

Research Article

Ensemble Classifiers for Arabic Sentiment Analysis of Social Network (Twitter Data) towards COVID-19-Related Conspiracy Theories

Abdullah Al-Hashedi, Belal Al-Fuhaidi , Abdulqader M. Mohsen , Yousef Ali, Hasan Ali Gamal Al-Kaf, Wedad Al-Sorori , and Naseebah Maqtary

Faculty of Computing and IT, University of Science and Technology Sana'a, Sana'a, Yemen

Correspondence should be addressed to Belal Al-Fuhaidi; belalarh@gmail.com

Received 19 December 2020; Revised 5 June 2021; Accepted 20 December 2021; Published 13 January 2022

Academic Editor: Ridha Ejbali

Copyright © 2022 Abdullah Al-Hashedi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sentiment analysis has recently become increasingly important with a massive increase in online content. It is associated with the analysis of textual data generated by social media that can be easily accessed, obtained, and analyzed. With the emergence of COVID-19, most published studies related to COVID-19's conspiracy theories were surveys on the people's sentiments and opinions and studied the impact of the pandemic on their lives. Just a few studies utilized sentiment analysis of social media using a machine learning approach. These studies focused more on sentiment analysis of Twitter tweets in the English language and did not pay more attention to other languages such as Arabic. This study proposes a machine learning model to analyze the Arabic tweets from Twitter. In this model, we apply Word2Vec for word embedding which formed the main source of features. Two pretrained continuous bag-of-words (CBOW) models are investigated, and Naïve Bayes was used as a baseline classifier. Several single-based and ensemble-based machine learning classifiers have been used with and without SMOTE (synthetic minority oversampling technique). The experimental results show that applying word embedding with an ensemble and SMOTE achieved good improvement on average of *F1* score compared to the baseline classifier and other classifiers (single-based and ensemble-based) without SMOTE.

1. Introduction

With the emergence of the coronavirus (COVID-19) disease in December 2019 in the Chinese city of Wuhan and the speed of its outbreak in most countries over the world, much talk has begun on social media about the causes of the emergence of the virus, its symptoms, ways to prevent it, and the efforts made by the developed countries and global research centers to find out the drug to recover from this virus.

The talk about a conspiracy theory about the presence of COVID-19 and its quickly spread also forms a significant portion of the daily content of the news and social media. Social media content forms a public opinion and trends that may not often depend on the facts but rather on anecdotal

and speculation, wishful thinking, or conspiratorial fantasy [1].

Since there is no scientific evidence from governmental sources or the World Health Organization to show the true reasons for the emergence of the COVID-19 and its outbreak, people turn to other sources such as social media to obtain information, exchange ideas and experiences, and discuss various issues related to this disease.

The media also deals extensively with this disease and reports on the actions taken by individuals in responding to the advice and instructions issued by official authorities related to committing to the ban, staying at home, social distancing, and a healthy diet.

Thus, news and social media have formed sources for obtaining information about COVID-19, whether it is based

on the facts and evidence or that misleading information is based on people's opinions. What made it more difficult for fact-checking organizations is the speed and tremendous amount of information that enters the information ecosystem so that most reputable sites and platforms feed massive amounts of information quickly and profusely which contain inaccurate information.

In light of the coronavirus crisis, there are many psychological and economic factors that affect the ability and willingness of the people to comply with the procedures for confronting the COVID-19. One of the most potent factors that may influence individuals' decision to commit to preventive measures is the belief in a conspiracy theory about COVID-19. In the past few months, there have been many claims that support the conspiracy theory of COVID-19; for example, the electromagnetic waves transmitted through the 5G phone masts caused COVID-19, and the government intentionally induces panic among people by inflating the disease in news media in order to impose a constraint on the people to control the population [2, 3]. Another example of conspiracy theory is that many people believed that COVID-19 had developed in the lab and that COVID-19 was Bill Gates' attempt to take over the medical industry [4]. A national survey in the United Kingdom found that 50% of the population supported the conspiracy theory to some extent [5].

Because of people's belief in conspiracy theory, people may resort to countermeasures to formal government measures such as rejection of social distancing and engaging in risky behavior that spreads the virus [6]. For example, some people set fire to the 5G phone masts in the United Kingdom [7].

In the last few years, sentiment analysis used to get critical information about individuals, groups, and communities' perceptions. This new computational approach is used to analyze textual data in order to classify and interpret public opinions towards any given issue [8].

Sentiment analysis has recently become closely associated with the analysis of textual data generated by social media that can be easily accessed, obtained, and analyzed [9]. This encouraged analysts to analyze the public's different opinions on social, health, political, and other issues. In the health sector, health institutions have begun to pay close attention to using the sentiments analysis approaches to measure people's opinions on important health issues. With the emergence of COVID-19, analysts began analyzing textual data in social media such as Twitter and Facebook to study the trends and opinions of the public on various topics, such as people's opinions about the conspiracy theory surrounding the emergence and spread of this disease. Unfortunately, most of the scientific work conducted on this topic dealt with the sentiment analysis of Twitter tweets in the English language. To the best of our knowledge, there are only a few researches conducted using sentiment analysis on tweets posted in the Arabic language.

The primary objective of this study is to analyze Arab people's attitudes towards COVID-19 conspiracy theories through sentiment analysis of Twitter data. This study was carried out on Twitter tweets using sentiment classification

with different machine learning algorithms for the analysis and visualization of the opinions regarding the COVID-19-related conspiracy theories such as COVID-19 made in a Chinese lab, COVID-19 related to 5G network, and Bill Gates supported the development of a global surveillance system by spreading vaccination widely using the COVID-19 pandemic. Most well-known classifiers are used in this study, including Support Vector Machine (SVM), MaxEntropy, Decisional Tree Classifier, Bagging, and Random Forest. In addition, an ensemble classifier was developed by combining several classifiers into a decision-making system based on an ensemble. The developed ensemble classifier synthesizes a more accurate Arabic sentiment classification procedure; it investigates the strengths and weaknesses of word embedding regarding COVID-19-related conspiracy theories with two pretrained models on our collected imbalanced dataset. Moreover, we apply SMOTE (Synthetic Minority Oversampling Technique) technique to the training data for comparison purposes and performance enhancement. The Naïve Bayes algorithm was used as a baseline. Finally, we explore the performance of combining classifiers using ensemble learning algorithms to enhance the accuracy of the model.

