IJWR

A Survey on Review Spam Detection Methods using Deep Learning Approach

Mahmoud Aliarab[a], Kazim Fouladi-Ghaleh[b*]

[a]Master of Information Technology Engineering, Deep Learning Research Lab, Faculty of Engineering, College of Farabi, University of Tehran, Iran; aliarab.m@gmail.com

[b]Assistant Professor, Department of Computer Engineering, Faculty of Engineering, College of Farabi, University of Tehran, Iran; kfouladi@ut.ac.ir

ABSTRACT

Review spam is an opinion written to promote or demote a product or brand on websites and other internet services by some users. Since it is not easy for humans to recognize these types of opinions, a model can be provided to detect them. In recent years, much research has been done to detect these types of reviews, and with the expansion of deep neural networks and the efficiency of these networks in various issues, in recent years, multiple types of deep neural networks have been used to identify spam reviews. This paper reviews the proposed deep learning methods for the problem of review spam detection. Challenges, evaluation criteria, and datasets in this area are also examined.

Keywords: ***Review Spam Detection; Opinion Spam; Deep Learning; Convolutional Neural Network; Long Short-Term Memory (LSTM); Literature Survey.***

1. Introduction

It can be said that in the world today, most people have experienced the use of Internet services in some way. Customer opinion is one of the most important resources for choosing a service or buying a product. User opinion is critical because it has a considerable impact on potential customers, and with the information it provides, users can decide whether or not to use a service. Given the importance of these opinions and the significant impact they can have on the sale of a product, some profiteers try to promote or destroy a brand or product by creating unrealistic opinions. Such opinions are called "deceptive opinions" or "review spam". These comments are written by people who do not have personal experience using the product and mislead users into making the right decision [1]. Therefore, developing a model for recognizing this type of comment is necessary.

Review spam detection is a relatively new topic, and unlike issues such as spam emails, which have a long history, it has been around since 2008 with research by Jindal et al. [2]. Review spam detection has received more attention in recent years, and much research has been done in this area. Models proposed in this field initially used statistical methods or traditional machine learning. However, in recent years, with the expansion of the use of deep learning and good efficiency of deep neural networks in various problems, researchers have also used these networks to identify spam opinions, and multiple methods based on deep learning have been proposed in this area. When it comes to identifying spam opinions, it is not easy to create a model that can be comprehensive and work well for different domains.

2. Review Spam Detection Datasets

Different supervised, semi-supervised, and unsupervised learning methods have been used to identify spam reviews. But it can almost be said that supervised methods have always worked better. However, due to the difficulty of labeling data in this area, labeled datasets are very small to detect spam reviews. The following will examine some of the most important datasets used in this field.

- **OpSpam:** This dataset, first published by Ott et al. [3] in 2011, is one of the most widely used datasets in the field of review spam detection. This dataset contains 1600 comments about 20 hotels in Chicago. This dataset is balanced. It has 800 spam comments and 800 real comments, including 400 positively polarized comments and 400 negatively polarized comments. The real comments in this dataset are from various sites such as TripAdvisor, Yelp, and the source of unreal comments (spam) is the Amazon Mechanical Turk (AMT) tool, which is a crowdsourcing tool. The statistical description of this dataset can be seen in Table 1.

- **YelpChi:** This dataset was first compiled by Mukherjee et al. [4]. As indicated in the dataset name, these comments were collected from yelp.com. This dataset contains 67,395 comments from reviews of Chicago area hotels and restaurants. This dataset contains 201 hotel and restaurant reviews written by 38,063 users. Comments in this dataset include information such as product information, user information, comment submission time, rating, comment text, and comment label.

doi http://dx.doi.org/10.22133/ijwr.2022.346586.1123

Citation. M. Aliarab and K. Fouladi, "A survey on review spam detection methods using deep learning approach," *International Journal of Web Research*, vol.5,no.1, pp.19-24, 2022.

