Covid-19 Fake News Detection: A Survey

Elena Shushkevich[1], Mikhail Alexandrov[2,3], John Cardiff[1]

[1] Technological University Dublin,
Dublin, Ireland

[2] Russian Presidential Academy of National Economy and Public Administration,
Moscow, Russia

[3] Autonomous University of Barcelona, Barcelona,
Spain

elena.n.shushkevich@gmail.com, John.Cardiff@TUDublin.ie, MAlexandrov@mail.ru

Abstract. The increase of fake news in social media, especially about Covid-19, poses a real threat to the mental and physical health of people. It is an important task to detect such news and to stop it spreading. In this article, we describe the main approaches for fake news about Covid-19 detection, including Classical Machine Learning models, models based on Neural Networks and models, which were created based on the other approaches and preprocessing steps. We analyze the results of the challenge "Constraint@AAAI2021 - COVID19 Fake News Detection", the main goal of which was the binary classification of news collected from social media for fake and real news. We analyze the best approaches, which were proposed by researchers during the challenge. In addition, we describe datasets of fake news related to Covid-19, which could be useful for the detection and classification of such news.

Keywords. Fake news, Covid-19, classical machine learning models, neural networks, text transformers.

1 Introduction

With the increase of the role of social media in our lives, the fake news, which appear in social media, have become a serious problem. Fake news are information hoaxes designed to deliberately mislead the reader in order to gain a financial or political advantage [1]. Such news can cause as much mental as physical harm, so it is very important to control fake news and prevent their spreading.

There are several very important works [2,3] which aim to highlight the main approaches used by researchers for fake news detection. Speaking of fake news here, we mean fake news related to different fields of human activity such as politics, economy, goods and services, travels and tourism, etc. It should be noted that the range of such approaches is quite wide and includes classical machine learning models, such as Text Transformers, and other models based on neural networks. Furthermore, we note the high importance of fake news detection and classification at speed, as such news should be detected promptly.

Especially important is to detect and prevent spreading of fake news about Covid-19 - an urgent and global problem, which appeared in the last two years with the pandemic and is growing at lightning speed. In [4] it was shown that the huge number of hoaxes and misinformation leads to an increase in the spread of the virus and a decrease of mental health of individuals. To prevent such dramatic consequences the authors suggest creating more close communications between mass media, healthcare organizations, and other important stakeholders and create a unified platform for actual true health-related news spreading. In addition, it is important to use AI, including natural language processing tools, to indicate and delete fake news of online content from all social media platforms and create special law enforcement measures to control such fake news spreading.

The main goal of this article is to highlight possible approaches, recent research and useful datasets for fake news detection about Covid-19,

which could help new researchers understand the topic better.

The article is structured as follows: in the Introduction section, we give a definition to fake news and describe the danger of such news and the fake news about Covid-19. In Section 2, we define the different approaches for fake news detection about Covid-19, including classical machine learning models, models based on neural networks and models, which were created according to other approaches. Section 3 is devoted to the challenge "Constraint@AAAI2021 - COVID19 Fake News Detection", and we study the task of the challenge, dataset and models which achieved the best results on it. In Section 4, we describe datasets for detection of fake news connected with Covid-19. Section 5 is the conclusion, in which we sum up the findings of the survey.

2 Approaches to Covid-19 Fake News Detection

When we speak about the best models for Covid-19 fake news detection, we should pay attention to the structure of the review of such models, and to combine them in some groups by their nature, to make the review easier to read. First of all, it should be noted that there is not a strict division for models' types: usually, researchers work with a big number of different models and combinations (ensembles) of models, with different preprocessing steps, trying to perform the best combinations.

Despite that fact, we can conditionally distinguish the main three groups of models: classical machine learning models, models based on neural networks, and models based on the other approaches. We review all three groups of models in this section in detail.

2.1 Classical Machine Learning Approach for Detection of Fake News about Covid-19

The most commonly used algorithms for the fake news detection, and fake news related with Covid-19 in particular, are classical machine learning algorithms. They are really helpful in case of binary classification, when we need to identify if a message is fake news or real news. In a range of classical machine learning algorithms it is reasonable to highlight Logistic Regression with the basic idea of linear classifier, which is able to divide a feature's space into two spaces by a hyperplane, and each half-space is reflecting each class of a binary classification. Support Vector Machines are very popular in the framework of classification problems, and this classifier is creating a hyperplane or set of hyperplanes in a multidimensional or infinite-dimensional space. For fake new detection, methods, which are based on classification trees, are also used. The main representatives of such methods are Gradient Boosting Classifier and Random Forest Classifier. The Random Forest Classifier in contrast to Gradient Boosting Classifier builds each tree independently and combines results at the end of the process.

