International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

FRARBiLSTM-A Novel Fake Review Authentication Model Using Afinn and Roberta

Vikas Attri[1], Isha Batra[2], Arun Malik[3], Vipin Kumar[4]

Submitted: 24/10/2023 **Revised:** 16/12/2023 **Accepted:** 26/12/2023

Abstract: Due to the rise in online transactions, fake customer review identification is attracting attention. Fake customer reviews are identified using features such as reviewer identification, product information, and review text. Recent research suggests that review semantics may be particularly pertinent for text classification. The reviewers' veiled feelings could also point to misleading information. Our neural network model combines word context, customer emotions, and the traditional bag-of-words to improve fake review detection. The algorithms use N-grams, dynamic word embeddings, and emotion indicators based on lexicons to learn document-level representation. We contrast the classification performance of the detection systems with several cutting-edge methods for fake review detection to demonstrate the value of the systems. No matter the sentiment polarity or product category, the suggested approaches on the present datasets outperform Afinn, RoBERTa, Ensemble, and hybrid models. The paper offers Hybrid/Ensemble-based strategies under the proposed model named FRARBiLSTM(Fake Reviews-AFINN RoBERTa using Bidirectional LSTM). This model performs better than previous classifiers in detecting false reviews with an accuracy of 97.31. When used with Ensemble and hybrid Learning, this model can exceed and attain superior performance compared to the most modern word embedding algorithms, particularly RoBERTa and AFINN.

Keywords*: Online Social Networks (OSN), E-commerce, Machine Learning, Word Embeddings, Ensemble Learning, Afinn, RoBERTa*

1. Introduction

Customers place a growing amount of faith in online product reviews due to the crucial information they supply [1]. The majority of online marketplaces give higher rankings to products with more positive reviews (an effect known as the "snowball effect"), which may benefit businesses that buy reviews. According to a meta-analysis of more than 20 empirical research [2], the volume and tone of reviews affect retail sales. High-involvement products can only be reviewed after consumption. Eighty percent of customers put the same amount of stock in online testimonials as they do in personal recommendations [3]. Fake reviews are becoming increasingly sought after by businesses. A company could get advantages from either fake positive or fake negative evaluations.

In fact, according to recent statistics, every third TripAdvisor review is fake [4]. The business is now concerned with online reviews. Therefore, platforms must identify and remove fake reviews and forbid fraudulent users to ensure fair competition. Manual or automatic detection of fake reviews is possible [5]. Manual fake review identification, however, is pricy, cumbersome, and unreliable [6]. Over the past ten years, automated bogus review detection has gotten better. Popular techniques for identifying fake reviews include SVMs and NNs [7, 8]. These techniques classify reviews as fake or genuine based on the review's text, user behaviour. A low false positive rate is significant because, in the absence of one, users of online platforms would not be able to read correct evaluations, and reliable people would face the consequences and lose interest in posting reviews. A list of words or phrases with weights (bag-of-words) or word categories (psycholinguistic or part-of-speech tagging) is produced by machine learning algorithms based on review content [9]. However, sparsity makes it challenging to explain customer reviews meaningfully. Using sentence representations, Ren and Ji [10] developed a gated recurrent NN model to detect false opinion spam. Semantically related words were mapped to word vectors using the continuous bag-of-words (CBOW) model [11, 12]. Thus, global semantic representation is feasible, resolving the data scarcity problem. CNN was incorporating sentence representations into a recommendation made by Li et al.[13].

Inspired by these cutting-edge models, we use word embeddings to describe customer feedback semantically [13, 10]. Since its word embeddings were trained on limited datasets of several hundred reviews, the CNN algorithm in [13] cannot identify fake customer reviews. Pre-trained word embeddings fared better in [10]. According to the authors, the CBOW model used in [10] cannot produce a generalisable context model [12]. In contrast to [10], we use a Skip- Gram Word2Vec model to construct word embeddings from customer reviews. [12]

[1,2,3] *Lovely Professional University,* [1,4]*Chandigarh University*
vikasmca09@gmail.com[1], isha.17451@lpu.co.in[2], arunmalikhisar@gmail.com[3], vksnp222@gmail.com[4]
** Corresponding Author Email: isha.17451@lpu.co.in*

Skip-Gram makes better use of word context than CBOW does. The above deep NN models solely consider word embeddings, neglecting consumer review mood indicators. [14] Review ratings rarely match the review content mood. Additionally, reviews with similar ratings might have very varied emotional strengths. Most notably, sentiment strength outperforms ratings in detecting bogus reviews [14]. We integrate word embeddings with bag-of-words and numerous lexicon-based emotion indicators to boost detection performance and overcome deep NN model issues. Recent research only considered positive and negative sentiments [15, 16, 17]. However, trust and aptitude are good indicators of online review helpfulness [18, 19]. The fake review detection algorithm uses deep learning to incorporate those emotion markers.

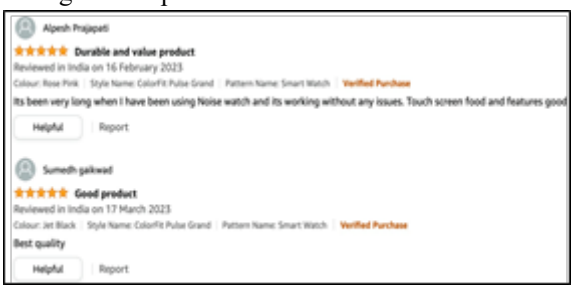


Fig. 1. Amazon Customer Reviews

Before making a purchase, customers can more accurately gauge the quality of a product by reading online consumer reviews (OCRs). In recent years, OCR has experienced increased levels of confidence. A recent poll found that roughly 80% of consumers trust OCRs as much as personal recommendations from friends or family and that over 90% read OCRs before making a purchase.

