Journal of Scientific & Industrial Research
Vol. 82, January 2023, pp. 120-127
DOI: 10.56042/jsir.v82i1.69937

Deep Learning Hybrid Approaches to Detect Fake Reviews and Ratings

N Deshai* & B Bhaskara Rao
Department of C.S.E.Gitam IT, Visakhapatnam, 530 045, Andhra Pradesh, India

Received 16 July 2022; revised 26 September 2022; accepted 06 October 2022

Nowadays, online reviews and ratings are the most valuable source of word-of-mouth, voice-of-customer, and feedback, also customers can make purchasing decisions on what to buy, where to buy, and what to select. Genuine online reviews are becoming popular, but unfortunately, we have an issue that might only sometimes be unbiased or accurate. Because most of the reviews are fake reviews and ratings, these could mislead innocent customers and highly influence customers' purchasing decisions in the wrong manner. This paper's primary goal is to accurately detect fake reviews and what is the main difference between them. The secondary goal is to detect fake ratings and actual ratings-based reviews across the online platform, especially Amazon datasets. The Paper proposes two novel deep-learning Hybrid techniques: CNN-LSTM for detecting fake online reviews, and LSTM-RNN for detecting fake ratings in the e-commerce domain. Both Hybrid models can outperform and achieve better performance with the most advanced word embedding techniques, Glove, and One hot encoding techniques. As per the experimental results, the first technique efficiently detects fake online reviews with the highest prediction accuracy. The second hybrid model is better than the existing models that detect fake online ratings with the most excellent precision of 93.8%. The experimental research efficiently revealed that the CNN-LSTM and LSTM-RNN methods are more efficient and practicable and might be better suited for optimal results and maximizing the efficiency of fake online review detection.

Keywords: CNN-LSTM, Glove, LSTM-RNN, One hot encoding

Introduction

In Today's digital world online reviews becoming essential to our day-to-day activities, and their value, importance, and reliability tend to influence people's purchasing decisions. However, their success and failure depend entirely on positive online reviews and high ratings because most of the reviews provide detailed information about various products and services posted by customers regarding their experiences, opinions, suggestions, and feedback.[1] Nowadays, a single online review plays a significant role in electronic business, deciding sales increase or decrease. Therefore, customers read a few online reviews before making a purchase decision for online products and services. But all online reviews and ratings are not positive because most online reviews are fake and harmful, which are available more in terms of text and numerical ratings. Still, fake online reviews have constantly raised a big challenge. There is a wide range of estimates for the percentage of fabricated reviews, ranging from 16.1%, 20%, 25%, and 33.3%. As early as 2012, around 10.3% of online products were found to have been the targets of manipulating reviews.[2] Therefore, decisions of customer for purchasing depend not only on money or merchant but also on online reviews and ratings. With the fast switch from conventional to digital dealing, which has transformed how individuals purchase, the matter of word-of-mouth couldn't be shifted down or ignored. Amazon, Trip Advisor, Yelp, Google play store, and IMDB provide platforms for customers to provide constructive feedback regarding their products and services.[3] This compel the customers to purchase 5-star-rated products and services over four and 3-star-rated products even if the price difference is 20 to 94 percent. Furthermore, a high numerical rating encourages customers to buy products and services confidently.[4] Finally, each customer can trust the highest rating because a customer saves time and effort while making a purchasing decision, but many highest ratings are fake. Star-rating systems were introduced in a paper in 2013.[(5,6)] We want to know how people utilize product reviews and then help them during purchasing. Customers need access to reviews and rating websites before making a purchase.

E-commerce enterprises have achieved popularity, credibility, product quality, and consumer recommendations as shown in Fig. 1. Fake reviews might be difficult to spot and detect, but with these seven pointers, we can navigate the sea of

