Applied Soft Computing 109 (2021) 107559

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

An effective fake news detection method using WOA-xgbTree algorithm and content-based features

Saeid Sheikhi

Department of Computer, Gorgan Branch, Islamic Azad University, Gorgan, Iran

ARTICLE INFO

Article history:
Received 7 October 2020
Received in revised form 23 May 2021
Accepted 28 May 2021
Available online 6 June 2021

Keywords:
Fake news
Fake news detection
Online fake news
Feature extraction
Machine learning

ABSTRACT

In recent years, with the fast development of the internet and online platforms such as social media feeds, news blogs, and online newspapers, deceptive reports have been universally spread online. This manipulated news is a matter of concern due to its potential role in shaping public opinion. Therefore, the fast spread of fake news creates an urgent need for automatic systems to detect deceitful articles. This motivates many researchers to introduce solutions for the automatic classification of news items. This paper proposed a novel system to detect fake news articles based on content-based features and the WOA-Xgbtree algorithm. The proposed system can be applied in different scenarios to classify news articles. The proposed approach consists of two main stages: first, the useful features are extracted and analyzed, and then an Extreme Gradient Boosting Tree (xgbTree) algorithm optimized by the Whale Optimization Algorithm (WOA) to classify news articles using extracted features. In our experiments, we considered the bases of the investigation on classification accuracy and the F1-measure. Then, we compared the optimized model with several benchmark classification algorithms based on a dataset that compiled over 40,000 various news articles recently. The results indicate that the proposed approach achieved good classification accuracy and F1 measure rate and successfully classified over 91 percent of articles.

1. Introduction

In recent years, fake news detection methods have attracted more attention since the circulation of online misinformation increases, and this is becoming a central concern of modern society [1]. In general, the idea of fake news is not a new issue. This concept has been around before the development of the internet. Many publishers use misleading information to further their interests [2]. Plenty of these publishers are spreading fake news through both traditional print news media and online platforms. Online platforms have a significant role in spreading fake news in society; these platforms, such as social media and online newspapers, provide access to a variety of publications in one sitting to users, which is more convenient and faster than traditional newspapers. Moreover, the nature of social networks provides a suitable platform for the rapid spreading of information in real-time, even with the reliability of this information, it has created severe problems in information authenticity [3].

Fake news not only negatively affects individuals but, over time, damages society as a whole. For instance, in the US 2016 presidential election, fake news had been widely spread on Facebook instead of the more reliable and popular mainstream news [4]. This shows that users may pay more attention to manipulated news than genuine ones. Social media users who engage in information manipulation could have various motives for the spread of this information over the internet, such as influence, political agendas, and manipulation, but while many of these users are real, those who are malicious and spread manipulated news may or may not be real users [5]. Since creating social media accounts are uncomplicated and inexpensive, many people design social accounts for malicious activities. If a computer algorithm controls a social media account, it is mentioned as a social bot [4]. These social bots can interact with people through social media and automatically generate content and spread it over the internet, which makes it very difficult for ordinary people to identify these manipulated contents [6].

Therefore, it is challenging to verify online content using manual ways since a massive volume of online information is generated and distributed over the internet in recent years. Moreover, many researchers have agreed on the necessity of automatic, computational fake news detection systems [7]. In general, fake news detection systems have been classified into two categories of "social context" and "news content" based on their data sources. The social context methods focus on social features and signals, features like interaction and engagement of users in social media with given news (retweeting it on Twitter or share or liking it on Facebook). These systems are referred to as "social-based"

E-mail address: S.Sheikhi@outlook.com.

https://doi.org/10.1016/j.asoc.2021.107559
1568-4946/

methods [8]. The second category is "news content" methods, and these methods try to verify the content of news and use features such as title, body text, and some additional metadata to detect fake news. These systems are referred to as "content-based" methods [8].

This paper is proposing a novel fake news detection model using content-based features that can automatically classify news headlines into fake and true ones to resolve the challenges mentioned earlier. In this model, first, the most useful content-based features were extracted. Then the hybrid proposed WOA-xgbTree algorithm has been used to implement the model to classify news articles with high detection performance. In this hybrid model, considering the WOA algorithm's strengths, we have optimized Extreme Gradient Boosting Tree (xgbTree) algorithm parameters using a WOA method, which has made an effective method that can classify news articles with high classification accuracy. The proposed method has the ability to work effectively offline with extracted features and only news content without other information.

The remaining paper is organized as follows. We review several recently introduced methods for fake news detection and discuss their methods and advantages in Section 2. Then, we explain the used dataset, extracted features method, WOA, and xgbTree algorithms as the methodology in Section 3. In Section 4, we introduce the WOA- xgbTree model for detecting fake news articles and the effectiveness of the model in Section 5. Finally, we discuss and conclude this paper in Section 6.

2. Related works

The purpose of news verification is to employ technology to identify manipulated unreliable online news content. It is an essential concern within specific streams of library and information science [9]. Therefore, numerous researchers try to solve fake news problems in diverse fields, especially online news. This section will review the different approaches used to detect fake news on online platforms, and we will briefly discuss their advantages and their result.

Castillo et al. (2011) [10] introduced a new system for identifying authentic and fake news stories. They have used a comprehensive set of linguistic features such as emoticon symbols, special characters, hashtags, and positive/negative words in order to classify news stories. Jin et al. (2013) [11] presented a method in order to identify information cascades in the Twitter online platform. In their approach, models were applied to analyze the patterns, which allowed them to distinguish between authentic and fake news. Wu et al. (2015) [12] introduced a graph kernel-based SVM classifier that learns high-order distribution patterns for fake news detection. Ma et al. (2016) [13] presented a new system that utilizes a recurrent neural network (RNN) for detecting fake news articles using linguistic features reached from a series of user comments.

