Article

Role of Artificial Neural Networks Techniques in Development of Market Intelligence: A Study of Sentiment Analysis of eWOM of a Women's Clothing Company

Zubair Nawaz [1,*], Chenhui Zhao [2,*], Fouzia Nawaz [3], Asif Ali Safeer [1,4] and Waseem Irshad [1]

[1] School of Management, Huazhong University of Science and Technology, Wuhan 430074, China; aasafeer@hust.edu.cn (A.A.S.); waseemirshad@hust.edu.cn (W.I.)
[2] Business School, Wuchang University of Technology, Wuhan 430223, China
[3] Department of Computer Sciences and Information Technology, University of Sargodha, Sargodha 40100, Pakistan; fouzia.nawaz1@gmail.com
[4] Business School, Huanggang Normal University, Huanggang 438000, China
[*] Correspondence: zubairnawaz@hust.edu.cn (Z.N.); z0001292@zuel.edu.cn (C.Z.)

Abstract: Web 2.0 technology enables customers to share electronic word of mouth (eWOM) about their experiences. eWOM offers great market insights to the organization, and important for organization's success. eWOM monitoring and management is one of the major contemporary challenges for the organization, because of high volume and frequency of the content. It is nearly impossible for an organization to manually monitor content generated by each user. In this paper, we propose sentiment analysis as an alternative method for analysis of emotions and behavioral intentions in real-time data. Sentiment analysis is performed on women's e-clothing reviews collected from the Kaggle data repository. The dataset consists of 23,486 reviews, comprising ten feature variables. This study applied artificial neural network techniques to determine polarity of the data in terms of positive or negative. Sentiment analysis was performed by using two artificial neural networks, Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM), to classify the review as recommended (positive) or not recommended (negative). The proposed models have been evaluated on these performance measuring parameters: accuracy, recall, specificity, F1-score and roc-curve. The LSTM method outperformed CNN and achieved classification accuracy of 91.69%, specificity 92.81%, sensitivity 76.95%, and 56.67% F1-score. Based on results of this study, LSTM technique is highly recommended for the sentiment analysis of unstructured text-based user-generated content.

Keywords: sentiment analysis; neural networks; market intelligence; LSTM; CNN

1. Introduction

Over the past decade, the significance of social media has enormously increased in e-commerce. Web 2.0 enables customers to express their opinions, experiences, criticism, and suggestions. Social media data offer great insights for organizations and potential customers. Organizations utilize this data to analyze customer behavior, and to identify emerging trends in the market; whereas, potential customers seek information about product experiences. Customers value user-generated information more than the information provided by the company [1]. Therefore, it is important for organizations to monitor how their products are discussed "in the wild" [2]. Social media has become an integral part of marketing strategies; organizations use social media for promotion and evaluation of market offerings [3].

Online consumer reviews of products and services is an important source of electronic words of mouth (eWOM), which refers to "any positive or negative verbal statement by potential, actual, and former customers of a product or company" [4]. eWOM is unstructured information, which offers great depth of consumers' sentiments. eWOM

has been regarded as the most credible source of product information, and potential customers often pay attention to eWOM during purchase decisions [5]. Positive eWOM significantly increase sales [6,7], brand equity [8], and company's reputation [9]; whereas, negative eWOM significantly lowers purchase intentions and brand trust [10]. eWOM is more influential as compared to conventional interpersonal interactions because of high virality, reliability, and anonymity [11]. Consequently, eWOM can make or break a company's reputation, and an organization must pay attention to the presence of eWOM on internet platforms.

One of the major contemporary problems of e-commerce is the detection and timely response to the critical comments. Daily, huge amounts of data are generated on social media platforms, and manual analysis of data is time consuming and costly for organizations. For large organizations, it is nearly impossible to monitor the data manually, because organizations may be discussed in user-generated content millions of times. An organization's timely reaction to the eWOM is important, especially for negative eWOM. Studies show that negative eWOM has more impact on behavioral intentions as compared to positive eWOM [5]. Hence, computer-based technologies need to be applied for refining, monitoring, and managing the data.

Over the past few years, sentiment analysis received wide attention from researchers and organizations. Sentiment analysis is defined as "the process of mining of data, view, review or sentence to predict the emotion of the sentence through Natural Language Processing (NLP)" [1]. Sentiment analysis develops marketing intelligence, which helps in the decision-making process. Studies reported a definite relationship of sentiment analysis of online reviews to the ratings and recommendations; consumers' sentiments are often reflected in the ratings, but sentiment analysis offers more depth of information [12]. Sentiment analysis helps in the classification of data on the basis of polarity, such as "positive" and "negative". Due to the fact that sentiment analysis helps in generating the polarity of good and bad emotions, decision makers can utilize sentiment analysis to identify the attractive features and improvements for products. Moreover, sentiment analysis also helps to read between the lines, and decision makers can consider the intensity of emotions beyond the customer's opinion [13].

Research studies found that artificial neural network techniques are extremely accurate in predictions as compared to traditional computer vision techniques [14]. Based on higher accuracy and efficiency, artificial neural networks are widely used in various fields of studies for prediction of a desired outcome variable [15]. Neural networks are robust in nature and capable of handling large data without going through system failure. Moreover, neural networks have self-learning abilities and do not need direct human supervision. Consequently, neural networks are able to detect trends in the data that are difficult to explore manually. Studies found that neural networks are highly effective for detection of trends and classification of the data [16–18].

There are several neural network techniques, such as, Radial Basis Function Neural Network, Modular Neural Network (MNN), Spiking Neural Network (SNN), Feedforward Neural Network (FNN), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM). CNN is considered as one of the most efficient neural network techniques for sentiment analysis; on sentence level as well as on a document level [19,20]. The main advantage of CNN compared to its predecessors is that it automatically detects the important features without any human supervision. Some studies also established the superiority of LSTM in sentiment analysis [21]. Studies also suggest that LSTM and CNN, with both neural network technique analyses combined, provide better understanding about sentiments [22,23].

This study aims to perform a sentiment analysis on unstructured data of women's e-clothing reviews. The main objective of the study is to classify the data on the basis of underlying emotions of the customers towards products and services of the organization. This study applied CNN and LSTM techniques on real time data to perform sentiment analysis. Furthermore, this study also aims to provide a comparison of CNN and LSTM

with previous proposed models [24,25]. Traditionally, an organization gathers feedback from consumers with the help of a survey; normally, a structured questionnaire is used as a tool for data collection. Survey data have limitations (i.e., response biasness) and cannot provide a true representation of the emotions. This paper proposes neural network techniques (CNN and LSTM) as an alternative method for analysis of emotions and behavioral intentions in real-time data. The proposed model is evaluated on the performance parameters of accuracy, recall, specificity, F1-score, and roc-curve.

2. Background

Emotions have great impact on attitudes, and also help in understanding behavioral intentions. Festinger [26] developed the cognitive consistency theory, which suggests that people are motivated towards actions which are consistent with their beliefs and perceptions. Furthermore, people also motivated to change their behavior if the experiences are not consistent with their perceptions. Therefore, consumers express their opinions about product experiences by rating, reviews, and recommendations. Social media has become an important part of the lifestyle, and consumers' share positive and negative eWOM without any hesitation.