*Author for Correspondence
E-mail: desaij4@gmail.com

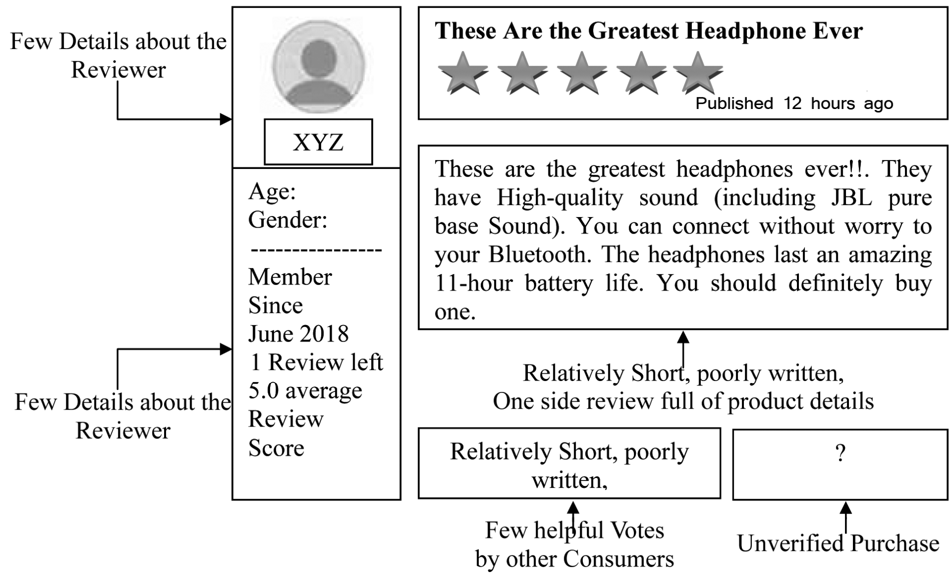

Fig. 1 — Fake review with detailed information

testimonials when doing your online shopping, as shown in Fig. 1. Look at the reviews' dates, 1. Look at the wording used in the reviews, 2. Be wary of evaluations using the same terminology, 3. Check out the reviewer's profile if you can, 4. Social media reviews should be avoided, 5. Check to see if the purchaser of the reviewer's product was confirmed, 6. There is a lack of specificity, 7. Finally, you should avoid anything that sounds too nice to be true. According to Gartner, 40% of business concepts fail to grasp their intended results due to poor data quality. Detecting fake reviews using deep learning techniques and hybrid models provides good outcomes. However, closely focused studies suggest that neural networks may be critical in this sector.[7] Conventional neural networks are the best for significantly training the data by a neural network, and recurrent networks have been recommended. Because this technique is necessary to differentiate between fake and genuine reviews, but this challenge is framed as a binary classification issue.

Related Work

Nowadays, customer online reviews and ratings are primary elements because every customer might look at and analyze before purchasing any product, item, or service. In addition, these online reviews are the most significant priority for current consumers as they can find insights to make purchasing decisions fast based on others' experiences over online platforms like Amazon, Yelp, Google play store, Trip Advisor, and social media. Therefore, Fake reviews drastically decide and significantly impact businesses' success and failures. Unfortunately, many online reviews encourage sellers to take advantage of client feedback and mislead them to boost their sales.[8] Therefore, fake review detection has become a significant challenge and a new research field in the current era. The digital buyer's volume will likely reach 2.14 billion by the decade's end. Text reviews have a higher market share than images or videos, at 43%. Up to 89% of customers in the worldwide marketplace could consult customer reviews before making a purchase decision. Nowadays, online customer reviews could influence 93% of buyers' purchasing decisions. More than a third of consumers read 2 to 3 reviews before making a purchase. Most customers believe a company's rating should have at least 50 reviews to be considered reliable. It's estimated that just half of the small business owners have enough time to focus on their web presence. Up to 70 percent more customers will buy from online platforms that publish customer reviews on their websites, according to new research.[9] Online reviews can significantly impact an organization's income in e-commerce. Since 2007, the analysis of review spamming has studied the problem of detecting fake reviews. For example, approximately 455 million users visit TripAdvisor each month, and 602 million reviews on 7.5 million hotels, restaurants, attractions, and more are on the site. In 2018, about 710 million people from all over the world made their way to Europe, making it the most popular travel destination. During the past ten years, numerous notable breakthroughs have been achieved in the field of detecting automatic bogus reviews. Support Vector Machines (SVMs) and Neural Networks (NN.) are machine learning technologies that have developed a reputation for being excellent in detecting bogus reviews.[10] In recent polls, 94% of respondents said they were more likely to see clear of a brand or organization after reading an online review. A fake review can damage a company's reputation and lead to financial losses. Deep learning methods to discover fake ratings and reviews have promising potential.[11] Because it is necessary to differentiate between fake and genuine reviews, binary classification has become a significant challenge. The Hybrid (LSTM+CNN) model employing Camem-BERT accomplished the most excellent performance in classifying the online reviews of French with 93.7% accuracy.[12] The performance of DNN, CNN, GRN, and HAN on the Tweep-Fake Dataset achieved an 89.7% accuracy.[13] BLSTM-2DPooling has 88.3% accuracy, while the other combination achieved 89.5% on the SST2 database is BLSTM-2DCNN. On IMDB and yelp2015, the SR-LSTM model was 44.0% & 63.9% accurate while the SSR-LSTM model was 44.3% & 63.8% accurate. According to the findings, combined (CNN-BiLSTM) achieves the highest accuracy of 90.66%. Researchers have demonstrated that a deep learning model can outperform a classic machine learning model.[13] On two Arabic datasets, the combination (CNN-LSTM) achieved the highest accuracy (85.38% and 86.88%).