Sampson et al. (2016) [14] proposed a new approach to detect trustworthy news articles. They used inevitable linkages among conversation fragments to identify reliable news stories. In another approach, Yang et al. (2012) [15] investigated the same set of user characteristics on the Sina Weibo social platform, the most famous Chinese microblogging website. Reganti et al. (2016) [16] presented a system to detect sarcasm tweets and product reviews automatically. They have utilized general features based on lexicon and baseline features. The features such as word n-grams, character n-grams, and word skip-grams are extracted and combined with lexicon features. Then, they classified these features using various algorithms, such as logistic regression, ensemble classifiers, random forest (RF), and decision tree.

Table 1
Description of the dataset.

News type	Records	Percentage
True news	21,417	47.70%
Fake news	23,481	52.30%
Total articles	44,898	100%

Buschmeier et al. (2014) [17] studied the importance of different features in the detection and classification of ironic and sarcastic reviews on various types of products. Firstly, they extracted features based on Bag-of-Words and lexicon-based features. Then, they utilized these extracted features on different classifiers such as SVM, Logistic Regression, Random Forest (RF), and Decision Tree. Kwon et al. (2017) [18] proposed a new method to detect the authenticity of the news items. They used a combination of the various set of features such as linguistic, User, temporal, and structural features in order to classify fake news.

Therefore, with the increased use of online platforms, it has become easy for all users to access information. According to the literature, under the broad and rapid deployment of information and the need for techniques to ensure such information's authenticity, many studies were conducted to develop methods to solve the problem of fake news. Many of these researchers still did not achieve enough performance in identifying news content, and this problem motivated us to propose a new method to automatically identify news articles using the useful linguistic and content-based features and the proposed hybrid WOA-xgbTree algorithm that can classify different news articles with high performance.

3. Methodology

Our goal is to classify news articles as authentic or fake; in this section, we first describe the dataset we used for our tests; then, we discuss the feature extraction process and describe used features for classification purposes. Finally, we present the hybrid model implemented for identifying fake news articles on the dataset.

3.1. Data collection

Online news items can be collected from various sources, such as social media, news agencies websites, and search engines. However, it is difficult and complicated to determine authentic online news items manually. This task normally needs people with satisfying expertise in the domain, making a careful investigation of claims using news context, additional evidence, and reports from reliable sources. Usually, online news data with classification could be collected in different ways, such as Fact-checking websites, Expert journalists, and Industry detectors.

Therefore, in this research, our experiments are conducted on the ISOT Fake News dataset, which contains over 44,000 news articles classified as legitimate and fake, and is publicly available in [19]. The used dataset is collected news from different sources that were cleaned and processed to be more reliable. Fake news articles were collected from unreliable websites flagged by Politifact and Wikipedia, while true news was collected using the Reuters website. The description of the used dataset is presented in Table 1.

3.2. Feature extraction and selection

Feature Extraction is one of the essential processes to decrease the number of required resources to represent a dataset. One of the significant problems which researchers are facing in analyzing complex data is the number of variables involved. Analysis with a large volume of variables usually needs high computational power and process time. Moreover, it could negatively affect classification performance and make the classification algorithm overfit the training samples and predict the class of new samples inaccurately. Hence, feature extraction has an essential role for the success of classification algorithms with a large number of variables.

Therefore, in order to have the best performance in classifying news articles, we have used a comprehensive set of features. Utilizing various features will help us to compare the various efficiencies and F-measures obtained for different features, as well as, a combination set of features. The features we have used for our study are summarized as follows:

■ **Presence of special symbols**- Presence of some special symbols in the text. Symbols such as "-" and ",.

■ **Start with articles**- We consider the Start with articles in the headline and text as a feature since it can use for detection of fake news. Articles such as "the", "a" and "an".

■ **Start with keywords**- Typically, the use of some suspicious keywords at the start of headlines to gain users' attention to the news can be used to detect fake news. Keywords such as "report:", "breaking:", "study:".

■ **Start with number**- Checks if the news headline or body starts with a number or not.

■ **Presence of special keywords**- The presence of some of the special keywords usually most formal news agencies avoid to use on their headlines and body can classify fake news articles effectively. Keywords such as "unsure", "coworkers", "sort of", etc.

■ **Presence of offensive swear words** - We consider the presence of these words in news as a feature to classify news headlines since formal news agencies usually avoid using these words in their headlines.

■ **Presence of abbreviations**-Checks if the news headline or body contains any abbreviations or not.

■ **Presence of ellipsis**- An ellipsis is a series of dots that normally show an intentional omission of a word. This feature checks if the news contains Ellipsis or not.

■ **Presence of uppercased words**- We consider the presence of uppercased words as a feature to identify fake news since many spammers use Uppercased words to attract users in the headline or body of news.

■ **Content length** - this feature contains the total length of news content including space, special characters, symbols, etc. The average authentic news headline used a mere 34 characters.

■ **Number of words**- This feature consists of the total number of words of a news headline and body. The ideal length of a news headline is six words and for the body is over 1000 words.

All of the presented features are extracted independently from both the headline and body content of news articles. Hence, each of the produced dataset records contains 22 extracted features. The 11 of them belong to headlines, and the other 11 are the body content of news articles. Table 2 shows how each feature value is extracted from legitimate and fake news articles based on headline content, while a similar process is performed for each news article's body content.