In the context of using classical machine learning algorithms to identify fake news about Covid-19 the authors [5] analyzed fake news connected with COVID pandemic. They collected a dataset from 150 users, extracting data from their social media accounts including Twitter, email, mobile, Whatsapp and Facebook for 4 months from March 2020 to June 2020. At the preprocessing stage they removed information not related to the Covid-19 data and deleted incomplete news. K-Nearest Neighbour was used to perform the classification. As a result, they achieved the best prediction results for June with 0.91 F1-score and the worst ones for March with 0.79 F1-score.

In [6,7] the researchers aimed to detect fake news connected with Covid-19 on small datasets. The researchers compared Logistic Regression, Support Vector Machine, Gradient Boosting and Random Forest on a limited dataset of 1,000 fake and real messages. Support Vector Machine and Random Forest classifiers performed the best results with 69% micro-F1 score. The authors noted that despite the fact that the results are not as strong as the results obtained on the full dataset of fake news about Covid-19, such an approach could be helpful for the researchers who do not have a big enough dataset or sufficient time to collect a big dataset, but where the classification decisions have to be taken in the shortest span of time possible.

In [8] the authors compare four machine learning baselines (Decision Tree, Logistic Regression, Gradient Boost and Support Vector Machine (SVM)) on the dataset devoted to fake news about Covid-19 detection. The authors collected themselves from Facebook, Twitter and other social media platforms. The best result - 0.93 F1-score - was achieved with the SVM model.

From the above, we can conclude that there are a lot of algorithms of machine learning, which showed quite good results on the binary classification for fake and real news.

2.2 Neural Networks for Covid-19 Fake News Detection

There is a large variety of algorithms based on neural networks which have been successfully applied to fake news detection, including that related to Covid-19. Researchers use a lot of interesting linguistic models, the most important of which is BERT [9], which stands for Bidirectional Encoder Representations from Transformers, and which can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks.

DistilBERT [10] model allows to reduce the size of a BERT model by 40%, while retaining 97% of its language understanding capabilities and being 60% faster.

COVID-Twitter-BERT (CT-BERT) model [11] was pretrained on a large corpus of Twitter messages on the topic of COVID-19 and has proved very useful in case of detecting such messages.

In addition, an important model which is often used is RoBERTa [12], which is a robust BERT trained on a larger dataset for a larger number of iterations and with a larger batch size.

Some interesting and useful models are ELECTRA [13] – Pre-training Text Encoders as discriminators rather than generators, and AlBert [14] – a Lite BERT for Self-supervised Learning of language representations.

There is the XLNet model [15], which is similar to BERT but learns bidirectional context along with autoregressive formulation.

Hierarchical Attention Networks (HAN) is based on LSTM and comprises four sequential levels –

word encoder, word-level attention, sentence encoder and sentence-level attention [16].

ELMO [17] is a deep contextualized word representation that models both complex characteristics of word use (e.g. syntax and semantics), and how these uses vary across linguistic contexts (i.e. to model polysemy).

A successful use of neural networks for Covid-19 fake news detection is shown in the paper [18]. Its authors collected 4.8K expert-annotated social media posts related to COVID-19 and 86 common misconceptions about the disease to evaluate the performance of misinformation detection related to the Coronavirus pandemic. Each message was labeled in a context of chosen misconceptions as "misinformative" (tweet is a positive expression of the misconception), "informative" (tweet disagrees with the misconception), or "irrelevant" (tweet is not relevant to the misconception).

For the modelling, the authors used TF-IDF and GloVe parameterizations to obtain vectorized representations. Researchers used the RoBERTa-base implementation with two models of textual similarity (cosine similarity computed by averaging over the token vectors, and BERTScore which involves adding cosine similarities between RoBERTa token embeddings) to obtain contextual word embeddings. In addition, the authors tuned the RoBERTa-base model using the dataset of tweets connected with COVID-19.

As a result, the domain-adapted BERTScore achieved the best results among the similarity models.