The rest of this paper is organized as follows. Section 2 introduces the related works. The proposed Arabic Sentiment Analyzer (ASA) framework is introduced in Section 3. The results and performance evaluation are presented in Section 4. Finally, conclusions and future works are depicted in Section 5.

2. Related Work

In the context of the COVID-19 pandemic, conspiracy theories may reinforce distrust of health authorities and their recommendations, which could seriously hamper efforts to end the epidemic. In the short term, respecting the guidelines for containment behavior (for example, hand washing and social distractions) is critical to reducing the spread of the epidemic, because developing treatment (including vaccine) may take months (WHO, 2020) [10]. However, in the long run, the development and distribution of a vaccine against COVID-19 may be a necessary step to end the pandemic. Bavel et al. [4] carried out a survey that discusses evidence from a selection of research topics relevant to COVID-19 pandemics, including conspiracy theories and other related topics. They provide some insights from the past century of work on related issues in the social and behavioral sciences. In [11], the review highlighted some of these challenges and suggests general measures to avoid information overload and disease in the connected world of the 21st century. The article in [12] attempted to understand the reasons for the widespread interest that COVID-19 has received. It briefly displayed what is known so far about SARS CoV2. Next, it explores whether the media has played a role in the widespread interest and possibly exaggerated targeting of COVID-19.

Several studies addressed the opinions of the people regarding COVID-19 conspiracy theories using surveys. In [5], the study carried out a nonprobability online survey with

2501 adults in England to estimate the prevalence of conspiracy thinking about COVID-19 pandemics and test associations with minimizing compliance to government guidelines. Results of this study showed that nearly 50% of this population showed little evidence of conspiracy thinking, 25% showed a degree of support, 15% showed a consistent pattern of support, and 10% had very high levels of support. Therefore, there is concrete support for conspiracy beliefs about coronaviruses in England. Conspiracy beliefs are associated with other forms of distrust and are associated with less compliance with government guidelines and an increased unwillingness to take tests and future treatment. In the same context, in another study [13], the authors proposed and tested a mediation model in which there is a rejection of COVID-19 conspiracy theories. The model mediates the relationship between analytic thinking and compliance with mandated social-distancing measures. The participants of this study consisted of an online, nationally representative sample of adults from the United Kingdom ($N = 520$, age $M = 45.85$ years). Exploratory factor analysis indicated that both of the new procedures were mono with sufficient internal consistency. The reciprocal relationships between the scores in all three measures were significant and positive. Mediation analysis indicated that analytic thinking and the rejection of conspiracy theories COVID-19, respectively, were significantly and directly related to compliance and that the median association was also important.

Bertin et al. in [14] conducted two cross-sectional studies to examine the relationships between COVID-19 conspiracy beliefs, vaccine attitudes, and the intention to vaccinate against COVID-19 when the vaccine becomes available. In an exploratory study 1 ($N = 409$), the two subdimensions of conspiracy theories COVID-19 were strong predictors of a negative attitude toward vaccine science. The results were repeated and expanded in a prerecorded study 2 ($N = 396$). Moreover, they found that the conspiracy beliefs of COVID-19 (among them the conspiracy beliefs about chloroquine), as well as the conspiracy mindset (i.e., willingness to believe in conspiracy theories), negatively predicted the intentions of those involved in vaccination against COVID-19 in the future. The study noticed an incredibly high level of COVID-19 vaccination frequency. In addition, pandemic effects and possible responses were discussed on the basis of the existing literature.

The study in [15] examined the cultural and psychological factors associated with intentions to reduce the prevalence of COVID-19. Seven hundred and twenty-four participants were recruited for this relational study. Participants completed ($n = 704$) individual group metrics, belief in conspiracy theories about COVID-19, feelings of impotence, and intentions to engage in behaviors that reduce the prevalence of COVID-19. The results showed that vertical individuality negatively predicted the intentions of engaging in social divergence, directly and indirectly through belief in conspiracy theories COVID-19. The vertical community positively predicted social intentions directly. The horizontal group positively predicted social intentions indirectly through a sense of helplessness.

Moreover, the horizontal group positively predicted hygiene intentions, either directly or indirectly through a reduced sense of powerlessness. These results suggest that collective enhancement may be a way to increase participation in efforts to reduce the prevalence of COVID-19.

The study in [16] aimed to ascertain the prevalence of conspiracy beliefs circulating in Cyprus and Greece via social media. The questionnaire was designed based on assessing their psychological pressures related to COVID-19, evaluating their confidence in science to solve COVID-19 problems, and assessing their willingness to adhere to measures regarding social exclusion and quarantine. The results estimated from one thousand and one ($N = 1001$) individuals adult showed that conspiracy theories are widely believed even among highly educated individuals. Chen et al. [17] provided the first evidence of conspiracy theory belief that COVID-19 was developed intentionally in a lab. This study conducted a survey of 252 healthcare staff in Ecuador for many factors including anxiety, distress, job satisfaction, and life satisfaction of healthcare staff. The study's result showed that 32.54% of health care staff suffered from distress disorder, and 28.17% had an anxiety disorder. Compared to other health care staff who were not sure about the source of the virus, those who believed the virus had been intentionally developed in a laboratory reported higher levels of anxiety and lower levels of job satisfaction and life satisfaction. Older healthcare workers and those who exercise are more comfortable with the job.

In the literature, many researchers analyze sentiment on social media, which helps user behaviors and situations in relation to a specific topic or condition around the world. These studies are based on sentiment analysis for posts on social media platforms. The authors in [18] have used different machine learning approaches to test carefully tweets extracted from Twitter regarding the COVID-19 outbreak. The various machine learning classifiers, such as Naïve Bayes, support vector machine (SVM), decision tree, MaxEntroPy, LogitBoost, and random forests, classified tweets as positive, negative, and neutral. The LogitBoost ensemble classifier is the most remarkable accuracy compared to other classifiers.

