Iraqi Journal for Computer Science and Mathematics

Journal Homepage: http://journal.esj.edu.iq/index.php/IJCM
e-ISSN: 2788-7421 p-ISSN: 2958-0544

Hard Voting Approach using SVM, Naïve Bays and Decision Tree for Kurdish Fake News Detection

Rania Azad M. San Ahmed
Computer Network Department, Technical College of Informatics, Sulaimani Polytechnic University, Sulaimani, Iraq,

*Corresponding Author: Rania Azad M. San Ahmed

DOI: https://doi.org/ 10.52866/ijcsm.2023.02.03.003
Received March 2023; Accepted May 2023; Available online July 2023

ABSTRACT: In today's age of excessive information, the proliferation of fake news has emerged as a significant challenge, eroding the trustworthiness of news sources and posing serious threats to society. With the advent of social media and the internet, false information spreads more easily, necessitating the development of effective methods to identify and prevent its dissemination.While machine learning models have been widely employed to classify text data as authentic or fake, the majority of existing research has primarily focused on well-resourced languages such as English, leaving low-resourced languages like Kurdish largely overlooked. To address this gap, our proposed work introduces a novel approach: a hard voting ensemble method that combines the insights of four weak learners. By fine-tuning these weak learners for optimal performance, our approach achieves enhanced accuracy compared to the current state-of-the-art methods. Specifically, our findings demonstrate the effectiveness of the hard voting approach using a combination of Support Vector Machines (SVM), Decision Trees (DT), and Naive Bayes (NB) classifiers, resulting in an impressive accuracy rate of 89.73%. This outperforms the individual SVM classifier and underscores the potential of ensemble techniques in fake news detection.

Keywords: Kurdish language, Ensemble Learning, Hard Voting, Fake News Detection, False News Identification

1. INTRODUCTION

From paper-based to web-based, media platforms like “Facebook”, “Twitter”, “Youtube”, “Instagram”, and others facilitated users to create, interact, participate, collaborate, and share information in addition to consuming content. Social media depend on user-generated decisions such as tweeting, commenting, liking, posting, sharing, and uploading, among others. As result, such platforms have greatly assistant to the spread of misinformation, disinformation, and fake news [1]. Fake news, also known as misinformation or disinformation, has become a pervasive problem in today's society. The spread of false information has the potential to cause harm to individuals and society as a whole, affecting politics, economics, and social issues [2]. Fake news can be spread through various media outlets, including traditional news sources and social media platforms, and has become more prevalent with the rise of the internet [1]. The spread of false information has become a growing concern, as it can have serious impacts on politics, economics, and social issues. For instance, fake news about COVID-19 has led to misinformation about its symptoms and treatments, causing panic and confusion among people [3] [4].

AI [5], machine learning (ML) [6] and deep learning (DL) [7]. approaches offer a promising solution to the problem of fake news detection, with the potential to accurately identify and classify fake news [3]. Machine learning algorithms can be trained on large datasets to classify news articles into different categories, such as fake or real, based on various features, such as text and images [4]. ML algorithms have been used with different languages, including English, French, and Spanish, to detect fake news [5]. However, these approaches have been neglected in low-resourced languages like Kurdish, making it imperative to address this gap in research [8]–[10]. Kurdish is a low-resourced language, spoken by approximately 30 million people primarily in the middle east [11]. which spans Iran, Iraq, Turkey, and Syria. Kurmangi and Sorani are the two main dialects of the Kurdish language. Sorani Kurdish alphabets have 33 letters with an Arabic-based writing script that reads from right to left as shown in figure 1 [12]. While the Kurdish language has a rich cultural and historical heritage, few resources exist in the field of natural language processing (NLP) and artificial intelligence(AI) which lead lack of addressing important problems, such as fake news detection in the Kurdish language[11].

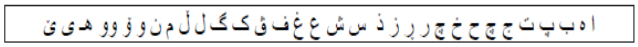


Figure 1**: Kurdish Alphabets [12]**

The aim of this study is to propose a new model for detecting fake news in the Kurdish language using an ensemble learning technique. This paper makes two key contributions to the field:

- Applying four base individual classifiers: Support Vector Machines (SVM), Logistic Regression (LR), Decision Trees (DT), and Naive Bayes (NB) classifiers with fine-tuning parameters to outperform the state-of-art.
- Propose a novel hard voting ensemble model on the same individual base classifiers to enhance the performance and improve fake news detection accuracy.

