

Article

Deep Learning Analysis for Reviews in Arabic E-Commerce Sites to Detect Consumer Behavior towards Sustainability

Nada Ali Hakami ¹ and Hanan A. Hosni Mahmoud ^{2,*} 

¹ Department of Computer Science, College of Computer Science and Information Technology, Jazan University, Jazan P.O. Box 45142, Saudi Arabia

² Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh P.O. Box 84428, Saudi Arabia

* Correspondence: hahosni@pnu.edu.sa

Abstract: Recently, online e-commerce has developed a major method for customers to buy various merchandise. Deep learning analysis of online customer reviews can detect consumer behavior towards sustainability. Artificial intelligence can obtain insights from product reviews to design sustainable products. A key challenge is that many sustainable products do not seem to fulfill consumers' expectations due to the gap between consumers' expectations and their knowledge of sustainable products. This article proposes a new deep learning model using dataset analysis and a neural computing dual attention model (DL-DA). The DL-DA model employs lexical analysis and deep learning methodology. The lexical analysis can detect lexical features in the customer reviews that emphasize sustainability. Then, the deep learning model extracts the main lexical and context features from the customer reviews. The deep learning model can predict customers' repurchase habits concerning products that favor sustainability. This research collected data by crawling Arabic e-commerce websites for training and testing. The size of the collected dataset is about 323,150 customer reviews. The experimental results depict that the proposed model can efficiently enhance the accuracy of text lexical analysis. The proposed model achieves accuracy of 96.5% with an F1-score of 96.1%. We also compared the proposed model with state of the art models, where our model enhances both accuracy and sensitivity metrics by more than 5%.

Keywords: deep learning; lexical analysis; reviews; Arabic e-commerce; sustainability



Citation: Ali Hakami, N.; Hosni Mahmoud, H.A. Deep Learning Analysis for Reviews in Arabic E-Commerce Sites to Detect Consumer Behavior towards Sustainability. *Sustainability* **2022**, *14*, 12860. <https://doi.org/10.3390/su141912860>

Academic Editors: Mehrbakhsh Nilashi and Rabab Ali Abumalloh

Received: 26 August 2022

Accepted: 2 October 2022

Published: 9 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the increase in the number of e-commerce websites, more customers intend to use e-commerce platforms. Related to physical shopping in real stores, customers have the ability to shop at any time period with less time and effort and the merchandise on e-commerce sites are diverse. Customers can purchase the desired merchandise without leaving home [1]. Nowadays, green practices are impacting consumer decisions for selecting merchandise even though there are a lot of challenges in the merchandise retailed on online forums, such as the discrepancy between the sustainability and the real merchandise, poor features, and deficient customer service [2]. Consequently, it is important to utilize lexical analysis of the merchandise evaluation from e-commerce forums.

Lexical analysis of customer assessment provides a beneficial reference for new customers who favor the sustainability features of the products, and it can help e-commerce forums to develop quality and customer satisfaction. Lexical analysis employs text analysis and views mining techniques. Lexical analysis automatically investigates individual comments with customers' expressive usage of color and generates the customers' emotional trends [3].

The main models of data mining are rule-based or deep learning, or involve a joint method of both. Rule-based models include lexical analysis models. Deep learning models

comprise conventional deep learning techniques, such as random fields and neural networks. Deep learning models are used in many fields, such as image classification [3–6], object recognition [6], object and vehicle tracking [7–9], optimization [9], fire detection [8–12], and medical fields [13]. Recently, many articles have utilized deep learning models with lexical analysis by building lexical datasets, which attained better performance [14–16].

The central phase of lexical analysis methodology is the construction of lexical datasets. The lexical dataset is built by choosing suitable lexical words, degree nouns, and negative verses. Word intensity and polarity are used to build the lexical dataset. The datasets are linked with the lexical words in the lexical dataset. The linked lexical words are scored and added to compute the lexical score of the input, thus deciding the lexical polarity of the input.

The rest of the article is divided as follows: Section 2 describes the background. Section 3 designates our proposed DL-DA model. Section 4 displays the experimental settings and the results of our testing of the DL-DA model. Section 5 presents the discussion of the experimental results. Section 6 concludes the paper.

2. Background

There are several models that utilize a feature map vector. Models such as WrDVec, FstTxt are examples of such models. Customary deep learning models manually extract the emotional attributes of the data from the input to build the text feature map and then employ the conventional deep learning techniques to classify the lexical features [15–17]. This models usually needs manual involvement to attain the lexical class of the input text. Conventional deep learning models utilize naïve Bayes modeling, support-vector machines (SVM), and random forest models [17–19].

Lexical analysis is usually used in natural language processing systems [20]. World-wide, people use e-commerce and usually add reviews to express their satisfaction for their transactions [21]. The remarkable development of lexical analysis techniques in the previous years has allowed lexical analysis models to help customers to select suitable e-commerce sites. An enhanced website usually employs rate prediction. However, customer review perception in busy surroundings face many problems [22–25].

Many studies have tackled the improvement of recommender systems that preprocess customer review speech and utilize this data to feed to the lexical analysis models and overpower customer dissatisfaction [26,27]. Akoglu et al. [26] proposed false information removal techniques utilizing solo and multi-deep learning platforms. The advantage of the method proposed in [27] is that they used a multiple deep learning false information removal to filter the true rating from false or gibberish ratings. The authors of [28] confirmed the advantages of the multi-deep learning models in restaurant settings ratings in five-star rating analyses with moderate accuracy but faster learning curve. The authors of [29] validated a simulated deep learning beam model with 5600 users for 9 weeks. Qualitative reports from the lexical analysis cases designated that the beam model created a higher recommendation quality and was favored to other e-commerce handling models in commercial environments. The authors of [30] evaluated a mixture of false information removal models (i.e., a noise removal model combined with directional attention algorithms implemented on the CP810 processor) on 16 experimental recommenders. The outcomes validated that the proposed model improved lexical analysis prediction accuracy in commercial environments. The model proposed in [31] used a post-beam dual attention model and presented an enhancement in rating prediction performances over previous beam models on lexical analysis. A summary of the state of the art models in the literature using intelligent models is depicted in Table 1.

Table 1. Summary of different machine and deep learning models that detect consumer behavior in different datasets.

Reference	Domain	Dataset	Model
[28]	Restaurant ratings in five stars rating analyses towards organic products	Otlab rating data	Multi-model deep learning models
[29]	Qualitative reports from the lexical analysis to create higher recommendation for green energy in Hotels	Dataset from 5600 users for 9 weeks	Simulated deep learning beam model
[30]	Evaluation of a mixture of false information removal models towards more sustainability (i.e., a noise removal model)	16 experimental recommenders	Combined directional attention algorithms
[31]	An enhancement in rating prediction performances over previous beam models on lexical analysis to extract lexicons related to green products	Social media posts	Post-beam dual attention model a
[32]	Consumer behavior anomaly detection in green energy usage	Dashboard	Statistical model
[33]	Consumer behavior and their willingness to spend more to attain green products detection using support-vector machines	Consumer rating data from real social media posts	Ant colony on support-vector machine
[34]	Detecting consumer behavior towards organic foods using machine learning	Social media posts and comments from different distances	Neural network trained with histogram of features occurrence
[35]	Detection of consumer behavior towards hotels with non-disposable products using IOT	Consumer rating data (star scores)	Deep learning and long-term/short-term memory
[36]	A deep learning model for consumer behavior towards sustainable products prediction	Real consumer rating data	Deep learning recurrent CNN
[37]	Consumer behavior towards the anomaly of spending more on green energy detection using 3D CNNs	Real-time consumer rating videos	3D CNN
[38]	Consumer behavior towards sustainability detection model	CAD-CVIS dataset	Deep learning model YOLO
[39]	Sentiment analysis model utilizing semantic segmentation	Arabic posts from book-ing.com crawled dataset	Autoencoder

Based on the published research, challenges and gaps can be defined that have to be faced in future research in the Arabic language context in consumer reviews and product sustainability [40–42]. It was depicted that most studies are exploratory, suggesting descriptive outputs without intelligent analysis. Therefore, the proposed research adopted a deep learning methodology to extract parameters that can influence consumers' acceptance of sustainable products. The proposed model primarily reports an intelligent methodological gap.

With these goals in mind, the proposed methodology is directed by the following research questions:

- (1) How can deep learning models be established to detect sustainability favoring in consumer reviews?
- (2) How can an objective comparison of the deep learning models be identified?
- (3) How can negative reviews be analyzed to help companies understand the sense of balance between sustainability and other factors?