Early work on sentiment analysis can be traced from the late 1970s. Carbonell [27] proposed a theory of "subjective understanding" and designed a process model named POLITICS. Later, Wilks and Bien [28] introduced a model of "computational beliefs" and provided a base for natural language processing. Hearst [29] introduced a direction-based text interpretation model (DTI), which provided the domain-independent text interpretation. Wiebe [30] designed an algorithm that helped in tracking certain viewpoints in fictional narrative texts. Sack [31] proposed an impressive model that can retrieve concise information from a longer text, which helped in retrieval and classification of texts.

Sentiment analysis gained burst popularity from the beginning of the Web 2.0 era; social media sites introduced a new mode of communication, which grasped interest of organizations and researchers. Kantrowitz [32] was among the pioneer scholars who analyzed emotions and its effects in unstructured text. Hu and Liu [33] introduced text mining techniques in customer reviews, which helped in attribute-based review summaries.

2.1. Sentiment Analysis of eWOM in Market Intelligence

Marketing intelligence refers to a continuous process of identifying, understanding, and analyzing the firm's internal and external environments associated with customers, competitors, and markets [34]. Assessment of consumer preferences and identification of current and future trends in the market is one of the major objectives of market intelligence. Conventionally, to develop market intelligence, the organizations collect primary data from actual and potential customers. Recent studies emphasize modern artificial neural network techniques, which are more accurate, efficient, and economical [20,21].

Liang et al. [35] classified customer data on the basis of sentiment analysis, and found that eWOM has strong association with sales of mobile phone applications. They also suggest that eWOM related to both product quality and service quality has significant impact on sales; however, in case of service attributes, the impact of eWOM is very substantial on purchase intentions. Schweidel and Moe [36] analyzed 7565 social media comments related to a company and highlighted the impact of venue on social media monitoring. The study suggested that sentiments vary in strength and subject matter for different platform venues of social media.

Market intelligence not only focuses on the identification and understanding of the needs of potential customers, but also emphasizes post-purchase behavior of the current customers. User's sentiment plays an important role in company reputation. Studies suggest that sentiment analysis of eWOM helps in measurement of corporate reputation. Corporate reputation is measured in terms of volume and valence. Volume refers to the frequency of "corporate brands" mentioned in consumers reviews; whereas, valence refers to positive or negative sentiments of reviewers [37]. Colleoni et al. [38] argued that

organizations are losing control in building corporate reputation by themselves; instead, user-generated content on social media platforms build or damage the corporate reputation. Aakash et al. [39] performed artificial neural network techniques on eWOM of customers of the hospitality industry and measured users' service satisfaction level.

eWOM helps in shaping potential customers' behavior, especially if they are facing uncertainties in the decision-making process [40]. Potential customers seek additional information from the user's comments and reviews in order to minimize the ambiguity. eWOM provides product need-matching information to potential customers, which ultimately reduces uncertainties in the decision making process [41]. Neural network techniques are also currently used for the product-need matching process, which provide a customized solution for the customers. For example, a study revealed that neural network techniques can be applied to provide customized offers to peer-to-peer tourists on the basis of sociodemographic characteristics [42].

2.2. Proposed Research Work

Due to the advent of social media, digital data are increasing daily and can be used to perform data mining analysis to extract useful information. Natural Language Processing (NLP) is the subdomain of Artificial Intelligence (AI) that processes natural languages based on perceptive activities, and extracts meaningful information. Sentiment analysis is a technique of NLP that depends on keywords, frequency of keywords, and co-occurrence of words, according to their context. In this research paper, a semantic analyzer is developed that determines the polarity of each review and assigns a value to the text: either positive or negative. The flow of the proposed work is presented in Figure 1.
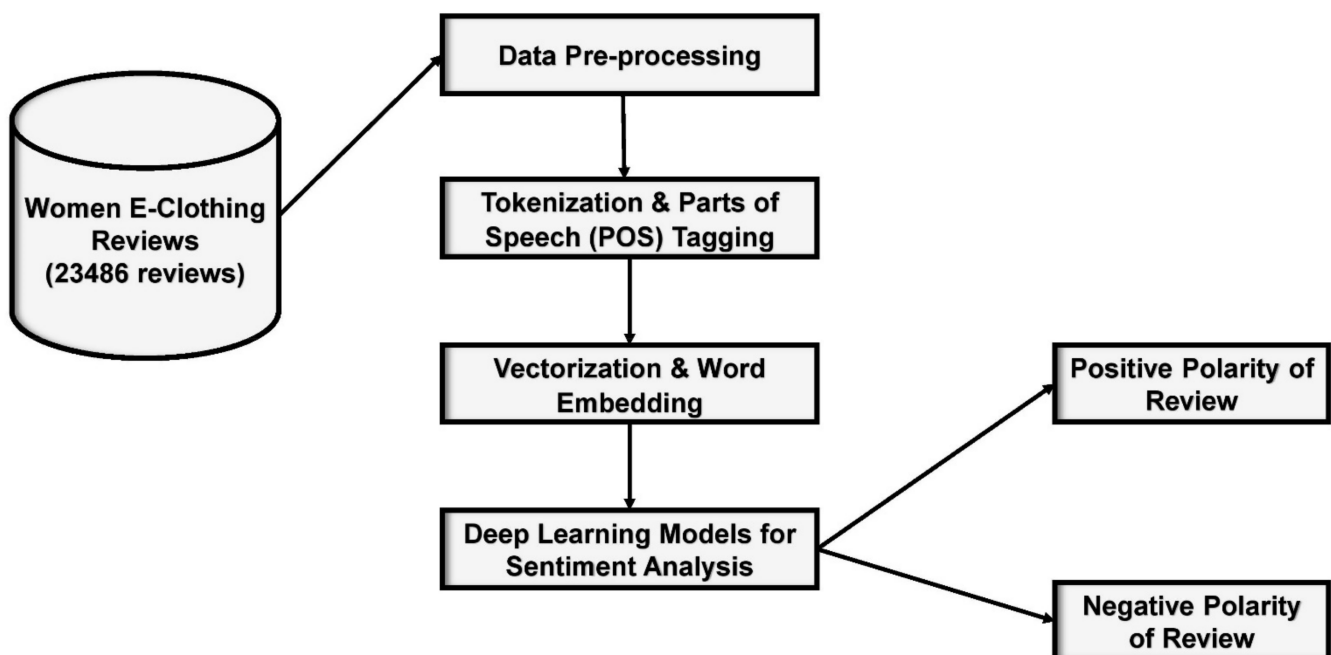

Figure 1. The flow diagram of the proposed architecture.

3. Material and Methods

3.1. Data Repository

Data from the Kaggle repository (https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews, accessed on 15 March 2021) was used in this research [43], which is an open-source platform of data science that helps to explore new dimensions of artificial intelligence by providing benchmark datasets. Kaggle allows the researchers to find and publish datasets, and explore and build new models in collaboration with other data scientists to provide solutions to data-related problems. Data used were re-

lated to women e-clothing reviews, and unstructured in nature. The dataset consisted of 23,486 reviews, along with ten feature variables: clothing ID, age, title, review text, rating, recommendation, positive feedback count, division name, department name, and class name. To maintain privacy, the company name in the text reviews was replaced by the word "Organization".

