Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 04 Issue 01
2022

Research Article
https://doi.org/10.56979/401/2022/85

Fake News Classification using Machine Learning: Count Vectorizer and Support Vector Machine

Sajid Khan[1], Mehmoon Anwar[1], Huma Qayyum[2], Farooq Ali[1]*, Marriam Nawaz[1]

[1]Department of Computer Science,University of Engineering & Technology Taxila,Pakistan.
[2]Department of Software Engineering, University of Engineering & Technology Taxila,Pakistan.
*Corresponding Author: Farooq Ali. Email: farooq.ali@uettaxila.edu.pk

Received: July 26, 2022 **Accepted:** November 29, 2022 **Published:** December 29, 2022.

Abstract: The quick advancement of the internet facility and rapid uptake of social networking sites like Twitter and Facebook has led to the generation of an extensive amount of data never seen previously in the history of humanity. Users are producing and disseminating more content than ever because of the widespread use of digital platforms, some of which are false and spreading wrong narratives. Accurate categorization of textual documents as fabrication or falsehood is a complicated job. The majority of the research emphasizes certain databases or topics, most notably the area of politics. As a result, the trained approaches perform effectively on a specific category of documents field and do not generate robust performance when evaluated to articles from other areas. We proposed a solution to the false news identification task in the presented study by combining NLP and machine learning technologies. In the first phase, the data is pre-processed by employing steps like removing null, duplicate, values, and punctuation marks. After this, the clean data is converted into numeric representation using a count vectorizer (CV) and Tf-Idf vectorizer. While for the classification task, we have used the SVM classifier. Our proposed solution performed well and catered to all possible fake news domains.

Keywords: NLP; Fake articles; Machine learning; SVM; Classification; Count vectorizer.

1. Introduction

The great revolution in machine learning (ML) and the internet has linked people socially around the globe with the help of various social sites like Facebook, Twitter, etc. Such platforms have caused a massive increase in the amount of generated data which is never observed before [1]. In addition to other applications, media organizations benefited from the extensive usage of social websites by giving their customers access to up-to-date news almost instantly [2]. News outlets, magazines, and periodicals gave way to digital news sources, websites, social feeds, and other online publishing mediums as the news channels changed. Users now have more access to the most recent information at their fingertips [3, 4]. 70% of visitors to newspaper sites come from Facebook recommendations. In the present era, these social networking sites are very effective and helpful for enabling people to talk, exchange, and start discussions about politics, schooling, and healthcare [5]. However, certain groups also exploit these channels negatively, frequently to obtain financial advantage and occasionally to influence public opinion, affect people's attitudes, or propagate irony or ridiculousness [6]. Fake news is the term used to describe the such situation. In the past ten years, there has been a sharp rise in the dissemination of misinformation, which was most visibly seen during the 2016 US elections [7]. Numerous issues have arisen due to the wide spreading of untrue information online, affecting not only political but also entertainment, healthcare, and research areas [8]. The stock markets are among such arenas where such news impacts because a rumor can have terrible results and even stop the trade in its tracks.

To cope with the issues of fake news propagation among people, researchers have proposed several approaches to recognize real and fake news correctly. One such method is presented in [9], where a model named the FNDNet is proposed that [1] uses the Kaggle fake news repository and shows better categorization performance for ML classifiers from the historical approaches. In this work, the technique emphasizes exclusively the vector space key points where the work used the GloVe approach for encoding the sentences in a feature descriptor of size 100 as an input for FNDNet architecture. The FNDNet architecture follows an attuned Convolutional Neural Network (CNN) network by joining three consecutive convolutional layers accompanying dense layers. The work shows better fake news recognition ability. However, suffering from a high computational cost. For the future, they suggest the use of multi-model word embeddings. Another approach to detecting fake news is presented in [10] that is compared with the Logistic Regression (LR), Naive Bayes (NB), SVM, Random Forest (RF), and Deep Neural Network (DNN) classifiers. They used the standard text pre-processing from the natural language processing (NLP) domain (such as stemming, removal of stop words, etc.) to perform experiments on the LIAR dataset. As a result, they confirmed that deep neural network-based techniques outperform traditional machine learning methods. Another deep learning-based approach is proposed by [11] that is also compared with various models using vector space representation as TF-IDF. They implemented CNN, LSTM, NB, Decision tree, RF, and K-Nearest Neighbours (KNN). The performance of the algorithms decreased respectively to the mentioned order. The best performance was reached when combining CNN and LSTM confirming the findings of good performance of deep learning models. For the data, they used a combination of Kaggle datasets.

