Fake review prediction using Neural Network with different activation functions

Aditi Bhat
Msc In Computing
Dublin City University
Dublin, Ireland
aditi.bhat2@mail.dcu.ie

Kashish Kishinchandani
Msc In Computing
Dublin City University
Dublin, Ireland
kashish.kishinchandani2@mail.dcu.ie

Manoj Kesavulu
PhD Usagae Analytics
Dublin City University
Dublin, Ireland
manoj.kesavulu@dcu.ie

***Abstract*—People commonly educate themselves before purchasing a product by reading internet reviews. Online reviews assist purchasers in gaining knowledge and experience about the goods and giving them an impression of the product. Buyers trust products with the most reviews since they appear to be the most purchased and rated goods. Online reviews are favourable and highly regarded; they help the business flourish. However, phoney reviews are made to deceive consumers into purchasing the products. As a result, fake review identification has grown increasingly vital to assist online shoppers in receiving the most legitimate opinions when purchasing items. In this paper, we try to figure out how to distinguish between false and authentic reviews by employing several word embeddings with varying activation functions using a neural network. Furthermore, the study compares how activation functions are essential in detecting bogus reviews. Various word embeddings and activation functions were used, and the best one was presented in this work.**

***Index Terms*—Fake reviews, Activation Functions, Word Embeddings, Neural Network**

I. Introduction

Individuals and businesses depend highly on online reviews to make purchasing and business decisions. Positive feedback may result in substantial financial advantages and public recognition for firms and people. Nevertheless, this creates enormous incentives for imposters to rig the system by publishing fictitious evaluations to promote or denigrate certain items or services. Individuals performing such activities are known as opinion spammers. Reviews play a vital role and are referred by consumers before buying any product online as consumers look for an honest and experienced view in the form of reviews.

Online reviews are becoming more trusted when purchasing things as internet usage grows. Knowing that product sales depend on online reviews, merchants frequently attempt to deceive buyers by writing biased evaluations about the product. Sellers frequently engage people to create favourable or false reviews to hurt a business's image. Amazon has filed a lawsuit against the owner of over 10,000 Facebook groups for posting fraudulent reviews on the e-commerce site. [10] According to Amazon, these organizations recruit people worldwide to publish fraudulent reviews in exchange for money. Amazon, Yelp, and other similar sites have been combating fake reviews for years. Many businesses buy positive reviews for cash, coupons, and promotions. Yelp.com has launched a sting operation against these businesses, trying to increase their sales through fake reviews. [11]

Amazon, Yelp.com, and other comparable sites have employed machine learning extensively to counteract fake reviews. The primary detection approach has been supervised learning. However, due to a scarcity of trustworthy false reviews, the current work is primarily based on ad hoc reviews and other computer-generated reviews. False reviews tend to have a pattern, and such patterns can be detected using neural networks. Neural networks replicate the activities of the human brain to identify correlations among massive volumes of data. The pattern that needs to be identified and trained are complex. Using activation functions in the neural network helps the network learn such complex patterns in the data. The activation function determines whether a neuron should be stimulated by generating a weighted sum and then adding bias. The activation function aims to induce non-linearity into a neuron's output. The activation function takes the output signal from the previous cell and passes it to the next cell in the form of some input. There are several types of activation functions, and they can play a vital role when identifying fake reviews.

The research questions that have been attempted to answer in this paper are below:

- How do individual activation functions combined with different word embedding techniques work in fake review prediction?
- What makes the difference in a specific combination of a word embedding and an activation function that increases accuracy compared to other combinations?

The rest of the paper is organized as follows. Section 2 examines the current state of the art in fake review identification based on past studies. Section 3 then outlines the approach that was used. Section 4 discusses the results and evaluation. Finally, Section 5 analyses the conclusion and future scope.

