UDC 070.16-048.445:004.85 **doi: 10.32620/reks.2023.4.01**

Halyna PADALKO[1, 2, 3], Vasyl CHOMKO[2],
Sergiy YAKOVLEV[4], Dmytro CHUMACHENKO[1]

[1] ***National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine***
[2] ***University of Waterloo, Waterloo, Canada***
[3] ***Balsillie School of International Affairs, Waterloo, Canada***
[4] ***Lodz University of Technology, Lodz, Poland***

ENSEMBLE MACHINE LEARNING APPROACHES FOR FAKE NEWS CLASSIFICATION

In today's interconnected digital landscape, the proliferation of fake news has become a significant challenge, with far-reaching implications for individuals, institutions, and societies. The rapid spread of misleading information undermines the credibility of genuine news outlets and threatens informed decision-making, public trust, and democratic processes. Recognizing the profound relevance and urgency of addressing this issue, this research embarked on a mission to harness the power of machine learning to combat fake news menace. This study ***develops*** *an ensemble machine learning model for fake news classification. The research is* ***targeted*** *at spreading fake news. The research* ***subjects*** *are machine learning methods for misinformation classification.* ***Methods:*** *we employed three state-of-the-art algorithms: LightGBM, XGBoost, and Balanced Random Forest (BRF). Each model was meticulously trained on a comprehensive dataset curated to encompass a diverse range of news articles, ensuring a broad representation of linguistic patterns and styles. A distinctive feature of the proposed approach is the emphasis on token importance. By leveraging specific tokens that exhibited a high degree of influence on classification outcomes, we enhanced the precision and reliability of the developed models. The empirical* ***results*** *were both promising and illuminating. The LightGBM model emerged as the top performer among the three, registering an impressive F1-score of 97.74% and an accuracy rate of 97.64%. Notably, all three of the proposed models consistently outperformed several existing models previously documented in academic literature. This comparative analysis underscores the efficacy and superiority of the proposed ensemble approach. In* ***conclusion****, this study contributes a robust, innovative, and scalable solution to the pressing challenge of fake news detection. By harnessing the capabilities of advanced machine learning techniques, the research findings pave the way for enhancing the integrity and veracity of information in an increasingly digitalized world, thereby safeguarding public trust and promoting informed discourse.*

Keywords: *fake news; classification; misinformation; disinformation; balanced random forest; XGBoost; LightGBM; WELFake; machine learning.*

Introduction

In the contemporary digital age, the proliferation of misinformation and disinformation has emerged as a pressing concern. Misinformation, defined as false or inaccurate information shared without malicious intent [1], and disinformation, which is deliberately disseminated to deceive, pose significant threats to the integrity of public discourse, informed decision-making, and the fabric of democratic societies [2]. The ubiquity of digital platforms and the rapid dissemination of information have exacerbated these challenges, making it urgent for scholars, policymakers, and technologists to address them [3].

The alarming spread of fake news is parallel to the challenges of misinformation and disinformation [4]. Fake news, which is often sensationalized and devoid of factual grounding, is not merely an informational concern but a societal one [5]. Its rapid dissemination can sway public opinion, influence electoral outcomes, and even incite real-world harm [6]. The viral nature of fake news, propelled by social media algorithms and human cognitive biases, underscores the need for effective countermeasures to ensure the veracity of information consumed by the public [7].

As the digital landscape becomes increasingly complex, innovative approaches are being developed to investigate and combat misinformation [8]. Digital forensics, data analytics, and machine learning algorithms are at the forefront of these efforts [9]. These tools allow researchers to trace the origins of misleading narratives, understand their propagation patterns, and develop strategies to counteract their spread [10]. Furthermore, collaborations among tech companies, academia, and civil society are fostering the development of tools that can detect and flag dubious content in real time, aiding users in discerning fact from fiction [11].

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

The task of classifying fake news has never been more relevant. Manual verification becomes an insurmountable challenge with the sheer volume of online information. Automated classification systems that leverage advanced computational techniques offer a promising solution [12]. By training models on vast datasets, these systems can discern patterns typical of fake news stories, enabling timely detection and mitigation. The actuality of such classification systems is underscored by their potential to safeguard information ecosystems, ensuring that individuals and communities are equipped with accurate and reliable information in an era of digital uncertainty.

Machine learning (ML) has demonstrated remarkable efficacy and potential in the detection of fake news [13]. Leveraging intricate algorithms and computational power, ML models can analyze and interpret vast data arrays at unprecedented speeds, identifying nuanced patterns and inconsistencies that are often imperceptible to human analysts. When trained on comprehensive datasets comprising both genuine and deceptive news articles, such models can achieve impressive accuracy rates in discerning factual reports from fabricated narratives [14]. Moreover, the adaptability of ML ensures that these models continuously evolve, refining their analytical capabilities in response to the ever-changing tactics employed by misinformation purveyors. Consequently, the integration of machine learning in the battle against fake news augments the efficiency of detection mechanisms and fortifies the resilience of information ecosystems against deceptive onslaughts.

The aim of this paper is to develop an ensemble machine learning model for fake news classification. The research is targeted at spreading fake news. The research subjects are machine learning methods for misinformation classification.

To achieve the aim of the study, the following tasks were formulated:

1. To analyze machine learning models and methods for fake news classification.
2. To develop the XGBoost model for fake news classification.
3. To develop Random Forest model for fake news classification.
4. To develop LightGBM model for fake news classification.
5. To evaluate the models' performance and classification results.

The promising contributions of this research are manifold and have significant implications for the domain of misinformation detection. First, this study comprehensively analyzes existing machine learning models tailored for fake news classification, offering a consolidated overview of the current state-of-the-art methodologies. Second, this research aims to push the boundaries of classification accuracy and efficiency by developing and fine-tuning models such as XGBoost, Random Forest, and LightGBM specifically for fake news detection. Such dedicated model development is anticipated to yield algorithms more attuned to the intricacies and nuances of deceptive narratives. Furthermore, the rigorous evaluation of these models' performance will furnish empirical evidence of their efficacy, potentially establishing new benchmarks in the field. Collectively, these contributions advance the technical prowess of fake news detection mechanisms and fortify the broader endeavor to maintain the integrity of information ecosystems in the digital age.

The further structure of the paper is the following: Section 1, Current research analysis, briefly describes the current state of machine learning application for fake news detection. Section 2, Materials and Methods, describes the developed models and their tuning. Section 3, Data, describes the dataset. Section 4, Results, provides results of the experimental study. Section 5, Discussion, discusses the obtained results. The conclusion describes the outcomes of this study.

1. Current Research Analysis

In recent years, the application of machine learning techniques to the challenge of fake news detection has garnered significant attention in academic and technological circles [15]. These techniques, harnessing the power of data-driven analytics, have demonstrated the ability to sift through vast textual corpora, extracting features and indicators that differentiate genuine news from deceptive narratives. Advanced models, such as deep neural networks and ensemble classifiers, have been particularly effective in this endeavor, offering nuanced insights into the structure and semantics of content. The continuous feedback loop inherent in machine learning allows for real-time refinement, ensuring that detection systems remain updated despite evolving misinformation strategies [16]. This fusion of computational prowess and adaptive learning underscores the potential of machine learning as a formidable tool in the quest to maintain informational authenticity in digital ecosystems.

The research paper [17] addresses the challenge of detecting fake news in Urdu, a domain that has been underexplored compared with its Western counterparts. While there have been numerous endeavors in fake news detection for Western languages, the Urdu language has faced a dearth of attention, primarily due to the limited availability of datasets. To bridge this gap and encourage further research, this study introduces "UrduFake'21", a dedicated track for Urdu fake news detection. The paper's primary contribution is introducing an ensemble machine learning model that uses a voting mechanism, combining three potent techniques to determine the authenticity of Urdu news articles. This model was evaluated using machine learning methods, three feature types (unigram, bigram, and trigram), and the newly released dataset. The empirical results underscore the superiority of the ensemble approach over singular methods. The proposed model achieved commendable macro average F1 and accuracy scores of 0.621 and 0.713, respectively.

