# Exploring COVID-19 Discourse: Analyzing Sentiments for Fake News Detection in Twitter Topics

Carmela Comito

National Research Council of Italy (CNR)

ICAR Institute

Via Pietro Bucci, 8-9 C, 87036 Rende (CS), Italy

Email:carmela.comito@icar.cnr.it

*Abstract*—The COVID-19 pandemic generated extensive and far-reaching discussions on social media platforms, effectively becoming a primary source for people to access and disseminate information regarding the outbreak. These social media conversations possess the potential to shape public opinions, but they also carry the risk of spreading panic and misinformation during crises like the COVID-19 pandemic.

Given this context, it becomes imperative to identify the prevailing topics under discussion on social media and gain insights into people's perceptions, opinions, and emotions by conducting sentiment analysis on user interactions concerning these topics. As a response to this challenge, this paper introduces a novel approach for conducting sentiment analysis that is sensitive to the topics under discussion with the aim of analysing online conversations on Twitter and identifying false information associated with COVID-19. To tackle this challenge, we have devised a semi-supervised methodology that combines false information detection with sentiment-aware topic modeling algorithm. Linguistic and sentiment features are initially extracted for subsequent analysis. Following this, a set of cutting-edge machine learning algorithms are employed to classify the COVID-19-related dataset. These algorithms are subsequently assessed using a range of metrics. The findings indicate that the incorporation of our model's fake news feature extraction resulted in a notable accuracy improvement, increasing it from 59.20% to 88.12%. and that the ensemble-based approaches excels, surpassing the other classifiers with an accuracy rate of 88.50% for XGBoost and 84% for AdaBoost.

*Index Terms*—COVID-19, Social Media Data, Topic Modeling, Sentiment Analysis, Fake News Detection

## I. INTRODUCTION

The rapid global spread of the COVID-19 pandemic has led to the extensive dissemination of pandemic-related information. Online social media platforms, such as Twitter and Facebook, have played a significant role in reporting and sharing news, events, updates, and public sentiment regarding COVID-19. Consequently, a vast amount of real-time data about the pandemic is circulating across social networks.

This abundance of data can be harnessed by medical and governmental institutions to comprehend the dynamics of the population and implement preventative and corrective measures to mitigate the effects of COVID-19. Consequently, there is a pressing need for efficient analytical methods and tools to make sense of this data.

As social media's influence continues to grow in our lives, the proliferation of fake news on these platforms has become a significant concern. Fake news refers to deliberate information hoaxes designed to mislead readers for financial or political gain. The impact of such news can be as detrimental to mental well-being as it is to physical health. Therefore, it is crucial to manage and prevent the dissemination of fake news.

Of particular importance is the identification and prevention of the spread of fake news related to COVID-19, an urgent global issue that has emerged over the past two years with the pandemic and is spreading rapidly. The COVID-19 pandemic has brought forth an urgent demand for tools to combat the dissemination of misinformation. Given that the pandemic impacts the global community, there is a vast and diverse audience seeking information on the subject. Unfortunately, this audience's safety is at risk due to the actions of adversarial entities with vested interests in spreading misinformation, often for political and economic motives. Furthermore, the intricacies of medical and public health matters make it challenging to maintain complete accuracy and factualness. This complexity leads to disputes that are exacerbated by the circulation of misinformation. Moreover, the rapid evolution of our understanding of the disease adds another layer of complexity. As researchers gain more insights into the virus, previously accepted statements may prove to be incorrect, and vice versa. Detecting and curbing the dissemination of pandemic-related misinformation has, therefore, emerged as a critical issue, drawing substantial attention from government and public health organizations.

This article primarily focuses on the analysis of conversations on social media, particularly Twitter, concerning COVID-

19, with the aim of identifying the main topics of discussion, the real sentiment and tone and exposing false information pertaining to COVID-19 with reference to specific topic of discussion. To tackle this challenge, we have employed a semi-supervised approach that combines the detection of false information with a topic detection algorithm guided by sentiment features. Traditional sentiment analysis techniques primarily focus on classifying the overall sentiment of a text without considering other essential attributes, such as the specific topics or aspects to which the sentiment pertains. This work aims to identify the most relevant topics that can aid in understanding the COVID-19 pandemic situation and subsequently assess users' opinions and sentiments towards those topics.

Specifically, this paper proposes a topic-level sentiment analysis model in which topics are extracted from tweets, and sentiment analysis is performed in reference to these extracted topics. This approach combines methodologies from two research areas, namely, the detection of topics in social media streaming data and sentiment analysis.