3.2. Text Cleaning

To obtain reliable results, text cleaning was performed. First, only valid entries are maintained by deleting the rows that contain missing values. Some unnecessary columns were also dropped. Contracting terms are replaced with proper words (i.e., I've → I have, I'm→ I am, and I'll → I will). Second, stop words were also removed as they do not provide any unique information that can be used in sentiment analysis. Finally, after performing necessary pre-processing, the text was converted into lowercase letters.

3.3. Tokenization and POS (Parts of Speech) Tagging

Tokenization and POS are considered as one of the basic steps in natural language processing (NLP); first, tokenization is performed by breaking the text into smaller units based on delimiters. Each smaller unit of text is named as a token. In text-mining, the only Bag of Words (BOW) analysis is insufficient as it fails to capture the syntactic relation between the words. For example, "I like you" is a sentence where "like" is a verb and depicts a positive sentiment; whereas, the phrase "I'm like you" depicts neutral sentiment, and "like" is a preposition here. To improve the BOW technique, POS is performed to obtain the most pertinent information. POS tagging is a technique where the words are matched to the corresponding part of the speech tag based on the context of the text. After POS, lemmatizing is performed to reduce the word into its root word.

3.4. Vectorization and Word Embedding

To build a deep learning model for text mining it is necessary to convert the words into real numbers. Humans have a very good understanding of words and their relationships, but it is not easy for machines to understand the meaning of words according to their context. The process of converting text into a real number vector is called vectorization. Word embedding is a technique that converts the word into representation; the form that imparts the human understanding of language into the machines. Previously, Continuous Bag of Words (CBOW) and TF-IFD techniques were utilized to find the co-occurrence of words in the corpus. In this research, GloVe (Global Vectors for word representation) is employed to obtain the co-occurrence relation between the words in the corpus.

3.5. Sentiment Analysis Based on Deep Learning Models

Neural networks are considered good at recognizing the patterns that are extracted from a real-world dataset like images, text, and sound. In this research, after performing lexical analysis of the text, state of the art deep neural networks are employed for the sentiment analysis of women's e-clothing reviews: Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM). Each word of the corpus is translated into the vector space; CNN and RNN are both deep learning models that learn the data through various deep layers. CNN learns the data representation through the words of the corpus and extracts the feature set that makes the semantic representation. The CNN sequential model is defined with four layers: (i) an input layer that contains the size of the corpus and a pre-trained embedding layer, (ii) the GlobalMaxPool1D layer is utilized to generate a feature map for each category of classification and it does not require any optimizing parameter that reduces the overfitting problem, (iii) the fully connected dense layer to learn the connections among the text, and (iv) the output layer to predict the polarity of the text. LSTM is a variant of RNN and is considered an expert in processing sequential data. Due to its feedback connections, it can process the entire block of data. LSTM is known as the best-suited algorithm for sequential text data as it has some memory that

remembers the long-distance dependency between the tokens of the corpus. Due to the internal memory of the RNN, they can remember the information about the input that helps them to perform precise prediction about incoming data. RNNs are the preferred algorithm for sentiment analysis; RNNs have a deeper understanding of the sequence and its context as compared to other neural networks. In this research, our LSTM architecture is assembled using five layers: (i) an input layer that is enriched with Glove pre-trained word embedding, which uses 100 length vectors to characterize each word of the corpus; (ii) the dropout layer; (iii) the next is the LSTM layer that uses 128 internal memory units; and (iv) the output layer with sigmoid activation function and binary cross entropy, which are used because it is a binary classification problem.

3.6. Experimental Setup

The semantic analyzer was developed using Python 3.7, Keras, and the natural language processing library PyTorch version 1.3 in a Core i7 Windows environment. To perform the text pre-processing, the NLTK, spacy, and en_core_web libraries were utilized. Adam optimizer was used along with a 0.001 learning rate sigmoid activation and binary cross-entropy loss function for both configurations. Adam optimizer was considered as the default optimization method for deep learning problems, and optimal learning rate was chosen to avoid divergent behavior. Binary cross-entropy function was employed for the binary classification problem. For LSTM based implementation, 128 batch size was used. The dataset was split into an 80:20 ratio of training and testing parts.

4. Results

To measure the average performance in this research, the performance of proposed models was evaluated on statistic performance measures of accuracy, sensitivity, specificity, and F1-score. The accuracy depicts the efficiency of the model in predicting the polarity of the review. AUC (Area Under the Curve) and ROC (Receiver Operating Characteristics) are considered as the most important performance measuring metrics. ROC is a probability curve that describes the operation of the classification model at all threshold values; the curve plots two parameters: the true positive rate is plotted on the y-axis and the false positive rate is plotted on the x-axis. The AUC curve reveals the degree of separability (how well a model distinguishes between the classes). It measures the entire flat area under the ROC curve. The greater value of ROC means that the system is proficient at classifying the polarity of the reviews. The ROC and AUC are collectively termed as AUROC. Precision assesses the ratio of actual negatives that are corrected identified. The formulas used for measuring the accuracy, specificity, and sensitivity follow:

$$Accuracy = \frac{TP + TN}{TN + TP + FN + FP} \tag{1}$$

$$Sensitivity/Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$F1 = \frac{2 * Recall * Precision}{Recall + Precision} \tag{5}$$

where *TP*, *FN*, *TN*, *FP* are the true positive, false negative, true negative, and false positive, respectively. We experimented with the CNN and LSTM on a real-time commercial dataset of women's e-clothing reviews; the training graphs of models are presented in Figures 2 and 3. The model was trained for 25 epochs together with a batch size of 50.
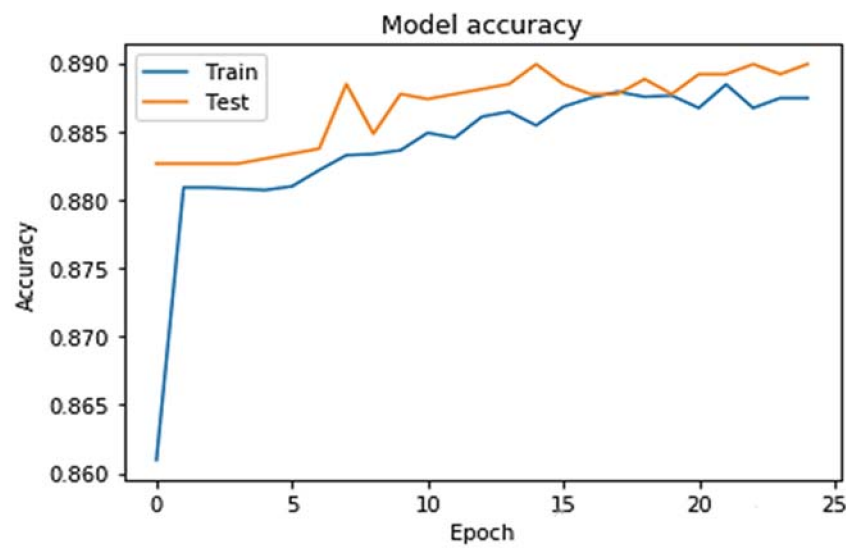
(a)
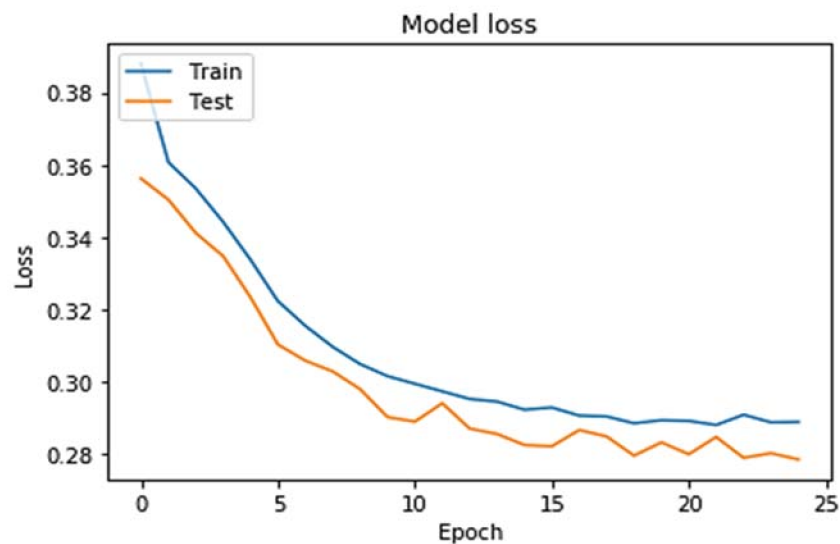
(b)