The study [18] delves into the challenges posed by the proliferation of fake news on social media platforms, emphasizing its potential to undermine democratic processes and the trustworthiness of information. Given the intentional design of fake news to mislead readers, its detection presents a unique challenge. While prior research has indicated the potential of machine learning in addressing this issue, the authors introduce a novel approach using the k-nearest neighbors (KNN) algorithm. A significant enhancement to their system is the integration of Genetic and Evolutionary Feature Selection (GEFeS), which bolstered the model's accuracy to 91.3%. Further, this study explored the realm of quantum machine learning by employing a quantum KNN (QKNN) model trained using features identified by GEFeS. This quantum-based approach yielded an accuracy of 84.4%, demonstrating its potential applicability in the domain of fake news detection.

The research study [19] examined the efficacy of evolutionary algorithms in fake news detection, explicitly focusing on genetic algorithms. This study presents a comparative analysis of several SVM, Naïve Bayes, Random Forest, and Logistic Regression classifiers applied to different datasets. The SVM classifier demonstrated notable performance in the Liar, Fake Job Posting, and Fake News datasets. In addition, this paper introduces a genetic algorithm-based fake news detection method, where SVM, Naïve Bayes, Random Forest, and Logistic Regression serve as the fitness functions. In this proposed algorithm, both the SVM and LR classifiers achieve 61% accuracy on the LIAR dataset. In comparison, the SVM and RF classifiers attain 97% accuracy on the Fake Job Posting dataset.

The research study [20] aims to develop an automated system for detecting fake news with superior accuracy compared with existing methodologies. Using a dataset comprising 7796 news articles evenly split between genuine and fake news, this study evaluates the performance of several machine learning classifiers, including Random Forest, Decision Tree, KNN, Logistics Regression, and SVM. The SVM classifier emerges as the most effective, achieving an accuracy of 93.61%, surpassing results from a recent research in the field. By leveraging the SVM classifier, this study offers a robust solution for distinguishing between authentic and fabricated news, thereby enhancing the reliability of information consumption.

This paper [21] addresses the challenge of fake news detection on social media platforms. While many existing detection methods primarily analyze the linguistic and compositional features of fake news, this study introduces a machine learning-based model that also considers user characteristics, news content, and social network dynamics rooted in social capital. This study employs the XGBoost model to determine the significance of each feature, thereby identifying key factors influencing fake news detection. Several machine learning classification models, including SVM, RF, LR, CART, and NNET, were then developed using these identified features. To ensure the robustness of these models, a cross-validation step is undertaken. Among the models, the RF classifier demonstrates superior performance with a prediction accuracy of approximately 94%, whereas the NNET model registers an accuracy of approximately 92.1%. The findings offer valuable insights for enhancing fake news detection systems, especially considering the evolving nature of fake news generation and dissemination.

The research paper [22] delves into the challenge of detecting fake news in low-resource languages, explicitly focusing on Kurdish. While numerous automated detection methods exist for widely spoken languages such as English and Arabic, there must be more solutions for languages such as Kurdish. To address this, the study utilizes the Kurdish fake news dataset (KDFND), which contains 100,962 news articles, roughly half of which are genuine and the other half is fabricated. The research employs three feature extraction techniques from news texts: word embedding, term frequency-inverse document frequency, and count vector. In addition, three classifiers, namely Random Forest, Support Vector Machine, and Convolutional Neural Networks (CNN), were tested for their efficacy in identifying fake news. The findings highlight the superior performance of CNN, achieving an F1-score of 95% and an accuracy exceeding 91%. This study underscores the potential of machine learning techniques in effectively detecting fake news in underrepresented languages such as Kurdish.

The paper [23] addresses the pressing issue of fake news detection, emphasizing its broad implications from local to global scales. While various methods exist for identifying fake news, this study introduces a unique machine learning model, Chaotic Ant Swarm with Weighted Extreme Learning Machine (CAS-WELM), which is tailored explicitly for Cybersecurity Fake News Detection and Classification. The CAS-WELM model begins with data preprocessing and applies the Glove technique for word embedding. Subsequently, an N-gram-based approach is employed to extract features

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

and generate feature vectors. The final classification of news as genuine or fake is achieved using the WELM model, where its weight parameters are optimally adjusted using the CAS algorithm. The model's efficacy was validated using a benchmark dataset, and its performance was evaluated across various metrics. The findings highlight the superiority of the CAS-WELM approach over existing methods in fake news detection.

The study [24] addresses the growing challenge of fake news, which can have significant repercussions, from tarnishing individual or organizational reputations to generating undue advertising revenue. While fake news is a global concern, this study specifically focuses on its impact within the Bengali-speaking community, which has seen limited research. This study explores the application of machine learning techniques to detect fake news in Bengali and investigates the potential of incorporating sentiment analysis as a feature. Through experimentation, the study found that the SVM yielded the highest accuracy, achieving a score of 73.20%. However, the research also determined that sentiment analysis does not offer significant value as a feature for fake news detection in this context. The paper's primary contributions lie in broadening the scope of fake news detection research for the Bengali language and evaluating the utility of sentiment analysis in this domain.

The paper [25] addresses the pressing challenge of the COVID-19 "infodemic", where the rapid spread of fake news exacerbates public health concerns. While previous studies have demonstrated the capability of machine learning models to detect COVID-19-related fake news based on article content, the potential of integrating biomedical information frequently present in such news remains unexplored. This study introduces an innovative approach that combines biomedical information extraction (BioIE) with machine learning models to enhance fake news prediction. By analyzing 1,164 COVID-19 news articles, this study employs advanced BioIE algorithms to derive 158 unique features. These features were subsequently used to train 15 machine learning classifiers. The random forest classifier emerges as the most effective, achieving an area under the ROC curve (AUC) of 0.882, significantly outperforming models reliant on conventional features. Moreover, the integration of BioIE-based features enhances the efficacy of a leading multi-modality model. The findings underscore the value of incorporating biomedical information in fake news detection, offering a potent tool to combat the COVID-19 infodemic.

The research paper [26] addresses the pervasive issue of fake news, intentionally crafted to mislead readers. The rise of the internet and social media platforms has exacerbated the spread of such misinformation, especially given the minimal oversight of online content. This spread is particularly concerning in contexts such as presidential elections and health-related matters such as the COVID-19 pandemic. Recognizing the challenges humans face in discerning the veracity of such news, this study advocates the deployment of machine learning algorithms for detection and classification. This paper explicitly explores using Term Frequency-Inverse Document Frequency (TF-IDF) for feature extraction coupled with two machine learning classifiers: Support Vector Machine (SVM) and Multinomial Naive Bayes. The findings indicate that the SVM and Multinomial Naive Bayes classifiers achieve 94.83% and 91.38% accuracy, respectively. This study underscores the potential of machine learning in effectively identifying fake news, as evidenced by the high accuracy rates achieved.

The pervasive challenge of fake news detection has been the focal point of numerous studies, as evidenced by the diverse range of papers analyzed. These papers underscore the issue's significance, from its impact on public health, as seen in the COVID-19 infodemic, to its influence on democratic processes, such as presidential elections. A common thread across these studies is the exploration of machine learning techniques to discern the authenticity of news articles. While some research has delved into the potential of specific classifiers such as SVM, Random Forest, and Convolutional Neural Networks, others have ventured into innovative territories, leveraging biomedical information extraction or evolutionary algorithms. The emphasis on languages, from widely spoken ones such as English to low-resource ones such as Kurdish and Bengali, further highlights the global scope of the challenge. Given this backdrop, the current research's aim to develop an ensemble machine learning model for fake news classification is timely and pertinent.

2. Data

For the experimental study, we used the WELFake open dataset [27]. The WELFake dataset is a rich compilation of 72,134 news articles, with 35,028 being genuine and 37,106 categorized as fake. This comprehensive collection was crafted by merging four distinguished news datasets: Kaggle, McIntire, Reuters, and BuzzFeed Political. These datasets were integrated strategically to mitigate the risk of classifier over-fitting and to ensure a broader text corpus for more effective machine learning training. Each entry in the dataset is uniquely identified by a serial number, starting from 0. In addition, every article is characterized by its headline, detailed content, and a label indicating its authenticity. Specifically, the label '0' denotes a fake news article, while '1' signifies a real one. It is noteworthy to mention that while the associated CSV file of the dataset boasts 78,098 entries, only 72,134 of these are accessible as per the data frame, hinting at potential missing or redundant data points. Overall, given its diverse and balanced composition, the WELFake dataset is a valuable asset for those venturing into the domain of fake news detection.