It is becoming increasingly usual for websites to provide fake reviews. Buying and selling false reviews can be profitable. However, their contributions to detecting bogus reviewers are limited. To begin, even though behavioural factors are necessary for detecting fake reviewers, most research concentrates on generating distinctive behavioural features, which requires a great deal of time-consuming and costly human labor and knowledge. Second, several text features such as n-grams, part-of-speech n-grams, and word embedding have been implemented to improve detection performance. The bag of words (BoW) assumption analyses random words and obtains word frequency-based characteristics. A sparse feature vector can negatively impact detection efficiency; therefore, it should be avoided if an online review is full of informal terminology, acronyms, and even obfuscated words. The use of POS n-grams enables online review sites to identify bogus reviewers. It's possible that such characteristics can't distinguish experienced fake reviewers. To give the impression that they are trustworthy, they employ terms and phrases virtually always found in fake evaluations rather than actual ones. Making excessive use of a few words might make fake reviews sound more convincing. N-grams aren't as effective as other methods for classifying false reviewers because there aren't many relevant terms. Therefore they might only be found in some of the fraudulent reviews. Word2Vec only captures a limited amount of semantic information because it employs a single embedding vector to represent a word in all possible usages. When reviews contain terms with many semantic interpretations depending on the context, specific tactics may decrease detection performance.

It has come to the attention of observers that customer reviews impact the choices potential customers could make. In other words, consumers decide whether or not to complete the buying journey based on the reviews they read on social media after making up their minds to buy the product before reading the reviews. As a result, customer reviews provide folks with a beneficial service. Reviews that are positive result in significant financial advantages, while unfavourable reviews frequently have the opposite effect on a company's bottom line. The open manner in which customers provide and use their feedback has been a contributing factor in problems that have arisen on websites containing customer reviews. Anyone can freely give criticism or critiques of any company at any moment without any duties or constraints, thanks to social media platforms like Twitter and Facebook, amongst others.

Deep Neural Networks have been utilized in recent studies with a great deal of success for a variety of spam detection tasks. These tasks include the detection of spam in email [20, 21] and the identification of spam in social media [22, 23, 24, 25]. A deep feed-forward neural network (DFFNN) and a convolutional neural network (CNN) is the two deep NN models used in this inquiry to extract the detailed properties buried within high-dimensional word, sentence, and emotion representations.

The following framework can be used for the sections of this task that remain to be completed. The section 2 will describe the relevant research on spotting fraudulent reviews. The mythology utilized for the model implementation is detailed in Section 3. Section 4, results and comparisons, and Section 5, the conclusion and future scope with the outcomes, are discussed. follow.

2. Literature Review

It has become widely recognized that among the most severe problems linked with online shopping is the prevalence of fake reviews. The purpose of positive and negative counterfeit reviews is to either promote or demote particular items to obtain a competitive advantage and to influence customers' decisions. Because customers cannot recognize false reviews, machine learning methods have been deployed to ensure that these evaluations are discovered as soon as possible. To train and test review classifiers on an annotated corpus of reviews, which includes labels indicating which classes the reviews belong to, automatic review classification can be achieved. Table

1 contains a selection of the many research articles that have been published over the past ten years on the topic of automatically detecting fake reviews. It comprises the characteristics and procedures used, the datasets, and the performance evaluation produced.

Jindal and Liu [26] published that tried to identify false product reviews based on the similarities between the aspects of the review and the product itself. To be more specific, the inclination of spammers to copy and paste their product reviews was exploited.

Wang et al. [27] developed a heterogeneous review graph to record the relationships between reviews, reviewers, and shops. This was done to identify spammers, who could then modify their behaviour. Therefore, the trustworthiness of reviewers and review content can be evaluated apart from one another, as can the sincerity of reviews and the dependability of stores.

The principle underpinning this technique served as the foundation for the probabilistic graph classifier created by Liu et al. [28]. Their model develops the multimodal embedded representation of nodes by utilising a bidirectional neural network in conjunction with an attention mechanism.

Ghai et al. [29] demonstrated that a thorough review with ratings significantly different from those of other reviews is proof of the existence of fraudulent reviews. Spam attacks are related to review ratings; hence, unusual temporal patterns in the ratings may signal that spam attacks occur.

Xue et al. [30] developed a way to identify the trustworthiness of users, reviews, and products by adding the variation in a user's perspective sentiment into a scoring system. This allowed the researchers to determine the reliability of users, reviews, and items. Word embeddings have only recently been used to obtain a semantic representation of thoughts. In the study referenced above (31), the authors tuned a pre- trained CBOW model by applying CNN to real review datasets, resulting in enhanced detection accuracy. In addition, a semi-supervised framework was built using the CBOW model in conjunction with relational features. Previous research found that when it came to the classification of the many different sorts of reviews, including fraudulent and honest reviews, the most popular method was machine learning approaches. Logical regression was one of the earliest traditional machine-learning methods. It was one of the first approaches because of its ability to give a probability estimate that precisely reflected the likelihood that a review was fake. However, conventional machine learning techniques like logistic regression and k-NN (k-nearest neighbor) have at least two drawbacks [21]. These solutions could be more effective when managing high-dimensional false review data. This is crucial since many word characteristics are often discovered in this data. Second, they require assistance to deal with the limited amount of data available successfully. This is of the utmost significance because testimonials often contain a few words or phrases. Other methods of machine learning, such as Naive Bayes (NB) [32] and support vector machines (SVMs) [33, 34], have lately gained favor as a means of detecting fake reviews as a way to get around the problems that were stated earlier. Similarly, evolutionary algorithms [35] and ensemble learning techniques [36, 37] have been used to combat the challenges of achieving convergence and overfitting, respectively. A thorough study [38, 39, 40] has been conducted on the issue of the typical machine learning algorithms that are utilized in the process of identifying fraudulent reviews.