**Coressponding Author(s)*

Article History: Received 10 June 2022; Revised 23 July 2022 ; Accepted 30 July 2022

- **YelpNYC & YelpZIP:** These two datasets were first collected by Rayan et al. [5]. The source of these two datasets is the yelp.com website. The YelpNYC dataset contains 359,052 reviews related to restaurants in New York City. The information contained in each of the comments in this dataset includes user information, product information, comment submission time, rating, comment text, and comment label. Comments on this collection are related to 923 restaurants written by 160,225 users. The YelpZIP dataset also includes 60,598 reviews of restaurants with zip codes starting in New York City and reviews of restaurants in the US Map area. These comments are for 5,044 restaurants written by 260,277 commenters, including user and product information, comment submission time, rating, comment text, and comment label.

- (**Hotel, Doctor, Restaurant (HDR)):** This dataset was first collected by Li et al. [6] and included opinions on doctors, restaurants, and hotels. This dataset contains three types of opinions: real opinions, which are customer opinions, spam comments, which are comments generated by AMT; and spam comments, which are tagged and generated by employees in each field (experts). This dataset is also one of the most widely used databases in the field of review spam detection. Table 2. shows the statistics of the data in this dataset.

3. Deep Models For Review Spam Detection And Existing Challenges

The issue of review spam detection is relatively new in the field of spam detection, which began in 2008 with a study by Jindal et al. [2]. This issue is more complex than other issues of text classification, and in addition to various challenges such as lack of labeled data, and time complexity that exist in deep learning issues, it also has various challenges that are specific to this issue. This section examines review spam detection challenges and some of the best models that use deep learning to identify spam reviews.

3.1 Challenges of Review Spam Detection

- **Singleton Spam reviews:** One of the most critical challenges in this area is the existence of singleton spam comments. A singleton spam review is a comment whose author only recorded one comment. It is a problem in models that rely on metadata such as user information, and comment submission time, because an author who has submitted only one comment does not give much information to the model, and for this reason, in most studies, this type of comment is dropped from the dataset.

- **Feature selection:** Another major problem in detecting spam reviews is selecting features that can be used to correct classification. These features should be a good representation of whether an opinion is a spam or not. The difficulty of choosing these characteristics is that it is also difficult for humans to distinguish whether an opinion is a spam or not [7].

Table 1. Statistics of Opspam Dataset

Source	*Num of comments per section*	*Total size*
TripAdvisor	400 reals (positive)	1600 comments
Priceline, Orbitz, Hotels.com, Expedia	400 reals (negative)	
AMT	400 spams (positive)	
AMT	400 spams (negative)	

Table 2. Statistics of Hotel, Doctor, Restaurant Dataset

Domain	*Turker*	*Expert*	*Customer*
Hotel (P/N)	400/400	140/140	400/400
Restaurant (P/N)	200/0	100/0	200/200
Doctor (P/N)	200/0	32/0	200/0

- **Model generalizability:** Another problem is model generalizability. It is difficult to provide a model that can be used in different domains and with different data in the issue of review spam detection and get a good result. The issue of review spam detection is highly dependent on the dataset. Articles that have tried to test the accuracy of the trained model on data such as doctor's offices for data related to hotel reviews have drastically reduced the accuracy of the model.

- **Limit on the number of features:** Datasets that exist to detect spam reviews have a limited number of features for model training. The reason for this is that each person who has collected data to train their model has collected this data in a completely tasteful manner, and usually, the existing dataset has a small number of features, or because of differences in the features of different datasets, all features of a dataset cannot be used [8].

3.2 Deep Models for Detecting Spam Reviews

Studies for the English Language

As mentioned, research into spam detection first began in 2008 with a study by Jindal et al. [2], but with the expansion of the use of deep neural networks in various issues and the excellent performance of these networks compared to traditional machine learning methods, these networks were also used in the problem of detecting spam reviews. As shown in Figure 1, research into the field of review spam detection using deep learning has intensified since 2015 and has grown exponentially in recent years. As mentioned in [9], the use of deep learning in spam detection has intensified since 2015. One of the essential researches in 2015 was a study conducted by Lie et al. [10], and they tried to do learning using convolutional neural networks (CNN) and use these networks to detect spam reviews. This study gives word vectors as input properties to the network, and spam reviews are directly detected using CNN. The results of this study showed a good performance of CNN networks in detecting spam. In research to identify spam reviews, the use of ensemble methods and the integration of several deep learning models have also been considered.