In [19] the authors created an ensemble of linguistic models XLNet, RoBERTa, XLM-RoBERTa, DeBERTa, ERNIE 2.0, and ELECTRA for the task of Covid-19 fake news detection. They also used a tweet-preprocessor library from Python [20] to filter out noisy data such as usernames, URLs, emojis, etc.

Furthermore, the authors implemented Heuristic Post-Processing, which takes Soft-voting prediction vectors into account. Thanks to the preprocessing step, the construction of such an ensemble, and by choosing the soft-voting through prediction vectors instead of the hard-voting approach, the researchers achieved an F1-score of 0.9831.

To deal with a problem of Covid-19 fake news detection, the authors of [21] implemented a

combination of topical distributions from Latent Dirichlet Allocation (LDA) with contextualized representations from XLNet.

For the implementation, the authors used the Transformers library[1] maintained by the researchers and engineers at Hugging Face [22], which provides the PyTorch interface for XLNet. The created model allowed researchers to achieve 0.967 F1-score on the test dataset. In addition, to compare the results the authors implemented SVM model with contextualized representations of the input using Universal Sentence Encoder (USE)[2], BERT with document-topic distributions from LDA, fine-tuned pretrained XLNet model, and the combination of BERT and BERT+topic models. During the comparison, the model of XLNet with topic distributions showed the best results.

2.3 Other features for Covid-19 Fake News Detection

Both classical machine learning models and neural networks play an important role in the problem of fake news detection connected with Covid-19, but there are a lot of other features and approaches, which are popular in natural language processing and which could be very helpful for fake news detection. For example, we can mention algorithms using n-grams of words or characters [23], GloVe [24], an unsupervised learning algorithm for obtaining vector representations for words, Fast-text [25], the classifier, which is often on par with deep learning classifiers in terms of accuracy, and many orders of magnitude faster for training and evaluation, Label Smoothing – a regularization technique that introduces noise for the labels [26]; adversarial training [27], a novel regularization method for classifiers to improve model robustness for small, approximately worst case perturbations; and tax2vec [28], a Semantic space vectorization algorithm.

There are some interesting researches describing the impact of fake news about Covid-19, for example [29], which is devoted to researching the social impact of fake news in a context of health information in social media. Researchers also try to define false information and information, which

is the evidence of the social impact shared in social media.

The authors analyzed data from three social media platforms: Twitter, Facebook and Reddit. The authors' research questions were about ways that social media messages focused on fake health information, and how interactions based on health evidence with social impacts help overpower fake health information.

The authors used a methodology of social impact in social media, which combined quantitative and qualitative content analysis of the collected messages. The messages' selection corresponded to three criteria: relevance of the number of active users in millions according to Statista 2019 data, availability of public messages, and suitability for online discussion. The selection data contained the hashtags "health", "vaccines", "nutrition" and "Ebola", which matched well with the interests of researchers.

The authors divided the data in four groups: ESISM (the message is an example of evidence of the social impact shared in social media), MISFA (the message is a fake news), OPINION (the message is a user's opinion) and INFO (the message is a fact or news).

As a result, the authors found that misinformation and fake news is higher in Twitter (19%) than in Facebook (4%) or Reddit (7%). The researchers also found that messages focused on false health information are mostly aggressive in the analyzed data, while messages based on evidence of social impact are more peaceful and respectful.

In addition, messages based on evidence of social impact successfully challenged false information in cases when the authors of the misinformation messages were polite and open to discussion. However, this is not the case when the authors of misinformation messages showed disrespect and an aggressive position against science.

In [30], the authors used natural language processing to analyze messages connected with COVID-19 and to indicate the main topics, which people discuss in social media in the frames of the pandemic. To build the dataset, the researchers

---

[1] https://github.com/huggingface/transformers

[2] https://tfhub.dev/google/universal-sentence-encoder-large/3

extracted messages and comments from Twitter, Youtube, Facebook and three online discussion forums.

At the preprocessing stage the authors removed URLs, HTML tags, number words, special characters that are not required for sentence boundary detection, hash tags, and mentions.

The authors expanded contractions, compressed words with repeated characters, and converted slangs to English words.

Afterwards, the researchers implemented the Keyphrase Extractor algorithm using Python, which contains 7 steps: (1) grammar determination (where all words are marked as corresponding parts of speech); (2) sentence breaking and tokenization (where documents are separated into sentences using Python NLTK's tokenize library and sentences into words or tokens); (3) POS tagging (where each token is assigned a POS tag); (4) lemmatization (where each token is converted to its root word); (5) chunking (where words are combined in chunks using special rules); (6) transformation and filtering (where stopword keyphrases are deleted); and (7) sentiment scoring and filtering (the last step where the authors assigned a sentiment score to each keyphrase using VADER lexicon-based algorithm).

