Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

Majority Voting Technique for Fake News Detection

Howaida Abdul Hadi Al ibraheemi[1] Mohammed Jabardi [2]
[1] computer science College of Education University of Kufa, Najaf, 54003, Iraq. edu@uokufa.edu.iq
[2] faculty of education, university of Kufa, Najaf, 54003 Iraq. edu@uokufa.edu.iq

[1] Howaidaa.alibraheemi@student.uokufa.edu.iq

https://doi.org/10.46649/fjiece.v3.2.34a.13.6.2024

Abstract. *A considerable number of people get their news from electronic sources. As there was a great increase in the usage of these platforms, more and more people are consuming the news from social media sources along with other websites. Thus, a variety of sites and sources has grown exponentially, allowing for easy and fast distribution of false information. Such deliberately generated lies fixing to deceive both the person and society are referred to as fake news Since the media plays an essential role in presenting fake information that alters public opinion and makes members of society take responsibility for unsupported facts. Currently the widespread of social media has worsened the level of fake news dissemination. To assess our methodology, we used popular machine learning classifiers such as Support Vector Machine (SVM), Random forest (RF), Logistic Regression(LR), Naive Bayes(NB), Gradient Boosting, AdaBoost, K-nearest neighbor (KNN), Decision tree (DT), and Extreme Gradient Boosting (XGBoost). We constructed a multi-model false news detection system using the Majority Voting approach, using the previously discussed classifiers to provide more accurate findings. Our strategy achieved an accuracy, of 0.9977%, recall of 0.998%, precision of 0.997%, and 0.997% of F1-measure, with a loss of 0.0022 according to the trial data. The assessment reveals that, in comparison to individual learning strategies, the Majority Voting approach produced more outcomes that were deemed acceptable.*

Keywords: *Fake news detection, natural language processing (NLP), Machine learning, TF_IDF, Majority Voting.*

1. INTRODUCTION

Misleading information is now widely disseminated due to the rise in popularity and speed of Internet use brought about by the advancement of information technology. News is any information intended to notify the public about local happenings that might affect their social or personal lives [1]. In recent years, online social media has become a well-liked medium for distributing news for political, economic, and entertainment reasons [2]. This component demonstrated both beneficial and detrimental effects [3]. People disseminate false information under the guise of legitimate news to amuse themselves and make money. The fake news circulating on Facebook in recent years has greatly influenced the 2016 US presidential election campaign [1]. In light of this tragedy, many businesses and research institutions are concentrating on comprehending the phenomenon and reducing fake news. Many scholars who examine false news and social influence have used terms like "rumors," "misinformation," and "fake news." Politics, the economy, and public opinion may all suffer from fake news. One well-known example of false news is the assertion that Barack Obama was hurt in the explosion that destroyed equities worth $130 billion [4]. Several false reports about the COVID-19 epidemic have sowed mistrust and concern, which has caused social media systems to crash. The most unsettling thing on the Internet is people's lack of confidence in one another and

bogus news. Classifying news, particularly fake news, is a complex task that has engaged the attention of many scholars. While some have approached it as a binary classification (i.e., real or fake) [5], others have

Found this approach problematic and have suggested multiclass classification [6], regression, or clustering [7] as alternative methods. In their research, these scholars have employed various techniques to identify.

And classify false information, each with its own criteria. The ever-growing rate of development and variety of approaches in fake news detection still comes with open issues, for instance, in how to enhance the models' transferability across languages and data and address the computational cost of the new and even more sophisticated models. Against this background, the present study provides important insights and outcomes regarding fake news detection with the help of ML algorithms.

We propose to detect bogus news items using a majority vote method. We've utilized many textual characteristics from both actual and phony news. We utilized a dataset of fictitious news from the Kaggle website. 43 MB of news content that is publicly accessible Out of the articles, 21417 are authentic (labeled as 1) and 123481 are false (labeled as 0). Also, We evaluated our method using popular machine learning classifiers such as Support Vector Machine (SVM), Random Forest (RF), XGBoost (XGB), Decision Tree (DT), Logistic Regression (LR), AdaBoost (AB), and Naive Bayes (NB). AdaBoost, Gradient Boosting, and K-nearest neighbor (KNN). For more accurate findings, Using the Majority Voting method and the previously stated classifiers, we created a multi-model fake news detection system. The trial results showed that our proposed technique yielded a loss of 0.0022 and an accuracy of 0.9977%, precision of 0.998%, recall of 0.998%, and F1-measure of 0.997%. The evaluation confirms that most of the The voting technique yielded more results that were judged acceptable than the individual learning approach. The following are the main contributions of this paper:

- A majority vote serves as the foundation for our method of spotting fake news. It was shown that the majority voting technique provided more results that were considered acceptable when compared to individual learning strategies.
- We have more precisely categorized the news as genuine or fraudulent by using a range of language traits.
- We evaluated the effectiveness of state-of-the-art machine learning classifiers using Kaggle's publicly available Fake News dataset. These classifiers included Decision Tree, Logistic Regression, XGBoost, Random Forest, AdaBoost, SVM, Naive Bayes, KNN, and Gradient Boosting. The classifier's performance is evaluated using recall, precision, detection accuracy, and the F1 measure.
- We created a multi-model learning system to detect fake news, and the results indicate a gain in accuracy: 0.997% F1-measure with 0.0022 loss, 0.998% precision, 0.9988% recall, and 0.9977% accuracy.

The remainder of this paper is organized as follows: Section 2 outlines the literature review, Section 3 outlines the method proposed, Section 4 highlights the results, and discussion Section 5 gives the conclusion.

2. LITERATURE REVIEW

Research into fake news detection has been ongoing in recent years. However, most of this research focuses on analyzing and detecting fake information recognition from Internet networks and articles. So, the authors proposed numerous strategies.where use Patel et al.[8] Used (SVM) with Natural Language Processing (NLP) to detect fake news. Where fake news data from Kaggle with true articles comprising 21,418 while fake articles made up 23,503. The demonstrated comparisons that the method SVM was the most effective yielding an average of 94.93%. Although the model proves very efficient, it is computationally intensive and, hence possibly not suitable for resource-limited environments.

