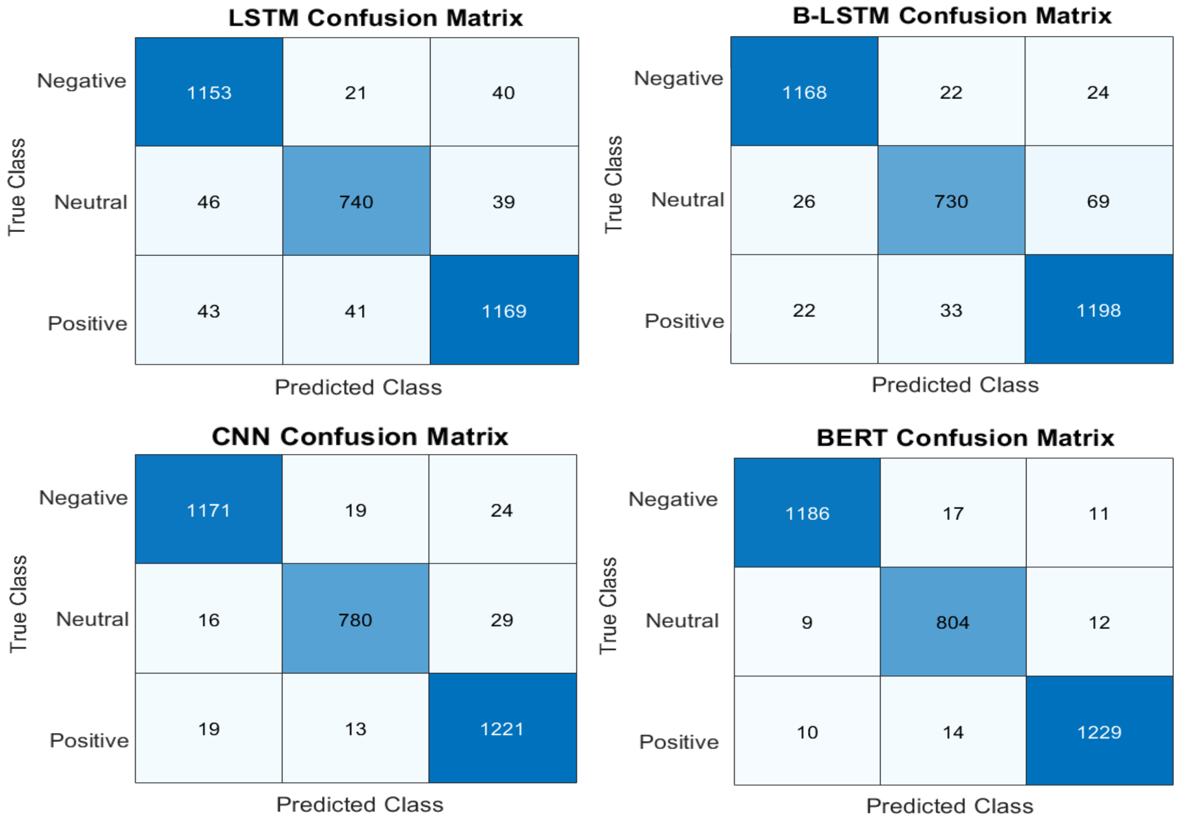

Figure 7. Confusion matrix of transformer and deep learning models

The results showed that the LSTM deep learning algorithm had 3062 correctly labeled data (TP+TN) and 230 incorrectly labeled data (FP+FN), while the B-LSTM deep learning algorithm had 3096 correctly labeled data and had 196 incorrectly labeled data. The CNN deep learning algorithm had 3172 correctly labeled data points and 120 incorrectly labeled data points; the BERT transformer algorithm had 3219 correctly labeled data points and 73 incorrectly labeled data points. The BERT transformer algorithm had the highest number of correctly labeled data. The BERT transformer algorithm was followed by CNN, B-LSTM, and LSTM algorithms. The performance of transformer and deep learning models is presented in Table 4.

