

COVID-19 Fake News Detection Using Machine Learning Techniques: A Comparative Study

Amor Ben El Khettab Lalmi, Abderrahim Djaballah and Mohamed Gharzouli*

Faculty of New Information and Communication Technologies, Abdelhamid Mehri Constantine 2 University, Constantine, Algeria

Abstract

Fake news has become one of the most serious issues in recent years, especially on social media. For example, during the covid-19 pandemic, a great deal of false information about the virus spread easily and quickly through the internet. In this area, researchers have given substantial answers to this problem utilizing various machine learning techniques. However, there are some gaps that need to be clarified. In the context of COVID-19 fake news detection, in this study, we present a comparison of four major machine learning algorithms: SVM, Nave Bayes, Logistic Regression, and Random Forest. We proposed four new machine learning models by combining these algorithms with two feature extraction techniques (TF-IDF and CountVectorizer). On three datasets, we tested the suggested models and analyzed their performance. According to the obtained results, we concluded that some properties of the used datasets can affect the obtained results. In addition, we find the best model overall.

Received on 05 May 2022; accepted on 19 June 2022; published on 24 June 2022

Keywords: Fake News Detection, COVID-19, Machine Learning, Artificial Intelligence, Natural Language Processing

Copyright © 2022 A.B.K Lalmi *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi:10.4108/eai.24-6-2022.174229

1. Introduction

Humans are born with the ability to convey their thoughts and feelings. Nowadays, with the use of social media, people can easily access the opinions of others. However, in parallel, the popularization of the Internet has facilitated the spread of fake news. This phenomenon is a crisis history of human inter-interaction even before the apparition of information and communication technologies. In human society, fake news creates conflict, discord, and understanding. This problem is gaining new traction in the era of digital communication and social networks. Generally, it becomes a big challenge when there are social and economic problems in a country or around the world [18]. Unfortunately, through health crises, fake news might have catastrophic consequences. During the coronavirus pandemic, online fake news (infodemic) regarding Covid-19 has caused many deaths worldwide. At the same time, numerous

individuals have been admitted to hospitals because of a single piece of fake news [17]. As a result, screening COVID-19 misinformation (online opinions) has become a need. In this context, several solutions were proposed to solve this purpose. Techniques of artificial intelligence like natural language processing and machine learning algorithms are applied for a crucial classification of fake news on social media [1], [16]. However, there are some gaps in the reported studies. According to a variety of commonly used classifiers and accompanying feature extraction methods, the best machine learning algorithm for detecting fake news has yet to be determined. Furthermore, the impact of some quantitative and qualitative proprieties of the used datasets is rarely examined in research papers. In our approach, we studied four main machine learning techniques to detect fake news: Random Forest (RF), Logistic Regression (LR), Naive Bayes (NB) and Support Vector Machine (SVM). We analyzed their performance by using parameters such as F1 score, recall, precision, and accuracy. In addition, we have chosen three datasets about the COVID-19 topic. The choice criteria were the size of the dataset, the length of the contained reviews

*Corresponding author. Email: mohamed.gharzouli@univ-constantine2.dz

and the diversity of data. The main objective of this study is to determine the best machine-learning model for fake news detection. In addition, we want to show how some properties of the used datasets can influence the obtained results and the performance of the studied models.

The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 presents our contribution. Section 4 shows the experimental results and the discussion. Finally, Section 5 concludes this study.

2. Related Work

Fake news is becoming more prevalent every day. People are quick to believe the false information and begin discussing it. The more they hear about fake news, the more likely they will believe it. Therefore, detecting false news is a vital step in preventing the spread of fake news. As a result, many intelligence algorithms are used to detect fake news. These algorithms are also utilized to solve a lot of real problems like sentiments analysis of users in social media, spam detection, in the domain of e-learning and other fields [3]. In the following, we will mention some of the techniques used for fake news detection. The bellow presented related work be divided into traditional machine learning-based detection and deep learning-based detection.

The authors of [2] proposed an approach of sentiment analysis based on different machine learning and deep learning models. The purpose of this study is to detect coronavirus misinformation on Twitter. After processing the machine learning-based models, it obtains that K-nearest neighbour, multilayer perceptron and random forest are the best classifiers through the four metrics: accuracy, precision, recall and F1 score. Furthermore, the deep learning-based classification algorithms LSTM, BiLSTM, GRU, RNN, and CNN are extremely accurate in categorizing fake news on a specific topic in terms of execution time and speed, as measured by the same four metrics. To study the dissemination of covid-19 fake news consequences on the psychology of Moroccan people, and to protect individuals from this fake news on Twitter, the researchers in [3] have used Apache Spark to implement a new technique by using machine and deep learning algorithms. After testing the different algorithms, which are LR, DT, RF, NB, GBT, LSTM, and MLP, the authors have obtained that the machine learning RF algorithm is the best through its accuracy has an estimated ratio of 0.79, so they have used this algorithm to classify the new tweets about Covid-19 topic Fake news has become a big concern on social media as a result of its widespread distribution and the difficulties of managing many accounts operated by humans or