The contributions of this research are as follows:

- The proposed research investigates the artificial intelligence techniques to advance the sustainable design of products by extracting positive lexical words related to green products.
- A novel lexical analysis methodology using a lexical technique, lexicon vectors, and a deep learning and attention model is proposed.
- Extensive experiments using 323,150 customer reviews from real Arabic e-commerce sites were performed. The experiments extracted positivity reviews from customers about green products.
- The impact of correlated parameters, such as the size of the built dictionary, the size of the input review, and the number of the epochs of the model, was investigated.

3. Methods and the Proposed DL-DA Model

The proposed research goal developed an intelligent model, incorporating automated lexical analysis techniques of online customer reviews. Lexical analysis, the deep learning (DL) model, and the attention model are employed to enhance the performance of prediction. The DL-DA model extracts lexical features from consumer reviews (concerning the satisfaction and the willingness to repurchase green and sustainable products). Then, DL is utilized to learn the main lexical and context attributes from the reviews and employ the attention model as a scoring scheme. The last stage classifies the scored lexical attributes. The proposed model has five layers: an embedded input layer, a convolutional deep layer, a Maxpooling layer, a dual attention layer, and fully connected layer followed by the Softmax classifier. The model is depicted in Figure 1.

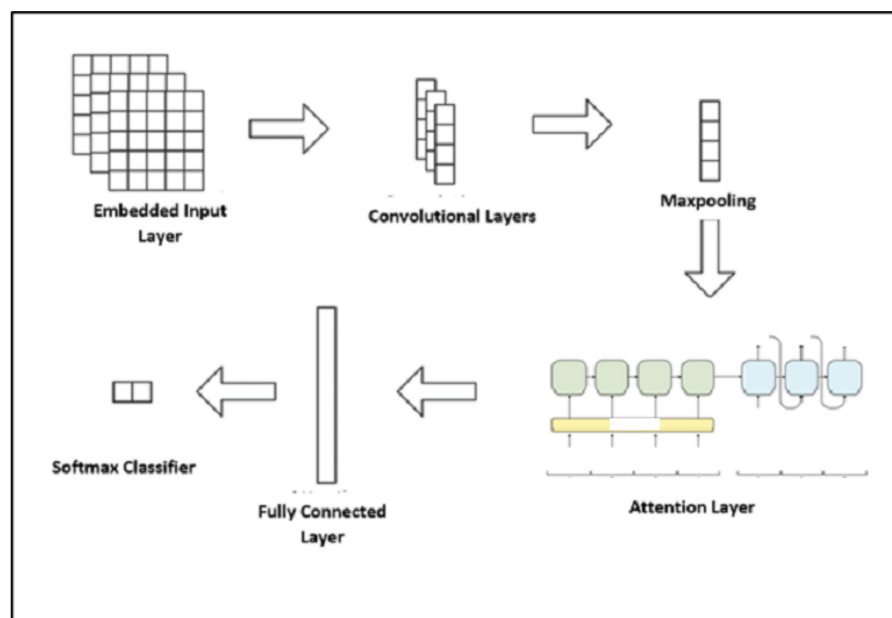


Figure 1. The stages of the DL-DA model.

Description of the DL-DA Model

The input review phrase is defined as $R = \{L_1, L_2, \dots, L_i, \dots, L_n\}$, where L_i denotes a lexicon in R . The objective of our proposed model is to classify the lexical polarity P_L of the review R .

A. Constructing a Lexical score

The function of the lexical scoring system is to attach each word L_i , in R with a lexical score $Score_L$.

The most used Arabic open source lexica are: ArabNet [26], the lexica ontology [27] and the streamlined Arabic polarity lexica of Alexandria University [28]. The most utilized Arabic-English dictionary is WordNet [29]. Our lexical scoring is founded on the emotional terminology library of Alexandria University. We eliminate the lexical words that denote impartiality and neutral sexes and keep the lexical words that denote derogatory values, that is, keep words with small value polarities. Lexical words are separated into five classes based on their lexical odd intensity, explicitly 1 to 9, with their intensity defining the lexical score. Lexica with negative polarity have their value multiplied by -1 .

The definition of the lexical score of any word L_i , is depicted as follows:

$$Score(L_i) = \begin{cases} S_{L_i} & \text{if } L_i \in SD \\ 1 & \text{if } L_i \neg \in SD \end{cases} \quad (1)$$

where L_i denotes a lexicon, S_{L_i} denotes the score of the word L_i in the lexical dataset, and SD denotes the lexical dataset.

B. Embedded Input Layer

The input layer symbolizes the review phrase R as a scored vector. In natural language processing, words are denoted by digital values, such as the one-time encoding scheme. The one-time encoding scheme syndicates the words in the dataset to generate one vector. The size of the vector is computed as the count of words in the dataset. The dimension of a word is equal to 1, and the other dimensions are equal to 0. The one-time vector represents a word as independent and does not define the association between words. In addition, when the count of words in the dataset is large, the size of such a vector is enormous. The one-time encoding scheme is depicted in Table 2 for 3 words.

Table 2. Example of the one-time encoding.

The words in the review phrase				
	Id		Word	
	1		One	
	2		Two	
	3		Three	
	4		Two	
	5		One	
	6		Three	
	7		One	
One-Time coding				
	Id	One	Two	Three
	1	1	0	0
	2	0	1	0
	3	0	0	1
	4	0	1	0
	5	1	0	0
	6	0	0	1
	7	1	0	0

Many papers presented a word vector encoding scheme to tackle the problematic size of the one-time code [5]. The main indication is to denote words as a short-dimensional nonlinear condensed vector. Moreover, similar words are mapped to same points in

the vector. Systems that employed word vector representations are WrdVec [30] and W-BRT [31].

The W-BRT representation is a pre-trained model introduced by Google that defines bidirectional mapping and outperforms other word representations. W-BRT is depicted in Figure 2, where it makes an analogy between man and woman.

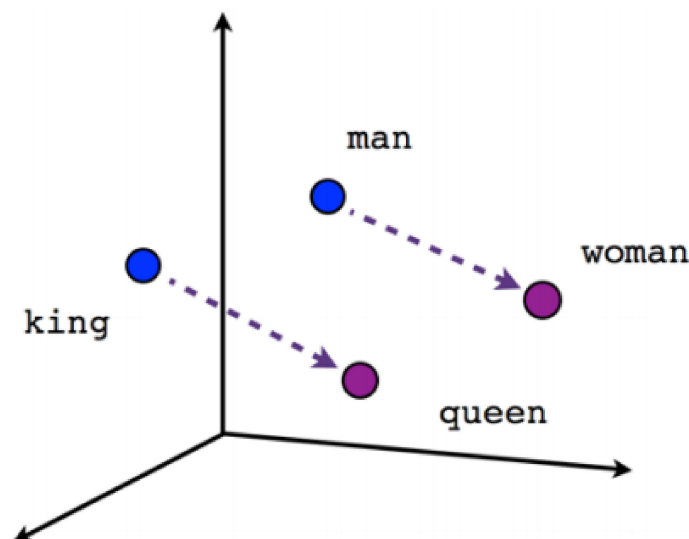


Figure 2. W-BRT model with an analogy between man and woman.

The proposed model employs the W-BRT model for deep learning of word vectors.

Each word $\{L_i\}$ in a review phrase R is transformed into a vector $\{V_i\}$ using a W-BRT representation. V_i is an m -dimensional vector, where $m = 700$. The rate of the word vector V'_i , utilizing the lexical score, is defined as follows:

$$\{V'_i = V_i \times \text{Score}(L_i)\} \quad (2)$$

The word vector score is used in the embedded layer's final output.

C. Neural Convolution Layer

The key objective of the convolution layer is to select the main local attributes of the input vector [32]. The vector representation of a word is always complete. Consequently, the convolution layer dimension takes the size of the word vector.

For the input vector $\{V' = [V'_1, V'_2, \dots, V'_i, V'_n]\}$, the convolution function V''_i is defined as follows:

$$V''_i = \text{ReLU}(S \cdot V[i + k + 1] + a) \quad (3)$$

where S denotes the score vector, a is the offset, and ReLU is the activation function.

When the convolution function is computed, the eigenvector E' is defined as follows:

$$E' = [V''_1, V''_2, \dots, V''_i, V''_{n-k+1}] \quad (4)$$

D. Maxpooling

The Maxpooling layer reduces the feature set attained from the convolution layer by selecting the main attributes. For lexical analysis, the pooling function selects a smaller number of words in the phrase.

The Maxpooling function P_i is:

$$P = [P_1, P_2, \dots, P_i, V''_{m-k+1}] \quad (5)$$

$$P_i = \max[V''_1, V''_2, \dots, V''_m, V''_{m+k-1}] \quad (6)$$

where m denotes the size of the vector V''_i and $\max$ is the Maxpooling operation.

E. Dual Attention Module

In each review phrase, each word has a unique effect on the lexical polarity of the entire phrase. Important words have a conclusive impact on the lexical polarity of the entire phrase. Less important words do not impact the phrase score. So, the attention model gives dissimilar scores to various words in a phrase.