3.3. The Whale Optimization algorithm (WOA)

The WOA algorithm is a nature-inspired optimization algorithm developed by Mirjalili et al. [20] in 2016. This algorithm is inspired by the humpback whale's feeding behavior, which is considered the largest animal on earth. This algorithm is classified as a metaheuristic algorithm. The main idea is that it produces better search results through a unique hunting behavior of whales named "bubble mesh feeding". This search method is based on a loop path or involute-shape path in the search space. The WOA algorithm search behavior structure for optimal searching can be described as follows:

■ Encircling prey:

In WOA, the exploration method begins by producing candidate solutions, and each humpback whale describes a possible solution. The current best candidate is considered the most optimal or close to the searching space's optimum solution. After the best available solution is found, all other individuals update their position towards the agent with an optimal solution. The mathematical model for this behavior described as follows:

$$X_{i+1} = X_i - A|CX_i^* - X_i| \tag{1}$$

Where i represents the current number of iteration, The X_i is a vector that gives the whale position, and the X_i^* indicates the best search candidate position, and the A and C are the coefficient vectors. The values of A and C vectors are computed as follows:

$$C = 2at_1 - a \tag{2}$$

$$C = 2t_2 \tag{3}$$

Where, a value is linearly decreasing between 2 to 0 during iterations, and the t_1 and t_2 is a random value between the 0 to 1.

■ Bubble-net hunting method:

In bubble-net preying humpback whale hunting, they move towards the prey in a shrink wrapping and spiral uprising. In WOA, the probability of selection within the two methods is 0.5. The local optimization solution can be produced using these two methods; the mechanism of shrinking encircling is performed by decreasing the value of a in Eq. (1)–(2). This stage is mathematically represented as follows:

$$X_{i+1} = f(x) = \begin{cases} X_i - A|CX_i^* - X_i|, & p < 0.5 \\ \left|CX_i^* - X_i\right| e^{bi}\cos(2\pi l)X_i^*, & p \geq 0.5 \end{cases} \tag{4}$$

Where b represents a constant value that defines the shape of the spiral, t indicates a random number from -1 to 1 while p shows a random number between the 0 and 1.

■ Searching for prey:

The WOA algorithm forced search agents away from the base whale by changing the A value in the exploitation phase. In this stage, the WOA updates individuals' positions to a random whale rather than the whale with the optimal solution to implement a global search. This process mathematical described as follows:

$$X_{i+1} = X_{rand} - A|CX_{rand} - X_i| \tag{5}$$

Where X_{rand} represents the randomly selected agent positions.

3.4. Extreme gradient boosting (XGBoost)

The XGBoost algorithm is an optimized gradient tree boosting ensemble method that generates decision trees sequentially. This algorithm developed by Chen et al. (2016) [21], is a scalable implementation of the gradient boosting framework that was initially proposed by Friedman (2001 and 2002); it is a useful algorithm for solving supervised learning problems such as classifications and regression [22,23]. Therefore, as signified by the name, the base idea of the XGBoost algorithm is boosting. This algorithm combines various weak learners to build a robust learner using additive training strategies and integrating techniques [24]. The boosting reduces overall errors and improves the classifier's classification performance by adding supplementary models based on previous iterations that make this method a useful technique for classification [25]. Furthermore, in XGBoost,

Table 2
Example of features values for legitimate and fake news contents.

Feature name	Headline text	
	"American richard thaler wins nobel economics prize"	"100% of teenagers huge f*****g a******s,' confirms study by sobbing, red-faced scientists"
Special_symbols	No	Yes
Start_by_articles	No	No
Start_by_keywords	No	No
Start_by_number	No	Yes
Special_keywords	No	No
Offensive_swear_words	No	Yes
Abbreviations	No	No
Ellipsis	No	No
Uppercased_words	No	No
Content length	50	90
Number of words	7	12

the objective function typically contains two sections, training loss and regularization. This function is defined as follows:

$$Obj\,(\theta) = L\,(\theta) + \Omega(\theta) \tag{6}$$

Where L represents the value of the training loss function, while Ω denotes the regularization term. The model's performance on training data was measured by using the training loss. The regularization term controls the model's complexity and overfitting problems [26]. There are many ways to express the complexity; the measurement of complexity for each tree usually is formulated as follows:

$$\Omega\,(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2, \tag{7}$$

Where T denotes the total number of leaves on the tree, and ω is a vector value representing scores on leaves. The XGBoost structure score is an objective function that is described as follows:

$$Obj = \sum_{j=1}^{T} \left[G_j \omega_j + \frac{1}{2}(H_j + \lambda)\omega_j^2 \right] + \gamma T, \tag{8}$$

In which the ω_j terms are separate and are independent of other ones. The formulate $G_j\omega_j + \frac{1}{2}(H_j + \lambda)\omega_j^2$ is also quadratic, the best ω_j for a given structure *q(x)*.

4. Proposed WOA-xgbTree model

Detection of fake news is a difficult task as it is intentionally written to falsify information. This research proposes a hybrid model that uses linguistic and content features to identify fake news. In the proposed system, fake news detection is formulated as a binary classification problem since our system categorizes news items into the two classes of Fake and Legitimate. Moreover, to achieve more accurate results in this model, we optimized its parameters using a WOA algorithm. The parameters were considered and optimized, including the number of boosting iterations (nrounds), the max depth (max_depth), step size shrinkage (eta), Minimum Loss Reduction (gamma), Sub-sample Ratio of Columns (colsample_bytree), Minimum Sum of Instance Weight (min_child_weight) and Subsample Percentage (subsample). The information of each parameter is shown in Table 3.