robots. Therefore, the authors of [4] proposed an unsupervised and domain-independent approach to analyzing factual and emotional tweets. According to the researchers' contribution, they analyzed tweets connected to the epidemic and based on the results of their investigation, they discovered that the majority of the factual tweets are ascribed to reliable profiles. While the majority of emotional tweets come from accounts whose trustworthiness has yet to be established. To investigate the impact of fake news on government and society, the authors of [5] proposed a model that employs a deep learning framework that combines neural networks and long-term memory architecture to distinguish fake from real news. In addition, they used a pre-processing model to improve the performance of the model that can be built. After the comparison of the proposed technique with other methods for the detection of fake news, they concluded that the model LSTM is the best one that achieves a good result. The proposed model is evaluated by the four metrics: accuracy, precision, recall, F1 score and support measures.

Fake news propagation, short text processing, and language reliance, as well as acquiring information from web search engines, are all difficult tasks. To address this issue, the authors of [6] suggested a model of fake news identification that employs a new integration strategy based on link2vec. They make a comparison between their proposed model and the others that are used on the two datasets (English and Korean). First, they discovered that the suggested model combined with SVM has an accuracy of 0.93 for the English dataset. The performance of the two comparison models is 0.8 and 0.89, respectively, for the text-based model and the white list-based model integrated with ANN. In the case of the Korean dataset, the best performance (0.81) was demonstrated by the proposed model combined with ANN and the performances of the comparison models were 0.77 for (text-based model+ on a white list combined with an ANN) and 0.78 for (text-based model combined with an SVM). The authors of the article [7] did a conventional study to evaluate the performance of multiple machine learning algorithms applicable to three different datasets in the context of the dissemination of fake news. They discovered that the investigated deep learning algorithm (Bi-LSTM) outperforms the machine learning method (Naive Bayes). Furthermore, the collected findings showed that pre-trained language models (RoBERTa) outperform Bi-LSTM and RoBERTa outperforms the two techniques. With the use of online social media, fake news cause enormous distress to people's social life. So, the study conducted by the authors of the article [8], tried to find the most essential procedures taken to discover fake news. For this, they proposed

an approach entitled "domain-antagonistic and neural network model of graphic attention DAGA-NN)" to meet the challenge. Following that, they reach the following outcomes: The EANN "event adversarial neural network" outperformed the other approaches by a little margin (excluding DAGA-NN), however DAGA-NN still did well in identifying fake news. People are extremely exposed to false information due to social media and broadcasting organizations around the world, all of which have a negative impact on collective views and government policies. To solve this problem, Reza et al. [9] provided three classifiers with distinct pre-trained models for embedding news articles as input, as well as a system for fake news detection based on contextual text representation and deep neural classification. As a result of their investigation, these researchers arrived at the following conclusion: First, for the LIAR dataset, Funnel Transformer and RoBERTa outperform BERT embedding results, with Funnel-CNN achieving the top results in the competition. As a consequence, using the ISOT dataset, Funnel Transformer outperformed other textual representation models, and the models described in this paper outperform other methods. However, in the COVID-19 dataset, the CNN classifier outperformed SLP and MLP. Finally, the Gaussian noise layer improved the learning process of RoBERTa-CNN on the LIAR and COVID-19 datasets for Roberta-GN-CNN. Because the spread of fake news on social media has the potential to harm public opinion and social development, the authors of [10] have modelled real-world news time evolution patterns as a graph evolving within the context of continuous-time dynamic broadcasting networks. After analyzing their methods, they came up with the following results: on the different datasets of reference, TGNF surpasses all baselines in terms of accuracy and F1 score with statistical significance, but BiGCN exceeds all baselines in three datasets in terms of false news detection. They also obtained that TGNF performs better than its variants without the GCL module, and the TDN+TGNF module performs better than GCL+TDN on three datasets. In the paper [11], the researchers presented a Bi-LSTM-GRU-dense deep learning model based on a set of classifiers to classify news as fake or real using LIAR dataset. After experimentations, the results showed that the proposed model achieved an accuracy of 0.898, a recall of 0.916, a precision of 0.913 and an F-score of 0.914, respectively. Furthermore, the results of the proposed models are dominant when compared to previous studies for fake news detection using the LIAR dataset. In [12], the authors suggested a deep triple network (DTN) that uses knowledge graphs to detect misinformation, and they applied low-level and high-level feature extraction to classify the input news article and provide explanations for the classification. The outcomes of the