Another approach using the idea of Ensemble frameworks is presented in [12], in which they used vector space representation and stylometric key points. In this case, the stylometric key points are distributed into 3 different feature subdivisions. The first one contained several distinctive instances, complications, the number of Gunning-Fog index, characters containing whitespaces, Flesch-Kincaid readability value, etc. The second subgroup contains key points from the employed database that are further distributed into given phases: Amount, Terminology, Syntax, Flesch-Kincaid value, Vagueness, etc. The third phase consisted of a write-print feature set for authorship in short manuscripts with features: Character, phrase, grammatical, Organizational, and content. For vector space representation, the work used several options bag-of-words (BoW), TF-IDF, continuous BoW that estimates the word from the available context utilizing neural network (NN), and skip-gram (SG) that also forecasts the coming related word, both using Word2Vec and FastText tools. Feature selection is used on both types of features. Stylometric features are selected by re-cursive elimination of the weakest key points. In word vector space features, vocabulary is lessened by lemmatization and stopping together with Chi-square tests for key points nomination. As classifiers, several conventional ML classifiers are used to perform the categorization task with the employment of the boosting approach to optimize the behavior of classifiers. The best overall accuracy is received with the Gradient boosting using CBoW Word2Vec embeddings, outperforming all the non-ensemble machine learning algorithms. However, it is noteworthy that CBoW representation improved the performance of non-ensemble algorithms.

Information credibility on Twitter is assessed by an approach presented in [13] that focuses on the trustworthiness of the information spread over the social network. The work [5] defines social networking trustworthiness as a characteristic of content that can be verified only utilzing information from social media. In [14] the focus on the use of cognitive psychology to determine the propagation of misrepresentation on social websites is performed. In the algorithm, they segregate retweets and draw a retweet graph for each of the sources. Then calculates, the Gini index for the measurement of dispersity and for the pattern of the retweets for the given source. A low Gini index (below 0.5 thresholds) means low reliability of the source. Lastly, they compute the acceptability score employing the PageRank algorithm, and it is looked at as true news if it is above the threshold.

Mykhailo Granik et al. [15] proposed an efficient model to classify fake news employing the naive Bayes approach. This approach is implemented as a software application and evaluated on a data set containing news from Facebook social websites regarding the domain of politics. The data is collected from 3 large Facebook communities, along with 3 nominative popular political news websites i.e., BBC news, Politico, and CNN, respectively. The work has attained a classification result of about 74%, which requires further improvements. The major reason for the decline of classification results is due to the imbalance data samples issues of the employed dataset, as only 4.9% of samples are manipulated. Another approach is discussed in [16], which presents a model employing various ML approaches that tackle different issues containing accuracy deficiency, time interval (BotMaker), and huge computational complexity to deal with extensive set of news in minimum time. Initially, a set of 400,000 news samples are taken from the HSpam14 dataset containing data from political area. The taken samples are classifier intro real and fake classes. In addition to the Top-30 terms from the Bag-of-Words approach that offer the largest performance gain, the work also deduced several lightweight properties. The model outperformed the prior approaches by about 18% and reached an accuracy of 91.65%.