To the best of our knowledge, this is the first approach of utilizing ensemble voting approach for detection Kurdish fake news.
The remaining sections are organized as follows: Part II provides a brief review of relevant work on the detection of false news. In Part III, the approach employed in this study is described. The results and evaluation of the proposed approach are presented in Section IV. Part V closes the paper and describes future research.

2. RELATED WORK

Countless previous studies on fake news identification have been conducted using machine learning algorithms and ensemble classifiers in English language but only one work until today focuses on the Kurdish language [11]., Thus, to develop reliable literature review section, it is important to focus on previous studies that apply ensemble methods on low-resourced languages similar to Kurdish language such Persian, Arabic and Kurdish.

According to [11], the first Kurdish dataset has been built to address the issue of fake news in the Kurdish language entitled the “KurdFake” that consists of 10000 news. Preprocessing was then done to prepare the data. Data cleaning, normalization and UTF-8 was performed first to remove any unnecessary data. Tokenization is then done to split the data into single words separated by a space. Removal of stop words is performed to remove stop words from a list that contains 240 Kurdish stop words. The final step is stemming to return each word to its original root. The outcome of the research shows that SVM had the best accuracy of 88.71% while LR classifier outperformed the other classifiers using auto-manipulated false news from the existing real news. Few other studies similar to this work has been conducted on the Persian language. Researchers in [13] built a Persian corpus for COVID-19 hoaxes Then, using a parallel CNN classifier and the XLM-RoBERT language model, then applied knowledge transfer across language and domain models to detect Persian COVID-19 false news. The results demonstrate a 2.39 percent improvement over the baseline. While [14] built a rumor Persian dataset entitled KNTUPT that contains 35049 tweets that have been used for training and testing, the Random Forest scored 99.7% which is the highest F-score compared to MLP, KNN, Decision Tree, Naïve Bayes, Rules. Part. On the other hand, [15] explored the possibility to use the content-based method based on lexical features applied to the Persian Tweets dataset. Results indicate that SVM outperforms the other classifiers by 87%.
Few other studies focus on applying ensemble methods on Arabic language, the authors in [16] used An Arabic dataset related to COVID-19 pandemic that contains real and fake articles is used by work. In data preprocessing a number of functions were applied to clean the data and prepare it for feature extraction. count vectorization and term frequency-inverse-document-frequency (TF-IDF) is used to extract features from the data in the feature extraction step. Five classifiers are then used (NB, SVM, LR, RF and GB). Results show that the Gradient Boosting classifier outperforms other classifiers with all the metrics except for recall. The authors in [4] collected a corpus that consists of 37 000 Arabic tweets that were fed to Logistic regression with the n-gram-level and TF-IDF as a feature. LR with TF-IDF achieved an F1-score of 87.8 on the manually annotated dataset while achieving 93.3% on the automatically annotated dataset an F1score of 93.3 was achieved using the same classifier with count n-gram.
Consequently, The work [17] applied four machine learning namely RF, DT, AdaBoost, and LR on 1862 Arabic tweets related to the Syrian crisis. The outcome reveals accuracy of 76% using AdaBoost and Logistic regression. In addition to this, [18] Employed soft voting on XGBoost, RF, and MLP on 4024 Arabic tweets during the FIRE 2019 forum, and the model scored 84.4% F1-points. The work [19] combined stacking classifier Genetic Algorithm Based Support Vector Machine and the Logistic Regression on the Arabic rumors dataset. The accuracy reached 92.63%.
While in the Urdu language, the authors [20] proposed to use the voting ensemble method on DT, LR, RF, NB, SVM, K-Nearest Neighbor, AdaBoost, Light-GBM, XGBoost the proposed model achieved 71.3% accuracy which bypass all the algorithm separately applied except AdaBoost 72.3%.
Two Turkish datasets were used by work [21], the research firstly uses a number of machine learning algorithms SVM,

Naïve Bayes, Naïve Bayes Multinomial, J48, Random Forest, and KNN. The results show that Multinomial Naïve Bayes had the best accuracy of 96.74% which was improved to 98.2% after applying the AdaBoost technique on the top of Multinomial Naïve Bayes.

The techniques suggested in the literature for detecting false news are summarized in Table 1. It can be seen from the table that various techniques for detecting false news in various languages have been employed by existing studies.