proposed approach are compared against DTN on the two datasets using three metrics: precision, recall rate, and F1-score. The comparative analysis was divided into two parts. First, they found that conventional text classification methods classify articles with good classification results compared to the TF-IDF+SVM method. In addition, it is observed that deep learning models, such as textCNN and textRNN perform significantly better than conventional text classification methods such as TF-IDF and SVM. Furthermore, feature-based attention networks make an interest that is the improvement of DTN performance. Second, they compare and contrast several DTN configurations: The first finding is that, when comparing several KG integration models, TransD is the best. Second, comparison investigations in various knowledge graphs show that DB4 is the finest of all the knowledge graphs evaluated. Third, the entity-based attention network (EAN) will be combined with the DTN paradigm. The experiment's findings suggest that the entity-based attention network is quite useful. Abdul Nasir et al. [13] proposed in their work a new hybrid deep learning model that combines convolutional and recurrent neural networks for the classification of fake news. They made a global synthesis via the comparison of their model with another one proposed by [Elhadad et al 2019a]. Therefore, the obtained results show that the proposed approach is comparatively better than the other method in both datasets ISOT and FA-KES. The authors of [14] created a manually annotated dataset of 10,700 social media posts and articles of true and false news about COVID-19. They performed a binary classification task (true or false) and evaluated the data with four machine learning algorithms: decision tree, logistic regression, gradient boost, and support vector machine (SVM). The results show that the SVM has the highest F1 score of 0.93, followed by logistic regression (LR) with an F1 score of 0.91. In [15], the authors compare various active learning strategies for different text classifiers, with a particular focus on Bayesian on various datasets. They have discovered that for the vast majority of tasks, the traditional dropout Monte Carlo strategy works well. The active strategy using Dropout MC and Deep Ensembles, on the other hand, achieved near-perfect performance for several datasets. The best results were obtained for the most recent embedding of RoBERTa. The authors of [18] presented an approach for fake news detection. They studied some machine learning techniques such as SVM, Naive Bayes and logistic regression. Their performances are analyzed using parameters such as F1 score, recall, precision, support and accuracy. After analyzing these different algorithms, they found that SVM and logic regression are the best performing classifiers compared to Naive Bayes.

3. Methods

In this section, we first introduce the datasets used in our study. Then we discuss the main steps needed for the system to operate. We present the data preprocessing and the different models (classifiers) that we investigated in this study. Finally, we describe the used libraries to develop the system.

3.1. Studied Datasets

In our study, we have chosen three datasets which are published in two repositories: Zenodo and Mendeley. They are described below.

Dataset1: COVID Fake News Dataset. The dataset [19] contains a list of COVID Fake News/Claims that have been widely circulated on the internet. Its content comprises headlines and outcomes. Where the shared headlines/facts are stored as string attributes and the outcome is binary data, with 0 indicating that the headline is false and 1 indicating that it is true.

Dataset2: Covid-19 News Dataset Both Fake and Real. This dataset [20] includes both fake and real news. There are 16898 distinct rows, indicating the number of news items. The dataset was created by combining two datasets: one from various CBC news sources (link: <https://zenodo.org/record/4722470>) and the other from various web portals (link: <https://zenodo.org/record/4282522>). The data set contains two columns: text that represents the news and outcome which is the status of the news (fake or real).

Dataset3: COVID-19 Fake News Dataset. This dataset [21] contains a collection of COVID-19-related true and false news. The news is collected from December 2019 to July 2020. Webhose.io was used to collect this data, which was then manually labelled. It is divided into three categories: false news, true news, and partially false news. Both partially false and fake news have been labelled 0 for classification purposes, whereas correct news has been labelled 1.

Table 1 shows the detailed statistics of these datasets.