After a thorough investigation of the literature, here we can formulate our problem statement of this research: Both supervised and unsupervised learning approaches have been used numerous times in the existing fake news database to classify the material. However, a large portion of the study concentrates on particular databases or subjects, most prominently the field of politics. As a result, the trained approaches perform better on a certain type of document's areas and do not produce effective classification scores when tested on documents from other areas. Because each domain's articles have a distinct written format, it is challenging to train a general approach that performs well on all types of news articles. In this study, we present a framework for false news identification by combining NLP and machine learning technologies. After performing the pre-processing step, we computed features by using two techniques named the count vectorizer (CV) and Tf-Idf vectorizer. The extracted features are used as input to the SVM classifier to perform the classification task. We have confirmed through the performed experiments that the CV features with the SVM classifier are robust to recognize real and fake news. Koloski et al. [17] proposed an approach for fake news classification by exploring how different document representations, ranging from simple symbolic bag-of-words to contextual, neural language model-based ones can be used for efficient fake news identification. The work concludes that knowledge graph-based representations alone are too generic for jobs where short texts are the primary input type. However, it has been demonstrated that adding more contextualized and analytical information about these texts improves performance.

The remaining manuscript is structured as follows: Section 2 comprises the description of the employed datasets and proposed approach, while the detailed experimentation analysis is discussed in Section 3. At the same time, the conclusion is discussed in Section 4.

2. Materials and Methods

With the ascent of misinformation and fake news flooding the web, it is exceptionally hard to figure out what is valid and what is not. Fake news is one of the significant threats plaguing the world order. Therefore, two datasets having information about fake news are found and used in this study. Selecting the proper methodology is one of the basic aspects of concluding the success of prediction analysis. This is because a proper assortment for the method of analysis will result in a system that is specification oriented and does not cause the disorder. Hence, initially, we have converted the employed data into numeric representation using count vectorizer (CV) and Tf-Idf vectorizer. While for the classification task, we have nominated the SVM classifier. We have performed a huge experimental analysis compared to other ML approaches like RF, LR, and DT that address fake news prediction tasks. This study has explored the ML methods used to identify the predictors that ensure the suitability of news. It has been found that the major issue faced by researchers with an abundance of information is time constraints. Additionally, because the process is data-driven, this method would make it easier to replicate for studies in other areas. Fig. 1 shows the research methodology used for this study.

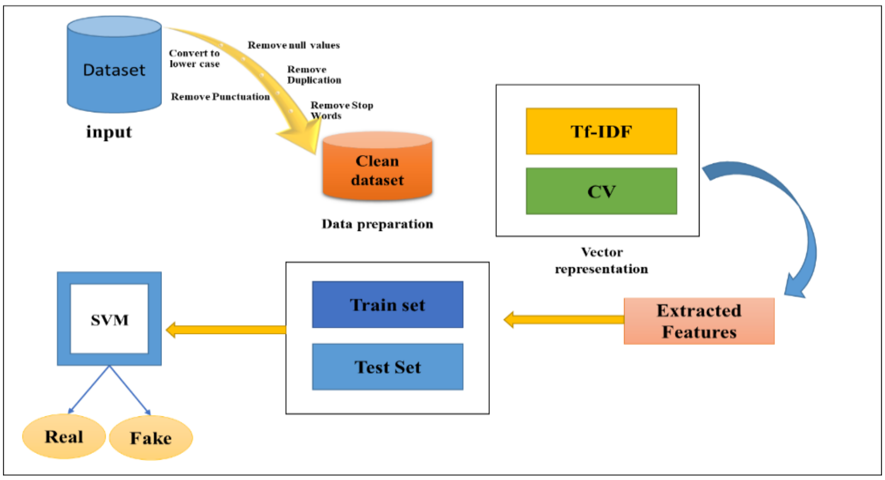

Figure 1. Methodology Workflow Process

2.1. Material

Two different datasets named the Fake News Kaggle and ISOT Fake News databases, are used in this study. Table 1 shows the characteristics of each dataset.