Figure 1 shows the distribution of the dataset.

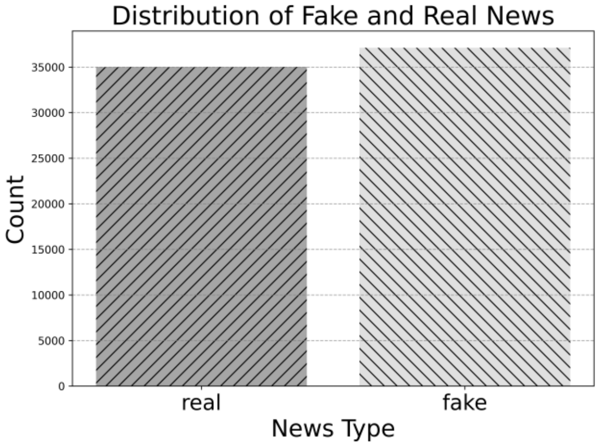

Fig. 1. The dataset distribution

Figure 2 shows the word cloud of the real text in the dataset. Figure 3 shows the word cloud of the fake text in the dataset. A word cloud, often called a tag cloud or text cloud, is a visual representation of text data whose size denotes the frequency of each word. It provides a macroscopic view of the most prominent terms in a dataset, allowing immediate recognition of dominant themes or patterns. In academic and research contexts, word clouds are effective preliminary tools for dataset analysis, offering a quick, intuitive snapshot of textual content. However, while they provide a general overview, deeper, more nuanced analyses are often required for comprehensive insights into the dataset. In the given study, the word clouds depicted in Figures 2 and 3 served as an initial step to discern the prevalent terms within the dataset's real and fake news segments.

Fig. 2. True text word cloud

The dataset was cleaned to eliminate potential noise sources, such as HTML tags and URLs, while retaining only the core textual information. Furthermore, we ensured text uniformity by converting all content to lowercase.

Fig. 3. Fake text word cloud

Between the data cleaning and vectorization processes, it is noteworthy that the dataset used was meticulously curated to ensure a balanced distribution between real and fake news articles. Such equilibrium is paramount in machine learning tasks because it prevents model biases and ensures that the classifier does not become predisposed to any particular class, thereby enhancing the generalizability and robustness of the subsequent analysis.

For vectorization, the Term Frequency-Inverse Document Frequency (TF-IDF) method was chosen. Unlike simple term frequency, TF-IDF quantifies a word's importance in a document relative to its frequency across all documents. This helps emphasize words that are more unique to a particular document, thereby aiding in differentiating real from fake news. The transformation was achieved using the TfidfVectorizer from scikit-learn, excluding English stop words to reduce dimensionality and remove commonly occurring words that do not contribute significantly to content differentiation.

3. Materials and Methods

3.1. Balanced Random Forest

The Balanced Random Forest (BRF) is an ensemble learning method designed to address the challenge of class imbalance in datasets [28]. Class imbalance is common in many real-world datasets where one class significantly outnumbers the other(s), leading to biased predictions. The BRF algorithm modifies the standard Random Forest (RF) approach by incorporating techniques to balance the class distribution before training each tree.

The Random Forest algorithm builds an ensemble of decision trees by bootstrapping samples from the dataset and training each tree on a different sample. The final prediction is obtained by aggregating the predictions from all trees, typically by majority voting for classification or averaging for regression. Before training each tree in the forest, the BRF algorithm undersamples the majority class to balance the class distribution. This ensures that each tree is trained on a dataset in which the classes are roughly equally represented.

Formulation:

Given a dataset with n samples and m features, the objective of a standard Random Forest is to minimize the generalization error by aggregating predictions from T decision trees. For a binary classification problem with classes C_1 (majority) and C_2 (minority), the BRF modifies the training set for each tree as follows:

1. From C_1, randomly select n_2 samples without replacement, where n_2 is the number of samples in C_2.

2. Combine the n_2 samples from C_1 with all samples from C_2 to form a balanced training set.

Each tree f_t in the forest is then trained on a different balanced training set. The final prediction for an instance x is given by:

$$\hat{y}(x) = \frac{1}{T}\sum_{t=1}^{T} f_t(x). \quad (1)$$

The class with the majority vote across all trees is chosen as the final prediction for classification.

Key Features:

- by training each tree on a different balanced subset of the data, BRF introduces diversity in the ensemble, often leading to improved generalization performance;

- the balancing mechanism reduces bias toward the majority class, ensuring that the minority class is adequately represented and recognized;

- like standard Random Forests, BRF can provide feature importance measures, indicating which features are most influential in making predictions.

The Balanced Random Forest is a robust ensemble method tailored for imbalanced datasets. By ensuring that each tree in the ensemble is trained on a balanced subset of the data, BRF provides a more equitable representation of all classes, leading to more accurate and unbiased predictions.

3.2. XGBoost Model

XGBoost, or eXtreme Gradient Boosting, is an advanced implementation of the gradient boosting algorithm designed for speed and performance [29]. It has gained significant attention in the machine learning community because of its efficiency and effectiveness in handling various predictive modeling tasks.

XGBoost is rooted in the concept of boosting, where weak learners (typically decision trees) are sequentially trained to correct the errors of their predecessors [30]. The final prediction is an ensemble of these weak learners. XGBoost specifically employs the gradient boosting framework, in which new models are trained to predict the residuals or errors of prior models. Mathematically, if y_i is the true label and $\hat{y}_i^{(t-1)}$ is the prediction from the ensemble of trees up to iteration t−1, the next model aims to predict the difference $y_i - \hat{y}_i^{(t-1)}$.

Given a dataset with n samples and m features, the objective function that XGBoost optimizes is:

$$Obj(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{i=1}^{T}\Omega(f_i), \quad (2)$$

where $l(y_i, \hat{y}_i)$ is the loss function that measures the difference between the true label y_i and the predicted label $\hat{y}_i$;

$\Omega(f_i)$ is the regularization term that penalizes the complexity of the model. It helps in preventing overfitting;

T is the total number of trees.

The regularization term is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T}\omega_j^2, \quad (3)$$

where γ and λ are regularization hyperparameters;

T is the number of leaves in the tree;

ω_j represents the score on the leaf j.

Key features:

- XGBoost incorporates L1 (Lasso regression) and L2 (Ridge regression) regularization terms in its objective function, which helps in reducing overfitting;

- XGBoost can automatically handle missing data during training and prediction;

- XGBoost is optimized for parallel processing and can be run on distributed environments, thereby enhancing its speed and efficiency;

XGBoost is a robust machine learning algorithm that leverages the principles of gradient boosting, combined with regularization, to provide a robust, efficient, and accurate predictive modeling tool.

3.3. LightGBM Model

LightGBM, which stands for Light Gradient Boosting Machine, is a gradient boosting framework specifically designed for speed and efficiency [31]. Developed by Microsoft, it introduces several innovations and optimizations that make it faster and more memory-efficient than other gradient boosting frameworks without compromising accuracy.

Similar to other gradient boosting methods, LightGBM builds an ensemble of decision trees sequentially, where each tree corrects the errors of its predecessor. The ensemble prediction is the cumulative sum of the predictions from individual trees. Unlike traditional gradient boosting methods that grow trees level-wise, LightGBM employs a leaf-wise growth strategy. It chooses the leaf with the maximum delta loss to grow, allowing for a reduction in more loss than level-wise algorithms at the expense of increased model complexity.

The objective function of LightGBM is the same as that of XGBoost and is presented in formula (2).