Table 1: Related Work Summary

Work	Suggested methodology	Performance
[11]	SVM	Accuracy = 88.71%
[13]	CNN classifier and the XLM-RoBERT	
[14]	Random Forest, MLP, KNN, Decision Tree, Naïve Bayes,	F-score RF = 99.7%
[15]	SVM	Accuracy = 87%.
[16]	NB, SVM, LR, RF and GB	Accuracy GB and SVM=94%
[4]	Logistic Regression with the n-gram-level and TF-IDF	F1-score =87.8%
[17]	RF, DT, AdaBoost, and LR	Accuracy of AdaBoost and Logistic regression=76%
[18]	Soft voting on XGBoost, RF, and MLP	F1-score=84.4%
[19]	Combined Stacking Classifier (LR) and GA-SVM	Accuracy of Stacking LR= 92.63%.
[20]	voting ensemble method on DT, LR, RF, NB, SVM, K-Nearest Neighbor, AdaBoost, Light-GBM, XGBoost	Accuracy= 71.3%
[21]	SVM, Naïve Bayes, Naïve Bayes Multinomial, J48, Random Forest, and KNN.	Adaboost = 98.2%

3. METHODOLOGY

The proposed model aims to investigate enhanced results to identify false and genuine news in Kurdish language. Therefore, A Kurdish Dataset is used. After cleaning and preprocessing the data, 5-fold cross validation was applied, then TF-IDF was utilized to extract features from the dataset. Finally, the classification experiment run into two stages.

Stage 1: Apply four base individual classifiers named: Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), and Naive Bayes (NB) on Kurdish Dataset.

Stage 2: In order to enhance the performance of the previous step. Three classifiers chosen from the previous step have been selected as input for the hard voting process.

3.1. DATASET

In this study, the "FakeKurdNews"[1] dataset was utilized, which is a novel dataset created for the purpose of detecting fake news in the Kurdish Sorani language [11]. The dataset is comprehensive and accessible to the public, covering a variety of domains. It consists of 10,000 news articles, half of which are fake news collected from untrustworthy Facebook pages, and the other half are legitimate news obtained from reputable Facebook pages using a Facebook scraper tool.

3.2. PREPROCESSING

The goal of preprocessing is to prepare the data in a format that is usable by the machine learning algorithms, and to

[1] FakeKurdNews - Fake Kurdish News Dataset | Kaggle

improve the performance of the fake news detection model. To optimize feature selection and improve the prediction results, the data must be pre-processed before being fed into the classifier using a Python script. Firstly, the dataset must be cleaned of hashtags, emojis, links, HTML tags, Latin characters, and digits, as these can cause issues for the machine learning process. Then, the 32 primary punctuation marks were removed and replaced with an empty string: '!"#$%&'()*+؛-./:<=>?@[]^_`{|}~'. The output was then normalized. Next, the Kurdish stopwords were eliminated. Stopwords are the most frequently occurring words in a text that do not provide valuable information. In the Kurdish language, a list of 240 stopwords was developed by [12] including "جا", "تەنها", "پاشان", "ئێمە", "هیچی", and "زوو". After these steps are completed, the entire dataset was tokenized, split into words by splitting from whitespaces, and finally, the tokens were stemmed to their roots using the KLPT Kurdish toolkit [23]. These preprocessing steps are crucial for improving the performance of fake news detection models, as they help to reduce noise, eliminate irrelevant information, and make the data more suitable for use by the machine learning algorithms.

3.3. DATASET SPLIT AND CROSS-VALIDATION

It is essential to divide the dataset to evaluate classification models. This involves separating the data to the training dataset, which is used to train the model, and the test dataset, which is used to test the model's accuracy. To enhance performance, the test set is typically smaller than the training set. The process of splitting the data is called k-fold cross-validation, which helps to prevent overfitting by resampling the dataset based on a parameter called k. The k parameter determines the number of groups into which the dataset will be divided. In this particular experiment, we employed 10-fold cross-validation.

3.4. FEATURE EXTRACTION

It is a crucial step in detecting fake news. It involves selecting relevant and informative words or phrases from text data to train a machine learning model. TF-IDF representation is a common method used, where each document is represented as a vector of weighted word frequencies. The most relevant and informative features are identified to differentiate between fake and real news. The choice of features and extraction method can impact the performance of the fake news detection model [22]. In this study, we used TF-IDF because of its simplicity and robustness since it can be applied to any language, making it appropriate for detecting fake news in languages with limited access to powerful natural language processing techniques.

