Natural Language Processing Approach for Fake News Detection Using Metaheuristics Optimized Extreme Gradient Boosting

Aleksandar Petrovic
Faculty of Informatics and Computing
Singidunum University
Belgrade, Serbia
0000-0003-3324-3909

Jasmina Perisic
Faculty of Informatics and Computer science
Union - Nikola Tesla University
Belgrade, Serbia
0000-0001-7412-7870

Luka Jovanovic
Faculty of Informatics and Computing
Singidunum University
Belgrade, Serbia
0000-0001-9402-7391

Miodrag Zivkovic
Faculty of Informatics and Computing
Singidunum University
Belgrade, Serbia
0000-0002-4351-068X

Milos Antonijevic
Faculty of Informatics and Computing
Singidunum University
Belgrade, Serbia
0000-0002-5511-2531

Nebojsa Bacanin
Faculty of Informatics and Computing
Singidunum University
Belgrade, Serbia
0000-0002-2062-924X

***Abstract*—The problem of fake news can be dangerous, as widely spread misinformation can cause serious physical and psychological harm. In the age of digitization and artificial intelligence (AI), particularly deep learning methods, which are able to generated synthetic text, being able to distinguish between real and fake news has become even greater challenge. Therefore, this research proposes a metaheuristics optimized machine learning (ML) approach with natural language processing (NLP) application for detecting fake news. The main classification model that was used is the eXtreme gradient boosting (XGBoost), which was tuned by the modified variable neighborhood search (VNS) algorithm. The NLP technique that was applied to transform text into the meaningful form for ML is the term frequency-inverse document frequency (TF-IDF). The performance of the VNS-tuned XGBoost was evaluated and compared with other cutting edge metaheuristics, where the proposed method obtained superior performance.**

***Index Terms*—XGBoost, TF-IDF, VNS, fake news detection, optimization.**

I. Introduction

Fake news is a term related to misinformation spread from illegitimate sources. Such information is usually with harmful intent, but not necessarily. The psychology of humans incites humourous behavior and cyberspace is the perfect place to perform practical jokes. Users in cyberspace do not have to reveal their identities and can easily find new groups to socialize with if something goes terribly wrong in terms of social behaviors. This creates a lack of responsibility which stimulates malicious behavior. If the problem had not escalated from the level of practical jokes this issue would have much lesser importance. Unfortunately, fake news are used to manipulate the masses, forward agendas, and even perform scams. The damage that can be done can even have the cost of human lives.

Elon Musk's acquisition of the social network formerly known as Twitter led to the changes in the use of the blue check mark which was used to verify the identity of the users. The check mark was mostly used by celebrities to differentiate them more easily from the fake and fan accounts. The problem appeared when the new owner of the social network made a decision to offer the blue check mark as a monthly subscription. The attempt to capitalize on identification had consequences. An example of this scenario is the Eli Lilly and Company. A malicious user created a copy of the company's page, obtained the blue check mark, and then shared the information that insulin would be free from now on. As a result, the stock of the company plummeted. Another example happened in New Hampshire, USA where AI was used to recreate the voice of one of the presidential candidates. Namely, the candidate would call voters and spread misinformation that their votes would not contribute during the general elections if they voted in primary elections. In this case, the attackers impersonated the victim with the goal of causing harm to that person.

Generative models are the latest trend in AI. The rapid expansion was led by ChatGPT, which popularized and brought closer generative models to the masses. Users can have access to powerful models for a relatively cheap monthly fee. The increased availability of such models has to lead to an increase in their misuse, especially when considering their capabilities. Depending on different governments, the laws of some of them hold the platform, like social networks, responsible for the actions of its users. It is important to have clear regulations for the use of generated content so that neither

979-8-3503-8459-8/24/$31.00 ©2024 IEEE

DOI: 10.1109/AIC.2024.42

the users nor the owners of the platform get harmed in any way. The increased adoption of AI in real-world scenarios indicates its huge potential. Many different branches of AI have emerged that tackle problems based on nature of the data. The detection of fake news can be performed by the combination of natural language processing (NLP) and AI.This work employs TF-IDF NLP technique for data preprocessing. The idea is to provide the importance of words based on its frequency and its frequency over all the documents into question. Consequentially, the context of a term is possible to be revealed.