To achieve this, we first extract and analyze all topics related to COVID-19 by clustering bursts of textual features found in tweets. Once a topic is detected, it is subjected to sentiment analysis, which classifies its polarity as positive, negative, or neutral. In this manner, topic-based sentiment analysis aims to discern sentiments towards specific topics by examining the tweets within those topics and a set of key textual features (e.g., hashtags) within a given time frame. After the topics have been identified, the fake news detection step is invoked. Specifically, for each tweet a supervised classifier determines its veracity by exploiting a small set o labelled data and the extracted features, both semantic and sentiment features, of the topic to which the tweet belongs.. This procedure yields comprehensive insights into the character and extent of false information within the analyzed dataset. It enables us to quantitatively assess the presence of false information within the primary topics under discussion by users.

A real-world dataset of COVID-19 tweets has been used to evaluate the proposed approach and a dataset of label COVID-19 tweets misinformation benchmark. Results shown that the method is effective in detecting topics of discussions related to COVID-19 and to compute the sentiment of the topics, outperforming traditional approaches. A set of cutting-edge machine learning algorithms are employed to classify the veracity of COVID-19-related data within the topics of discussion identified. The findings indicate that the incorporation of our model's fake news feature extraction resulted in a notable accuracy improvement, increasing it from 59.20% to 88.12%. and that the ensemble-based approaches excels, surpassing the other classifiers with an accuracy rate of 88.50% for XGBoost and 84% for AdaBoost.

The rest of the paper is organized as follows. Section II overviews related work. Section III formulates the problem, introducing the key aspects of the approach, while Section IV presents the topic sentiment detection algorithm and the fake news detection approach. The results of the evaluation performed over the real-world dataset of tweets are shown in Section **??**. Section VIII concludes the paper.

## II. RELATED WORK

Starting from the first days of COVID-19 outbreaks several research activities exploiting social media data to address COVID-19 issues have been launched. Some of such activities focused on the study of human behavior and reactions to COVID-19 [1], [9], [12], [13]. Conspiracy detection is another topic widely explored [6], [14] as well as misinformation propagation about COVID-19 [16], [11].

COVID-19 topic detection is another well studied research line. By the best of our knowledge the majority of the proposals in literature aiming at detecting COVID-19 topics on social networks exploits the *Latent Dirichlet Allocation (LDA)* model[2], like the following references [10] [1] [16] [15] [7]. $LDA$ is a topic model where words and documents are linked by means of latent topics. The model computes for each textual content a probability distribution over topics, which are distributions over words. The main relevant drawback of LDA is that the number of topics should be fixed in advanced and the data structure and size has also to be fixed in advanced. To overcome the limitations of current state-of-the-art, in [3] has been proposed an approach, presented in [4], for COVID-19 topic dection in Twitter that combines peak detection and clustering techniques. Space-time features are extracted from the tweets and modeled as time series. After that, peaks are detected from the time series, and peaks of textual features are clustered based on the co-occurrence in the tweets. Each cluster obtained is then associated to a topic.

## III. PROBLEM FORMULATION AND DATA MODEL

In this paper, we investigate how discovering the topics discussed in a set of tweets can be used to improve the sentiment classification of Twitter users and determine the veracity of the information reported in the tweet. Accordingly, the study focuses on analyzing user-generated content on Twitter to detect and explore false information related to COVID-19. To achieve this, we have developed a semi-supervised approach that combines topic modeling. sentiment analysis and fake news detection to create a topic-centric representation of false information.

Specifically, we used unlabeled data to identify the primary topics of discussion related to COVID-19 within social media conversations. For this step, we used a topic detection algorithm proposed in .... after that we applied a supervised classifier for sentiment analysis to exploit within a supervised classifier for determine whether the content of a given post is real or false,also computing a veracity score for each topic. This process enables us to quantitatively assess the impact of false information on Twitter discussions about COVID-19 from a topic-specific perspective. Consequently, by adopting this approach, we can pinpoint the key discussion topics that are most susceptible to false information, quantitatively evaluate their impact, and provide concrete examples of user-generated content that is misinformed on these topics.

A textual content of a tweet usually refers to one or more topics. Therefore, grouping the tweets on the basis of the topic of discussion expresses better the content of the discussions taking place on social media and enhances the sentiment extraction process. Motivated by these challenges, this paper proposes an online learning model for sentiment analysis of topics. Specifically, the aim is to detect the topics in an online manner to deal with streaming data and then perform topic-level sentiment analysis.

