Comparing Traditional Machine Learning Methods for COVID-19 Fake News

1st Sattam Almatarneh
Faculty of Information Technology
Zarqa University
Alzarqa, Jordan
0000000269704888

2nd Pablo Gamallo
CiTIUS
Universidade de Santiago de Compostela
Galiza, Spain
0000000258192469

3rd Bassam ALshargabi
Faculty of Information Technology
Middle East University
Amman, Jordan
0000000269480431

4th Yazan AL-Khassawneh
Faculty of Information Technology
Zarqa University
Alzarqa, Jordan
0000-0002-8683-6698

5th Raed Alzubi
Department of Computer Science
King Faisal University
Hofuf, Saudi Arabia
0000-0002-6651-8342

2021 22nd International Arab Conference on Information Technology (ACIT) | 978-1-6654-1995-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ACIT53391.2021.9677453

***Abstract*—This article describes some supervised classification techniques for COVID-19 fake news detection in English, where the sources of data are annotated posts from various social media platforms such as Twitter, Facebook, or Instagram. The main objective is to examine the performance of traditional machine learning techniques of COVID-19 fake news detection. In this situation, models trained with Support Vector Machine and Na¨ıve Bayes algorithms outperformed all other strategies.**

***Index Terms*—Fake News, COVID-19, Supervised Machine Learning, Natural Language Processing.**

I. Introduction

The widespread use of Social media platforms such as Facebook, Instagram, and Twitter pushed people to rely on them as the main source for information. Unfortunately, an abnormal amount of fake news, rumors, and disinformation has overflowed Social media, with the aim of drawing the attention of the Social media's users to shape their opinions and judgments [1]. Fake news and all kind of disinformation can have dramatic effects on countries, businesses, and people on various levels, whether political or economically [2].

Nowadays, the whole world is facing a critical crisis because of the COVID-19 Pandemic, and the amount of fake news about COVID-19 has swarmed Social media, which led to the spread of fear and panic among people all over the world. This only happens because people are misinformed due to fake news related to COVID-19 cures, vaccines, and the spread of infection. additionally, the World Health Organization shed light on addressing the spread of misleading information about COVID-19 cures and infections especially after the announcement of the British Government about the new mutation of COVID-19 that was discovered at the end of the year 2020 [3].

Many approaches have been proposed to identify the authenticity of published news on social media. Some of these approaches rely on the users of Social media. For example, Facebook urges their users to report suspicious news or comments [4], and even makes use of professionals to manually check the reported comments and news published on their platform. The manual fact-checking process also has been used by many other fact-checkers, journals and organizations to discover questionable news, however, this manual method is a waste of human efforts because of the huge amount of news published every second on social media [5]. Accordingly, automating the detection of fake news has caught the attention of researchers in academia and industry particularly after the incident of the American elections in 2016. Social media major players such as Facebook, Google, and Twitter employed many researchers to develop techniques that can help to automate the detection of fake news published on social media.

The use of Natural Language Processing (NLP) techniques with machine learning and deep learning methods for the detection of fake news can help to stop or at least reduce the misinformation about COVID-19 and its cures. The use of traditional machine learning techniques for the detection of fake news has received much attention recently [1], [6], [7]. The main tasks for automated fact-checking are evidence retrieval and claim verification. As the CONSTRAINT-2020 shared sub task is focused on COVID-19 Fake News Detection in English. The sources of data are annotated posts from various social media platforms such as Twitter, Facebook, or Instagram [8], the objective of this sub task is what is known as *claim verification* in fact-checking domain.

In this paper, in order to detecting COVID-19 fake news (claim verification), we compared traditional machine learning techniques, such as Support Vector Machine (SVM) and Naive Bayes (NB) classifiers. The purpose of using the aforementioned methods is to assess its effectiveness in detecting COVID-19 fake news in English. Moreover, we also tried to pinpoint the importance of feature selection methods that considerably affect the classifier's accuracy in detecting COVID-19 fake news.

The remaining of the paper is organized as follows. The

review of recent related works is presented in section II. In section III, we describe the proposed methods. Experimental results and discussion are addressed in section IV, while section V reports the conclusion of the current study as well as future work.

II. Related Work

Most strategies for automated fact checking are supervised, i.e., where the classifiers trained over a provided labeled data [9]. As supervised machine learning methods are required labeled training documents with pre-defined categories (*fake* or *real*), they are very well suited to the task of claim verification task, which is easily modeled as a binary classification task. In addition, it is important to point out that the quality of machine/deep learning approaches for fact checking relies on selecting the appropriate method to extract features from evidences previously retrieved from reliable sources, so as to detect if the checked claims are fake or not.

