Indonesian Journal of Electrical Engineering and Computer Science
Vol. 36, No. 3, December 2024, pp. 1816~1826
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v36.i3.pp1816-1826

A robust method for detecting fake news using both machine and deep learning algorithms

Enas Ahmad Alikhashashneh[1], Khalid M.O. Nahar[2,3], Mohammed Abual-Rub[4], Hedaya M. Alkhaldy[1]
[1]Department of Information Systems, Faculty of Information Technology and Computer Science, Yarmouk University, Irbid, Jordan
[2]Faculty of Computer Studies, Arab Open University, Riyadh, Saudi Arabia
[3]Department of Computer Science, Faculty of Information Technology and Computer Sciences, Yarmouk University, Irbid, Jordan
[4]Department of Management Information Systems, School of Business, King Faisal University, Hofuf, Saudi Arabia

Article Info

Article history:

Received Feb 15, 2024
Revised Aug 10, 2024
Accepted Aug 26, 2024

Keywords:

Deep learning
Fake news
Information technology
Machine learning
Social media
TF-IDF
Word2Vec

ABSTRACT

Spreading fake news and false information on social media is very common and can be done effortlessly due to the huge number of users of each of the various social media platforms. Another reason for having such a speedy spread of fake news (which makes about 40% of the information published on social media platforms) is the inability of these platform to verify the authenticity of the news before allowing it to be published. This research will use information technology to detect fake news/ false information and change this kind of technology from being the cause of the problem to a tool to solve it. This research provides a method that uses both machine learning (ML) and deep learning (DL) algorithms to detect fake information versus real information and compare the performance of the algorithms. The results of this research indicate that the algorithms that use term frequency inverse document frequency (TF-IDF) have achieved better results than the algorithms that use Word2Vec. Long short-term memory (LSTM) algorithm, however, has achieved the best performance; of 99% accuracy -when using TF-IDF, and 94% -when using Word2Vec.

Corresponding Author:

Enas Ahmad Alikhashashneh
Department of Information Systems, Faculty of Information Technology and Computer Science
Yarmouk University
Irbid, 21163, Jordan
Email: enas.a@yu.edu.jo

1. INTRODUCTION

Given the tremendous development we are witnessing in the field of information technology, which has facilitated and accelerated the access of information to all people, and because information is the true essence that we seek as students, researchers, and even teachers, it is significant to ensure the reliability of this information that flows to us from everywhere. We have recently witnessed an alarming increase in the use of social media sites by people of all ages and cultures. This has caused the rapid and easy spread of information regardless of its accuracy and reliability. A recent study has found that fake news is spread six times faster than real news, and that 70% of this fake news is republished by users via social media [1].

The danger of fake news stems from the fact that it is created to mislead or deceive the reader or to influence people's opinions, whether for political or financial purposes [2]. People who spread fake news use several methods to deceive readers and make them believe this news, such as using bots to create multiple accounts on social media sites and increase the number of "shares" of such news. Another method to spread fake news is by creating advertisements on fake websites that resemble the official news sites [3].

Some methods have been used to detect the accuracy and reliability of the news posted on these sites. A method that has shown accurate results in this field is machine language. This method branches off from artificial intelligence; it uses data as the core of its work. The method collects data and then processes and analyzes it to use for prediction. Algorithms mimic the way in which humans learn; the thing that helps produce very accurate results.

Machine learning (ML) algorithms have many uses in forecasting fields such as fire incidents visualization and pattern recognition [4], a wind turbine with a yaw mechanism [5], prediction of early-stage diabetes [6], Weather prediction [7], the nexus of corruption and non-performing loan [8], Keylogger detection [9], and many other areas. However, there are many ML algorithms that can be utilized to identify fake news. The accuracy of each algorithm varies for the following reasons: the specific problem, type and size of data, and the experience of data scientists. This research provides a comparison of different ML; algorithms that detect fake news; namely logistic regression (LR), Gaussian Naïve Bayes (NB), multinomial NB, support vector machine (SVM), long-short-term memory (LSTM), and capsule network. A dataset containing fake news and real news was used in this comparison.

Accuracy should be met when comparing ML algorithms to get precise results. The research conducted by Sharma *et al.* [10], it was found that the best accurate algorithm is LR-with an accuracy of 65%. The researchers have increased the accuracy by improving the search parameter, the accuracy of the algorithm reached 75%. The accuracy of ML algorithms in classifying text data varies depending on the feature extraction method and the specific algorithm. The faster the extraction method, the more efficiently it converts the text into a machine-readable format. This also applies to ML algorithms, where the accuracy of the results varies depending on the training data. This has been observed by the research conducted by Tiwari *et al.* [11]; the results this research have reached indicate that the LR algorithm has the highest accuracy among other algorithms-achieving 71%. This high accuracy percentage is reached due to the use of term frequency inverse document frequency (TF-IDF) vector.