3.2. System operating

After a deep study of the problem of fake news detection, and according to several related works (some examples are presented in section 2), we have concluded that the use of machine learning approaches to solve this problem, is very sufficient to give the best model performance to build with good predictions and exacting for the classification of the text during a short period. First, we implemented the four machine learning classifiers which are: «RF, SVM, LR and NB» on the datasets presented in the previous subsection. Our objective is to find the optimum method selection that

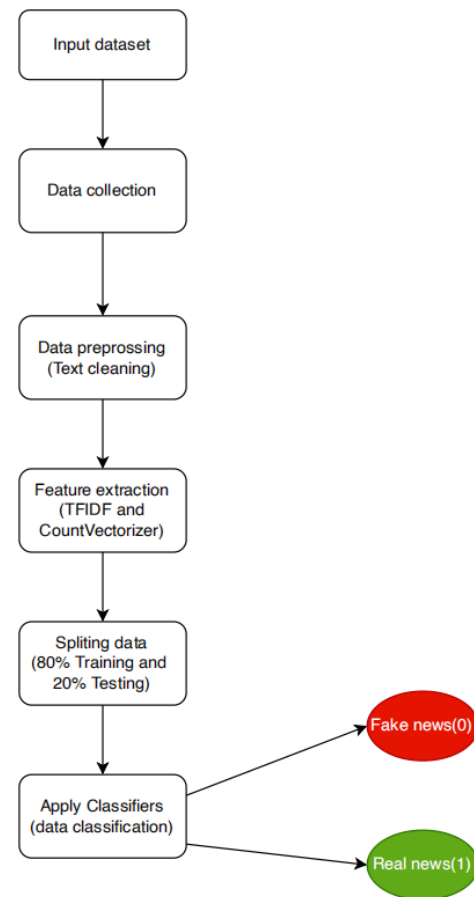


Figure 1. Architectural Diagram.

provides greater accuracy on these datasets. In addition, we have used Python as a language programming and we have employed different Python libraries to develop the proposed system. The latter (presented in Figure 1) follows the below steps:

Datasets input and data division. First, the dataset is entered using the Pandas library [24]. Therefore, the data is extracted from the dataset by separating it into two parts: the characteristic portion, which comprises textual information, and the label part, which contains binary values '0' and '1' or 'fake' and 'real'.

Data preprocessing. First, we use text cleaning to make minor adjustments to the collected data. This phase is divided into two steps:

- Elimination of symbols: symbols have been deleted via the 'Regular Expression (RE)' Python library.
- Elimination of stopwords, suffixes and prefixes: stopwords are removed by importing them from the 'NLTK.corpus' library [22], and the suffixes and prefixes are deleted from the 'Porter

Table 1. Properties of datasets.

Dataset	Total data	Fake news	Real news	Topic	Date
Dataset1	10201	9727	474	COVID-19	Nov 20, 2020
Dataset2	16989	9727	7262	COVID-19	Apr 27, 2021
Dataset3	3119	1058	2061	COVID-19	Feb 23, 2021

Table 2. Results of dataset 1.

	F1-score	Recall	Precision	Accuracy
(RF, CountVectorizer)	0.96	0.97	0.96	0.97
(SVM, TF-IDF)	0.96	0.97	0.97	0.97
(LR, TF-IDF)	0.95	0.96	0.96	0.96
(NB, CountVectorizer)	0.96	0.95	0.96	0.95

Table 3. Results of dataset 2.

	F1-score	Recall	Precision	Accuracy
(RF, CountVectorizer)	0.93	0.93	0.93	0.93
(SVM, TF-IDF)	0.95	0.95	0.95	0.95
(LR, TF-IDF)	0.94	0.94	0.94	0.94
(NB, CountVectorizer)	0.93	0.93	0.93	0.93

Stemmer' class that was imported from the 'NLTK.stem.porter'.

Table 4. Results of dataset 3.

	F1-score	Recall	Precision	Accuracy
(RF, CountVectorizer)	0.79	0.80	0.80	0.80
(SVM, TF-IDF)	0.80	0.81	0.81	0.81
(LR, TF-IDF)	0.79	0.80	0.80	0.80
(NB, CountVectorizer)	0.81	0.80	0.81	0.80

Feature extraction and data splitting. The Feature Extraction process allows the transformation of the textual data into numerical values (Vectorization). Following the literature review, we have chosen TF-IDF and CountVectorizer as the two extraction methods that produced the best results. Both TF-IDF and CountVectorizer are imported from the 'sklearn.feature_extraction.text' 'SK-LEARN' library [23]. The data splitting allows the division of the datasets into two parts: 80% training and 20% test. To develop this task, we have used the method 'train_test_split' of the same library [23].

Model construction (apply Classifier). In this step, we combined the chosen classifiers (LR, SVM, RF, NB) with the feature extraction methods. We applied TF-IDF with SVM and LR, and CountVectorizer with RF and NB. Then four obtained models are following: (SVM, TF-IDF), (LR, TF-IDF), (RF, CountVectorizer) and (NB, CountVectorizer).