Research presented in this article employs extreme gradient boosting (XGBoost) machine learning (ML) model for tackling fake news detection challenge, which was selected due to its high performance, especially when coupled with TF-IDF technique. However, the true potential of ML classification can be achieved through optimization. Therefore, this work also utilizes one variation of the variable neighborhood search (VNS) metaheuristics for tuning the XGBoost model. The ML algorithms are controlled by their unique set of hyper-parameters and to achieve optimal results tuning is required for each use case as a different combination of hyper-parameters is optimal (or sub-optimal) for different cases.

The key points of this research can be summarized as:

- Application of NLP for the fake news detection problem;
- Improvements of VNS metaheuristics towards better XGBoost's hyper-parameters' optimization and
- Fake news detection framework that is based on the XGBoost and modified VNS algorithm.

The organization of topics in this work is given as: Section II provides fundamental information on the applied techniques in this work, Section III provides the basis of the work, Section IV gives information regarding the performed experiments, as well simulation outcomes and disucssion. The final remarks about the proposed research are given in Section V.

II. Related Works

The spread of misinformation is especially troublesome on social media [1]. In this environment, the information spreads at enormous speeds. This happens because of the nature of the social networks as well as the nature of the fake news that is spread. The concept of social networks is to socialize with other people and first of all share interesting media. When a user encounters such information they usually instinctively forward it as it is always interesting. However, the cost of such actions can even result in loss of lives, for example during cases like pandemics.

Yildirim [13] proposes a novel method based on the threading principle. This work tackles the problem of detecting fake news in social media. This term references media that is shared in virtual communities, which does not have to exclusively mean social networks. A notable feature is the supervisor thread which is used to monitor and optimize the work of other threads. The other type of threads that are used are the worker threads which use shared objects to provide feedback to the supervisor thread.

Al-Ahmad et al. [2] explored the k-nearest neighbors (kNN) approach for detecting fake news during the COVID-19 pandemic. The work explored swarm metaheuristics for optimization of the kNN which was used as the main predictor. The spread of fake news during pandemics increases and overwhelms information sources even quicker than in cases where general fear is not present. For this reason, the importance of creating a robust fake news detection system is paramount.

Seddari et al. [12] employ a two-step NLP procedure for feature extraction. The model takes into account the linguistic features of the text along with fact-checking. The authors report significant improvements in the random forest algorithm application along the dual NLP mechanism in comparison to use cases that only employ one of the two.

Deepak and Chitturi [9] explore a deep learning approach for fake news detection with NLP techniques. The authors performed two experiments where the performance of deep learning techniques was tested with the application of text mining and without it. In the first set feed-forward neural network was applied, and in the second one the long short-term memory model. The authors applied words, word2vec, and GloVe embedders in the preprocessing of data. Text mining was performed with the HTML parser Beautiful Soup and through its use more data was provided on the articles that were analyzed. The results indicate significant improvements with the use of text mining along with deep learning techniques.

The use of metaheuristics in combination with NLP and ML techniques indicates a literature gap in the problem of fake news detection. Optimization techniques adhere to the no free lunch (NFL) theorem, and the best possible results are still to be achieved.

A. XGBoost

The performance in XGBoost is its differentiating factor in terms of machine learning methods [6], [7]. However, to achieve the highest possible performance the parameters of the model need to be tuned. The application of many weaker predictors in combination with a precise prediction is the main principle of the XGBoost solution. When applied with regularization and gradient boosting, optimization can result in significant performance improvements. The predictions are made from the observed patterns with the goal of complex input and target dependencies understanding.

The tuning of the XGBoost's hyper-parameters is not feasible through the trial and error method due to their large number. For this reason, an optimization algorithm needs to be introduced which is tasked with predicting the best possible combination for the hyper-parameters. The optimization process goals are speed, accuracy, and generalization.

