Benha Journal of Applied Sciences (BJAS) **print: ISSN 2356–9751**
Vol. (7) Issue (5) (2022), (265-268) **online: ISSN 2356–976x**
http://bjas.journals.ekb.eg

Misinformation Detection in Arabic Tweets: A Case Study about COVID-19 Vaccination

Nsrin Ashraf, Hamada Nayel and Mohamed Taha
Department of Computer Science, Faculty of Computers and artificial Intelligence, Benha University, Benha, Egypt
E-mail: {nisrien.ashraf19, hamada.ali, mohamed.taha}@fci.bu.edu.eg

Abstract

Misinformation about COVID-19 overwhelmed our lives due to the tremendous usage of social media, especially Twitter. Spreading misinformation caused fear and panic among people affecting the national economic security of many countries. Vaccination is the crucial key to limiting the pandemic spread of COVID-19. Therefore, researchers start to detect and fight against the spread of misinformation taking it as a new challenge. This paper illustrates a model for misinformation detection in Arabic tweets using Natural Language Processing (NLP) techniques. A machine learning-based system has been developed regarding COVID-19 vaccination tweets. Term Frequency-Inverse Document Frequency (TF-IDF) has been used as vector space model for feature extraction. Support Vector Machines classification algorithm has been used for implementation the proposed system. Evaluation of the system, using different metrics, has been implemented on Arcov-19Vac, a dataset of Arabic tweets related to COVID-19 vaccination. The results reported by the illustrated model show that the performance of our model is promising.

Keywords:
COVID-19 Misinformation detection,Machine Learning, Social media analysis

1. Introduction

The coronavirus disease (COVID-19) has lately impacted our lives, across the world there was a large number of infection cases and deaths every day. With a few resources and knowledge about how to deal with the symptoms because they differ from one person to another. People though they have similar symptoms they start to share their quarantine behaviors and how to deal with the virus without making sure if this data is true or not. Invention of COVID-19 vaccines has been a significant improvement in the battle against the virus. In December 2020, many pharmaceutical companies proposed vaccines that have been found to be safe and effective in generating an immune reaction [1].

Sixty-five percent of approximately 3.8 billion internet users are currently informed about top news stories from social media platforms with the spread of misleading information's such as Twitter, Facebook, and Reddit rather than traditional news outlets [2]. The spread of misinformation causes fear and panic between people affecting the national economic security of many countries and lead to refuse to get vaccinated or suggest vaccination to others of their social circle. A few months researchers start to detect and fight against the spread of this information taking it as a new challenge. Natural Language Processing (NLP) aims at analyzing text from more than one direction. NLP can help in studying the effect of people's thoughts and the spread of COVID-19 related information on countries which leads to making the consequences of wrong decisions, through developing different models to easily detect rumors [3].

Different NLP tasks such as detecting rumors have been formulated as a text classification problem and various approaches have been applied to automatically detect rumors and fake news [7,8]. Rumors regarding COIVID-19 has been widely spread and research work to detect such rumors have been explored for English [9]. For Arabic language, ArCov-19 dataset investigated the impact of fake news for COVID-19 [4]. This dataset comprises of 2.7M tweets that covering the period of January 2020 till January 2021. Authors also provided the propagation networks of a the most-retweeted and -liked tweets. Ensemble based approach has been used to develop a model to classify COVIC-19 related rumors [6]. A similarity measure has been used to track the source of rumors. The proposed approach flags the tweet with the highest similarity as the source of rumor to purpose a new combine method genetic algorithm based SVM model with TF-IDF tokenizer and different word embedding techniques.

ArCovVac is the first Arabic tweets dataset constituting about COVID-19 vaccine campaign covered in many Arabic countries [5]. It comprises different levels of annotation, involving informativeness, fine-grained contents and stance towards vaccination. In addition, advanced analysis has been conducted including the popularity of different vaccines, trending hashtags and topics.

In this paper we aimed to show the advancements made in the field of fake news detection. This work aims at automatically detecting the fake tweets related to COVID-19 vaccine. A range of *unigrams and bigrams* Bag of Words (BoW) model have been used and TF-IDF vectorizer to for tweet representation. The proposed system has been developed using machine learning-based model and different classification algorithms namely, Support Vector Machines (SVMs), Multinomial Naïve Bayes (MNB), Stochastic Gradient Descent (SGD), Multi-Layer Perceptron (MLP), Random Forest (RF), voting classifier, K-Nearest Neighbor (KNN) and AdaBoost classifiers. All of these algorithms have been evaluated using ArCovVac dataset.