Shahsavari et al. [1] have estimated narrative framework discovery that described the underlying structure of COVID-19-related conspiracy theories in the social media corpus (4Chan, Reddit) and news reports corpus (GDELT). This framework analyzed the interplay between the two corpora using automated machine learning methods. The results of the social media analysis show that the conversations concentrate on the four COVID-19-related conspiracy theories: COVID-19 was made in a Chinese lab, COVID-19 was related to 5G network, COVID-19 is not more dangerous than mild flu, and Bill Gates supported the development of a global surveillance system by spreading vaccination widely using the COVID-19 pandemic. The Naïve Bayes machine learning model was used by Alhajji et al. [19] to perform Twitter postanalyses of Arabic feeling. It collected and analyzed the tweets containing Hashtags for seven government-imposed public health measures. In the same context, authors in [20] carried out sentiment analysis

using the Naïve Bayes algorithm through Twitter data. They measured customer opinions and perceptions using this model.

Ahmed et al. [21] analyzed the content of Twitter data and web resources in the United Kingdom to about the motives for the 5G COVID-19 conspiracy theory and strategies to deal with misinformation. This study analyzed users by social network graph clusters. In this cluster, the degree of centralization points represents the size of the nodes, and the graph headers are grouped by cluster using the Clauset-Newman-Moore algorithm. The analysis results show that, from 233 tweets, 34.8% contained opinions that COVID-19 and 5G were related, 32.2% disapproved conspiracy theory, and 33.0% did not express any personal opinions. Pulido Rodríguez et al. [22] compared Twitter's tweets and Weibo posts regarding coronavirus containing either real scientific information or misinformation. The messages have been carried out through Python programming language from all social media classified and compared. The results of this study show that false news is published on Twitter more than on Sina Weibo, and at the same time, real scientific information is published on Twitter more than Weibo but less than misinformation.

The authors in [23] designed a dashboard based on Twitter's news to track misinformation. The dashboard scans the social media discussions about the COVID-19 pandemic frequently and the shared information updated over time. In addition, it provided an analysis of public sentiments about intervention policies using lexical sentiment extraction, and it tracked topics, ratings, and feelings emerging across countries. The dashboard maintains a sophisticated list of misleading information chains, feelings, and trends emerging over time; it classified tweets as positive, negative, and neutral based on a large German data set.

Boberg et al. [24] analyzed the actual basics of such fears in alternative news media's on Facebook during the early Corona pandemic. This study used an unsupervised learning algorithm such as latent Dirichlet allocation (LDA) that can discover latent topics inductively based on patterns of terms occurring together. The access, interactions, actors, and topics were examined messages, as well as the use of misinformation and conspiracy theories. The analysis revealed that alternative news media remain valid for the types of

messages. Alam et al. [25] designed and released a dataset for the research community through social media to analyze the misinformation about the COVID-19 pandemic. The collected dataset covers English and Arabic languages. This study provided a detailed analysis of the annotations, reporting the label distribution for different questions, as well as the correlation between different questions, along with other statistics. To explore Arab views against the closure, Mohsen et al. [26] proposed an ML model in which they investigated various classifiers, data balancing, and feature extraction techniques. In addition, Al-Sorori et al. [27] applied a machine learning (ML) model to analyze and categorize the tweets related to fear and anxiety regarding the COVID-19 outbreak.

Looking closer to what has been so far presented in the previous studies, we find that most studies were published since COVID-19-related conspiracy theories to understand the public sentiments and opinions and study the impact of the pandemic on their lives, thus enabling decision-makers to adopt the appropriate strategies and preventing measures to respond to the COVID-19 pandemic in case of conspiracy theories. Most studies use the Twitter and Facebook platforms to analyze posts written in the English language. With regard to Arabic language, there is a lack of studies that addressed this issue and this study attempts to fill this gap. Moreover, in case of COVID-19-related conspiracy theories, most studies did not include the sentiment analysis model based on classifier ensemble.

A description of the proposed model that is considered in this study is presented in the next section.

3. Proposed Arabic Sentiment Analyzer (ASA) Model

The proposed model has several phases, including data collection, preprocessing techniques, feature extraction, and model generation, as shown in Figure 1.

3.1. Data Collection. The tweets are generated with the Tweepy API using tracking-related keywords as shown in the following.

حرب بيولوجي كوفيد, حرب بيولوجي كورونا, فيروس مصنع, حرب بيولوجي كورونا,
حرب بيولوجي سارس, حرب بيولوجي جائحة, مؤامرة جائحة, مؤامرة كوفيد, حرب
بيولوجي مؤامرة, فيروس بيولوجي, فيروس بيولوجي مؤامرة, مؤامرة كورونا,
فيروس صيني, مؤامرة سارس

We selected the most effective and perfect keywords connected to our targets (label) which are as displayed above. The collection process was inspired by the pre-screening process used in the systematic literature review such as [28]. It is a collective process. First, there are 1824 tweets used to build the corpus1 as shown in Figure 2. The

tweets were collected using streaming Tweepy API and a list of Hashtags related to conspiracy theories. Then, we remove the duplicated tweets, retweets, and filter the non-Arabic tweets to build the filtered dataset with a size of 953 tweets. Moreover, we add other new tweets with a size of 1829 to build corpus2. Then, we remove the duplicated tweets and

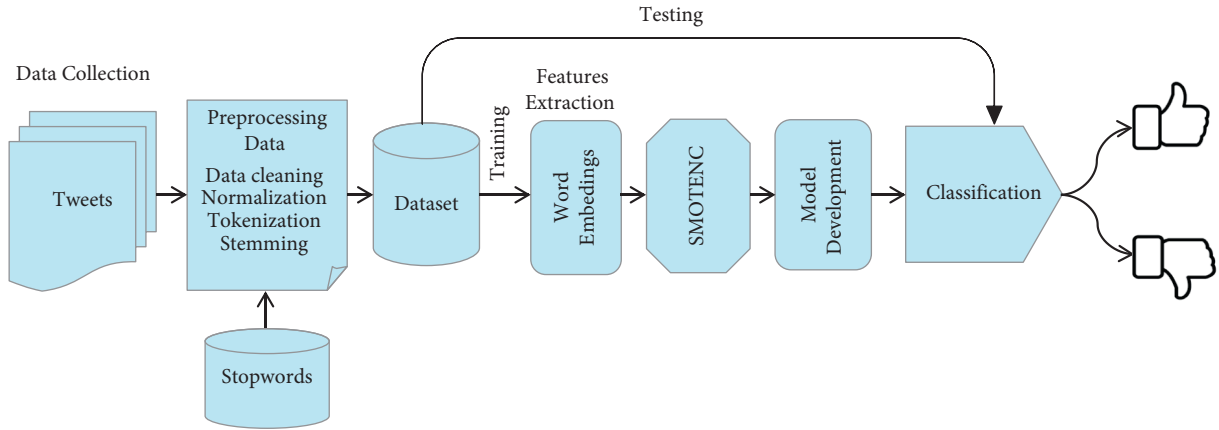


FIGURE 1: The proposed ASA model.