The final architecture of the proposed model is shown in Fig. 1.

As shown in Fig. 1, the proposed system contains a set of processes. The different steps involved in the proposed system are: Firstly, the raw news body and headlines contents are preprocessed using several preprocessing methods; then, features are extracted and represented using the content of headlines and body of the news. Further, the extracted features are used to build a final dataset for classification. Finally, after implementing the model using the developed dataset, the proposed model's performance has been investigated in classification accuracy.

In order to make the final WOA-xgbTree model ready for classification the following steps were followed: first, an initial xgbTree model was developed; then, the WOA algorithm optimized the xgbTree parameters. In the optimization section, every whale makes an exploration to find the optimal values of model parameters. For each position, the WOA algorithm measures each searching agent's fitness based on a fitness function, and then it updates the positions of other searching agents. A corresponding accuracy value was computed for each of the parameters, and the best fit model corresponds to the highest accuracy.

5. Experiments and result analysis

This section describes the experiment's details for evaluating the proposed fake news detection system, presenting the performance metrics, and discusses the classification results.

5.1. Experiment setup

Fake news detection is a complex process, and a reasonable system needs several aspects to detect manipulated news effectively. This is the main reason we used a combination of the xgbTree boosting method and the WOA algorithm in the proposed system. Moreover, unlike other introduced solutions, our proposed system used only 22 features to detect fake news articles that can detect most of the fake news in short process time with low complexity and processing cost. Therefore, to better evaluate the model performance, we have performed two sets of experiments on the produced dataset. First, we studied the proposed model performance and compared it with other classification algorithms such as Random Tree, Naive Bayes, K-nearest Neighbors (KNN), Support Vector Machines (SVM), Multilayer Perceptron (MLP), and Radial basis function (RBF) neural network on the same dataset. In the second part of the experiments, we performed a comparative study between the WOA algorithm and several metaheuristics algorithms to optimize xgbTree parameters. The algorithms applied in this study are Particle Swarm Optimization [27], Grey Wolf Optimization (GWO) [28], Sine Cosine Algorithm (SCA) [29], and Artificial Bee Colony (ABC) [30]. The proposed model and considered algorithms were implemented on a system with 16 GB RAM using the R and Caret package. The default parameter settings are applied in all of the metaheuristic algorithms used in this research, with the same value for population size and number iterations parameters. Accordingly, the number of population size was set to 10, and the number of iterations is set to 100. Furthermore, in this research, after extracting content-based features and constructing a new dataset, we split the dataset into training and testing sets. Therefore, 80 percent

Table 3
Information about parameters of the xgbTree model.

Parameters	Default value	Range	Explanation
nrounds	100	[10, ∞]	Number of boosting iterations
max_depth	6	[0, ∞]	Maximum depth of a tree
eta	0.3	[0,1]	learning rate
gamma	0	[0, ∞]	Minimum loss reduction
colsample_bytree	1	(0,1]	Subsample ratio of columns
min_child_weight	1	[0, ∞]	Minimum sum of instance weight
subsample	1	(0,1]	Subsample ratio of training