The following is a list of the primary contributions that this paper makes:

(1) New deep learning hybrid techniques: CNN-LSTM is suggested to predict fake reviews, and LSTM-RNN is recommended to detect fake ratings. LSTM is utilized for data prediction, and CNN is used for data extraction in this method. Online reviews and ratings can be fully utilized to achieve more accurate predictions.

(2) In this research, we compare the performance indices of CNN-LSTM with those of multilayer perceptron, CNN, RNN, and LSTM. The results significantly reveal that the LSTM-RNN hybrid model has great accuracy on prediction and is higher than appropriate for false rating predicting than CNN-LSTM.

Proposed Hybrid Deep Learning Framework

Pre-processing

Nowadays, online reviews are enormous, and they have a collection of noise like hyperlinks, HTML tags, unofficial comments, etc., and many words don't significantly impact the feeling of the review. Therefore, we keep text preprocessing to a minimum, and each online review is essential to extract more meaningful knowledge. To achieve this, we utilize standard python libraries to eliminate capitalizations, stop words, and punctuation, as shown in Fig. 2. Therefore, the text must be transformed into an appropriate format so deep learning methods can accurately understand, utilize, and efficiently classify each online review.

Feature Extraction

N-grams

These are very popular and useful in text classification tasks. N-grams are all types of ticket letters, numerals, and symbols. The most popular tokenization methods are whitespace, Expression, and unigram tokenization might be performed across the sentence, comment, or character levels.[14] For example, unigrams have only one token. At the same time, bigrams will have a variety of tickets in sequence. Similarly, trigrams will have a combination of three passes together.

Normalization

Normalization is a technique for dropping the number of unique tokens and eliminating variances in a text. As well as cleansing content by deleting useless data stemming and lemmatization are two strategies for normalizing. A lexeme denotes a basic form in linguists and Natural Language Processing.

Stemming

Mainly to convert words to their root form, that stemming can introduce ambiguity. However, the benefit of stemming is that it is typically robust to spelling errors, as the correct root may still be inferred correctly example, porter stemmer.

Lemmatization

Unfortunately, stemming isn't a great way to get things normalized because this can create meaningless terms that aren't in the dictionary. Lemmatization is

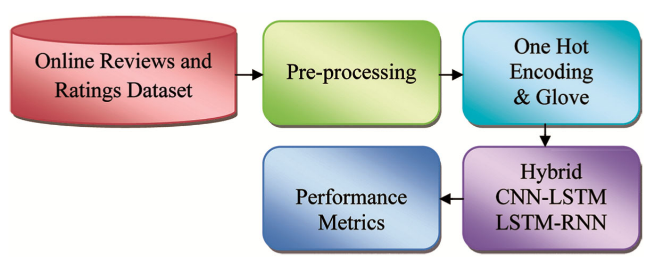

Fig. 2 — Proposed Deep learning hybrid methodology

the well-regulated elimination of a token's inflectional shape and changes into a lemma.[15]

Vectorization

To have a few distinct characteristics out of the text for the model to train on by transferring text to numerical vectors. Count vectorizer Term Frequency (TF) and Inverse Document Frequency (IDF) are both examples of standard vectorizing procedures.[16] To calculated word frequency from number of times it occurs in a sample and dividing the total number of words in that sample.[16] The IDF is calculated by taking the logarithm of the number of pieces included data set and dividing it by the number of examples in which the term is found. The mathematical equation for TFIDF is as follows:

$$idf(t)^n = \log\left[\frac{n}{df(t)}\right] + 1 \quad \ldots(1)$$

$$(\mathrm{tf} - \mathrm{idf}(\mathrm{t}, \mathrm{d}) = \mathrm{tf}\,(\mathrm{t}, \mathrm{d})\ \mathrm{idf}(\mathrm{t})) \quad \ldots(2)$$

Word2vec

One of the most popular tools in modern NLP and revealing hidden semantic relations. Users can select Word2Vec between the CBOW architectural style or the Continuous Skip-gram architectural style. The CBOW algorithm forecasts the vector of a word made independently based on the context vector of the words close to it. On the other hand, the skip-gram method makes predictions about the nearby context vector based on the word in the center. It is a prediction model and analyses the context of each words around the target words to establish what that word is.