Figure 2. (**a**) CNN training accuracy graph and (**b**) CNN training loss graph.

The CNN model was trained for 25 epochs together with a batch size of 50. The batch size and epoch are control hyperparameters, which are associated with the optimization. The number of epochs defines the number of times the entire training dataset passed through the learning algorithm, which means that each data sample obtains an opportunity to update the internal model parameters. The dataset is grouped into the batches. Each batch contains a batch size. The impact of batch size is typically computational as it usually affects the training time and not so much the test performance. The smaller batch size is used for both configurations as it results in quick convergence and more stability, which can be observed in Figures 2 and 3. The graphs are plotted to visualize the model's training accuracy and loss trends, as can be seen in Figures 2 and 3. In Figure 2a, an interesting trend is observed—the accuracy curve rises as the number of epochs increases. Initially the validation accuracy value is less, but the validation accuracy curve is higher than the training curve. In Figure 2b, the validation loss curve, which is generally downward, shows that the validation loss or error decrease as the model improves in terms of learning. Similarly, in Figure 3a, the rising trend of the accuracy curve is observed, and Figure 3b shows degradation of model loss or error.
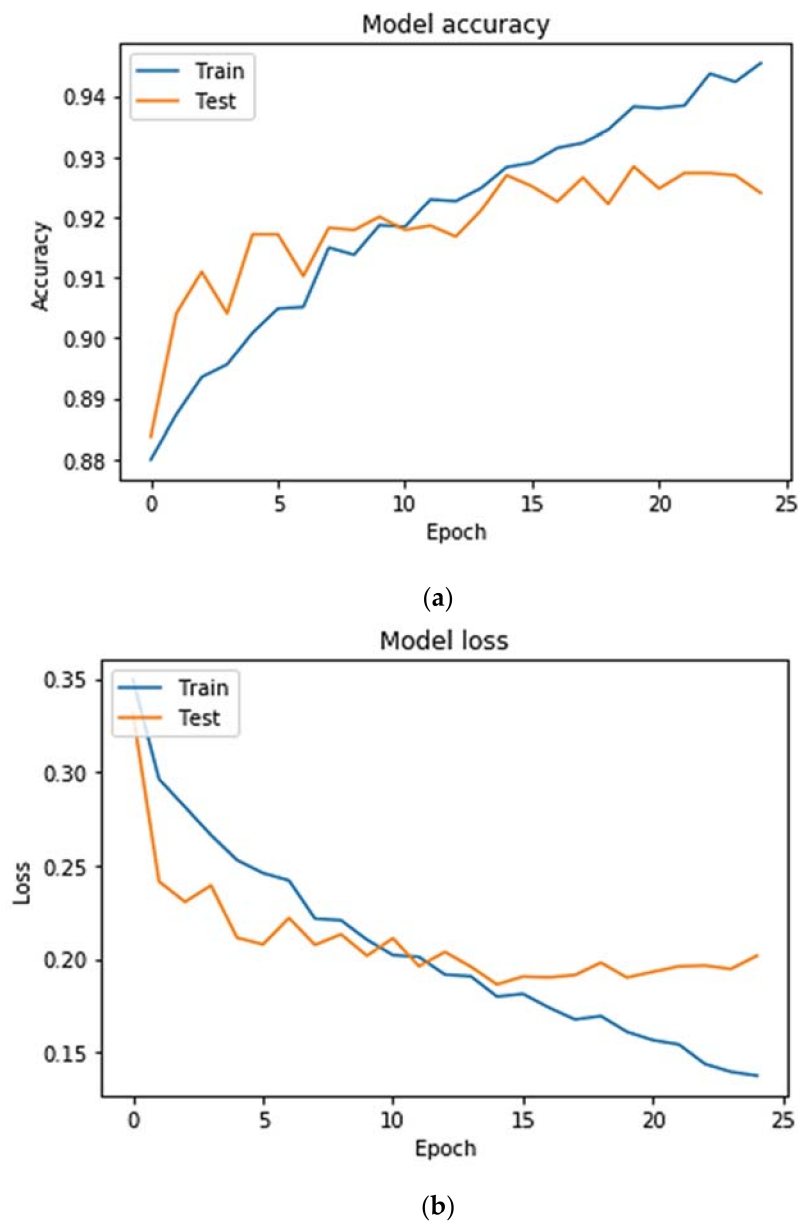
(**a**)

(**b**)

Figure 3. (**a**) LSTM training accuracy graph and (**b**) LSTM model loss graph.