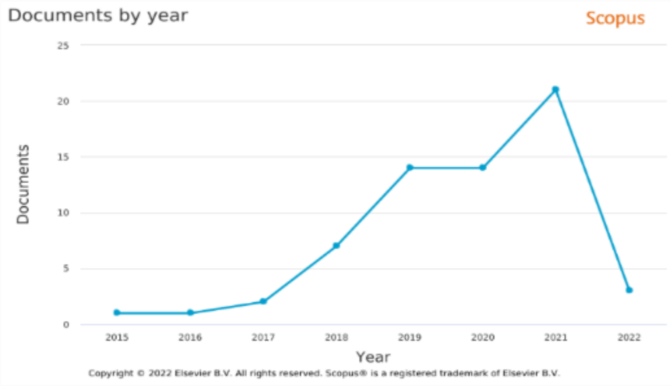

Figure. 1. Number of articles on review spam detection using deep learning (extracted from scopus.com)

In 2016, an article [11] was written by Ren and Zhang that used document-level learning to identify spam reviews. In this paper, a document is first given to the model, and using CNN combined with an RNN-Gated neural network, the sentences, their structure, and how they are represented are learned, and the document vectors are extracted using this method. These vectors are then used directly for learning to detect spam reviews. Another study was conducted in 2017 by Zhao et al. [12]. In this study, they tried to use a new method called word order-preserving in the convolutional layers and the CNN network instead of using the usual pooling layer in the convolutional network. They are trying to improve CNN networks to address the issue of review spam detection. The use of the attention mechanism was also considered in a study conducted in 2017 by Wang et al. [13]. They present a model in this paper that dynamically conducts learning on behavioral and linguistic characteristics, as well as a model for attention-based learning that combines these methods to examine and identify spam reviews.

In 2017, Li et al. Conducted a study [23] on the detection of spam. In this study, they used a method called SWNN. What is essential in this method is that in addition to giving weight to the words according to their importance in the spam or not of an opinion, it also gives weight to the sentences according to their coefficient of impact on the label of a comment. In this method, convolutional neural networks (CNN) are used, and in the pooling layer, a coefficient α is used, which will be related to the weight of each sentence. The evaluation of the proposed model is performed on the HDR dataset, and for mixed-domain mode, they obtained to F1 score of 86.1%.

In 2018, Wang et al. [14] tried to detect spam opinions using long short-term memory networks (LSTM). A noteworthy point in this article is that by comparing different methods of machine learning and comparing them with newer methods such as deep learning models, they concluded that among the different methods, deep learning methods are better for this problem, and LSTM networks are more suitable than support vector machines (SVMs). If we look at the research of recent years, most of these studies use deep learning methods to try to solve the problem of review spam detection. In a study [24] conducted in 2018 by Zhang et al., a model called DRI-RCNN was proposed. In this research, it is mentioned that it is better to vectorize words for spam and real opinions separately and train on both in the embedding stage. By stating that the words used in a spam review are closer to each other in the embedded space and the words used in the actual comments are also closer to each other in this space, their model uses two types of vectorizations for each comment and uses pair learning for each comment. OpSpam and HDR datasets have been used to train and evaluate this model.

Another study was conducted in 2019 by Shahariar et al. [15]. This study tried to use a model to detect any type of spam. This article uses different models such as multilayer perceptron (MLP), CNN, and LSTM. They used both labeled and unlabeled data for training their model. Also, this model has been implemented using traditional machine learning methods such as k-nearest neighbor (KNN) and support vector machine (SVM), and in the end, the performance of these models has been compared with each other.

Another study [26] by Stanton and Irissappane in 2019 used a semi-supervised GAN to detect spam reviews. In this study, a combination of labeled and unlabeled data was used to teach the model, which yielded the best performance for a situation where 50% of the data was labeled, and 50% was unlabeled. In this study, they named their model spamGAN-50, and by evaluating the model on the opSpam dataset, they reached F1 85.6%.