For the hidden layer value h_i ($h_i = b_h S_h$), the previous layer produces the score b_h , which is calculated as follows:

$$b_h = \text{Softmax}(S_h, \tanh(S_h)) \quad (7)$$

while the polarity of the phrase O_j is defined as follows:

$$O_j = h_j^T S_r \tanh(\text{Phrase}_j) \quad (8)$$

where $\tanh$ is the hyperbolic tangent function with attained parameters.

F. Fully Connected (FC) Layer

The key operation of the FC layer is the classification of the input.

Its output is defined as the rank of the j^{th} word (R_j) in the phrase, calculated as follows:

$$R_j = S_{h_j} \quad (9)$$

The following formula computes the polarity of the phrase ($Polr_j$):

$$Polr_j = h_j^T S_h b_h \quad (10)$$

where $Polr$ defines the sigmoid function, S is the score matrix, and b is the score.

The FC layer associates the input feature with a score in the range $[0, 1]$. The smaller the score, the closer the lexical polarity of R to the negative trend. On the contrary, if the score is closer to 1, then the input phrase polarity is positive.

The pseudo code for the proposed algorithm is depicted as Algorithm 1:

Algorithm 1: the pseudo code for the proposed algorithm

DL-DA Model (input $R = \{L_1, L_2, \dots, L_i, \dots, L_n\}$, SD : the lexical dataset), (output: Score: value range from 0 to 1)

Start

For each consumer review R

Do

{

1: Construct the **Lexical score**:

1.1: Attach to each word L_i a lexical score Score_{L_i} .

$$\text{Score}(L_i) = \begin{cases} S_{L_i} & \text{if } L_i \in SD \\ 1 & \text{if } L_i \notin SD \end{cases}$$

2: Transform each word $\{L_i\}$ in a review phrase R into to a vector $\{V_i\}$ using a W-BRT representation.

$$\{V'_i = V_i \times \text{Score}(L_i)\}$$

3: **Compute the** convolution function V''_i for the input vector $\{V' = [V'_1, V'_2, \dots, V'_i, V'_n]\}$, $V''_i = \text{ReLU}(S \cdot V[i+k+1] + a)$.

4: Define the eigenvector E' .

$$E' = [V''_1, V''_2, \dots, V''_i, V''_{n-k+1}]$$

5: Perform Max pooling using the function P_i .

$$P_i = \max [V''_1, V''_2, \dots, V''_m, V''_{m+k-1}]$$

6: Compute the score b_h using the dual attention model.

7: Compute the polarity $Polr_j$ of the phrase O_j .

8: The FC layer associates the input feature with a score in the range $[0, 1]$. The smaller the score, the closer the lexical polarity of R to the negative trend. On the contrary, if the score is closer to 1, then the input phrase polarity is positive.

}

End

4. Experiments Results

We performed extensive experiments to test our model and evaluate its performance for the lexical analysis process. Table 3 depicts the lexicons extracted from online surveys answered by buyers of green products. The topics in Table 3 are green product, environmental aspects, durability, clean energy, and material recycling. Many lexicons are extracted from consumer surveys. Lexicons include green, plastic use, durable, recycling, and clean energy, among other lexicons. From these surveys, we selected important lexicons that matter to sustainability to extract from consumer reviews about products.

Table 3. The lexicons extracted from online surveys answered by buyers of green products.

Investigated Topics	Related Parameters From Experiments	Examples of Lexicon-Vector	Score
Green product	Liking green product. Positive lexicon with score higher than 3. Willing to repurchase. Recommend to others. Willing to pay for sustainable product.	Satisfied + green	Score 1 = Totally disagree 5 = Totally agree
Environmental aspects	The knowledge the consumer has about the impact of the green products and their impacts.	not + use plastic	Score 1 = No knowledge 5 = Totally know
Durability	Customer's feeling towards durable products impacting the tendency to repurchase.	well-made + durable	Score 1 = Totally disagree 5 = Totally agree
Clean energy	Customer's feeling towards using clean energy and the tendency to repurchase.	Important + produced + clean + energy?	Score 1 = Not important 5 = Very important
Material recycle	Facilities that inspire recycling behaviors.	Awareness + recycling + dangers + not recycling.	Score 1 = Not important 5 = Very important

A. Dataset

A list of sustainability-related lexicons to look for in customer reviews includes the following topics:

1. Green product;
2. Environmental aspects;
3. Durability;
4. Clean energy;
5. Material recycling.

The review dataset was collected from Egyptian e-commerce through the web crawler technique. The proposed model identifies the products by one of the five listed sustainable topics. The customer reviews in the dataset are partitioned into five ranks using a five-star metric. The algorithm partitions the ranks into two classes, with less than two stars denoted as negative and the class from three to five stars denoted as positive. The collected dataset has 323,150 reviews, half of the data is positive, and the other half is negative. The description of the dataset is defined in Table 3. The distribution of the number of stars is depicted in Table 4 and Figure 3.

Table 4. The features of the collected dataset.

Feature	Number	Feature	Number
Number of reviews	323,150	Median reviews per website	170
Number of websites	2858	The least number of reviews per websites	51
Avg. reviews per product	334	Number of users	40,189
The highest number reviews per website	6493	Average number of reviews per user	13.7
Average discarded reviews per website	12	Number of tokens	1,320,186

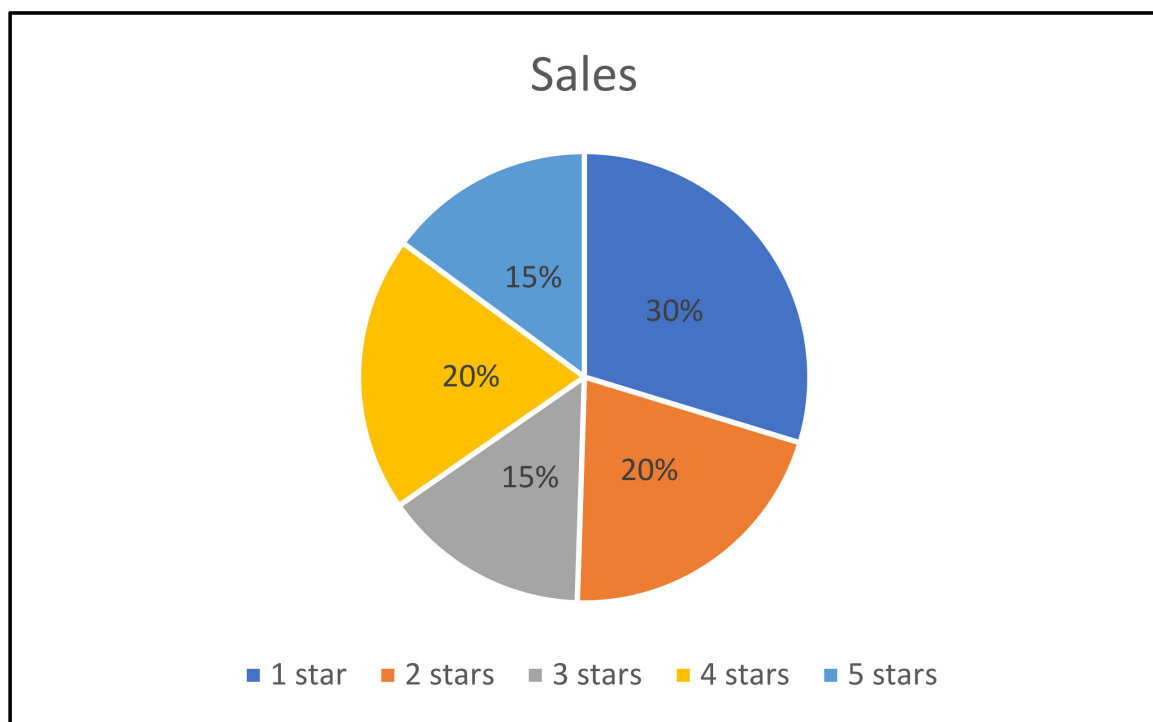
**Figure 3.** The distribution of the number of stars.

Table 5 describes the variables representing consumers answering surveys about buying sustainable products. The variables emphasize social impacts of the consumers' choice as well as the inability of the consumers to choose a sustainable product (maybe because it is more expensive). Additionally, variables representing the willingness of the consumer to favor environmental features are included. CIP represents the willingness to experience novel products and the acceptance of green products.

Table 5. Description of the variables of the crawled data from consumer reviews.

Variable	Code	Definition
Social impact	SI	The social pressure to carry out a behavior.
Observed behavioral control	OBC	Consumer's perception on the difficulty to perform the behavior of interest.
Sustainability concern	SC	Insights of environmental issues when making buying decisions.
Consumer innovation pursuing	CIP	The openness of consumers to novel products
Green consumption acceptance	GA	Consumption choice of an products with green impact motivation.