4. Results and discussion

To evaluate the four models using the three selected datasets, each model was predicted based on the calculation of the four parameters (F1-Score, Recall, Precision and Accuracy). The obtained results for each dataset are shown in tables 2, 3 and 4. According to the results obtained from dataset 1 (shown in table 2), we can see that (RF, CountVectorizer) and (SVM, TF-IDF) give the best results with an accuracy of 0.97. However, from the outcomes of datasets 2 and 3 (presented in tables 3 and 4), we conclude that the model (SVM, TF-IDF) is the best one, with an accuracy of 0.95 and 0.81, respectively.

Furthermore, figures 2, 3, 4 and 5 display the performances (F1 score, recall, precision and accuracy) of each model with the three studied datasets. As a

first finding, we observe that the performances of each model decrease gradually from dataset1 to dataset3. Then, the four models perform well with dataset1, which has a medium size (total data of 10201) but contains a variety of data. This factor enables the optimization of the model's training. Furthermore, with the shortest dataset, the four models produce the worst results (dataset3 with a total data of 3119). The latter has a lack of variation in the contained data, and the majority of the texts are lengthy. Despite being the largest dataset, the diversity of information in the second dataset is average when compared to datasets 1 and 3. Then, we can conclude that the three elements (the size of the dataset, the diversity of the contained information and their lengths) have an impact on the obtained results of the applied machine learning algorithms.

Finally, to determine the best machine learning model that gives good performance on prediction in the textual classification of Covid-19 fake news, we compare the obtained results as shown in the figures 6, 7 and 8. Based on the results obtained with dataset1, we can conclude that the (RF, CountVectorizer) and (SVM, TF-IDF) models provide superior results than the other models. Furthermore, for dataset2, we can see that (SVM, TF-IDF) and (LR, TF-IDF) exceed (RF, CountVectorizer) and (NB, CountVectorizer). In addition, we can judge that the four models give almost the same results for dataset3. Finally, we find that (SVM, TF-IDF) is the best model overall, based on its excellent results with the three analyzed datasets.

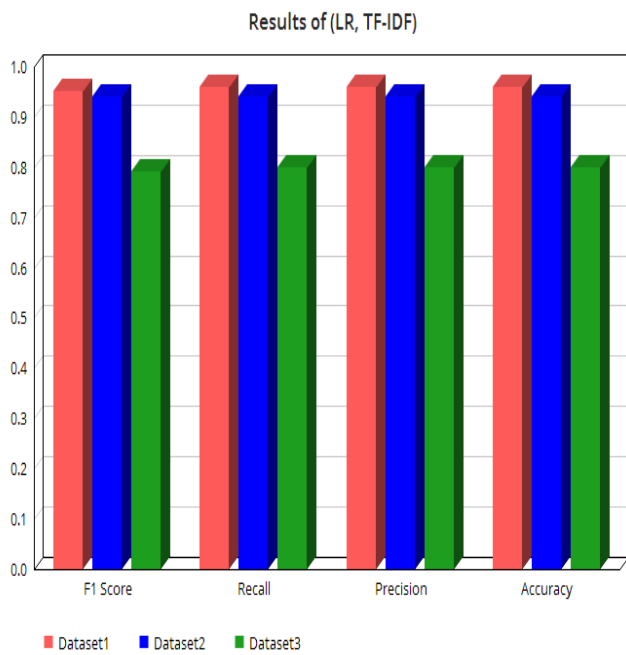


Figure 2. Results of (LR, TF-IDF).

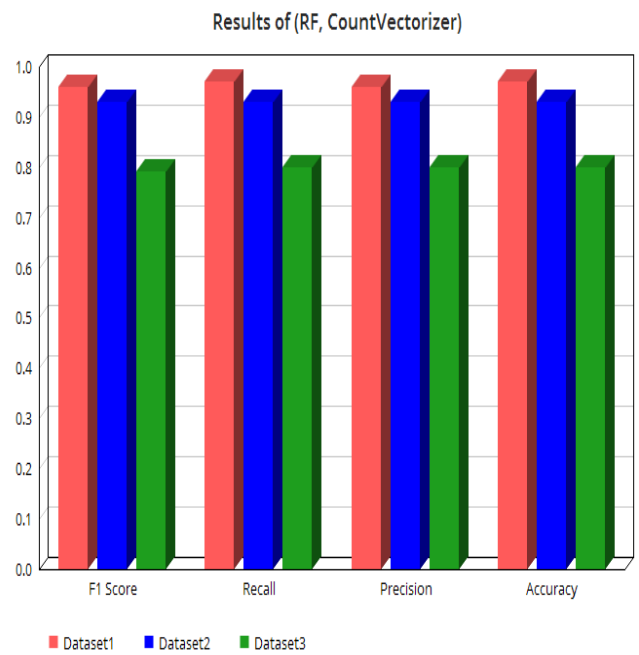


Figure 4. Results of (RF, CountVectorizer).