Convolutional neural networks were utilized by Li et al. [48] to construct a neural network model that could learn document representation to recognize misleading spam perspectives. The model presented used the term "vector" as an input at some point during those two phases. A sentence-weighted neural network model has been constructed to capture each phrase and document included in the review precisely. The suggested model has an architecture that is made up of two convolutional layers. These layers are referred to as the sentence layer and the document layer. While it is the responsibility of the sentence layer to create a composition of the sentence, it is the responsibility of the document layer to turn the sentence vector into a document vector.

The "cold-start problem," which happens when a new reviewer submits a review to detect fake reviews based on behavioural and textual elements, was addressed by Wang et al. [49], who devised a technique to overcome the problem. CNN was used to model the review text because it can capture the rich semantic information that is extremely difficult to portray using more convenient features such as unigram and LIWC. CNN was used to model the review text. DRI-RCNN, a model for identifying false reviews of items, was proposed by Zhang et al. [50]. This model has two components that are used to identify potentially fraudulent reviews. These components include a recurrent convolutional neural network and word contexts.

A convolutional layer has been built to educate the overall vector in the direction of adequately portraying a word. An implementation of a recurrent neural layer that can learn right and left for a false and actual context vector of a comment has been developed. The structure that has been suggested has a total of four layers. On both the AMT and Deception datasets, the proposed model was put through its paces and given a thorough analysis. The investigation revealed that the proposed model successfully generated an accuracy of 82.9%.

Chang et al. [51] were the ones who initially conceived of the concept for the solution that is now commonly known as X-BERT. The BERT embedding will be fine-tuned for this approach's overall goal. As a result, the problem that

occurred in the past with BERT has been resolved. This specific embedding achieved a precision rate of 68 percent, and it was subsequently broadened by a large number of BERT models that were fine-tuned to attain remarkable performance.

The "BAKE" model that Jwa et al. developed to detect fake news utilized BERT embedding [52]. The term "exBAKE" refers to BAKE's unlabelled news contributions. The models used algorithms to decide what constituted fake news. These models functioned adequately with the FNC-1 dataset because they analyzed headlines and the body text of news items. They outscored their rivals by 0.125 and 0.137 points in the Formula One competition. The incorporation of news into the pre-training phase might make this better.

Buyukoz et al. [53] came up with the idea of domain adaptive fine-tuning, which is a simple method for applying unsupervised labelling to brand-new domains. This method was offered as a way to simplify the process. Domain adaptive fine-tuning is the name given to the method in discussion here. The contextualized embeddings in the text were modified by employing masked language modeling within the body of the writing. The ELMo and BERT embeddings experienced considerable gains due to some fine- tuning, which ultimately yielded great results of 83%.

Wang et al. [54] conducted a series of controlled tests to comprehensively explore traditional word embeddings in addition to contextualized versions of these embeddings for text classification. The tests were designed to compare the performance of the two types of word embeddings. BERT performs noticeably better than ELMo, mainly when dealing with lengthy document collections.

The fakeRoBERTa model based on GPT2 was presented by Joni Salmine et al. [55] as having the highest-performing accuracy among the other models. This particular model achieved a score of 96.64%.

Table 1. Deep Learning Approaches

Study	Model Classifier	Review based features	Dataset	Accuracy
[41]	deep feed-forward neura network	unigrams, bigrams, trigrams, Skip-Gram word embeddings	Hotels	89%
[42]	SWNN (sentence weighted neural network), CNN	weights of sentences, POS, pronouns	Hotels, Doctors, Restaurants	84%
[43]	CNN, GRNN	CBOW	Hotels, Restaurants, Doctors	84%
[44]	LSTM ensemble	Middle context First and last sentence	Hotels, Restaurants, Doctors	83%
[45]	Adaboost	Polarity of words, LDA . Ngrams	Yelp	F-Score 81
[46]	BERT	word embeddings-Skip-Gram, capitalized words,review length, polarity of words	Yelp	89%
[47]	Deep feed-forward neura network	Skip-Gram, ngrams, word embeddings	Hotels	89%
[51]	X-BERT	Parabel(linear)and attention XML(Neural)	Wiki Dataset	76.95%
[55]	fakeRoBERTa	GPT2	OSF dataset	96.64%

3. Methodology

The model we have provided has been disassembled into its component elements, as seen in Figure 2. The gathering of data and its preliminary analysis are both components of the project's initial phase. The data has been processed using NLP (natural language processing) methods that involve eliminating stop words and punctuation, changing them into lowercase English characters, and stemming. Other NLP approaches include stemming and transforming them into lowercase English letters.

These techniques are used to standardize the data. These are also considered to be NLP operations. Embedding data and enriching text with sentiment polarity is part of the second phase, which ensures the high quality of the dataset in the feature selection process. This procedure will employ Word Embedding techniques like Roberta to represent texts as numerical values. Embedding data is also a part of the second phase. The phrase that detects spam is the last one in the model that we have presented. During this process stage, traditional machine learning and deep learning classifiers sort reviews into two categories: spam and ham.