Key features:

- LightGBM uses histogram-based techniques, where continuous feature values are bucketed into discrete bins, thereby reducing the number of split points to consider and speeding up the training process;

- LightGBM bundles mutually exclusive features into a single feature, thereby reducing the dimensionality of the data and further speeding up the training;

- LightGBM samples a subset of the data to handle large datasets based on gradients. It retains instances with large gradients and randomly samples those with small gradients, ensuring both efficiency and accuracy;

LightGBM is an advanced gradient boosting framework that, through its innovative techniques, offers a balance between computational efficiency and predictive accuracy, making it particularly suitable for large datasets or scenarios where computational resources are limited.

3.4. Hyperparameter Optimization and Models Validation

Given the multitude of hyperparameters in gradient boosting frameworks such as XGBoost or LightGBM, an efficient optimization technique is necessary. We used Optuna [32], a state-of-the-art hyperparameter optimization library:

- Bayesian Optimization: Instead of traditional grid or random search methods, Optuna employs Bayesian optimization. By modeling the objective function (in our scenario, the validation error) with Gaussian processes, Optuna updates its beliefs after each trial, guiding the search toward hyperparameters that may minimize the error;

- Pruning: Enhancing efficiency, Optuna's pruning mechanism terminates trials early if a hyperparameter set appears unlikely to surpass the current best.

We particularly fine-tuned the following XGBoost hyperparameters:

- n_estimators: Ranging between 100 and 1000, dictating the amount of boosting rounds;

- learning_rate: Varied logarithmically between 1E-3 to 1E-1, defining the step size at each iteration;

- subsample & colsample_bytree: Both sampled between 0.5 and 1, determining the fraction of data and features used in each boosting round respectively;

- max_depth: An integer between 3 and 9, specifying the maximum tree depth;

- regularization parameters: Including gamma (0 to 1), lambda (log scale between 1E-3 to 10), and alpha (log scale between 1E-3 to 10), which work collectively to prevent overfitting.

For LightGBM, we fine-tuned the following hyperparameters:

a) num_leaves: Dictating the number of leaves in a tree, we experimented with values ranging from 2 to 256, understanding that larger numbers can improve accuracy but might also lead to overfitting;

b) learning_rate: This parameter determines the step size at each boosting iteration. We varied it logarithmically between 1E-3 and 1E-1, ensuring a balance between model convergence speed and accuracy;

c) bagging_fraction & feature_fraction:

- bagging_fraction: Dictates the fraction of data used for each boosting round. We sampled values between 0.4 and 1.0, ensuring both diversity and sufficient data for each round;

- feature_fraction: Represents the proportion of features considered for each boosting round. Similarly, this was adjusted between 0.4 to 1.0, allowing the model to focus on different subsets of features for different rounds;

d) max_depth: An essential parameter that sets the maximum depth of trees in LightGBM, we allowed it to vary between 3 and 9. A deeper tree can capture more intricate patterns but also risks overfitting.

e) regularization parameters:

- lambda_l1: This L1 regularization term was adjusted on a log scale, spanning from 1E-8 to 10. It helps in feature selection and prevent overfitting;

- lambda_l2: Acting as the L2 regularization term, we also fine-tuned it on a log scale ranging from 1E-8 to 10. It reduces model complexity and prevent overfitting.

With the optimal hyperparameters identified, we proceeded to the final training phase:

- Retraining on the Full Dataset: The best hyperparameters from the optimization phase were used to train the models on the entire training dataset. This approach ensures that the model benefits from all available data, leveraging patterns and relationships that might be missed in a subset;

- Validation: After training, the model's performance was evaluated on a separate test set. This provided an unbiased assessment of its ability to generalize to new, unseen data.

This systematic approach, which combines meticulous data preprocessing with the advanced optimization

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

capabilities of Optuna, produces a model proficient at discerning real from fake news. The model's hyperparameters and performance metrics are documented in the subsequent sections of this study.

4. Results

To classify fake news, we underwent rigorous training for three prominent machine learning models: XGBoost, LightGBM, and Balanced Random Forest. Using our dataset, we allocated 80% for training, enabling the models to discern patterns and intricacies within the news articles. The remaining 20% was reserved for testing, providing an unbiased measure of each model's classification prowess.

While designed to address class imbalances, the Balanced Random Forest (BRF) model was particularly adept in our context, even with a balanced dataset. By incorporating a strategy that under-samples during each tree's training phase, the BRF ensured that genuine and fake news were treated equally. Despite the class disparity, this approach fortified the model's resilience against potential overfitting. It ensured that the ensemble's collective output was nuanced and reliable in its classification.

The XGBoost model was instrumental, leveraging its gradient boosting mechanism to focus on previously misclassified instances iteratively. This methodical refinement ensured that the model differentiated between authentic and fabricated news content.

The LightGBM model, employing its distinctive leaf-wise growth strategy, exhibited remarkable efficiency in terms of computational speed and classification accuracy. Given the intricacies and potential class disparities in fake news datasets, this model's emphasis on maximizing loss reduction was particularly beneficial.

Table 1 shows the models' performance evaluation.

To gain a deeper understanding of the classification models' performance and identify areas of potential improvement, we include the confusion matrix in our evaluation metrics. A confusion matrix, often used in machine learning and statistical classification, comprehensively visualizes a classifier's performance. It is a tabular representation that delineates the actual versus predicted classifications. The matrix typically comprises four main components: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These components offer insights into the correct predictions made by the model and the errors it commits. By analyzing the confusion matrix, researchers can derive several performance metrics, such as accuracy, precision, recall, and the F1-score, each offering a unique perspective on the model's efficacy. The matrix is an invaluable tool for understanding the nuances of a classifier's performance beyond mere accuracy, allowing for a more holistic evaluation. Furthermore, the confusion matrix aids in pinpointing specific areas where the model may struggle, such as scenarios leading to false positives or false negatives. This granular insight can guide subsequent refinements of the model, ensuring that it achieves high accuracy and minimizes specific types of misclassifications that may be particularly detrimental in certain applications.

Figure 4 shows the confusion matrix for the balanced random forest. Figure 5 shows the confusion matrix for XGBoost model. Figure 6 shows the confusion matrix for LightGBM model.

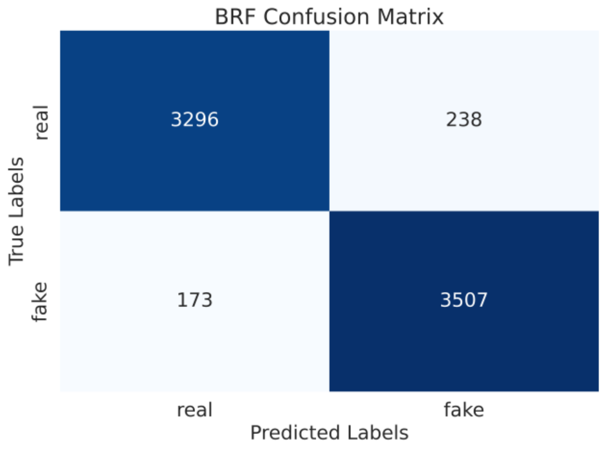

Fig. 4. Confusion matrix for BRF model

There is a noticeable number of false positives and false negatives in the BRF model despite it achieving a commendable classification in most instances. This suggests that while the model is generally effective, specific instances or patterns might be misinterpreted, leading to these misclassifications.

The XGBoost model, on the other hand, demonstrates a more refined classification with fewer false positives and false negatives than the BRF model. This indicates a higher precision and recall, suggesting that the model effectively identifies fake news and ensures that genuine news is not mistakenly classified as fake.

The LightGBM model exhibited the best performance, with the least false positives and false negatives. This implies an optimal balance between precision and recall, ensuring that the model is accurate in its predictions and sensitive to the nuances of the data.

While all three models exhibit robust classification capabilities, the LightGBM model stands out in terms of its precision and recall, making it the most effective in distinguishing between real and fake news in this dataset.