Traditional supervised machine learning methods, such as Na¨ıve Bayes(NB), Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), etc, mainly rely on TF-IDF and n-gram, or word embeddings for feature selection process [10]. Several classification models were built In [6] using traditional machine learning strategies and features selected from n-grams and embeddings, in order to detect tweets with hate speech. Authors In [11], compare three different traditional machine learning algorithms, namely NB, SVM and RF, to identify fake news, giving rise to similar results.

Deep learning methods also have been used to detect fake news, where the use of ensemble methods of Convolutional Neural Network (CNN) and bidirectional Long Short-Term Memory (LSTM) revealed comparable results in the detection of fake news [1], [12].

In [13], the authors compared three different deep learning techniques to discriminate fake news on the Internet and among them the one based on BERT, which achieved state-of-the-art results. Another work using BERT for claim detection is reported in [14]. A comparison of three fake new detection models for fake news of COVID-19 presented in [15]. The three models are the long short-term memory (LSTM) networks, multichannel convolutional and neural network (MC-CNN); and k-nearest neighbors (KNN). The result showed that the KNN showed a better performance compared to LSTM and MC-CNN. A new technique to detect fake COVID-19 news on Twitter is proposed in [16]. The technique relies on classifying tweets based on extracting tweets features using syntactic and semantic features of tweets alongside machine learning, and deep learning. The proposed technique was implemented in parallel with apache spark. Their results show that the random forest with SVM has an accuracy equal to 79 which is higher when compared to MLP and NB. Eight machine-learning and deep learning algorithms were exploited to detect fake news of COVID-19 in [17], the machine learning algorithms were Naive Bayesian, Adaboost, -nearest neighbors, random forest, logistic regression, decision tree, neural networks, and support vector machine and four deep learning CNN, LSTM, RNN, and GRU. The models were trained and evaluated using performance measures such as confusion matrix, classification rate, true positives rate, etc. Their results show that the Random Forest and MLP achieved a higher accuracy. Another paper evaluated ten machine learning algorithms alongside different feature extraction methods to study their effects on the performance of machine learning algorithms to detect fake Covid-19 news [18]. They also used a voting ensemble machine learning classifier. They used 5-fold cross-validation to evaluate the validity of the collected dataset and the machine learning algorithms.

III. Method

We aim to compare different classification algorithms and build the corresponding classifiers using the same training data in a supervised strategy. The characteristics of labeled posts are encoded as features in vector representation. These vectors and the corresponding labels feed the classifiers. Linguistic features are the most important and influential factor in increasing the efficiency of classifiers for any task. We included linguistic features for fake news detection. The main linguistic features we will use and analyze are n-grams. More precisely, for all classification algorithms, we deal with n-grams based on the occurrence of unigrams. Unigrams is valuable to detect specific domain-dependent expressions. We assign a weight to all terms by using Term Frequency-Inverse Document Frequency (TF-IDF) representations.

SVMs are supervised learning methods used for classification and regression, working effectively in high dimensional spaces. SVM classifiers show excellent performance on the text classification task. In our experiments, we chose Lin- ear SVC from the scikit-learn library[1]. Concerning NB, we used the system we implemented in previous work for the task of bot detection [19], which turned out to behave better than the implementation based on scikit-learn library. It is worth noting that the extraction of n-grams is performed with Linguakit [20], a multilingual toolkit with a specific tokenizer that might be the determining element in relation to the differences between the two implementations of the Bayesian classifier.

IV. Experiments

A. CONSTRAINT-2020 Dataset

The CONSTRAINT-2020 Shared Task is aimed at detecting hostile posts. It calls for participating in two subtasks, one of them being COVID19 Fake News Detection in English. It focuses on the detection of COVID19-related fake news in English. More precisely, given a post on social media, the system has to classify it into either fake or real statements. The data sources consist of several social networks posts such as Instagram, Facebook, Twitter, etc. This task is a specific case of claim verification in fact-checking. The dataset consists of three partitions: training, validation, and test [8]. Training

[1]https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html?highlight=linearsvcsklearn.svm.LinearSVC

consists of 6420 labeled posts from social media, containing 180K words. Validation and test datasets consist of 2140 posts each, the first being used for development and the second for final evaluation and official results see Table I.

Split	Real	Fake	Total
Training	3360	3060	6420
Validation	1120	1020	2140
Test	1120	1020	2140
Total	5600	5100	10700

TABLE I
DISTRIBUTION OF DATA ACROSS CLASSES AND SPLITS ON CONSTRAINT-2020 DATASET.