II. Literature Review

G. M. Shahariar et al. focus on detecting deceptive reviews [1], making use of both labelled and unlabelled data and developed deep learning algorithms for spam review detection, including the Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), and a variation of the Recurrent Neural Network (RNN) known as Long Short-Term Memory (LSTM). Additionally, they used some classic machine learning classifiers to detect spam reviews, such as Naive Bayes (NB), K Nearest Neighbour (KNN), and Support Vector Machine (SVM), and compared the performance of both traditional and deep learning classifiers.

In [2], Khan H et al. propose using a supervised machine learning technology called support vector machine (SVM) to distinguish between fake (spam) and authentic (ham) text. The supervised learning technique for classifying spam and genuine reviews begins with entering the input review, preprocessing it, and then using the SVM classifier to classify it as fraudulent (SPAM) or genuine (HAM) on fake benchmark reviews. The proposed supervised machine learning technique for detecting false reviews is compared to previous work and other supervised machine learning classifiers.

The paper's primary goal by S. M. Anas and S. Kumari [3]is to combat spammers by building a sophisticated model on millions of reviews. This research used the "Amazon Yelp dataset" to train the models. Its short dataset is used for training on a small scale, but it may be scaled up to achieve great accuracy and flexibility. The fake review dataset is trained using two Machine Learning (ML) models that estimate the accuracy of how authentic the reviews in a dataset are. When relying on product reviews for items discovered online on various websites and applications, the rate of fraudulent reviews in the E-commerce business and even on other platforms is increasing. This model may be used by websites and applications with tens of thousands of users to forecast the legitimacy of reviews, allowing website owners to take appropriate action.

I. Amin and M. Kumar Dubey [4] focus on maliciously identifying views or feedback in the comments. The spam detection framework has improved efficiency over traditional machine learning. It works well with unique models, providing better answers to challenging real-life scenarios, thanks to a surge in the practical uses of soft computing. Most current research on review spam detection uses the standard bag-of-words model to recognize text analysis characteristics and traditional machine learning models such as Support Vector Machines and other such classifiers. They have three proposed methodologies: Feature Selection based on Linguistic Approach, Feature Selection based on Behavioural Approach of Reviewers related with its Products and Soft Computing Techniques.

Barbara Probierz et al. [5] notice that a severe problem is disinformation in the form of fake news. The researchers have done some initial news analysis by its title. The purpose of the researchers was to create a novel model for initial news analysis and rapid detection of false news based solely on the headline rather than analyzing the complete article content. Furthermore, the ability to balance precision and recall as categorization quality measures was developed, allowing for better news selection. After analyzing the text with NLP approaches, the researchers have used the Adaptive goal function of ant colony optimization algorithms to find fake news. This research hypothesized that employing the goal-oriented ACDT method and a confined term matrix allowed for better classification than traditional algorithms. Experiments have shown that it is possible to conduct a preliminary analysis of fake and authentic news based on the collected data. Experiments showed that it is possible to undertake a preliminary examination of fake and genuine news using a constrained word matrix. Adopting goal-oriented ACDT by the researchers has allowed them to significantly increase the recall of real news and the precision of fake news. Using decision tables helps the researchers gain better results and increase model accuracy. Researchers think it is worth looking into the impact of the number of words in the choice table on the outcomes and classification coverage. Following a comprehensive examination, it may be worthwhile to construct a two-stage verification system in the future – the first based on a restricted number of words from the news headline, and the second based on the complete content, but only for news that could not be categorized earlier. The method given here can identify false news by looking at the titles of stories on the internet.

Rozita Talaei Pashiri et al. propose a feature selection-based method [6] based on the sine–cosine algorithm to reduce spam detection inaccuracy (SCA). Feature vectors are updated by the proposed technique [6]. The precision, accuracy, and sensitivity of the proposed technique for the Spambase dataset in MATLAB were 98.64 per cent, 97.92 per cent, and 98.36 per cent, respectively. In other words, when it came to spam identification, the proposed method outperformed the multilayer perceptron (MLP) neural network, Bayesian network, decision tree, and random forest classifiers. According to the test findings, the feature selection error in the MLP neural network was reduced by approximately 2.18 per cent employing the SCA.