3.5. HYPERPARAMETERS AND FINE-TUNING

In order to obtain better results while training individual classifiers, fine-tuning was applied by selecting randomly parameters as shown in table 2 to produce the best performance against the state-of-art

Table 2: Classifiers' Hyper-parameters

Classifier	**Hyper-parameters**
SVM	kernel='rbf', C=100.0, probability=True
NB	alpha=0.1, class_prior=None, fit_prior=True
LR	C=5,penalty='l2', solver='saga'
DT	criterion='entropy', random_state=10, min_samples_split=5, min_samples_leaf=1

The selection of parameters in the SVM algorithm is essential for detecting fake news. The 'rbf' kernel is used to capture delicate data patterns and correlations, while setting the regularization parameter (C) to 100.0 emphasizes accurate classification of training points. Probability estimation (probability=True) is also enabled to derive probability estimates for the predicted class labels, providing valuable insights into the confidence or likelihood associated with the model's predictions. This parameter configuration indicates the importance of precise classification for distinguishing false news from authentic news.

While, The Naive Bayes classifier, with alpha=0.1, class_prior=None, and fit_prior=True, is effective in fake news detection. The alpha parameter enables better generalization, class_prior=None estimates class probabilities automatically, and fit_prior=True learns class priors from the data. These parameters help the classifier adapt well to fake news detection, handling uncertain data, accurately estimating class probabilities, and learning effectively from the training data.

Logistic regression classifier with parameters C=5, penalty='l2', and solver='saga' can have a positive impact on fake news detection. The regularization parameter C controls the extent of regularization, while the 'l2' penalty encourages the model

to use smaller coefficients. The'saga' solver is well-suited for large-scale problems and performs well with L1 and L2 penalties. These parameters help the classifier learn the patterns and features indicative of fake news, leading to accurate predictions and robust performance in identifying instances of misinformation. In meanwhile, The Decision Tree classifier with the parameters criterion='entropy', random_state=10, min_samples_split=5, min_samples_leaf=1 has a positive impact on fake news detection. The 'entropy' criterion prioritizes information gain and focuses on identifying the most informative features for classification. Setting the random_state to 10 ensures reproducibility of results. The min_samples_split=5 parameter specifies that a node must have at least 5 samples to be considered for further splitting. The min_samples_leaf=1 parameter ensures that each leaf node must have at least one sample. These parameters collectively help the Decision Tree classifier effectively classify fake news by selecting relevant features, reducing overfitting, and capturing both general and specific patterns in the data.

3.6. CHOOSING TOP 3 CLASSIFIER

Only three individual classifiers were combined at one time named and fed it into voting classifier to avoid equal votes.

3.7. UTILIZING HARD VOTING CLASSIFIER

The Hard Voting Classifier is a technique in machine learning that is extensively used in various applications due to its effectiveness. It is based on the principle of combining the predictions of multiple individual classifiers. The ultimate prediction of the ensemble is determined by the majority vote of these individual classifiers. The predicted class is chosen based on the class that gets the highest number of votes which can result in improved accuracy.

3.8. EVALUATION METRICS

In classification tasks, accuracy is commonly used as a performance indicator for assessing the model. However, relying solely on accuracy can lead to a flawed evaluation of the model's effectiveness. Hence, it is common to use multiple evaluation metrics to obtain a comprehensive assessment. A model with superior performance would exhibit high scores for accuracy, precision, recall, and F1-score [24] [25]. In this study, accuracy is the main metric of evaluation for the proposed model. Accurate prediction means the number of correct predictions divided by the total number of predictions. The equation for accuracy is:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (1)$$

Precision is how close the measurements [25] are:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

The recall is how many correctly actual positive [24] is defined:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

4. RESULTS AND DISCUSSIONS

This section summarizes the result of applying voting ensemble model using three individual classifiers on divided dataset using 10-fold cross-validation.

The first step of the experiment was applied fine-tuning parameters on SVM, NB, LR and DT. The classifier SVM achieved the best accuracy score at 89.5%, as well as the best recall score at 90.34% comparing to the other algorithms as shown in table 3.