Every algorithm should preprocess the data so that it is easier to deal with inappropriate data. Pandey *et al.* [12] have proposed a model that proves that the algorithm with the best accuracy was LR-with 90.46%, then KNN with 98.98%, followed by SVM with 89.33%, followed by NB with 86.89%, and finally decision tree with 73.33%. Katsaros *et al.* [13], however, indicates that the longer the training time for the data is, the higher the accuracy of the algorithms is achieved. Katsaros *et al.* [13] have evaluated eight algorithms: namely regression, support vector classification, multiclass perceptron, NB, decision trees, polynomial, random forests, and convolutional neural networks (CNNs) using three sets of fake news data; different in the efficiency and speed of training each algorithm. The CNN algorithm has shown the best results in detecting fake news-despite of the fact that it requires a relatively large training time.

Abdulaziz Al-Bahr and Marwan Al-Bahr have collected a public dataset containing fake and false news. They have applied four main ML algorithms: random forest, NB, neural network, and decision trees to detect fake news. They have found out that the NB algorithm is the best in terms of accuracy. The reason why this algorithm obtained the highest accuracy -as explained by these researchers- is the high rate in calculating conditional probabilities based on the occurrence of the text or news individually, as it achieved a rate of 99% [14]. Jain and Kasbe [15], on the other hand, have used the NB algorithm to detect fake news among Facebook posts. They have confirmed that the NB algorithm has shown an accuracy of 80.6% for titles, and 91.2% accuracy for text titles. When using the n grams technique, however, the text accuracy has increased to 93.1%. The titles' accuracy, on the other hand, has increased slightly to 80.7%.

The main motivation for using ML to detect fake news is to provide a powerful solution to users and help them not fall into the trap of misinformation and lies. Nagaraja *et al.* [16] have applied ML algorithms on a set of data using semantic measurement. Their research indicates the following results: the SVM algorithm is the most accurate- with an accuracy of 75%- while the accuracy of NB is 63%. The reason behind these results- as stated by Nagaraja *et al.* [16] is that NB separates the measurement features. In addition, SVM has shown excellent results in detecting fake news. This is confirmed in the research paper conducted by Baarir and Djeffal [17]. These researchers have used a bag of words for the most common words in addition to the ngram technique. To obtain a good data recognition rate, Shaikh and Patil [18] have proposed a model that applies the following methods and to achieve the following accuracy rates: SVM indicates the highest accuracy of 95.05%, passive aggressive has achieved 92.9%, and NB has achieved 84.056.

X (formerly known as Twitter) is another platform that can contribute to spreading news rapidly since it achieves widespread circulation and is used by international celebrities. For this reason, Mahir *et al.* [19] have collected data from the X platform by using ML algorithms that SVM, NB method, LR, and recurrent neural network (RNN) separately. They have compared the accuracy of each algorithm and have stated the following accuracy rates: NB: 89.6%, SVM: 89.34%, LR: 69.47%, the neural network: 74% -by RNN, and LSTM:78%. NB and the SVM algorithm -as indicated by the results- have obtained the best results.

Jain *et al.* [20] propose a model that uses the following ML algorithms: NB, SVM, and NLP. This model has achieved a high accuracy rate- of 93.6%. Ahmed *et al.* [21] have proved that using types of algorithms leads to obtaining high accuracy. A group of ML algorithms have been used to detect fake news and have achieved the following accuracy rates: random forest algorithm and Perez-LSVM have achieved 99%, linear SVM, multi-layer perceptron, bagging classifiers, and boosting classifiers have achieved 98% accuracy. Moreover, there were many differences in the results for each group, all algorithms were applied to them individually and collectively on 4 datasets. According to their results, in all performance measures, the algorithms in the ensemble were better compared to the single algorithm, but at the single algorithm level, LR had the best accuracy, with an average accuracy of more than 90%.

In their study, Khanam *et al.* [22] have used the following algorithms on a set of data: XGBoost, random forests, NB, K-nearest neighbors (KNN), decision tree, and SVM. The main goal was to try to obtain a classification model that is highly accurate in detecting true and fake news. The accuracy results of all algorithms are very close; the SVM algorithm has high accuracy, as it obtained a rate of 73%, which is the second after the algorithm that obtained the highest accuracy, XGBoost, which achieved 75%, while NB obtained a rate of 68%.

Ahmed and Ramasamy [23] predict that SVM would be the best algorithm -since it does not use probability distributions, but rather uses the concept of hyperplanes. This is also seen in the NB algorithm that has conducted an experiment using twenty thousand samples of fake news (collected from several sources). This experiment shows that the neural network model has shown better accuracy results than the rest of the algorithms. NB algorithm requires high computational requirements, as it has achieved an accuracy of 92%, SVM, on the other hand, has obtained 76.8%, and NB has obtained 68.4%. The performance of ML algorithms in detecting fake news significantly outperforms human ability-according to Waikuma and Goswam [24]. These researchers have successfully used a combination of machine data algorithms and have reached the following accuracy rates: the LIAR ensemble, the bagging classifier, and XGBoost algorithms have obtained accuracy of 39%, and random forest has achieved an accuracy of 37%. However, Gradient Boost has achieved an overall accuracy of 36% -in terms of recall and f1. However, after reclassifying, using binary classification that attempts to simplify the data, and then re-running the algorithms, the percentages changed, with Bagging and AdaBoost at 70%, and random forest at 65%. But both Extra Trees and XGBoost reached 62%.

Ouassil *et al.* [25], the authors used a combination of different word embedding techniques and a hybrid CNN and bidirectional long short-term memory (Bi-LSTM) model to detect fake news. They trained their approach on the WELFake dataset. The best accuracy of the model was 97%. The best model was generated by combining a pre-trained Word2Vec CBOW model and a Word2Vec skip-word model with a CNN on Bi-LSTM layers.