The spam detection error is affected by several elements: the ANN inputs and the selection of crucial attributes that best represent spam and emails. Because each feature has varying relevance, using all of them for training the ANN raises the inaccuracy. On the other hand, including all features increases the problem and data size, increasing execution time. The results of the testing revealed that the suggested technique outperformed other learning methods in terms of accuracy, precision, and sensitivity in spam detection, including the MLP neural network, Bayesian network, decision tree, and random forest. In this aspect, the MLP neural network came in second.

K. Archchitha and E.Y.A. Charles [7] has proposed that CNN is offered to distinguish genuine opinion reviews from opinion spam. The proposed model was trained to utilize opinion text represented as word vectors by the pre-trained

GloVe word embedding model. This method varies from prior methods in two respects. Most earlier efforts relied on classifiers such as Support-Vector Machines and Nave Bayes to describe review text characteristics using the classic bag-of-words paradigm. The suggested CNN model outperformed existing techniques when evaluated on the Deceptive opinion spam corpus. The trials demonstrated that additional factors other than textual semantics must be investigated to identify deceptive viewpoints in reviews successfully. The suggested model confirmed its capabilities by enhancing accuracy even more. The suggested model's performance may be enhanced by employing a more significant data set, adding new characteristics such as behavioural information, and fine-tuning the CNN model's hyper-parameters. The classification method is like that used in many text-based machine learning applications. As a result, this model may also be used for sentiment analysis, auto-tagging client questions, and text segmentation into predetermined subjects.

Ting-You Lin et al. [8] have created a framework for detecting fraudulent reviews and conducted a thorough analysis of the efficacy of various variables and their combinations using three classifiers. A diverse selection of popular benchmark datasets from Amazon review data and Yelp review data was used in the simulated trials. It has been discovered that readability and subject characteristics are more successful than sentiment analysis (sentiment features) in detecting bogus reviews. The essential feature group emerged as readability features. When paired with FOG or FK's readability characteristics, topic features become even more crucial. The researchers investigate other significant elements for detecting phoney reviews and compare them to other research studies. Researchers have suggested and investigated making use of deep learning models.

Petr Hajek et al. suggested two deep NN models for identifying fake consumer reviews in [9], utilizing an integrated framework of n-gram, Skip-Gram, and emotion models. The experimental findings on four real-world fake review datasets in this paper indicate the efficacy of the suggested models. Notably, the suggested models beat current baseline techniques and state-of-the-art fake review identification systems regarding the accuracy, AUC, and F-score. The experimental findings of the researchers also revealed that the suggested integrated models were the most successful for more enormous datasets with coupled polarity, hinting that they might be used in real-world circumstances. The suggested detection methods were also successful in terms of time complexity and detection time, which the researchers revealed using the ARR multicriteria measure. It was suggested that future research should integrate the presented models with graph-based techniques based on review information. In the future, researchers want to combine the benefits of the DFFNN and CNN models to create a hybrid deep NN structure akin to the Network in Network. A hybrid model like this might help the CNN model generalize even more. Another shortcoming of the suggested approach is that sentence weights were omitted due to their domain-specific character, unlike the CNN model.

III. Implementation and Methodology

A. Platform

The programming language used for this research is Python 3, which is the latest Python Programming Language. Jupyter Notebook and Google Colab were heavily used throughout this research.

- Jupyter Notebook is a free, open-source web application for creating and sharing documents containing live code, equations, visualizations, and text.
- In contrast, Google Colab allows you to construct Python scripts for machine learning and data analysis.
- Our research uses python libraries extensively, and some of them used are as below:
  - Pandas: Used to store data in the form of data frame.
  - Numpy: Used to perform mathematical operations on the data.
  - Scikit-learn: Used for machine learning and statistical operations like vectorization, data scaling and training the data.
  - Matplotlib and Seaborn: Used in exploratory analysis to attain informative statistical graphs.
  - Natural Language Tool Kit: Used for test processing libraries such as tokenization, stemming and lemmatization.
  - Gensim: Used for unsupervised topic modelling and contains parallelized implementations such as fastText, doc2vec and word2vec.
  - Keras and Tensorflow: Provides high-level API for building and training models.