Also Wotaifi et al [9]. Applied a technique that was a hybrid of modified Random Forest with Fuzzy logic for the better forecast of fake news in Arabic. Further to the data from Twitter, the researchers gathered. Additional information thus increasing the size of snap-shot to 3000 tweets. For feature extraction, they used the TF-IDF techniques and to determine the important features they employed fuzzy logic. The analysis of the results indicated that the proposed enhanced model yielded an accuracy of 89.5 %.

Granik and Mesyura.[10] proposed a simple method for fake news detection by a Naive Bayes Technique, gathered data from the Facebook news post content, which comprised 1771 pieces. Next to this, used the Bag of Words and TF-IDF methods for feature extraction. The findings achieved an accuracy of 75.4%. Despite all these, the model is quite simple yet its outcome is acceptable though this may be enhanced in many ways

Additionally, Lyu and Lo. [11] Applied Decision Tree for the identification of fake news. Collected articles from Gossipcop and PolitiFact amounted to 24,556 but after data pre-processing where the data was reduced by features acquired by doc2vec such as URL, text, author, and title, they were left with 14,641. The proposed model got a mean accuracy of 95.54%.To identify fake news Johnson et al. [12] implemented a Logistic Regression with the help of NLP methods. Used a Kaggle dataset that is a collection of 20,000 news articles categorized as real or false. 50 features are being extracted from text data using the TF-IDF method of feature extraction. Thus, the model yielded an accuracy of 97.90%, precision of 96.59%, recall of 99.32%, and score of 97.94%. This approach was superior to KKN and Naïve Bayes.Kesarwani et al.[13] introduced a simple method to identify fake news on social media by developing (a KNN) classifier. Data were obtained from the Buzz Feed News organization which was 22,82,000 posts with social features such as count of sharing and comment count. With the preprocessing and feature extraction they got a classification accuracy of 79% in their KNN model.

Selva Birunda and Kanniga Devi . [14] proposed a mechanism for fake news classification using Gradient Boosting. was employed Kaggle's dataset of 2050 articles, in which they used TF-IDF and site_url analysis to feature extract. Accuracy has achieved a high level of 99.5%.

Haumahu et al.[15] utilized(XGBoost) in making fake news classifications. Where employed 500 Indonesian news articles, which one's genuine and which one's a hoax. The model was trained and tested after preprocessing of the data and feature extraction using TF-IDF to obtain an accuracy of 89 %, precision of 90 %, and recall of 80 %. Holla and Kavitha.[16] proposed a fake news detection model with hybrid TF-IDF and AdaBoost. On the WELFake dataset containing 72,134 articles, they obtained 98.98% accuracy, 99.00% precision, and 99.00% recall, while the F1 score was 99.00%. As shown in the table1

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

Table 1. Summary of Related Work of Fake News Detection

author	(Used or methods) features	Extraction Techniques	Dataset	Classification Method	Performance Metrics
Patel et al(2021) [8]	NLP features (tokenization, stemming, stop	Data preprocessing techniques (cleaning, tokenization, stemming, stop word removal, feature	Kaggle dataset (21418 true news, 23503 fake news)	SVM (Support Vector Machine)	Accuracy: 94.93%, Precision: 93.98%, Recall: 96.04%, F1 Score: 94.99%
Wotaifi et al. (2022) [9]	Text features (TF-IDF), User features	Fuzzy model, TF-IDF	Twitter dataset (expanded from 1862 to 3000 tweets)	Modified Random Forest	Accuracy: 0.895
Granik and Mesyura (2017) [10]	Text feature	ag of Words (BoW), TF-IDF	Facebook posts dataset (1771 articles)	Naive Bayes	Accuracy: 75.40%, Precision: 0.71, Recall: 0.13
Lyu and Lo(2020) [11]	URL, text, author, title	doc2vec	Gossipcop and PolitiFact (24556 articles, filtered to 14641)	DT	Accuracy: 95.54%
Johnson et al (2023) [12]	Text features (TF-IDF)	Data preprocessing techniques (cleaning, tokenization, stop word removal, TF-IDF vectorization)	Kaggle dataset (20,000 news articles)	LR	Accuracy: 97.90%, Precision: 96.59%, Recall: 99.32%, F1 Score: 97.94%
Kesarwani et al.(2020) [13]	Social engagement features (share count, comment count, reaction count)	Data mining techniques	BuzzFeed dataset (2282 posts)	KNN	Accuracy: 79%, Precision: 0.75, Recall: 0.79
Selva Birunda and Kanniga Devi(2021) [14]	Text-based features, site_url feature	TF-IDF, site_url feature extraction	Kaggle dataset (2050 news article)	Gradient Boosting	Accuracy: 99.5%, Precision: 1.00, Recall: 0.99, F1 Score: 0.99
Haumahu et al(2021) [15]	Text features (TF-IDF)	Text preprocessing (case folding, tokenization, filtering)	Indonesian news dataset (500 articles: 250 valid, 250 hoax)	XGBoost	Accuracy: 89%, Precision: 90%, Recall: 80%, F1 Score: 0.92
Holla and Kavitha.(2024) [16]	Text features (TF-IDF, Word2Vec	Hybrid TF-IDF, feature selection (LASSO)	WELFake dataset (72,134 articles)	AdaBoost (ensemble of ID-3, Random Forest, Naive Bayes	Accuracy: 98.98%, Precision: 99.00%, Recall: 99.00%, F1 Score: 0.99

3. THE PROPOSED METHODOLOGY

This study used machine learning techniques on fake news through text analysis using natural language (NLP) and extraction techniques. Texts were converted to digital representations using TF-IDF technology. Then, several machine learning models were trained, including SVM, RF, Logistic Regression, Naive Bayes, Gradient Boosting, AdaBoost, KNN, DT, and XGBoost. Finally, enter to the majority voting technique and then the evaluation process. The model's performance was evaluated. It was applied to improve fake news detection operations.

A-Data Description

The database is from the Kaggle website by (BHAVIK JIKADARA[1]); it is a data collection of 43 MB of news items with the following features: title, text, subject, and data. Of the articles, 21417 are accurate (the label is 1), and 23481 are false (the label is 0).