The proposed solution is designed as a novel hybrid sentiment classification model combining sentence and topic-based approaches. The main rationale behind this choice is that predicting an overall score for a tweet is not suitable since a tweet can mention different topics. Therefore, it is more effective to perform sentence-level sentiment of each tweet and multi-topic sentiment classification, predicting different ratings for each topic discussed in the tweet rather than an overall rating.

In the following are reported some examples to make clearer many of the above concepts.

The sentiment of words used in a tweet are often dependent on the topic/s of that tweet. For example let's consider the tweet "*Happy!Remdisivir is working symptoms disappeared fortunately as I dont trust in vaccines*". The tweet belongs to two topics, the "*Medicine*" topic and the "*Vaccine*" topic. The tweet expresses a positive sentiment towards the "Medicine" topic while it expresses a negative sentiment toward the "Vaccine" topic. If we would consider the general sentiment of the tweet would be wrongly fully positive.

Another example of tweet with controversial sentiment is the following: "*Stay safe from Coronavirus. There is currently no antiviral treatment or vaccine to prevent COVID-19*". In fact, the bigrams "*stay safe*" could determine a positive sentiment whereas the overall sentiment of the tweet is determined by the absence of a treatment or vaccine to prevent COVID-19.

To address the above challenges and issues, the paper introduces the *Topic-specific Sentiment Classifier* (TSC) algorithm, which is only trained on tweets of the same topic and is formulated as follows:

*Definition 1:* (Topic Sentiment classification ) Given a stream of tweets and a set of textual features associated to them, the aim is to determine the topics discussed and the sentiment of each topic.

In the following are introduced the concepts used to model the topic-specific sentiment classifier approach, mainly the *tweet* and the *topic*.

A tweet $tw$ is defined as a data structure storing relevant features of the textual content of the posts:

*Definition 2:* (Tweet) It is defined as a tuple $tw = (id, fv, sfv, s)$ where $id$ is the tweet identifier, $fv$ is a vector of textual features extracted from the tweet, mainly words and hashtags, $sfv$ is a vector of textual features expressing sentiments and $s$ is the overall sentiment of the tweet.

A topic is composed by the set of tweets discussing about it and a sentiment classification. Accordingly, we define a topic $T$ as a data structure similar to the one introduced to represent a tweet, where in the feature vector $fv$ are stored the textual features of each tweet $tw$ assigned to the topic, and in the sentiment feature vector $sfv$ are stored the semantic features of the tweets, as formalized in the following:

*Definition 3:* (Topic) A topic is a tuple $T = (t, \mathcal{TW}, fv, sfv, S)$, where $t$ is the topic label, $fv$ and $sfv$ are the textual and sentiment feature vectors, respectively, analogous to the ones defined for the tweet, $S$ is the topic overall sentiment.

## IV. METHODOLOGY

In the subsequent sections, we offer an in-depth explanation of the primary steps in our approach.

In contrast to prevailing methods that address the issue of false information with reference to single post concerning COVID-19 discussions, our approach offers a more nuanced analysis allowing us to delve deeper and scrutinize the impact of false information on distinct topics that emerge during these discussions.

By incorporating NLP techniques and sentiment analysis for false information detection and topic modeling, our approach not only aids in pinpointing specific instances of false information but also sheds light on the underlying factors and mechanisms that contribute to its proliferation. This comprehension is pivotal for formulating targeted interventions and strategies aimed at effectively countering the spread of false information.

### A. Topic-aware Sentiment analysis

The proposed methodology is designed as a set of concatenated steps. The first step consists of identifying the topics of discussion. After that the sentiment of the topics has been determined and used by the fake news detector that establish whether or not the tweet is real or fake. Those two steps are implemented by means of the *Topic-specific Sentiment Classifier* (TSC) algorithm.

The *Topic-specific Sentiment Classifier* (TSC) algorithm is formulated as a pipeline model combining two main machine learning methods: an unsupervised learning approach for topic modeling and a supervised classification algorithm to perform sentiment analysis with reference to the topics.