Most previous studies have tried to detect fake news using traditional ML techniques. As far as we know, none of them have compared traditional techniques like SVM, LR, Gaussian NB, and multinomial NB with newer approaches like capsule networks and LSTM. The comparison was made after applying either the TF-IDF or Word2Vec to achieve the best accuracy in detecting fake news. Furthermore, this paper distinguishes itself from previous studies by using a new dataset for training and evaluating both traditional and newer techniques. The main contributions of this research are:

a) Modern technologies have been used such as artificial intelligence, natural language processing (NLP), and data analytics to detect fake information, which produced excellent results.
b) This research pushes the boundaries of fake news and real news classification beyond traditional approaches using ML and DL algorithms. These advanced techniques allow ML from large datasets, which in turn allows for high-accuracy fake news detection. This is largely based on different linguistic and semantic features.
c) The use of Word2Vec and TF-IDF led to enhanced semantic understanding of textual content, which greatly contributed to a more complex analysis of the characteristics of fake news. Word2Vec embeddings were used that capture the semantic similarity between words, identifying the importance of Terms measured by TF-IDF weights, which focus on the frequency of terms in documents.
d) This research includes various evaluations of the performance of ML and deep learning (DL) models in detecting fake news. It provides metrics such as accuracy, precision, recall, and F1-score. Therefore, this research contributes to the scientific understanding of the effectiveness of these methods in real-world scenarios. It proves its great ability to perform such tasks, as it has achieved high accuracy.

The remaining sections of this paper our are presented as follows: section 2 explains the basic methodology. In section 3 presents and discusses the results obtained after implementing the proposed method. Finally, section 4 presents the conclusions of our proposed model and future works.

2. METHOD

The method we propose and use to address the main contributions of this study is divided into four main stages as shown in Figure 1. The details of the four stages are explained in the following subsections.

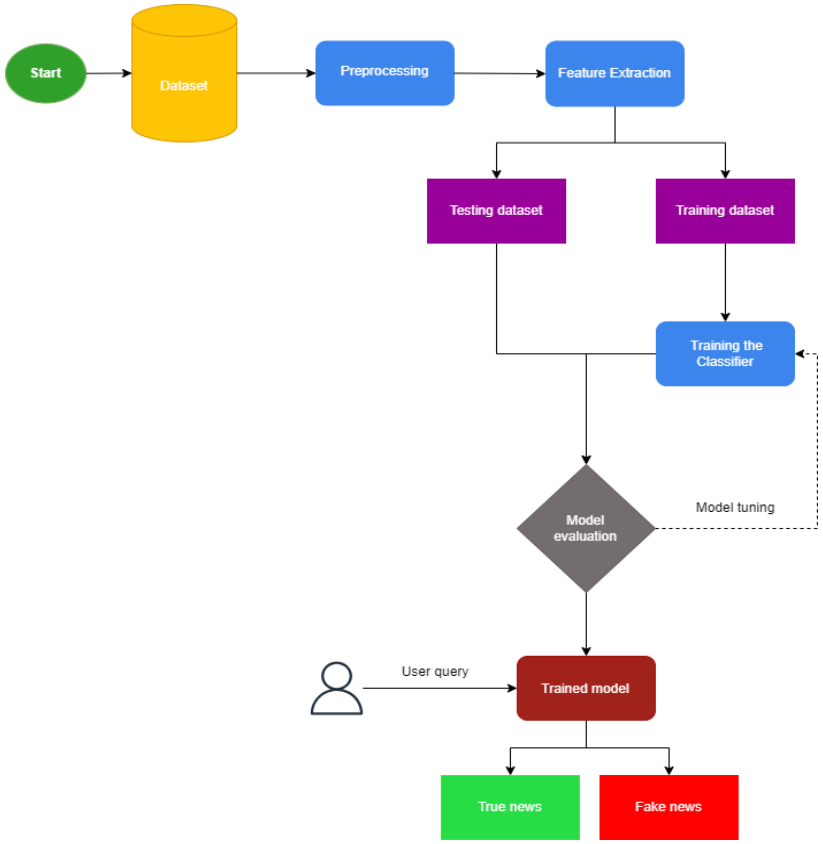

Figure 1. The stages of the proposed method

2.1. Dataset

The dataset used in this proposed method is the unbiased WELFake dataset, which consists of 72,134 news articles containing 37,106 fake news and 35,028 real news. This dataset has been carefully curated, classified and checked by journalists to ensure the validity of the data. The WELFake dataset is a combination of four distinct sources: Reuters, McIntyre, Cagle, and BuzzFeed Policy [26]. The goal of the merging strategy is to avoid overfitting classifiers.

2.2. Preprocessing

The data is prepared in this step, as the data is usually in a format that is not analyzable, so the pre-processing stage plays an important role in converting this data into a format that is easy to handle and raising the quality of the input data, thus giving ML and DL algorithms an opportunity to perform more powerfully and accurately, as this was done in this step. The stage cleans the texts by removing all punctuation and stop words, converting all texts to lowercase, and then creating a new "fulltext" column that contains each text combined with its title.