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

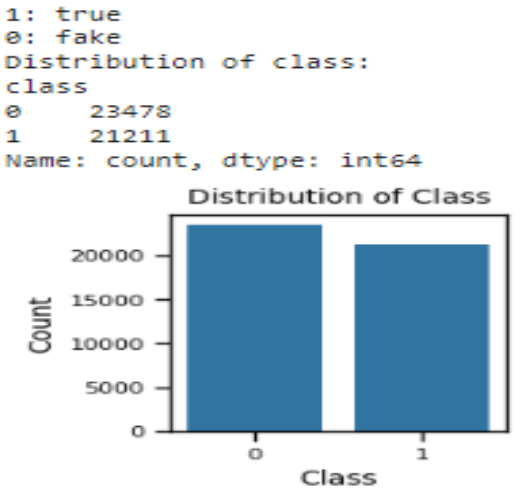


Figure 1. Dataset Labels of Ratio

The distribution of classes in our data is shown in Figure (1); the sample is reasonably balanced, with 23,468 instances of false information and 21,202 cases of real news. This balance is essential to our research because it avoids model bias toward any specific class during training. A balanced guarantee that the performance acquired during training is accurate.

B- Data Pre-processing

In the initial pre-processing stage, we obtained the dataset from the Kaggle website. It is available in two files, one true and one false. These two files were merged into one for this study to include then several pre-processing procedures. Including removing unnecessary features as well as those containing Arabic letters, also Removing HTML tags, removing "stop" words, eliminating special characters, Lemmatization, and Stemming: and Normalizing After these modifications, then split the dataset into three parts: 70% for training, and 15% for validation and 15% for test data. For the second stage of transformation using the BERT technique, the texts were divided into hashed word-based tokens using a unique symbol. Sequence transformation and padding operations were then used to transform the texts into digital sequences containing zeros at the end so that the length of the sequence was identical. Numeric strings of uniform length were produced as a result of this procedure, with zeros added where necessary. The preprocessing techniques employed included:

- Removing HTML tags: Often overlooked, HTML components introduced during web scraping must be eliminated as they are irrelevant to text analysis.
- Removal of "stop" words: Words like "a," "an," "the," "of," and "is" are taken out to speed up processing and draw attention to the text's main ideas.
- Eliminating special characters: Characters like &, %, and $ that are neither letters nor numbers are taken out because they don't add useful information.
- Stemming and Lemmatization: This method breaks down words into their basic forms, which makes sure that the meaning of the text is always the same.
- Normalization: All the text is changed to lowercase so natural language processing (NLP) tools can handle it.

More extensive input paragraphs are encoded or reduced to a single word. Figures 1 show the data before and After preprocessing; we used only the text of the article and the text.

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

	title	text	subject	date	class
0	Donald Tru...	Donald Trum...	News	December 31...	0
1	Drunk Brag...	House Intel...	News	December 31...	0
2	Sheriff Da...	On Friday, ...	News	December 30...	0
3	Trump Is S...	On Christma...	News	December 29...	0
4	Pope Franc...	Pope Franci...	News	December 25...	0

A. Dataset before cleaning

	text	class
4893	You know how we often jest, largely because it...	0
17864	When did our Washington DC Mall become a stagi...	0
6464	Now that the side-show is essentially over and...	0
18366	(Reuters) - An opposition leader who said she ...	1
11029	William Stanford Nye or Bill Nye is an America...	0

B. Dataset after cleaning

Figure 2. Data partition preprocessing

C- Feature Extraction

Text can be converted to vectors using various methods, including word2vec, TF-IDF, and N-gram. Our study used the TF-IDF technique, which performs well on non-large datasets. TF-IDF vector: One of the most popular feature extraction techniques is inverse document frequency (TF-IDF)[17]. This method has two stages: the term frequency (TF) is calculated in the first stage, and the inverse document frequency (IDF) is calculated in the second. A statistical analytic technique called TF-IDF analyses a word's significance within a corpus or group of documents. It is based on the number of times a term appears in the article and the term frequency-inverse document frequency (TF-IDF). TF divides the number of times a word appears in the article by the total number of words to avoid bias against lengthy papers. With fewer documents containing a word indicating a strong word's capacity to discriminate, IDF refers to the frequency of reverse documents. Regarding longer texts, TF-IDF excels in RNN and other neural network models in text feature extraction [18].

D- models classifier

Our model was developed using nine well-known machine-learning techniques:

- **(LR)** assesses categorical issues, the binary results of the popular version of the LR model are true/false, yes/no, and others. LR is also offered with several outcomes in place of this multinomial. LR reads the input vector and maps it to the proper category using the logistic or sigmoid function. Due to its flexibility and robustness as a classification approach, we utilized LR in this article for assessment [19]

- **(DT)** uses recursive partitioning Of all features in the training dataset to predict the final class. With nodes standing in for features, branches for decisions, and leaves for outcomes, the dataset is depicted

as a tree. We started with the data as input and gradually divided it into smaller chunks until the output indicated whether the data was real or fake [20].

- **(RF)** combines several trees, which is why it's called a forest. It is applicable to use cases involving both regression and classification. RF uses ensemble learning to prevent over-fitting and merges numerous DTs to increase the model's accuracy. We employed this classifier to expedite our suggested model's training and learning process. Since this study focused on binary problems, the outcome of RF is determined by tallying the votes cast by each tree, which can be either 0 or 1. The highest number of votes determines the outcome of RF. Where $\hat{r}= \frac{1}{n}\sum_{i=1}^{n}\hat{r}(a)$, Here, "N" stands for the total number of trees in the forest, "i" for the current tree, and "a" for the training data. $\hat{r}$ is the tree prediction [21].

- **(NB)** uses maximum conditional probability to classify news as authentic or phony. The foundation of it is "Bayes' Theorem.

$$P(X|Y) = \frac{P(Y|X) * P(X)}{P(Y)} . \quad (1)$$

When X and Y represent two occurrences. Since the NB classifier is easy to use and reasonably priced to compute, we used it for text classification. Compared to other classifiers, NB requires less data for training [22].