For detection of topics from streaming data, an online clustering algorithm is proposed. Standard approaches to topic modeling like LDA requires that the number of topics must be set beforehand as well as the data structure and dimension. Clearly, in dynamic and evolving settings like social media this it is not feasible. To overcome this drawback in the proposed approach we extract a set ok key features from the tweets that are incrementally grouped through a clustering algorithm relying on co-occurences of bursty textual features (e.g., hashtags and words). To find the sentiments of the detected topics, a classifier that considers the overall sentiment of all the tweets in the topics is proposed. The algorithm,

given a timestamp $t$, first extracts a set of features from the tweets for the $k$ preceding temporal values of $t$ and model each feature as a time series. Then, an ad-hoc peak detection method analyzes the time series to find peaks/bursty. After that, a clustering algorithm will group the bursty textual features to identify trending topics of discussion related to the COVID-19 outbreak. The clustering approach is based on co-occurences of the keywords in the tweets. A preliminary version of the topic detection approach has been presented in [4].

The proposed approach overcomes the limitations of current state-of-the-art that mainly focus on extracting the sentiment from each individual tweets. In fact, even if sentiment classification is a widely studied research line, sentiment analysis of topics is not investigated adequately. Current research focused mainly on performing sentiment analysis at document and sentence level. Document-level sentiment analysis extracts the overall polarity of a whole document while sentence level focuses on fine-grained analysis where each sentence is treated as an independent element and the approach is based on the assumption that the sentence refers to one opinion. Accordingly, traditional sentiment analysis focuses on classifying the overall sentiment expressed in a text (that could be a document or a sentence) without specifying what the sentiment is about. This may not be enough if the text is simultaneously referring to different topics, possibly expressing different sentiments towards different topics.

### B. Fake News Detector

In this section is reported the approach used to identify fake tweets within each topic of discussion. The methodology exploits supervised classifiers that are trained with already labeled data and the features characterizing the topics. Specifically, both content and sentiment features are used. For each tweet the classifier explores the features associated to the topics to which the tweets belongs and produces a classification about the veracity of the tweets. After this phase, each topic is characterized also by the thruth label of the tweets composing it. The algorithm takes a tweet and explits only the featues of the topics to which the tweet belongs, both content and sentiment features are considered. It is also important to note that also the labeled tweets are gouprd in to topic in order to improve the detection phase.

A significant aspect of this research involves the extraction of features from the textual content and subsequently utilizing these features for the detection of fake news, rather than analyzing the raw text. In order to optimize the performance of machine learning algorithms, a specific subset of features, encompassing linguistic and sentiment characteristics, is extracted using topic modeling. In particular, the features used are the ones obtained by the topic detector, both linguistic and content features a well as the set of sentiment features extracted after the sentiment analysis step. Subsequently, these extracted features are employed in the training of various state-of-the-art machine learning algorithms, including the XG-Boost, random forest classifier, AdaBoost classifier, decision tree classifier, and KNN classifier. An analysis of the machine learning algorithms' performance is conducted both before and after the feature extraction process.

## V. TWITTER DATASETS

The approach has been validated on a set of tweets posted in the United States. Specifically, two different datasets has been used. The first dataset is the one collected within the CoronaVis project [7] and accessible from the Github repository (https://github.com/mykabir/COVID19). The dataset consists of over 200 million tweets related to COVID-19 posted by 30.070 unique users in the period March 2020 - June 2021. Among the keywords used to filter COVID-19 tweets there were *corona, pandemic, lockdown, quarantine, virus, pneumonia, outbreak*, etc.

Since some days in the faced period were missing due to connectivity issues, the dataset in [7] has been integrated with tweets collected in the GeoCOV19Tweets dataset [8]. Geo-COV19Tweets consists of 675,104,398 tweets (available from the web site https://ieee-dataport.org/open-access/coronavirus-covid-19-geo-tagged-tweets-dataset) and contains IDs and sentiment scores of the geo-tagged tweets related to the COVID-19 pandemic.

While the data provided in the CoronaVis dataset [7] is already preprocessed, the GeoCOV19Tweets set required a preliminary preprocessing step to remove reetweets, punctuation, stop words and emojis. Moreover, stemming has also been performed.

## VI. QUALITATIVE RESULTS

The section presents a quantitative analysis of the experimental results, outlining the most prevalent discussion topics, users' sentiment towards these topics, and which of these topics are impacted by the spewading of fake news.

*1) Sentiment scores:* To compute the sentiment of the topics and, thus, determine the score of the sentiments, we refer to a set of widely adopted metrics used for sentiment analysis. In particular, we used the package in [5], which classifies each sentiment as either positive (+1 score), negative (-1 score) or neutral (0 score).

Let $N$ be the total number of tweets containing a keyword $k$, $N_{pos}$, $N_{neg}$, and $N_{neutral}$ the number of positive, negative, and neutral tweets regarding $k$, respectively.