As a result, the authors using keyphrases and experts assessments revealed 34 negative and 20 positive categories. The top-5 negative themes included concerns about social distancing and isolation policies, misinformation, political influence, financial issues, and poor governance. The top-5 positive themes connected with the pandemic were public awareness, spiritual support, encouragement, charity, and entertainment.

3 Constraint@AAAI2021 - COVID19 Fake News Detection

Despite the fact that today some challenges include Covid-19 related fake news in their datasets, like IberLEF-2021 shared task [31], which includes 237 Covid-19 fake news, we have found only one challenge devoted exclusively to the detection of fake news about Covid-19. In this section, we discuss the challenge

"Constraint@AAAI2021 Fake News Detection" [32], which is very important in a context of fake news related with Covid-19 detection, because it proposed the participants a dataset of such fake news and the participants of the challenge showed a big variety of models aimed to detect fake news about Covid-19. We describe the task, the dataset and the best models, which were created in the frames of the challenge.

3.1 Challenge and Dataset

The main goal of the mentioned challenge was to create a system for binary classification of fake news and real news. The challenge contained the same task for two languages: English and Hindi, but in this survey, we concentrated our attention only on the English task and dataset.

The dataset for Covid-19 fake news detection contains 10,700 messages, with 5,100 of them real and 5,600 of them fake.

Real news was collected from reliable sources such as the World Health Organization (WHO), Centers for Disease Control and Prevention (CDC), and others, whereas fake news was collected from social media such as Facebook posts, Twitter tweets, Instagram posts, etc. The dataset contains 37,503 unique words, and the principal statistics are presented in Table 1.

Table 1. Numeric features of the dataset

Attribute	Fake	Real	Combined
Unique words	19728	22916	37503
Avg. words per post	22	32	27
Avg. chars per post	143	218	183

The interesting observation from the dataset is that on average real news tend to be longer than the fake ones by approximately 10 words and the number of characters in a fake message is lower than in the real one by nearly 75 characters.

Some examples of real and fake news from the dataset (which were collected from Twitter) are presented in Table 2.

Table 2. Examples of fake and real news

Label	Text
Fake	*1972 #Watch Italian Billionaire commits suicide by throwing himself from 20th Floor of his tower after his entire family was wiped out by #Coronavirus #Suicide has never been the way, may soul rest in peace May God deliver us all from this time*
Fake	*2 Trump announced that Roche Medical Company will launch the vaccine next Sunday and millions of doses are ready from it !!! The end of the play*
Fake	*China Muslims hidden at Bihari mosque has been taken to corona virus test by Bihari police. Erode police has caught Thailand Muslim mullahs infected with corona virus. Today Salem Police has caught 11 Indonesian Muslim mullahs at Salem mosque. This video shows that they are applying and putting saliva on spoons plates and utensils and also they are in the intention of spreading corona virus disease. Nobody knows what's happening in the Nation*
Real	*Almost 200 vaccines for #COVID19 are currently in clinical and pre-clinical testing. The history of vaccine development tells us that some will fail and some will succeed-@DrTedros #UNGA #UN75*
Real	*14 new cases of #COVID19 have been confirmed in Nigeria: 2 in FCT 12 in Lagos Of the 14 6 were detected on a vessel 3 are returning travellers into Nigeria; 1 is close contact of a confirmed case As at 7:35 pm 26th March there are 65 confirmed cases 3 discharged 1 death*
Real	*Currently most cases of #COVID19 in the US are in California and Washington State. However many other communities are also dealing with cases of COVID-19. See CDC recommendations for preventing spread of COVID-19 in communities.*

The shared task's participants' submissions were ranked according to their weighted average F1-score. F1-score was calculated for fake news' class and real news' class, while the average is weighted by the number of true instances for that class.

3.2 Best Approaches from the Challenge

There were 166 teams who took part in the challenge with the English dataset, and 114 of them beat the baseline of 93% F1-score. The best results are very close to each other and are higher than 98% F1-score. The top three results are presented in Table 3.