B. Experimental Metrics

The experiment performance metrics are accuracy, sensitivity, specificity, and F1-score, which are the most used metrics in performance studies.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (11)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (12)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (13)$$

$$F1 - \text{Score} = \frac{2 * TP}{2 * TP + FP + FN} \quad (14)$$

where

- TP is defined as the number of comments predicted as positive and they really are positive.
- FP is defined as the number of comments that are predicted as negative but they really are positive.
- TN is defined as the number of comments that are predicted as negative but they really are negative.
- FN is defined as the number of comments that are predicted as negative but they really are positive.

C. Preprocessing

Python phrase tokenization tool P-Tok is used to implement phrase segmentation of the comments.

The software eliminates stop words and non-Arabic characters (including Latin characters) from the phrase segmentation results.

The computed parameters are: the word count, word frequency, number of characters in the largest word, and number of words in each customer review. The mean customer review length is utilized as the length of the review.

D. Evaluation Results

The computed parameters captured from the dataset are depicted in Table 6.

Table 6. The computed parameters captured from the dataset.

Parameters of the Training Model	Value
The average length of the input customer review	120
The word vector dimension	112
Built dictionary size	14,000
First Convolution Layer	$64 \times 5 \times 5$
Hidden Layers	256
Second Convolution Layer	$32 \times 3 \times 3$
Hidden Layers	256
Max pooling	$9 \times 2 \times 2$
Dropout	0.37

The 12-fold cross-validation process partitions the dataset to 70% training, 15% validation, and 15% testing [30]. In the 12-fold validation technique, we partitioned the data into 12 portions, using 8 partitions for training, one at a time, and 4 partitions for testing and validation. The experiments compute the average of these 12 results as the performance result of the proposed model. The 8-2-2cross-validation model randomly partitions the data into three partitions. The model utilizes one of them as the training set, and the remaining partitions as the testing subsets. The mean value of the eight results is employed in the

performance results. Table 7 depicts the performance results of the DL-DA model using the 12-fold validation and 8-2-2cross-validation.

Table 7. The 12-K cross-validation results.

Method	Accuracy%	Sensitivity%	Specificity%	F1-Score%
12-fold	93.4	92.8	93.8	93.2
8-2-2	93.8	93.1	94.2	93.9

As the size of the review phrase in the dataset is variant, the size of the phrase is defined as the value for the model. We chose the extreme review size and the average review size in the dataset to perform our experiments. The results are depicted in Table 8. The utilization of the average review size as the fixed size of the input review causes loss in the context properties for long sentences. This loss can impact the accuracy of the model.

Table 8. The impact of the fixed size of the input review on the performance of the model.

Phrase Length	Accuracy%	Sensitivity%	Specificity%	F1-Score%
30	91.4	92.6	91.6	91.2
42	92.8	92.1	91.2	92.9
130	93.4	92.9	93.9	93.2
205	95.8	95.1	94.9	95.9

In the experimental results, it is noted that the count of words in the dictionary has an influence on the accuracy of the model. The experiments begin with the count of words in the dictionary starting from 20,000 words with word frequency abridged from those with the least occurrences. Experiments are repeated for every 10,000 lexicons. The results are depicted in Table 9 and Figure 4. The accuracy of the model is best when the count of words is approximately 60,000 words. As the count of words in the dictionary grows or shrinks, the accuracy of the model drops.

Table 9. The impact of the built dictionary size.

Dictionary Size	Accuracy%	Sensitivity%	Specificity%	F1-Score%
80,000	92.2	91.6	91.6	91.2
70,000	91.8	91.1	91.2	92.9
60,000	95.4	94.6	94.6	94.2
50,000	93.8	93.1	93.2	92.9
40,000	92.8	92.9	92.9	93.1
30,000	91.8	91.1	91.9	91.9
20,000	90.4	90.6	90.6	90.2

In the experiment setting, various epochs of the proposed model impact the accuracy of the model. As the count of the epochs of the model grows, the accuracy of the model increases in the beginning and then decreases. Table 10 and Figure 5 depict that the accuracy of the model grows with the count of epochs up to 40 epochs. When the count of the epochs of the model exceeds 40 times, the model progressively over-fits, causing the model's performance to decrease.

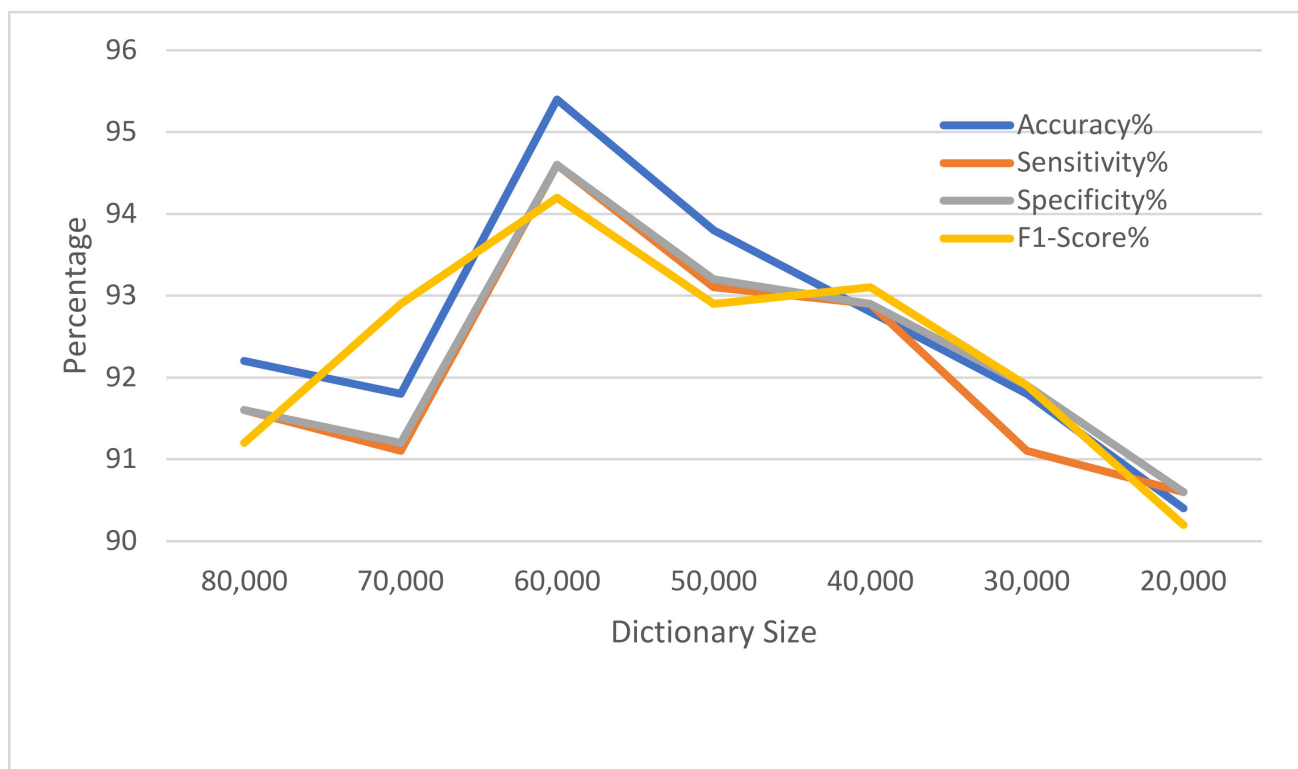


Figure 4. The impact of the constructed dictionary size on the model.

Table 10. The impact of the count of the epochs on the model.

Epoch	Accuracy%	Sensitivity%	Specificity%	F1-Score%
10	90.5	90.9	90.9	90.2
20	91.7	91.1	91.9	91.9
30	92.7	92.9	92.9	93.1
40	95.5	93.9	93.9	93.2
50	90.5	90.9	90.9	90.2
60	88.5	88.8	88.8	88.2
70	88.7	88.8	88.8	88.8
80	88.5	88.8	88.8	88.2
90	88.7	88.8	88.8	88.8

The generalization of the proposed model is impacted by the dropout value. Different dropout thresholds were selected for the experiments. The proposed model reaches the highest generalization aspect when the dropout is between 0.4 and 0.5. The dropout results are depicted in Table 11 and Figure 6.

Table 11. The impact of the dropout on the performance of the proposed model.