Table 1

Models' performance metrics

Metric	**BRF**	**XGBoost**	**LightGBM**
F1-score (test)	0.944557408	0.968226763	0.97741164
Accuracy (test)	0.943300756	0.966588105	0.976431443
Recall (test)	0.949577542	0.984186545	0.985794693
Precision (test)	0.939590076	0.952776336	0.96916996
Loss (test)	2.043647915	0.095977571	0.123440667
F1-score (train)	1	0.988855286	1
Accuracy (train)	1	0.98844161	1
Recall (train)	1	0.996497727	1
Precision (train)	1	0.981329177	1
Loss (train)	2.22E-16	0.055085406	1.68E-06

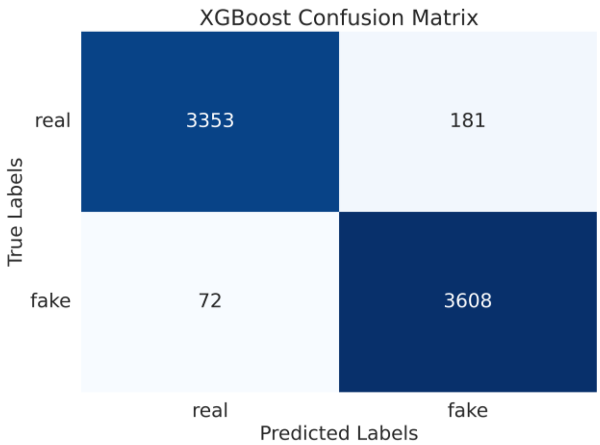

Fig. 5. Confusion matrix for XGBoost model

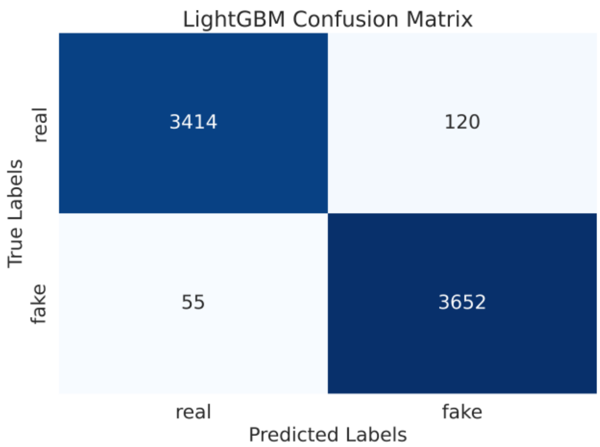

Fig. 6. Confusion matrix for LightGBM model

We incorporate the Receiver Operating Characteristic (ROC) curve in our evaluation framework to elucidate the trade-offs between sensitivity and specificity across various threshold values. The ROC curve is a graphical representation that plots the true positive rate (sensitivity) against the false positive rate (1-specificity) for a binary classifier system as its discrimination threshold varies. The curve provides insights into the classifier's performance across all threshold levels, with the area under the curve (AUC) serving as a singular metric to quantify the model's ability to distinguish between the positive and negative classes. A model with perfect discriminatory power will have an AUC of 1, whereas a model with no discriminatory power will have an AUC of 0.5. By including the ROC curve, we can assess the model's robustness and capacity to balance detecting true positives and avoiding false positives.

Figure 7 shows ROC curve for the balanced random forest model.

Figure 8 shows ROC curve for XGBoost model.

Figure 9 shows ROC curve for LightGBM model.

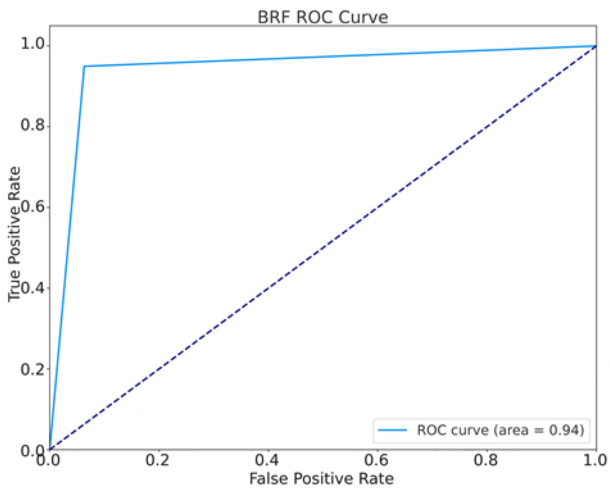

Fig. 7. ROC curve for BRF model

We observed distinct performance metrics for each model. The F1-scores on the test set were notably high for all models, with LightGBM achieving the highest at approximately 0.977, followed closely by XGBoost at around 0.968 and BRF at 0.945. This indicates a harmonious balance between precision and recall for each model, especially for LightGBM. The accuracy metrics

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

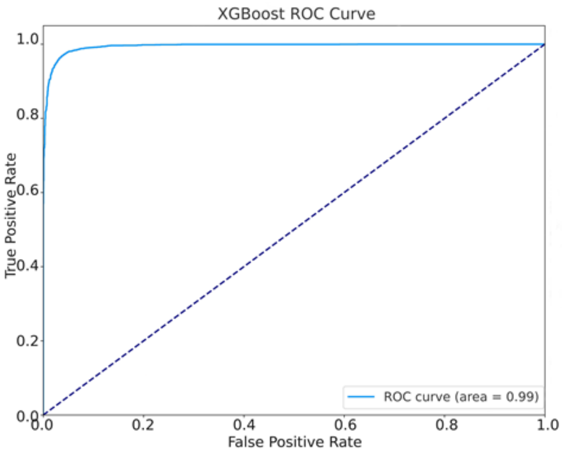

Fig. 8. ROC curve for XGBoost model

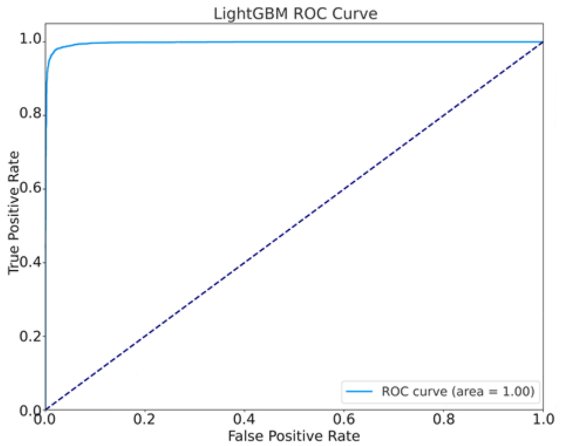

Fig. 9. ROC curve for LightGBM model

further corroborate this, with values nearing or exceeding 94% for all models on the test set. Interestingly, the recall for XGBoost and LightGBM was exceptionally high, suggesting that these models were adept at identifying most fake news instances. However, in terms of precision, LightGBM slightly outperformed XGBoost, indicating fewer false positives. The loss values on the test set provide insights into the models' deviations from the actual outcomes, with BRF having a notably higher loss than the other two. When examining the training metrics, both LightGBM and BRF achieved perfect scores across F1, accuracy, recall, and precision, suggesting a potential overfitting to the training data. In contrast, XGBoost, while still exhibiting high metrics, showed a slight divergence from perfection, which might indicate a more generalized model. Overall, these results underscore the efficacy of the models in fake news classification, with each model presenting its unique strengths and areas for further exploration.

5. Discussion

The findings have provided noteworthy insights into the exploration of fake news classification using ensemble machine learning models. The overarching aim of this research was to harness the power of ensemble techniques to enhance the accuracy and reliability of fake news detection, which is a pressing concern in today's digital age.

Table 2 shows a comparison of the performance of the developed models and other studies that used the same dataset.

The performance metrics derived from the confusion matrix and the ROC curve offer a comprehensive understanding of each model's strengths and potential areas of improvement. LightGBM demonstrated remarkable efficiency, especially when considering its speed and accuracy balance. XGBoost showcased its prowess in iterative refinement, emphasizing misclassified instances from previous iterations. The Balanced Random Forest presented a robust classification mechanism that reduced inherent biases.

The XGBoost model exhibits a commendable F1-score of approximately 0.968 on the test set, suggesting a harmonious balance between precision and recall. The accuracy of 0.966 further underscores its capability to correctly classify news articles. Notably, its recall of 0.984 indicates a strong ability to identify most fake news instances. However, a precision of 0.953 suggests a slightly higher rate of false positives than the other models.