The best results are achieved through an iterative process [6], and the Eq. 1 provides the objective function of the XGBoost.

$$\mathrm{obj}(\Theta) = L(\theta) + \Omega(\Theta), \tag{1}$$

in which the regularization term and loss function are combined. The hyperparameter set is represented by Θ, the loss function by $L(\Theta)$, and the regularization term which manages the model's complexity as $\Omega(\Theta)$.

For the loss function, the mean square error (MSE) is applied which is provided in Eq. 2.

$$\mathrm{L}(\Theta) = \sum_i (y_i - \hat{y_i})^2, \quad (2)$$

in which the predicted value is indicated by y_i, while the value of the predicted target variable is indicated by $\hat{y_i}$ for each iteration i.

The differentiation process of actual and predicted values is described in Eq. 3. Overall loss function minimization increases classification values.

$$L(\Theta) = \sum_i [\, y_i \ln (1 + e^{-\hat{y}_i}) + (1 - y_i) \ln (1 + e^{\hat{y}_i})]. \quad (3)$$

B. TF-IDF

Information search and text mining are the primary uses for TF-IDF. The method uses weight calculation for the analysis of feature importance for a term in a file. The described process is provided in Eq. 4.

$$TF - IDF(t, d, D) = tf(t, d) * idf(t, D) \quad (4)$$

in which term frequency is given as TF and it represents a specific term's t frequency in a document. The number of repetitions in a file indicates term frequency. The TF value of a single term has infinite divergence as it rises. The inverse document frequency is indicated by the IDF and it is calculated by dividing the total document number by the number of documents in which the term appears in. By applying this, terms that are general and appear across all documents can have diminished importance. Furthermore, this emphasizes the importance of rarely used terms that are potentially more informative. Document frequency is calculated according to Eq. 5.

$$DF(t, D) = \frac{t(n)}{D(n)} \quad (5)$$

in which the number of documents in which the term t appears in are marked as t, the number of all documents AS D, while the term frequency in documents is marked as DF. The increase in the term's use over documents indicates a lack of discrimination in that term. The words that hold no information like "a", "the", "and", and similar words should have their importance reduced. Therefore, DF needs to be calculated to indicate informative words. The logarithm value of DF is used in IDF to avoid infinite importance for words that appear 0 times. This behavior is described by Eq. 6.

$$IDF(t, D) = log(\frac{D(n)}{t(n)}) \quad (6)$$

The higher values of $TF - IDF$ are provided for terms that have high single document frequency and low overall document frequency.

C. Metaheuristics optimization

Algorithms of metaheuristic nature are applied for various cyber-threat purposes as predictors [10], as well as optimizers [15], [16]. The field of swarm intelligence is distinguished among other metaheuristic-based approaches due to its high capabilities for optimization. The algorithms of this type are highly adaptable as well. However, the full potential of these solutions is achieved through the application of a technique called hybridization. The process consists of improving an existing solution with another excelling in the shortcomings of the first solution. The nature of swarm intelligence is stochastic population-based with two distinct phases. The first is the exploration phase in which the solution is looked for in the search space globally. The second phase is exploitation which focuses on a specific prominent region and employs a more precise search.

The algorithms from the field of swarm intelligence are widely applied in the real world. Various examples include load prediction of cloud platforms [5], fraud detection [11], intrusion detection [3], spam detection [14], and even detection of Parkinson's disease [4].

III. Methods

A. The Original Variable Neighbourhood Search Algorithm

The VNS is a local search strategy with the purpose of combinatorial optimization. The basic principle of the VNS algorithm is the systematic neighborhood structure change. What differentiates VNS from other local search methods is that it does not follow a trajectory. The method explored increasingly distant neighborhoods instead, where the transition from one neighborhood to another is performed only when improvements are achieved.

It is typical for the basic version of VNS to alternate between local search and shake procedures until a stopping criterion is reached. The local search is used to find local minima, while the shake technique enhances diversification. The general version of the VNS includes an iterative process that is repeated until all neighborhoods have been evaluated. During the process, a random point is generated from which local search is employed. If the next evaluated solution is better move to it and continue searching in the same neighborhood. Otherwise, move to the next neighborhood.