Dropout Value	Accuracy%	Sensitivity%	Specificity%	F1-Score%
0.25	93.3	93.8	93.8	93.2
0.35	92.7	92.1	92.2	92.9
0.45	95.1	95.5	95.5	95.1
0.55	91.7	91.1	91.9	91.9
0.75	90.3	90.8	90.8	90.2

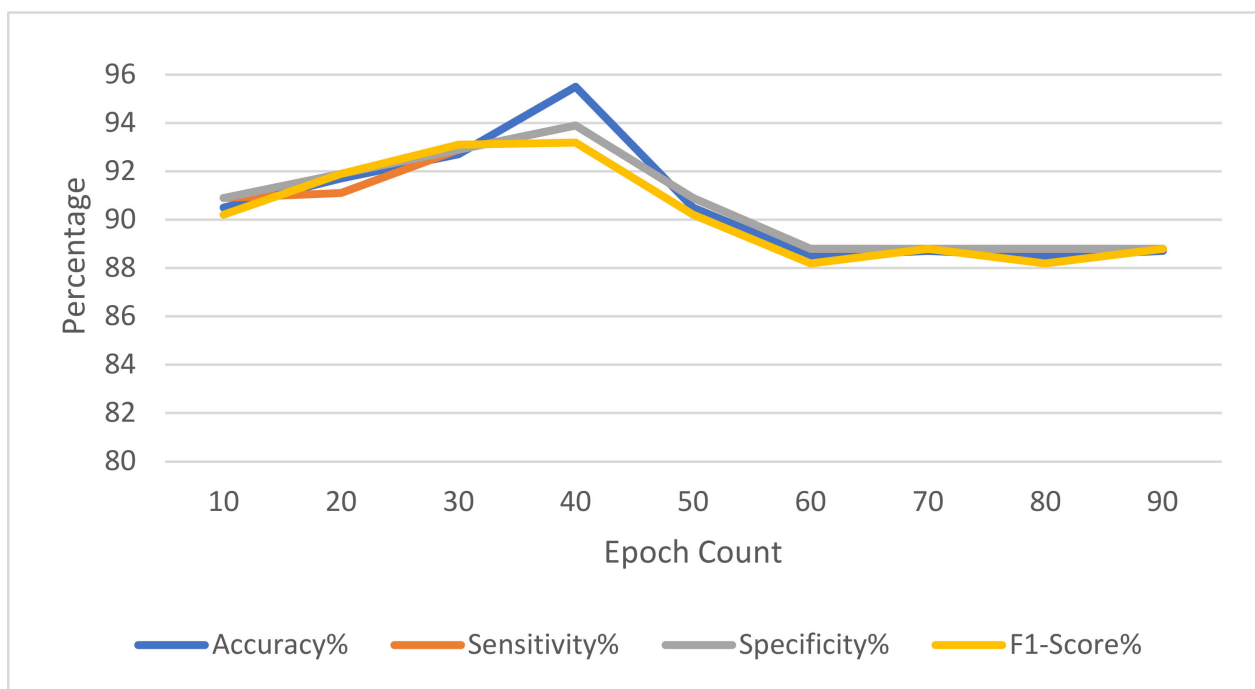


Figure 5. The impact of the count of epochs on the performance of the model.

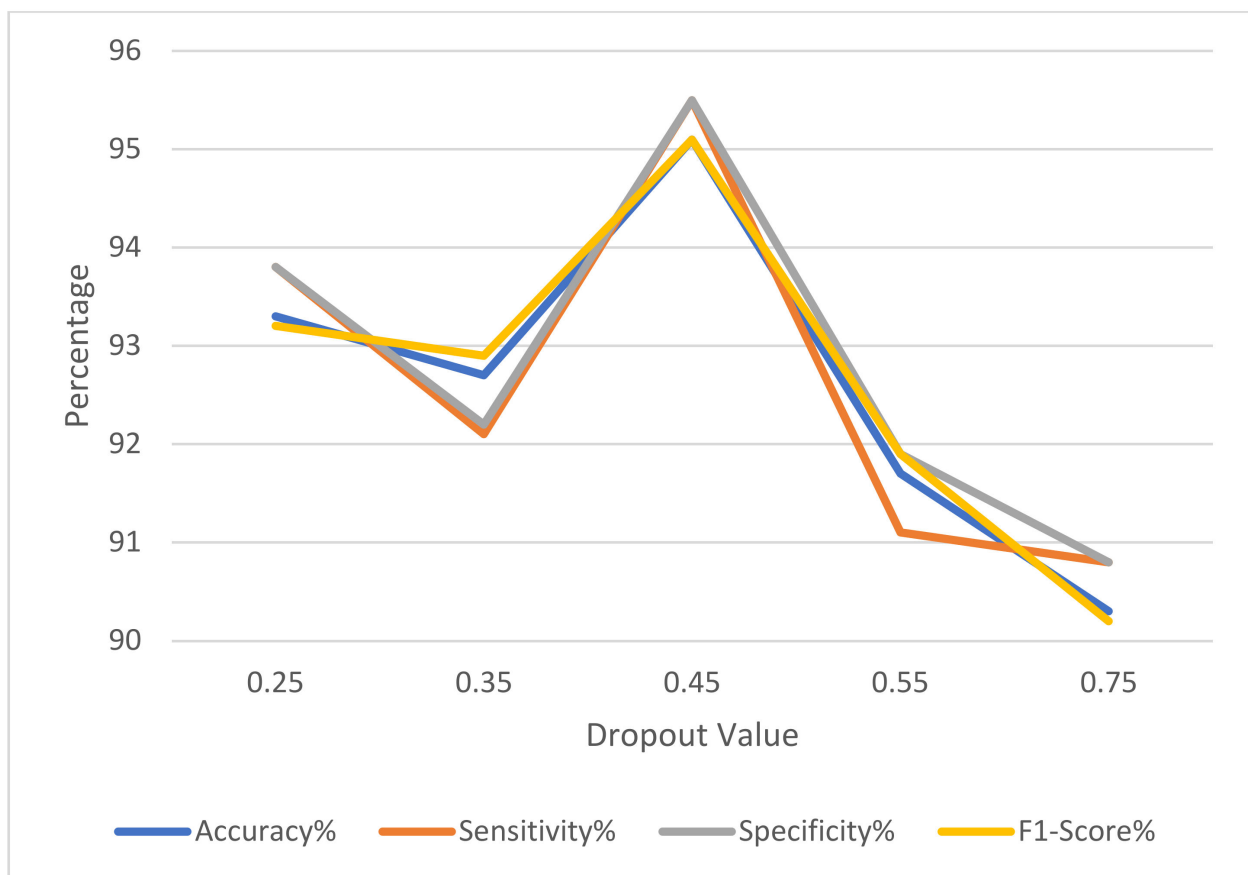


Figure 6. The impact of the dropout on the performance of the proposed model.

The impact of the word vector score depends on the lexical dictionary. The scored and unscored vectors were utilized as independent parameters in the experiment. The

experimental results are depicted in Table 12. The vector score of the lexical data can improve the lexical properties of the review. The proposed model attains higher accuracy from the scored word vector than the unscored word vector.

Table 12. The impact of the word vector score on the performance of the proposed model.

	Accuracy%	Sensitivity%	Specificity%	F0-Score%
Scored Word Vector	95.2	95.1	94.5	94.2
Unscored vector	90.7	90.2	90.9	90.9

A comparison of the proposed DL-DA model with the state of the art models is depicted in Table 13. The results display that the prediction accuracy of the deep models CNN and BiCNN is considerably higher than the deep models naïve Bayes and support vector machine (SVM). Additionally, incorporation of the attention feature model enhances the prediction performance of those models. The prediction accuracy of the DL-DA model is higher than other models.

Table 13. Comparison of state of the art models versus our model.

Model	NaïveBayes [11]	Support-Vector Machine SVM) [14]	CNN [10]	CNN and Attention Model	BiCNN [11]	BiCNN and Attention Model	Our DL-DA Model
Accuracy%	60.5	71.7	91.7	91.5	91.5	91.5	96.7
Sensitivity%	61.1	71.1	89.1	91.1	91.1	91.1	95.1
Specificity%	60.5	70.4	90.4	91.1	91.4	91.7	95.9
F1-Score%	61.1	71.1	89.1	91.1	91.1	91.4	96.1

5. Discussion

We propose a novel lexical analysis deep learning model (DL-DA). A constructed lexical dictionary is utilized to score the lexical vector in the review phrase depicted in Appendix A Table A1. The deep learning neural network selects the significant features in the input vector. The BiCNN reflects the order data of the input and the text context properties. The attention model allocates different rates of various input properties and selects the lexical properties of the review. The fully connected layers are utilized to classify the properties. The proposed model improves the selection of the lexical properties of the customer review phrases and incorporates the significant properties to improve the predication accuracy.

The experiments studied the impact of the input size on the accuracy. The training dataset size impacts the accuracy of the classification. The experimental results prove that the accuracy of the model is the highest with a medium-sized input.