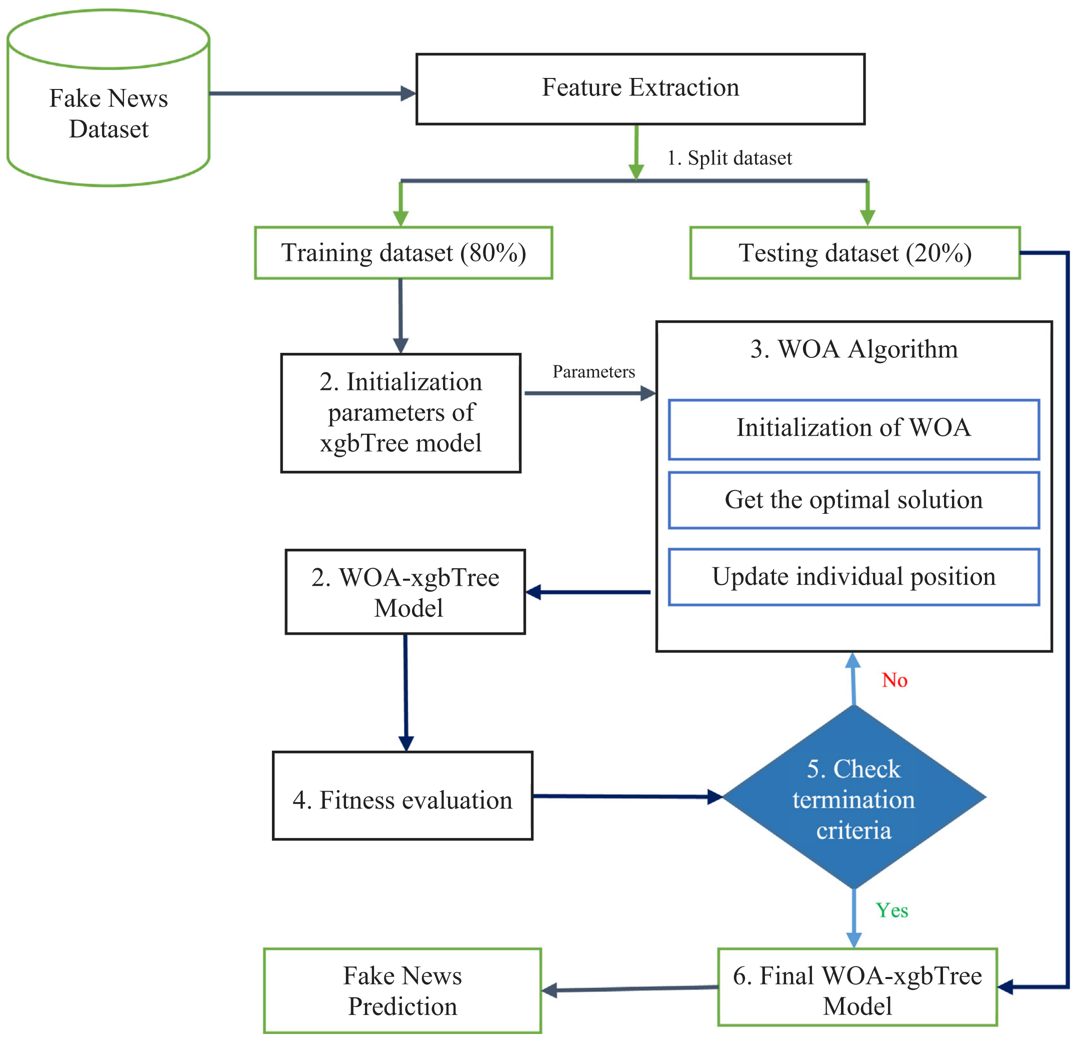

Fig. 1. Scheme of the development of the WOA and WOA-xgbTree model.

of the data was used for training, while 20 percent was used for testing the model. Furthermore, to reduce the model's error, we utilized a well-known 10-fold cross-validation technique for training and testing purposes.

5.2. Evaluation metrics

Many researchers have developed systems for classifying news items. They measured their proposed systems' performance by a variety of metrics, such as accuracy, recall, precision, and F-Measure. Accuracy is a measure that calculates the percentage of news items predicted in the correct classes during a classification process. Though, in the case of unbalanced classes, the accuracy metric could produce false results. In these cases, F-measure is a better metric for classification. F-measure is a measure that combines precision and recall using the measurement of the harmonic mean of recall and precision. Furthermore, the confusion matrix illustrates as follows (see Table 4):

Where,

True Positive Rate (TP): denotes the percentage of Fake news items that were correctly classified using the proposed model.

True Negative Rate (TN): describes the percentage of legitimate news items that were correctly classified using the proposed model.

False Positive Rate (FP): represents the percentage of legitimate news items that were incorrectly classified as Fake news by the proposed model.

False Negative Rate (FN): describes the percentage of Fake news items that were incorrectly classified as legitimate news by the proposed model.

The classification accuracy estimates the closeness among forecasted records and the total number of legitimate and fake

Table 4
The confusion matrix.

Actual	Predicted	
	Fake	Legitimate
Fake	TP	FP
Legitimate	FN	TN

Table 5
Comparisons between the proposed method and several classification techniques.

Algorithm	Accuracy	Kappa	Precision	Recall	F1
Random tree	89.26	93.30	0.946	0.946	0.946
KNN	89.07	78.12	0.891	0.891	0.891
SVM	77.95	55.44	0.789	0.780	0.776
Naive Bayes	70.17	41.56	0.777	0.702	0.685
MLP	86.72	73.6	0.879	0.867	0.867
RBF	84.91	69.72	0.849	0.849	0.849
WOA-xgbTree	91.86	83.70	0.923	0.920	0.921

news articles. The value of accuracy is measured by Eq. (9).

$$AC = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{9}$$

The F-measure measured by the following formula:

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{10}$$

Where,

Precision: denotes the percentage of news items that were classified correctly as Fake by the proposed model. The value of Precision is calculated using Eq. (11).

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

Recall: represents the percentage of Fake news articles that were correctly classified as Fake news. It presents the completeness. It was measured using Eq. (12).

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

Therefore, many past researchers have used the F-Measure metric for evaluation of their systems since in classification literature as an overall assessment of a classifier's performance as it considers both precision and recall [21]. In evaluating the proposed method Performance, we considered Accuracy and F-measure as base evaluation metrics that can demonstrate the algorithm's capability to detect fake news correctly.

5.3. Result analysis

In the first set of experiments, we compared the proposed model's experimental results with some well-known classification algorithms on the dataset. For a fair comparison, the proposed model and considered algorithms are implemented on a system with similar characteristics and tested on the same dataset. The model's best training performance was achieved with the following parameters: eta is 0.4, max_depth is 3.0, gamma is 0, colsample_bytree is 0.8, min_child_weight is 1.0, subsample is 0.75 and nrounds is 150. Then the optimized model is applied to the testing set. Table 5 exhibits the proposed approach's results and other well-known classification techniques on produced fake news dataset.

As presented in Table 5, the classifiers are compared and evaluated based on overall accuracy, Precision, Recall, and F-Measure. According to the represented results, the proposed approach overall performance was further compared to other classification techniques on the extracted features. It successfully classified 91.86% of news articles, respectively. After the proposed model, the Random Tree, KNN, and MLP obtained the highest performance, and they successfully classified 89.26%, 89.07%, and 86.72% of news articles, but there is over 2 percent gap between them and the proposed model. As observed in Table 5, the Naive Bayes and SVM algorithms produced the worst results in classifying different news articles. Furthermore, the majority of considered classification algorithms successfully identified the class of news articles, which indicates the extracted features perform satisfactorily and significantly impact the detection of the class of news in short process time.

In the second set of experiments, we analyzed and compared the proposed model results using the xgbTree algorithm and some other metaheuristics algorithms to optimize xgbTree parameters for classification. Fig. 2 shows a comparison between the model results and other metaheuristics algorithms optimized xgbTree algorithm.

