ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/358004087

Fake News Detection Using Naïve Bayes and Long Short Term Memory algorithms

Article *in* IAES International Journal of Artificial Intelligence (IJ-AI) · June 2022

DOI: 10.11591/ijai.v11.i2.pp746-752

CITATIONS
29

READS
3,135

2 authors:

Rania Azad M. San Ahmed
Sulaimani Polytechnic University

13 PUBLICATIONS 79 CITATIONS

SEE PROFILE

Sarra Senhadji
Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf

7 PUBLICATIONS 43 CITATIONS

SEE PROFILE

All content following this page was uploaded by Rania Azad M. San Ahmed on 25 March 2022.

The user has requested enhancement of the downloaded file.

Fake news detection using naïve Bayes and long short term memory algorithms

Sarra Senhadji[1], Rania Azad San Ahmed[2,3]
[1]Computer Science Department, Faculty of Mathematics and Informatics, University of Science and Technology Mohamed Boudiaf, Oran, Algeria
[2]Computer Network Department, College of Informatics, Sulaimani Polytechnic University, Sulaimani, Iraq
[3]The American University of Iraq, Sulaimani, Iraq

Article Info

Article history:

Received Jul 24, 2021
Revised Dec 22, 2021
Accepted Jan 11, 2022

Keywords:

Deep learning
Fake news detection
LSTM
Machine learning
Naïve bayes

ABSTRACT

Information and communication technologies have revolutionized the numerical world by offering the freedom to publish and share all types of information. Unfortunately, not all information circulated on the internet is accurate, which can have serious consequences, including misleading readers. Detecting false news is a complicated task to overcome. Massive studies focus on using machine and deep learning techniques in an attempt to classify the news as authentic or not. The goal of this research is an attempt to glance and evaluate how naïve bayes (NB) and long short-term memory (LSTM) classifiers can be used to positively identify fake news. The outcomes of this experiment reveal that LSTM achieves an accuracy of 92 percent over naive bayes. Moreover, the findings of the proposed approach's results outperform the related work results.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Rania Azad San Ahmed
Computer Network Department, College of Informatics, Sulaimani Polytechnic University
Sulaimani, Iraq
Email: Rania.azad@spu.edu.iq

1. INTRODUCTION

With the emergence of social media platforms, publishing and sharing information has become much easier and quicker. Individuals are now only one click away from global information. Social media platforms can serve two purposes in terms of news consumption: they can be used to keep the community informed of breaking news and, on the other hand, they can be used to spread misinformation [1]. Despite the many benefits of social media, it has presented some challenges like misleading information, fraudulent reviews, phony adverts, rumors, false political remarks, satires, and fake news due to information's low-cost, easy-to-access, and rapid distribution of news and information.

Fake news has a harmful impact on the community in a variety of categories, including social, financial, and even political dimensions, as demonstrated by the 2016 US Presidential election and the Brexit referendum events [2]. Until today, no comprehensive definition is provided for fake news, many explanations, definitions, and ambiguities exist. Understanding what the term represents depends on the purpose and aim of the definer. For example, in [2], it is considered as "news stories that have no factual basis but are presented as fact". In [3] "fake news, or hoax news, refers to intentionally false information or propaganda published under the disguise of being authentic."

According to studies, human's ability to detect false information without special assistance is only 54% [4]. Therefore, there is a necessity for a computerized fake and real news classification that is accurate. Automatically detecting fake news is considered a major challenge, due to the dynamic nature of social media platforms [5] and the complexity and diversity of natural languages [6], and the scarcity of high-quality training data complicates the process of developing supervised learning methods. In the light of these circumstances, both industrial and academic authors are taking an active role in combating internet fake news like Google, Facebook, and Twitter [1].

2. RELATED WORKS

Until today, the issue of "Detecting Fake News on Social Media" was researched extensively, and countless models were developed to help address the problem. Various machine learning or deep learning models along with natural language processing techniques were used [7]. This section contains a brief discussion of potential works in this domain.