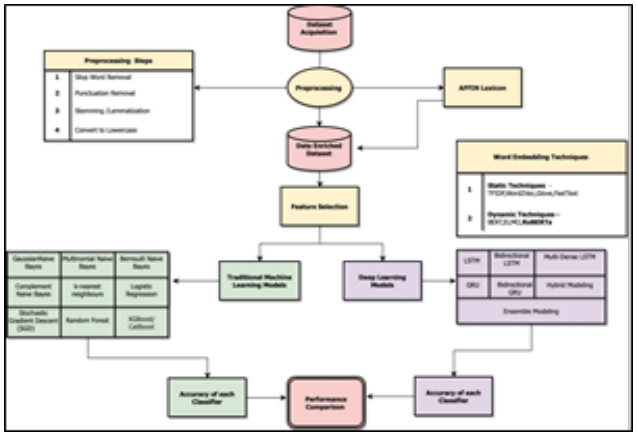


Fig. 2. Proposed Methodology

• Data Acquisition: Only a select few datasets include authentic reviews of high quality and reviews that are written to mislead readers. First, you must construct a massive database containing fake and genuine customer reviews. The dataset needs to be well-balanced, meaning it needs to have an equal amount of genuine and fraudulent reviews. This particular research used the OSF Fake Review Dataset, which can be located at https://osf.io/tyue9/.

• Data Preprocessing: At this point, the labeled instances retrieved from the OSF Dataset must go through preprocessing. Tokenization, the use of lowercase letters, and the elimination of commas, periods, and colons were all incorporated throughout the preprocessing phase of the project. This phase also included the usage of lowercase letters. It is common practice to regard the data preprocessing stage as among the most critical steps of any machine learning method.

• Data Enrich Text using AFINN: To enrich the text columns, our team suggested utilizing the AFINN lexicon to assess the polarity of the expressed sentiment. Using a computer program based on a dictionary, such as the AFINN model, is one alternative you have. One can associate a specific emotion with every word in the dictionary. When we tokenize a statement, we assign a score to each word that shows the degree to which it contributes positively or adversely to its overall meaning. AFINN provides an overall rating for the statement. The result of the calculation will disclose whether positive or negative terms dominate a sentence.

• Feature Selection: The third phase of the model we have provided comprises the selection of features. The RoBERTa is being put to use in the construction of our suggested model. Facebook developed a new tactic known as RoBERTa, essentially a retraining of BERT based on enhanced training methodologies. A variation of the BERT method that is more reliable and efficient is known as RoBERTa. In addition, dynamic masking has been incorporated, which ensures that the token that is being concealed will evolve as the training epochs continue. Part of speech (POS) and Linguistics inquiry and word count(LIWC) also integrated with RoBERTa to enhance the results.

• Fake Review Detection: The five deep-learning algorithms we have utilized to recognize fake reviews with hybrid and Ensemble modeling are LSTM, Bidirectional LSTM, multi-dense LSTM, GRU, and Bidirectional GRU.

• Long-Term Short-Term Memory (LSTM): The challenges of vanishing and exploding gradients, which occur in normal RNNs, can be overcome with the help of LSTM. When the RNN attempts to learn something complicated, it encounters this obstacle. This problem becomes apparent when the RNN's weights must be changed more than usual. The LSTM model is an essential component of deep learning. Memory cells are to blame for this ability, enabling the model to store and update information over time. Memory cells are accountable for this ability. In the relatively recent past, the field of LSTMs has witnessed several improvements.

• Bidirectional LSTM (BLSTM): The conventional LSTM architecture is expanded with the BLSTM algorithm, which processes the input sequence in both the forward and the backward directions. As a result, the model can represent the linkages between the current input and both the previous and the upcoming inputs. Because of this, it is beneficial for activities such as speech recognition, natural language processing, and video analysis, which require a more in-depth knowledge of the input sequence. When a BLSTM is used, one of the LSTM layers will process the line forward, while the other layer will process the string in reverse order.

• Multi-dense LSTM Model: In deep learning, a neural network design known as a Multi-dense LSTM Model is characterized by combining numerous dense (fully linked) LSTM layers. The thick layers carry out processing operations on the input features and extract useful information. LSTM layers process sequential input and store the memory of previous events so that it can be used to guide future predictions. Because it can effectively capture both types of relationships within the data thanks to the combination of dense and LSTM layers, the Multi-dense LSTM Model is an effective tool for tasks involving sequential and non-sequential information, such as sentiment analysis and time series prediction. This is because the Multi-dense LSTM Model can effectively capture both types of relationships within the data; due to this, sequential and non-sequential relationships within the data effectively.

• GRU: Deep Learning uses a subclass of RNNs, GRUs, designed to learn complex data sets. The GRU was developed to process sequential information and to keep a constant remembrance of events that have occurred in the past to impact predictions. This was achieved to fine-tune the accuracy. In contrast to more conventional LSTM and RNN systems, GRUs are outfitted with two gates that control data flow into and out of the hidden state. These gates are responsible for GRUs' ability to learn. One can consider these gates to be entrances and exits to the facility. Because of this, GRUs can improve their computing efficiency, as well as the ease with which they may be trained. Additionally, they can better capture the connections between the objects that appear in sequential data.