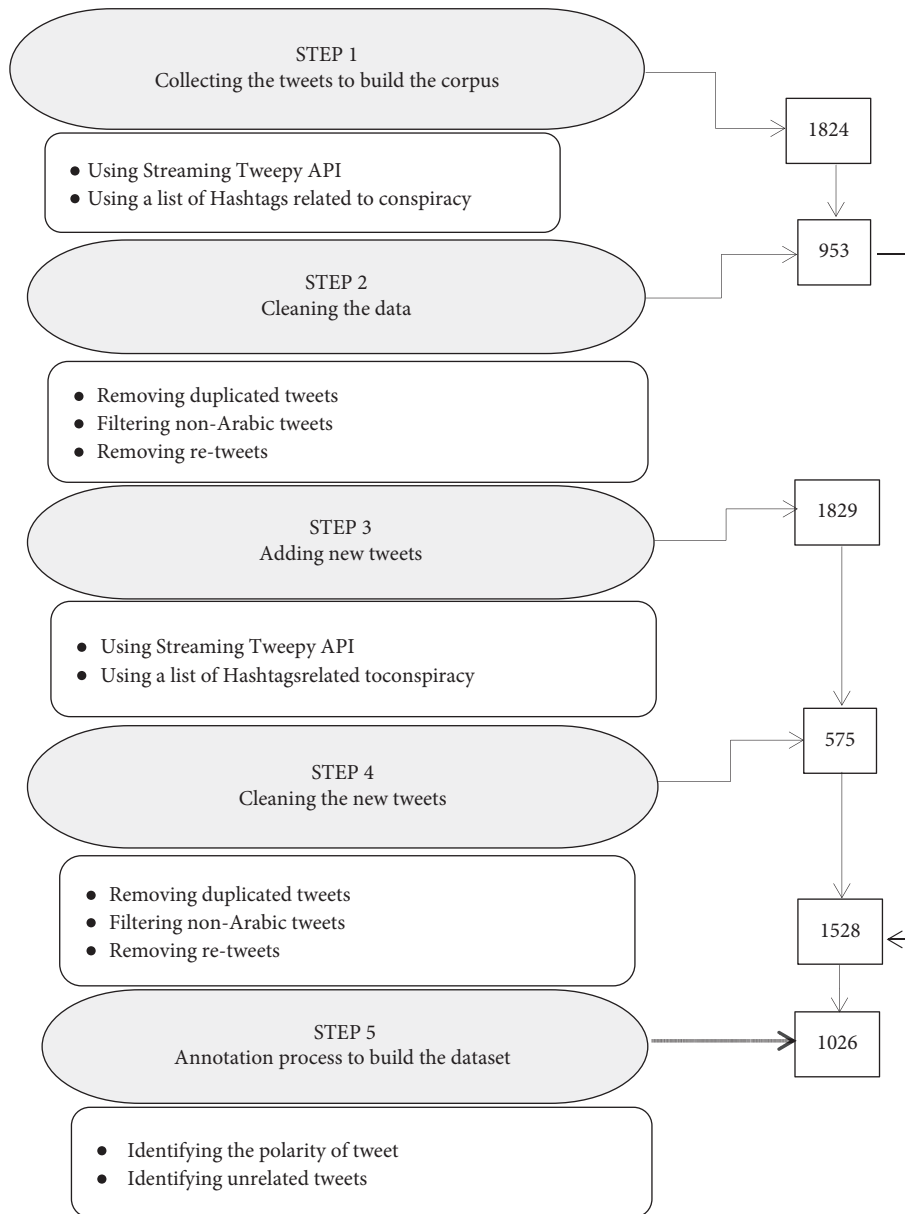


FIGURE 2: The conspiracy dataset derived from the corpus.

retweets and filter the non-Arabic tweets to build the filtered dataset with a size of 575 tweets. After that, we combine the filtered dataset of two corpora with a size of 1528 tweets. Finally, the data is classified into binary classes, including positive for tweets supporting a conspiracy and negative for those not supporting a conspiracy.

3.2. Preprocessing Step. The preprocessing is a required step in order to remove the unwanted special and Latin characters. Different techniques were conducted for further analysis as described as follows.

- (1) Removing noise from Arabic data: several steps have been conducted in order to remove the noisy complexity of collected data, including removing the punctuations, Arabic diacritics, Numbers, emotions, and no letters from the word.
- (2) Normalization: there were several rules applied to normalize the text. For example, each Alif with different forms such as (آ, ا, إ) is replaced with a bare Alif character “ا.” Besides, the characters repeated with more than two frequencies and Tatweel “_” were removed.
- (3) Tokenization: the tokenization procedure tokened the cleaned text into words in order to filter out unnecessary tokens. The tokenization step is vital before transforming to vectors that are used as input for classifiers.
- (4) Streaming: an Arabic light streamer used to remove prefixes and suffixes resulting in reducing the dimensionality of the text data.
- (5) Remove stop words: stop words are a set of commonly used words in a natural language occurring frequently and carry less meaning. The natural languages toolkit (NLTK) library has an extensive list of Arabic stop words. The Arabic stop words are integrated and stored in a file. These stop words would take up space in our dataset and valuable processing time; therefore, they were removed.