Table 1. Datasets details

Fake News Kaggle Dataset	ISOT Fake News Dataset
This repository comprises following characteristics: identifier: a distinct id to identify each news document title: indicating the topic of each news document author: indicating authorship of each news document text: indicating the content of each news document label: indicating the class associated with each news document 1: unreliable 0: reliable Samples count = 25113	The ISOT Fake News repository comprises thousands of real and manipulated documents collected from various legal news websites and social networking posts. Samples count = 6335

2.2. Data preparation

We have applied several pre-processing steps over the data in this phase to make it more appropriate for the classification task. Pre-processing is essential in enhancing text categorization's effectiveness through machine learning. Its phases involve removing null, punctuations, stop words, and duplicate values from the datasets. Moreover, we have converted all lower case values to upper case to make the sentence structure more effective for key points nomination. A visual description of all steps in the pre-processing stage is given in Fig. 2.

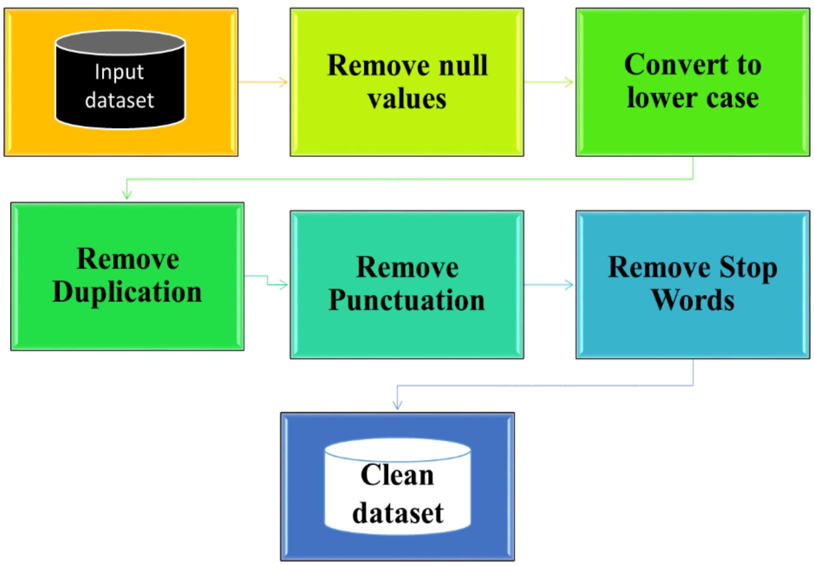

Figure 2. Pre-processing steps

2.3. Feature selection

Data cleaning, the key points from the database appropriate for the classifier are extracted and utilized to train and test the model for this research. A classifier can only comprehend numerical data; hence, it must transform text expression into numerical data representation. Several approaches are available to transform the textual data into a numeric format. For our case, we have nominated the count vectorizer (CV) [18] and Tf-Idf vectorizer [19] for the said purpose due to their high effectiveness in the area of NLP.

2.3.1. CV

CV, also known as term frequency, establishes the token's rate value. It records every distinct token's frequency in the document; the larger the token's score, the more frequently it appears in the text.

2.3.2. Tf-Idf

Tf-Idf, also known as frequency counts and inverted word count, is a subsequent stage in the vector data representation. With the uniqueness of the tokens into consideration, it establishes the occurrence of a token in the text. For instance, Tf-Idf grants a smaller score to a token if it appears more frequently across all texts. The tokens have a greater value which is more common in some texts than others.

2.4. Classification

To perform the classification task, we have chosen the SVM [20] classifier because of its ability to tackle the over-fitted training data better. The SVM falls under the category of supervised learning models. As the proposed problem belongs to the supervised classification model. Consequently, the choice of SVM for the fake news classification system is natural. As shown in Fig. 3, the SVM generates a hyper line when a set of labeled data is given for classification, and new examples are generated from it. SVM is used when the data is linearly separable. SVM tries to maximize the margin. In this work, the linear kernel is used. The discriminant function of the hyperplane can be described by Eq (1).

$$g(x) = w^T x + b \tag{1}$$

Where x is the input and w is the coefficient vector. If the points in SVM $g(x) \geq 0$, then the points belong to one class, and if the points $g(x) < 0$, then it belongs to another class.