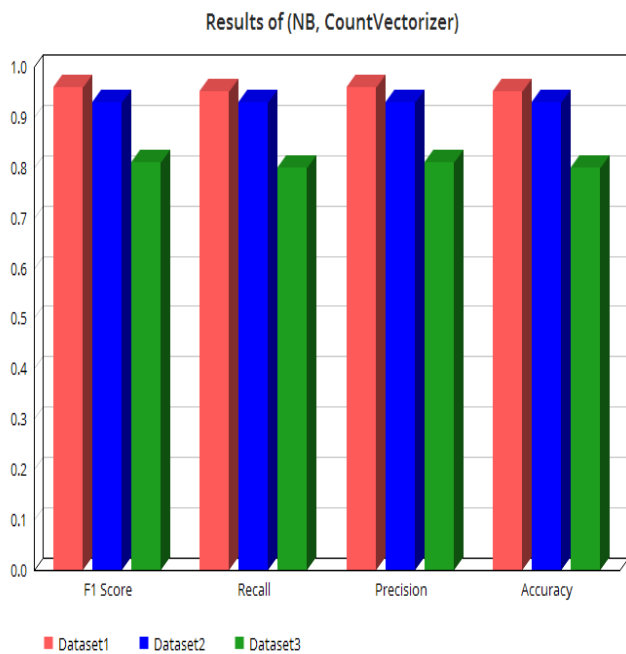


Figure 3. Results of (NB, CountVectorizer).

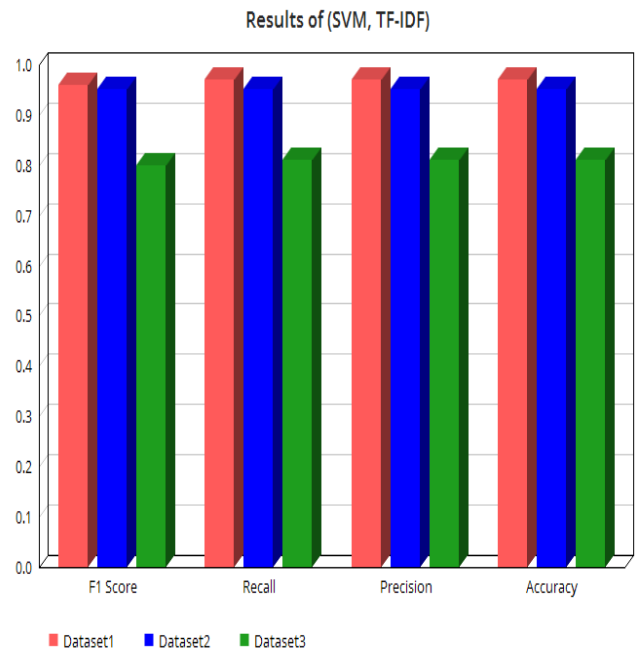


Figure 5. Results of (SVM, TF-IDF).

5. Conclusion

Many research works have suggested employing machine learning algorithms for fake news detection. However, there are some gaps that need to be clarified.

In this paper, we investigated the performances of four machine learning classifiers (Random forest, SVM, Nave Bayes, and Logistic Regression) for COVID-19 fake news identification. We combined these algorithms with two main feature extraction techniques (TF-IDF

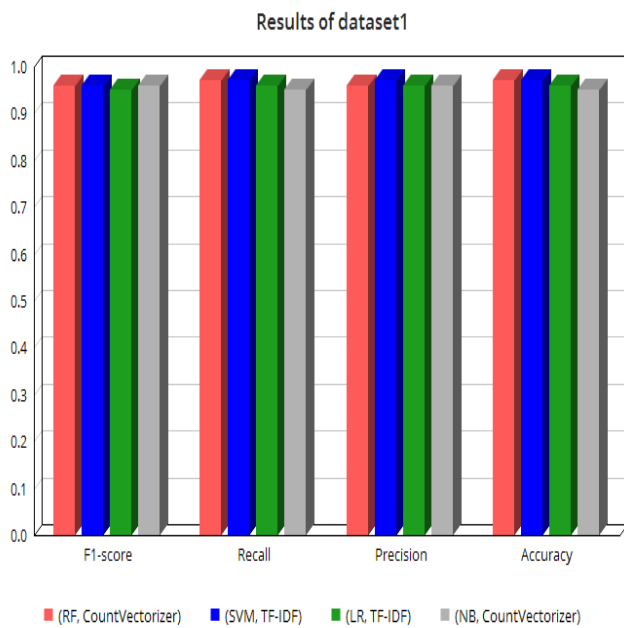


Figure 6. Results of dataset1.