• GRU for Bi-Directional: Bidirectional GRU, or BiGRU, is a version of the GRU network that analyses sequential input in both the forward and the backward directions. Deep Learning developed it. A BiGRU network will read the sequence in forward and backward movements to comprehensively represent the information. After this, the hidden states that are produced will be concatenated. This bidirectional processing improves the capability of BiGRUs to collect contextual information and dependencies in series, which enables them to be effective in sequential data processing applications, like sentiment analysis and speech recognition. BiGRUs can also improve the ability of other systems to capture contextual information and dependencies in series. Because they combine the benefits of GRUs with bidirectional processing, BiGRUs is a more robust and versatile option for managing sequential data than regular GRUs or other forms of RNNs. This is because they incorporate the strengths of both types of processing. This is achieved by utilising the benefits of RNNs in conjunction with GRUs. strength

Algorithm 1: Deep Learning Process for Spam Review Detection

```
loadData();
preprocessData();
shuffleData();
splitData();
wordEmbedding();
loadModel();
foreach epoch in epochNumber do
    foreach batch in batchSize do
        logit = model(feature);
        loss = crossEntropy(logit, target);
        loss.backward();
        Evaluation(trainData, model, bestTrainAccuracy);
        Evaluation(testData, model, bestTestAccuracy);
    end
end
Function Evaluation(data, model, best):
    correct = 0;
    foreach batch in batchSize do
        logit = model(feature);
        loss = crossEntropy(logit, target);
        correct += (predict.data == target.data ).sum();
    end
    accuracy = (correct/totalData)*100;
    best = max(best, accuracy);
return
```

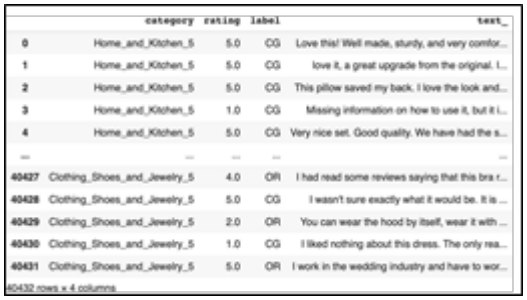
Fig. 3. Deep Learning Process

• Ensemble Modelling: A model that is more reliable and accurate is developed as a result of the process of ensemble modeling, which involves the merging of the predictions of numerous models. The concept underlying ensemble modeling is that the potential benefits of any model can be maximized by combining several other models. This is the central notion behind ensemble modeling. To arrive at a single conclusion that is the one that is most usually reached by all of the models, a Voting Classifier combines the predictions of multiple models and votes on which one is the most accurate. The bootstrap aggregating process, also called "bagging," involves producing numerous samples of the training dataset and then training several models on each piece individually. Boosting is an approach that progressively introduces a series of models, with each succeeding model aiming to rectify the flaws caused by the model that came before it in the training process. This strategy is used to improve the accuracy of the model's predictions.

• Hybrid Modelling: Hybrid modeling aims to generate a more accurate model by merging several distinct models or a wide range of data sources into a single modeling framework. One way to implement mixed modeling to identify false reviews is to combine different features, including text-based features, metadata features (such as review rating, timestamp, and reviewer profile information), and behavioral characteristics (such as click stream data and purchase history).

4. Experimental Setup

Database Used: The dataset needs to be well- balanced, which means it needs to include an equal number of authentic and fake reviews. The dataset can be downloaded from several websites, including Amazon, Yelp, TripAdvisor, and Open-Source Foundation (OSF). The Open-Source Foundation Fake Review Dataset (https://osf.io/tyue9/) was consulted for this study. The information about the dataset of fraudulent reviews is presented in Figure 4. Experiments utilising machine learning and deep learning classifiers are carried out here to improve performance benchmarks.



	category	rating	label	text_
0	Home_and_Kitchen_5	5.0	CG	Love this! Well made, sturdy, and very comfor...
1	Home_and_Kitchen_5	5.0	CG	love it, a great upgrade from the original. I...
2	Home_and_Kitchen_5	5.0	CG	This pillow saved my back. I love the look and...
3	Home_and_Kitchen_5	1.0	CG	Missing information on how to use it, but it i...
4	Home_and_Kitchen_5	5.0	CG	Very nice set. Good quality. We have had the s...
...	...	...	...	...
40427	Clothing_Shoes_and_Jewelry_5	4.0	OR	I had read some reviews saying that this bra r...
40428	Clothing_Shoes_and_Jewelry_5	5.0	CG	I wasn't sure exactly what it would be. It is ...
40429	Clothing_Shoes_and_Jewelry_5	2.0	OR	You can wear the hood by itself, wear it with ...
40430	Clothing_Shoes_and_Jewelry_5	1.0	CG	I liked nothing about this dress. The only rea...
40431	Clothing_Shoes_and_Jewelry_5	5.0	OR	I work in the wedding industry and have to wor...

40432 rows x 4 columns

Fig. 4. Fake Review Dataset

Enriching text columns: We proposed using the AFINN lexicon to determine the polarity of sentiment to enrich text columns. The AFINN lexicon is a collection of English phrases manually rated by Finn Rup Nielsen for their valence level between 2009 and 2011. To represent the level of positivity or negativity exhibited by each concept, an integer number ranging from -5 indicating a negative to positive was assigned to the term.

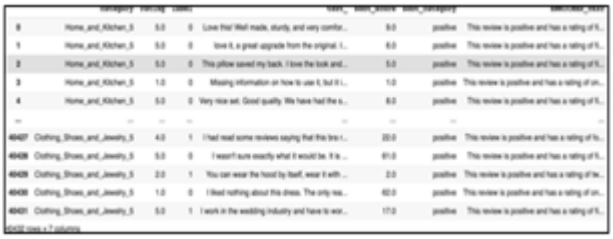


Fig. 5. Fake Review Enriched Text Dataset

The next step was implementing the Deep Learning Algorithms (Epoch=100) with and without RoBERTa on the AFINN-based dataset. These algorithms can learn for themselves and come to intelligent conclusions. Table 2 demonstrates that the accuracy was improved by utilising the RoBERTa algorithm.