Table 4. The performances of transformer and deep learning models

Algorithm	Label	RCL	PRC	F1	Kappa Statistic	Overall ACC
LSTM	**Negative**	0.928	0.949	0.938	0.893	93.01%
	Neutral	0.922	0.897	0.909		
	Positive	0.936	0.933	0.934		
B-LSTM	**Negative**	0.960	0.962	0.961	0.909	94.05%
	Neutral	0.929	0.884	0.906		
	Positive	0.928	0.956	0.941		
CNN	**Negative**	0.971	0.964	0.967	0.944	96.35%
	Neutral	0.960	0.945	0.953		
	Positive	0.958	0.974	0.966		
BERT	**Negative**	0.984	0.976	0.980	0.966	97.78%
	Neutral	0.962	0.974	0.968		
	Positive	0.982	0.981	0.981		

When Table 4 is examined, the RCL, PRC, F1, kappa, and overall ACC values of all algorithms are greater than 0.80. Additionally, the statistical value of kappa between 0.6 and 0.8 indicates a significant degree of agreement. The results showed that the BERT transformer model-based solution was more accurate and efficient than deep learning models. The BERT transformer model achieved the highest overall ACC performance. The BERT transformer model was followed by CNN (96.35%), B-LSTM (94.05%), and LSTM (93.01%). The performance results of the BERT transformer model were 97.78% overall ACC and 0.966 kappa. The BERT transformer model is a bi-directional encoder and can create a representation of each word by taking into account both the left and right context of the text. Therefore, it achieved a higher overall ACC than other models. The results of this study were discussed in relevant studies on text classification. The objectives, input data, the algorithms, and the overall ACC value of the relevant studies for text classification are given in Table 5.

Table 5. Comparative analysis of recent studies for text classification.

Researchers	Objectives	Input data	Algorithms	Results
Siering et al., (2018) [36]	Evaluation of airline passenger reviews	Passenger reviews taken from the airlinequality.com	SVM	ACC: 80.80%
Shihabeldeen (2019) [37]	Predicting exchange rates	Social media and news data	MLP	ACC: 60.26%
Afzaal et al., (2019) [38]	Evaluation of tourist opinions	Datasets containing tourist reviews on websites and social media (TripAdvisor, Expedia, Zomato, Facebook)	ECC + BR classifiers	ACC: 90.00%
Lucini et al., (2020) [39]	Airline customer satisfaction analysis	Online passenger reviews for airlines taken from airlinequality.com	Regression analysis	ACC: 79.95%
Tsai et al., (2020) [40]	Analysis of online hotel customer reviews	Customer reviews from the TripAdvisor.com website	RF	ACC: 81.20%
Guerreiro & Rita (2020) [41]	Predicting explicit recommendations in online reviews	Customer reviews from different restaurants were selected from the Academic Yelp Dataset.	RF	ACC: 66.83%
Pratama et al. (2022) [23]	Classify e-commerce beauty product reviews	Dataset obtained from beauty product review data on Amazon	SVM	ACC: 80.06%
Zhou (2023) [42]	Sentiment analysis of consumer reviews in the social media environment.	E-commerce products (compassing electronic products, food, clothing, books, and other categories) reviews in the social media environment.	BERT+ B-LSTM	RCL: 90.32% PRC: 92.64% F1: 91.46%
The proposed model	Sentiment analysis of consumer reviews on an e-commerce platform	Consumer reviews of technological products, taken from the electronics category	BERT	ACC: 97.78%

MLP: Multilayer perceptron; ECC: Ensemble classifier chains; BR: Binary relevance.

When the input data used in other studies in the literature on the classification of texts are examined, user reviews from social media [37, 38] or web pages [23, 39-40] are generally used. Similarly, consumer reviews on the most well-known e-commerce platform were used as data in this study. Algorithm selection is critical to the performance of models. In this study, a comparative analysis of the BERT transformer model, LSTM, B-LSTM, and CNN deep learning algorithms was performed. The BERT transformer model, which was compared with deep learning models, achieved the highest performance. When studies on text classification are examined, the success of other text classification studies is generally lower, ranging from 60 to 90% [23, 36-41]. In the results, a very high classification success was achieved, with an overall ACC of 97.78% using the BERT transformer model. The BERT transformer model can better capture the semantic relationships and context of words in the text. Therefore, the BERT can better understand the general meaning of the text and the connections between words. Additionally, BERT can be adapted for different types of texts and tasks. It can be applied in various areas, including sentiment analysis, text summarization, question answering, and natural language generation. It is expected that BERT will be utilized in numerous new applications and research in the future, continuing to make noteworthy contributions.