The shaking procedure randomly selects a neighborhood $\mathcal{N}_l(x)$, from the search space $\mathcal{N} = \mathcal{N}_1, ..., \mathcal{N}_{k_{max}}$ where each $\mathcal{N}$ maps x agent to a neighborhood. The order in the population determines which neighborhood will be evaluated first. The switch between neighborhoods can be performed with different techniques where the solutions are moved to explore the search space. Some of the neighborhood change steps include the sequential step where the neighborhood is revised if a better solution is found, the cyclic step where the list is iterated irrespective of the improvements, and the pipe step where the search is continued in the current neighborhood as long as improvements are made.

The local neighborhoods $\mathcal{N}(x)$ of the current solution x are explored in each iteration and replaced through the improvement protocol with a better solution x'. The first and best improvement strategies are the most common. The first strategy picks the new solution as soon as a better one is obtained, and the best strategy after the evaluation of the population.

B. Adaptive Variable Neighbourhood Search Algorithm

The modifications to the general VNS are towards a better balance between exploration and exploitation. The algorithm favors exploitation due to its nature and a mechanism for balancing exploration is needed. This can lead to premature convergence and homogenous populations. To tackle this issue, a σ parameter is introduced. This is a depreciating factor, favoring exploitation in later phases. As the factor decreases the algorithm should favor its exploitation principles instead of the exploration creating a mechanism that balances these two phases. The name of the solution is the adaptive VNS (AVNS) and the pseudocode is provided in Alg. 1.

Algorithm 1 Pseudo-code of the AVNS

```
Function VNS(x, k_max, t_max, sigma, sigmaStep)
t = 0
repeat
    k = 1
    repeat
        if random() < sigma then
            x' = Shake(x, k)
        else
            x' = x
        end if
        x'' = BestImprovement(x')
        x = NeighbourhoodChange(x, x'', k)
        sigma = max(sigma − sigmaStep, 0)
        k = k + 1
    until k = k_max
    t = CpuTime()
until t > t_max
return x
```

IV. Experimental Setup, Outcomes and Discussion

This research applies ISOT fake news dataset which is publicly available at the link: https://www.kaggle.com/datasets/emineyetm/fake-news-detection-datasets. The included data is of real and fake nature. The https://www.reuters.com/ was crawled for truthful articles, while the fake news where collected from different sources due to their nature. The websites for fake news articles were obtained by referencing Politifact, which is a USA-based fact-checking website, and Wikipedia as well. The majority of fake news articles were from the field of politics, but overall the specter of topics is wide.

The data is separated into two CSV files. The first contains 12.600 truthful articles. The second one consists of almost double the amount of fake news in comparison to the first one. The text, type, date, and title of the article are provided as features. The experiments were executed in 30 independent runs, throughout 15 iterations and 10 units in the population.

The true positive (TP) to all predictions ratio is depicted by the precision metric. The metric is provided in Eq. (7).

$$P = \frac{TP}{TP + FP} \tag{7}$$

The ratio between correct positive predictions and all positive predictions including false negative (FN) is depicted by recall in Eq. (8).

$$R = \frac{TP}{TP + FN} \tag{8}$$

The f1 score depicts the harmonic mean of precision and recall depicted by Eq.9. A perfect precision is depicted by value 1 and the range of values is $[0, 1]$.

$$F1 = 2 * (precision * recall)/(precision + recall) \tag{9}$$

Finally, the XGBoost hyper-parameters' which were subjected to optimization are shown in Table I.

TABLE I
XGBoost parameter constrains for the optimization.

Parameter	Lower Bound	Upper Bound
Learning Rate	0.1	0.9
Min child weight	1	10
Subsample	0.1	1.0
Colsample by tree	0.1	1.0
Max depth	1	3
Gamma	0	0.8

To tackle the imbalances in the true and fake classes of the dataset Cohen's kappa indicator [8] is used as the objective function. The metric measures inter-rater and intra-rater reliability in categorical use cases. The results of the objective function experiments are provided in Table II. The proposed XG-AVNS obtains the best results for best, worst, mean, and median dominating these categories. However, the best results for standard deviation and variance are obtained by the XG-PSO.