The proposed model classified more than 98% of the instances accurately and mapped them to the correct lexicon. To validate our results and verify the proposed model, the specific classification accuracy rates of the different output classes, including green products, environmental aspects, durability, clean energy, and material cycles, are shown in Figure 7.

To prove that the proposed model is robust, the experimental results are compared with state of the art models. Figure 8 depicts the comparison of the proposed model with deep learning models for the consumer behavior classifier. The proposed model has a higher performance than state of the art models by more than 8% in accuracy with moderate recall.

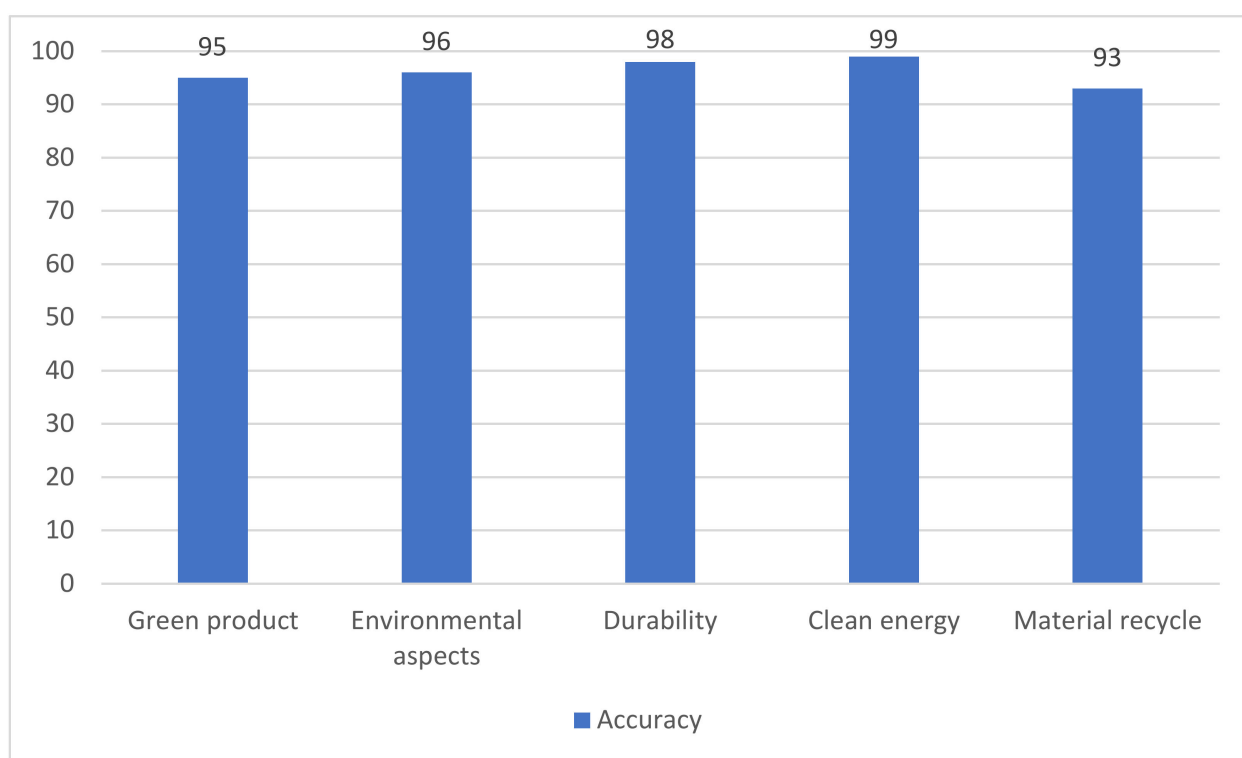


Figure 7. Classification accuracy rate for the specific lexical classes.

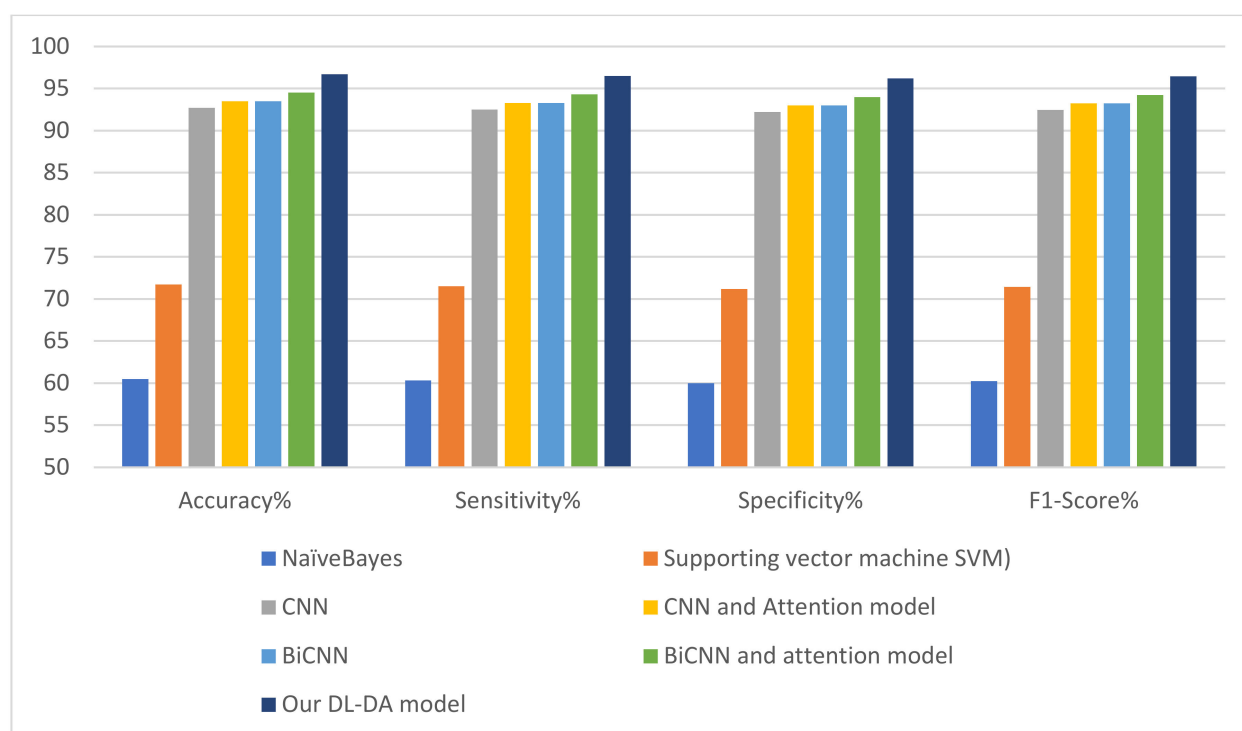


Figure 8. Comparison of state of the art models versus our model.

The incorporation of the proposed model with sustainable input reviews can enhance their efficiency to induce more sustainable efforts in the products.

6. Conclusions

The speedy progress of e-commerce sites attenuates the lexical investigation of customer reviews. The proposed DL-DA model utilizes lexical training models of customer reviews to build a lexical dictionary. The word dataset is utilized to improve the lexical properties in the customer reviews. The employed deep learning extracts the key features of the customer reviews. The attention model scores the consumer reviews. Finally, the scored lexical properties undergo classification. By studying the experimental results, it can be proven that the classification model has higher accuracy than other lexical techniques. By utilizing our model to predict input reviews, the model can help customers gain instant and timely feedback to enhance amenity quality.

The constant augmentation of the lexical dictionary and the data size enhance the prediction performance. We propose a deep learning model for lexical classification. The proposed model augmentation was based on a deep learning technique with parameter tuning. On the crawled Arabic dataset, the precision of this model was assessed using various performance metrics, including accuracy, precision, recall, and F1-score. The experimental results were compared to similar state of the art models. The experimental results were more accurate than comparable approaches. In the future, this work can involve other deep learning architectures, such as transfer learning and auto encoders.

Our model is limited to binary classification of reviews into positive and negative classes. This can be extended to multi-classification in the future. Thus, the future work will concentrate on lexical and context input fineness.

Author Contributions: Conceptualization, H.A.H.M. and N.A.H.; methodology, H.A.H.M.; software, H.A.H.M.; validation, H.A.H.M., N.A.H.; formal analysis, H.A.H.M.; investigation, H.A.H.M.; resources, H.A.H.M.; data curation, H.A.H.M.; writing—original draft preparation, H.A.H.M.; writing—review and editing, H.A.H.M.; visualization, H.A.H.M.; supervision, H.A.H.M.; project administration, H.A.H.M.; funding acquisition, N.A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R113), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

Appendix A

Table A1. An example of collected data from Arabic reviews from online shopping forums with English translation.