One hot Encoding

It is a deep learning technique to be applied to sequential classification problems. The main strategy is to convert to a numerical vector. Essentially the representation of categorical variables as binary vectors to be more expressive and get a better prediction.

Glove

GloVe combines global statistics (word co-occurrence) and local statistics (local context information) to create word vectors. As a result, the glove assigns less weight to word pairings used frequently. Statistics could be utilized to determine how the words relate to one another. The solution may be parallelized more easily. It is an unsupervised method. It concentrates on word co-occurrences in the whole corpus. The Glove model is trained via least squares using the cost function:

$$J = \sum_{i=1, j=1}^{V} \mathrm{f}(X_{ij})\left(u_i^T \boldsymbol{v}_j - \log(X_{ij})\right)^2 \quad \ldots(3)$$

Hybrid Classifier

Deep learning neural network models are used for analyzing, identifying, and categorizing fraudulent reviews based on glove and one-hot-encoding elements of review text as fake or real review.[17] For example, these models may be found in the following section: These models are applied to evaluate reviews to establish whether they are genuine.[17] To be more exact, it suggests utilizing a combination of CNN-LSTM, and LSTM- RNN, which increases the accuracy of the recommended hybrid models for fake rating and fake review, as shown in Fig. 2.

CNN-LSTM Model

Embedding Layer: The CNN-LSTM model's primary layer performed a major role in significantly transforming every word into an actual-valued vector in the training dataset. Dropout Layer's main task is to keep away from the model overfitting. The primary function is deactivating arbitrary neurons in a review text sentiment word where every neuron denotes the dense exemplification. The CNN-LSTM convolution layer is employed to obtain features from the input matrix. It uses n convolutional filters to find convolutions for every series. In addition, Max Pooling Layer down samples input sequences' spatial dimensions as shown in Fig. 3. It examines each filter kernel's max input value. LSTM Layer can learn long-

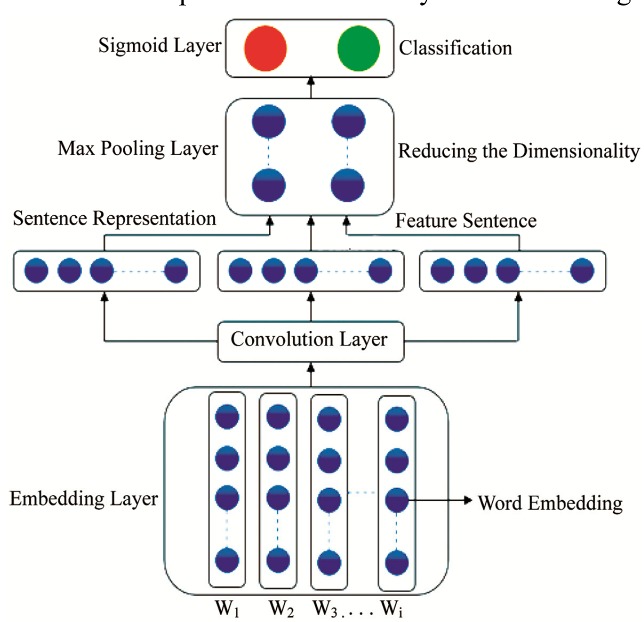

Fig. 3 — An instance of unreliable online reviews described by LSTM-RNN models

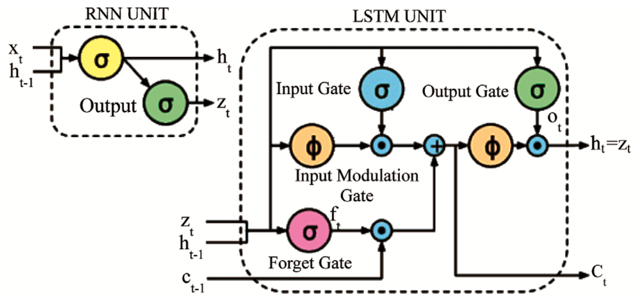

Fig. 4 — RNN – LSTM model

term reliance. More than a conventional LSTM, the extraction of features eliminates aggregate features. LSTM precalculated the input sequence before producing a network output. In every cell, four discrete computations are performed using four gates: input, forget, candidate, and output. In Fig. 4 LSTM's structure is shown. Where, sig and tanh are the sigmoid and tangent activation functions, X is the input data, W and b are the weight and bias factor, Ct is the cell state, c t is the candidate gate, and ht is the LSTM cell's output.