2.3. Feature extraction

Feature extraction is a vital process that enables us to convert the data into digital features so that it can be processed without affecting the original data set and thus meaningful information is extracted from pre-processed textual data. In our proposed method, we have used advanced and most widely used feature extraction techniques, and this includes but is not limited to:

– TF-IDF vector: text data is converted into digital vectors and explains the importance of words in distinguishing between real news and fake news.

– Word2Vec for word embedding: each word is represented as a vector and thus can be used to represent the semantic relationships between words.

2.4. Building the method

After sorting the data into two categories, training and testing, the proposed method model was built using ML and DL algorithms (LR, Gaussian NB, Polynomial NB, SVM, Capsule neural network (CapsNet), LSTM, and the model was built once using TF-IDF and once using Word2Vec.

2.4.1. Logistic regression

LR is a supervised ML algorithm that predicts the probability of a binary outcome by analyzing the relationships between independent variables. The following formula is the basic formula for LR by (1):

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

this represents a LR equation, where *e* represents the natural logarithm base, and *x* represents the transformed numerical input [27]. LR was chosen because it is simple and easy to interpret, in addition to its suitability for binary classification tasks.

2.4.2. Gaussian NB

Gaussian NB is a type of continuous probability distribution and is used in classification tasks. Therefore, it was chosen in the proposed method as it is characterized by simplicity, efficiency, and high ability to deal with data. Its importance comes from the central limit theorem, where the average of many Samples from a random variable with a given mean and variance are themselves a random variable - their distribution as the number of samples increases converges to a normal distribution. The following formula is the basic formula for Gaussian NB by (2):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_{y,i}^2}} e^{-\frac{(x_i-\mu_{y,i})^2}{2\sigma_{y,i}^2}} \tag{2}$$

where x_i represent the specific input value and yi represent class label [28].

2.4.3. Multinomial NB

The multinomial NB algorithm is a learning method based on the principle of probability that is mostly used in NLP. It is an algorithm that is based mainly on Bayes' theory, where it predicts a text tag such as an email or a newspaper article. It is very effective in dealing with data, especially data with multiple categories, and is characterized by simplicity in modeling, so it was chosen in the proposed method. It calculates the probability of an event occurring based on its prior knowledge of the circumstances that are related to the event. It is based on the following formula that helps calculate the probability of textual tags by (3):

$$F(A|B) = \frac{F(A)\,.\,F(B|A)}{F(B)} \tag{3}$$

we calculate the probability of class A when predictor B is already provided.

2.4.4. SVM

SVM are supervised learning algorithms used for classification, regression, and outlier detection. The primary goal of SVMs is to find a hyperplane that best separates data into different classes. The SVM was chosen in the proposed method due to its ability to work well in high-dimensional spaces and its flexibility in dealing with different data, and it recorded high performance in data classification. The following formula is the basic formula for linear SVM by (4):

$$f(x) = sign(\sum_{i=1}^{n} w_i.x_i + b)) \tag{4}$$

where "x" represents the input data, "w" represents the weight for each feature, and "b" is the bias term [28].

2.4.5. Word2Vec

Word2Vec is a neural network-based algorithm that learns word embedding through digital representations of words that capture their semantic and syntactic relationships. It creates vectors of words that are distributed digitally. The embedding process helps create an association of words in the context with similar meaning through vectors. The following formula is the basic formula for Word2Vec by (5):

$$P(c|w) = \frac{\sum_{i=1}^{v} \exp(v_i.v_w)}{\exp(v_c.v_w)} \quad (5)$$

where $P(c|w)$ represents the conditional probability of the context word c according to the target word w. So v_c represents the context word c and v_w represents the target word w, but $exp(x)$ represents the exponential function.

2.4.6. Capsule neural network

Capsule neural network is a type of artificial neural networks created to enhance the modeling of hierarchical relationships, and its goal is to organize biological neural systems so closely and robustly that they can understand biologically complex visual hierarchies and has a very high ability to recognize complex patterns and relationships in data. Therefore, they are chosen in the proposed method. The following formula is the basic formula for CapsNet by (6):

$$u_{i,j} = W_{i,j} . x_i \quad (6)$$

where $u_{i,j}$ represents the prediction vector from capsule i to capsule j, x_i represents the input data, and w_i represents the transformation matrix.

2.4.7. LSTM

LSTM is an algorithmic RNN algorithm designed to address the vanishing gradient problem inherent in RNNs. It has a great ability to maintain sensitivity to long-term dependencies and is responsible for managing the flow of information within the network. It is built from an A cell and three gates (input, output, and forget). It was chosen in the proposed method because it is very suitable for NLP, as it has the ability to understand and process complex sequences of information. The following formula is the basic formula for LSTM by (7):

$$h_i = \sigma(W . [h_{i-1}, x_i] + b \quad (7)$$

where x_i represents the input data, w represents the weight matrix, h_{i-1} represents the previous hidden state, and b represents the bias vector. Therefore, using a set of ML methods that are characterized by their simplicity and interpretability, and using a set of DL that helps to identify complex patterns and dependencies within textual data, will enable us to detect fake news with accuracy and high performance, and explore the impact of TF-IDF and Word2Vec on performance and the comparison between them. For the performance evaluation of the proposed method, the dataset was split by 75 to 25 in terms of training to testing news.