Rate 1 = Strongly disagree 2 = Disagree 3 = Neither agree or disagree 4 = Agree 5 = Strongly Agree 1 = أرفض بشدة 2 = لا أوافق 3 = لا أوافق أو لا أوافق 4 = موافق 5 = أوافق بشدة	The Original Question in Arabic	Translated Question
4	Most people around me think green purchasing is an effective way to save the earth	يعتقد معظم الناس من حولي أن الشراء الأخضر هو وسيلة فعالة لإنقاذ الأرض
5	Most people who are important to me think engaging in green purchasing is good	يعتقد معظم الأشخاص المهمين بالنسبة لي أن الانخراط في الشراء الأخضر أمر جيد
3	My family highly values the environmental friendliness of a product	تقدر عائلتي بشدة الصداقة البيئية للمنتج
4	People in my circle of friends highly value the environmental friendliness of a products	الأشخاص في دائرة أصدقائي يقدرون تقديراً عالياً الملاءمة البيئية للمنتجات
4	The decision whether or not to consume green products is within my control	إن قرار استهلاك المنتجات الخضراء أم لا هو تحت سيطرتي
4	For me to do green purchasing is do-able	بالنسبة لي ، فإن الشراء الأخضر هو أمر ممكن
4	I have the ability to do green purchases	لدي القدرة على القيام بعمليات شراء صديقة للبيئة
5	I have the ability to do green practices	لدي القدرة على القيام بالممارسات الخضراء
5	I believe that my decision in consuming green products have a direct influence on the environmental as a whole	أعتقد أن قراري في استهلاك المنتجات الخضراء له تأثير مباشر على البيئة ككل
3	When deciding what to buy, consumers should balance the product's price with the best interest of the environment	عند تحديد ما يجب شراؤه ، يجب على المستهلكين موازنة سعر المنتج مع المصلحة الفضلى للبيئة

Table A1. Cont.

Rate 1 = Strongly disagree 2 = Disagree 3 = Neither agree or disagree 4 = Agree 5 = Strongly Agree 1 = أرفض بشدة 2 = لا أوافق 3 = لا أوافق أو لا أوافق 4 = موافق 5 = أوافق بشدة	The Original Question in Arabic	Translated Question
4	Those who consume more, should be the most responsible in protecting the environment	أولئك الذين يستهلكون أكثر ، يجب أن يكونوا أكثر مسؤولية في حماية البيئة
5	Consumers should consider the environment as one of their stakeholders when making decisions	يجب على المستهلكين اعتبار البيئة كأحد أصحاب المصلحة عند اتخاذ القرارات
4	To be environmentally responsible, consumers need to make purchases that is best for the environment	لكي يكون المستهلكون مسؤولين بيئيًا ، يحتاجون إلى إجراء عمليات شراء تناسب البيئة
2	Consumers should be concerned about maintaining a good place to live	يجب أن يهتم المستهلكون بالحفاظ على مكان جيد للعيش فيه
5	Consumers should have interest in the well-being of the community in which they live in	يجب أن يهتم المستهلكون برعاية المجتمع الذي يعيشون فيه
1	I often seek out information about new products and brands	غالبًا ما أبحث عن معلومات حول المنتجات والعلامات التجارية الجديدة
2	I frequently look for new products and services	كثيرًا ما أبحث عن منتجات وخدمات جديدة
5	I seek out situations in which I will be exposed to new and different sources of product information	أبحث عن المواقف التي أتعرض فيها لمصادر جديدة ومختلفة لمعلومات المنتج
3	I am continually seeking new product experience	أنا أبحث باستمرار عن تجربة منتج جديد
1	When I go online shopping, I find myself spending very little time checking out new products and brands	عندما أذهب للتسوق ، أجد نفسي أقضي القليل من الوقت في التحقق من المنتجات والعلامات التجارية الجديدة

Table A1. Cont.

Rate 1 = Strongly disagree 2 = Disagree 3 = Neither agree or disagree 4 = Agree 5 = Strongly Agree 1 = أرفض بشدة 2 = لا أوافق 3 = لا أوافق أو لا أوافق 4 = موافق 5 = أوافق بشدة	The Original Question in Arabic	Translated Question
5	I take advantage of the first available opportunity to find out about new and different products	أستفيد من أول فرصة متاحة للتعرف على المنتجات الجديدة والمختلفة
3	I prefer purchasing of energy saving products (e.g., low energy lamps, energy efficient fridge)	(مثل المصابيح منخفضة الطاقة والثلاجة الموفرة للطاقة) أفضل شراء المنتجات الموفرة للطاقة
	I prefer purchasing water saving products (e.g., water savings shower heads/washing machine)	(الغسالات الموفرة للمياه / مثل رؤوس الدش) أفضل شراء منتجات توفير المياه
5	I prefer purchasing product free from Chlorofluorocarbons (CFC) (e.g., CFC-free aerosols propellant, CFC-free inhalers)	(على سبيل المثال ، بخاخات بخارية خالية من مركبات الكربون الكلورية فلورية ، وأجهزة استنشاق خالية من مركبات الكربون الكلورية فلورية) أفضل شراء منتج خالٍ من مركبات الكربون الكلورية فلورية
3	I prefer purchasing recycled paper products (e.g., recycled paper note pads, recycled face tissue)	(على سبيل المثال ، المناديل الورقية المعاد تدويرها ، مناديل الوجه المعاد تدويرها) أنا أفضل شراء المنتجات الورقية المعاد تدويرها
4	I prefer purchasing green/organic foods and vegetables (e.g., vegetables grown without chemical, meat that don't contain hormones)	(مثل الخضروات المزروعة بدون مواد كيميائية ، واللحوم التي لا تحتوي على هرمونات) العضوية / أفضل شراء الأطعمة والخضروات الخضراء
5	I prefer purchasing environmentally friendly building material (e.g., green roof, bamboo flooring)	(مثل الأسطح الخضراء وأرضيات الخيزران) أفضل شراء مواد بناء صديقة للبيئة
5	I prefer purchasing green furniture (e.g., bamboo or rattan furniture)	(مثل أثاث الخيزران أو الخيزران) أفضل شراء أثاث أخضر

Table A1. Cont.

Rate 1 = Strongly disagree 2 = Disagree 3 = Neither agree or disagree 4 = Agree 5 = Strongly Agree 1 = أرفض بشدة 2 = لا أوافق 3 = لا أوافق أو لا أوافق 4 = موافق 5 = أوافق بشدة	The Original Question in Arabic	Translated Question
3	I make every effort to buy paper products made from recycle papers	أبذل قصارى جهدي لشراء المنتجات الورقية المصنوعة من أوراق إعادة التدوير
4	When I purchase products, I always make a conscious effort to buy those products that are low in pollutants	عندما أشتري المنتجات ، أبذل دائماً جهداً واعياً لشراء تلك المنتجات التي تحتوي على نسبة منخفضة من الملوثات
2	I try to buy products that can be recycled	أحاول شراء منتجات يمكن إعادة تدويرها
5	I make every effort to recycle paper	أبذل قصارى جهدي لإعادة تدوير الورق
3	I make every effort to recycle plastic	أبذل قصارى جهدي لإعادة تدوير البلاستيك
4	I make every effort to recycle glass	أبذل قصارى جهدي لإعادة تدوير الزجاج