- **SVM** is a supervised approach based on machine learning mostly applied to classification issues. The SVM algorithm aims to identify the best hyperplane in an N-dimensional space or optimal line in a 2D space for classifying the points. To tell the difference between the two groups of points, this kind of hyperplane is used to make the gap between them as big. There are different ways to group data points that are on either side of the line into different groups. The main goal of the SVM method is to make the space between the data points and the hyperplane bigger. The hinge loss function is the one that makes the gap bigger. The equation for the hyperplane is

$$w^x + b = 0 \quad [23]. \quad (2)$$

Where the bias (b) and weight vector (w) are expressed.

$$l(w) = \sum(max\ (0,1 - y_i[w^t\ x_i + |b|]) + \lambda||w||^2 \quad [24]. \quad (3)$$

Any errors resulting from data points closer to the categorization boundary than the margin are calculated using the loss function, which is the first term. The second term, the regularization function, is employed to prevent overfitting.

- **Gradient Boosting:** Super Gradient Boosting (XGBoost) is a boosting classifier designed for supervised machine learning methods. To create a model that is more reliable and accurate
This approach involves many weak learners, such as decision trees. Its basic idea is to train multiple models one by one with reinforcement. Each model attempts to handle errors Where learners are brought together to make the final prediction. Thus, it provides the durability and high accuracy to improve grading results. XG and SF comprise the two properties of this classifier. The first characteristic was used in this study[25].

- **KNN:** This approach, which comes from supervised machine learning techniques, is known for its simplicity and ease of use. It was initially applied to resolving regression and classification issues in the early 1970s. The similarity between neighbors' ideas serves as the foundation for this method. Cases are categorized based on the majority of votes cast by their neighbors, as the similarity principle is dependent on the value of K. This happens due to nearby, comparable occurrences [26] [13].

- **AdaBoost:** techniques combine multiple weak classifiers to make a robust classifier The suggested approach combined many decision tree classifiers with various hyperparameters using AdaBoost to

produce a more reliable and accurate classification. Weak classifiers are iteratively trained using AdaBoost, and their weights are modified based on their performance on the training set. The weak classifiers are then combined using a weighted majority vote to create the final classifier [27].

• **XGBoost**: This is well-liked by rivals in the machine learning space and practitioners in the field. Gradient-boosting decision trees are implemented using XGBoost, which offers speed and performance. Y. For binary and multiclass classification tasks, it does rather well. XGBoost is employed in the two regressions.

For our classification task, classification issues are predicted to perform well. Resolution XGBoost Due to its gradient-boosting nature, it typically performs better than random forests [28]

- **Majority Voting:** Voting is the most basic way to work as a group, and it usually works pretty well. It can be used for both classification and regression problems. It splits a model into two or more submodels. The majority vote method is used to combine the results from each sub-model. Figure 3 shows how the majority vote works. It is a meta-classifier that uses a majority vote to find machine learning classifiers that are theoretically similar or different. We guess the final class label by voting for the most votes, which is how classification algorithms usually guess the class label. We can guess the class name y by using equation (4) and the vote of the majority of each classifier, Cj. [29][30].

$$y = mode\{C1(x), C2(x), \ldots . Cm(x)\} . \quad (4)$$

where,
y = predicted label of class and
$C1(x), C2(x), \ldots, Cm(x)$= models for classifying.

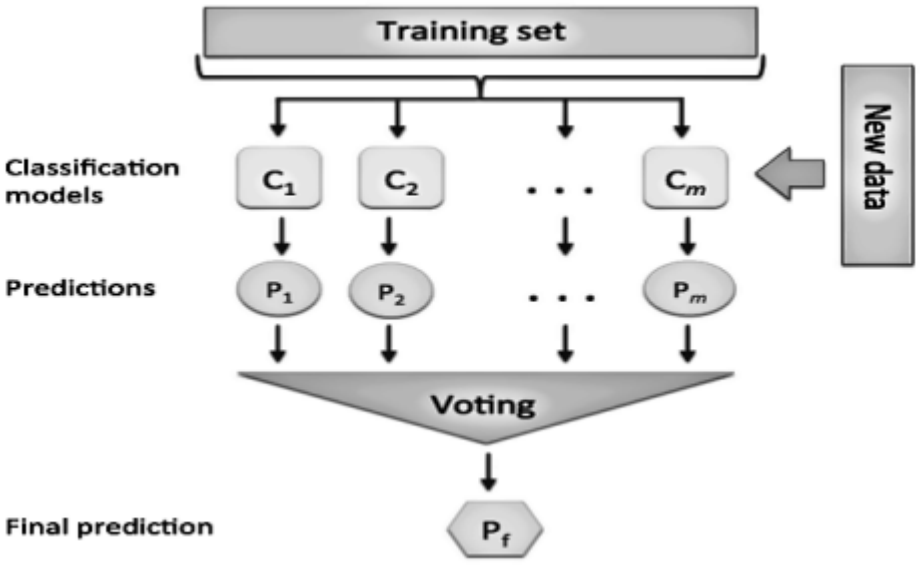


Figure 3. Classifier for Majority Voting [31].

A. Methodology

This study used different ML techniques to identify fake news, including SVM, LR, RF, Naive Bayes Gradient Boosting, AdaBoost, KNN, DT, and XGBoos. Data were collected and pre-cleaned, as well as extracting important features using TF-IDF to enter the classification process using the aforementioned.

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

Techniques, after the process of dividing the data into 70% training, 15% verification, and 15% testing. Finally, enter to the majority voting technique and then the evaluation process. The proposed model combines several machine-learning techniques to classify whether the news is fake or real. The block diagram of the proposed model is shown in Figure 4.