Saumya et al. [16] proposed a hybrid model in 2020 using LSTM and autoencoder networks. In this study, they stated that one of the main problems in review spam detection was the lack of labeled data, and therefore tried to create an unsupervised model using a combination of long short-term memory networks (LSTM) and autoencoder using the text features of the comment without label of comments. They used a criterion called the Matthews Correlation Coefficient (MCC) to evaluate their model, and in comparison with other similar works [17], the performance of the proposed model seems acceptable.

Another study conducted in 2020 by Mahalakshmi et al. [18] also used long short-term memory networks (LSTMs). They used the OpSpam dataset in this study and obtained better accuracy (about 80 to 85%) than traditional machine learning models such as the support vector machine and k-nearest neighbor. In 2021, Neisari et al. [19] identified spam reviews by combining a method called self-organizing maps with the CNN model. They have used an innovative method in this work. In their research, the words in the text of the comments are put together by examining the degree of similarity and their relationship with each other as an image. Each word in the comment text is considered as a pixel of the image, and the juxtaposition of these words forms an image of the words. This image is then given to a CNN network. Due to the nature of CNN networks commonly used to work with images, this generated image is given to a CNN network, and a label of comment is generated.

Another study was conducted in 2021 by LIU et al. [20] that uses hierarchical networks based on the attention mechanism. In this research, two layers are used to extract the properties of comments. They extract important sentence properties in the first layer using CNN and the N-gram method. In the second layer, using a combination architecture of convolutional neural networks and BiLSTM, they extract the semantic properties and general dependencies of the document. Finally, by extracting these features, they have identified spam reviews.

In 2021, Bhuvaneshwari et al. [27] used convolutional neural networks, self-attention mechanisms, and long short-term memory networks to detect spam reviews. They first vectorize the comments in the model they present using the word embedding mechanism. These vectors are given to one layer of the attention mechanism, and then the vectors are given to several convolutional neural networks. The outputs of different convolutional neural networks are concatenated to form a vector. The formed vector is then given to a long short-term memory network, and the output of this network passes through a fully connected network to produce an output label. This model is evaluated on the YelpZip dataset and reached 87.3% accuracy on this dataset.

Research [25] was conducted in 2021 by Salunkhe. A BiLSTM network and an Attention layer are used in the model presented in this research. Comments are first vectorized by passing through an embedding layer, then these vectors are passed to the BiLSTM network, the output is given to an attention layer, and weighted vectors are output from the network. The weighted vectors are exiting the Attention layer pass through a Softmax layer to produce the final label. In this research, the OpSpam dataset has been used, and by evaluating the proposed model on this dataset, they have reached 90.25% accuracy.

A 2021 study by Alsubari et al. [28] combined CNN and LSTM to create a model that detects spam using an embedding layer. The model presented in this research works in such a way that first, the comments are converted into vectors using an Embedding layer, and then the vectors are given to a CNN network. The output of this network is given to the LSTM network, and the output of the LSTM network produces the final label by passing through a Sigmoid layer. This model has been evaluated in two forms: In-domain and Cross-domain on Amazon, Yelp, OpSpam datasets, and restaurant-related comments. For In-domain mode, 77%, 85%, 86%, and 87% accuracy were obtained for Yelp, Restaurant, OpSpam (Hotel), and Amazon datasets, respectively, while for Cross-domain mode, the accuracy was 89%.

In 2022, Salminen et al. [29] conducted a study that first generated spam from the Amazon Store database using two models, GPT-2 and ULMFiT, which are a subset of GAN networks. After the number of comments generated using these networks reached an acceptable value, a comparison was made between the human and machine to detect these spam comments. Finally, the results of this study show that the machine (machine learning models) can detect spam better than humans, and the authors state that "the machine can fight the machine."

In a study [30] conducted in 2022 by Cao et al., a model called ST-MFLC was developed to detect spam reviews. The model presented in this research uses three networks CNN, LSTM, and Self-Attention, separately. It means that each of these networks is a complete and separate part for detecting spam reviews, which itself includes the embedding layer, the main network layer of the fully connected layer, and the Sigmoid layer. The input data is given to each of these segments, the output of the fully connected layer of all the segments is concatenated together, the final generated vector is passed to a fully connected layer, and the output label is generated. This model is evaluated on the HDR dataset, and for In-domain mode for Hotel, Doctor, and Restaurant domain data, 88%, 90.3%, and 85% accuracy are obtained, respectively.