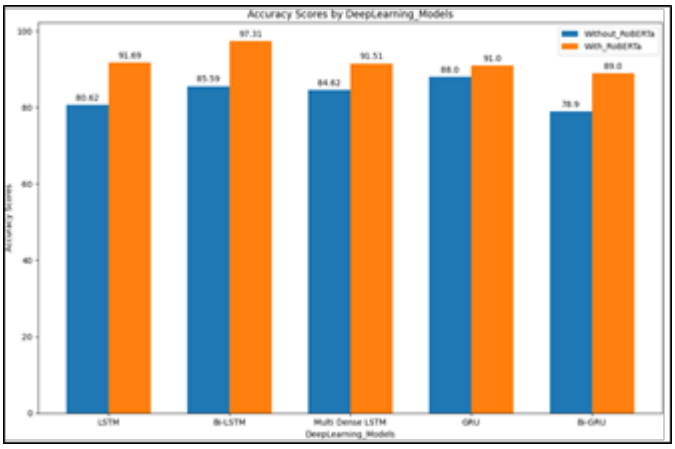


Fig. 6. Accuracy Scores by Deep Learning Models

Table 2. Deep Learning Models Performance on Enriched Text Dataset (AFINN)

Deep Learning Models	Accuracy without RoBERTa (POS+LIWC)		Accuracy Enhanced with RoBERTa+POS+LIWC	
	Training	Validation	Training	Validation
LSTM	90.71	80.62	93.71	91.69
FRARBiLSTM	99.9	85.59	**99.9**	**97.31**
Multi-dense LSTM	91.69	84.62	93.69	91.51
GRU	96.24	88.00	97.71	91.00
Bidirectional GRU	98.78	78.90	99.1	89.00

Therefore, in this research, we proposed our first Hybrid Model, which we referred to as FRARBiLSTM. This model surpasses the rest of the deep learning models and is considered superior, as per Figure 6. In the present environment, making up good online evaluations for a product or service is a serious issue that has become more pervasive and difficult to detect. This problem is because it needs to send the right message to potential customers. In this part, we will compare the results created by the deep learning- based hybrid model named FRARBiLSTM, with the results created by the ML-based and deep learning-based models, respectively. These comparisons will be made using the effects produced by the hybrid model.

Table 3. Comparative Results of Proposed Model with existing models

Models	Accuracy	
	Previous Study:	This Study Model: FRARBiLSTM (Fake Reviews-AFINN RoBERTa using Bidirectional LSTM) +POS+LIWC
Joni Salmine et al. [55]	96.64 (RoBERTa)	**97.31**

Deshai et al. [57]	93.8%(CNN-LSTM and LSTM- RNN)	**97.31**
Qadir et al. [56]	87.81% (BERT+ML Model)	**97.31**

5. Conclusion

In the present research, the importance of reviews and how they influence nearly every aspect of the information obtained on the internet was the primary focus of our attention. The dependability of the reviews and ratings that individuals find online strongly influences people's choices, even if the reviews and ratings in question are fraudulent. The application of a vast number of approaches categorised this dataset. The proposed Model FRARBiLSTM performs significantly better than other classifiers in identifying false reviews, with an accuracy of 97.31 percent. Because of this, this algorithm can efficiently classify reviews as genuine or false by taking into account merely the text content of the reviews and analyzing the characteristics of the sentiments expressed in the reviews. Nevertheless, the work that will be done in the future might consider an integrated strategy of ensemble modeling with deep learning models and other dynamic word embedding techniques.

References

[1] Rout, Jitendra Kumar, Amiya Kumar Dash, and Niranjan Kumar Ray. "A framework for fake review detection: issues and challenges." 2018 international conference on information technology (ICIT). IEEE, 2018.

[2] Floyd, Kristopher, et al. "How online product reviews affect retail sales: A meta-analysis." Journal of retailing 90.2 (2014): 217-232.

[3] BrightLocal (2018) Local consumer review survey 2018. https:// www.brightlocal.com/research/local-consumer-review- survey/. Accessed 8 Nov 2019

[4] The Times (2018) 'A third of TripAdvisor reviews are fake' as cheats buy five stars. The Times September 22, 2018. https:// www.thetimes.co.uk/article/hotel-and-caf-cheats-are-caught-try ing-to-buy-tripadvisor-stars-027fbcwc8. Accessed 22 Jan 2019

[5] Ott M, Cardie C, Hancock JT (2013) Negative deceptive opinion spam. In: 2013 conference of the North American chapter of the association for computational linguistics: human language technologies, ACL, pp 497–501

[6] Harris C (2012) Detecting deceptive opinion spam using human computation. In: Workshops at AAAI on artificial intelligence, AAAI, pp 87–93

[7] Atefeh, Heydari, et al. "Detection of review spam: A survey." Expert Systems with Applications 42.7 (2015): 3634-3642.