$Polarity$ is the ratio between the number of positive tweets and the number of tweets that express a sentiment about $k$.

$$polarity = \frac{N_{pos}}{N_{pos} + N_{neg}} \quad (1)$$

$Subjectivity$ gives the fraction of not neutral tweets with respect to the total number of tweets.

$$subjectivity = \frac{N_{pos} + N_{neg}}{N} \quad (2)$$

$Positive$ is the ratio between the number of positive tweets and the total number of tweets.

$$positive = \frac{N_{pos}}{N} \quad (3)$$

*Negative* is the ratio between the number of negative tweets and the total number of tweets.

$$negative = \frac{N_{neg}}{N} \qquad (4)$$

*2) Results:* In the rest of the section we present the results obtained by using the topic-specific classifier, aiming at analyzing people sentiment concerning COVID-19 as expressed in Twitter.

We start the analysis by showing in Table I the top 12 topics of discussion detected from the considered dataset of tweets. Table shows for each topic a brief description, the top frequent content features and top frequent sentiment features.

As can be noted from Table I, the top topics of discussion span from the initial days of the outbreaks with conversations about the novel virus appeared in China and spreading worldwide, characterized of sentiments of fear, alert and stress caused by COVID-19, due to its quick spread , and continue with a variety of topics of different natures and expressing a variety of sentiments. We can observe that there are topics discussing the preventive measures adopted to fight COVID-19 and limit its diffusion like *Lockdown, Quarantine, Medical supplies* characterized by sentiments of approval/ disapproval, efficacy or inadequacy of the adopted measures; the curative measures used to fight COVID-19 (e.g., hospital overwhelmed, testing center, drugs) expressing again agreement/disagreement, efficacy but also fear, worry, skepticism; similar opposite sentiments characterize the topics arisen around the research efforts started to study and fight the pandemics (e.g., vaccines, virus research); topics concerning the virus tracking and surveillance like *Cases, Deaths, Testing Center* with sentiments of fear, worry, alarm, disappointment; other topics discussing about the consequences brought in people life like *Remote working, Crisis* with negative sentiments of worry and fear. The topic *Medical supplies* concerns tweets about the importance of facial masks and gloves as prevention measures to reduce the outbreak and also their shortage in several countries. Tweets about quarantining people infected or suspected to have COVID-19 are grouped in the topic *Quarantine*.

As evidenced by the table, the topics exhibit varying distributions of sentiment. For instance, in certain topics like "vaccination," there is no prevailing sentiment. On one hand, there is a positive sentiment, reflecting the population's hope and trust in vaccines. On the other hand, there is concern among some individuals regarding the potential side effects of vaccines, such as with the "anti-vax" sentiment, where no prevalent score is observed.

In the following are reported some examples of tweets, together with the sentiment, about some of the top popular topics shown in the above Table.

- Topic: Covid-19 outbreak. Tweet: "The coronavirus was created in a lab in Wuhan, China. It was a bioweapon that was accidentally released. The Chinese government is covering up the truth. #coronavirus #china #bioweapon". Sentiment: Negative

- Topic: Medicine. Tweet: "I have been taking hydroxychloroquine for two weeks and I feel great. It is a miracle drug that cures COVID-19. The mainstream media and the WHO are lying to you. #covid19 #hydroxychloroquine #miracle". Sentiment: Positive.

- Topic: Lockdonw. Tweet: "Please stay home and stay safe. Wear a mask when you go out. Wash your hands frequently. We can beat this virus together. #covid19 #stayhome #staysafe". Sentiment: Positive

- Topic: Testing Center. Tweet: "I tested positive for COVID-19 yesterday. I have mild symptoms and I am isolating at home. I hope I recover soon. Please pray for me and my family. #covid19 #positive #pray". Sentiment: Neutral

As a final analysis, in the following are reported some examples of false or positive tweets about some of the top popular topics shown in the above Table.

- False Claim: "COVID-19 Vaccine Contains Microchips for Government Surveillance"
  Explanation: This false claim suggests that COVID-19 vaccines, particularly those developed by pharmaceutical companies, contain microchips that can be used for government surveillance purposes. This conspiracy theory implies that individuals who receive the vaccine are unknowingly being tracked and monitored.
  Real Information: COVID-19 vaccines do not contain microchips for surveillance. These vaccines have undergone rigorous testing and scrutiny by health authorities worldwide, including the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA). The vaccines are designed to stimulate an immune response against the virus, not to track individuals.
  Impact: False claims like this can contribute to vaccine hesitancy and undermine public trust in vaccination efforts. Such misinformation can have serious consequences during a global pandemic, as it may deter people from getting vaccinated, which is crucial for achieving herd immunity and controlling the spread of the virus.
  This example underscores the importance of fact-checking and relying on trusted sources for accurate information during the COVID-19 pandemic. It's essential to combat the spread of such false claims to ensure public health and safety.