In machine learning, a massive number of studies have proposed different models of machine learning algorithms like naïve Bayes (NB), support vector machine (SVM), logistic regression (LR), and many other models [2], [8]–[11], Our focus is on naïve Bayes model, many fake news detections have been implemented using NB as in the second model of the work done by [12], NB achieved 56% accuracy among 7 other classifiers using linguistic and user features. The work [13] shows the model NB classified the second in accuracy after SVM with 55.85%, authors in [14] achieved the highest accuracy of 80% using bayesian classifier. Smartly, other researchers used different ensemble methods techniques to get better performance [15], [16].

On the other hand, deep learning models have made great development in recent years and are currently considered promising methods for depicting and detecting online fake news using long short-term memory (LSTM) models in a single approach like [17] where LSTM is applied to capture dynamic changes in forwarding content to identify rumors. Verma *et al.* [18] successfully applied recurrent neural network (RNN), grated recurrent units (GRU), and LSTM, the LSTM model achieved an accuracy score of up to 94 percent. Other studies used LSTM in hybrid approaches, such as the work [19] that combined LSTM-RNN and convolutional neural network (CNN) to identify false news from Twitter posts. The plain LSTM model had the best performance, unlike LSTM with dropout regularization due to under-fitting. Asghar *et al.* [20] suggested a rumor classification model proposed by merging bi-long-short term memory (BiLSTM) with CNN. Moreover, Singh *et al.* [12] proposed an attention-based LSTM network that distinguishes rumor and non-rumor tweets using tweet text and 30 different linguistic and user features. The performance of the model reached 88% against the other conventional machine and deep learning models. According to [21], who implemented a hybrid model of CNN and LSTM, with dimensionality reduction methods such as principal component analysis (PCA) and chi-square, intending to determine whether the headlines of a news article agree with the text body, the highest accuracy of this study was 97.8 percent in less time.

Similar to our objective, the following studies did comparisons between different models in machine learning (ML) and deep learning (DL) models, including NB and LSTM. As shown in [14], the paper provided evaluation with different algorithms and achieved an accuracy of 80% using a bayesian classifier and 77% using a hybrid-LSTM. Similarly, Han and Mehta [22] demonstrated a benchmark between LSTM and NB in terms of accuracy and performance, with 82.29% against 67.12%. Likewise, Alameri and Mohd [23] aimed to evaluate classifiers such as NB, SVM, NN-Keras, NN-TensorFlow, and LSTM. Deep learning models, such as LSTM, outperformed all other models by 91.31%, while NB received only 67.12%. While Agarwal and Dixit [16] presented a technique for detecting false news that analyzes the context of brief sentences and news to generate a credibility score for both the news and its author. By extracting features and generating credibility scores from textual data. With a 97 percent accuracy rate, LSTM outperformed SVM, CNN, KNN, and NB.

3. RESEARCH METHOD

The contribution of this work is to investigate the efficiency between classical classifiers of NB and LSTM. In this section, an overview of the dataset is explained. Then we demonstrate the preprocessing techniques and features that were used to eliminate unessential plain text and convert the text into vectors of features. This is followed by a detailed section about the used classifiers and their parameters. Figure 1 represents the steps to achieve the objective of this study.

3.1. Dataset description

A specific dataset has been used in this work to address the fake news classification which was proposed by Kaggle and is openly available. The dataset contains 20,800 news about presidential elections in the US 2016, that are labeled with 1 if the article is true (real) and 0 if it is false (fake) with the following attributes: id, title, author, text, and label. The dataset is split into two CSV's files: training data (67%) and test data (33%).