$$J(w) = \frac{1}{2} \| w \|^2 \tag{2}$$

In linear SVM, $(2/||w||2)$ should be used for the minimization of the cost function. $y_i(w^T x_i + b) \geq 1$, $i = 1,2,3 \ldots n$ and $y_i = \{-1, +1\}$ represent class labels. In quadratic optimization function with the set of linearly inequality constraints. From Karush-Kuhn-Tucker conditions, the Lagrange function is defined as follows:

$$L_p(w,b,a) = \frac{1}{2} \| w^2 \| - \sum_{i=1}^{N} a_i \{ y_i (w^T x_i + b) - 1 \} \tag{3}$$

Where a_i are Lagrange multipliers, and L_p should be minimized to find optimal w and b. The optimization function can be defined as follows:

$$Maximize \left(\sum_{i=1}^{N} a_i - \frac{1}{2} \sum_{i,j=1}^{N} a_i a_j y_i y_j x_i^T x_j \right) \tag{4}$$

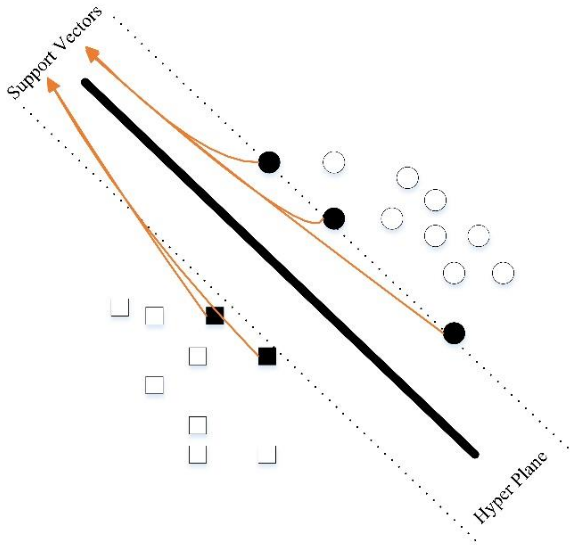

Figure 3. Support vector machine separates hyperplane

3. Results

In this chapter, the attained results with both feature descriptors are discussed. Moreover, the complete process that is used to apply the ML algorithm with attained features from datasets is explored. Further, the outcomes of comparison against each algorithm are discussed.

3.1. Evaluation Parameters

The used approach is evaluated using a variety of performance assessment criteria. The different evaluation factors for text categorization are used in this research. Classification measuring parameters like Recall, Accuracy, AUC, Precision and F-Measure, etc., are used as these measures are widely used by numerous investigators for classification.

3.1.1. Accuracy

The accuracy metric [21] is utilized to determine how many of the classification algorithm's predictions are accurate. However, this metric is not very convincing when applied to an unbalanced sample because it makes incorrect predictions for the other classes while making biased predictions for the high-frequency group. Accuracy is defined as:

$$\text{Accuracy} = \frac{\text{True}_{+} + \text{True}_{-}}{\text{True}_{+} + \text{True}_{-} + \text{False}_{+} + \text{False}_{-}} \quad (5)$$

3.1.2. Precision

An approach's precision metric (Nazir et al., 2021) is measured as constancy and replication, demonstrating how frequently repeated assessments under the same conditions yield identical results. It is mathematically defined as:

$$\text{Precision} = \frac{\text{True}_{+}}{\text{True}_{+} + \text{False}_{+}} \quad (6)$$

3.1.3. Recall

The portion of the total amount of relevant instances that are really retrieved is known as recall [22]. In this way, the organization and percentage of relevancy affect recall and precision. The recall is typically used to assess the smaller group's coverage in extreme learning cases. It is mathematically defined as:

$$\text{Recall} = \frac{\text{True}_{+} + \text{True}_{-}}{\text{True}_{+} + \text{False}_{-}} \quad (7)$$