Fig. 2 shows that the optimized xgbTree algorithm using the WOA algorithm on extracted features produced the best performance compared with the other metaheuristic methods. The proposed model has produced overall accuracy and F-measure of 0.9186 and 0.9219, respectively. After the WOA, the artificial bee colony (ABC) performed better than other benchmark algorithms with overall accuracy and F-measure of 0.9086 and 0.9128. The PSO algorithm also performed almost similar results with the ABC and obtained accuracy and F-measure of 0.9061 and 0.9101. The GWO algorithms achieved good performance with the identification of over 87% of news articles correctly. The SCA algorithm achieved the worst results and classified 83.56 of articles. Overall, the comparative study outcomes demonstrate that the WOA-xgbTree achieved higher accuracy and F-measure than considered metaheuristic algorithms, and was very successful in distinguishing different news articles. In order to show the model's effectiveness, the result of the receiver operating characteristic (ROC) curve is presented in Fig. 3.

The ROC curve is a helpful tool for the model's visualization performance, deciding whether a Model is suitable regarding cost sensitivity. In the represented (ROC) curve, the curve's x-axis denotes the false positive rate while the y-axis describes the false negative. The area found under the curve (AUC) also represents the model's separability degree. The AUC-Roc curve proposed model is presented in Fig. 3, with the value of (0.9767). The AUC shows the model's performance in separating different classes and marking the model's separability degree. The AUC near the 1 shows model successfully distinguished different classes. The presented AUC (0.9767) also shows that the current model is a suitable model that has a strong capability to recognize different news article classes. In this study, we extracted effective features for detecting fake news articles, and it is necessary to present the importance of the model's input features. In the next test, we marked the importance of each extracted feature used as the model input. Fig. 4 summarizes each feature's importance on the dataset ranked by the importance estimated from the trained model.

Fig. 4 explains the contributions of each feature for the prediction performance of the model. The features started with the letter "T" are extracted features from news headlines, while features that begin with "B" belong to the news articles' body. We can observe in Fig. 4 the "T_Number_of_Words", "T_Content_Length", and "B_Abbreviations" are the top three with the highest contributions. The existing two headline features between the top three features indicate that the news headline has a significant role in detecting fake news articles. It also shows the length and number of words used in the headline could positively impact the detection of class news articles.

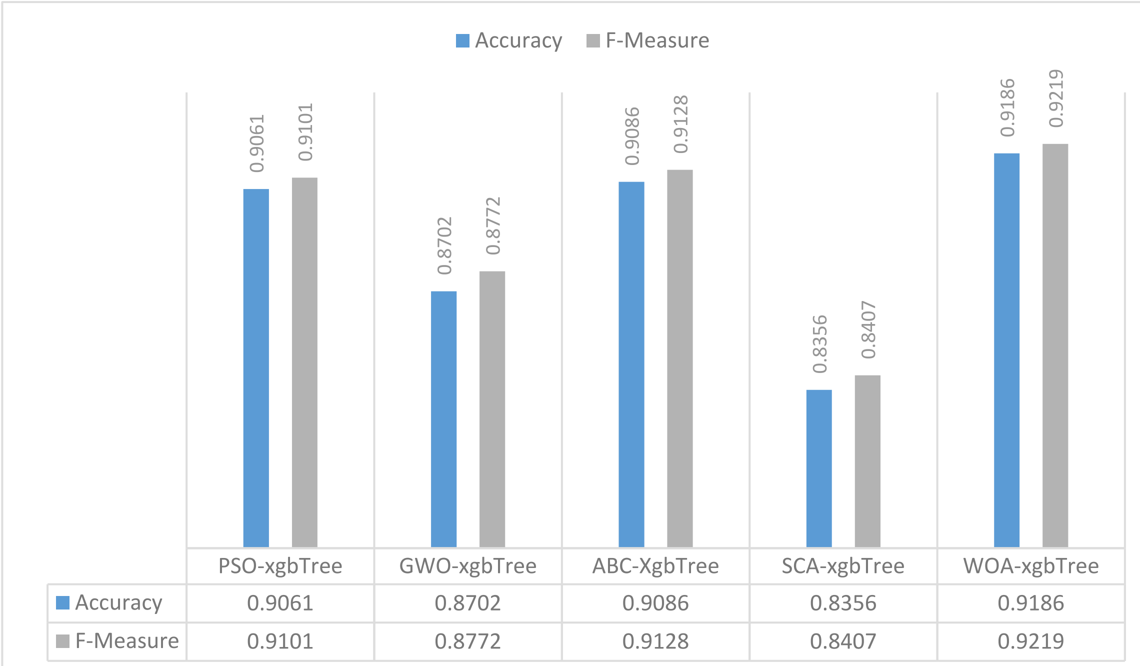

Fig. 2. Comparison of the proposed model with other optimization algorithms.