$$Ft = sig\,(Wfxt + Ufht - 1 + bf) \quad \ldots (4)$$

$$It = sig\,(Wixt + Uiht - 1 + bi) \quad \ldots (5)$$

$$Ot = sig\,(Woxt + Uoht - 1 + bo) \quad \ldots (6)$$

$$c\sim t = \tanh\,(wcxt + Ucht - 1 + bc) \quad \ldots (7)$$

$$Ct = (ftoct - 1 + itoc\sim t) \quad \ldots (8)$$

$$Ht = Oto \times \tanh\,(Ct) \quad \ldots (9)$$

$$Tanh\,(x) = 1 - e2x1/e2x \quad \ldots (10)$$

Dense Layer (Fully Connected Layer)

CNN-hidden LSTM's layer and five hundred twelve artificial neurons connect an entire network of neurons. This layer's function is the rectified linear unit:

Sigmoid Activation Function

Primary layer that recognizes output classes (positive or negative sentiment). The equation for the sigmoid function: (10)

$$S(x) = 1 + \frac{1}{1+e^{-x}} \quad \ldots (11)$$

RNNs obtain data from sequencing or time-series data. They can handle variable-size inputs, outputs, and sequences (like DNA sequences) or time-series data. RNN loops networks, LSTM is a form of recurrent neural network (RNN) and constructed utilizing Keras of TensorFlow, a deep learning toolkit and API written in python that enables defining and training deep neural network models.

Table 1 — The details of two dataset

Dataset	First Dataset	Second Dataset
Name of the Dataset	Hotel Dataset	Amazon_ Unclocked_Mobile
No. of Attributes	1. Head 2. Body 3. Label 4. Features	1. Product name 2. Brand Name 3. Price 4. Rating 5. Reviews 6. Review Votes
Total records	24903	413841
No. of Positives	800 267450	24903 268340
Feature extraction	Glove One hot Encoding	Glove One hot Encoding

Datasets

It shows the Dataset from the Amazon_ Unlocked_Mobile Dataset for the second hybrid model, CNN-RNN, to detect fake online ratings shown in Table 1. This Amazon_Unlocked_Mobile data set has six attributes with max_features = 20000, EMBEDDING_DIM = 100, VALIDATION_SPLIT = 0.2, maxlen = 30, batch size = 32 presents the datasets utilized in the research carried out to identify fake reviews and ratings. When utilizing Amazon Dataset, training a CNN can take anything from 280 to 706 seconds, with CNN requiring the least amount of time out of the three options. LSTM outperforms the other two models in the test performance shown in Table 2 by achieving the highest value from the lowest to the highest point on the scale. The information reflects the statistical averages of the characteristics established for 4790 fake and 4666 honest reviews.

Experimental Analysis

The recommended hybrid deep learning model is created with the help of the sequential model included in the Keras deep learning Python toolkit. This model was constructed by utilizing the library's resources.[17] The Sequential model includes these neural network layers: Keras embedding was the first layer of the ANN. This is the input layer, which uses pre-trained word embedding by delivering the embedding matrix. Training terminates after the model is trained with training data. Next, Conv1D, a one-dimensional CNN layer, extracts local features using 128 5-pixel filters. Rectified Linear Unit (ReLU) activation is employed again.[18]

Table 2 — Accuracy of a new hybrid model

Classification	Accuracy	Precision	Recall	F1-score
Word embedding model				
Linear SVM + TF IDF	92%	92%	90%	90%
Decision tree + TF IDF	92.80%	92.30%	92.30%	92.80%
1st Proposed Model CNN-LSTM +GloVe	93.07%	93.09%	93.07%	93.07%
2nd Proposed Model LSTM-RNN + One Hot Encoding	93.09%	93.09%	93.08%	93.08%