B. Data Analysis

The dataset used for training the model is collected from Kaggle. The dataset has a wide scope as it includes negative and positive sentiments. It is a labelled dataset for both fake and real reviews. The dataset has real positive reviews, real negative reviews, fake positive reviews, and fake negative reviews. This dataset allows the model to train on all four types.

On further data analysis, the following splitting of the dataset was noted. Looking at the below table, we know that the dataset is unbiased.

Polarity	Review	Percent
Positive	Real	25%
Positive	Fake	25%
Negative	Real	25%
Negative	Fake	25%

TABLE I
Overview of the Dataset.

We have explored the data using count plot to understand the spread of real and fake reviews based on different parameters like polarity and source of the data. Below are the visualizations for the exploratory data analysis.

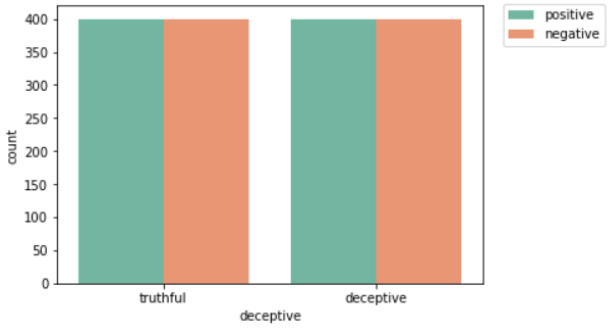

Fig. 1. Count Plot for real and fake reviews based on polarity

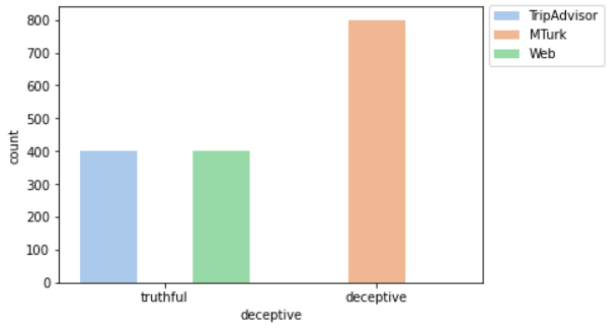

Fig. 2. Count Plot for real and fake reviews based on source

C. Pre-processing the Data

Data preprocessing occurs in four steps: *tokenization, lemmatization, stop word removal*, and *normalization*.

Tokenization is the process of splitting the text into meaningful bits of data called tokens. We performed tokenization of the reviews using nltk and spacy libraries in Python. In addition to tokenization, we perform POS tagging to assign each word to a syntactical category.

Lemmatizing is performed on the tokenized words. In lemmatization, different forms of the same word are put under the same category. For example, after lemmatizing, words like "troubled", "troubling", and "troubles" are reduced to "trouble".

In the next step, we perform *stop word removal*. Words like "a", "the", and "is" are called stop words i.e. words that occur very frequently but do not add much information to the text. These words are moved as they do not add semantic value to the classification.

Lastly, we perform *normalizationon* the tokenized words. This is done in order to reduce randomness in the text, thereby the amount of different information the computer has to process. This improves the computational efficiency.

D. Word Embedding

Deep learning algorithms require numeric inputs. Hence, the reviews are vectorized using word embedding techniques. The *Doc2Vec* (Document to Vector) method and *GloVe* (Global Vector) method are the word embedding techniques used.

Doc2Vec, as the name suggests, is a word embedding model used to convert documents to vectors. It is an extension of the popular Word2Vec model. It produces a pair of vectors that have minimal distance between each other in the vector space. A review may contain more than one paragraph. Doc2Vec deals with multiple paragraphs by including another vector called a Paragraph ID, making Doc2Vec ideal for the vectorization of reviews. Additionally, on empirical analysis [14], Doc2Vec proved to outperform Word2Vec in terms of robustness.