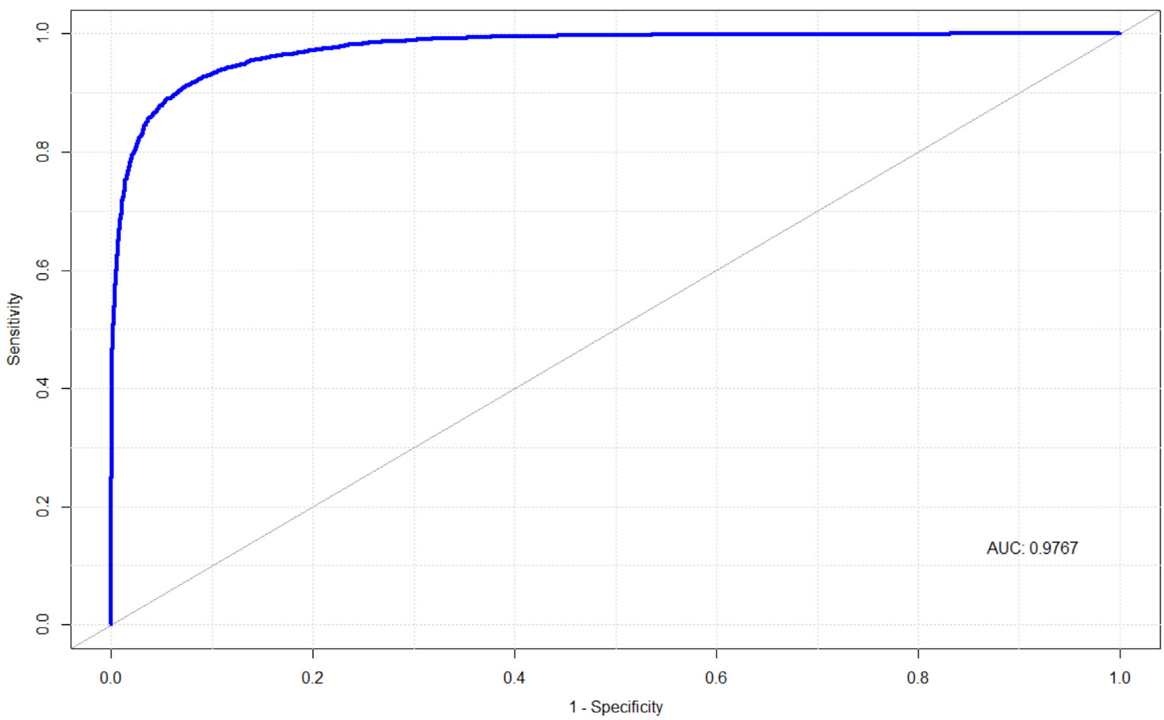

Fig. 3. The ROC curve of the proposed model.

6. Conclusions

Today, fake news is one of the most significant problems producing substantial negative impacts on individual users and society, making fake news detection a big challenge currently. In this research, we first introduced the fundamental concepts of fake news in both traditional and online platforms. Then, we reviewed some of the current fake news detection methods and discussed their weakness and advantages. Finally, we developed a new, accurate model to identify fake news items based on the xgbTree method and WOA algorithm. The proposed model's achieved result has been investigated and compared with well-known classifications and metaheuristics algorithms with regard to classification accuracy and f1 measures toward the same dataset. Furthermore, one of our research focuses was on extracting features from news articles to help the fake news detection system have a higher classification accuracy and shorter process time. The model achieved results that demonstrated extracted features made a positive impact on the performance of the proposed fake news detection system. This positive impact

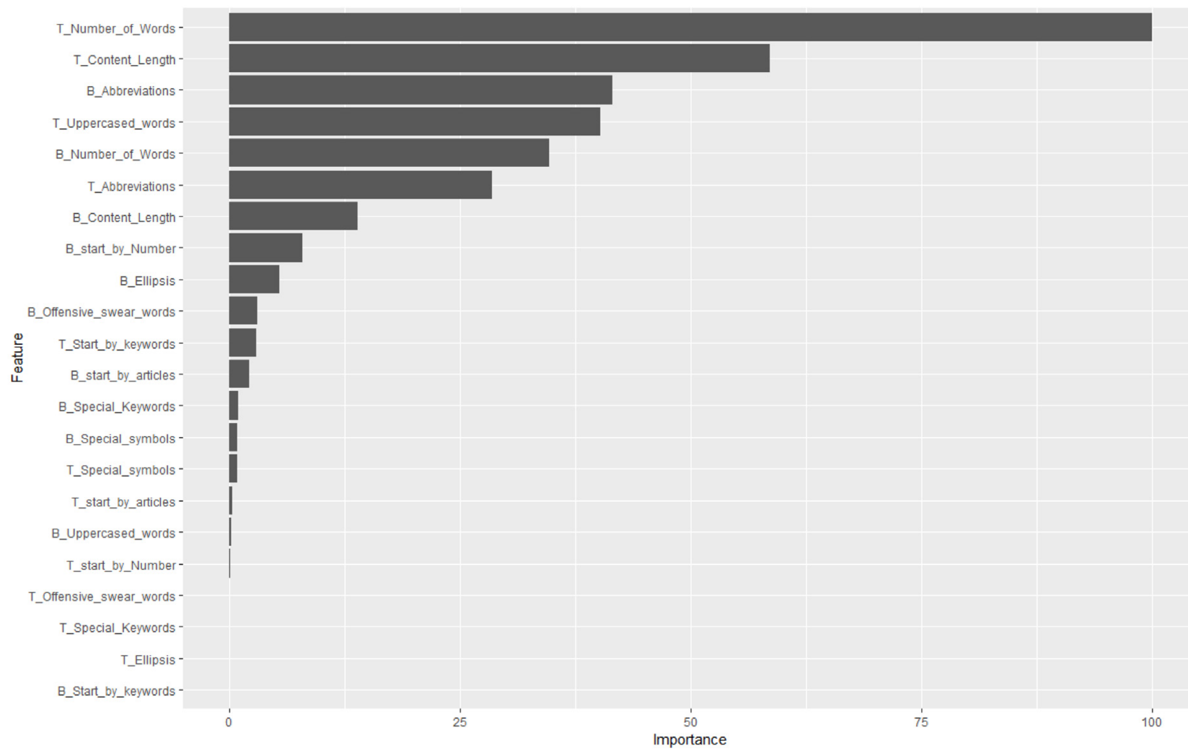

Fig. 4. Rank of Features contributions in the proposed model result.

improved the classification accuracy of the system and reduced its process time.