[8] Hussain N, Turab Mirza H, Rasool G, Hussain I, Kaleem M (2019) Spam review detection techniques: a systematic literature review. Appl Sci 9(5):987. https://doi.org/10.3390/ app9050987

[9] Crawford M, Khoshgoftaar TM, Prusa JD, Richter AN, Al Najada H (2015) Survey of review spam detection using machine learning techniques. J Big Data 2(1):1–23. https://doi.org/10. 1186/s40537-015-0029-9

[10] Ren Y, Ji D (2017) Neural networks for deceptive opinion spam detection: an empirical study. Inf Sci 385:213–224. https:// doi. org/10.1016/j.ins.2017.01.015

[11] Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: International conference on machine learning, JMLR, vol 32, pp 1188–1196

[12] Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, NIPS, vol 26, pp 3111–3119

[13] Li L, Qin B, Ren W, Liu T (2017) Document representation and feature combination for deceptive spam review detection. Neurocomputing 254:33–41. https://doi.org/10.1016/ j.neucom.2016. 10.080

[14] Peng Q, Zhong M (2014) Detecting spam review through senti- ment analysis. J Softw 9(8):2065–2072. https://doi.org/ 10.4304/ jsw.9.8.2065-2072

[15] Barbado R, Araque O, Iglesias CA (2019) A framework for fake review detection in online consumer electronics retailers. Inf Process Manag 56(4):1234–1244. https://doi.org/10.1016/j. indmarman.2019.08.003

[16] Kennedy S, Walsh N, Sloka K, McCarren A, Foster J (2019) Fact or factitious? Contextualized opinion spam detection. In: Proceedings of the 57th annual meeting of the association for computational linguistics: student research workshop, ACL, pp 344–350. https://doi.org/10.18653/v1/p19-2048

[17] Liu Y, Pang B, Wang X (2019) Opinion spam detection by incorporating multimodal embedded representation into a probabilistic review graph. Neurocomputing 366:276–283. https://doi. org/10.1016/j.neucom.2019.08.013

[18] Felbermayr A, Nanopoulos A (2016) The role of emotions for the perceived usefulness in online customer reviews. J Interact Mark 36:60–76. https://doi.org/10.1016/j.intmar.2016.05.004

[19] Malik, M. S. I., and Ayyaz Hussain. "Helpfulness of product reviews as a function of discrete positive and negative emotions." Computers in Human Behavior 73 (2017): 290-302.

[20] Barushka A, Hajek P (2016) Spam filtering using

regularized neural networks with rectified linear units. In: Adorni G, Cagnoni S, Gori M, Maratea M (eds) Conference of the Italian association for artificial intelligence, vol 10037. Lecture notes in computer science. Springer, Cham, pp 65–75. https:// doi.org/10.1007/978- 3-319-49130-1_6

[21] Barushka A, Hajek P (2018) Spam filtering using integrated distribution-based balancing approach and regularized deep neural networks. Appl Intell 48(10):3538–3556. https:// doi.org/ 10.1007/s10489-018-1161-y

[22] Barushka A, Hajek P (2018) Spam filtering in social networks using regularized deep neural networks with ensemble learning. In: Iliadis L, Maglogiannis I, Plagianakos V (eds) Artificial intelligence applications and innovations. AIAI 2018, vol 519. IFIP advances in information and communication technology. Springer, Cham, pp 38–49. https://doi.org/ 10.1007/978-3-319- 92007-8_4

[23] Jain G, Sharma M, Agarwal B (2018) Spam detection on social media using semantic convolutional neural network. Int J Knowl Discov Bioinform (IJKDB) 8(1):12–26. https://doi.org/ 10.4018/ IJKDB.2018010102

[24] Jain G, Sharma M, Agarwal B (2019) Spam detection in social media using convolutional and long short term memory neural network. Ann Math Artif Intell 85(1):21–44. https:// doi.org/10. 1007/s10472-018-9612-z

[25] Madisetty S, Desarkar MS (2018) A neural network-based ensemble approach for spam detection in Twitter. IEEE Trans Comput Soc Syst 5(4):973–984. https://doi.org/ 10.1109/TCSS. 2018.2878852

[26] Jindal N, Liu B (2007) Analyzing and detecting review spam. In: 7th IEEE international conference on data mining, ICDM 2007, IEEE, pp 547–552. https://doi.org/10.1109/icdm.2007.68

[27] Wang G, Li C, Wang W, Zhang Y, Shen D, Zhang X, Henao R, Carin L (2018) Joint embedding of words and labels for text classification. In: Proceedings of the 56th annual meeting of the association for computational linguistics, ACL, pp 2321–2331.https://doi.org/10.18653/v1/p18-1216

[28] Liu Y, Pang B, Wang X (2019) Opinion spam detection by incorporating multimodal embedded representation into a prob- abilistic review graph. Neurocomputing 366:276–283. https://doi. org/10.1016/j.neucom.2019.08.013

[29] Ghai R, Kumar S, Pandey AC (2019) Spam detection using rating and review processing method. In: Panigrahi B, Trivedi M, Mishra K, Tiwari S, Singh P (eds) Smart innovations in com- munication and computational sciences. Springer, Singapore, pp 189–198. https://doi.org/ 10.1007/978-981-10-8971-8_18

[30] Xue H, Wang Q, Luo B, Seo H, Li F (2019) Content-aware trust propagation toward online review spam detection. J Data Inf Qual (JDIQ) 11(3):11. https://doi.org/10.1145/3305258

[31] Xue H, Wang Q, Luo B, Seo H, Li F (2019) Content-aware trust propagation toward online review spam detection. J Data Inf Qual (JDIQ) 11(3):11. https://doi.org/10.1145/3305258

[32] Li F, Huang M, Yang Y, Zhu X (2011) Learning to identify review spam. In: International joint conference on artificial intelligence (IJCAI 2011), pp 2488–2493

[33] Li H, Chen Z, Mukherjee A, Liu B, Shao J (2015) Analyzing and detecting opinion spam on a large-scale dataset via temporal and spatial patterns. In: 9th international AAAI conference on web and social media (ICWSM 2015), AAAI, pp 634–637