Jeffrey Pennington et al. [15] proposed a novel approach towards vectorization called GloVe method. GloVe method has shown a higher accuracy than skip-gram and CBOW overall. This is due to the incorporation of global statistics, i.e, word co-occurrence, for vectorization. It performs better than other vectorization models in tests including named entity recognition, word analogies, and word similarity.

E. Modelling

For fake review prediction, we are using a supervised classification model. A hybrid *CNN-BiLSTM* model is used to classify the reviews as fake or real. In this model, the input is fed into the CNN, and the output of the *CNN* is fed in as the input of *BiLSTM*.

The vectorized data is fed in as input to the *Convolution layer* of the Convolution Neural Network. Here, the vector is summarized into a smaller vector. The *dropout layer* is used after the convolution layer to prevent overfitting of the model by nullifying some neurons while operating the rest. *Max Pooling layer* is employed after the dropout layer to decrease the dimension of the vector, thereby reducing the computational load. Different *activation functions* are used to set the bias and weights for the neurons. The *Bidirectional LSTM* is fed with the output of the CNN. A dropout layer is employed again. *Softmax* function is used to set the weight of the neurons. The output is obtained, and the accuracy is calculated against various activation functions and the results are shown in Table II.

Activation Functions used for modelling:

- *ReLU* stands for Rectified Linear Unit and is a non-linear activation function. ReLU activation function is used widely as it does not activate all the neurons at the same time.

$$f(x) = max(x, 0)$$

  The neurons in the ReLU activation function will only be deactivated if the output of the linear transformation is less than 0. Hence, ReLU is computationally better when compared to sigmoid and tanh.
- *Sigmoid* is a non-linear activation function and transforms the value between the range of 0 and 1. Multiple neurons having sigmoid activation function will have a non-linear output. The mathematical expression for sigmoid is below:

$$f(x) = \left(\frac{1}{1 + e^{-x}}\right)$$

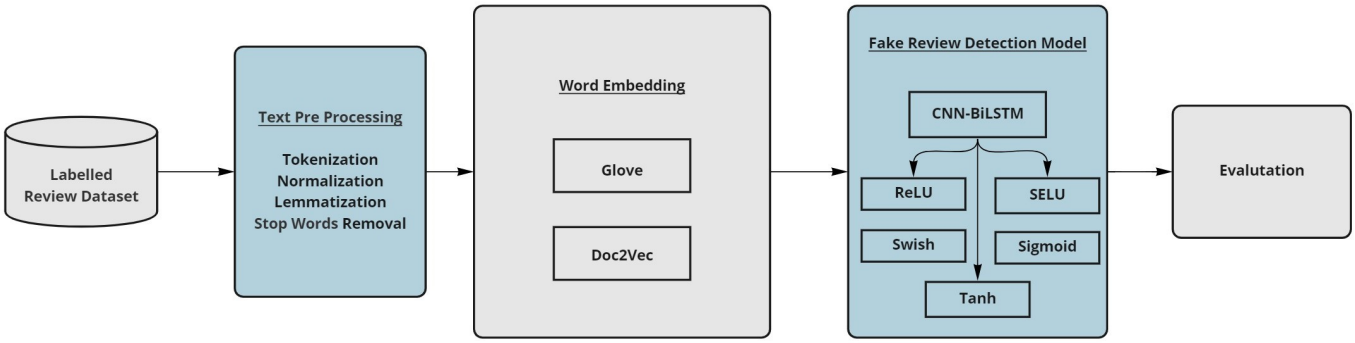

Fig. 3. Proposed Methodology

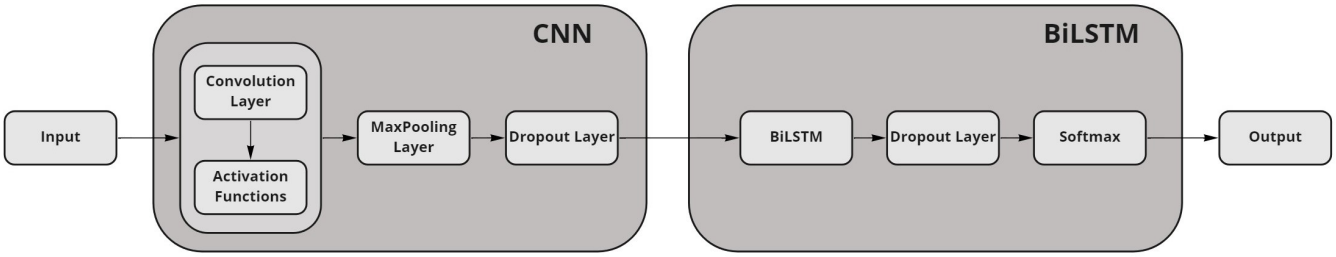

Fig. 4. Fake Review Detection Model

- *Tanh* activation function is symmetric around the origin, with values ranging from -1 to 1, implying that the following layer's inputs will not always be of the same sign. It is defined as below:

$$tanh(x) = 2sigmoid(2x) - 1$$

- *SELU* also known as Scaled Exponential Linear Unit induce self normalizing properties. The mathematical expression for SELU is given as below:

$$f(x) = \lambda x, \text{ if } x \geq 0$$

$$f(x) = \lambda\alpha(e^x - 1) \text{ if } x < 0$$

- *Swish* activation function is a smooth and non-monotonic function that performs well with different domains. Swish activation function provides better results than ReLU in previous studies. [12] [13]. The mathematical expression for swish is:

$$f(x) = xsigmoid(\beta x)$$

$$f(x) = \left(\frac{x}{1 + e^{-\beta x}}\right)$$

IV. Experimental Results

We employed a convolution neural network with a bi-directional LSTM model in our experiment to estimate the accuracy of false reviews using different activation functions. We utilized 90 percent of the dataset for training and 10 percent for testing. The accuracy is calculated as follows:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}, \text{ where}$$

$$TP = TruePositives$$

$$TN = TrueNegatives$$

$$FP = FalsePositives$$

$$FN = FalseNegatives$$

Table II displays the results with the accuracy for each activation function.

Epochs and iterations are important part of training the neural network. Using higher number of epoch and iterations require high computational power. We have made use 10 epochs and 4 iterations with minimal computational power to get the best results. Sigmoid activation function when used with Doc2Vec and GloVe gives the lowest accuracy. The output for the sigmoid activation function saturates for a

Activation Functions	Word Embeddings	
	Doc2Vec	GloVe
ReLU	91.25%	90.15%
Sigmoid	77.31%	77.81%
TanH	83.43%	89.22%
SELU	89.37%	88.91%
Swish	91.38%	90.21%

TABLE II
ACCURACY TABLE

large positive or large negative number, hence the accuracy of sigmoid is low. As it can be seen from Table II, swish activation function when used with Doc2Vec gives the best results of 91.38 %. ReLU when used with Doc2vec and GloVe provides 91.25% and 90.15% respectively.

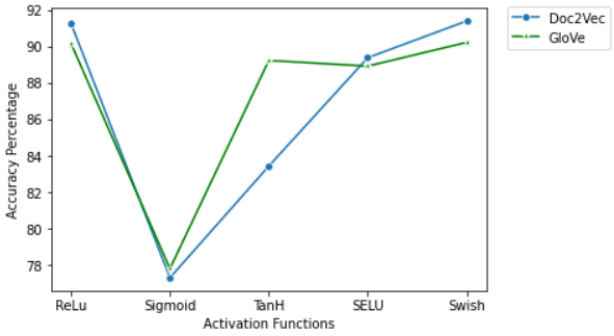

Fig. 5. Accuracy graph with Different Activation Functions

V. CONCLUSION

In this paper, the focus is on identification of fake reviews using different activation functions with two word embeddings, Doc2Vec and GloVe. Doc2Vec and GloVe are taken into consideration as they are widely used. The model used for this study is the CNN-BiLSTM model as it has shown better accuracy from previous studies. [16] [17] Different activation functions such as the ReLU, Sigmoid, Tanh, SELU and Swish functions were used in gathering the results. Hyper parameters such as dropout and changing the layers in the model was carried out to get the best optimal configuration.