- False Claim : "5G Technology Causes COVID-19"
  Explanation: This conspiracy theory falsely suggests that 5G technology is responsible for causing or spreading the COVID-19 virus. It alleges that the rollout of 5G networks led to the pandemic.
  Real Information: There is no scientific evidence supporting the idea that 5G technology is connected to the COVID-19 virus. COVID-19 is caused by the SARS-CoV-2 virus, a novel coronavirus, and its transmission occurs through respiratory droplets.
  Impact: This false claim has led to vandalism of 5G infrastructure and fueled unnecessary fear and confusion

**TABLE I:** Top 12 topics of discussion about COVID-19 in US.

| Topics | Content features | Sentiment features |
|---|---|---|
| Covid-19 outbreak | 2019-nCov, novel coronavirus, corona, Covid-19, outbreak, China ,pandemic | spread, emergency,fear, worry, alert, stress , scared |
| Lockdown | measure, lockdown, quarentine, prevention, stop, stayhome | ability, assurance, safe alarm, abuse, burden |
| Cases | case, cases, confirmed suspect, lag | advantage, accept, accuse accusation, crisis, bailout, fear |
| Death | dead, died, kill, death, sick, ill | fear, alarm, worry, relieved, stress |
| Hospital overwhelmed | hospital full, nursing, doctors, ill, ventilators, no bed, shed | help,care, alarm, alert, worry, effective, manage, readiness |
| Quarantine | home, close, lockdown, alone, inside, stay inside | benefit, assure, respect, follow, escape, avoid, anti, stressed |
| Medical supplies | mask, disinfection, protection, shortage, gloves, gel | effective, mandatory, annoying, tired, difficult |
| Virus research | detection, research, laboratory, treatment, vaccine | hope, efficacy, confident, negative, difficult |
| Testing center | result, test, tested, positive, negative, quantitative, symptoms | pain, relief, worried, scared |
| Vaccines | vaccine, immune, antibody, mouse, death vaccination, animal, protective, shot | against, pro, no-vax, skeptical, effective, anti,danger |
| Remote Work | remote, work, remote working, office close, cancel school, home | admit, agree, mandatory, annoying, tired, difficult |
| Crisis | emergency, crisis, pandemic, policy, strategy, economic, loss | prevent, capacity, effectively, manage, strengthen, support, readiness, anger |
| Medicines | Remdesivir, Tocilizumab, Clorochine, treats, hydroxychloroquine | efficacy, concern skeptical, danger |

about the pandemic.

- False Claim : "COVID-19 is Just a Flu, and the Pandemic is a Hoax"
  Explanation: Some false narratives minimize the severity of COVID-19 by equating it to the seasonal flu and falsely claim that the pandemic is a hoax or exaggerated by governments.
  Real Information: COVID-19 is significantly more contagious and can lead to severe respiratory illness and death. It has overwhelmed healthcare systems worldwide and is a genuine public health crisis.
  Impact: Downplaying the pandemic's seriousness can discourage safety measures and hinder efforts to control the virus's spread.
- False Claim : "Drinking Silver Colloidal Kills COVID-19"
  Explanation: Some false information suggests that drinking colloidal silver can cure or prevent COVID-19. Colloidal silver is promoted as a home remedy to treat various health issues.
  Real Information: Colloidal silver is not a proven treatment for COVID-19 and can be harmful. The World Health Organization (WHO) and the U.S. FDA advise against its use.
  Impact: Believing in such false claims can lead to unsafe practices and delay seeking appropriate medical care.

These examples demonstrate the importance of critical thinking, reliable sources, and fact-checking when encountering information related to COVID-19. Misinformation can have serious consequences, and it's crucial to rely on credible health organizations and experts for accurate information about the pandemic.

## VII. PERFORMANCE RESULTS

In this section are presented the results of the experiments carried out to evaluate the performance of the proposed fake news detection approach.