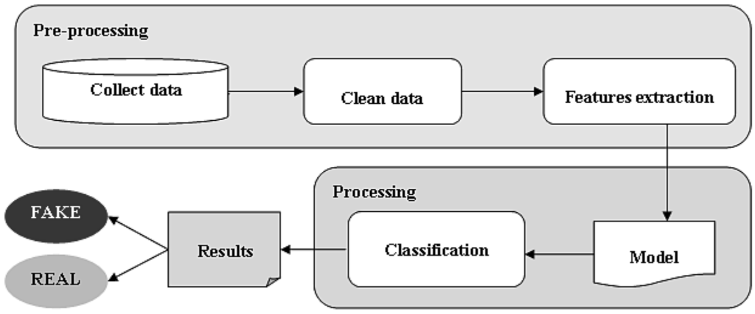

Figure 1. Fake news detection model

3.2. Pre-processing data

In the preprocessing step, it is important to transform the data text into a clean corpus before feeding the classifiers [24], To accomplish this, punctuation and special characters such as @, percent, &..., URLs, foreign language terms, emoji's, and unnecessary spaces from data were removed, followed by a list of common stop words used in the English language such as “the”, “a”, “an”, and “in” to reduce the size of the corpus used. Later, the library's PortStemmer has been imported to stem and transform the words to their base or origin form, taking into account a list of common prefixes and suffixes. Lastly, text should be divided into single words separated by white space. This step is known as tokenization.

3.3. Classification models

For classification techniques, we have chosen two algorithms: naïve bayes and long short-term memory LSTM. We explain the main idea and the results of each one. The following sections explain NB and LSTM details.

3.3.1. Naïve Bayes

The NB classifier is founded on the bayes theorem. Its prediction is on the erroneous assumption that all features are contained within themselves. In the Bayes theorem, P(c|x) is calculated using (1), where c denotes the class of possible outcomes and x denotes the specific occurrence to be classified [21]. The class is either 0 or 1, depending on the data, with 0 indicating false news and 1 indicating real news. We shall calculate both P(real news|N) and P(false news|x) for a given instance of news N. If P(real news|N) is greater than P(false news|N), the algorithm predicts that the news is true. Otherwise, the news will be deemed to be fake.

Table 1 represents the NB’s parameters used in this work. The count vectorizer model has been used for the embedding to generate a vector representation for each news with n-gram between 1 and 6 words. The max features were fixed to 5,000.

$$P(x) = P(c) * \frac{P(c)}{P(x)} \quad (1)$$

Table 1. Parameters used for NB

Parameters	Embedding	Max_features	n-gram_range
Values	count vectorizer	5,000	(1.6)

3.3.2. Long short-term memory (LSTM)

LSTM was introduced by [25] to address a mutual issue that RNNs frequently face. Because the RNN has restricted access to previous events, it must decide or predict an event that will not occur in the future. Simply put, RNN can only go back to a few states, whereas LSTM can trace back several states and observe what occurred, resulting in an effective prediction of what will happen in the future.

As depicted in Figure 2, the long-term state c(t-1) passes via forget-gate, which deletes some memories and replaces them with those chosen by the input gate. Following that, the result c(t-1) equals c (t). Additionally, c(t-1) is copied and transmitted via the tanh function, which filters the output gate to calculate the short-term state h(t), which is the output of the cell at the t time step, yt. In general, a basic RNN cell contains a single fully connected layer. However, the LSTM cell contains three additional gate controller layers, the output of which ranges from 0 to 1 due to the logistic activation function. When the zeros are released, the gate is closed; when one is released, the gate is open. In the long run, what should be deleted is determined by the forgotten door. The input-gate decides how the long-term status should be added. The output-gate determines which shares of the long-run state in the current time step should be read and output. Let x be a representation of the input sequence vector and W be the weights of each matrix element.

According to Table 2, each text was converted to a fixed dimensional vector using the word embedding “One Hot.” The model's first layer will be the embedding layer, which will receive input with a vocabulary size of 5,000 numbers; any list longer than 5,000 numbers will be truncated. For lists, less than 5,000 words in length, 0's have been added to the beginning of the list to maintain the same length size. Subsequently, a 100-neuron LSTM layer was added and employed sigmoid activation in the final layer. Later, the model was compiled using binary cross-entropy for loss function, 10 for epochs, “Adam” optimizer, and 64 for the batch size.