The rest of the paper is organized as follows; section 2 introduces material and methods have been used in this study. Results and discussion are given in section 3. Section 4 concludes the work have been done and suggests the future work.

2. Material and Methods

2.1. Data

ArCovidVac is the first Arabic twitter dataset designed to gather tweets concerning people reviews after vaccine constituting about 10k tweet [5]. They used a set of keywords (vaccine, vaccination, تطعيم and لقاح) to collect tweets and manually annotated COVID-19 vaccine infodemic from Jan 5th to Feb 3rd, 2021 using twarc search API. The collected tweets covering different Arabic countries and labeled with one of the predefined set of the following classes: 'Info-news', 'Celebrity', 'Plan', 'Requests', 'Rumors', 'Advice', 'Restrictions', 'Personal', 'Advice', 'Restrictions', 'Personal', 'Unrelated' and 'others'. Tweets are classified in two stances positive (pr-vaccination), and negative (against vaccination), we focused on the main task of our research which is misinformation classification tasks including fake tweets from non-informative ones, and real tweets from Info-news, trusted organizations [5].

2.2. Data Pre-processing

The first stage of the proposed system is preprocessing, which comprises a set of tasks such as:

- Hashtags removal, where the hashtag itself will be removed while the word after each hashtag was kept
- Removal of URLs, mentions, and whitespaces. These kind of noise data is uninformative
- Repeated characters were removed
- Normalize Arabic words such as ("[إأآا]" to "ا", "گ" to "ك", etc.)
- Non-Arabic characters were removed

2.3. Feature Extraction

After preprocessing and before feeding the data into machine learning classifier, the TF-IDF has been used as a vector space model for feature extraction. This phase aimed at converting all tweets into a vector of real numbers, this process is called vectorization. To explore the efficiency of this phase, we used BoW model with ranges of *unigrams* and *bigrams* . The higher dimension of *n*-gram may increase the feature dimension without adding discriminatory information. Feature extraction is a crucial phase, where the performance of algorithm mainly depends on how the quality features are. To reduce the dimensionality of features, we applied aforementioned preprocessing tasks. These tasks eliminate the unwanted and duplicated information.

2.4. Training the Model

After vectorization, we applied different machine learning classification algorithms to explore the efficiency of each. These algorithms are SVMs, MNB, SGD, MLP, RF, voting classifier, KNN and AdaBoost classifiers as shown in table 1.

2.5. Experimental Setup

We used SkLearn package (https://scikit-learn.org/stable/) to implement the different algorithms. Each algorithm has various parameters, we set most of these parameters as the default values defined in the implementation. The values of parameters that have been used while implementation are given in Table 2. For all classifiers we set the random_state parameter at 42, to achieve the same results during development phase.

Table (1) Arcov-19Vac Dataset statistics.

Class	Train	Dev	Test
Rumors	79	15	24
Others	6921	985	1976

Table (2) Parameters settings.

Classifier	Parameter	Value
SVM	Kernel	linear
	Gamma	2
	C	1
SGD	Loss	hinge
	Penalty	l2
	Alpha	1e-3
	Max_iter	5000
MLP	Hidden Layer	20
	Activation	Logistic
RF	n_estimators	300
PA	Max_iter	110
	loss	hinge
	C	1

2.6. Evaluation metrics

Precision, recall, accuracy and F1-score have been used to evaluate the performance of our model. F1-score is often used for evaluating information retrieval systems such as search engines by combining the model's precision and recall through the following formula:

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

3. Results and Discussion

The results of our method for different algorithms for unigram and bigram models are given in Table 3 and Table 4, respectively. In terms of F1-score, unigram model outperforms bigram model. In terms of accuracy, SVM outperforms all other classifiers in both unigram and bigram models. That may be clear because of the nature of SVM, where SVM is a binary classifier. The imbalance of the dataset reported a gap between accuracy and F1-score in all classifiers for both unigram and bigram models. In terms of F1-score, KNN outperforms all other classifiers for both unigram and bigram models. KNN reported the highest recall which is an important factor in imbalanced data. All other classifiers reported a moderated result. Both MNB and SGD classifiers reported the least performance measure in both unigram and bigram models.