Studies for the Persian Language

Research on the Persian language in the review spam detection field is very limited, and also the existing methods have not used deep learning methods to detect spam and have only used traditional machine learning methods.

One of the researches that have been done on the Persian language is [21]. This study was conducted by safarian et al. in 2019 and tried to examine the various features used to teach the model in the problem of review spam detection. They have used different models such as simple Bayesian methods, decision trees, and support vector machines, and each of these models with different features such as overall product rating, the sentiment of comments, and POS tags, trained. In this research, the Digikala dataset has been used.

Another study to identify spam reviews in Persian was the study [22] conducted by Basiri et al. In this research, they tried to use different machine learning methods such as the naive Bayes model, decision tree, support vector machine, and using various features that are extracted from the text of the comment, as well as using other metadata that is available in the dataset to detect spam in Persian. In this research, they used the supervised learning method and the Digikala dataset for training and evaluation. In this study, they used balanced and unbalanced data to teach their model, and finally, based on the results they obtained, they concluded that for unbalanced data, the best model is a support vector machine, and for unbalanced data, the best model is a decision tree. A comparison of the most crucial research conducted to detect spam reviews using deep learning is shown in Table 3.

4. Performance Measurement Metrics For Review Spam Detection Models

Complex criteria for evaluation are not used in the problem of review spam detection, but the critical point is that the model's effectiveness should be evaluated based on several different criteria. In some studies, only one criterion (e.g., accuracy) is used, which alone cannot show the model's efficiency well. Especially if the dataset used is not balanced, it is necessary to use several different criteria to measure the model's efficiency.

Most researches on the problem of detecting spam reviews use four criteria: Accuracy, Precision, Recall, and F1 Score. The details of calculating each of these criteria are shown in Eq. (1) through (4).

$$Accuracy= (TP+TN)/ (TP+TN+FP+FN) \tag{1}$$

$$Precision= TP/ (TP+FP) \tag{2}$$

$$Recall= TP/ (TP+TN) \tag{3}$$

$$F1=2\times (Precision\times Recall)/ (Precision+Recall) \tag{4}$$

In calculating these criteria, we mean TP (True Positive) data that was positive and correctly detected by the model, TN (True Negative) data that was negative and correctly detected, and FP (False Positive) and FN (False Negative) data that were positive and negative and misdiagnosed by the model.

Table 3. Comparison Of Research Conducted To Identify Spam Using Deep Learning

Paper	*Pub. year*	*Dataset*	*Model*	*Best perf.*
[10]	2015	HDR	CNN	Acc: 79.5
[11]	2016	HDR	CNN+Gated RNN	Acc: 84.1, F1:83.9
[12]	2017	OpSpam	Word order-perserving CNN	Acc: 70.02
[14]	2018	Mobile01.com (Crawled)	LSTM	Acc: 79.6
[15]	2019	OpSpam	LSTM+CNN+MLP	Acc: 96.75
[16]	2020	Youtube.com	LSTM autoencoder	MCC: 98
[18]	2020	OpSpam	LSTM	Acc:80-85
[19]	2021	HDR	CNN	Acc:87.1, F1: 87
[20]	2021	HDR	CNN + BilSTM	Acc: 86.5, F1: 89.3
[27]	2021	YelpZip	Self-attention + BiLSTM + CNN	Acc: 87.3
[25]	2021	OpSpam	Attention + BiLSTM	Acc: 90.25
[28]	2021	Multidataset (Yelp, Amazon, ...)	CNN + LSTM	Acc: 87 for restaurant dataset
[30]	2022	HDR	CNN + LSTM + Self-Attention	Acc: 90.3 for doctor domain

5. Conclusion

The purpose of this study is to review the work done in the field of spam detection based on deep learning. In this article, we tried to examine the essential researches that have been done to detect spam and in which deep neural networks have been used, and the strengths and weaknesses of each of them have been examined. The study also showed that research in the field of review spam detection using deep learning is growing exponentially. The research reviewed in this study shows that in recent years, researchers use CNN and LSTM networks more to detect spam reviews, and these two networks have better performance than other deep learning methods. The most important dataset used in the field of review spam detection is also presented in this study, and the details of each of them were examined. Examining the research conducted in the field of spam detection, it can be concluded that the spam detection problem still has much work to do, and there is a need to do more studies on different parts of model production and data collection. Given that this issue is more complex than other issues of text classification, such as toxic word recognition, it seems that more effort is needed to

Declarations

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Authors' contributions

MA: Study design, acquisition of data, interpretation of the results, analysis, drafting the manuscript; KF: Study design, revision of the manuscript; supervision, drafting the manuscript.