The LightGBM model exhibited the highest F1-score and accuracy on the test set, exceeding 0.976. This suggests that LightGBM classifies most of the instances correctly and maintains an optimal balance between precision and recall. The recall value 0.986 is impressive, implying that the model rarely misses any fake news instances. The precision of 0.969 further accentuates its ability to minimize false positives.

Finally, the BRF model, designed to address class imbalances, yields an F1-score of 0.945 on the test set. Although this is slightly lower than the other two models, it still indicates a robust performance. The accuracy of 0.943 and recall of 0.950 further support its efficacy. However, the precision of 0.940, although commendable, suggests a slightly higher propensity for false positives. The loss value for the test set is notably higher than that of the other models, which might indicate areas for optimization.

While all three models demonstrated strong capabilities in fake news classification, LightGBM was the top performer in this evaluation.

Models from the literature such as KNN, SVM, and NB [27] have lower F1-Scores, ranging from 89.24% to 96.56%. Notably, the SVM model from [27] has a recall metric of 98.61%, which is comparable to the proposed models, indicating its high sensitivity. However, its precision is slightly lower at 94.60%.

Table 2

Models' performance metrics

Metric	F1-score (%)	Accuracy (%)	Precision (%)	Recall
LightGBM (proposed)	97.74	97.64	96.92	98.58
XGBoost (proposed)	96.82	96.66	95.28	98.42
SVM [27]	96.56	96.73	94.60	98.61
Adaboost [27]	95.02	95.32	91.81	98.46
Bagging [27]	95.00	95.31	91.78	98.46
BRF (proposed)	94.46	94.33	93.96	94.96
NB [27]	91.85	92.12	91.45	92.25
KNN [27]	89.78	90.16	89.02	90.55
DT [27]	89.24	89.92	86.10	92.62
GaussianNB [33]	-	74.00	-	-
BernoulliNB [33]	-	86.00	-	-
MLPClassifier [33]	-	92.00	-	-
Random Forest [33]	-	89.00	-	-
XGBClassifier [33]	-	94.00	-	-
N-Gram with TF-IDF and LSTM [34]	-	96.00	-	-
N-Gram with TF-IDF and BERT [34]	-	96.8	-	-

Models from source [33], including GaussianNB, BernoulliNB, and MLPClassifier, provide only accuracy metrics, making a comprehensive comparison challenging. Nevertheless, their accuracy rates, ranging from 74.00% to 92.00%, are generally lower than those of the proposed models.

Lastly, the source [34] models, which use N-Gram with TF-IDF along with LSTM and BERT has accuracy metrics of 96.00% and 96.8%, respectively. These figures are competitive, but a holistic assessment is limited without additional metrics such as precision and recall.

In summary, the proposed models, especially LightGBM and XGBoost, demonstrate superior performance in fake news classification compared with several models from the existing literature. Their high precision and recall metrics underscore their effectiveness in identifying and correctly classifying fake news instances.

Understanding the intricacies of a model's decision-making process is crucial for ensuring its transparency, reliability, and interpretability. This becomes especially vital in applications such as fake news detection, where the stakes are high and the implications of misclassification can be profound. We analyzed token importance graphs for models XGBoost and LightGBM to identify specific words or phrases that play a pivotal role in the classification process. These graphs not only offer a visual representation of the significance of each token but also provide a deeper understanding of the patterns and characteristics that the models deem crucial for distinguishing between real and fake news. In essence, token importance graphs justify the model's decisions, allowing researchers, stakeholders, and end-users to gain confidence in the model's outputs and comprehend the underlying linguistic cues that drive its determinations. Moreover, by discerning these key linguistic indicators, we can refine our models further, ensuring that they are attuned to the most salient features of the data, thereby enhancing their robustness and precision in real-world scenarios.

Figure 10 shows token importance for XGBoost model. Figure 11 shows token importance for LightGBM model.

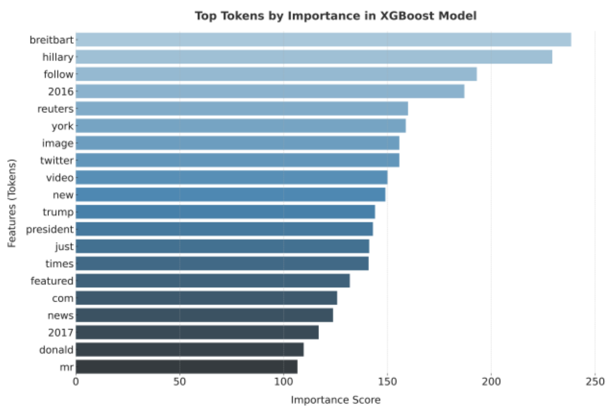

Fig. 10. Token importance for XGBoost model

In fake news detection, understanding the significance of specific tokens or features is paramount to deciphering the patterns that models identify as indicative of misinformation. Analyzing the top features by importance for the XGBoost and LightGBM models offers a window into the linguistic patterns and keywords these models deem crucial for distinguishing between real and fake news.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

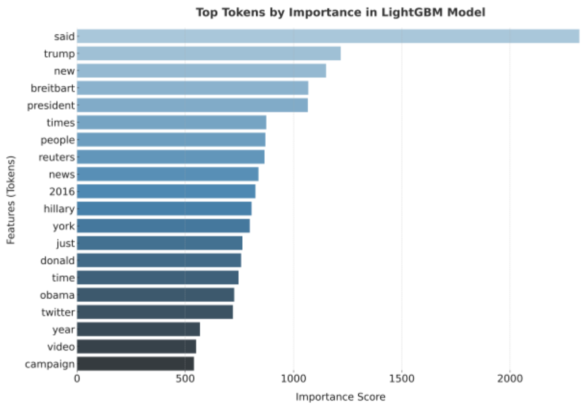

Fig. 11. Token importance for LightGBM model

For the XGBoost model, terms such as “breitbart”, “hillary”, and “follow” emerged as the most influential features. The prominence of terms such as “breitbart” and “hillary” suggests that specific news sources or political figures might be more frequently associated with specific types of news, whether genuine or deceptive. Other terms such as “2016”, “reuters”, and “york” might be indicative of the temporal or geographical context of the news articles, while terms such as “image”, “twitter”, and “video” could reflect the multimedia nature or dissemination platforms of the news.

On the other hand, the LightGBM model highlighted “said” as its most influential feature, followed by terms such as “trump” and “new”. The high importance of the term “said” might indicate a specific linguistic structure or reporting style that the model associates with either genuine or fake news. Similarly, terms such as “president”, “times”, and “people” might be indicative of the thematic content of the news articles. The recurrence of specific terms, such as “trump”, “hillary”, “reuters”, and “york”, in both models underscores their potential significance in news classification. In essence, these feature importance scores provide a window into the linguistic and thematic elements that both XGBoost and LightGBM models deem crucial in fake news classification. They highlight the nuanced interplay of sources, political figures, multimedia elements, and common linguistic constructs in the intricate landscape of misinformation detection.

Conclusions

In the contemporary digital age, the proliferation of fake news poses significant challenges to the integrity of information dissemination, with potential ramifications on public opinion, policy decisions, and societal trust. This study embarked on a journey to address this pressing issue by harnessing the power of machine learning, specifically focusing on developing and evaluating ensemble models for fake news classification. Through rigorous experimentation, the proposed models, namely LightGBM, XGBoost, and BRF, demonstrated commendable efficacy in discerning genuine news from fabricated narratives. Their performance, benchmarked against several models from the existing literature, underscores their potential as robust tools in the fight against misinformation.

The study’s findings contribute to the growing body of knowledge on fake news detection and pave the way for future research endeavors. By leveraging advanced machine learning techniques and continuously refining model architectures, the academic and tech communities can collaboratively create a more trustworthy digital information ecosystem. As the landscape of fake news evolves, so must our strategies and tools to counteract its spread. This research affirms the possibilities of harnessing technology for the greater good, ensuring that truth and authenticity remain at the forefront of our shared digital narratives.