In our experiments, we did not use any external source of knowledge or other annotated datasets to train the systems. All models were built by using either the training data or merging both training and validation data.

B. CoVID19-FNIR dataset

This dataset is a CoVID-19-specific dataset consisting of fact-checked fake news scraped from Poynter, and true news from the verified Twitter handles of news publishers. The data samples were collected from India, The United States of America, and European regions and consisted of online posts from social media platforms between February 2020 to June 2020. The dataset went through prepossessing steps that included removing special characters, and non-vital information [21]. The dataset consists of 3795 fake news and 3793 true news tweets of trusted news sources.

C. Systems Configuration

Since we are dealing with a text classification problem, any existing supervised learning methods can be applied. We decided to utilize scikit-learn, which is an open-source machine learning library for Python programming language [22]. We chose SVM, Näıve Bayes (NB), Decision Tree (DT), k-nearest neighbors (KNN), and Random Forest (RF) as our classifiers for all experiments. Hence, in this study, we will compare, summarize, and discuss the behavior of these learning models on the datasets with the linguistic features introduced above. In order to provide a comprehensive comparison between classifiers, we adopted the default values for all classifiers on all experiments.

D. Evaluation

According to the results shown in Tables II and III of all classifiers SVM and NB clearly outperform the remaining classifiers in both datasets see fig 1 and fig2.

Figure 3 shows how SVM outperforms the other classifiers in terms of F1 values across the two datasets. The plot shows that the performance is stable with all classifiers among the two datasets.

E. Discussion

In our experiments with all datasets of the COVID19 Fake News Detection in English task, traditional machine learning techniques show excellent performance. SVM achieves the best results in the two datasets. The classifier's behavior is similar to SVM in terms of stability, but its performance tends to be lower than SVM on both datasets.

The superficial similarities detected by traditional strategies seem sufficient to obtain decent results (more than 93% accuracy and F1).

It is worth noting that the scores are very similar in the two datasets. This is not surprising because the label's distribution is the same and composed of the duplicate entries number. The two datasets are also very homogeneous in terms of subject matter and language style (social media).

Methods	precision	recall	f1
SVM	**93**	**93**	**93**
KNN	88	86	86
DT	85	85	85
RF	90	89	89
NB	91	91	91

TABLE II
RESULTS IN PRECISION, RECALL, AN F1 OF FIVE TRADITIONAL CLASSIFIERS ON CONSTRAINT-2020 DATASET.

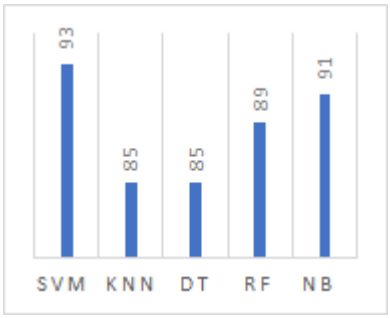

Fig. 1. Example of a figure caption.

Methods	precision	recall	f1
SVM	**90**	**90**	**90**
KNN	86	86	86
DT	81	81	81
RF	87	87	87
NB	89	89	89

TABLE III
RESULTS IN PRECISION, RECALL, AN F1 OF FIVE TRADITIONAL CLASSIFIERS ON COVID19-FNIR DATASET.

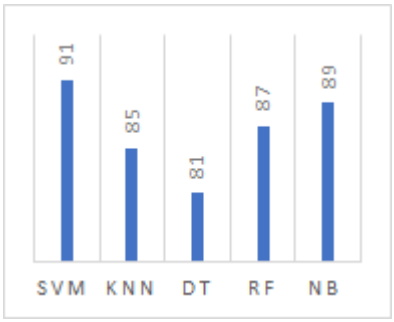

Fig. 2. Example of a figure caption.

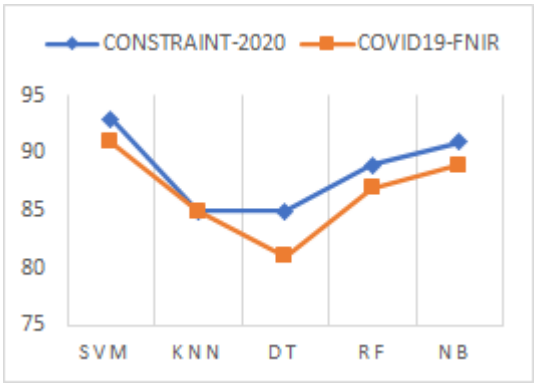

Fig. 3. Results obtained by all classifiers for all collections in terms of F1 scores.