References

1. Alsubari, S.N.; Deshmukh, S.N.; Al-Adhaileh, M.H.; Alsaade, F.W.; Aldhyani, T.H. Development of integrated neural network model for identification of sustainability reviews in e-commerce using multi domain datasets. *Appl. Bionics Biomech.* **2021**, *2021*, 1–11. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Li, Y.; Feng, X.; Zhang, S. Detecting reviews concerning sustainability utilizing semantic and emotion model. In Proceedings of the 2016 3rd International Conference on Information Science and Control Engineering, Beijing, China, 6 July 2016; pp. 317–320.
3. Hu, X.; Tang, J.; Gao, H.; Liu, H. Product sustainability detection with sentiment information. In Proceedings of the 2014 IEEE International Conference on Data Mining, Shenzhen, China, 1 June 2014; pp. 180–189.
4. Long, F.; Zhou, K.; Ou, W. Sentiment analysis of text based on bidirectional LSTM with multi-head attention for product sustainability. *IEEE Access* **2019**, *7*, 141960–141969. [\[CrossRef\]](#)
5. Feng, V.W.; Hirst, G. Detecting green product opinions with profile compatibility. In Proceeding of the 6th International Joint Conference on Natural Language Processing, Nagoya, Japan, 21 January 2013; pp. 14–18.
6. Delany, S.J.; Buckley, M.; Greene, D. SMS green product sustainability issues filtering: Methods and data. *Expert Syst. Appl.* **2012**, *39*, 9899–9908. [\[CrossRef\]](#)
7. Li, L.; Qin, B.; Ren, B.W.; Liu, T. Document representation and feature combination for sustainability review detection. *Neurocomputing* **2016**, *254*, 33–41. [\[CrossRef\]](#)
8. Sarika, S.; Nalawade, M.S.; Pawar, S.S. A survey on detection of shill reviews by measuring its linguistic features for green product reviews. *Int. J. Emerg. Trends Technol. Comput. Sci.* **2014**, *3*, 269–272.
9. Peng, Q. Store review sustainability detection based on review relationship. In *Advances in Conceptual Modeling*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 287–298.
10. Hussain, N.; Mirza, H.T.; Hussain, I.; Iqbal, F.; Memon, I. Sustainability review detection using the linguistic and spammer behavioral methods. *IEEE Access* **2020**, *8*, 53801–53816. [\[CrossRef\]](#)
11. Shojaei, S.; Murad, M.; Azman, A.B.; Sharef, N.M.; Nadali, S. Detecting green product reviews using lexical and syntactic features. In Proceedings of the 2013 13th International Conference on Intelligent Systems Design and Applications, Salangor, Malaysia, 27 May 2013; pp. 53–58.
12. Heydari, A.; Tavakoli, M.A.; Salim, M.N.; Heydari, Z. Detection of customer review in sustainability: A survey. *Expert Syst. Appl.* **2015**, *42*, 3634–3642. [\[CrossRef\]](#)
13. Zhang, D.; Leng, J.; Li, X.; He, W.; Chen, W. Three-Stream and Double Attention-Based DenseNet-BiLSTM for Fine Land Cover Classification of Complex Mining Landscapes. *Sustainability* **2022**, *14*, 12465. [\[CrossRef\]](#)
14. Crawford, M.; Khoshgoftaar, T.M.; Prusa, J.D.; Richter, A.N.; Al Najada, H. Survey of review of green products detection using machine learning techniques. *J. Big Data* **2015**, *2*, 1–24. [\[CrossRef\]](#)
15. Noekhah, S.; Fouladfar, E.; Salim, N.; Ghorashi, S.H.; Hozhabri, A.A. A novel approach for opinion spam detection in e-commerce. In Proceedings of the 8th IEEE International Conference on E-Commerce with Focus on E-Trus, Mashhad, Iran, 25 January 2014; pp. 1–8.
16. Ott, M.; Cardie, C.; Hancock, J.T. Negative sustainability opinion spam. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Atlanta, GA, USA, 9–14 June 2013; pp. 497–501.
17. Alsubari, S.N.; Shelke, M.B.; Deshmukh, S.N. Sustainability reviews identification based on deep computational linguistic features. *Int. J. Adv. Sci. Technol.* **2020**, *29*, 3846–3856.
18. Goswami, K.; Park, Y.; Song, C. Impact of reviewer social interaction on online consumer review for sustainability fraud detection. *J. Big Data* **2017**, *4*, 1–19. [\[CrossRef\]](#)
19. Jindal, N.; Liu, B. Opinion of green products in customer review. In Proceedings of the 2008 International Conference on Web Search and Data Mining, Palo Alto, CA, USA, 17 July 2008; pp. 219–230.
20. Mukherjee, A.; Venkataraman, V.; Liu, B.; Glance, N. What yelp sustainability review filter might be doing. In Proceedings of the International AAAI Conference on Weblogs and Social Media, MA, USA, 4 December 2013; pp. 409–418.
21. Ahmed, H.; Traore, I.; Saad, S. Detecting opinion of sustainability and green product news using text classification. *Secur. Priv.* **2018**, *1*, 1–15.
22. Savage, D.; Zhang, X.; Yu, X.; Chou, P.; Wang, Q. Detection of opinion of sustainability based on anomalous rating deviation. *Expert Syst. Appl.* **2015**, *42*, 8650–8657. [\[CrossRef\]](#)
23. Fitzpatrick, E.; Bachenko, J.; Fornaciari, T. Automatic detection of verbal deception. *Comput. Linguist.* **2015**, *43*, 269–271.
24. Zhang, D.; Zhou, L.; Kehoe, J.L.; Kilic, I.Y. What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of sustainability reviews. *J. Manag. Inf. Syst.* **2016**, *33*, 456–481. [\[CrossRef\]](#)
25. Luca, M.; Zervas, G. Fake it till you make it: Reputation, competition, and Yelp review fraud. *Manag. Sci.* **2016**, *62*, 3412–3427. [\[CrossRef\]](#)
26. Akoglu, L.; Chandy, R.; Faloutsos, C. Opinion sustainability detection in online reviews by network effects. In Proceedings of the ICWSM, Cambridge, MA, USA, 8–11 July 2013; pp. 2–11.
27. Barbado, R.; Araque, O.; Iglesias, C.A. A framework for fake review detection in online consumer electronics retailers. *Inf. Process. Manag.* **2019**, *56*, 1234–1244. [\[CrossRef\]](#)

28. Hajek, P.; Barushka, A.; Munk, M. Fake consumer review for sustainability detection using deep neural networks integrating word embeddings and emotion mining. *Neural Comput. Appl.* **2020**, *32*, 17259–17274. [\[CrossRef\]](#)
29. Narayan, R.; Rout, J.K.; Jena, S.K. Review of sustainability detection using opinion mining. In *Progress in Intelligent Computing Techniques: Theory, Practice, and Applications*; Springer: Singapore, 2018; pp. 273–279.
30. Shirsat, V.S.; Jagdale, R.S.; Deshmukh, S.N. Sentence level sentiment identification for sustainability from news articles using machine learning techniques. In *Computing, Communication and Signal Processing*; Springer: Singapore, 2019; pp. 371–376.
31. Louppe, G.; Wehenkel, L.; Suter, A.; Geurts, P. Understanding variable importance in forests of randomized trees. In *Advances in Neural Information Processing Systems*; Lafayette, IEEE, Tahoe, CA, USA, May, 26, 2013; Volume 26, pp. 431–439.
32. Toke, P.S.; Mutha, R.; Naidu, O.; Kulkarni, J. Enhancing text mining using side information for sustainability studies. *Int. J. Adv. Res. Comput. Commun. Eng.* **2016**, *5*, 793–797.
33. Akhtar, M.J.; Ahmad, Z.; Amin, R.; Almotiri, S.H.; Al Ghamdi, M.A.; Aldabbas, H.J.C. An efficient mechanism for product data extraction from e-commerce websites. *Comput. Mater. Contin.* **2020**, *65*, 2639–2663. [\[CrossRef\]](#)
34. Alantari, H.J.; Currim, I.S.; Deng, Y.; Singh, S. An empirical comparison of machine learning methods for text-based sentiment analysis of online consumer reviews. *Int. J. Res. Mark.* **2021**, *39*, 1–19. [\[CrossRef\]](#)
35. Yadav, V.; Verma, P.; Katiyar, V. E-commerce product reviews using aspect based Hindi sentiment analysis. In *Proceedings of the 2021 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 27–29 January 2021; pp. 1–8.
36. Desai, Z.; Anklesaria, K.; Balasubramaniam, H. Business Intelligence Visualization Using Deep Learning Based Sentiment Analysis on Amazon Review Data. In *Proceedings of the 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 6–8 July 2021; pp. 1–7.
37. Mohbey, K.K. Sentiment analysis for product rating using a deep learning approach. In *Proceedings of the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, Coimbatore, India, 25–27 March 2021; pp. 121–126.
38. Darokar, M.S.; Raut, A.D.; Thakre, V.M. Methodological Review of Emotion Recognition for Social Media: A Sentiment Analysis Approach. In *Proceedings of the 2021 International Conference on Computing, Communication and Green Engineering (CCGE)*, Pune, India, 23–25 September 2021; pp. 1–5.
39. Yousefpour, A.; Ibrahim, R.; Hamed, H.N.A. Ordinal-based and frequency-based integration of feature selection methods for sentiment analysis. *Expert Syst. Appl.* **2017**, *75*, 80–93. [\[CrossRef\]](#)
40. Aljuhani, S.A.; Alghamdi, N.S. A Comparison of Sentiment Analysis Methods on Amazon Reviews of Mobile Phones. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*, 608–617. [\[CrossRef\]](#)
41. Nguyen, H.; Veluchamy, A.; Diop, M.; Iqbal, R. Comparative Study of Sentiment Analysis with Product Reviews Using Machine Learning and Lexicon-Based Approaches. *SMU Data Sci. Rev.* **2018**, *1*, 7.
42. Iqbal, A.; Amin, R.; Iqbal, J.; Alroobaea, R.; Binmahfoudh, A.; Hussain, M. Sentiment Analysis of Consumer Reviews Using Deep Learning. *Sustainability* **2022**, *14*, 10844. [\[CrossRef\]](#)