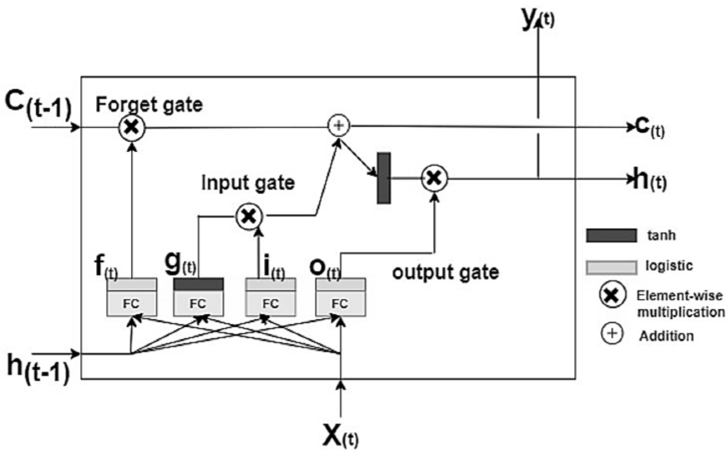

Figure 2. LSTM cell [26]

Table 2. Parameters used for LSTM

Parameters	Embeddings	Voc_size	Embeddings Vector Features	Activation Function	Batch size	epoch
Values	One Hot	5000	40	sigmoid	64	10

3.4. Evaluation metrics

Various metrics are used to assess the performance of this work. These metrics are based on the confusion matrix. This matrix depends on four important parameters: TP, TN, FP, and FN, which are explained in Table 3. Moreover, the accuracy (a) was the main performance metric selected to measure the model.

– Accuracy: It is the proportion of correctly classified classes to the total number of classes.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

– Recall (sensibility): It is the proportion of correctly classified positive cases to the total of positive cases that are correctly classified and negative cases that are incorrectly classified.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

– Precision: It is the proportion of correctly classified positive cases to the total of the positive cases that are correctly and incorrectly classified.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

– F1_score: It establishes the model's accuracy for each class.

$$F1_{score} = \frac{2}{\frac{1}{precision}+\frac{1}{recall}} = \frac{2*(precision*recall)}{precision+recall} \quad (5)$$

Table 3. Confusion matrix parameters

TP	FP	TN	FN
When a real article is correctly predicted	When a real article is wrongly predicted	When a fake article is correctly predicted	When a fake article is wrongly predicted

4. RESULTS AND DISCUSSION

Figure 3 represents confusion matrix with fake and real news using the NB algorithm in (a) and LSTM in (b). As shown in Figure 3(a), the results of the confusion matrix using the NB algorithm indicate that 355 real articles are classified as fake, while 246 fabricated articles are classified as authentic. Additionally, we observe that 3034 fake articles are truly fake, while 2,400 real articles are effectively legitimate. According to the confusion matrix in Figure 3(b) generated by the LSTM algorithm, 316 real articles are classified as fake, while 218 fraudulent articles are classified as real. Additionally, we observe that 3,118 fake articles are truly fake and 2,383 real articles are effectively legitimate.