Conflict of interest

The authors declare that there is no conflict of interest.

References

[1] Q. Peng, and M. Zhong, "Detecting Spam Review through Sentiment Analysis", *J. Softw.*, vol. 9, no. 8, pp. 2065-2072, 2014.

[2] N. Jindal, and B. Liu, "Opinion spam and analysis", In *Proceedings of the 2008 International Conference on Web Search and Data Mining, WSDM '08*, Association for Computing Machinery, New York, NY, USA, , 2008, pp. 219–230.

[3] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination", arXiv preprint arXiv:1107.4557, 2011.

[4] A. Mukherjee, B. Liu, and N. Glance, "Spotting fake reviewer groups in consumer reviews", In *Proceedings of the 21st International Conference on World Wide Web, WWW '12*, Association for Computing Machinery, New York, NY, USA, , 2012, pp. 191–200

[5] S. Rayana, and L. Akoglu, "Collective Opinion Spam Detection: Bridging Review Networks and Metadata", In *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining*, Association for Computing Machinery, New York, NY, USA, 2015, pp. 985–994.

[6] J. Li, M. Ott, C. Cardie, and E. Hovy, "Towards a general rule for identifying deceptive opinion spam". In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, vol. 1: Long Papers, 2014, pp. 1566–1576.

[7] A. Rastogi, and M. Mehrotra, "Opinion spam detection in online reviews", *Journal of Information & Knowledge Management*, vol. 16, no. 04, p. 1750036, 2017.

[8] N. Hussain, M. Turab, R. Hamid, G. Rasool, I. Hussain, and M. Kaleem, "Spam review detection techniques: A systematic literature review", *Applied Sciences*, vol. 9, no. 5, p. 987, 2019.

[9] Y. Ren, and D. Ji, "Learning to detect deceptive opinion spam: A survey", *IEEE Access*, vol. 7, pp. 42934-42945, 2019.

[10] L. Li, W. Ren, B. Qin, and T. Liu, "Learning Document Representation for Deceptive Opinion Spam Detection", In *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*, Springer, Cham, 2015, pp. 393–404.

[11] Y. Ren, and Y. Zhang, "Deceptive opinion spam detection using neural network", In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, Osaka, Japan, The COLING 2016 Organizing Committee, Dec. 2016, pp. 140–150.

[12] S. Zhao, Z. Xu, L. Liu, and M. Guo, J. Yun, "Towards accurate deceptive opinion spam detection based on word order-preserving CNN", *Mathematical Problems in Engineering*, vol. 2018, pp. 1-9, 2018.

[13] X. Wang, K. Liu, and J. Zhao, "Detecting Deceptive Review Spam via Attention-Based Neural Networks", In *National ccf conference on natural language processing and chinese computing*, Springer, Cham, 2018, pp. 866–876.

[14] C. C. Wang, M. Y. Day, C. C. Chen, and J. W. Liou, "Detecting spamming reviews using long short-term memory recurrent neural network framework", In *Proceedings of the 2nd International Conference on E-commerce, E-Business and E-Government*, 2018, pp. 16–20.

[15] G. M. Shahariar, S. Biswas, F. Omar, F. M. Shah, and S. Binte Hassan, “Spam review detection using deep learning”, In *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2019, pp. 0027–0033.

[16] S. Saumya, and J. P. Singh, “Spam review detection using LSTM autoencoder: an unsupervised approach”, *Electronic Commerce Research*, vol. 22, pp. 113-133, 2022.

[17] T. C. Alberto, J. V. Lochter, and T. A. Almeida, “Tubespam: Comment spam filtering on youtube”, In *2015 IEEE 14th international conference on machine learning and applications (ICMLA)*, IEEE, 2015, pp. 138–143.