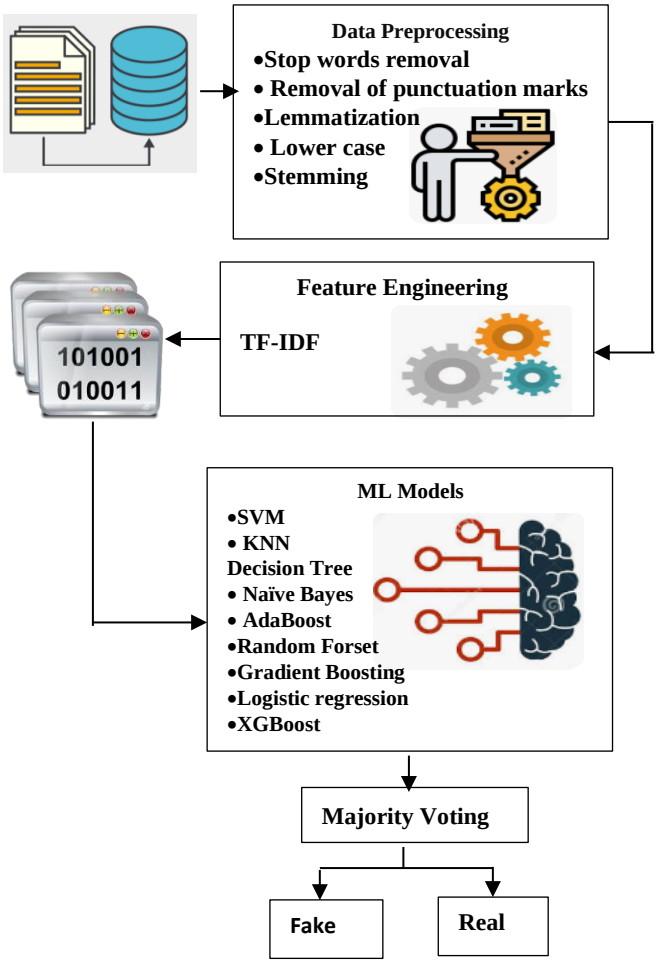


Figure 4. Proposed framework for our method to detect fake news

B. Performance Evaluation Matric

We use accuracy (A), precision (P), recall (R), and F1-score (F) as assessment measures to compare and evaluate our model. Equations 1 calculate accuracy, and 2 And 3 calculate precision and Recall. On the other hand, as stated in equation 4, the F1-score is the harmonic mean for recall and precision.

- **Accuracy:** The classification accuracy is the number of samples in the source data set that were properly matched with each sample in the data set. A "TP" result is an output from the classifier that has a positive forecast and a real class. The (TN) is what a classifier gives you when the expected class is also a true negative. When the classifier guesses a positive result when the real result is a negative one, this is called a false positive (FP) or false negative (FN) classification error [32]. Eq. (1)1Explains the accuracy calculation.

$$A = \frac{TP + TN}{TP + FN + FP + TN} \quad (5)$$

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

- **Precision:** Precision is the ratio of correctly categorized positive class values to the sum of incorrectly classified positive class values. It provides information about the model's factual accuracy [33]

$$P=\frac{\textbf{True Positive}}{\textbf{True Positive + False Positive}} \quad (6)$$

- **Recall:** The ratio of correctly categorized positive class values to the total of correctly classified positive class values and incorrectly classified negative class values is known as the recall rate. It provides information on the model's completeness [34]

$$R=\frac{True\ Positive}{TruePositive + FalseNegative} \quad (7)$$

- **F1-Measure:** The F1-score harmonically represents recall and precision [35]. Score F1 is explained in Equation (4).

$$F1= 2*\frac{\textbf{precision} \cdot \textbf{recall}}{\textbf{precision} + \textbf{recall}} \quad (8).$$

Where it is characterized by efficiency and fast, deals with large data, does not allow for Preventing over-fitting of the data, handles missing values, and performs tree pruning to enhance the predictive models making it ideal for predictive modeling.

4. Results and Discussion

This section provides the last conclusion and recommendation of this research work. Classification had been done with the help of various machine learning classifiers and TF-IDF is the method employed for word embedding. From Table 2, it is evident that the accuracy level attained for most of the informative techniques applied in the classification process such as (RF and DT) was 0. 99. SVM also proved to be very effective because of the possibility to separate one class from another within high-dimensional space based on the analysis of the most effective decision planes, which is important for distinguishing fake news. The performance assessment also showed that the classifier (KNN) had the worst models among the ones that the study explored because is noisy, has slow performance on large datasets, and has high dimensional data where distances between the points are not distinct.

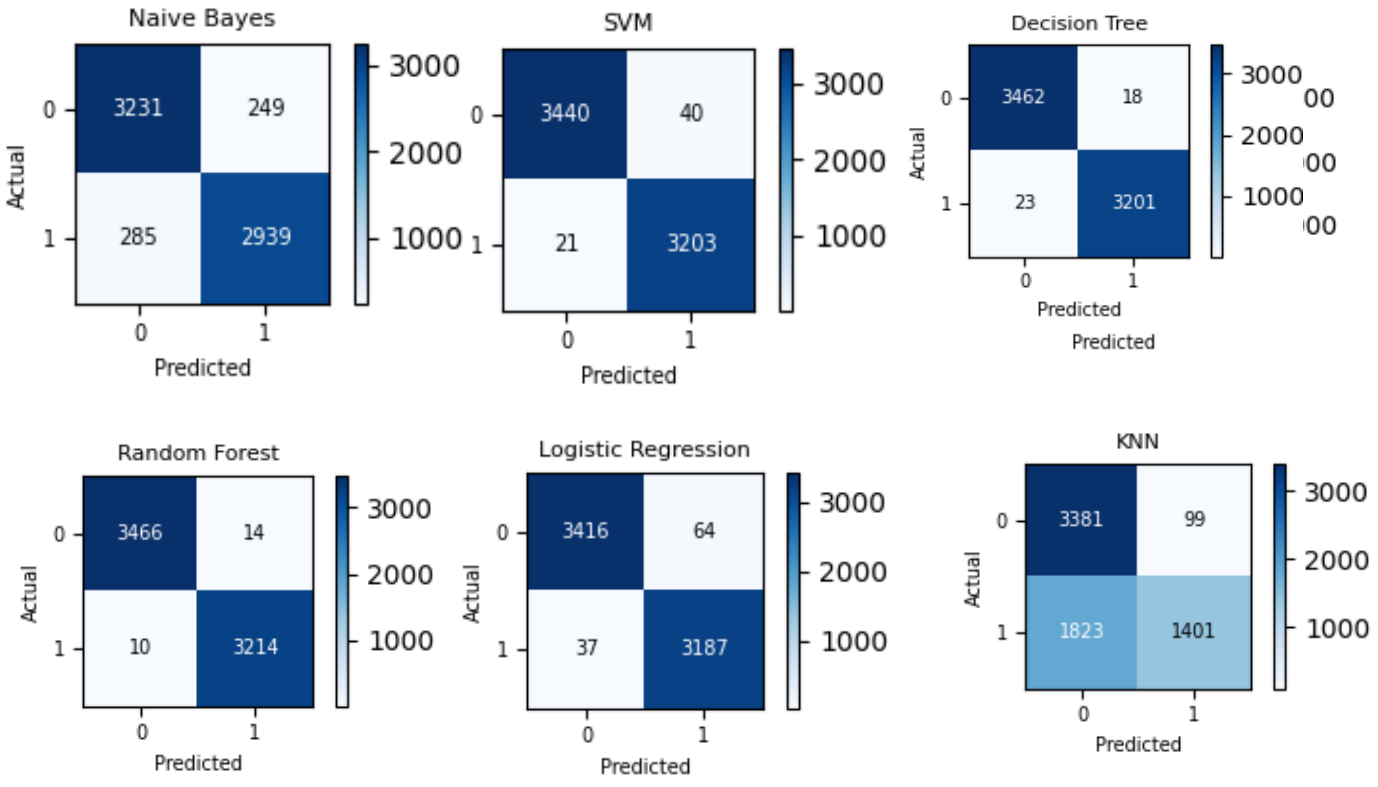
Finally, found that the best technique among the other methods was our model (Majority Voting), which achieved the highest results of 0.9977 with the lowest loss of 0. 0022%. Because is distinguished by its efficiency and speed of work and due to its better ability to deal with large data sets, it does not prevent data duplication and is good at dealing with missing data. The reason that makes suitable for use in predictive modeling.