4. Conclusion

In conclusion, understanding and analyzing consumer thoughts and emotions about products is crucial for manufacturers. In the online environment, extracting insights from a large amount of data requires thorough data analysis. Performing the analysis, which was previously done manually, by machines, both reduces the company's cost and provides an unbiased analysis. Sentiment analysis extracts information from data, automates the process of utilizing that information, and minimizes workforce requirements. Reviews in unstructured text format were transformed into a structured form during the data preprocessing stage, and the performance of the transformer and deep learning models was compared on this structured data. When the overall ACC values of these algorithms were examined, the highest overall ACC belonged to the BERT transformer model algorithm. Based on the model evaluation criteria of this algorithm, the results were 96.6% kappa and 97.78% overall ACC. The closeness of these values to one confirms that the model does not yield random results. Considering that approximately 95% of the data encountered in daily life is unstructured [13], the automatic analysis of data containing text files is crucial. As the number of reviews on the products increases, it becomes increasingly complex to keep track of this data over time. For this reason, there is a need for systems that can automatically analyze reviews in online environments, which both users and company owners can utilize. These systems will benefit both the producer and the consumer by increasing the efficiency of automatic review evaluation, which can process millions of reviews without the need to read them individually, thereby facilitating their use and report preparation. It also means increased competition in the e-commerce market, providing personalized services to customers, and ultimately enhancing customer satisfaction.

In future studies, it may be recommended to expand the sample by increasing the number of documents and comparing the performance of transformer models with that of traditional feature extraction techniques, such as bag-of-words, n-gram, and TF-IDF. Moreover, the proposed model is designed to distinguish three types of emotional tones: positive, negative, and neutral. However, e-commerce reviews may exhibit a broader range of emotional tone types, for example, admiration, disappointment, anger, or surprise. Therefore, in future research, it is essential to adapt the proposed model to encompass a broader range of emotional tones. This will enable the model to perform a more detailed and accurate analysis of emotional tone. Therefore, in future research, it is essential to adapt the proposed model to encompass a broader range of emotional tones, which will enable the model to perform a more detailed analysis.

References

[1] A. Andonov, G. P. Dimitrov, and V. Totev, "Impact of e-commerce on business performance, " *TEM Journal*, vol. 10, no. 4, pp. 1558-1564, 2021.

[2] Ö. Alkan and Y. C. Bayhan, "COVID-19 pandemic and use of e-commerce: generational differences," *Journal of Economics and Administrative Sciences*, vol. 24, no. 2, pp. 170-185, 2023, doi: 10.37880/cumuiibf.1179003.

[3] TUBISAD, "The transformative power of economy: e-commerce impact analysis," Jun. 2021. [Online]. Available: https://www.tubisad.org.tr/en/images/pdf/the-transformative-power.pdf. [Accessed: Feb. 5, 2025].

[4] J. Mosteller and C. Mathwick, "Consumers helping consumers; the role of psychological need fulfillment in an online reviewer community," in Let’s get engaged! crossing the threshold of marketing’s engagement era, M. Obal, N. Krey, and C. Bushardt, Eds. Cham, Switzerland: Springer, 2016, pp. 106, doi: 10.1007/978-3-319-11815-4_106.

[5] S. C. Nune and I. Kozhıkode, "Characterizing and predicting early reviewers for effective product marketing on ecommerce websites," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 14, pp. 2870-2877, 2021.

[6] M. Yang, Y. Ren, and G. Adomavicius, "Understanding user-generated content and customer engagement on Facebook business pages," *Information Systems Research*, vol. 30, no. 3, pp. 839-855, 2019, doi: 10.1287/isre.2019.0834.

[7] S. L. Shengli and L. Fan, "The interaction effects of online reviews and free samples on consumers’ downloads: An empirical analysis," *Information Processing and Management*, vol. 56, pp. 1-12, 2019, doi: 10.1016/j.ipm.2019.102071.

[8] D. Gavilan, M. Avello, and G. Martinez-Navarro, "The influence of online ratings and reviews on hotel booking consideration," *Tourism Management*, vol. 66, pp. 53-61, 2018, doi: 10.1016/j.tourman.2017.10.018.

[9] P.Y. Chen, S. Dhanasobhon, and M. D. Smith, "All reviews are not created equal: The disaggregate impact of reviews and reviewers at Amazon.com," *SSRN Electronic Journal*, 2008, doi: 10.2139/ssrn.918083.

[10] R. Thakur, "Customer engagement and online reviews," *Journal of Retailing and Consumer Services*, vol. 41, pp. 48-59, 2018, doi: 10.1016/j.jretconser.2017.11.002.