3.3. Word Embedding. The function of word embedding is to map closely the words that have a similar meaning or common contexts in the space using word vectors. It efficiently computes word vector representation in the high dimensionality of vector space. The most two common methods used for producing word embedding in NLP are Word2Vec and Global Vectors. The Word2Vec uses a combination of two neural network architectures, including continuous bag-of-words (CBOW) and skip-gram.

Word2Vec effectively computes representations of the word vector in high-dimensional vector space. Word vectors are located in the vector space where terms that have similar semantics and share common contexts are represented in space close to each other. Besides syntactic information, the similarity of word representations extracts semantic features [29].

In the Global Vectors for Word Representation or GloVe, the calculation is an augmentation to the Word2Vec strategy for effectively learning word vectors, created by Pennington et al. [29] at Stanford. Classical vector space model portrayals of words were produced utilizing matrix factorization strategies, for example, Latent Semantic Analysis (LSA), which complete a great job of utilizing global text statistics, yet they are not in the same class as the educated techniques like Word2Vec at catching importance and exhibiting it on undertakings like figuring analogies.

Practically, both methods are used extensively in the literature, one of which works better on some data sets than the other and vice versa. They both do very well at capturing the semantics of analogy, and therefore, we selected Word2Vec in this study for word embedding.

3.4. SMOTENC. We integrated SMOTENC (synthetic minority oversampling technique for nominal and continuous) on the training data to balance an imbalanced dataset. SMOTENC method uses SMOTE technique by synthesizing new minority samples. The categories of newly generated samples are determined by selecting the most frequent category of the nearest neighbors.

3.5. Model Generation. In this section, a brief description of the suitable classification model for enhancing the process of building a model is presented.

3.5.1. Naïve Bayes (NB). Naïve Bayes (NB) classifier is a probabilistic classifier that uses the properties of Bayes theorem assuming the strong independence between the features. One of the advantages of this classifier is that it requires a small amount of training data to calculate the parameters for prediction. Instead of calculating the complete covariance matrix, only the variance of the feature is computed due to feature independence. This classifier is widely used as a baseline classifier [30]. In this paper, we selected Naïve Bayes as baseline (BNB).

3.5.2. Support Vector Machine (SVM). SVM is a machine learning classification technique that uses a function called kernel to map a space of data points in which the data is not linearly separable onto a new space, with allowances for erroneous classification. It has been successfully utilized on Arabic sentiments analysis [31, 32].

3.5.3. Logistic Regression (LR). The logistic regression classifier (LR) calculates the conditional probability distribution of a class relied on the dataset. It supposes no former knowledge. The ratio of the training dataset puts a constraint on the conditional distribution, thus forcing the classifier to find the ME distribution that is consistent with the constraint [33]. Since the classification is binary and our dataset is small in size, LRCV is utilized with a solver (liblinear) to discover the weights of the parameter to minimize a cost

function. In this paper, the experiments are conducted using 10-fold cross-validation.

3.5.4. Stochastic Gradient Descent (SGD). Stochastic gradient descent (SGD) is an optimization technique for converging on a problem solution by choosing an arbitrary solution. It measures the goodness of fit under a loss function and iteratively takes steps in the direction that minimizes loss. According to [28], this classifier has some advantages such as efficiency and ease of implementation. Thus, it is applied in this paper.

3.5.5. Random Forest (RF). Random forest is a type of supervised machine learning algorithm based on ensemble learning. It combines multiple decision trees and then produces a forest of trees, hence the name “random forest.” The random forest algorithm can be used for both regression and classification tasks.

3.5.6. Voting. A voting classifier is an ensemble classification method that has the advantage of combining the predictions by majority voting from multiple machine learning algorithms. It exploits the features of each algorithm [34]. Several combinations of single classification methods are used with the majority vote in this paper. The voting classifier combined RF, SGD, SVM, BNB, and LR.

4. Experiments and Results

Once the data were labeled manually, the data showed highly imbalance where the clear majority class is positive, while the minority of class is negative. The related tweets (Twitter posts) were crawled with the Tweepy API. Simultaneous streams were collected to build the dataset. The data were annotated on one level, which is the document level in binary classes, including positive and negative. Two specialists in the Arabic language have annotated the data. For preprocessing, we removed all URLs, emails, and newlines, as they are not informative for language or textual analysis. After labeling the generated tweets, the label values (i.e., positive and negative) were not balanced after data collection. To solve this problem, we integrated SMOTENC (synthetic minority oversampling technique for nominal and continuous) on the training data to balance an imbalanced dataset. The positive class and negative class are 636 and 363 tweets, respectively, as shown in Table 1. Then, the dataset was divided into two datasets: one, which forms 90% used as a training set, while the other, which forms 10% used as a test set. In addition, the type of cross-validation used is 10-fold cross-validation.

We used different evaluation metrics including accuracy, precision, recall, and F1 score. The metrics are defined as follows.

TABLE 1: The conspiracy dataset distribution.

Dataset	Testing	Training	Total
Positive	66	597	663
Negative	36	327	363
Total	102	924	1026

$$\begin{aligned}
 \text{Accuracy (Acc)} &= \frac{TP + TN}{TP + TN + FP + FN}, \\
 \text{precision (Pr.)} &= \frac{TP}{TP + FP}, \\
 \text{recall (Re.)} &= \frac{TP}{TP + FN}, \\
 \text{F1 - score} &= \frac{2 * \text{Re.} * \text{Pr.}}{\text{Re.} + \text{Pr.}},
 \end{aligned} \tag{1}$$

where TP, FP, TN, and FN denote true positive, false positive, true negative, and false negative, respectively. SMOTENC was employed on the imbalanced training set to evaluate the computational models utilizing synthesized samples. It is implemented in Python by an imbalanced-learn toolbox.

Two types of classification methods were investigated and evaluated. The first type is a single-based classifier, which includes SVM, LSVM, LR, SGD, and BNB. The second type is an ensemble-based classifier [35], which includes several combinations of classifiers with hard voting which considers the majority vote. We investigated several resampling techniques in our proposed model. According to the results of applying each technique, we selected the best one to be used in our model. We used RF and voting. The ensemble combination of a single classifier is shown in Table 2.