CRediT authorship contribution statement

Saeid Sheikhi: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Visualization, Investigation, Writing - review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Bessi, M. Coletto, G.A. Davidescu, A. Scala, G. Caldarelli, W. Quattrociocchi, Science vs conspiracy: Collective narratives in the age of misinformation, PLoS One 10 (2) (2015) e0118093.
[2] G. Gravanis, A. Vakali, K. Diamantaras, P. Karadais, Behind the cues: A benchmarking study for fake news detection, Expert Syst. Appl. 128 (2019) 201–213.
[3] S.M. Alzanin, A.M. Azmi, Detecting rumors in social media: A survey, Procedia Comput. Sci. 142 (2018) 294–300.
[4] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news detection on social media: A data mining perspective, ACM SIGKDD Explor. Newsl. 19 (1) (2017) 22–36.
[5] A. Bondielli, F. Marcelloni, A survey on fake news and rumour detection techniques, Inform. Sci. 497 (2019) 38–55.
[6] E. Ferrara, O. Varol, C. Davis, F. Menczer, A. Flammini, The rise of social bots, Commun. ACM 59 (7) (2016) 96–104.
[7] X. Zhang, A.A. Ghorbani, An overview of online fake news: Characterization, detection, and discussion, Inf. Process. Manage. 57 (2) (2020) 102025.
[8] M.L. Della Vedova, E. Tacchini, S. Moret, G. Ballarin, M. DiPierro, L. de Alfaro, Automatic online fake news detection combining content and social signals, in: 2018 22nd Conference of Open Innovations Association, FRUCT, IEEE, 2018, pp. 272–279.
[9] D.K. Vishwakarma, D. Varshney, A. Yadav, Detection and veracity analysis of fake news via scrapping and authenticating the web search, Cogn. Syst. Res. 58 (2019) 217–229.
[10] C. Castillo, M. Mendoza, B. Poblete, Information credibility on twitter, in: Proceedings of the 20th International Conference on World Wide Web, March, 2011, pp. 675–684.
[11] F. Jin, E. Dougherty, P. Saraf, Y. Cao, N. Ramakrishnan, Epidemiological modeling of news and rumors on twitter, in: Proceedings of the 7th Workshop on Social Network Mining and Analysis, August, 2013, pp. 1–9.
[12] K. Wu, S. Yang, K.Q. Zhu, False rumors detection on sina weibo by propagation structures, in: 2015 IEEE 31st International Conference on Data Engineering, April, IEEE, 2015, pp. 651–662.
[13] J. Ma, W. Gao, P. Mitra, S. Kwon, B.J. Jansen, K.F. Wong, M. Cha, Detecting rumors from microblogs with recurrent neural networks, 2016.
[14] J. Sampson, F. Morstatter, L. Wu, H. Liu, Leveraging the implicit structure within social media for emergent rumor detection, in: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, October, 2016, pp. 2377–2382.
[15] F. Yang, Y. Liu, X. Yu, M. Yang, Automatic detection of rumor on sina weibo, in: Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, August, 2012, pp. 1–7.
[16] A.N. Reganti, T. Maheshwari, U. Kumar, A. Das, R. Bajpai, Modeling satire in english text for automatic detection, in: 2016 IEEE 16th International Conference on Data Mining Workshops, ICDMW, December, IEEE, 2016, pp. 970–977.
[17] K. Buschmeier, P. Cimiano, R. Klinger, An impact analysis of features in a classification approach to irony detection in product reviews, in: Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, June, 2014, pp. 42–49.
[18] S. Kwon, M. Cha, K. Jung, Rumor detection over varying time windows, PLoS One 12 (1) (2017) e0168344.
[19] ISOT Fake News Dataset https://www.uvic.ca/engineering/ece/isot/datasets/fake-news.
[20] S. Mirjalili, A. Lewis, The whale optimization algorithm, Adv. Eng. Softw. 95 (2016) 51–67.
[21] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August, 2016, pp. 785–794.
[22] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Statist. (2001) 1189–1232.
[23] J.H. Friedman, Stochastic gradient boosting, Comput. Statist. Data Anal. 38 (4) (2002) 367–378.
[24] M. Adnane, M. El, S. El Fkihi, R.O.H. Thami, Prediction demand for classified ads using machine learning: an experiment study, in: Proceedings of the 2nd International Conference on Networking, Information Systems & Security, March, 2019, pp. 1–6.

[25] P. Carmona, F. Climent, A. Momparler, Predicting failure in the US banking sector: An extreme gradient boosting approach, Int. Rev. Econ. Finance 61 (2019) 304–323.
[26] W. Gao, W. Wang, D. Dimitrov, Y. Wang, Nano properties analysis via fourth multiplicative ABC indicator calculating, Arab. J. Chem. 11 (6) (2018) 793–801.
[27] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of ICNN'95-International Conference on Neural Networks, vol. 4, November, IEEE, 1995, pp. 1942–1948.
[28] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, Adv. Eng. Softw. 69 (2014) 46–61.
[29] S. Mirjalili, SCA: a sine cosine algorithm for solving optimization problems, Knowl.-Based Syst. 96 (2016) 120–133.
[30] B. Akay, D. Karaboga, Solving integer programming problems by using artificial bee colony algorithm, in: Congress of the Italian Association for Artificial Intelligence, December, Springer, Berlin, Heidelberg, 2009, pp. 355–364.