[18] P. Mahalakshmi, V. Sampreeth, and C. Venkataramana. Detection of opinion spam using LSTM networks, *International Journal of Advanced Science and Technology*, vol. 29, no. 6, pp. 67 – 75, 2020.

[19] A. Neisari, L. Rueda, and S. Saad, “Spam review detection using self-organizing maps and convolutional neural networks”, Computers & Security, vol. 106, p. 102274, 2021.

[20] Y. Liu, L. Wang, T. Shi, and J. Li, “Detection of spam reviews through a hierarchical attention architecture with n-gram CNN and bi-LSTM”, *Information Systems*, vol. 103, pp. 101865, 2021.

[21] N. Safarian, M. E. Basiri, and H. Khosravi, “Feature ranking for Persian review spam detection”, *Journal of Soft Computing and Information Technology*, vol. 8, no. 2, pp.1–16, 2019.

[22] M. E. Basiri, N. Safarian, and H. K. Farsani, A supervised framework for review spam detection in the Persian language. In *2019 5th International Conference on Web Research (ICWR)*, IEEE. 2019, pp. 203–207.

[23] L. Li, B. Qin, W. Ren, and T. Liu, “Document Representation and Feature Combination for Deceptive Spam Review Detection”, *Neurocomputing*, vol. 254, pp. 33-41, 2017. doi:10.1016/j.neucom.2016.10.080. 2017

[24] W. Zhang, Y. Du, T. Yoshida, Q. Wang, “DRI-RCNN: An approach to deceptive review identification using recurrent convolutional neural network”, *Information Processing & Management*, vol. 54, Issue 4, pp. 576-592, 2018.

[25] A. Salunkhe, “Attention-based Bidirectional LSTM for Deceptive Opinion Spam Classification”, *arXiv preprint arXiv:2112.14789*, 2021. doi:10.48550/ARXIV.2112.14789.

[26] G. Stanton, and A. A. Irissappane, “GANs for Semi-Supervised Opinion Spam Detection”, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, 2019, pp. 5204–5210. doi:10.24963/ijcai.2019/723.

[27] P. Bhuvaneshwari, A. N. Rao, and Y. H. Robinson, “Spam review detection using self attention-based CNN and bi-directional LSTM”, *Multimedia Tools and Applications*, vol. 80, no. 12, pp. 18107-18124, 2021. doi:10.1007/s11042-021-10602-y.

[28] S. N. Alsubari, S. N. Deshmukh, M. H. Al-Adhaileh, F. W. Alsaade, and T. H. Aldhyani, “Development of Integrated Neural Network Model for Identification of Fake Reviews in E-Commerce Using Multidomain Datasets”, *Applied Bionics and Biomechanics*, 2021. doi:10.1155/2021/5522574.

[29] J. Salminen, C. Kandpal, A. M. Kamel, S. G. Jung, and B. J. Jansen,, “Creating and detecting fake reviews of online products”, *Journal of Retailing and Consumer Services*, vol. 64, p. 102771, 2022. doi:10.1016/j.jretconser.2021.102771.

[30] L. Gutierrez-Espinoza, F. Abri, A. S. Namin, K. S. Jones, and D. R. Sears, “Fake Reviews Detection through Ensemble Learning”, *arXiv preprint arXiv:2006.07912*, 2020.

Mahmoud Aliarab received the B.Sc. degree in Computer Engineering from Shahid Beheshti University in 2018, and M.Sc. degree in Information Technology Engineering from University of Tehran in 2021. He currently works at Deep Learning Research Lap, University of Tehran. His current research interests include deep learning, natural language processing and data mining.

Kazim Fouladi-Ghaleh received the M.Sc. and Ph.D. degrees in Electrical and Computer Engineering / Artificial Intelligence and Robotics from School of ECE, College of Engineering, University of Tehran, Iran. He is currently an assistant professor of Computer Engineering at University of Tehran, College of Farabi, Faculty of Engineering and the director of the Cyberspace Research Lab and Deep Learning Research Lab. His research interests include AI, Image Processing & Computer Vision, Deep Learning, Cybernetics and Cyberspace Studies.