3.1.4. F-measure

Both recall and precision cannot provide the whole performance results on their own. There can be excellent precision but terrible recall, or perhaps poor precision but amazing memory. So, F-measure provides a method for conveying the two issues using just one score. Recall and precision can be combined to figure out the F-Measure once they have been calculated for a 2- or multi-classification problem. [23]. It is mathematically defined as:

$$\mathrm{F-measure} = \frac{2*\mathrm{precision}*\mathrm{recall}}{\mathrm{precision}+\ \mathrm{recall}} \quad (8)$$

3.2. Experimental Setup

The proposed approach is implemented in an Intel core i5 CPU @ 1.80GHz, 64-bit operating system, the x64-based processor machine with windows 10. The coding is performed in Python (version 3.6.5) using the Scikit-learn library. The Scikit-learn is a publically available library to accomplish ML-based tasks. We used the Scikit-learn unit with NumPy, a Python module used for scientific programs. The database was distributed in a ratio of 7:3 for the train and test splits. 70% of the samples are used for framework training, while the remaining 30% are used for model testing.

3.3. Assessment of the presented model

The experiments are accomplished to assess the performance of the presented work. In this section, we have discussed the performance results of the SVM classifier for both feature descriptors, namely the CV and Tf-Idf vectorizer, over both employed datasets. Initially, the results of the proposed approach for the Tf-Idf vectorizer are presented in Table 2. Results in terms of accuracy, recall, precision, and F1-score with justification against each dataset are given in Table 2. It depicts that when NLP features are extracted, and TF-Idf features are used to detect fake news, the highest accuracy is 92.48%. with an F1-Score of 90.04%. Further, it is quite visible from the results that the NLP performed effectively on large datasets, i.e., the Kaggle dataset having more than 25K instances. Using the NLP algorithm, the Accuracy, Precision, Recall, and F1-score of 87.22%, 88.02%, 81.32%, and 84.62%, respectively, for Fake News Kaggle Dataset have been achieved. Similarly, for the ISOT Fake News database, the Accuracy, Precision, Recall, and F1-score of 92.48%, 92.52%, 91.83%, and 90.04%, respectively, have been achieved.

Table 2. Results of Tf-Idf vectorizer with the SVM classifier

Performance	Fake News Kaggle Dataset	ISOT Fake News Dataset
Accuracy (%)	87.22	92.48
Precision (%)	88.02	92.52
Recall (%)	81.32	91.83
F1-Score (%)	84.62	90.04

Similarly, for the CV features along with the SVM classifier, the attained performance results over both datasets are given in Table 3. The values shown in Table 3 are present that the proposed technique with CV features performs robustly for both datasets. More clearly, for the Fake News Kaggle Dataset, the proposed approach has attained a classification accuracy of 89.66%, which is 99.57% for the ISOT Fake News Dataset, which clearly depicts the effectiveness of the CV features for the fake news classification task.

Table 3. Results of CV features with the SVM classifier

Performance	Fake News Kaggle Dataset	ISOT Fake News Dataset
Accuracy (%)	89.66	99.57
Precision (%)	91.06	99.45
Recall (%)	84.23	99.67
F1-Score (%)	87.51	99.56

3.4. Comparative analysis

In this section, we have performed the result analysis of the proposed approach with other ML classifiers. As the work has attained the best results with the CV feature descriptor, therefore, for comparison, we have considered it only with the SVM classifier. To perform a fair comparison, we have chosen state-of-the-art ML classifiers, namely the LR [24] and RF [25], and attained values are elaborated in Table 4. The value in Table 4 indicates that our method outperforms other ML classifiers for both datasets. The LR achieved 58.74% accuracy, RF obtained 62.24% accuracy, and SVM showed 89.66% accuracy using Fake News Kaggle Dataset. For ISOT Fake News Dataset, LR has obtained 99.12% accuracy, SVM has shown 99.57% accuracy, and RF has achieved 99.27% accuracy.