In Figures 4b and 5b, confusion matrix graphs are plotted for both configurations to determine the overall performance. Each matrix row represents actual class instances, and each column represents the predicted class instances. The confusion matrix is shaded for quick understanding of the model's performance. The first group is named as true positive values, which means that the model predicts positive and the reviews are indeed positive. The second group represents the false positive values, which depicts that the model predicts positive, but the reviews are actually negative. The third group represents the false negative values, the model predicted the positive, but the reviews are actually not positive. The last group, shaded navy blue, represents the true negative values, which means the model predicted the reviews as negative and are indeed negative. These values are further employed for calculating the accuracy and other performance metrics. The CNN model achieved 88.91% accuracy, 88.91% specificity, 17.85% F1-score, and 57.89% sensitivity. LSTM achieved 91.69% accuracy, 92.81% specificity 56.67%, F1-score, and 76.95% sensitivity.
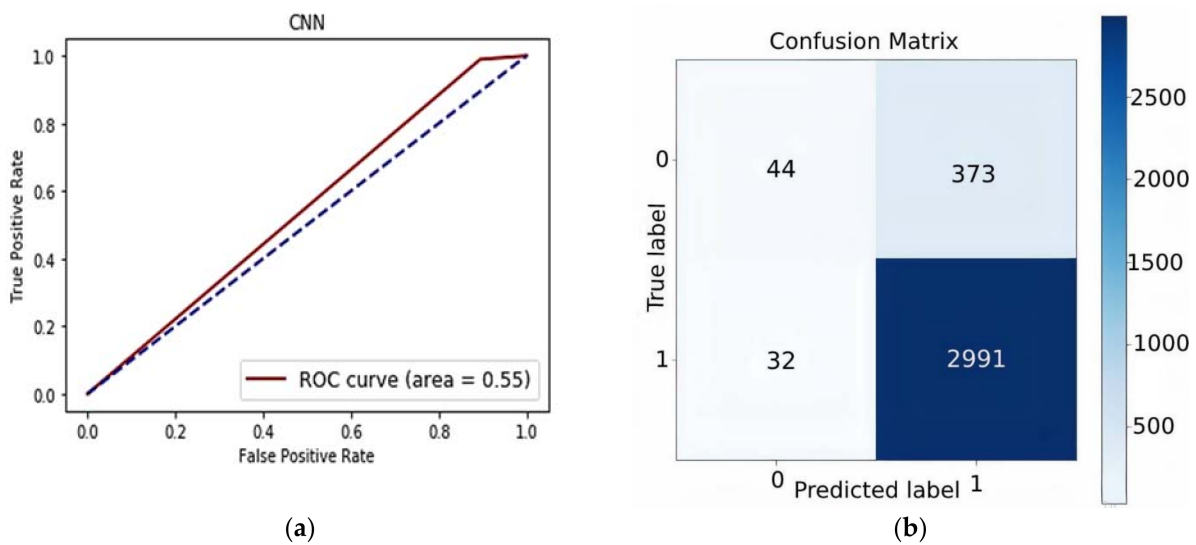
(**a**)

(**b**)

Figure 4. (**a**) AUROC curve of the CNN-based model and (**b**) the confusion matrix of proposed CNN-based model.
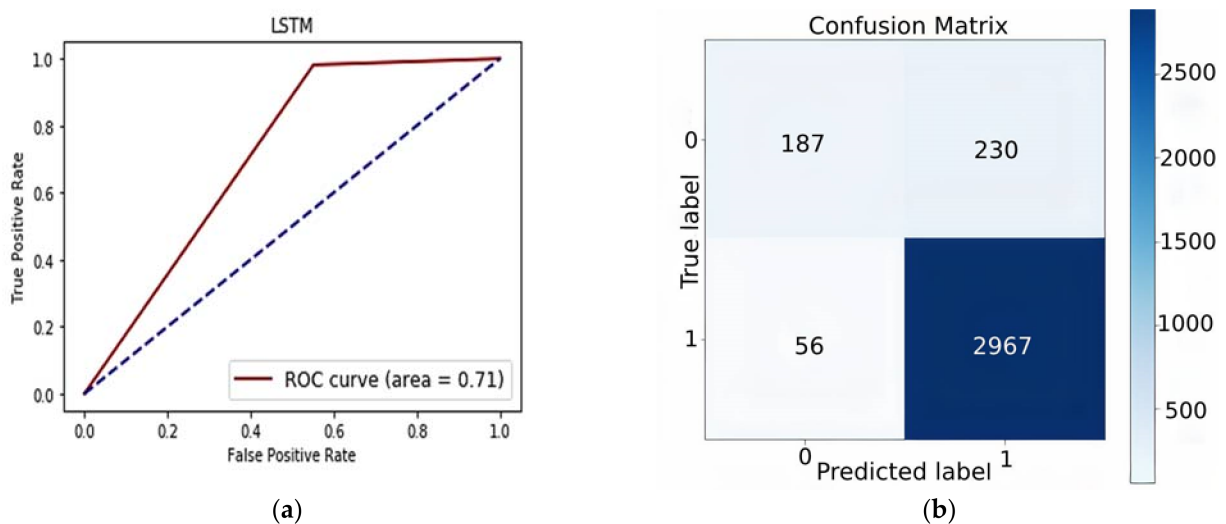

(**a**)

(**b**)

Figure 5. (**a**) AUROC curve of LSTM-based model and (**b**) the confusion matrix of the LSTM-based model.

To measure the performance of the proposed model, AUROC graphs are presented in Figures 4 and 5, respectively. AUROC is measured to determine the distinguishable property of the classes. The values of AUC vary from 0 to 1. When the model achieves a value near 0, it indicates that the model has the worst ability to distinguish between the classes. An excellent model has an AUC near 1, which shows that the model has separability ability. In Figure 5a, LSTM obtained an AUC value of 0.71, which shows that model has the ability to distinguish between the polarity of classes.

A comprehensive performance evaluation report of CNN and LSTM is presented in Table 1. By comparing the performance of CNN with LSTM, it is clear that the LSTM outperforms CNN. CNN is good at learning the local features of the words in the corpus; whereas, LSTM learns the temporal dependencies of the text. Therefore, LSTM improved the performance of the semantic analyzer and achieved 91.69% accuracy.

Table 1. Performance evaluation report.

Model	Accuracy	Specificity	Recall	F1-Score
LSTM	91.69%	92.81%	76.95%	56.67%
CNN	88.23%	88.91%	57.89%	17.85%