3. RESULTS AND DISCUSSION

This section demenostrates the results of applying the proposed method to the dataset, which -as mentioned in the previous section- is a news data set divided into real news and fake news. The dataset was divided into training (75%) and testing (25%). Figure 2 shows the distribution of the classes in the dataset after the preprocessing that is performed, where the size of the dataset has changed from the shape (72134, 4) -before the preprocessing process– to the shape (71537, 4) -after the preprocessing process has been applied. Furthermore, Figure 2 demonstrates that there is no imbalance problem in the dataset. Figure 3 depicts the distribution of the news word count.

In the proposed method, the following algorithms are used: SVM, LR, Gaussian NB, multinomial NB, LSTM, and capsule network to predict fake news using TF-IDF and using Word2Vec. A comparison of the performance of differetnt algorithms has been done using accuracy, precision, recall, and F1-score. The difference of perforamcne is observed when using TF-IDF and Word2Vec. In addition, cross-validation has been used to reduce overfitting. Table 1 shows a comparison of the performance of the algorithms when using TF-IDF, and Table 2 shows a comparison of the performance of the algorithms when using Word2Vec.

From Tables 1 and 2, we can conclude that the performance of the six classifiers when using TF-IDF compared to using Word2Vec is significantly different. Figure 4 and Table 1 depict that among the

six classifiers, LSTM gives the most prominent result, with precision, recall, and F1-score all reaching 99%. However, it achieved 98% in terms of precision. Likewise, capsule network has achieved 97% accuracy, SVM has achieved 96% accuracy, LR has achieved 95%, multinomial NB achieved 85%, and Gaussian NB has achieved 83% -when using TF-IDF. Table 2 and Figure 5, demonstrate that when using Word2Vec, LSTM, and capsule network an accuracy level of 95% has been achieved. SVM has achieved 93%, LR has achieved 63% accuracy, multinomial NB, and Gaussian NB have achieved 58%.

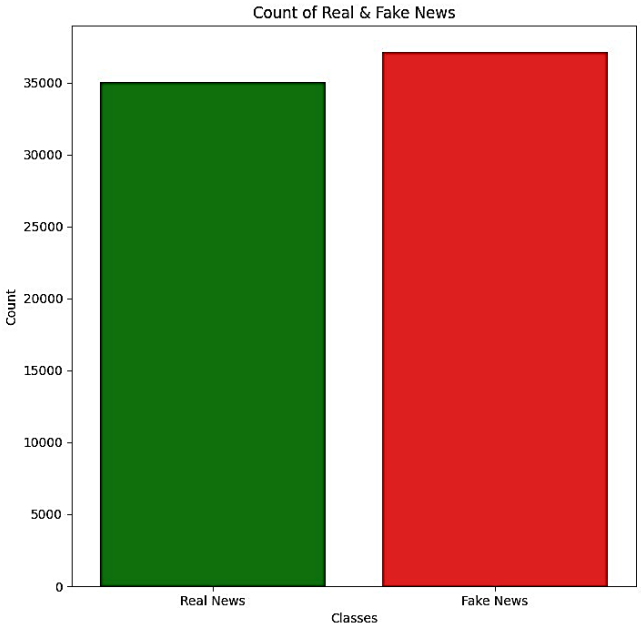

Figure 2. The distribution of the data in the data set

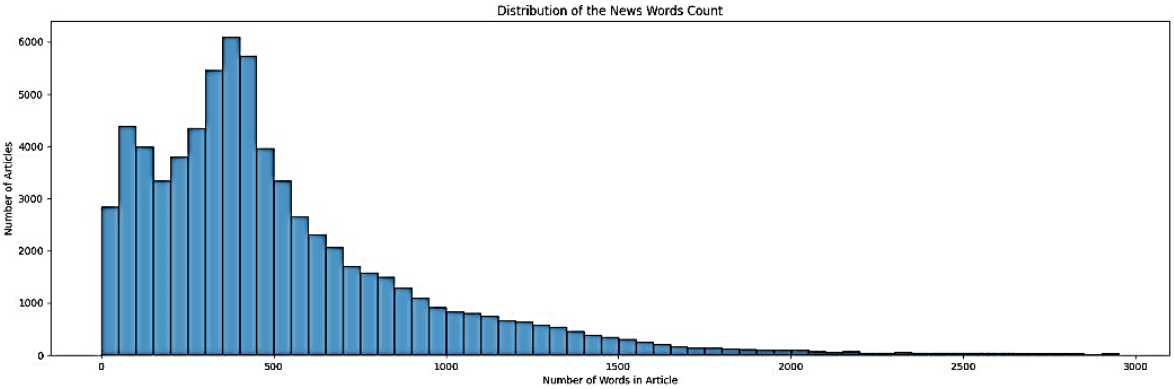

Figure 3. The distribution of the news word count

Table 1. Comparison of the performance of all algorithms using TF-IDF

Classifier	Accuracy	Precision	Recall	F1-score
SVM	0.96	0.96	0.96	0.96
Logistic regression	0.95	0.95	0.95	0.95
Gaussian NB	0.83	0.83	0.83	0.83
Multinomial NB	0.85	0.85	0.85	0.85
LSTM	0.99	0.98	0.99	0.99
Capsule network	0.97	0.95	0.99	0.97