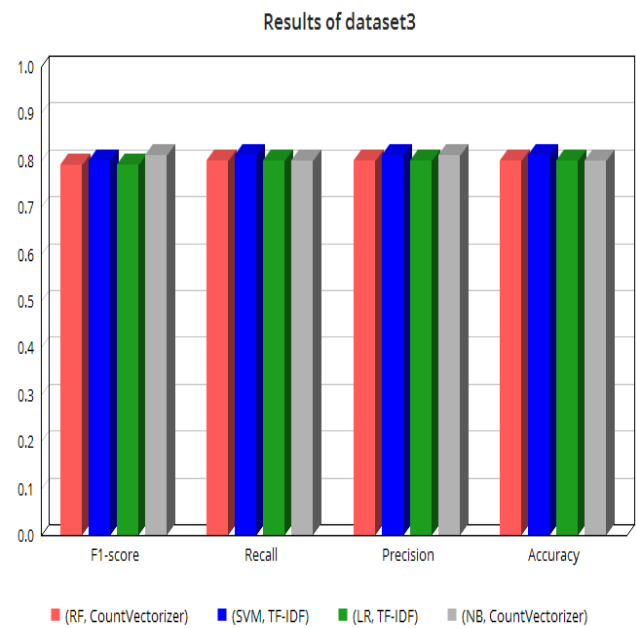


Figure 8. Results of dataset3.

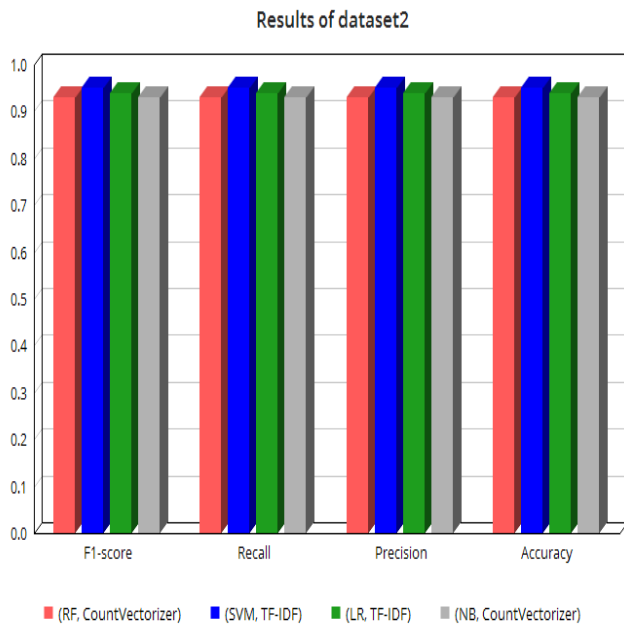


Figure 7. Results of dataset2.

and CountVectorizer) to propose four different machine learning models. In addition, we tested these models using three different COVID-19 datasets. These datasets are chosen based on three factors: the dataset’s size, the diversity of data it contains, and the length of

the texts. In our analysis, we discovered that all the models perform well with the dataset that contains diverse data and short texts. However, SVM with TF-IDF outperformed the other models on all three datasets, with the best recall, F1 score, precision, and accuracy. Finally, in the future, we may test deep learning algorithms, and we intend to present a hybrid strategy to detect online fake news that combines two or more machine learning techniques.

References

- [1] CARDOSO E.F, SILVA R.M. and ALMEIDA T.A. *Towards automatic filtering of fake reviews*. Neurocomputing 309 (2018) 106–116.
- [2] BANGYAL W.,QASIM R.,REHMAN N., AHMAD Z.,DAR H.,RUKHSAR L.,AMAN Z.and AHMAD J. *Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches*. Computational and Mathematical Methods in Medicine, 2021, (2021), 5514220, 1-14, doi:https://doi.org/10.1155/2021/5514220.
- [3] MADANI Y.,ERRITALI M.and BOUIKHALENE B. *Using artificial intelligence techniques for detecting Covid-19 epidemic fake news in Moroccan tweets*. Results in Physics ,25 (2021) , 104266,1-10 , doi:https://doi.org/10.1016/j.rinp.2021.104266.
- [4] ANSAR W. and GOSWAMI S. *Combating the menace: A survey on characterization and detection of fake news from a data science perspective*. International Journal of Information Management Data Insights ,1, (2021), 100052, 1-20, doi:https://doi.org/10.1016/j.jjimei.2021.100052.