Cross Comparisons

Organizations vigorously observe performance of "company's offerings" in the market to make efficient decisions. Cross comparisons enable organizations to compare data classes relative to feature variables. In order to visualize the traits of the dataset, the relationship between four feature class variables was determined. Figure 6 represents comparison of product categories on the basis of average ratings received from online reviewers. Three product categories—trend, chemises, and bottoms—received below average ratings. Nine product categories—knits, formal dresses, swim suits, blouses, leg wear, sleep wear, sweaters, and shorts—received moderate ratings ranging from 4.3 to 4.4. The remaining products—fine gauges, skirts, jackets, pants, layering wears, lounge wears, intimates, and jeans—are top performing categories of the organization. Similarly, Figure 7 represents average age of the users related to each product category. As shown in Figure 7, it can be concluded that mature users, with ages ranging from 35 to 45 years, are the target segment of the organization. Causal bottoms is the only product category which targets young consumers, i.e., those having an average age of 25 years.
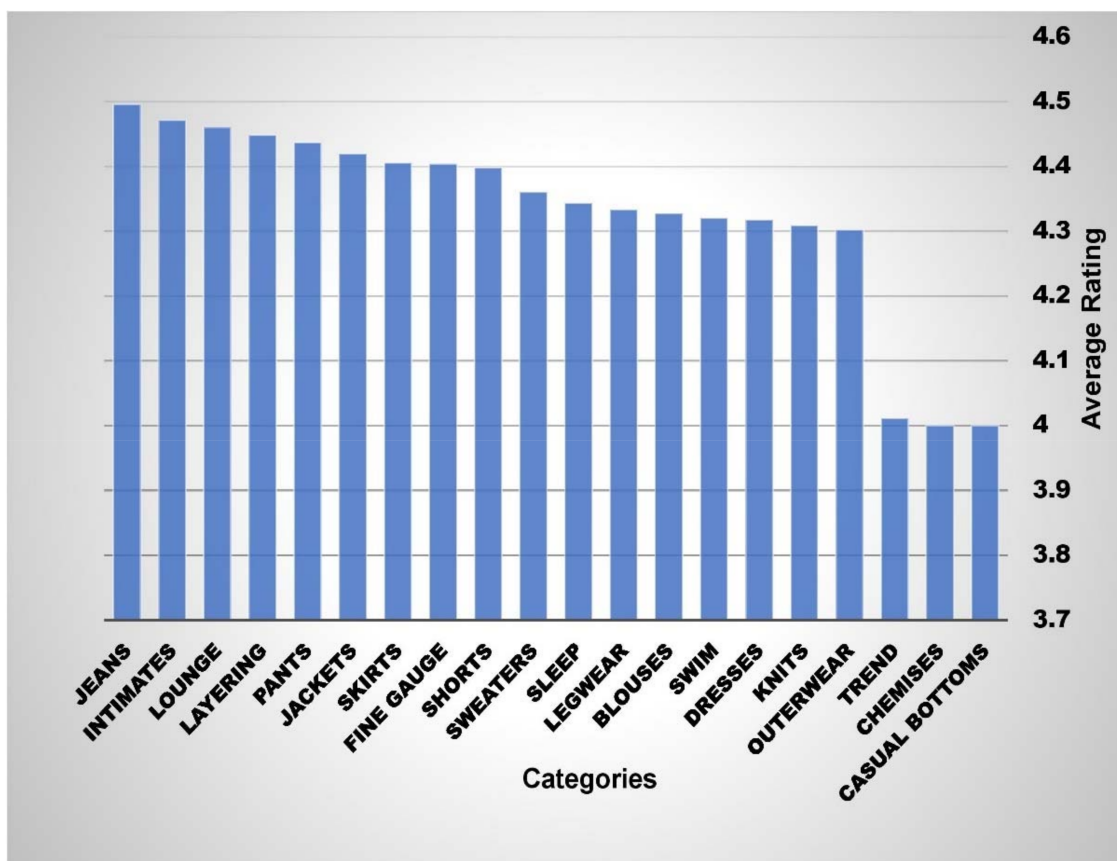

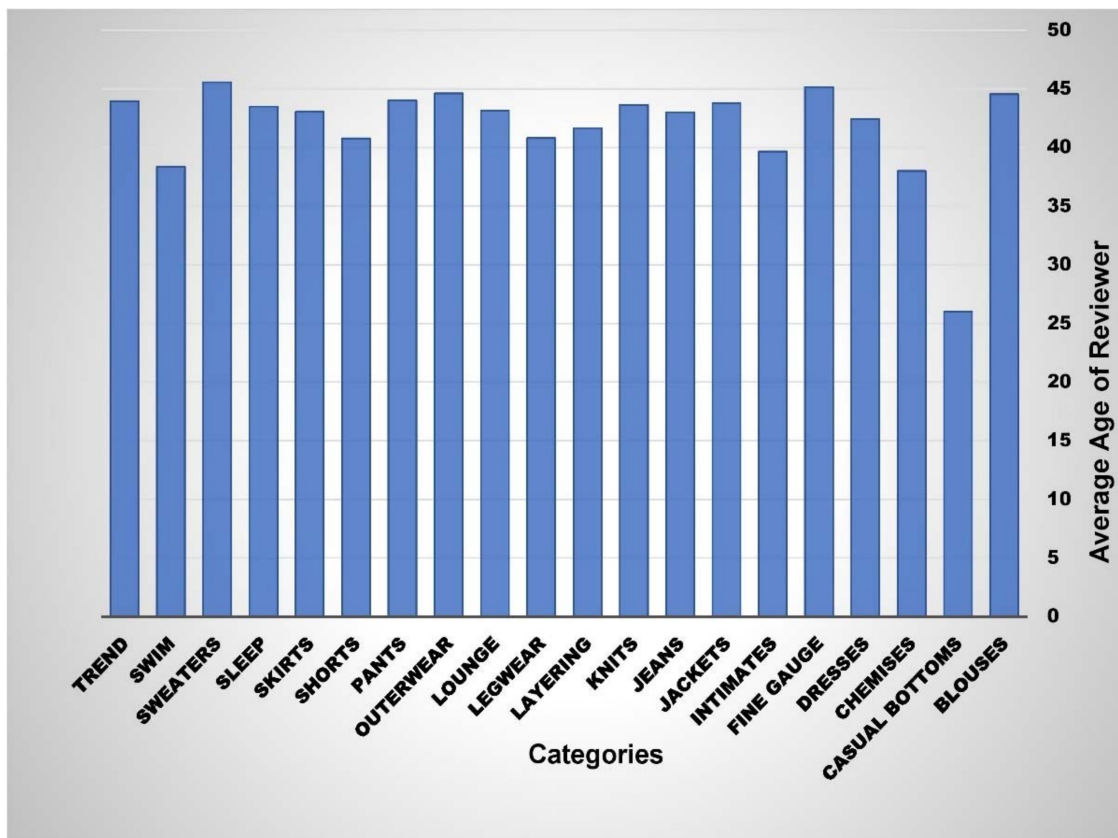**Figure 6.** Product categories comparison based on ratings.

Figure 7. Product categories comparison based on age.

5. Discussion

Sentiment analysis of eWOM offers valuable insights that help organizations to develop market intelligence. eWOM not only helps in understanding market needs, but also builds the organization's reputation. Traditional methods, i.e., survey data analysis and experimental research designs, are commonly used for sentiment analysis. These methods are inefficient, costly, and have many inherited drawbacks (e.g., sample biasness). The major problem in sentiment analysis is data handling because of the high speed, frequency, and volume of eWOM. This study proposed using latest neural network techniques for sentiment analysis, which are robust, efficient, accurate, and cost effective. CNN and LSTM techniques were used to perform sentiment analysis on real-time data. Findings suggest that LSTM is a better technique for sentiment analysis because of its high accuracy, specificity, recall, and F1-score. This study achieved 91.6 percent accuracy, which is very impressive; studies shows that even 52 percent accuracy is considered acceptable in behavioral sciences [43].

Machine learning techniques and deep learning techniques are rooted in the artificial neural network domain. Deep learning techniques are considered more effective in Natural Language Processing Techniques (NLP) as compared to machine learning techniques. Deep learning techniques are effective and efficient in processing extensively large data. A deep learning model contains many layers and each layer processes and transmits information to successive layers; thus, deep learning models do not need direct human supervision. Deep learning architectures are robust in nature and capable of handling data complexities. Furthermore, these models automatically infer hierarchical structure, which is effective in classification and prediction of tasks. Therefore, this study used CNN and LSTM deep learning techniques.