Table 2. Comparison of the performance of all algorithms using Word2Vec

Classifier	Accuracy	Precision	Recall	F1-score
SVM	0.93	0.93	0.93	0.93
Logistic regression	0.63	0.62	0.63	0.62
Gaussian NB	0.58	0.58	0.58	0.57
Multinomial NB	0.58	0.58	0.58	0.57
LSTM	0.94	0.93	0.96	0.94
Capsule network	0.94	0.94	0.95	0.94

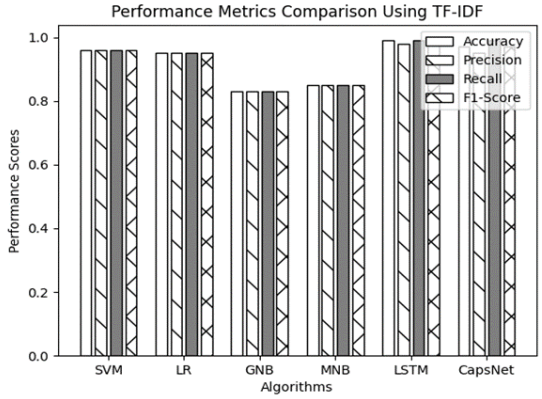

Figure 4. The performance metrics comparison using TF-IDF

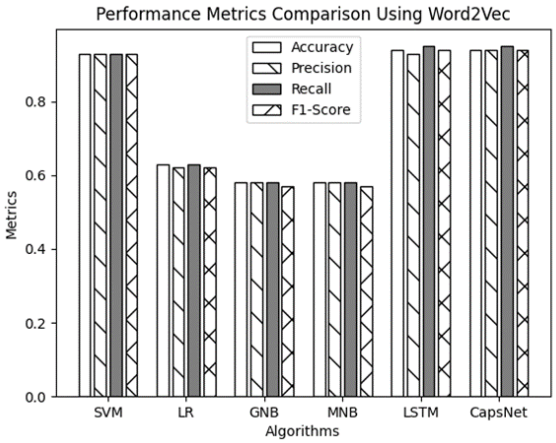

Figure 5. The performance metrics comparison using World2Vec

From Figure 6 we can conclude that using TF-IDF with algorithms has achieved higher accuracy than Word2Vec - as indicated in Tables 1 and 2. This is due to the following reasons: TF-IDF, (which depends on the frequency of the word in the text) focuses on the importance of the word, while Word2Vec depends on the semantic relationships between words and creates continuous representations for all words. The representations created by this method are less dynamic than the representations created by TF-IDF. The performance of Word2Vec is affected by the size of the vector, while the performance of TF-IDF does not depend on the size of the vector.

Moreover, we carried out a comparison between our proposed method which is classification using ML classifiers and DL. Additionally, we compared our results with some other research articles which worked on the same dataset. In Ouassil *et al.* [25], researchers achieved an accuracy of 97% by combining a pre-trained Word2Vec CBOW model and Word2Vec Skip-Word model with CNN on BILSTM layers. Our proposed method provided an improved result for the LSTM with TF-IDF.

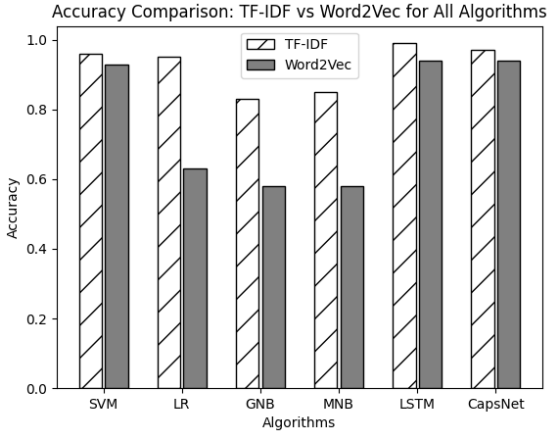

Figure 6. Accuracy comparison: TF-IDF versus Word2Vec for all algorithms

Finally, the experimental results show that we achieved the goal of this research, which is to reach the highest accuracy in detecting fake news. Word2Vec with algorithms has been used instead of TF-IDF. However, using TF-IDF with algorithms has achieved higher accuracy than Word2Vec - as indicated in Tables 1 and 2. In comparing the performance of the algorithms that have been used with TF-IDF, the following results are found: LSTM has achieved the highest accuracy rate of 99%, followed by capsule network with 97% accuracy, SVM achieving 96%, LR achieving 95%, multinomial NB achieving 85%, and Gaussian NB achieving 83%. When using Word2Vec, the accuracy has decreased significantly in some algorithms. LSTM and capsule network have achieved 94% accuracy, and SVM has achieved 93% accuracy. The rest of the algorithms, on the other hand, have not performed well; LR has achieved 63% accuracy, and both multinomial NB and Gaussian Naive have achieved 58% Bayes. The LSTM algorithm achieved the best accuracy in both experiments, so it is considered the best. Capsule network and SVM have not achieved high accuracy rates when using Word2Vec although their performance is better than other algorithms.