The scientific novelty of this paper lies in its integrative approach to fake news classification, which leverages the strengths of ensemble machine learning models. While previous studies have explored individual algorithms for misinformation detection, this research uniquely combines the capabilities of LightGBM, XGBoost, and BRF, offering a more holistic and robust solution. Furthermore, this study introduces a refined feature importance analysis, shedding light on the pivotal tokens and linguistic patterns that predominantly influence the classification process. This research provides a comprehensive performance benchmark by juxtaposing the proposed models against a diverse array of existing methodologies, enriching the discourse on fake news detection. This innovative exploration advances the state-of-the-art in misinformation classification and sets a precedent for future investigations in the realm of digital information integrity.

The practical novelty of this study is manifested in its actionable insights for real-world applications in the domain of fake news detection. By harnessing the combined strengths of LightGBM, XGBoost, and BRF, this study offers a robust ensemble model that can be readily deployed in digital platforms to enhance the accuracy of misinformation identification. The detailed feature importance analysis provides a roadmap for content curators and platform developers, enabling them to prioritize specific linguistic cues and patterns when designing misinformation filters. Furthermore, comparative performance evaluation against established methodologies equips stakeholders with a clear understanding of the efficacy of various approaches, facilitating informed decision-making. This research, therefore, not only contributes theoretically but also paves the way for tangible advancements in the tools and technologies employed to combat the spread of fake news in today’s digital landscape.

Future research directions stemming from this study can be multifaceted and expansive. First, a compelling case exists for exploring the integration of more advanced natural language processing techniques, such as transformer-based models like BERT, to refine the fake news classification process further. With their deep contextual understanding, these models might offer nuanced insights into the subtleties of misinformation. Second, while this study focused on textual data, the proliferation of multimedia misinformation, including manipulated images and videos, necessitates exploring multi-modal models that can concurrently analyze text, image, and video data. Another promising avenue is the investigation of real-time fake news detection, which would be pivotal in curtailing the rapid spread of misinformation on platforms with high user engagement.

Additionally, understanding the psychological and sociological underpinnings of why specific fake news stories gain traction can complement technical solutions. This would involve a multidisciplinary approach, merging computational methods with social sciences. Finally, as misinformation tactics evolve, there is a continuous need to update and expand datasets, ensuring that models are trained on contemporary examples and not blindsided by novel misinformation strategies.

Contributions of authors: conceptualization – **Halyna Padalko, Vasyl Chomko, Dmytro Chumachenko**; methodology – **Halyna Padalko, Vasyl Chomko**; formulation of tasks – **Halyna Padalko, Sergiy Yakovlev**; analysis – **Halyna Padalko, Vasyl Chomko, Dmytro Chumachenko**; development of model – **Halyna Padalko, Vasyl Chomko**; verification – **Halyna Padalko, Vasyl Chomko, Dmytro Chumachenko**; analysis of results – **Halyna Padalko, Dmytro Chumachenko**; visualization – **Halyna Padalko, Vasyl Chomko**; writing – original draft preparation – **Halyna Padalko, Dmytro Chumachenko**; writing – review and editing – **Vasyl Chomko, Sergiy Yakovlev**.

All authors have read and agreed with the published version of this manuscript.

References

1. Vraga, E. K., & Bode, L. Defining Misinformation and Understanding its Bounded Nature: Using Expertise and Evidence for Describing Misinformation. *Political Communication*, 2020, vol. 37, iss. 1, pp.136–144. DOI: 10.1080/10584609.2020.1716500.

2. Ó Fathaigh, R., Helberger, N., & Appelman, N. The perils of legally defining disinformation. *Internet Policy Review*, 2021, vol. 10, iss. 4. 26 p. DOI: 10.14763/2021.4.1584.

3. van der Linden, S. Misinformation: susceptibility, spread, and interventions to immunize the public. *Nature Medicine*, 2022, vol. 28, iss. 3, pp. 460-467. DOI: 10.1038/s41591-022-01713-6.

4. Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. The Science of Fake News. *Science*, 2018, vol. 359, iss. 6380, pp. 1094-1096. DOI: 10.1126/science.aao2998.

5. Zakharchenko, A., Peráček, T., Fedushko, S., Syerov, Y., & Trach, O. When Fact-Checking and ‘BBC Standards’ Are Helpless: ‘Fake Newsworthy Event’ Manipulation and the Reaction of the ‘High-Quality Media’ on It. *Sustainability*, 2021, vol. 13, iss. 2, article no. 573. DOI: 10.3390/su13020573.

6. Pennycook, G., & Rand, D. G. The psychology of fake news. *Trends in Cognitive Sciences*, 2021, vol. 25, iss. 5, pp. 388-402. DOI: 10.1016/j.tics.2021.02.007.

7. Kim, B., Xiong, A., Lee, D., & Han, K. A systematic review on fake news research through the lens of news creation and consumption: Research efforts, challenges, and future directions. *PLOS ONE*, 2021, vol. 16, iss. 12, article no. e0260080. DOI: 10.1371/journal.pone.0260080.

8. Thompson, R. C., Joseph, S., & Adeliyi, T. T. A Systematic Literature Review and Meta-Analysis of Studies on Online Fake News Detection. *Information*, 2022, vol. 13, iss. 11, article no. 527. DOI: 10.3390/info13110527.

9. Santos, F. C. C. Artificial Intelligence in Automated Detection of Disinformation: A Thematic Analysis. *Journalism and Media*, 2023, vol. 4, iss. 2, pp. 679-687. DOI: 10.3390/journalmedia4020043.

10. Wawrzynski, T. Artificial intelligence and cyberculture. *Radioelectronic And Computer Systems*, 2020, no. 3, pp. 20-26. DOI: 10.32620/reks.2020.3.02.

11. Badi, H., Badi, I., Moutaouakil, K. E., Khamjane, A., & Bahri, A. Sentiment analysis and prediction of polarity vaccines based on Twitter data using deep NLP techniques. *Radioelectronic and Computer Systems*, 2022, no. 4, pp. 19-29. DOI: 10.32620/reks.2022.4.02.

12. Lai, C.-M., Chen, M.-H., Kristiani, E., Verma, V. K., & Yang, C.-T. Fake News Classification Based on Content Level Features. *Applied Sciences*, 2022, vol. 12, iss. 3, article no. 1116. DOI: 10.3390/app12031116.

13. Kumar, S., & Arora, B. A Review of Fake News Detection Using Machine Learning Techniques. *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2021, pp. 1-8. DOI: 10.1109/ICESC51422.2021.9532796.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

14. Capuano, N., Fenza, G., Loia, V., & Nota, F. D. Content-Based Fake News Detection With Machine and Deep Learning: a Systematic Review. *Neurocomputing*, 2023, vol. 530, pp. 91-103. DOI: 10.1016/j.neucom.2023.02.005.

15. Babaiev, V. M., Kadykova, I. M., Husieva, Yu. Yu., & Chumachenko, I. V. The method of adaptation of a project-oriented organization's strategy to exogenous changes. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, 2017, vol. 2, pp. 134-140.

16. Yakovlev, S., Bazilevych, K., Chumachenko, D., Chumachenko, T., Hulianytskyi, L., Meniailov, I., & Tkachenko, A. The Concept of Developing a Decision Support System for the Epidemic Morbidity Control. *CEUR Workshop Proceedings*, 2020, vol. 2753, pp. 265-274.

17. Akram, H., & Shahzad, K. Ensembling Machine Learning Models for Urdu Fake News Detection. *CEUR Workshop Proceedings*, 2022, vol. 3159, pp. 1142-1149.

18. Tian, Z., & Baskiyar, S. Fake News Detection using Machine Learning with Feature Selection. *Proceedings of the 2021 6th International Conference on Computing, Communication and Security, ICCCS 2021*, 2021, pp. 1-6. DOI: 10.1109/ICCCS51487.2021.9776346.

19. Choudhury, D., & Acharjee, T. A novel approach to fake news detection in social networks using genetic algorithm applying machine learning classifiers. *Multimedia Tools and Applications*, 2022, vol. 82, iss. 6. DOI: 10.1007/s11042-022-12788-1.