The indicator functions indicate the behavior as in Table II. The results of error minimization used for this purpose are reported in Table III. The XG-AVNS obtained the best results for best, worst, mean, and median categories, while the XG-PSO again provided the best standard deviation and variance. It is important to note that XG-VNS which applies the unmodified version of the AVNS does not outperform the proposed solution in any of the categories.

The violin and error box plot diagrams are provided in Fig. 1. The XG-AVNS obtains the best performance from both of the graphs. In the box blot, the XG-AVNS had the narrowest interquartile range and the lowest value obtained, while in the violin plot, the solution obtained the highest values and exhibited narrow distribution.

TABLE II
Fake news detection objective function results.

Method	Best	Worst	Mean	Median	Std	Var
XG-AVNS	**0.759801**	**0.749584**	**0.754999**	**0.755070**	0.003210	1.03E-05
XG-VNS	0.756862	0.745006	0.750850	0.750457	0.003140	9.86E-06
XG-GA	0.752911	0.740558	0.748978	0.750240	0.003613	1.31E-05
XG-PSO	0.753440	0.747499	0.750399	0.750413	**0.001814**	**3.29E-06**
XG-ABC	0.750202	0.733546	0.745639	0.747199	0.004529	2.05E-05
XG-SCA	0.753279	0.746055	0.750038	0.749913	0.002357	5.56E-06
XG-SCHA	0.754757	0.748222	0.751094	0.750095	0.002374	5.63E-06
XG-COLSHADE	0.759239	0.746472	0.751377	0.750139	0.003468	1.20E-05

TABLE III

FAKE NEWS DETECTION INDICATOR FUNCTION RESULTS.

Method	Best	Worst	Mean	Median	Std	Var
XG-AVNS	**0.120564**	**0.125835**	**0.123051**	**0.123014**	0.001673	2.80E-06
XG-VNS	0.122123	0.128211	0.125167	0.125316	0.001571	2.47E-06
XG-GA	0.124053	0.130438	0.126154	0.125501	0.001851	3.43E-06
XG-PSO	0.123979	0.126875	0.125419	0.125390	**0.000911**	**8.29E-07**
XG-ABC	0.125390	0.134076	0.127847	0.127060	0.002337	5.46E-06
XG-SCA	0.123831	0.127691	0.125561	0.125575	0.001236	1.53E-06
XG-SCHA	0.123237	0.126503	0.125108	0.125612	0.001208	1.46E-06
XG-COLSHADE	0.120861	0.127468	0.124900	0.125575	0.001791	3.21E-06

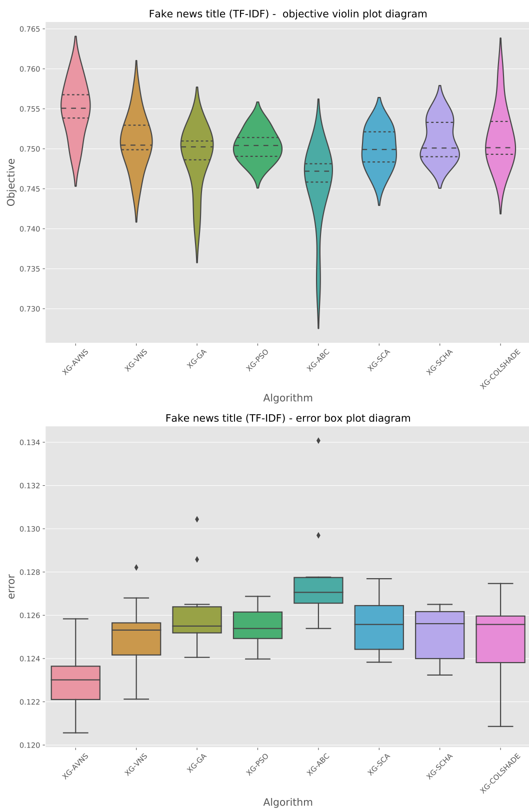

Fig. 1. Fake news detection distribution plots.