After this stage, CNN pools its effective feature vectors in a MaxPooling1D layer with a 2-pixel window. This reduces the number of feature vectors, parameters, and computations without affecting the network's efficiency. Finally, RNN LSTM is performed after pooled feature maps. This inputs trains the LSTM, which outputs long-term dependent attributes of the input feature maps while remembering them. The output is 32 bytes f(x) = x is Keras' built-in linear activation function. Finally, a Dense layer classifies the training feature vectors.[19] At this layer, the output space dimension is reduced to a single label (i.e., fake, or not fake). This network layer performs Sigmoid activation. Adam calculates the learning rate during each iteration of model training. Binary cross-entropy is used as the loss function and accuracy measures model performance. Each training batch has 64 participants over ten epochs.[20]

Metrics for Model Performance

Because favorable evaluations may significantly influence a company's bottom line, some dishonest sellers have begun producing fake product reviews to either promote their items or denigrate their competitors. From the confusion matrices, it is feasible to create a variety of performance measures by basing them on the rates of false-positive and false-negative items. The success of the CNN and LSTM models that have been presented for identifying and forecasting fraudulent and accurate reviews and therefore, using these measures, their ability to be utilized in such a manner was evaluated. Specificity, precision, recall, F1-score, and accuracy of an evaluation are calculated using confusion matrices. The confusion matrices used to determine the evaluation's specificity, precision, recall, F1 score, and accuracy are depicted in Table 2. These evaluation metrics are formulated as follows.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \text{X}100 \quad \ldots (12)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \text{X}100 \quad \ldots (13)$$

$$\text{Precision} \ \frac{\text{TP}}{\text{TP} + \text{FP}} \text{X}100 \quad \ldots (14)$$

$$\text{F1 score} = \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \times 10 \quad \ldots (15)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}} \times 100 \quad \ldots (16)$$

Comparative Analysis

In the modern world, a fake online review is a serious issue that has grown in scope and is challenging to detect. In this section, we will compare the results generated by the deep learning-based hybrid models CNN-LSTM and LSTM-RNN that were provided before with the results that the machine and deep learning-based models created. In addition, it compares the various transformation techniques, Glove, and one hot Encoding that was applied to convert the words of the text into numerical form vectors. This was done to understand better how the data was generated. For example, using the TF-IDF approach, the word in the review text is converted into a single-column vector. However, the Word2Vec method converts the term into N-dimensional vectors. The popular pre-trained word embeddings contain adequate vocabulary since they were trained with a large corpus to yield satisfactory results. Therefore, Glove is prevalent in all pre-trained word embedding techniques paired with CNN-LSTM to evaluate how well these methods identify false reviews and how well they perform in terms of giving decent results when processing text. At the same time, to improve performance, the hybrid model CNN-LSTM was created. CNN can extract valuable data properties, while LSTM can discover time series data dependencies.[20] This approach can increase the accuracy of fake review predictions.

Analytical characteristics may be extracted from both kinds of neural networks. In addition to this, convolutional neural networks are not very effective. CNN is a long-term memory network that is only capable of extracting local information and is unable to represent context information more effectively. Therefore, a text classification technique based on a