[11] J. H. Huang and Y. F. Chen, "Herding in online product choice," *Psychology and Marketing*, vol. 23, no. 5, pp. 413-428, 2006, doi: 10.1002/mar.20119.

[12] A. Batta, A. K. Kar, and S. Satpathy, "Cross-platform analysis of seller performance and churn for ecommerce using artificial intelligence," *Journal of Global Information Management (JGIM)*, vol. 31, no. 1, pp. 1-21, 2023.

[13] K. Adnan, R. Akbar, and K. S. Wang, "Information extraction from multifaceted unstructured big data," *International*

Journal of Recent Technology and Engineering (IJRTE), vol. 8, pp. 1398-1404, 2019, doi: 10.35940/ijrte.B1074.0882S819.

[14] T. Jo, Text mining Concepts, Implementation, and Big Data Challenge, New York: Springer, 2019.

[15] H. Göker, H. I.Bülbül, and E. Irmak, "The estimation of students' academic success by data mining methods." *In 12th International Conference on Machine Learning and Applications*, vol. 2, pp. 535-539, 2013, doi: 10.1109/ICMLA.2013.173.

[16] W. Hong, C. Zheng, L. Wu, and X. Pu, "Analyzing the relationship between consumer satisfaction and fresh e-commerce logistics service using text mining techniques," *Sustainability*, vol. 11, no. 13, pp. 1-16, 2019, doi: 10.3390/su11133570.

[17] Z. Shen, "Mining sustainable fashion e-commerce: social media texts and consumer behaviors," *Electronic Commerce Research*, pp. 1-23, 2021, doi: 10.1007/s10660-021-09498-5.

[18] E. Sezgen, K. J. Mason, and R. Mayer, "Voice of airline passenger: A text mining approach to understand customer satisfaction," *Journal of Air Transport Management*, vol. 77, pp. 65-74, 2019, doi: 10.1016/j.jairtraman.2019.04.001.

[19] P. Sobkowicz, M. Kaschesky, and G. Bouchard, "Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web," *Government Information Quarterly*, vol. 29, no. 4, pp. 470-479, 2012, doi: 10.1016/j.giq.2012.06.005.

[20] M. O. Aftab, U. Ahmad, S. Khalid, A. Saud, A. Hassan, and M. S. Farooq, "Sentiment analysis of customer for ecommerce by applying AI," *In 2021 International Conference on Innovative Computing (ICIC)*, Nov. 2021, pp. 1-7, doi: 10.1109/ICIC53490.2021.9693026.

[21] J. Jabbar, I. Urooj, W. JunSheng, and N. Azeem, "Real-time sentiment analysis on E-commerce application," *In 16th International Conference on Networking, Sensing and Control (ICNSC)*, May 2019, pp. 391-396, doi: 10.1109/ICNSC.2019.8743331.

[22] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, "A comparative study of support vector machine and Naive Bayes classifier for sentiment analysis on Amazon product reviews," *In International Conference on Contemporary Computing and Applications (IC3A)*, Feb. 2020, pp. 217-220, doi: 10.1109/IC3A48958.2020.233300.

[23] M. R. Pratama, F. A. G. Soerawinata, R. R. Zhafari, and H. N. Irmanda, "Sentiment analysis of beauty product e-commerce using support vector machine method," *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, vol. 6, no. 2, pp. 269-274, 2022, doi: 10.29207/resti.v6i2.3876.

[24] S. Paul and S. Saha, "CyberBERT: BERT for cyberbullying identification: BERT for cyberbullying identification," *Multimedia Systems*, vol. 28, no. 6, pp. 1897-1904, 2022, doi: 10.1007/s00530-020-00710-4.

[25] M. Panda, "Developing an efficient text pre-processing method with sparse generative Naive Bayes for text mining," *International Journal of Modern Education and Computer Science*, vol. 11, no. 9, pp. 11-19, 2018, doi: 10.5815/ijmecs.2018.09.02.

[26] S. Vijayarani and R. Janani, "Text mining: Open source tokenization tools-an analysis," *Advanced Computational Intelligence: An International Journal (ACII)*, vol. 3, no. 1, pp. 37-47, 2016, doi: 10.5121/acii.2016.3104.