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

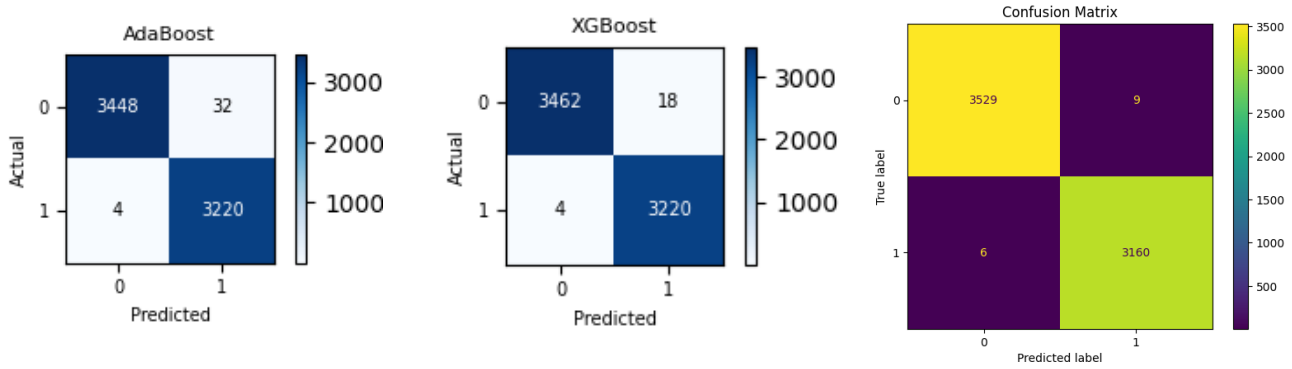
Table 2: Classification Result of Different Machine Learning Methods.

Classifier	Evaluation metrics				
	Accuracy	*Loss*	*Recall*	*Precision*	*F-score*
LR	**0.9846**	**0.015364**	**0.985605**	**0.981523**	**0.9835**
SVM	**0.9920**	**0.007906**	**0.992642**	**0.990425**	**0.9915**
RF	**0.9964**	**0.003580**	**0.997761**	**0.994579**	**0.9961**
Naive Bayes	**0.9279**	**0.072047**	**0.91170**	**0.93228**	**0.9218**
Gradient Boosting	**0.9953**	**0.004624**	**0.998401**	**0.99173**	**0.9950**
KNN	**0.7127**	**0.287291**	**0.420026**	**0.92075**	**0.5768**
DT	**0.9956**	**0.004326**	**0.995521**	**0.99520**	**0.9953**
AdaBoost	**0.9947**	**0.005221**	**0.995841**	**0.99298**	**0.9944**
XGBoost	**0.9967**	**0.003282**	**0.99840**	**0.99458**	**0.9964**
Proposed Approach	**0.9977**	**0.0022**	**0.99810**	**0.9971**	**0.9976**

The Python code that runs the algorithm code on the Anaconda platform automatically obtains the confusion matrix using the cognitive learning module.



Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985



Majority Voting

Figure 5. Results of confusion matrices for techniques

Figure (6) (7) shows the accuracy and loss results of all techniques. The results showed that the best technique for this study, which used data from Kaggel.