The Swish activation function provides the best accuracy with 91.38% with Doc2Vec and 90.15% with GloVe. The number of epochs configured while performing this model was 10, and the iterations were 4. The data used during this research had 1600 reviews, and minimal computational power was utilized. An attempt was made to use more epochs and iterations along with a more extensive set of reviews. However, results could not be obtained since the computational power was limited.

In future, a larger dataset needs to be used to train the model more effectively. As the size of the dataset increases, higher computational power will be required. The methodology used in this research can be applied in different domains to detect fake reviews.

REFERENCES

[1] G. M. Shahariar; Swapnil Biswas; Faiza Omar; Faisal Muhammad Shah; Samiha Binte Hassan, "Spam Review Detection Using Deep Learning," in IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2019.

[2] Khan H., Asghar M.U., Asghar M.Z., Srivastava G., Maddikunta P.K.R., Gadekallu T.R, "Fake Review Classification Using Supervised Machine Learning," in International Conference on Pattern Recognition, 2021.

[3] S. M. Anas and S. Kumari, "Opinion Mining based Fake Product review Monitoring and Removal System," in 6th International Conference on Inventive Computation Technologies (ICICT), 2021.

[4] I. Amin and M. Kumar Dubey, "An overview of soft computing techniques on Review Spam Detection," in 2nd International Conference on Intelligent Engineering and Management (ICIEM), 2021.

[5] Probierz B., Kozak J., Stefański P., Juszczuk P., "Adaptive Goal Function of Ant Colony Optimization in Fake News Detection," in International Conference on Computational Collective Intelligence, 2021.

[6] Rozita Talaei Pashiri, Yaser Rostami and Mohsen Mahrami, "Spam detection through feature selection using artificial neural network and sine–cosine algorithm," in Math Sci 14, 2020.

[7] K. Archchitha, E.Y.A. Charles, "Opinion Spam Detection in Online Reviews Using Neural Networks," in IEEE, 2019.

[8] Ting-You Lin, Basabi Chakraborty, Chun-Cheng Peng, "CNN is offered as a method for distinguishing genuine opinion reviews from opinion spam.," in International Conference on Data Analytics for Business and Industry (ICDABI), 2021.

[9] Petr Hajek, Aliaksandr Barushka and Michal Munk, "Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining," in Neural Comput and Applic 32, 17259–17274 (2020), 2020.

[10] Simonetti, I. (2022, July 19). nytimes. Retrieved from The New York Times: https://www.nytimes.com/2022/07/19/business/amazon-fake-reviews-lawsuit.html

[11] Streitfeld, D. (2012, October 12). nytimes. Retrieved from The New York Times: https://www.nytimes.com/2012/10/18/technology/yelp-tries-to-halt-deceptive-reviews.html

[12] Avenash, R., and P. Viswanath. "Semantic Segmentation of Satellite Images using a Modified CNN with Hard-Swish Activation Function." VISIGRAPP (4: VISAPP). 2019.

[13] Mercioni, Marina Adriana, and Stefan Holban. "P-swish: Activation function with learnable parameters based on swish activation function in deep learning." 2020 International Symposium on Electronics and Telecommunications (ISETC). IEEE, 2020.

[14] Jey Han Lau and Timothy Baldwin. 2016. An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation. In Proceedings of the 1st Workshop on Representation Learning for NLP, pages 78–86, Berlin, Germany. Association for Computational Linguistics.

[15] Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "Glove: Global vectors for word representation." Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014.

[16] W. Yue and L. Li, "Sentiment Analysis using Word2vec-CNN-BiLSTM Classification," 2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS), 2020, pp. 1-5, doi: 10.1109/SNAMS52053.2020.9336549.

[17] Lu, W., Li, J., Wang, J. et al. A CNN-BiLSTM-AM method for stock price prediction. Neural Comput Applic 33, 4741–4753 (2021). https://doi.org/10.1007/s00521-020-05532-z