Table 3: Four Classifiers Performance with tuning parameters

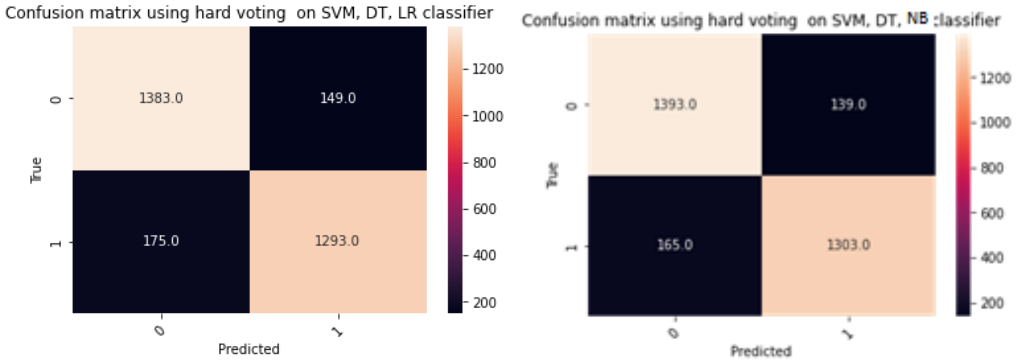
Classifiers	Accuracy (%)	Precision	Recall	F1 score
NB	89.4	**0.8955**	0.8951	**0.8952**
SVM	**89.5**	0.87.94	**0.9034**	0.8913
DT	80.7	0.8154	0.7854	0.8001
LR	88.67	0.8774	0.8895	0.8834

The second step of the experiment was applied hard voting method on combination of three traditional classifiers from the previous experiment. The results of the study, as shown in Table 4, reveal (SVM, Decision Tree, and Naive Bayes) achieved an accuracy of 89.87%, recall of 90.36% and F1-score of 89.55%. It can be seen that the voting of SVM, DT and LR combination resulted in an accuracy of 89.2%, precision of 88.08%, recall of 0.89.67%, and F1-score of 88.87%. Similarly, the voting of LR, DT and NB combination resulted in an accuracy of 89.47%, precision of 88.71%, recall of 89.61%, and F1-score of 89.19%. The voting of SVM, LR and NB combination resulted in an accuracy of 89.47%, precision of 88.08%, recall of 90.17%, and F1-score of 89.11%. These results indicate that the voting ensemble method can enhance the performance of traditional classifiers and provide a more reliable prediction of fake news.

Table 4: Applying Hard voting on three classifiers combinations

Classifiers	Accuracy (%)	Precision	Recall	F1-score
Voting (SVM+DT+LR)	89.2	0.8808	0.8967	0.8887
Voting(SVM+DT+NB)	**89.87**	**0.8876**	**0.9036**	**0.8955**
Voting (DT+LR+NB)	89.47	0.8871	0.8961	0.8919
Voting(SVM+LR+NB)	89.47	0.8808	0.9017	0.8911

The experiment's findings suggest that the voting ensemble model comprising SVM, DT, and NB is more accurate, with higher recall and F1-score than the SVM weak learner in detecting fake news in the Kurdish language. The precision and recall of the voting ensemble can be determined by analyzing the confusion matrix. High precision indicates that the model accurately predicts true positives, whereas high recall suggests that the model can effectively identify all actual positives.



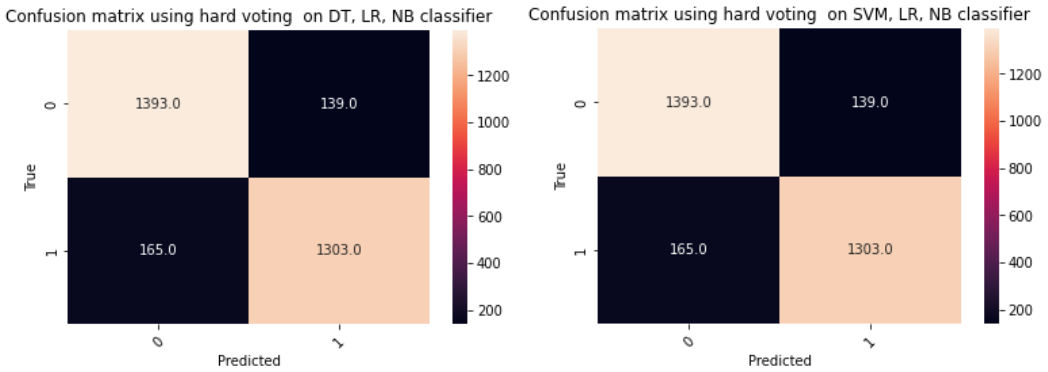


Figure 2: **Confusion matrix for hard voting**

Table 5: Benchmarking of the existing work with state-of-art

Current work				
Classifiers	**Accuracy (%)**	**Precision**	**Recall**	**F1 score**
NB	89.4	**0.8955**	0.8951	**0.8952**
SVM	**89.5**	0.87.94	**0.9034**	0.8913
DT	80.7	0.8154	0.7854	0.8001
LR	88.67	0.8774	0.8895	0.8834