4. CONCLUSION

The world is currently suffering from a massive spread of fake news, which has negative effects on societies and may lead to significantly negative consequences. In this research, a proposal was presented to detect fake news and true news using both DL and ML algorithms. The following methods have been used in this research: SVM, LR, Gaussian NB, multinomial NB, LSTM, and capsule network. This research indicates that using algorithms with TF-IDF gives better accuracy results than using Word2Vec; LSTM has achieved the highest accuracy of 99% using TF- IDF. In addition, LSTM and capsule network have achieved the highest accuracy of 94% using Word2Vec. The experimental results show that the goal of this research, which is to reach the highest accuracy in detecting fake news, was achieved. For future studies, the researchers of this paper will aim to add semi-supervised learning to improve and strengthen the proposed method.

ACKNOWLEDGEMENTS

The authors extend their appreciation to the Arab Open University for funding this work.

REFERENCES

[1] P. Dhiman, A. Kaur, C. Iwendi, and S. K. Mohan, "A scientometric analysis of deep learning approaches for detecting fake news," *Electronics*, vol. 12, no. 4, p. 948, 2023, doi: 10.3390/electronics12040948.

[2] F. Zimmer, K. Scheibe, M. Stock, and W. G. Stock, "Fake news in social media: bad algorithms or biased users?," *Journal of Information Science Theory and Practice*, vol. 7, no. 2, pp. 40-53, 2019, doi: 10.1633/JISTaP.2019.7.2.4.

[3] N. Hajli, U. Saeed, M. Tajvidi, and F. Shirazi, "Social bots and the spread of disinformation in social media: the challenges of artificial intelligence," *British Journal of Management*, vol. 33, no. 3, pp. 1238-1253, 2022, doi: 10.1111/1467-8551.12554.

[4] J. R. Asor, J. L. Lerios, S. B. Sapin, J. O. Padallan, and C. A. C. Buama, "Fire incidents visualization and pattern recognition using machine learning algorithms," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 22, no. 3, pp. 1427-1435, 2021, doi: 10.11591/ijeecs.v22.i3.pp1427-1435.

[5] S. B. Abhi *et al.*, "An intelligent wind turbine with yaw mechanism using machine learning to reduce high-cost sensors quantity," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 31, no. 1, pp. 10-21, 2023, doi: 10.11591/ijeecs.v31.i1.pp10-21.

[6] R. M. H. L. Flores, L. H. Macatangay, S. M. G. Garcia, and J. R. Melo, "Comparative analysis in the prediction of early-stage diabetes using multiple machine learning techniques," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 32, no. 2, p. 13, November 2023, doi: 10.11591/ijeecs.v32.i2.
[7] B. Govardhan and E. A. Babu, "Weather prediction using random forest machine learning model," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, 2021, pp. 1208-1215, doi: 10.11591/ijeecs.v22.i2.pp1208-1215.
[8] S. Masrom, R. A. Rahman, N. H. M. Shukri, N. A. Yahya, M. Z. A. Rashid, and N. B. Zakaria, "The nexus of corruption and non-performing loan: machine learning approach," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 32, no. 2, pp. 838-844, 2023, doi: 10.11591/ijeecs.v32.i2.pp838-844.
[9] M. S. Alsubaie, S. H. Atawneh, and M. S. Abual-Rub, "Building machine learning model with hybrid feature selection technique for keylogger detection," *International Journal of Advances in Soft Computing and Its Applications*, vol. 15, no. 2, 2023, doi: 10.15849/ IJASCA.230720.03.
[10] U. Sharma, S. Saran, and S. M. Patil, "Fake news detection using machine learning algorithms," *International Journal of Creative Research Thoughts (IJCRT)*, vol. 8, no. 6, pp. 509-518, 2020, doi: 10.22214/ijraset.2020.6125.
[11] V. Tiwari, R. G. Lennon, and T. Dowling, "Not everything you read is true! Fake news detection using machine learning algorithms," in *2020 31st Irish Signals and Systems Conference (ISSC)*, 2020: IEEE, pp. 1-4, doi: 10.1109/ISSC49989.2020.9180206.
[12] S. Pandey, S. Prabhakaran, N. S. Reddy, and D. Acharya, "Fake news detection from online media using machine learning classifiers," in *Journal of Physics: Conference Series*, vol. 2161, no. 1, p. 012027, 2022, doi: 10.1088/1742-6596/2161/1/012027.
[13] D. Katsaros, G. Stavropoulos, and D. Papakostas, "Which machine learning paradigm for fake news detection?," in *IEEE/WIC/ACM International Conference on Web Intelligence*, 2019, pp. 383-387.
[14] A. Albahr and M. Albahar, "An empirical comparison of fake news detection using different machine learning algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 9, 2020, doi: 10.1145/3350546.3352552.
[15] A. Jain and A. Kasbe, "Fake news detection," in *2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, 2018, IEEE, pp. 1-5, doi: 10.1109/SCEECS.2018.8546944.
[16] A. Nagaraja, K. N. Soumya, A. Sinha, J. V. R. Kumar, and P. Nayak, "Fake news detection using machine learning methods," in *International Conference on Data Science, E-learning and Information Systems* 2021, 2021, pp. 185-192, doi: 10.1145/3460620.3460753.
[17] N. F. Baarir and A. Djeffal, "Fake news detection using machine learning," in *2020 2nd International workshop on human-centric smart environments for health and well-being (IHSH)*, 2021, IEEE, pp. 125-130, doi: 10.1109/IHSH51661.2021.9378748.
[18] J. Shaikh and R. Patil, "Fake news detection using machine learning," in *2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)*, 2020, IEEE, pp. 1-5, doi: 10.1109/iSSSC50941.2020.9358890.
[19] E. M. Mahir, S. Akhter, and M. R. Huq, "Detecting fake news using machine learning and deep learning algorithms," in 2019 *7th International Conference on Smart Computing and Communications (ICSCC)*, 2019, IEEE, pp. 1-5, doi: 10.1109/ICSCC.2019.8843612.
[20] A. Jain, A. Shakya, H. Khatter, and A. K. Gupta, "A smart system for fake news detection using machine learning," in 2019 *International conference on issues and challenges in intelligent computing techniques (ICICT)*, 2019, vol. 1, IEEE, pp. 1-4, doi: 10.1109/ICICT46931.2019.8977659.
[21] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake news detection using machine learning ensemble methods," *Complexity*, vol. 2020, pp. 1-11, 2020, doi: 10.1155/2020/8885861.
[22] Z. Khanam, B. Alwasel, H. Sirafi, and M. Rashid, "Fake news detection using machine learning approaches," in *IOP Conference Series: Materials Science and Engineering*, 2021, vol. 1099, no. 1: IOP Publishing, p. 012040, doi: 10.1088/1757-899X/1099/1/012040.
[23] F. Ahmad and R. Lokeshkumar, "A comparison of machine learning algorithms in fake news detection," *International Journal on Emerging Technologies*, vol. 10, no. 4, pp. 177-183, 2019.
[24] L. Waikhom and R. S. Goswami, "Fake news detection using machine learning," in *Proceedings of International Conference on Advancements in Computing and Management (ICACM)*, 2019, doi: 10.2139/ssrn.3462938.
[25] M.-A. Ouassil, B. Cherradi, S. Hamida, M. Errami, O. E. Gannour, and A. Raihani, "A fake news detection system based on combination of word embedded techniques and hybrid deep learning model," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, 2022, doi: 10.14569/IJACSA.2022.0131061.
[26] P. K. Verma, P. Agrawal, I. Amorim, and R. Prodan, "WELFake: word embedding over linguistic features for fake news detection," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 4, pp. 881-893, 2021, doi: 10.1109/TCSS.2021.3068519.
[27] M. K. Chung, "Introduction to logistic regression," *arXiv preprint arXiv:2008.13567*, 2020, doi: 10.48550/arXiv.2008.13567.
[28] F. Pedregosa *et al.*, "Scikit-learn: machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825-2830, 2011.