*1) Performance metrics:* Different evaluation metrics such as precision, recall and f-measure, were used to assess these results:

- *Recall*: $R = \frac{TP}{TP+TN}$. It is the percentage of ground truth topics successfully classified by the method.
- *Precision*: $P = \frac{TP}{TP+FP}$. It is the fraction of correctly classified ground truth topics out of the total number of topics classified.
- *F-measure*: $F\text{-}measure = \frac{2RP}{R+P}$ is the harmonic mean of precision and recall metrics.

*2) Results:* Table II and Table III report the results of the fake news detection.

Table II clearly illustrates that the ensemble classifier like XGBoost and Ada-boost outperformed other machine learning algorithms, such as Decision Tree (DT) and K Nearest Neighbor (KNN), across various performance metrics, including accuracy, precision, recall, and F1-Measure scores. Prior to feature extraction, the Ada-boost classifier achieved the highest prediction accuracy of 79.88% among all machine learning classifiers. Furthermore, XGBoost demonstrated precision, recall, and F1-Measure scores of 76.76%, 86.36%, and 81.82%, respectively. In contrast, DT and KNN exhibited prediction accuracies of 67.81% and 62.06%, along with precision, recall, and F1-Measure scores as follows: DT - 70.51%, 62.50%, and 66.26%; KNN - 72.91%, 39.77%, and 51.47%. It is noteworthy that traditional machine learning algorithms performed exceptionally well in comparison to the proposed deep learning models detailed in Section 3. Consequently, the introduction of novel text-based features is expected to significantly enhance the performance of our proposed model.

Table III shows the performance of the ML algorithms after feature extraction. Again the ensemble based classifiers outperform the other approaches. XGBoost reaches an accuracy of 88.74%, precision 89.20%, recall 89.28% and f-score 89.24%, the Ada-boost classifier achieved a remarkable prediction accuracy of 82.75%, while DT and KNN exhibited prediction accuracies of 77.58% and 69.54%, respectively. Furthermore, Ada-boost demonstrated impressive precision, recall, and F1-

**TABLE II:** Performance metrics of COVID-19 fake news detection using machine learning classifiers, before feature extraction

| Method | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| XGBoost | 79.88 | 76.76 | 86.36 | 81.82 |
| Adaboost | 67.81 | 70.51 | 62.50 | 66.26 |
| Decision Tree | 66.5 | 63 | 60 | 62 |
| KNN | 62.06 | 72.91 | 39.77 | 51.47 |
| Random Forest | 63.64 | 63.11 | 73.48 | 67.90 |

Measure scores of 79%, 89.77%, and 84.04%, respectively. DT displayed precision, recall, and F1-Measure scores of 72.47%, 89.77%, and 80.20%, whereas the KNN classifier achieved precision, recall, and F1-Measure scores of 62.41%, 100%, and 76.85%, respectively. Post feature extraction, which encompassed both deep learning (DL) and machine learning (ML) techniques, it becomes evident that deep learning outperforms machine learning algorithms in terms of precision, recall, F1-Measure, and prediction accuracy.

**TABLE III:** Performance metrics of COVID-19 fake news detection using machine learning classifiers, after feature extraction

| Method | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| XGBoost | 88.74 | 89.20 | 89.28 | 89.24 |
| Adaboost | 82 | 84 | 76 | 80 |
| Decision Tree | 77.58 | 72.47 | 89.77 | 80.20 |
| KNN | 69.54 | 62.41 | 100 | 76.85 |
| Random Forest | 82.75 | 79.00 | 89.77 | 84.04 |

The results section clearly demonstrates that the performance of machine learning algorithms improves when trained with features extracted from the COVID-19 fake news dataset. The raw fake news dataset contains numerous words that do not significantly impact the classification results. Training the machine learning algorithms using the raw dataset, without feature extraction, increases the likelihood of these algorithms being influenced by common words in the text that do not contribute to classification outcomes. These extracted features were subsequently employed in the training of machine learning algorithms. A comparison of the results before and after feature extraction demonstrates a clear enhancement in the performance of the machine learning algorithms. Given the dataset's size, comprising approximately 1,100 records, the choice was made to employ machine learning algorithms for classification rather than deep neural network-based approaches.

The research aimed to assess the efficiency of diverse machine learning techniques for COVID-19 detection. Although prior studies have demonstrated the prowess of such models models in recognizing COVID-19-related information, there has been relatively limited exploration of how these models can be integrated with topic-focused architectures for COVID-19 detection. Consequently, this study extends the current body of research on machine lerning classifier fake newd detector by investigating their performance in the realm of COVID-19 detection alongside topic modeling and sentiment analysis structures. The outcomes of this study align with prior research regarding the efficacy machine learning models, including ensemble approaches.