V. Conclusions

In this paper, several classification algorithms have been compared for the fake news detection task on social media posts conveying information about COVID-19. More precisely, we compared traditional machine learning methods. Experiments were performed without considering external sources of knowledge or other annotated datasets. We observed that traditional models in machine learning performed very well in detecting fake news about COVID-19.

In future work, we will try to analyze the results obtained to establish what factors determine the significant difference between the neuronal and traditional approaches with the datasets evaluated. We will also look for other sources of information providing reliable COVID-19 claims to improve the results.

References

[1] Kumar, Sachin, et al. "Fake news detection using deep learning models: A novel approach." Transactions on Emerging Telecommunications Technologies 31.2 (2020): e3767.

[2] Apuke, Oberiri Destiny, and Bahiyah Omar. "Fake news and COVID-19: modelling the predictors of fake news sharing among social media users." Telematics and Informatics 56 (2021): 101475.

[3] Nguyen, Van-Hoang, et al. "Fang: Leveraging social context for fake news detection using graph representation." Proceedings of the 29th ACM International Conference on Information Knowledge Management. 2020.

[4] Umer, Muhammad, et al. "Fake news stance detection using deep learning architecture (CNN-LSTM)." IEEE Access 8 (2020): 156695-156706.

[5] Giachanou, Anastasia, and Paolo Rosso. "The battle against online harmful information: The cases of fake news and hate speech." Proceedings of the 29th ACM International Conference on Information Knowledge Management. 2020.

[6] Almatarneh, Sattam, et al. "Supervised classifiers to identify hate speech on English and Spanish tweets." International Conference on Asian Digital Libraries. Springer, Cham, 2019.

[7] Gilda, Shlok. "Notice of violation of IEEE publication principles: Evaluating machine learning algorithms for fake news detection." 2017 IEEE 15th student conference on research and development (SCOReD). IEEE, 2017.

[8] Patwa, Parth, et al. "Fighting an infodemic: Covid-19 fake news dataset." International Workshop on Combating Online Ho stile Posts in Regional Languages during Emergency Situation. Springer, Cham, 2021.

[9] Thorne, James, and Andreas Vlachos. "Automated fact checking: Task formulations, methods and future directions." arXiv preprint arXiv:1806.07687 (2018).

[10] Gravanis, Georgios, et al. "Behind the cues: A benchmarking study for fake news detection." Expert Systems with Applications 128 (2019): 201-213.

[11] Cus¸maliuc, CIPRIAN-GABRIEL, LUCIA-GEORGIANA Coca, and A. D. R. I. A. N. Iftene. "Identifying Fake News on Twitter using Naive Bayes, SVM and Random Forest Distributed Algorithms." Proceedings of The 13th Edition of the International Conference on Linguistic Resources and Tools for Processing Romanian Language (ConsILR-2018) pp. 2018.

[12] Ko, Hoon, et al. "Human-machine interaction: A case study on fake news detection using a backtracking based on a cognitive system." Cognitive Systems Research 55 (2019): 77-81.

[13] Rodr´ıguez, A´ lvaro Ibrain, and Lara Lloret Iglesias. "Fake news detection using Deep Learning." arXiv preprint arXiv:1910.03496 (2019).

[14] Soleimani, Amir, Christof Monz, and Marcel Worring. "Bert for evidence retrieval and claim verification." Advances in Information Retrieval 12036 (2020): 359.

[15] Alenezi, Mohammed N., and Zainab M. Alqenaei. "Machine Learning in Detecting COVID-19 Misinformation on Twitter." Future Internet 13.10 (2021): 244.

[16] Madani, Youness, Mohammed Erritali, and Belaid Bouikhalene. "Using artificial intelligence techniques for detecting Covid-19 epidemic fake news in Moroccan tweets." Results in Physics 25 (2021): 104266.

[17] Bangyal, Waqas Haider, et al. "Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches." Computational and Mathematical Methods in Medicine 2021 (2021).

[18] Elhadad, Mohamed K., Kin Fun Li, and Fayez Gebali. "Detecting misleading information on covid-19." Ieee Access 8 (2020): 165201-165215.

[19] Gamallo, Pablo, and Sattam Almatarneh. "Naive-Bayesian Classification for Bot Detection in Twitter." CLEF (Working Notes). 2019.

[20] Gamallo, Pablo, et al. "LinguaKit: a Big Data-based multilingual tool for linguistic analysis and information extraction." 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS). IEEE, 2018.

[21] Julio A. Saenz, Sindhu Reddy Kalathur Gopal, Diksha Shukla. (2021). "Covid-19 Fake News Infodemic Research Dataset (CoVID19-FNIR Dataset)." Web.

[22] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." the Journal of machine Learning research 12 (2011): 2825-2830.