Artificial neural network techniques have been used in text analysis [24,25,44,45], and a few studies proposed the use of CNN and LSTM networks. Thus, it would be mandatory to compare the results of the proposed model with previous models. In this

regard, the proposed work is compared with three hybrid deep-learning models. The VD-CNN is actually a CNN-LSTM hybrid model that uses a k-fold approach. Five folds were selected during the training phase; the models started overfitting in the first fold and substantially led toward degraded accuracy [46]. Another research study [25] used a combination of CNN and LSTM on three different datasets. The accuracy of all three models is noted in Table 2. By observing the values presented in Table 2, it is evident that proposed model achieved better values not only in terms of accuracy but also in terms of specificity and recall. The model shows some degraded values in the F1-score. F1-score is also known as positive predictive value. F1-score mainly depends on the class ratio of the dataset; therefore, the value achieved may vary in different class ratio as the trend of varying values can be observed in Table 2. Overall, the LSTM model achieved better performance measurement parameters compared to the other machine learning and deep learning models. Thus, for the selected dataset, the proposed LSTM model performs better than the other hybrid models.

Table 2. Comparison with previous artificial neural network models.

Model	Year	Accuracy	Specificity	Recall	F1-Score
Machine Learning Models					
Proposed Model	**–**	**0.916**	**0.928**	**0.76**	**0.57**
Naïve Bayes [24]	2019	0.43	0.84	0.38	0.52
Decision Tree [24]	2019	0.80	0.88	0.88	0.88
Decision Tree Regression [24]	2019	0.79	0.89	0.86	0.87
Random Forest [24]	2019	0.86	0.87	0.97	0.92
Random Forest Regression [24]	2019	0.67	0.93	0.65	0.76
Ada Boost [24]	2019	0.86	0.91	0.92	0.91
Logistic Regression [24]	2019	0.88	0.90	0.95	0.93
KNN [24]	2019	0.83	0.83	0.99	0.91
Logistic Regression [25]	2020	0.63	0.58	0.63	0.58
Random Forest [25]	2020	0.58	0.51	0.59	0.48
Stochastic Gradient Descent [25]	2020	0.63	0.58	0.63	0.59
Support Vector [25]	2020	0.63	0.58	0.63	0.58
Voting Classifiers [25]	2020	0.63	0.58	0.63	0.58
Deep Learning Models					
Proposed model	**–**	**0.916**	**0.928**	**0.76**	**0.57**
VD-CNN [46]	2016	0.319	0.43	0.28	0.34
CNN–BiLSTM [44]	2017	0.614	0.32	0.22	0.25
CNN-LSTM [25]	2020	0.781	0.74	0.85	0.79

Cross comparisons provide depth of information regarding customers' emotion, particularly related to desired features that help decision makers. In this dataset, we provide cross comparison on the basis of product categories and gender. Findings suggest that organizations need to develop strategies for under-performing categories of causal bottoms, trend, and chemises. Furthermore, we found that organizations should target the customer group having ages ranging from 35 to 45 years. Similarly, organizations can develop datasets on the basis of demographics and product categories, or can even classify on the basis of brands.

6. Conclusions

AI neural network techniques provide efficient and insightful sentiment analysis on real-time data. This study applied neural network techniques (CNN and LSTM) and achieved good scores for performance measurement parameters of sentiment analysis; therefore, artificial neural network techniques are highly recommended for sentiment analysis of online reviews. The main objective of this study was to compare performance of artificial neural network techniques in real-time data. Results clearly established the

superiority of the LSTM technique over the CNN technique for sentiment analysis. LSTM sentiment analysis achieved accuracy of 91.69%, which is considered impressive in behavioral sciences.

Sentiment analysis of eWOM is challenging for organizations because of high volume and complexity of the data. Although some previous studies utilized machine learning techniques for sentiment analysis, there are some inherited drawbacks in machine learning techniques. The deep learning techniques proves its superiority over machine learning techniques by efficiently processing extensive data and automatic feature extraction from the raw data. This study found that the LSTM model proposed is efficient, accurate, and robust. Moreover, LSTM has a special ability to efficiently process the sequential text data due to presence of internal memory that memorizes the long-distance dependency of the text. It remembers and makes connections between the data that helps to make better and more precise predictions about the unseen data. Thus, LSTM is considered as the best suited algorithm for natural language processing application as it has better understanding of sequence and context of the data. This study presented the architectural design and complete procedural guidelines for the application. It is highly recommended to apply LSTM technique for text-based user-generated content.

Despite theoretical and practical contributions of this study, there are a few limitations. The study mainly focuses on detection and classification of comments on the basis of underlying emotions, and does not consider textual analysis. Future studies may consider textual analysis techniques (i.e., word frequencies analysis, thematic analysis, word correlations, and networks) in order to develop in-depth understanding of emotions. Data were gathered from a single platform and eWOM were related to post-purchase experiences. Thus, future studies may integrate data from various platforms (i.e., Facebook, Twitter, web portals, and blogs) for sentiment analysis. Future studies may also consider bi-LSTM for sentiment analysis and compare the findings with this study.

References

1. Gopi, A.P.; Jyothi, R.N.S.; Narayana, V.L.; Sandeep, K.S. Classification of tweets data based on polarity using improved RBF kernel of SVM. *Int. J. Inf. Technol.* **2020**. [CrossRef]
2. Gensler, S.; Völckner, F.; Egger, M.; Fischbach, K.; Schoder, D. Listen to your customers: Insights into brand image using online consumer-generated product reviews. *Int. J. Electron. Commer.* **2015**, *20*, 112–141. [CrossRef]
3. Arora, A.S.; Sanni, S.A. Ten years of 'social media marketing'research in the Journal of Promotion Management: Research synthesis, emerging themes, and new directions. *J. Promot. Manag.* **2019**, *25*, 476–499. [CrossRef]
4. Hennig-Thurau, T.; Gwinner, K.P.; Walsh, G.; Gremler, D.D. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? *J. Interact. Mark.* **2004**, *18*, 38–52. [CrossRef]
5. Shin, E.; Chung, T.; Damhorst, M.L. Are negative and positive reviews regarding apparel fit influential? *J. Fash. Mark. Manag. Int. J.* **2020**. [CrossRef]

6.   Roy, G.; Datta, B.; Basu, R. Effect of eWOM valence on online retail sales. *Glob. Bus. Rev.* **2017**, *18*, 198–209. [CrossRef]

7.   Lee, C.; Xun, X.; Chia-Chun, L. Using Online User-Generated Reviews to Predict Offline Box-Office Sales and Online DVD Store Sales in the O2O Era. *J. Theor. Appl. Electron. Commer. Res.* **2019**, *14*. [CrossRef]

8.   Sijoria, C.; Mukherjee, S.; Datta, B. Impact of the antecedents of eWOM on CBBE. *Mark. Intell. Plan.* **2018**, *36*, 528–542. [CrossRef]

9.   Crijns, H.; Cauberghe, V.; Hudders, L.; Claeys, A.-S. How to deal with online consumer comments during a crisis? The impact of personalized organizational responses on organizational reputation. *Comput. Hum. Behav.* **2017**, *75*, 619–631. [CrossRef]

10.   Bhandari, M.; Rodgers, S. What does the brand say? Effects of brand feedback to negative eWOM on brand trust and purchase intentions. *Int. J. Advert.* **2018**, *37*, 125–141. [CrossRef]

11.   Wu, T.; Yang, X. Electronic word of mouth in online social networks. In Proceedings of the 2010 Second International Conference on Communication Systems, Networks and Applications, Hong Kong, China, 29 June–1 July 2010; Volume 2, pp. 249–253.