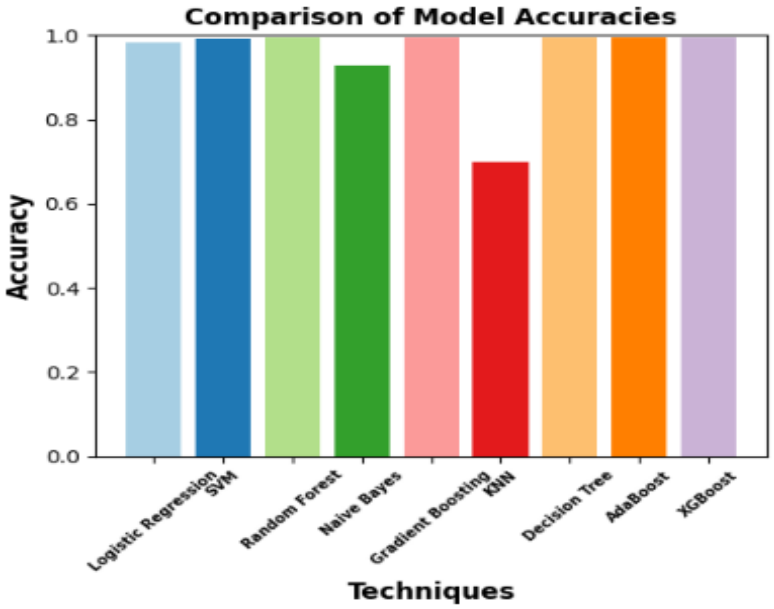


Figure 6. Results accuracy of all techniques

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

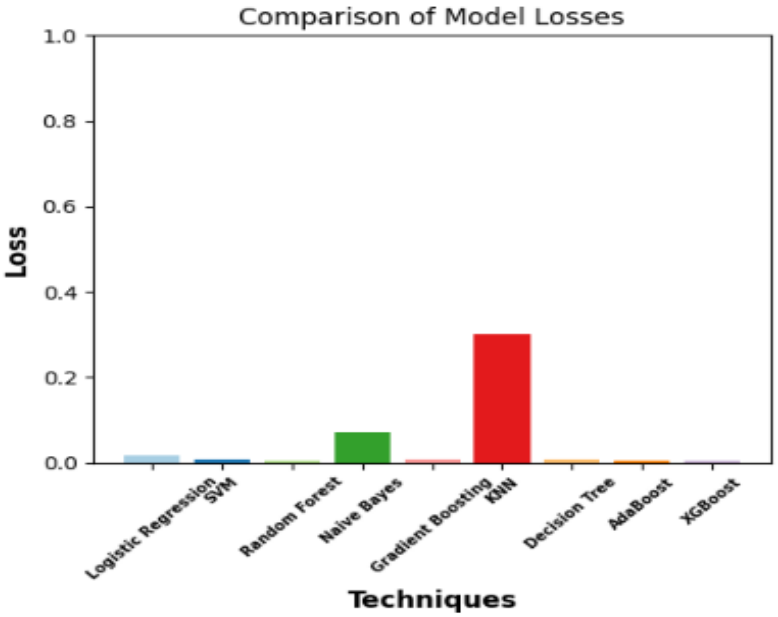


Figure 7. Results in the loss for different techniques

5. Conclusion and Future Works

In this study, we have put forward a majority voting-based method for identifying fake news. We utilized the 43 MB from the Kaggle website for the News dataset, which is publicly accessible. Which includes about 21417 of real and 23481 of fake news, with the binary numbers 0 and 1. To analyze the dataset, we have used popular machine learning classifiers such as Support Vector Machine (SVM), Random forest (RF), Logistic Regression(LR), Naive Bayes(NB), Gradient Boosting, AdaBoost, K-nearest neighbor (KNN), Decision tree (DT), and Extreme Gradient Boosting (XGBoost). We have used the CountVectorizer and TF-IDF features. To get more accurate results, we developed a multi-model false news detection system using the Majority Voting approach, using the previously described classifiers and attributes. According to the experimental findings, our suggested method produced results with an accuracy of 0.9977%, precision of 0.9971%, recall of 0.99810%, and F1-measure of 0.9976% with a 0.0022 loss. When compared to individual learning strategies, the examination demonstrates that the majority voting methodology produced more accurate outcomes.

References

[1] Y. Yang, L. Zheng, J. Zhang, Q. Cui, Z. Li, and P. S. Yu, "TI-CNN: Convolutional Neural Networks for Fake News Detection," 2018, [Online]. Available: http://arxiv.org/abs/1806.00749

[2] S. Gundapu and R. Mamidi, "Transformer-based Automatic COVID-19 Fake News Detection System," pp. 1–12, 2021, [Online]. Available: http://arxiv.org/abs/2101.00180

[3] G. Shrivastava, P. Kumar, R. P. Ojha, P. K. Srivastava, S. Mohan, and G. Srivastava, "Defensive modeling of fake news through online social networks," *IEEE Trans. Comput. Soc. Syst.*, vol. 7, no. 5, pp. 1159–1167, 2020, doi: 10.1109/TCSS.2020.3014135.

[4] Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science. 2018 Mar 9;359(6380):1146-51 [5] A. Roy, K. Basak, A. Ekbal, and P. Bhattacharyya, "A Deep Ensemble Framework for Fake News Detection and Classification," 2018, [Online]. Available:

http://arxiv.org/abs/1811.04670

[5] A. Roy, K. Basak, A. Ekbal, and P. Bhattacharyya, "A Deep Ensemble Framework for Fake News Detection and Classification," 2018, [Online]. Available: http://arxiv.org/abs/1811.04670

[6] A. Al Mamun Sardar, S. A. Salma, M. S. Islam, M. A. Hasan, and T. Bhuiyan, "Team sigmoid at CheckThat!2021 Task 3a: Multiclass fake news detection with Machine Learning," *CEUR Workshop Proc.*, vol. 2936, pp. 612–618, 2021.

[7] A. Altheneyan and A. Alhadlaq, "Big Data ML-Based Fake News Detection Using Distributed Learning," *IEEE Access*, vol. 11, no. March, pp. 29447–29463, 2023, doi: 10.1109/ACCESS.2023.3260763.

[8] A. Patel, A. K. Tiwari, and S. S. Ahmad, "Fake News Detection using Support Vector Machine," no. Icacse 2021, pp. 34–38, 2022, doi: 10.5220/0010562000003161.

[9] T. A. Wotaifi and B. N. Dhannoon, "Improving Prediction of Arabic Fake News Using Fuzzy Logic and Modified Random Forest Model," *Karbala Int. J. Mod. Sci.*, vol. 8, no. 3, pp. 477–485, 2022, doi: 10.33640/2405-609X.3241.

[10] Institute of Electrical and Electronics Engineers. Ukraine Section, Institute of Electrical and Electronics Engineers. Region 8, European Microwave Association, and Institute of Electrical and Electronics Engineers, "2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON) : conference proceedings : May 29-June 2, 2017, Kyiv, Ukraine," pp. 900–903, 2017.

[11] K. Shu, S. Wang, and H. Liu, "Understanding User Profiles on Social Media for Fake News Detection," *Proc. - IEEE 1st Conf. Multimedia. Inf. Process. Retrieval, MIPR 2018*, pp. 430–435, 2018, doi 10.1109/MIPR.2018.00092.