20. Fahad, N., Goh, K.-S., Hossen, I., Shopnil, K. M. S., Mitu, I. J., Alif, A., & Connie, T. Stand up Against Bad Intended News: An Approach to Detect Fake News using Machine Learning. *Emerging science journal*, 2023, vol. 7, iss. 4, pp. 1247-1259. DOI: 10.28991/esj-2023-07-04-015.

21. Park, M., & Chai, S. Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques. *IEEE Access*, 2023, vol. 11, pp. 71517-71527. DOI: 10.1109/ACCESS.2023.3294613.

22. Salh, D., & Nabi, R. M. Kurdish Fake News Detection Based on Machine Learning Approaches. *Passer journal of basic and applied sciences*, 2023, vol. 5, iss. 2, pp. 262-271. DOI: 10.24271/psr.2023.380132.1226.

23. Kumar Dutta, A., Qureshi, B., Albagory, Y., Alsanea, M., Al Faraj, M., & Rahaman Wahab Sait, A. Optimal Weighted Extreme Learning Machine for Cybersecurity Fake News Classification. *Computer Systems Science and Engineering*, 2023, vol. 44, iss. 3, pp. 2395-2409. DOI: 10.32604/csse.2023.027502.

24. Tohabar, Md. Y., Nasrah, N., & Samir, A. M. Bengali Fake News Detection Using Machine Learning and Effectiveness of Sentiment as a Feature. *2021 Joint 10th International Conference on Informatics, Electronics & Vision (ICIEV) and 2021 5th International Conference on Imaging, Vision & Pattern Recognition (icIVPR)* 2021. DOI: 10.1109/icievicivpr52578.2021.9564138.

25. Fifita, F., Smith, J., Hanzsek-Brill, M. B., Li, X., & Zhou, M. Machine Learning-Based Identifications of COVID-19 Fake News Using Biomedical Information Extraction. *Big Data and Cognitive Computing*, 2023, vol. 7, iss. 1, article no. 46. DOI: 10.3390/bdcc7010046.

26. Yousif, S. A., & Jehad, R. Classification of Covid-19 fake news using machine learning algorithms. *AIP Conference Proceedings*, 2022, vol. 2483, iss. 1, article no. 070007. DOI: 10.1063/5.0117133.

27. Verma, P. K., Agrawal, P., Amorim, I., & Prodan, R. WELFake: Word Embedding Over Linguistic Features for Fake News Detection. *IEEE Transactions on Computational Social Systems*, 2021, vol. 8, iss. 4, pp. 881-893. DOI: 10.1109/tcss.2021.3068519.

28. Agusta, Z. P., & Adiwijaya, A. Modified balanced random forest for improving imbalanced data prediction. *International Journal of Advances in Intelligent Informatics*, 2019, vol. 5, iss. 1, pp. 58-65. DOI: 10.26555/ijain.v5i1.255.

29. Chen, T., & Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 2016, pp. 785-794. DOI: 10.1145/2939672.2939785.

30. Natekin, A., & Knoll, A. Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 2013, vol. 7, article no. 21. DOI: 10.3389/fnbot.2013.00021.

31. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 3149-3157.

32. *Optuna - A hyperparameter optimization framework*. Optuna, 2023. Available at: https://optuna.org (Accessed 26 Apr. 2023).

33. Gupta, A., Batla, A., Kumar, C., & Jain, G. Comparative Analysis of Machine Learning Models for Fake News Classification. *3rd International Conference on Intelligent Technologies (CONIT)*, 2023, pp. 1-5. DOI: 10.1109/CONIT59222.2023.10205870.

34. Kausar, N., AliKhan, A., & Sattar, M. Towards better representation learning using hybrid deep learning model for fake news detection. *Social Network Analysis and Mining*, 2022, vol. 12, iss. 1. DOI: 10.1007/s13278-022-00986-6.

Received 14.05.2023, Accepted 20.11.2023

АНСАМБЛЕВІ ПІДХОДИ МАШИННОГО НАВЧАННЯ ДЛЯ КЛАСИФІКАЦІЇ ФЕЙКОВИХ НОВИН

Галина Падалко, Василь Чомко, Сергій Яковлев, Дмитро Чумаченко

У сучасному цифровому середовищі поширення фейкових новин стало значущим викликом з далекосяжними наслідками для осіб, установ та суспільств. Швидке поширення оманливої інформації підриває авторитет справжніх новинних джерел і створює загрози свідомому прийняттю рішень, громадській довірі та демократичним процесам. Враховуючи глибоку актуальність і невідкладність вирішення цієї проблеми, дослідження розглядає використання можливостей машинного навчання для боротьби з небезпекою фейкових новин. **Метою** дослідження є розробка ансамблевих моделей машинного навчання для класифікації фейкових новин. **Об’єктом дослідження** є поширення фейкових новин. **Предметом дослідження** є методи машинного навчання для класифікації дезінформації. **Методи:** були застосовані три методи машинного навчання: LightGBM, XGBoost та Balanced Random Forest (BRF). Кожен з цих алгоритмів був навчений на відкритому наборі даних, який включав різноманітний спектр новинних статей, забезпечуючи широке представлення лінгвістичних шаблонів та стилів. Запропонований підхід також акцентував уваги на важливості токенів. Для підвищення точності та надійності моделей використовувались конкретні токени, які виявили високий ступінь впливу на результати класифікації. Емпіричні **результати** показали високу продуктивність моделей. Модель LightGBM виявилася найкращою серед трьох, показавши F1-score у 97,74% та точність у 97,64%. Варто відзначити, що всі три запропонованих нами моделі послідовно перевершували кілька існуючих моделей, раніше задокументованих у науковій літературі. Цей порівняльний аналіз підкреслює ефективність та перевагу нашого ансамблевого підходу. **Висновок:** це дослідження пропонує надійне, інноваційне та масштабоване рішення для актуального виклику виявлення фейкових новин. Використовуючи можливості передових технік машинного навчання, отримані результати відкривають шлях для підвищення інтегральності та достовірності інформації у цифровому світі, тим самим захищаючи громадську довіру та сприяючи обізнаному діалогу.

Ключові слова: фейкові новини; класифікація; місінформація; дезінформація; збалансований випадковий ліс; XGBoost; LightGBM; WELFake.

Падалко Галина Анатоліївна – здобувачка каф. математичного моделювання та штучного інтелекту, Національний аерокосмічний університет ім. М. Є. Жуковського «Харківський авіаційний інститут», Харків, Україна; запрошена дослідниця лабораторії повсюдних технологій охорони здоров’я, Університет Ватерлоо, Ватерлоо, Канада; здобувачка, Школа міжнародних відносин Balsillie, Ватерлоо, Канада.

Чомко Василь Дмитрович – здобувач каф. системного проєктування, Університет Ватерлоо, Ватерлоо, Канада.

Яковлев Сергій Всеволодович – д-р фіз.-мат. наук, проф., проф. інституту інформаційних технологій, Університет технологій Лодзю, Лодзь, Польща.

Чумаченко Дмитро Ігорович – канд. техн. наук, доц., доц. каф. математичного моделювання та штучного інтелекту, Національний аерокосмічний університет ім. М. Є. Жуковського «Харківський авіаційний інститут», Харків, Україна.

Halyna Padalko – Applicant of Mathematical Modelling and Artificial Intelligence Department, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine; Guest Researcher of Ubiquitous Health Technology Lab, University of Waterloo, Waterloo, Canada; Applicant, Balsillie School of International Affairs, Waterloo, Canada,
e-mail: galinapadalko95@gmail.com, ORCID: 0000-0001-6014-1065.

Vasyl Chomko – Ongoing MSc in System Design Engineering, University of Waterloo, Waterloo, Canada,
e-mail: vchomko@uwaterloo.ca, ORCID: 0009-0009-4419-6651.

Sergiy Yakovlev – Dr. Sc. in Physics and Mathematics, Professor of Institute of Information Technology, Lodz University of Technology, Lodz, Poland,
e-mail: svsyak7@gmail.com, ORCID: 0000-0003-1707-843X.

Dmytro Chumachenko – PhD in Artificial Intelligence, Associate Professor of Mathematical Modelling and Artificial Intelligence Department, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine,
e-mail: dichumachenko@gmail.com, ORCID: 0000-0003-2623-3294.