12.   De Albornoz, J.C.; Plaza, L.; Gervás, P.; Díaz, A. *A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 55–66.

13.   Clavel, C.; Callejas, Z. Sentiment analysis: From opinion mining to human-agent interaction. *IEEE Trans. Affect. Comput.* **2015**, *7*, 74–93. [CrossRef]

14.   Abiodun, O.I.; Jantan, A.; Omolara, A.E.; Dada, K.V.; Mohamed, N.A.; Arshad, H. State-of-the-art in artificial neural network applications: A survey. *Heliyon* **2018**, *4*, e00938. [CrossRef]

15.   Capizzi, G.; Sciuto, G.L.; Napoli, C.; Woźniak, M.; Susi, G. A spiking neural network-based long-term prediction system for biogas production. *Neural Netw.* **2020**, *129*, 271–279. [CrossRef]

16.   Nawaz, F.; Ramzan, M.; Mehmood, K.; Khan, H.U.; Khan, S.H.; Bhutta, M.R. Early Detection of Diabetic Retinopathy Using Machine Intelligence through Deep Transfer and Representational Learning. *CMC Comput. Mater. Contin.* **2021**, *66*, 1631–1645.

17.   Borovkova, S.; Tsiamas, I. An ensemble of LSTM neural networks for high-frequency stock market classification. *J. Forecast.* **2019**, *38*, 600–619. [CrossRef]

18.   Capizzi, G.; Sciuto, G.L.; Woźniak, M.; Damaševicius, R. *A Clustering Based System for Automated Oil Spill Detection by Satellite Remote Sensing*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 613–623.

19.   Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; pp. 1–6.

20.   Hyun, D.; Park, C.; Yang, M.-C.; Song, I.; Lee, J.-T.; Yu, H. Target-aware convolutional neural network for target-level sentiment analysis. *Inf. Sci.* **2019**, *491*, 166–178. [CrossRef]

21.   Zhang, Y.; Tiwari, P.; Song, D.; Mao, X.; Wang, P.; Li, X.; Pandey, H.M. Learning interaction dynamics with an interactive LSTM for conversational sentiment analysis. *Neural Netw.* **2020**, *133*, 40–56. [CrossRef]

22.   Alayba, A.M.; Palade, V.; England, M.; Iqbal, R. *A Combined CNN and LSTM Model for Arabic Sentiment Analysis*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 179–191.

23.   Wang, Y. Building the linkage between project managers' personality and success of software projects. In Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement, Lake Buena Vista, FL, USA, 15–16 October 2009; pp. 410–413.

24.   Dey, U.K.; Noor, A. Comparative Exploration Of Prediction Algorithms For Sentiment Analysis Using NLP. In Proceedings of the 2019 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 23–25 January 2019; pp. 1–6.

25.   Umer, M.; Ashraf, I.; Mehmood, A.; Kumari, S.; Ullah, S.; Sang Choi, G. Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model. *Comput. Intell.* **2021**, *37*, 409–434. [CrossRef]

26.   Festinger, L. *A Theory of Cognitive Dissonance*; Stanford University Press: Palo Alto, CA, USA, 1957; Volume 2, ISBN 0-8047-0911-4.

27.   Carbonell, J.G. *Subjective Understanding: Computer Models of Belief Systems*; Yale Univ New Haven Conn Dept of Computer Science: New Haven, CT, USA, 1979.

28.   Wilks, Y.; Bien, J. Beliefs, points of view, and multiple environments. *Cogn. Sci.* **1983**, *7*, 95–119. [CrossRef]

29.   Hearst, M.A. Automatic acquisition of hyponyms from large text corpora. In Proceedings of the Coling 1992 volume 2: The 15th international conference on computational linguistics, Nantes, France, 23–28 August 1992.

30.   Wiebe, J.M. Tracking point of view in narrative. *arXiv* **1994**, arXiv:Cmp-Lg9407019.

31.   Sack, W. On the computation of point of view. In Proceedings of the AAAI, Seattle, DC, USA, 31 July–4 August 1994; p. 1488.

32.   Kantrowitz, M. Method and Apparatus for Analyzing Affect and Emotion in Text. U.S. Patent No 6,622,140, 16 September 2003.

33.   Hu, M.; Liu, B. Mining and summarizing customer reviews. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, DC, USA, 22–25 August 2004; pp. 168–177.

34.   Wright, S.; Calof, J.L. The quest for competitive, business and marketing intelligence. *Eur. J. Mark.* **2006**, *40*, 453–465. [CrossRef]

35.   Liang, T.-P.; Li, X.; Yang, C.-T.; Wang, M. What in Consumer Reviews Affects the Sales of Mobile Apps: A Multifacet Sentiment Analysis Approach. *Int. J. Electron. Commer.* **2015**, *20*, 236–260. [CrossRef]

36.   Schweidel, D.A.; Moe, W.W. Listening in on social media: A joint model of sentiment and venue format choice. *J. Mark. Res.* **2014**, *51*, 387–402. [CrossRef]

37.   Amblee, N.; Bui, T. Can brand reputation improve the odds of being reviewed on-line? *Int. J. Electron. Commer.* **2008**, *12*, 11–28. [CrossRef]

38. Colleoni, E.; Arvidsson, A.; Hansen, L.K.; Marchesini, A. Measuring corporate reputation using sentiment analysis. In Proceedings of the 15th International Conference on Corporate Reputation: Navigating the Reputation Economy, New Orleans, LA, USA, 18–20 May 2011.

39. Aakash, A.; Tandon, A.; Gupta Aggarwal, A. How features embedded in eWOM predict hotel guest satisfaction: An application of artificial neural networks. *J. Hosp. Mark. Manag.* **2021**, 1–22. [CrossRef]

40. Ahmad, S.N.; Laroche, M. Analyzing electronic word of mouth: A social commerce construct. *Int. J. Inf. Manag.* **2017**, *37*, 202–213. [CrossRef]

41. Chen, Y.; Xie, J. Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Manag. Sci.* **2008**, *54*, 477–491. [CrossRef]

42. Moral-Cuadra, S.; Solano-Sánchez, M.Á.; López-Guzmán, T.; Menor-Campos, A. Peer-to-Peer Tourism: Tourists' Profile Estimation through Artificial Neural Networks. *J. Theor. Appl. Electron. Commer. Res.* **2021**, *16*, 1120–1135. [CrossRef]

43. Agarap, A.F. Women's E-Commerce Clothing Reviews. 4 February 2018. Available online: https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews (accessed on 15 March 2021).

44. Rantanen, A.; Salminen, J.; Ginter, F.; Jansen, B.J. Classifying online corporate reputation with machine learning: A study in the banking domain. *Internet Res.* **2019**, *30*, 45–66. [CrossRef]

45. Cliche, M. BB_twtr at SemEval-2017 task 4: Twitter sentiment analysis with CNNs and LSTMs. *arXiv* **2017**, arXiv:170406125.

46. Conneau, A.; Schwenk, H.; Barrault, L.; Lecun, Y. Very deep convolutional networks for text classification. *arXiv* **2016**, arXiv:160601781.