In our study, two pretrained CBOW models were investigated. The first model is cc.ar.300 which was built by [36]. This model is character n-grams of length 5 and dimension of 300 with a learned embedding size of 2,000,000 vocabularies. The second model is Arabic.news with a learned embedding size of around 159,178 vocabularies.

There are several models of sentiment analysis based on genetic algorithms and deep learning techniques such as [9, 37], but not toward COVID-19-related conspiracy theories and not conducted on sentiment analysis on the Arabic language. Most studies use the Twitter and Facebook platforms to analyze posts written in the English language. With regard to the Arabic language, there is a lack of studies that addressed this issue, and this study attempts to fill this gap. Moreover, in the case of COVID-19-related conspiracy theories, most studies do not include the sentiment analysis model based on classifier ensemble. Furthermore, our model was built on ensemble machine learning methods, not on deep learning. Besides, to the best of our knowledge, there are no previous studies conducted in sentiment analysis on the Arabic language.

TABLE 2: An ensemble combination of single classifier.

Combination	Representative
LR, RF, BNB soft	Voting 1
LR, RF, BNB hard	Voting 2
LR, RF, SGD soft	Voting 3
LR, RF, SGD hard	Voting 4
RF, SVM, SGD hard	Voting 5
CV, SGD, SVM hard	Voting 6

TABLE 3: The results of the ensemble-based and single-based classifiers using two pretrained models with and without SMOTENC.

	Model CC.AR.300								Model Arabic.news							
	Without SMOTENC				With SMOTENC				Without SMOTENC				With SMOTENC			
	Acc	F1	Pr.	Re.	Acc	F1	Pr.	Re.	Acc	F1	Pr.	Re.	Acc	F1	Pr.	Re.
RF	78.07	84.48	79.73	90.19	94.54	89.63	98.14	83.16	79.44	84.79	80.07	87.92	94.82	88.77	97.42	84.39
SGDC	71.26	79.26	76.28	79.68	88.84	77.20	92.65	66.56	71.07	80.57	75.35	86.57	88.52	76.73	90.30	68.40
Linear SVC	72.81	79.56	76.49	82.06	92.27	84.61	96.16	78.02	72.13	78.97	76.68	81.15	92.18	83.46	91.68	79.22
Nu SVC	75.04	79.59	80.97	77.22	96.58	93.48	97.81	90.26	69.76	75.42	77.54	72.69	96.34	92.40	96.35	89.63
LR	71.83	77.79	76.37	77.68	94.68	89.59	96.86	86.76	72.42	78.48	75.12	78.88	94.32	88.27	92.37	86.34
Bernoulli NB	64.73	70.87	76.84	66.36	80.87	65.19	70.04	62.63	63.75	69.35	76.76	63.79	80.57	66.74	63.64	71.65
Voting 1	71.45	77.63	79.17	76.77	86.07	72.53	84.29	65.73	68.62	74.58	78.44	72.08	87.76	75.48	83.70	70.69
Voting 2	74.28	82.20	73.72	91.55	86.56	69.73	95.60	56.80	76.81	83.75	76.10	92.30	90.78	78.68	94.17	67.38
Voting 3	74.75	80.47	77.35	81.45	94.67	89.18	98.09	84.98	74.37	80.09	76.33	81.30	94.32	87.83	93.15	83.48
Voting 4	74.96	82.10	76.41	84.32	94.42	89.26	99.09	84.57	76.13	81.77	75.66	85.68	94.19	87.48	93.86	83.96
Voting 5	78.17	83.40	81.80	86.73	95.56	91.22	99.09	86.74	77.19	83.43	81.15	84.31	95.20	89.24	96.20	84.42
Voting 6	73.10	79.13	77.76	80.40	95.18	90.94	97.33	87.61	73.88	79.76	76.73	83.20	94.83	88.95	93.64	86.82

The classification experiments were conducted using several methods which form either single-based or ensemble-based classifiers. We used SVM, LSVM, SGD, LRCV, and BNB for single-based classifiers while for ensemble-based classifiers, we used RF and voting. With regard to voting-based ensemble classifiers, several combinations of single-based classifiers were evaluated. The first combination included LRCV, RF, and BNB with soft and hard voting. The second combination included three single-based classifiers, namely, LRCV, RF, and SGD with hard and soft voting (voting3 and voting4). The third combination included RF, SVM, and SGD with hard voting (voting5). The last combination included LRCV, SGD, and SVM with hard voting too (voting6).

The results obtained using single-based and ensemble-based classifiers and two pretrained models with and without applying SMOTENC are shown in Table 3. Four evaluated measures were used: accuracy (Acc), precision (Pr.), recall (Re.), and F1 score (F1). For the purpose of performance comparison between the different classifiers, the F1 score was considered for two reasons: first, the dataset is imbalanced. Second, F1 is a more informative score because it considers both precision and recall measures.

The baseline considered in this study is the performance of BNB without applying SMOTENC. To evaluate our model, two pretrained models (i.e., ArabNews model and CC.AR.300 model) with and without SMOTENC are applied. The highest F1 score obtained for baseline using the CC.AR.300 model without SMOTENC is 70.87%.

Most single-based classifiers were affected positively in terms of F1 score after applying SMOTENC with both

pretrained models, except SGDC and BNB. The highest results are that obtained with the CC.AR.300 model for SVM (93.48%), RF (89.63%), LR (89.59%), and LSVM (84.61%). For the ArabNews model, the highest results in terms of F1 score are those obtained for SVM (92.40%).

For ensemble-based classifiers, it can be generally noticed that the performance of classifiers with the CC.Ar.300 model is better than that achieved by classifiers with the ArabNews model. For ensemble-based classifiers with cc.AR.300 model, it can be noticed that most of the combinations have improved significantly after applying SMOTENC in terms of F1 score, except the ensemble-based classifiers with voting1 (LR, RF, BNB soft) and voting2 (LR, RF, BNB hard). The best result (91.94%) in terms of F1 score is that obtained for voting5 (RF, SVM, SGD hard) followed by (90.94%) obtained for voting6 (CV, SGD, SVM hard). On the other hand, it can be noticed also that most combinations of ensemble-based classifiers with the ArabNews model after applying SMOTENC have experienced significant improvement in terms of F1 score, except voting2 (LR, RF, BNB hard). The highest results were also those obtained for voting5 (89.24%) and voting6 (88.95%). For all sets of combinations with both models cc.AR.300 and ArabNews, it is clear that voting2 which includes LR, RF, and BNB did not achieve any significant improvement and this is maybe attributed to BNB which has the worst performance.