Similarly, other evaluation metrics are given in detail in Table 4. Fig. 4 shows the graphical view of accuracy achieved by algorithms for both datasets. It is quite evident from the scores given in Fig. 4 that our approach is more effective in fake news classification as compared to other ML classifiers and exhibited performance gains of 29.17% and 0.375% in terms of the accuracy scores over the Fake News Kaggle, and ISOT Fake News datasets respectively. The major reason for the better fake news classification performance of the proposed approach is that the comparative classifiers are suffering from the issues of vanishing gradient and result in model overfitting. In comparison, the SVM classifier can better tackle such issues and shows better classification results.

Table 4. Comparison with other ML classifiers

Performance analysis	Fake News Kaggle Dataset			ISOT Fake News Dataset		
	LR	RF	SVM	LR	SVM	RF
Accuracy (%)	58.74	62.24	89.66	99.12	99.57	99.27
Precision (%)	53.95	68.69	91.06	99.14	99.45	99.18
Recall (%)	28	22.5	84.23	99.08	99.67	99.01
F1-Score (%)	36.87	33.8	87.51	99.10	99.56	99.47

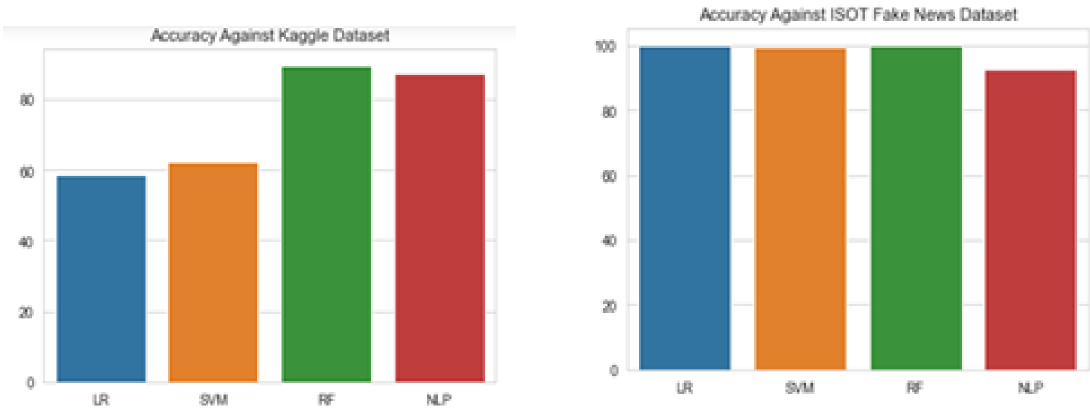

Figure 4. Accuracy comparison of ML algorithms

4. Conclusion

Computer-aided categorization of textual documents as real or disinformation is complex. Even experts from various domains have to explore various characteristics before deciding on the documents' reliability. Several researchers have employed conventional ML, either supervised or unsupervised approaches, and assessed to recognize the original and fake news articles. In this research, we have analyzed the previous methods for fake news identification and found that most of the techniques are only used for political domain datasets of fake news prediction. We have used the datasets having news about all possible aspects/domains. Moreover, we have compared the Tf-Idf and CV vectorizers with the SVM classifier for fake news prediction. We have attained the best results with the CV feature descriptor and the SVM classifier. We have performed a huge performance analysis on two challenging datasets, the Fake News Kaggle and ISOT Fake News datasets, to depict the proposed framework's efficacy for categorizing fake news. In future work, we plan to test other ML classifiers for fake news classification for further results enhancement.

Funding: This research received no external funding from any organization for this work.