[11]				
Classifiers	**Accuracy (%)**	**Precision**	**Recall**	**F1 score**
NB	88.58	**0.8878**	0.8858	0.8856
SVM	**88.71**	0.8769	**0.8975**	**0.8871**
DT	80.44	0.8054	0.8044	0.8042
LR	87.58	0.8761	0.8757	0.8757

In addition, Table 5 contains a comparison of the proposed model to the literature's benchmark studies. The hyper-tuned SVM outperformed the previous studies with an increase of 1.2%.

5. CONCLUSION

The rise of fake news and misinformation poses a significant threat to society by eroding the credibility of news sources. To combat this problem, it is crucial to develop effective methods for detecting and preventing the spread of fake news, especially in the era of widespread internet access and social media usage. While AI and ML offer promising solutions, much of the existing research focuses on well-resourced languages, neglecting low-resourced languages like Kurdish.
This study presents a novel approach to detecting fake news in the Kurdish language using machine learning techniques. The proposed method utilizes an ensemble hard voting classifier that combines four weak learners. These weak learners were optimized to enhance performance, and their outputs were aggregated through a hard voting classifier to make the final prediction. The results indicate that SVM performed the best among the individual weak classifiers, achieving an accuracy of 89.53% after parameter tuning. The hard voting classifier, consisting of SVM, DT, and NB, slightly surpassed the performance of the individual classifiers, achieving an accuracy of 89.7%.

This paper serves as a starting point for the development of more robust and accurate fake news detection systems in Kurdish and other low-resourced languages. It highlights the importance of addressing fake news detection in low-resourced languages and demonstrates the potential of machine-learning approaches in solving this critical problem.

Funding

None

ACKNOWLEDGEMENT

On behalf of our research team, I would like to express my special thanks of gratitude to MURATA Science Foundation, Ministry of Higher Education Malaysia and UTP Foundation for funding this research in order to innovate new approaches of teaching and learning engineering and technology related courses.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] S. Rastogi, & D. Bansal (2023). A review on fake news detection 3T's: typology, time of detection, taxonomies. International Journal of Information Security, 22(1), 177–212. https://doi.org/10.1007/s10207-022-00625-3

[2] S. Senhadji and R. A. S. Ahmed, "Fake news detection using naïve Bayes and long short term memory algorithms," IAES Int. J. Artif. Intell., vol. 11, no. 2, pp. 746–752, 2022, doi: 10.11591/ijai.v11.i2.pp746-752.

[3] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," Inf. Process. Manag., vol. 57, no. 2, p. 102025, 2020, doi: 10.1016/j.ipm.2019.03.004.

[4] A. R. Mahlous and A. Al-Laith, "Fake News Detection in Arabic Tweets during the COVID-19 Pandemic," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 6, pp. 778–788, 2021, doi: 10.14569/IJACSA.2021.0120691.

[5] E. Saquete, D. Tomás, P. Moreda, P. Martínez-Barco, and M. Palomar, "Fighting post-truth using natural language processing: A review and open challenges," Expert Syst. Appl., vol. 141, p. 112943, 2020, doi: 10.1016/j.eswa.2019.112943.

[6] N. Aslam, I. Ullah Khan, F. S. Alotaibi, L. A. Aldaej, and A. K. Aldubaikil, "Fake Detect: A Deep Learning Ensemble Model for Fake News Detection," Complexity, vol. 2021, 2021, doi: 10.1155/2021/5557784.

[7] X. Zhou and R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," ACM Comput. Surv., vol. 53, no. 5, 2020, doi: 10.1145/3395046.

[8] M. Alkhair, K. Meftouh, K. Smaïli, and N. Othman, "An Arabic Corpus of Fake News: Collection, Analysis and Classification," Commun. Comput. Inf. Sci., vol. 1108, no. October, pp. 292–302, 2019, doi: 10.1007/978-3-030-32959-4_21.

[9] E. M. B. Nagoudi, A. R. Elmadany, M. Abdul-Mageed, T. Alhindi, and H. Cavusoglu, "Machine generation and detection of arabic manipulated and fake news," 2020.