Figure 3 shows the average F1 score for single-based and ensemble-based classifiers with and without SMOTENC using the two pretrained models.

From Figure 3, it can clearly be noticed that the performance of ensemble-based classifiers with imbalanced datasets (i.e., without SMOTENC) is more robust than

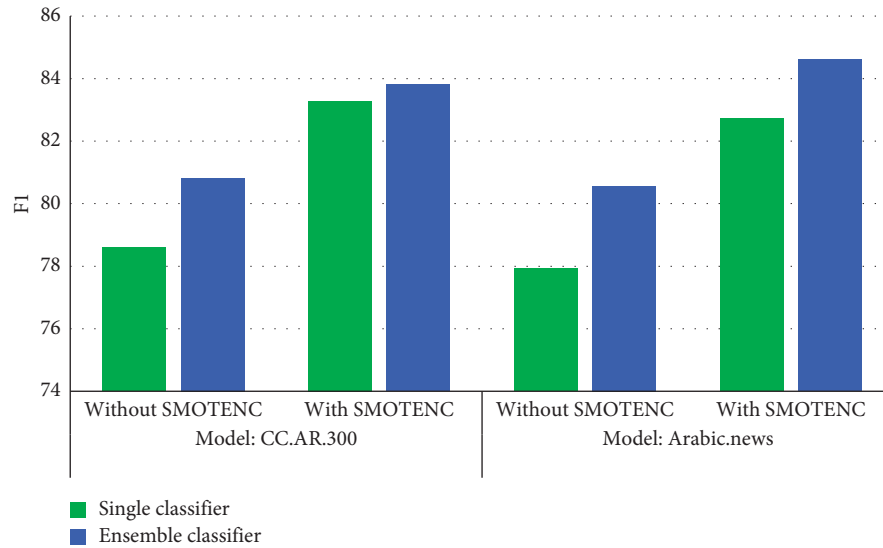


FIGURE 3: Average of $F1$ score for single-based and ensemble-based classifiers with and without SMOTENC.

single-based classifiers with both models. The average $F1$ score is 80.82% for ensemble-based classifiers and 80.56% for single-based classifiers with cc.ar.300, while it is 78.59% and 77.93% for single-based classifiers with both models. On the other hand, the overall average of $F1$ score of ensemble-based classifiers is also more robust than single-based classifiers with balanced datasets (i.e., with SMOTENC) with both models cc.ar.300 (83.81%) and Arabic.news (84.61%).

5. Conclusion

Most studies use the Twitter and Facebook platforms to analyze posts written in the English language. With regard to the Arabic language, there is a lack of studies that addressed this issue and this study attempted to fill this gap. In the case of COVID-19-related conspiracy theories, the proposed sentiment analysis model was built on ensemble machine learning methods. This study presents a model for Arabic sentiment analysis toward the COVID-19 conspiracy. The study relied on Arabic reviews collected from Twitter's tweets.

A popular and widely used learning approach Word2Vec is applied for word embedding which formed the main source of features. Two pretrained CBOW models are investigated. Moreover, the SMOTENC technique is utilized to address the problem of the imbalanced dataset by oversampling the minority class by adding synthetic samples. Several machine learning methods (single-based and ensemble-based) are used to evaluate the effectiveness of applying SMOTENC to word embedding features. Ensemble learning method, voting, is applied to increase the performance of single-based classifiers.

The experimental results show that applying SMOTENC significantly improves the performance of both types of classifiers (single and ensemble) with both pretrained models. As future work, further experiments can be conducted using other ensemble learning methods such as bagging, boosting, and stacking to evaluate the performance on other datasets and types of features.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] S. Shahsavari, P. Holur, T. R. Tangherlini, and V. Roychowdhury, "Conspiracy in the time of corona: automatic detection of covid-19 conspiracy theories in social media and the news," 2020, <https://arxiv.org/abs/2004.13783>.
- [2] A. Depoux, S. Martin, E. Karafillakis, R. Preet, A. Wilder-Smith, and H. Larson, "The pandemic of social media panic travels faster than the COVID-19 outbreak," *Journal of Travel Medicine*, vol. 27, 2020.
- [3] R. Kouzy, J. Abi Jaoude, A. Kraitem et al., "Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on twitter," *Cureus*, vol. 12, no. 3, Article ID e7255, 2020.
- [4] J. J. V. Bavel, K. Baicker, P. S. Boggio et al., "Using social and behavioural science to support COVID-19 pandemic response," *Nature Human Behaviour*, vol. 4, no. 5, pp. 460–471, 2020.
- [5] D. Freeman, F. Waite, L. Rosebrock et al., "Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England," *Psychological Medicine*, Cambridge University Press, Cambridge, UK, 2020.
- [6] G. Pennycook and R. M. Ross, "Commentary: cognitive reflection vs. calculation in decision making," *Frontiers in Psychology*, vol. 7, p. 9, 2016.
- [7] J. Waterson and A. Hern, "How false claims about 5G health risks spread into the mainstream," *The Guardian*, 2020, <https://www.theguardian.com/technology/2020/apr/07/how-false-claims-about-5g-health-risks-spread-into-the-mainstream>.
- [8] P. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1-2, 2008.