References

1. S. Mishra, P. Shukla, and R. Agarwal, "Analyzing machine learning enabled fake news detection techniques for diversified datasets," Wireless Communications Mobile Computing, vol. 2022, 2022.
2. J. Cui, K. Kim, S. H. Na, and S. Shin, "Meta-Path-based Fake News Detection Leveraging Multi-level Social Context Information," in Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022, pp. 325-334.
3. M. Masood, M. Nawaz, A. Javed, T. Nazir, A. Mehmood, and R. Mahum, "Classification of Deepfake videos using pre-trained convolutional neural networks," in 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2), 2021, pp. 1-6: IEEE.
4. M. Masood, M. Nawaz, K. M. Malik, A. Javed, A. Irtaza, and H. Malik, "Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward," Applied Intelligence, pp. 1-53, 2022.
5. Y. Zhu et al., "Memory-guided multi-view multi-domain fake news detection," IEEE Transactions on Knowledge Data Engineering, 2022.
6. S. Garg and D. K. Sharma, "Linguistic features based framework for automatic fake news detection," Computers Industrial Engineering, vol. 172, p. 108432, 2022.
7. S. Raza, C. J. I. J. o. D. S. Ding, and Analytics, "Fake news detection based on news content and social contexts: a transformer-based approach," International Journal of Data Science Analytics, vol. 13, no. 4, pp. 335-362, 2022.
8. P. M. S. Freire, F. R. M. da Silva, and R. R. Goldschmidt, "Fake news detection based on explicit and implicit signals of a hybrid crowd: An approach inspired in meta-learning," Expert Systems with Applications, vol. 183, p. 115414, 2021.
9. R. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "FNDNet- A Deep Convolutional Neural Network for Fake News Detection," Cognitive Systems Research., 2020.
10. C. K. Hiramath and G. C. Deshpande, "Fake News Detection Using Deep Learning Techniques," 2019 1st International Conference on Advances in Information Technology, pp. 411-415, 2019.
11. R. K. Kaliyar, "Fake News Detection Using A Deep Neural Network," 2018 4th International Conference on Computing Communication and Automation (ICCCA), pp. 1-7, 2018.
12. H. Reddy, N. Raj M, and M. Gala, "Text-mining-based Fake News Detection Using Ensemble Methods," International Journal of Automation and Computing., 2020.
13. C. Castillo, M. Mendoza, and B. Poblete, "Information Credibility on Twitter," in Proceedings of the 20th International Conference on World Wide Web, Hyderabad, India, 2011, pp. 675-684.
14. K. P. Kumar and G. Gopalan, "Detecting misinformation in online social networks using cognitive psychology," Human-centric Computing and Information Sciences, 2014.
15. M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," in 2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON), 2017, pp. 900-903: IEEE.
16. U. Sharma, S. Saran, and S. M. Patil, "Fake news detection using machine learning algorithms," International Journal of Creative Research Thoughts, vol. 8, no. 6, pp. 509-518, 2020.
17. B. Koloski, T. S. Perdih, M. Robnik-Šikonja, S. Pollak, and B. Škrlj, "Knowledge graph informed fake news classification via heterogeneous representation ensembles," Neurocomputing, 2022.
18. T. Turki and S. S. Roy, "Novel Hate Speech Detection Using Word Cloud Visualization and Ensemble Learning Coupled with Count Vectorizer," Applied Sciences, vol. 12, no. 13, p. 6611, 2022.
19. S. Wehnert, V. Sudhi, S. Dureja, L. Kutty, S. Shahania, and E. W. De Luca, "Legal norm retrieval with variations of the bert model combined with tf-idf vectorization," in Proceedings of the eighteenth international conference on artificial intelligence and law, 2021, pp. 285-294.
20. M. Nawaz et al., "Melanoma localization and classification through faster region-based convolutional neural network and SVM," Multimedia Tools Applications, vol. 80, no. 19, pp. 28953-28974, 2021.
21. M. Nawaz et al., "Melanoma segmentation: A framework of improved DenseNet77 and UNET convolutional neural network," International Journal of Imaging Systems Technology, 2022.
22. S. Albahli and M. Nawaz, "DCNet: DenseNet-77-based CornerNet model for the tomato plant leaf disease detection and classification," Frontiers in Plant Science, vol. 13, 2022.
23. M. Nawaz, T. Nazir, and M. Masood, "Glaucoma detection using tetragonal local octa patterns and SVM from retinal images," Int. Arab J. Inf. Technol., vol. 18, no. 5, pp. 686-693, 2021.
24. R. E. Wright, "Logistic regression," 1995.
25. G. Biau and E. Scornet, "A random forest guided tour," Test, vol. 25, no. 2, pp. 197-227, 2016.