[10] T. Thaher, M. Saheb, H. Turabieh, and H. Chantar, "Intelligent detection of false information in arabic tweets utilizing hybrid harris hawks based feature selection and machine learning models," Symmetry (Basel)., vol. 13, no. 4, 2021, doi: 10.3390/sym13040556.

[11] R. Azad, B. Mohammed, R. Mahmud, L. Zrar, and S. Sdiq, "Fake News Detection in low-resourced languages 'Kurdish language' using Machine learning algorithms," Turkish J. Comput. Math. Educ., vol. 12, no. 6, pp. 4219–4225, 2021.

[12] A. M. Mustafa and T. A. Rashid, "Kurdish stemmer pre-processing steps for improving information retrieval," J. Inf. Sci., vol. 44, no. 1, pp. 15–27, 2018, doi: 10.1177/0165551516683617.

[13] M. Ghayoomi and M. Mousavian, "Deep transfer learning for COVID-19 fake news detection in Persian," Expert Syst., no. 99009204, pp. 1–12, 2022, doi: 10.1111/exsy.13008.

[14] S. D. Mahmoodabad, S. Farzi, and D. B. Bakhtiarvand, "Persian Rumor Detection on Twitter," 9th Int. Symp. Telecommun. With Emphas. Inf. Commun. Technol. IST 2018, no. December, pp. 597–602, 2019, doi: 10.1109/ISTEL.2018.8661007.

[15] M. Z. Asghar, A. Habib, A. Habib, A. Khan, R. Ali, and A. Khattak, "Exploring deep neural networks for rumor detection," J. Ambient Intell. Humaniz. Comput., vol. 12, no. 4, pp. 4315–4333, 2021, doi: 10.1007/s12652-019-01527-4.

[16] D. Mohdeb, M. Laifa, and M. Naidja, "An Arabic Corpus for COVID-19 related Fake News," pp. 1–5, 2021, doi: 10.1109/icrami52622.2021.9585909.

[17] G. Jardaneh, H. Abdelhaq, M. Buzz, and D. Johnson, "Classifying Arabic tweets based on credibility using content and user features," 2019 IEEE Jordan Int. Jt. Conf. Electr. Eng. Inf. Technol. JEEIT 2019 - Proc., no. 1, pp. 596–601, 2019, doi: 10.1109/JEEIT.2019.8717386.

[18] M. Khalifa and N. Hussein, "Ensemble learning for irony detection in Arabic tweets," CEUR Workshop Proc., vol.

2517, pp. 433–438, 2019.
[19] S. N. Qasem, M. Al-Sarem, and F. Saeed, "An ensemble learning based approach for detecting and tracking COVID19 rumors," Comput. Mater. Contin., vol. 70, no. 1, pp. 1721–1747, 2021, doi: 10.32604/cmc.2022.018972.
[20] H. Akram and K. Shahzad, "Ensembling Machine Learning Models for Urdu Fake News Detection Ensembling Machine Learning Models for Urdu Fake News Detection," no. October 2021, 2022.
[21] M. BOZUYLA, "AdaBoost Ensemble Learning on top of Naive Bayes Algorithm to Discriminate Fake and Genuine News from Social Media," Eur. J. Sci. Technol., no. 28, pp. 459–462, 2021, doi: 10.31590/ejosat.1005577.
[22] E. Elsaeed, O. Ouda, M. M. Elmogy, A. Atwan, and E. El-Daydamony, "Detecting Fake News in Social Media Using Voting Classifier," IEEE Access, vol. 9, no. Ml, pp. 161909–161925, 2021, doi: 10.1109/ACCESS.2021.3132022.
[23] S. Ahmadi, "KLPT – Kurdish Language Processing Toolkit," pp. 72–84, 2020, doi: 10.18653/v1/2020.nlposs-1.11.
[24] H. Dalianis, "Evaluation Metrics and Evaluation," Clin. Text Min., no. 1967, pp. 45–53, 2018, doi: 10.1007/978-3-319-78503-5_6.
[25] J. Shaikh and R. Patil, "Fake News Detection Using Machine Learning Ensemble Methods," Proc. - 2020 IEEE Int. Symp. Sustain. Energy, Signal Process. Cyber Secur. iSSSC 2020, vol. 2020, 2020, doi: 10.1109/iSSSC50941.2020.9358890.