The detailed metrics are given in Table IV. The results indicate the dominance of the proposed XG-AVNS across almost all categories. The algorithm was outperformed only in the following cases: real close recall metric, fake class precision metric, and macro and weighted averages precision, which were all obtained by the XG-PSO. The PSO solution still proves relevant and indicates possible inspiration for further improvement of the proposed solution.

The convergences of the solutions compared are provided in Fig. 2. The XG-AVNS obtained the best values for objective and error convergences. Other algorithms provide slightly worse performance with the XG-COLSHADE variant following behind the closest to the XG-AVNS.

The best-obtained hyper-parameters by each metaheuristic-based tested solution are provided in Table V.

The precision-recall (PR) curve and the confusion matrix are given in Fig. 3. The PR curve indicates high scores with both classes, where the true news class had the score of 0.925 and

TABLE IV

FAKE NEWS DETECTION DETAILED METRICS.

Method	metric	real	fake	accuracy	macro avg	weighted avg
XG-AVNS	precision	**0.828791**	0.939374	**0.879436**	0.884082	0.886627
	recall	0.941790	**0.822569**	**0.879436**	**0.882180**	**0.879436**
	f1-score	**0.881684**	**0.877100**	**0.879436**	**0.879392**	**0.879287**
XG-VNS	precision	0.822886	0.944957	0.877877	0.883921	0.886731
	recall	0.948016	0.813911	0.877877	0.880963	0.877877
	f1-score	0.881030	0.874552	0.877877	0.877791	0.877642
XG-GA	precision	0.823930	0.938193	0.875947	0.881062	0.883691
	recall	0.941012	0.816608	0.875947	0.878810	0.875947
	f1-score	0.878588	0.873188	0.875947	0.875888	0.875764
XG-PSO	precision	0.814276	**0.955123**	0.876021	**0.884700**	**0.887941**
	recall	**0.958755**	0.800568	0.876021	0.879661	0.876021
	f1-score	0.880629	0.871042	0.876021	0.875836	0.875615
XG-ABC	precision	0.823940	0.934740	0.874610	0.879340	0.881890
	recall	0.937432	0.817317	0.874610	0.877375	0.874610
	f1-score	0.877029	0.872094	0.874610	0.874562	0.874448
XG-SCA	precision	0.826135	0.935244	0.876169	0.880689	0.883200
	recall	0.937743	0.820014	0.876169	0.878879	0.876169
	f1-score	0.878408	0.873847	0.876169	0.876127	0.876022
XG-SCHA	precision	0.818985	0.948675	0.876763	0.883830	0.886815
	recall	0.952062	0.808091	0.876763	0.880077	0.876763
	f1-score	0.880524	0.872758	0.876763	0.876641	0.876462
XG-COLSHADE	precision	0.827709	0.940335	0.879139	0.884022	0.886614
	recall	0.942879	0.821008	0.879139	0.881944	0.879139
	f1-score	0.881548	0.876629	0.879139	0.879089	0.878976
	support	6425	7045			