[34] Mukherjee A, Venkataraman V, Liu B, Glance N (2013) What yelp fake review filter might be doing?. In: 7th international AAAI conference on weblogs and social media, AAAI, pp 409–418

[35] Pandey AC, Rajpoot DS (2019) Spam review detection using spiral cuckoo search clustering method. Evol Intell 12(2):147–164. https://doi.org/10.1007/s12065-019-00204-x

[36] Barbado R, Araque O, Iglesias CA (2019) A framework for fake review detection in online consumer electronics retailers. Inf Process Manag 56(4):1234–1244. https://doi.org/10.1016/j. indmarman.2019.08.003

[37] Rout JK, Dalmia A, Choo KKR, Bakshi S, Jena SK (2017) Revisiting semi-supervised learning for online deceptive review detection. IEEE Access 5:1319–1327. https://doi.org/ 10.1109/ ACCESS.2017.2655032

[38] Crawford M, Khoshgoftaar TM, Prusa JD, Richter AN, Al Najada H (2015) Survey of review spam detection using machine learning techniques. J Big Data 2(1):1–23. https://doi.org/10. 1186/s40537-015-0029-9

[39] Patel NA, Patel R (2018) A survey on fake review detection using machine learning techniques. In: 2018 4th international confer- ence on computing communication and automation (ICCCA), IEEE, pp 1–6. https://doi.org/10.1109/ ccaa.2018.8777594

[40] Vidanagama DU, Silva TP, Karunananda AS (2019) Deceptive consumer review detection: a survey. Artif Intell Rev. https:// doi. org/10.1007/s10462-019-09697-5

[41] Barushka, Aliaksandr, and Petr Hajek. "Review spam detection using word embeddings and deep neural networks." Artificial Intelligence Applications and Innovations: 15th IFIP WG 12.5 International Conference, AIAI 2019, Hersonissos, Crete, Greece, May 24–26, 2019, Proceedings 15. Springer International Publishing, 2019.

[42] Li L, Qin B, Ren W, Liu T (2017) Document representation and feature combination for deceptive spam review detection. Neurocomputing 254:33–41. https://doi.org/10.1016/j.neucom.2016. 10.080

[43] Ren Y, Ji D (2017) Neural networks for deceptive opinion spam detection: an empirical study. Inf Sci 385:213–224. https:// doi. org/10.1016/j.ins.2017.01.015

[44] Zeng ZY, Lin JJ, Chen MS, Chen MH, Lan YQ, Liu JL (2019) A review structure based ensemble model for deceptive review spam. Information 10(7):243. https://doi.org/10.3390/ info10070243

[45] Barbado R, Araque O, Iglesias CA (2019) A framework for fake review detection in online consumer electronics retailers. Inf Process Manag 56(4):1234–1244. https://doi.org/10.1016/j. indmarman.2019.08.003

[46] Kennedy, Stefan, et al. "Fact or factitious? Contextualized opinion spam detection." arXiv preprint arXiv:2010.15296 (2020).

[47] Barushka, Aliaksandr, and Petr Hajek. "Review spam detection using word embeddings and deep neural networks." Artificial Intelligence Applications and Innovations: 15th IFIP WG 12.5 International Conference, AIAI 2019, Hersonissos, Crete, Greece, May 24–26, 2019, Proceedings 15. Springer International Publishing, 2019.

[48] 48.L.Li,W.Ren,B.Qin,andT.Liu,''Learning document representation for deceptive opinion spam detection,'' in Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Nanjing, China: Springer, 2015, pp. 393–404.

[49] 49.X. Wang, K. Liu, and J. Zhao, ''Handling cold-start problem in review spam detection by jointly embedding texts and behaviors,'' in Proc. 55th Annu. Meeting Assoc. Comput. Linguistics, vol. 1, 2017, pp. 366–376.

[50] 50.W. Zhang, Y. Du, T. Yoshida, and Q. Wang, ''DRI-RCNN: An approach to deceptive review identification using recurrent convolutional neural network,'' Inf. Process. Manage., vol. 54, no. 4, pp. 576–592, 2018.

[51] Chang, Wei-Cheng, et al. "Taming pretrained transformers for extreme multi-label text classification." Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. 2020.

[52] Jwa, Heejung, et al. "exbake: Automatic fake news detection model based on bidirectional encoder representations from transformers (bert)." Applied Sciences 9.19 (2019): 4062.

[53] Büyüköz, Berfu, Ali Hürriyetoğlu, and Arzucan Özgür. "Analyzing ELMo and DistilBERT on socio-political news classification." Proceedings of the Workshop on Automated Extraction of Socio-political Events from News 2020. 2020.

[54] Wang, Congcong, Paul Nulty, and David Lillis. "A comparative study on word embeddings in deep learning for text classification." Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval. 2020.

[55] Salminen, J., Kandpal, C., Kamel, A. M., Jung, S., & Jansen, B. J. (2022). Creating and detecting fake reviews of online products. Journal of Retailing and Consumer Services, 64, 102771.

[56] Qadir Mir, Abrar, Furqan Yaqub Khan, and Mohammad Ahsan Chishti. "Online Fake Review Detection Using Supervised Machine Learning And BERT Model." arXiv e-prints (2023): arXiv-2301.

[57] Deshai, N., and B. Bhaskara Rao. "Deep Learning Hybrid Approaches to Detect Fake Reviews and Ratings." Journal of Scientific & Industrial Research 82.1 (2022): 120-127.