- [5] CHAUHAN T.,M E.,PALIVELA H. and PHD. *Optimization and improvement of fake news detection using deep learning approaches for societal benefit*. International Journal of Information Management Data Insights ,1 ,(2021) ,100051 ,1-11 ,doi:https://doi.org/10.1016/j.jjime.2021.100051.
- [6] SHIM J.,LEE Y. and AHN H. *A link2vec-based fake news detection model using web search results*. Expert Systems With Applications, 184, (2021), 115491, 1-15, doi:https://doi.org/10.1016/j.eswa.2021.115491.
- [7] KHAN J.,KHONDAKER T.,AFROZ S. ,UDDIN G. and IQBAL A. *A benchmark study of machine learning models for online fake news detection*. Machine Learning with Applications ,4 ,(2021) , 100032, 1-12, doi:https://doi.org/10.1016/j.mlwa.2021.100032.
- [8] YUAN H.,ZHENG J.,YE Q. ,QIAN Y. and ZHANG Y. *Improving fake news detection with domain-adversarial and graph-attention neural network*. Decision Support Systems 1, 51, (2021), 113633, 1-11, doi:https://doi.org/10.1016/j.dss.2021.113633.
- [9] SAMADI M.,MOUSAVIAN M. and MOMTAZI S. *Deep contextualized text representation and learning for fake news detection*. Information Processing and Management ,58 ,(2021) , 102723, 1-13, doi:https://doi.org/10.1016/j.ipm.2021.102723.
- [10] SONG C.,SHU K. and WU B. *Temporally evolving graph neural network for fake news detection*. Information Processing and Management, 58 ,(2021) ,102712, 1-18, doi:https://doi.org/10.1016/j.ipm.2021.102712.
- [11] ASLAM N.,KHAN I.,ALOTAIBI F. ,ALDAEJ L. and ALDUBAIKIL A. *Fake Detect: A Deep Learning Ensemble Model for Fake News Detection*. Complexity ,2021 ,(2021) ,5557784 , 1-8, doi:https://doi.org/10.1155/2021/5557784.
- [12] LIU J.,WANG C.,LI C. ,LI N.,DENG J. and PAN J. *DTN: Deep Triple Network for Topic Specific Fake News Detection*. Web Semantics: Science, Services and Agents on the World Wide Web ,70 ,(2021) ,100646 , 1-18, doi:https://doi.org/10.1016/j.websem.2021.100646.
- [13] ABDUL NASIR J.,KHAN O. and VARLAMIS I. *Fake news detection: A hybrid CNN-RNN based deep learning approach*. International Journal of Information Management Data Insights 1, (2021), 100007, 1-13, doi:https://doi.org/10.1016/j.jjime.2020.100007.
- [14] PATWA P.,SHARMA S.,PYKL S.,GUPHA V. ,KUMARI G.,AKHTAR S.,EKBAL A.,Das A. and CHAKRABORTY T. *Fighting an Infodemic: COVID-19 Fake News Dataset*. Communications in Computer and Information Science ,1402 ,(India,2021) ,pp.21-29 .
- [15] SAHAN M.,SMIDL V. and MARIK R. *Active Learning for Text Classification and Fake News Detection*. International Symposium on Computer Science and Intelligent Controls (ISCSIC), (Rome,Italy,2021), pp.87-94.
- [16] MANZOOR S.,SINGLA J. and NIKITA. *Fake News Detection Using Machine Learning approaches: A systematic Review*. Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI), (Tirunelveli,India,2019), pp.230-234.
- [17] ISLAM S. ET AL.*COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis.*, The American Journal of Tropical Medicine and Hygiene, 103(4), 2020, 1621–1629. DOI: https://doi.org/10.4269/ajtmh.20-0812
- [18] VORA A. and SHEKOKAR N. *Fake News Detection Using Intelligent Techniques*. 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), (Noida, India,2021), pp.1-4.
- [19] BANIK S. *COVID Fake News Dataset*. [Dataset], (2020) DOI:10.5281/zenodo.4282522
- [20] SHAGOTO R., RAIHAN M., LABONI A., RAIHAN M.S. *Covid-19 News Dataset Both Fake and Real*. [Dataset], (2021), DOI:10.5281/zenodo.4722484
- [21] KOIRALA A. *COVID-19 Fake News Dataset*. [Dataset], (2021), Mendeley Data, V1, doi: 10.17632/zwfdmp5syg.1
- [22] BIRD S., KLEIN E., LOPER E. *Natural Language Processing with Python*. (2009). O’Reilly Media Inc. https://www.nltk.org/book/
- [23] PEDREGOSA F. ET AL. *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research (12), (2011). 2825-2830.
- [24] THE PANDAS DEVELOPMENT TEAM *pandas-dev/pandas: Pandas*, Zenodo (2020), doi: 10.5281/zenodo.3509134