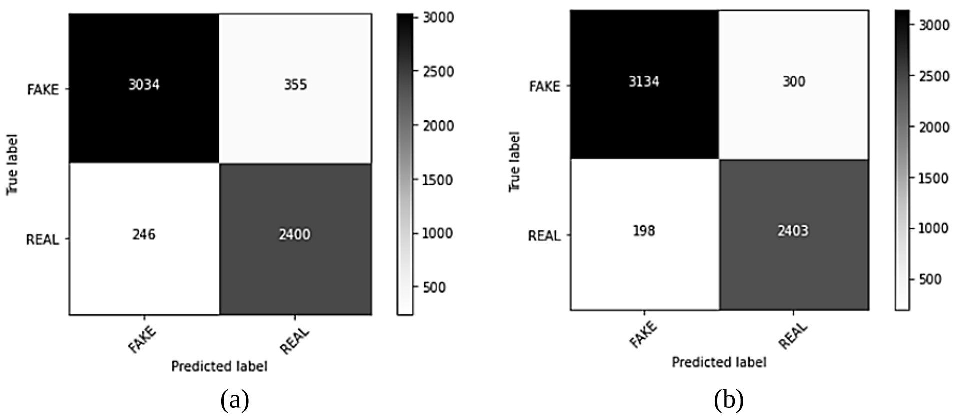

Figure 3. Confusion matrix of fake and real news using (a) NB algorithm and (b) LSTM algorithm

The LSTM model has an accuracy of 92 percent, which is only 2 percent higher than the NB model. The overall precision, recall, and F1-score for all classes are 91 percent, 94 percent, and 93 percent, with a 1 or 2 percent difference when compared to the NB model. Table 4 displays the comprehensive statistical results of our proposed models.

Table 4. NB and LSTM results

	NB (%)	LSTM (%)
Accuracy	90	92
Precision	90	91
Recall	93	94
F1-Score	91	93

As reflected in Figure 4, we observe that LSTM produces more true positive results than NB, implying that true news was predicted. While NB produces fewer false positives than LSTM, this implies that real news was predicted as fake. According to Table 5, the LSTM's primary strength is its ability to accurately analyze the dependency between sentences. Finally, we conclude that LSTM is more accurate than NB at detecting fake news.

In accordance with this study, a comparison with the existing studies [23] and [22] mentioned in the related work was done and summarized in Table 6. Both studies [22], [23] agreed that LSTM is a better classifier than naive bayes for detecting false and non-authentic news. Moreover, this research highlights the fact that our model is more effective than what was proposed in the compared studies [22], [23]. The difference might be due to one of the following reasons. First, none of the studies worked on a common dataset. Second, each study used a different classifier’s parameters. Third, multiple types of features extracted have been applied. This study can be extended in the future to apply our approach to different datasets and different feature extractions to validate the model.

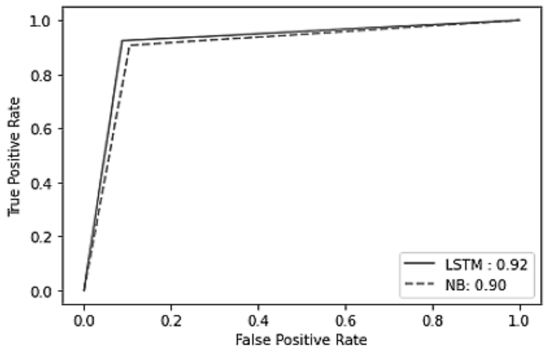

Figure 4. ROC (LSTM and NB)

Table 5. Comparison between NB and LSTM

	NB	LSTM
Features	Syntaxic (sequence of words)	Semantic (sequence of sentences)
Representation	vector	layer
Accuracy	weak	strong

Table 6. Comparative parameters and accuracy with some existing works results

Classifier	Parameters	[23]	[22]	Current study
LSTM	Accuracy	91.31%	82.29%	92%
	Embedding	Word2Vec	Not Mentioned	one hot
	epoch	5	Not Mentioned	10
Naïve Bayes	Accuracy	67.12%	67.89%	90%
	Input size vector	300	Not Mentioned	300
	Embedding	Doc2Vec	Not Mentioned	count vectorizer

5. CONCLUSION

The key objective of this study is to evaluate the performance and robustness of the classical machine learning model naïve bayes and the deep learning model LSTM. As a result, two models for detecting fake news were investigated: LSTM and NB. The results indicate that the LSTM algorithm outperformed better than the naive bayes algorithm. Finally, our proposed model established itself to be the finest model for false news identification compared with the existing state of art results on similar classifiers. In future research, our proposed model should be investigated further to determine how effectively the classifiers' parameters can be adjusted. Additionally, it is necessary to replicate our approach using different social media datasets in different languages, which necessitates the use of pre-processing methods to appropriately affirm and generalize our model's results.