BIOGRAPHIES OF AUTHORS

Enas Ahmad Alikhashashneh is an assistant professor at the College of Information Technology and Computer Science, Yarmouk University, Jordan. She received a B.Sc. degree in computer science from the Al Balqa Applied University, Jordan, an M.Sc. degree in Computer and Information Systems from the Yarmouk University, Jordan, another M.Sc. degree in Science from the Purdue University/ IUPUI, USA, and a Ph.D. degree in Computer Science from Purdue University/ IUPUI, USA. Her research areas are machine learning, deep learning, NLP, and pattern recognition. She is Assistant Dean of the College of Information Technology and Computer Science at Yarmouk University from 2023 to now. She can be contacted at email: enas.a@yu.edu.jo.

Dr. Khalid M.O. Nahar is an associate professor in the Department of Computer Sciences-Faculty of IT, he received his BS and MS degrees in computer sciences from Yarmouk University in Jordan, in 1992 and 2005 respectively. He was awarded a full scholarship to continue his Ph.D. in Computer Sciences and Engineering from King Fahd University of Petroleum and Minerals (KFUPM), KSA. In 2013 he completed his PhD and started his job as an assistant professor at Tabuk University, KSA for 2 years. In 2015 he went back to Yarmouk University and became the assistant dean for quality control. From 2020 to 2022, he was assigned the chairman of the training department at the accreditation and quality assurance center. Research Interests include: artificial intelligence, speech recognition, arabic computing, natural language processing, multimedia computing, content-based retrieval, machine learning, IoT, and data science. He can be contacted at email: khalids@yu.edu.jo.

Mohammed Abual-Rub is an assistant professor at the Department of Management Information Systems, Business School, King Faisal University, KSA, where she has been a faculty member since 2011. Abual-Rub graduated with degree in computer Sciences from Yarmouk University, Jordan, in 1996, and an M.Sc. in Computer Science from USM University, Malaysia in 2006. He then completed His Ph.D. Computer Science from USM University, in 2011. His research interests are primarily in bioinformatics, machine learning, and artificial intelligence, where he is the author/co-author of over 10 research publications. He can be contacted at email: mabualrub@kfu.edu.sa.

Hedaya Mohammed Alkhaldy is studying a master's degree in the Department of Computer Information Systems at the College of Information Technology and Computer Science at Yarmouk University, Jordan. She obtained a bachelor's degree in the Department of Computer Information Systems at the College of Information Technology and Computer Science at Yarmouk University, Jordan. Her research focuses include machine learning, deep learning, and NLP. She can be contacted at email: 2022930012@ses.yu.edu.jo.