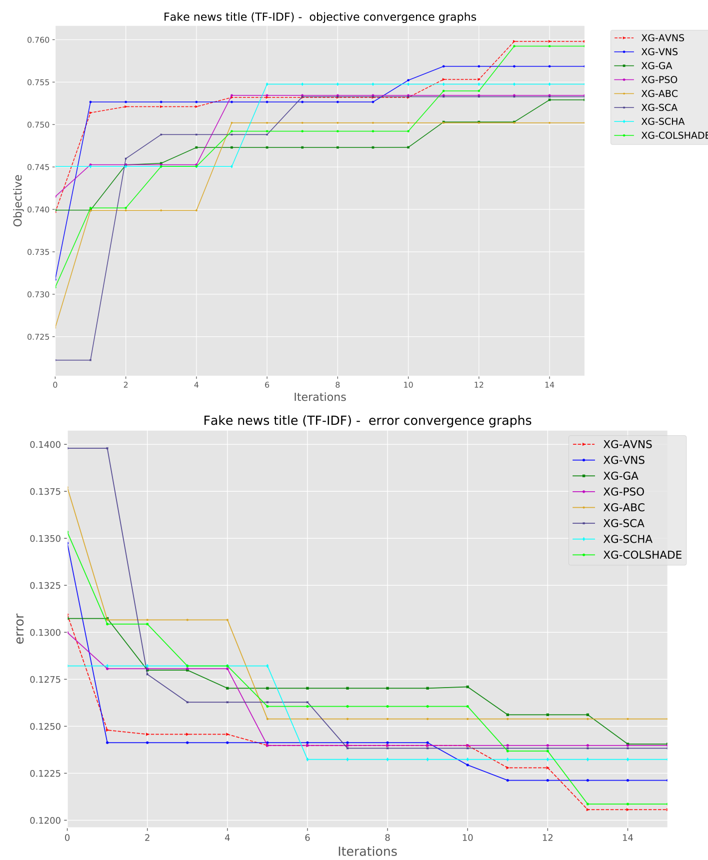

Fig. 2. Fake news detection convergence graphs.

TABLE V

PARAMETER SELECTIONS MADE BY EACH OPTIMIZER FOR THE RESPECTIVE BEST-PERFORMING MODELS.

Method	Learning Rate	Min Child W.	Subsample	Col by Tree	Max depth	Gamma
XG-AVNS	0.900000	1.285605	0.717967	0.163904	10	0.363737
XG-VNS	0.900000	1.000000	0.948218	0.259884	10	0.738255
XG-GA	0.900000	1.000000	0.573455	0.245407	10	0.800000
XG-PSO	0.900000	1.057344	1.000000	0.153352	10	0.557096
XG-ABC	0.897490	4.476197	0.637297	0.193511	10	0.047998
XG-SCA	0.900000	2.463144	0.825472	0.182204	10	0.074978
XG-SCHA	0.900000	2.249915	0.994411	0.251738	10	0.423137
XG-COLSHADE	0.900000	1.201168	0.866329	0.163116	10	0.000000

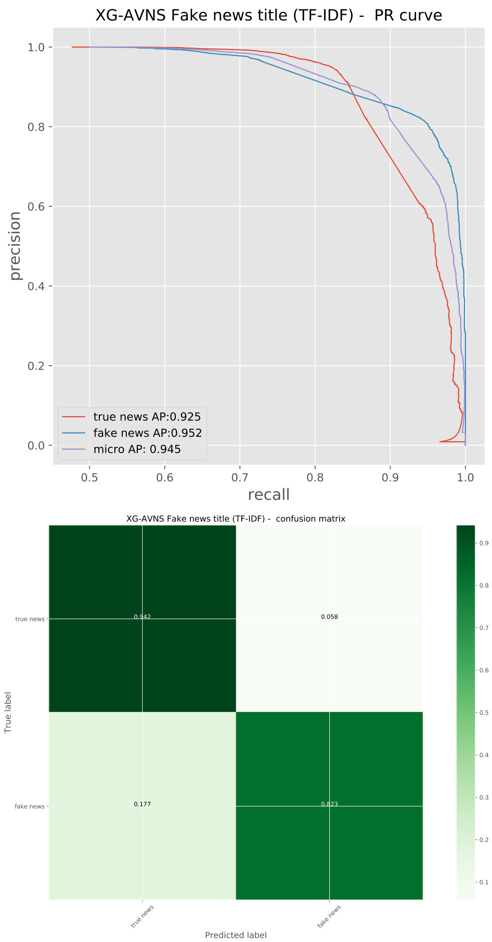

Fig. 3. Fake news detection PR and convergence matrix.

the fake news class 0.952, with a micro average of 0.945. The confusion matrix indicates high stability in the predictions of the proposed solution. The false positive rate can be observed from Fig. 3 and it is 5.8%, while the false negative rate is 17.7%.

V. Conclusion

This work tackled an important issue, the detection of fake news, which is crucial due to its implications and malicious intents such as fraud. The paper tackles this issue by providing a robust classifier of fake news which is based on the XG-Boost model. The NLP approach was applied to determine the importance of terms called TF-IDF. Furthermore, due to the necessity of XGboost hyper-parameters' optimization, a metaheuristic solution is applied. The VNS algorithm is tested for this purpose, and an improved version of this algorithm, called AVNS, is applied to establish improvements in determining proper set of XGBoost hyper-parameters for fake news detection.