REFERENCES

[1] X. Zhang and A. A. Ghorbani, "An overview of online fake news: characterization, detection, and discussion," *Information Processing and Management*, vol. 57, no. 2, Mar. 2020, doi: 10.1016/j.ipm.2019.03.004.
[2] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 211–236, May 2017, doi: 10.1257/jep.31.2.211.
[3] P. Rumman and M. Svärd, "Combating disinformation: detecting fake news with linguistic models and classification algorithms," *KTH Skolan för Datavetenskap Och Kommunikation*, pp. 1–31, 2017.
[4] N. Aslam, I. Ullah Khan, F. S. Alotaibi, L. A. Aldaej, and A. K. Aldubaikil, "Fake detect: a deep learning ensemble model for fake news detection," *Complexity*, vol. 2021, pp. 1–8, Apr. 2021, doi: 10.1155/2021/5557784.
[5] J. Nørregaard, B. D. Horne, and S. Adalı, "NELA-GT-2018: a large multi-labelled news dataset for the study of misinformation in news articles," in *Proceedings of the 13th International Conference on Web and Social Media, ICWSM 2019*, 2019, pp. 630–638.
[6] A. Thota, P. Tilak, S. Ahluwalia, N. Lohia, S. Ahluwalia, and N. Lohia, "Fake news detection: a deep learning approach," *SMU Data Science Review*, vol. 1, no. 3, pp. 1–20, 2018.
[7] R. Kibble, "Introduction to natural language processing." University of London, London, pp. 1–113, 2013.
[8] V. Rubin, N. Conroy, Y. Chen, and S. Cornwell, "Fake news or truth? using satirical cues to detect potentially misleading news," in *Proceedings of the Second Workshop on Computational Approaches to Deception Detection*, 2016, pp. 7–17, doi: 10.18653/v1/w16-0802.
[9] J. Thorne, M. Chen, G. Myrianthous, J. Pu, X. Wang, and A. Vlachos, "Fake news stance detection using stacked ensemble of classifiers," in *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*, 2017, pp. 80–83, doi: 10.18653/v1/W17-4214.
[10] R. Oshikawa, J. Qian, and W. Y. Wang, "A survey on natural language processing for fake news detection," in *LREC 2020 - 12th*