[27] S. Alav and K.S. Benli, "Turkish stance detection on social media using bert models: A case study of stray animals law," *Sakarya University Journal of Computer and Information Sciences*, vol. 8, no. 1, pp. 76-88, 2025, doi: 10.35377/saucis...1564138.

[28] J. Su, Q. Dai, F. Guerin, and M. Zhou, "BERT-hLSTMs: BERT and hierarchical LSTMs for visual storytelling," *Computer Speech & Language*, vol. 67, no. 101169, pp. 1-14, 2021, doi: 10.1016/j.csl.2020.101169.

[29] M. Bulut, "Improving deep learning forecasting model based on LSTM for Türkiye's hydro-electricity generation," *Sakarya University Journal of Computer and Information Sciences*, vol. 7, no. 3, pp. 325-337, 2024, doi: 10.35377/saucis...1503018.

[30] S. Varsamopoulos, K. Bertels, and C. G. Almudever, "Designing neural network based decoders for surface codes," *IEEE Transactions on Quantum Engineering*, vol. 3, pp. 1-13, 2022, doi: 10.1109/TQE.2022.3195723.

[31] H. Göker and A. Said, " Spectral analysis and Bi-LSTM deep network-based approach in detection of mild cognitive impairment from electroencephalography signals," *Cognitive Neurodynamics*, vol. 18, no.2, pp. 597-614, 2024, doi: 10.1007/s11571-023-10010-y.

[32] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing,* vol. 337, pp. 325-338, 2019, doi: 10.1016/j.neucom.2019.01.078.

[33] S. Soni, S. S. Chouhan, and S. S. Rathore, "TextConvoNet: A convolutional neural network based architecture for text classification," *Applied Intelligence*, vol. 53, no. 11, pp. 14249-14268, 2023, doi: 10.1007/s10489-022-04221-9.

[34] X. Guo, Q. Liu, Y. Hu, and H. Liu, "MDCNN: Multi-teacher distillation-based CNN for news text classification," *IEEE Access*, vol. 13, pp. 56631-56641, 2025, doi: 10.1109/ACCESS.2025.3555224.

[35] A. B. Cantor, "Sample-size calculations for Cohen's kappa," Psychological Methods, vol. 1, no. 2, pp. 150-151, 1996, doi: 10.1037/1082-989X.1.2.150.

[36] M. Siering, A. V. Deokar, and C. Janze, "Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews," *Decision Support Systems*, vol. 107, pp. 52-63, 2018, doi: 10.1016/j.dss.2018.01.002.

[37] H. Y. A. Shihabeldeen, "Using text mining to predicate exchange rates with sentiment indicators," *Journal of Business Theory and Practice*, vol. 7, no. 2, pp. 60-75, 2019, doi: 10.22158/jbtp.v7n2p60.

[38] M. Afzaal, M. Usman, A. C. Fong, and S. Fong, "Multiaspect-based opinion classification model for tourist reviews," *Expert Systems*, vol. 36, no. 2, e12371, 2019, doi: 10.1111/exsy.12371.

[39] F. R. Lucini, L. M. Tonetto, F. S. Fogliatto, and M. J. Anzanello, "Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews," *Journal of Air Transport Management*, vol. 83, no. 101760, pp. 1-12, 2020, doi: 10.1016/j.jairtraman.2019.101760.

[40] C. F. Tsai, K. Chen, Y. H. Hu, and W. K. Chen, "Improving text summarization of online hotel reviews with review helpfulness and sentiment," *Tourism Management*, vol. 80, no. 104122, pp. 1-13, 2020, doi: 10.1016/j.tourman.2020.104122.

[41] J. Guerreiro and P. Rita, "How to predict explicit recommendations in online reviews using text mining and sentiment analysis," *Journal of Hospitality and Tourism Management*, vol. 43, pp. 269-272, 2020, doi: 10.1016/j.jhtm.2019.07.001.

[42] X. Zhou, "Sentiment analysis of the consumer review text based on BERT-BiLSTM in a social media environment," *International Journal of Information Technologies and Systems Approach (IJITSA)*, vol. 16, no. 2, pp. 1-16, 2023.

Article Information Form

Conflict of Interest Notice

We declare no conflict of interest.

Ethical Approval

This paper does not include any studies with human or animal subjects.

Availability of data and material

The dataset used is available from the author upon request.

Artificial Intelligence Statement

No artificial intelligence tools were used while writing this article.

Plagiarism Statement

This article has been scanned by iThenticate™.