This research was limited by the high computational resource requirements and larger datasets could not be used. However, future experiments can be expanded in this direction. Furthermore, for the same reason, more algorithms were not compared which could possibly indicate a better solution for this challenge due to the NFL.

References

[1] E. Aïmeur, S. Amri, and G. Brassard, "Fake news, disinformation and misinformation in social media: a review," *Social Network Analysis and Mining*, vol. 13, no. 1, p. 30, 2023.

[2] B. Al-Ahmad, A. Al-Zoubi, R. Abu Khurma, and I. Aljarah, "An evolutionary fake news detection method for covid-19 pandemic information," *Symmetry*, vol. 13, no. 6, p. 1091, 2021.

[3] N. Bacanin, A. Petrovic, M. Antonijevic, M. Zivkovic, M. Sarac, E. Tuba, and I. Strumberger, "Intrusion detection by xgboost model tuned by improved social network search algorithm," in *International Conference on Modelling and Development of Intelligent Systems*. Springer, 2022, pp. 104–121.

[4] N. Bacanin, A. Petrovic, L. Jovanovic, M. Zivkovic, T. Zivkovic, and M. Sarac, "Parkinson's disease induced gain freezing detection using gated recurrent units optimized by modified crayfish optimization algorithm," in *2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)*. IEEE, 2024, pp. 1–8.

[5] N. Bacanin, V. Simic, M. Zivkovic, M. Alrasheedi, and A. Petrovic, "Cloud computing load prediction by decomposition reinforced attention long short-term memory network optimized by modified particle swarm optimization algorithm," *Annals of Operations Research*, pp. 1–34, 2023.

[6] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[7] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou *et al.*, "Xgboost: extreme gradient boosting," *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.

[8] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[9] S. Deepak and B. Chitturi, "Deep neural approach to fake-news identification," *Procedia Computer Science*, vol. 167, pp. 2236–2243, 2020.

[10] F. A. Ozbay and B. Alatas, "A novel approach for detection of fake news on social media using metaheuristic optimization algorithms," *Elektronika ir Elektrotechnika*, vol. 25, no. 4, pp. 62–67, 2019.

[11] A. Petrovic, M. Antonijevic, I. Strumberger, L. Jovanovic, N. Savanovic, and S. Janicijevic, "The xgboost approach tuned by tlb metaheuristics for fraud detection," in *Proceedings of the 1st international conference on innovation in information technology and business (ICIITB 2022)*, vol. 104. Springer Nature, 2023, p. 219.

[12] N. Seddari, A. Derhab, M. Belaoued, W. Halboob, J. Al-Muhtadi, and A. Bouras, "A hybrid linguistic and knowledge-based analysis approach for fake news detection on social media," *IEEE Access*, vol. 10, pp. 62 097–62 109, 2022.

[13] G. Yildirim, "A novel hybrid multi-thread metaheuristic approach for fake news detection in social media," *Applied Intelligence*, vol. 53, no. 9, pp. 11 182–11 202, 2023.

[14] M. Zivkovic, A. Petrovic, N. Bacanin, M. Djuric, A. Vesic, I. Strumberger, and M. Marjanovic, "Training logistic regression model by hybridized multi-verse optimizer for spam email classification," in *Proceedings of International Conference on Data Science and Applications: ICDSA 2022, Volume 2*. Springer, 2023, pp. 507–520.

[15] M. Zivkovic, A. Petrovic, K. Venkatachalam, I. Strumberger, H. S. Jassim, and N. Bacanin, "Novel chaotic best firefly algorithm: Covid-19 fake news detection application," in *Advances in Swarm Intelligence: Variations and Adaptations for Optimization Problems*. Springer, 2022, pp. 285–305.

[16] M. Zivkovic, C. Stoean, A. Petrovic, N. Bacanin, I. Strumberger, and T. Zivkovic, "A novel method for covid-19 pandemic information fake news detection based on the arithmetic optimization algorithm," in *2021 23rd international symposium on symbolic and numeric algorithms for scientific computing (SYNASC)*. IEEE, 2021, pp. 259–266.