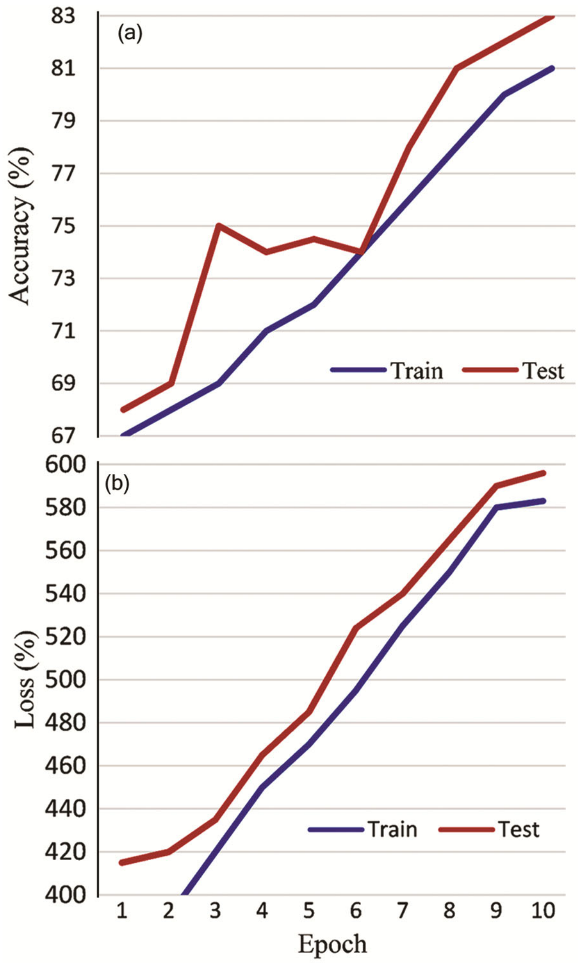

Fig. 5 — (a) Model accuracy; (b) Model loss

hybrid CNN-LSTM & LSTM-RNN model has been suggested. LSTM could obtain perspective dependencies, and the classification effect is substantial, considering the training time. Epochs = 10 Embedding = max_ features, 128) SpatialDropout1D rate = 0.2 LSTM 128, dropout = 0.2, recurrent_dropout = 0.2)) Dense nb_classes Activation sigmoid sgd = optimizers' (lr=0.10) as shown in Fig. 5(a) and Fig. 5(b).

The suggested Hybrid CNN-RNN technique outperforms all others in accuracy, precision, recall, and F1 score. Customer comments on Amazon.com are the number of the first Dataset in the final test. When utilizing Amazon Dataset, training a CNN can take anything from 280 to 706 seconds, with CNN requiring the least amount of time out of the three options. LSTM outperforms the other two models in the test performance shown in Fig. 5 (a) and Fig. 5 (b) by achieving the highest value from the lowest to the highest point on the scale. The confusion matrices are shown how well each model fared in predicting fake reviews, with the models that received higher scores performing more effectively. According to Table 2, the recommended models performed better than previous research based on the Amazon online review Dataset. An LSTM-RNN + One-hot model with a score of F1 is accurate, exact, and recallable to 90.95%. The CNN-LSTM + GloVe model finished in second place, with a precision score of 93.07% and accuracy, recall, and F1-score values of 93.09%. Gains of 90.08% or more in accuracy, recall, and F1-score for the CNN + Glove model, together with a 93.09% increase in precision.

Conclusions

In this paper, we proposed two hybrid models CNN-LSTM and LSTM-RNN for detecting online fake ratings and fake reviews. During the phase where the data is being preprocessed, we employed one-hot Encoding and Glove to transform each base into a matrix of the same size by padding it. According to our research findings, the proposed technique was effectively trained on the training dataset. Therefore, it had a performance superior to that of the previous models when applied to the test dataset. According to the findings, the deep learning hybrid CNN-LSTM and LSTM-RNN models have potentially been utilized and done online review detection and classification tasks with improved performance 93.07% by CNN-LSTM + GloVe and 93.09% by LSTM-RNN + One Hot Encoding 93.09%. Compared to other approaches, the results of our proposed strategy have proven to be successful.

References

1 He L, Wang X, Chen H & Xu G, Online spam review detection: a survey of the literature, *Human-Centric Intell Syst*, **2** (2022) 14–30, DOI: 10.1007/s44230-022-00001-3.

2 Halder S, Dutta S, Banerjee P, Mukherjee U, Mehta A & Ganguli R, A survey on online spam review detection, in *Emerg Techonol Data Min Inf Secur*, **1300** (2021) 717–724, DOI:10.1007/978-981-33-4367-2_68.

3 Ren Y & Zhang Y, Deceptive opinion spam detection using neural network, in *Proc COLING 2016, the 26th Int Conf Comput Linguist: Technical Papers,* (Osaka, Japan), 2016, 140–150, DOI:10.1007/s10489-022-03427-1.

4 Desale K S, Shinde S & Magar N A, Fake review detection with concept drift in the data: a survey, in *Proc Seventh Int Cong Inf Commun Technol, Lect Notes Netw Syst*, vol 448, edited by X S Yang, S Sherratt, N Dey & A Joshi (Springer, Singapore) (2022) 719–726, https://doi.org/10.1007/978-981-19-1610-6_63.

5 Jian Y, Chen X & Wang H, Fake restaurant review detection using deep neural networks with hybrid feature fusion method, in *Database Systems for Advanced Applications,*

DASFAA 2022, Lecture Notes in Computer Science, vol 13247, edited by, *et al*. (Springer, Cham) 2022, 133–148, https://doi.org/10.1007/978-3-031-00129-1_9.