## VIII. Conclusion

Topic modeling, when coupled with sentiment analysis, helps to find the ongoing topics being discussed on social media platforms improving sentiments understanding. Based on the user's sentiment towards the detected topics, public authorities and governments could make informed and appropriate decisions and devise the most suitable strategies.

In this work a topic-level sentiment analysis approach based on topic modeling and supervised machine learning classification models has been proposed. The proposed approach supports scalable and dynamic topic detection over streaming social media data and performs sentiment analysis at topic-level.

To discover relevant COVID-19 topics of discussion from Twitter, the work presented an approach that combines identification of bursty features and online clustering. Topics are incrementally obtained by grouping the bursty textual features based on the co-occurrence in the tweets. Each cluster obtained is then associated to a topic. After a topic has been detected, a multi-label classifier computes its sentiment. The novelty of the proposed approach is that it works at the topic level to extract sentiments from social media data, differently to literature approaches that instead extract the sentiments at the tweet-level. The proposed approach has the ability to detect multiple topics and associated sentiments from streaming social media posts enabling false information detection allowing for misinformation detection.

In order to address this challenge, we have developed a semi-supervised methodology that integrates false information detection with a sentiment-aware topic modeling algorithm. Initially, linguistic and sentiment features are extracted for subsequent analysis. Following this, a set of state-of-the-art machine learning algorithms are employed to classify the COVID-19-related dataset. Results, performed over a real-world dataset of tweets, shown the feasibility of the method that is able to detect a large number of relevant COVID-19 topics and extract false information from such topics.

## References

[1] A. Abd-Alrazaq, D. Alhuwail, M. Househ, M. Hamdi, and Z. Shah. Top concerns of tweeters during the covid-19 pandemic: Infoveillance study. *J Med Internet Res*, 22(4):e19016, Apr 2020.

[2] D. M. Blei, A. Y. Ng, M. I. Jordan, and J. Lafferty. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:2003, 2003.

[3] C. Comito. How covid-19 information spread in us the role of twitter as early indicator of epidemics. *IEEE Transactions on Services Computing*, pages 1–1, 2021.

[4] C. Comito, D. Falcone, and D. Talia. A peak detection method to uncover events from social media. In *2017 IEEE International Conference on Data Science and Advanced Analytics, DSAA*, pages 459–467. IEEE, 2017.

[5] I. Feinerer, K. Hornik, and D. Meyer. Text mining infrastructure in r. *Journal of Statistical Software*, 25(5):1–54, 3 2008.

[6] E. Ferrara. What types of covid-19 conspiracies are populated by twitter bots? *First Monday*, May 2020.

[7] M. Y. Kabir and S. Madria. Coronavis: A real-time covid-19 tweets analyzer, 2020.

[8] R. Lamsal. Coronavirus (covid-19) geo-tagged tweets dataset.

[9] L. Li, Q. Zhang, X. Wang, J. Zhang, T. Wang, T. Gao, W. Duan, K. Tsoi, and F. Wang. Characterizing the propagation of situational information in social media during covid-19 epidemic: A case study on weibo. *IEEE Transactions on Computational Social Systems*, 7(2), Apr. 2020.

[10] C. Ordun, S. Purushotham, and E. Raff. Exploratory analysis of covid-19 tweets using topic modeling, umap, and digraphs, 2020.

[11] K. R, A. J. J, K. A, E. A. MB, K. B, A. E, Z. J, T. C, A. EW, and B. K. Coronavirus goes viral: Quantifying the covid-19 misinformation epidemic on twitter. *Cureus*, 12(3), 2020.

[12] M. T. Rashid and D. Wang. Covidsens: A vision on reliable social sensing for covid-19, 2020.

[13] L. Schild, C. Ling, J. Blackburn, G. Stringhini, Y. Zhang, and S. Zannettou. "go eat a bat, chang!": An early look on the emergence of sinophobic behavior on web communities in the face of covid-19, 2020.

[14] 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.

[15] K. Sharma, S. Seo, C. Meng, S. Rambhatla, and Y. Liu. Covid-19 on social media: Analyzing misinformation in twitter conversations, 2020.

[16] L. Singh, S. Bansal, L. Bode, C. Budak, G. Chi, K. Kawintiranon, C. Padden, R. Vanarsdall, E. Vraga, and Y. Wang. A first look at covid-19 information and misinformation sharing on twitter, 2020.

764