4. Conclusion and Future Work

We provided a machine learning-based framework for misinformation detection regarding COVID-19 vaccine. Different classification algorithms have been implemented to explore their performance. Due to the imbalanced dataset that has been used there is a gap between different evaluation metrics. In future work, we can implement other framework such as deep learning framework. In other side, vectorization can be performed using other techniques such as word embedding. These improvements may reduce the gap between various metrics.

Table (3) Results of unigram model for all algorithms.

Classifier	Evaluation metric			
	Precision	**Recall**	**F1-score**	**Accuracy**
SVM	**0.695**	0.475	0.522	**82.90%**
MNB	0.454	0.344	0.321	80.50%
SGD	0.84	0.35	0.33	80.80%
MLP	0.628	0.528	0.56	81.20%
RF	0.692	0.424	0.455	82.20%
Voting Classifier	0.687	0.414	0.441	81.90%
KNN	0.626	**0.554**	**0.575**	82.40%
AdaBoost	0.583	0.408	0.418	80.70%

Table (4) Evaluation metrics of bigram model for all algorithms.

Classifier	Evaluation Metric			
	Precision	**Recall**	**F1-score**	**Accuracy**
SVM	**0.681**	0.425	0.457	**82%**
MNB	0.464	0.349	0.331	80.60%
SGD	0.547	0.346	0.322	80.80%
MLP	0.634	0.474	0.513	**82%**
RF	0.669	0.415	0.442	81.70%
Voting Classifier	0.672	0.403	0.425	81.50%
KNN	0.583	**0.504**	**0.53**	80.80%
AdaBoost	0.478	0.393	0.394	81.20%

Reference

[1] **Bose P., Roy S. and Ghosh P.** "A Comparative NLP-based study on the current trends and future directions in COVID-19 Research", IEEE Access, 9, 78341–78355. https://doi.org/10.1109/access.2021.3082108, **2021.**

[2] **Elhadad M. K., Li, K. F. and Gebali F**. "Covid-19-fakes: A Twitter (Arabic/English) dataset for detecting misleading information on COVID-19", Advances in Intelligent Networking and Collaborative Systems, 256–268. https://doi.org/10.1007/978-3-030-57796-4_25, **2020**.

[3] **Yan C., Law M., Nguyen S., Cheung J. and Kong, J.** "Comparing public sentiment toward covid-19 vaccines across Canadian cities: Analysis of comments on Reddit", Journal of Medical Internet Research, 23(9). https://doi.org/10.2196/32685, **2021**.

[4] **Fatima Haouari, Maram Hasanain, Reem Suwaileh, and Tamer Elsayed**. "ArCOV-19: The First Arabic COVID-19 Twitter Dataset with Propagation Networks", In Proceedings of the Sixth Arabic Natural Language Processing Workshop, pages 82–91, Kyiv, Ukraine (Virtual). Association for Computational Linguistics, **2021**.

[5] **Mubarak H., Hassan S., Chowdhury S. A. and Alam F.** "ArCovidVac: Analyzing Arabic Tweets About COVID-19 Vaccination". arXiv preprint arXiv:2201.06496, **2022**.

[6] **Noman Qasem S., Al-Sarem M. and Saeed F.** "An ensemble learning based approach for detecting and tracking Covid19 rumors", Computers, Materials & Continua, 70(1), 1721–1747, https://doi.org/10.32604/cmc.2022.018972, **2022**.

[7] Zhou X. and Zafarani R. "A survey of fake news: Fundamental theories, detection methods, and opportunities", ACM Computing Surveys (CSUR), 53(5), 1-40, https://doi.org/10.1145/3395046, **2020**.

[8] **Hamada A. Nayel** "NAYEL@APDA: Machine Learning Approach for Author Profiling and Deception Detection in Arabic Texts", In Working Notes of FIRE 2019 - Forum for Information Retrieval Evaluation, Kolkata, India, December 12-15, 2019, volume 2517 of CEUR Workshop Proceedings, pages 92–99, **2019**.

[9] **A. Shi, Z. Qu, Q. Jia and C. Lyu**, "Rumor Detection of COVID-19 Pandemic on Online Social Networks" IEEE/ACM Symposium on Edge Computing (SEC), 2020, pp. 376-381, https://doi:10.1109/SEC50012.2020.00055, **2020**.