International Conference on Language Resources and Evaluation, Conference Proceedings, 2020, pp. 6086–6093.
[11] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, “Some like it hoax: automated fake news detection in social networks,” in *CEUR Workshop Proceedings*, Apr. 2017, pp. 1–12, [Online]. Available: http://arxiv.org/abs/1704.07506.
[12] J. P. Singh, A. Kumar, N. P. Rana, and Y. K. Dwivedi, “Attention-based LSTM network for rumor veracity estimation of tweets,” *Information Systems Frontiers*, Aug. 2020, doi: 10.1007/s10796-020-10040-5.
[13] R. Azad, B. Mohammed, R. Mahmud, L. Zrar, and S. Sdiq, “Fake news detection in low- resourced languages ‘Kurdish language’ using machine learning algorithms,” *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 6, pp. 4219–4225, 2021.
[14] Abdullah-All-Tanvir, E. M. Mahir, S. M. A. Huda, and S. Barua, “A hybrid approach for identifying authentic news using deep learning methods on popular Twitter threads,” in *2020 International Conference on Artificial Intelligence and Signal Processing (AISP)*, Jan. 2020, pp. 1–6, doi: 10.1109/AISP48273.2020.9073583.
[15] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, “Fake news detection using machine learning ensemble methods,” *Complexity*, vol. 2020, pp. 1–11, Oct. 2020, doi: 10.1155/2020/8885861.
[16] A. Agarwal and A. Dixit, “Fake news detection: an ensemble learning approach,” in *Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020*, May 2020, pp. 1178–1183, doi: 10.1109/ICICCS48265.2020.9121030.
[17] Y. Liu, X. Jin, and H. Shen, “Towards early identification of online rumors based on long short-term memory networks,” *Information Processing and Management*, vol. 56, no. 4, pp. 1457–1467, Jul. 2019, doi: 10.1016/j.ipm.2018.11.003.
[18] A. Verma, V. Mittal, and S. Dawn, “FIND: fake information and news detections using deep learning,” in *2019 Twelfth International Conference on Contemporary Computing (IC3)*, Aug. 2019, pp. 1–7, doi: 10.1109/IC3.2019.8844892.
[19] O. Ajao, D. Bhowmik, and S. Zargari, “Fake news identification on Twitter with hybrid CNN and RNN models,” in *Proceedings of the 9th International Conference on Social Media and Society*, Jul. 2018, pp. 226–230, doi: 10.1145/3217804.3217917.
[20] M. Z. Asghar, A. Habib, A. Habib, A. Khan, R. Ali, and A. Khattak, “Exploring deep neural networks for rumor detection,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 4, pp. 4315–4333, Apr. 2021, doi: 10.1007/s12652-019-01527-4.
[21] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, and B.-W. On, “Fake news stance detection using deep learning architecture (CNN-LSTM),” *IEEE Access*, vol. 8, pp. 156695–156706, 2020, doi: 10.1109/ACCESS.2020.3019735.
[22] W. Han and V. Mehta, “Fake news detection in social networks using machine learning and deep learning: performance evaluation,” in *2019 IEEE International Conference on Industrial Internet (ICII)*, Nov. 2019, pp. 375–380, doi: 10.1109/ICII.2019.00070.
[23] S. A. Alameri and M. Mohd, “Comparison of fake news detection using machine learning and deep learning techniques,” Jan. 2021, doi: 10.1109/CRC50527.2021.9392458.
[24] A. Jain, G. Kulkarni, and V. Shah, “Natural language processing,” *International Journal of Computer Sciences and Engineering*, vol. 6, no. 1, pp. 161–167, Jan. 2018, doi: 10.26438/ijcse/v6i1.161167.
[25] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
[26] T. Jiang, J. P. Li, A. U. Haq, A. Saboor, and A. Ali, “A novel stacking approach for accurate detection of fake news,” *IEEE Access*, vol. 9, pp. 22626–22639, 2021, doi: 10.1109/ACCESS.2021.3056079.

BIOGRAPHIES OF AUTHORS

Sarra Senhadji Associate Professor since 2015, Computer Science Department, Faculty of Mathematics and Informatics, USTO, Oran, Algeria. Member of LSSD Laboratory (Laboratory Signal, System and Data). Her research interest in Data Mining, Data Replication, Big Data, and Machine Learning. She can be contacted at email: sarra.senhadji@univ-usto.dz.

Rania Azad San Ahmed A Lecturer at the Computer Network Department, Sulaimani Polytechnic University (SPU), Iraq since 2013. she received a Bachelor's Degree in Math and Computer Science from the University of Science and Technology of Oran (USTO) in 2007 and Master's Degree from the same University in 2009 with a thesis on "Localization of MAC address using SNMP Protocol" with the rank 8th. From 2016 to 2020, she was Director of Kurdistan Journal of Applied Research (KJAR). In Parallel, she served as Library and Quality Assurance Coordinator at the American University of Iraq since 2013. her current research interests are broad and varied, but generally fall into these areas: Fake news detection, fact-checking, web credibility, and Sentiment analysis. Previously, she have worked on Network Management and Security. She can be contacted at email: rania.azad@spu.edu.iq. ResearcherID: C-2826-2018.

View publication stats