6 Rehman A U, Malik A K, Raza B & Waqar A, A Hybrid CNN-LSTM model for improving the accuracy of movie reviews sentiment analysis, *Multimed Tools Appl*, **78(6)** (2019), 26597–26613, DOI:10.1007/s11042-019-07788-7.

7 Habbat N, Anoun H & Hassouni L, LSTM-CNN deep learning model for french online product reviews classification, in *Advanced Technologies for Humanity, ICATH 2021, Lecture Notes on Data Engineering and Communications Technologies*, vol 110, edited by R Saidi, B El Bhiri, Y Maleh, A Mosallam & M Essaaidi (Springer, Cham) 2022, 228–240, https://doi.org/10.1007/978-3-030-94188-8_22.

8 Singh M, Jakhar A K & Pandey S, Sentiment analysis on the impact of coronavirus in social life using the BERT model, *Soc Netw Anal Min*, **11** (2021) 1–11, DOI:10.1007/s13278-021-00737-7.

9 Tesfagergish S G, Damaševičius R & Kapočiūtė-Dzikienė J, Deep fake recognition in tweets using text augmentation, word embeddings and deep learning, in *Computational Science and Its Applications – ICCSA 2021. ICCSA 2021, Lecture Notes in Computer Science*, vol 12954, edited by *et al*. (Springer, Cham) 2021, 523–538, https://doi.org/10.1007/978-3-030-86979-3_37.

10 Sagnika S, Mishra B S P & Meher S K, An attention-based CNN-LSTM model for subjectivity detection in opinion-mining, *Neural Comput Appl*, **33** (2021) 17425–17438, https://doi.org/10.1007/s00521-021-06328-5.

11 Rehman U A, Malik A K, Basit R & Waqar A, A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis, *Multimed Tools Appl*, **78** (2019) 26597–26613, https://doi.org/10.1007/s11042-019-07788-7.

12 Alzahrani M E, Aldhyani T H H, Alsubari S N, Althobaiti M M & Fahad A, Developing an intelligent system with deep learning algorithms for sentiment analysis of e-commerce product reviews, *Comput Intell Neurosci*, **2** (2022) 1–10, DOI:10.1155/2022/3840071.

13 Jiang C, Zhang X & Jin A, Detecting online fake reviews via hierarchical neural networks and multivariate features, in *Neural Information Processing, ICONIP 2020, Lecture Notes in Computer Science*, vol 12532, edited by H Yang, K Pasupa, A CS Leung, J T Kwok, J H Chan & King I (Springer, Cham), 2020, 730–742, https://doi.org/10.1007/978-3-030-63830-6_61.

14 Alton Y K C & Xiaoyu C, Online helpful Lies: An empirical study of helpfulness in fake and authentic online reviews, in *Information for a Better World: Shaping the Global Future, iConference 2022, Lecture Notes in Computer Science*, vol 13192, edited by M Smits (Springer, Cham) 2022, 91–99, DOI:10.1007/978-3-030-96957-8_10.

15 Deshai N & Bhaskara Rao B, A detection of unfairness online reviews using deep learning, *J Theor Appl Inf Technol*, **100(13)** (2022) 4738–4779.

16 Deshai N, Sekhar B V D S, Prasad Reddy P V G D & Chakravarthy V V S S S, Processing real-world datasets using big data hadoop tools, *J Sci Ind Res*, **79(7)** (2020) 631–638.

17 Srivastava A, Gupta A & Anand R, optimized smart system for transportation using RFID technology, *Math Eng Sci Aerosp*, **12(4)** (2021) 953–965.

18 Raghavan R, Verma D C & Pandey D, Optimized building extraction from high-resolution satellite imagery using deep learning, *Multimed Tools Appl*, **81(2)** (2022) 42309–42323, DOI:10.1007/s11042-022-13493.

19 Bakshi G, Shukla R, Yadav V, Dahiya A, Anand R, Sindhwani N & Singh H, An optimized approach for feature extraction in multi-relational statistical learning, *J Sci Ind Res*, **80(6)** (2021) 537–542, DOI: 10.56042/jsir.v80i6.43632.

20 Sule O B, Ipek K & Sachin K M, Detecting fake reviews through topic modeling, *J Bus Res*, **149 (**2022) 884–900, DOI: 10.1016/j.jbusres.2022.05.081.