[12] J. Adeleke Adeyiga, P. Gbounmi Toriola, T. Elizabeth Abioye, and A. Esther Oluwatosin, "Fake News Detection Using a Logistic Regression Model and Natural Language Processing Techniques," pp. 1–18, 2023, [Online]. Available: https://doi.org/10.21203/rs.3.rs-3156168/v1

[13] A. Sharma, I. Singh, and V. Rai, "Fake News Detection on Social Media," *2022 2nd Int. Conf. Adv. Comput. Innov. Technol. Eng. ICACITE 2022*, pp. 803–807, 2022, doi: 10.1109/ICACITE53722.2022.9823660.

[14] S. Selva Birunda and R. Kanniga Devi, "A Novel Score-Based Multi-Source Fake News Detection using Gradient Boosting Algorithm," *Proc. - Int. Conf. Artif. Intell. Smart Syst. ICAIS 2021*, pp. 406–414, 2021, doi: 10.1109/ICAIS50930.2021.9395896.

[15] J. P. Haumahu, S. D. H. Permana, and Y. Yaddarabullah, "Fake news classification for Indonesian news using Extreme Gradient Boosting (XGBoost)," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1098, no. 5, p. 52081, 2021, doi: 10.1088/1757-899x/1098/5/052081.

[16] L. Holla and K. S. Kavitha, "An Improved Fake News Detection Model Using Hybrid Time Frequency-Inverse Document Frequency for Feature Extraction and AdaBoost Ensemble Model as a Classifier," *J. Adv. Inf. Technol.*, vol. 15, no. 2, pp. 202–211, 2024, doi: 10.12720/jait.15.2.202-211.

[17] S. Ghosh and M. S. Desarkar, "Class Specific TF-IDF Boosting for Short-text Classification," pp. 1629–1637, 2018, doi: 10.1145/3184558.3191621.

[18] K. Li, "HAHA at FakeDeS 2021 : A Fake News Detection Method Based on TF-IDF and Ensemble Machine Learning," no. September, 2021.

[19] P. K. Verma, P. Agrawal, V. Madaan, and R. Prodan, "MCred: multi-modal message credibility for fake news detection using BERT and CNN," *J. Ambient Intell. Humans. Comput.*, vol. 14, no. 8, pp. 10617–10629, 2023, doi: 10.1007/s12652-022-04338-2.

[20] R. K. Kaliyar, A. Goswami, and P. Narang, "Multiclass Fake News Detection using Ensemble Machine Learning," *Proc. 2019 IEEE 9th Int. Conf. Adv. Comput. IACC 2019*, pp. 103–107, 2019, doi: 10.1109/IACC48062.2019.8971579.

[21] "1806 @ Arxiv.Org." [Online]. Available: https://arxiv.org/abs/1806.08790

[22] M. Villagracia Octaviano, "Fake News Detection Using Machine Learning," *ACM Int. Conf. Proceeding Ser.*, pp. 177–180, 2021, doi: 10.1145/3485768.3485774.

[23] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A hybrid deep model for fake news detection," *Int. Conf. Inf. Knowl. Manag. Proc.*, vol. Part F1318, pp. 797–806, 2017, doi: 10.1145/3132847.3132877.
[24] S. Bajaj, "' The Pope Has a New Baby !' Fake News Detection Using Deep Learning," *Cs 224N*, pp. 1–8, 2017.
[25] M. A. Taha, H. D. A. Jabar, and W. K. Mohammed, "Fake News Detection Model Basing on Machine Learning Algorithms," *Baghdad Sci. J.*, 2024, doi: 10.21123/bsj.2024.8710.
[26] K. Alkhatib, H. Najadat, I. Hmeidi, and M. K. A. Shatnawi, "Stock Price Prediction Using K-Nearest Neighbor Algorithm," *Int. J. Business, Humanit. Technol.*, vol. 3, no. 3, pp. 32–44, 2013.
[27] A. K, S. Kumar P., N. Prasath E., and V. M. Sneha, "Fake News Detection using Decision Tree and Adaboost," *Eur. Chem. Bull*, vol. 12, no. S3, pp. 570–582, 2023, doi 10.31838/ECB/2023.12.s3.065.
[28] R. S. Utsha, M. Keya, M. A. Hasan, and M. S. Islam, "Qword at CheckThat! 2021: An extreme gradient boosting approach for multiclass fake news detection," *CEUR Workshop Proc.*, vol. 2936, pp. 619–627, 2021.
[29] "60c8d7653052d61a5defdc86e5728af672ea9570 @ rasbt.github.io." [Online]. Available: https://rasbt.github.io/mlxtend/user_guide/classifier/EnsembleVoteClassifier/
[30] Stratified K-Folds, "Sklearn @ Scikit-Learn.Org." [Online]. Available: https://scikit-learn.org/stable/modules/cross_validation.html#stratified-k-fold
[31] D. R. Patil, "Fake News Detection Using Majority Voting Technique," 2022, [Online]. Available: http://arxiv.org/abs/2203.09936
[32] S. A. Abdul Kareem and Z. F. Rasheed, "A Machine Learning Model for Cancer Disease Diagnosis using Gene Expression Data," *J. Kufa Math. Comput.*, vol. 10, no. 2, pp. 179–185, 2023, doi: 10.31642/jokmc/2018/100227.
[33] "S0306457309000259 @ doi.org." [Online]. Available: https://doi.org/10.1016/j.ipm.2009.03.002
[34] Real Academia de la Lengua, "Search @ Www.Google.Com," *2023*. 2024. [Online]. Available: https://www.google.com/search?q=corcetes&oq=corcetes&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTIMCAEQABgKGLEDGIAEMg8IAhAAGAoYgwEYsQMYgAQyCQgDEAAYChiABDIJCAQQABgKGIAEMgkIBRAAGAoYgAQyCQgGEAAYChiABDIJCAcQABgKGIAEMgkICBAAGAoYgAQyCQgJEAAYChiABNIBCDE1MzRqMGo3qAIAsAIA&sourc
[35] A. Botchkarev, "A new typology design of performance metrics to measure errors in machine learning regression algorithms," *Interdiscip. J. Information, Knowledge, Manag.*, vol. 14, no. January, pp. 45–76, 2019, doi: 10.28945/4184.