- [9] M. E. Basiri, M. Abdar, M. A. Cifci, S. Nemati, and U. R. Acharya, "A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques," *Knowledge-Based Systems*, vol. 198, Article ID 105949, 2020.
- [10] "World Health Organization," 2020, <https://www.who.int>.
- [11] F. Rathore and F. Farooq, "Information overload and infodemic in the COVID-19 pandemic," *Journal of the Pakistan Medical Association*, vol. 70, 2020.
- [12] G. M. Abaido and A. A. Takshe, "COVID-19: virus or viral conspiracy theories?" *American Journal of Biomedical Science & Research*, vol. 8, no. 2, pp. 122–124, 2020.
- [13] V. Swami and D. Barron, "Rational thinking style, rejection of coronavirus (COVID-19) conspiracy theories/theorists, and compliance with mandated requirements: direct and indirect relationships in a nationally representative sample of adults from the United Kingdom," *Journal of Pacific Rim Psychology*, vol. 15, 2020.
- [14] P. Bertin, K. Nera, and S. Delouvée, "Conspiracy beliefs, rejection of vaccination, and support for hydroxychloroquine: a conceptual replication-extension in the COVID-19 pandemic context," 2020.
- [15] M. Biddlestone, R. Green, and K. M. Douglas, "Cultural orientation, powerlessness, belief in conspiracy theories, and intentions to reduce the spread of COVID-19," *British Journal of Social Psychology*, vol. 59, no. 3, pp. 663–673, 2020.
- [16] M. Constantinou, A. Kagialis, and M. Karekla, "COVID-19 scientific facts vs. conspiracy theories: 0-1: science fails to convince even highly educated individuals," *Science Fails to Convince Even Highly Educated Individuals*, 2020.
- [17] X. Chen, S. X. Zhang, A. A. Jahanshahi et al., "Belief in conspiracy theory about COVID-19 predicts mental health and well-being: a study of healthcare staff in Ecuador," 2020, <https://www.medrxiv.org/content/10.1101/2020.05.26.20113258v2>.
- [18] R. Muthusami, A. Bharathi, and K. Saritha, "COVID-19 outbreak: tweet based analysis and visualization towards the influence of coronavirus in the world," 2020, https://www.researchgate.net/publication/340791904_COVID-19_Outbreak_Tweet_based_Analysis_and_Visualization_towards_the_influence_of_Coronavirus_in_the_World.
- [19] M. Alhajji, A. Al Khalifah, M. Aljubran, and M. Alkhalifah, "Sentiment analysis of tweets in Saudi Arabia regarding governmental preventive measures to contain COVID-19," 2020, <https://www.preprints.org/manuscript/202004.0031/v1>.
- [20] K. Suppala and N. Rao, "Sentiment analysis using naive Bayes classifier," *International Journal of nInnovative Technology and Exploring Engineering*, vol. 8, no. 8, pp. 265–269, 2019.
- [21] W. Ahmed, J. Vidal-Alaball, D. Joseph, and F. L. Segui, "COVID-19 and the 5G conspiracy theory: social network analysis of twitter data," *Journal of Medical Internet Research*, vol. 22, 2020.
- [22] C. Pulido Rodríguez, B. Villarejo Carballido, G. Redondo-Sama, M. Guo, M. Ramis, and R. Flecha, "False news around COVID-19 circulated less on Sina Weibo than on twitter. How to overcome false information?" *International and Multidisciplinary Journal of Social Sciences*, vol. 9, 2020.
- [23] K. Sharma, S. Seo, C. Meng, S. Rambhatla, and Y. Liu, "COVID-19 on social media: analyzing misinformation in twitter conversations," 2020, <https://arxiv.org/abs/2003.12309>.
- [24] S. Boberg, T. Quandt, T. Schatto-Eckrodt, and L. Frischlich, "Pandemic populism: facebook pages of alternative news media and the corona crisis—a computational content analysis," 2020, <https://arxiv.org/abs/2004.02566>.
- [25] F. Alam, S. Shaar, F. Dalvi et al., "Fighting the COVID-19 infodemic: modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society," 2020, <https://arxiv.org/abs/2005.00033>.
- [26] A. Mohsen, Y. Ali, W. Al-Sorori, N. A. Maqtary, B. Al-Fuhaidi, and A. M. Altabeeb, "A performance comparison of machine learning classifiers for COVID-19 Arabic quarantine tweets sentiment analysis," in *Proceedings of the 2021 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, Sana'a, Yemen, 2021.
- [27] W. Al-Sorori, A. M. Mohsen, Y. Ali et al., "Arabic sentiment analysis towards feelings among COVID-19 outbreak using single and ensemble classifiers," in *Proceedings of the 2021 International Conference on Intelligent Technology, System and Service for Internet of Everything (ITSS-IOE)*, Sana'a, Yemen, 2021.
- [28] A. Ghallab, A. Mohsen, and Y. Ali, "Arabic sentiment analysis: a systematic literature review," *Applied Computational Intelligence and Soft Computing*, vol. 2020, Article ID 7403128, 21 pages, 2020.
- [29] J. Pennington, R. Socher, and C. D. Manning, "Glove: global vectors for word representation," in *Proceedings of the 2014 Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014.
- [30] G. Vinodhini and R. Chandrasekaran, "Sentiment analysis and opinion mining: a survey," *International Journal*, vol. 2, no. 6, pp. 282–292, 2012.
- [31] N. A. Abdulla, M. Al-Ayyoub, and M. N. Al-Kabi, "An extended analytical study of Arabic sentiments," *International Journal of Big Data Intelligence*, vol. 1, no. 1-2, pp. 103–113, 2014.
- [32] R. M. Duwairi, R. Marji, N. Sha'ban, and S. Rushaidat, "Sentiment analysis in Arabic tweets," in *Proceedings of the 2014 5th International Conference on Information and Communication Systems (ICICS)*, pp. 1–6, Irbid, Jordan, 2014.
- [33] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of Biomedical Informatics*, vol. 35, no. 5-6, pp. 352–359, 2002.
- [34] M. Swamynathan, *Mastering Machine Learning with Python in Six Steps: A Practical Implementation Guide to Predictive Data Analytics Using Python*, Apress, New York, NY, USA, 2019.
- [35] A. Onan, S. Korukoğlu, and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification," *Expert Systems with Applications*, vol. 57, pp. 232–247, 2016.
- [36] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, and T. Mikolov, "Learning word vectors for 157 languages," in *Proceedings of the 2018 International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, 2018.
- [37] A. Onan and S. Korukoğlu, "A feature selection model based on genetic rank aggregation for text sentiment classification," *Journal of Information Science*, vol. 43, no. 1, pp. 25–38, 2017.