Table 3. Top-3 results of the Constraint@AAAI2021 - COVID19 Fake News Detection Shared Task (in %)

Rank	Team	Precision/Recall/F1-score
1	g2tmn	98.69 / 98.69 / 98.69
2	saradhix	98.65 / 98.64 / 98.65
3	xiangyangli	98.60 / 98.60 / 98.60

The winner of the challenge g2tmn team [33] achieved 98.69% F1-score using the ensemble of three pretrained CT-BERT models with random seed values and with different data splitting into training and validation samples. In addition, at the preprocessing step, the researchers used the Python emoji library to replace the emoji with short textual descriptions [34], URLs tokenizations (a replacement with URL token), and converting texts of messages to lowercase.

The second result 98.65% F1-score was achieved by the saradhix team [35]. During the research, the authors used several classical machine learning methods such as Naive Bayes, Logistic Regression, Random Forest, XGBoost, Support Vector Machine, and also several Transformer models, including BERT, DB-BERT, RoBERTa, Electra and XLNet. The best results were obtained using a RoBERTa-based model with 12-layer, 768-hidden, 12-heads, 125M parameters. Researchers noted that RoBERTa is a Robust BERT model, which has been trained on a much larger dataset and for a much larger number of iterations with a larger batch size.

The third place with 98.6% F1-score was obtained by the xiangyangli team [36]. In this case,

the authors also used Text Transformers for their research. Additionally, the authors used a Pseudo Label Algorithm to do data augmentation, because the dataset was quite small. As a test data was predicted with a probability greater than 0.95, the authors assumed that the data was predicted correctly with a relatively high confidence and added it into the training set. The authors used Text Transformers to create the ensemble, which is made of BERT, Ernie, XL-Net, RoBERTa and Electra models, and cross validation with Pseudo Label Algorithm to achieve their best result.

To summarize the best approaches from the Constraint@2021 Fake News Detection open shared task, we see that:

1. The most successful models were created as ensembles of Text Transformers, and

2. The most important step was to fine-tune such Transformers, while the preprocessing steps, which are usually very important for the classical machine learning models, did not play a significant role in these cases.

4 Datasets for Detection of Covid-19 Fake Mews

Previously, Constraint@AAAI2021 - COVID19 Fake News Detection challenge's dataset was described, which is possible to use for the purpose of binary classification for fake news and real news. In this section, we describe some additional datasets for fake news detection about Covid-19.

The CoAID dataset (COVID-19 Healthcare Misinformation Dataset) [37] contains diverse COVID-19 healthcare misinformation, including fake news on websites and social platforms, along with users' social engagement about such news. CoAID includes 4,251 news, 296,000 related user engagements, 926 social platform posts about COVID-19, and ground truth labels.

The first multilingual dataset for Covid-19 fake news is FakeCovid. The Multilingual Cross-domain Fact Check News Dataset for COVID-19 contains news from 150 countries in 40 languages [38]. The dataset includes 5,182 fact-checked news articles for COVID-19, collected between 04/01/2020 and 15/05/2020. The authors used 92 different fact-checking websites having received references from Poynter and Snopes for the articles' collection. The authors indicated 11 different categories of the fact-checked news according to their content.

The possible way to obtain Covid-19 fake news is to use Elasticsearch[3] to retrieve validated fake news from FakeHealth[4] [39]. We also want to highlight COVID-19 Infodemic Twitter dataset [40], where the authors not only prepared the data consisting of tweets annotated with fine-grained labels related to disinformation about COVID-19, but also proposed an annotation schema and detailed annotation instructions for the creation of such datasets.

Finally, in [41] the author describes dataset, where tweets with the hashtag #covid19 are collected using the Twitter API and a Python script. Collection started on 25/7/2020, with an initial 17k batch and is ongoing on a daily basis. It allows this collection to be used for trends analysis or creation of other Covid-19 fake news dataset.

5 Conclusion

The problem of detecting fake news in social media, especially fake news about Covid-19, is now more important than ever before. In this article, we described the main approaches for such news detection, including classical machine learning models, models which are based on neural networks and models, which were created based on other approaches.

We showed that there are effective and useful models and methods in each group. We analyzed the Constraint@AAAI2021 - COVID19 Fake News Detection challenge and the dataset, which the organizers created for it. The results of the challenge are showing that for the task of binary classification for fake news and real news the best approach is to create ensembles of Transformers, such as BERT and CT-BERT.

In addition, we described datasets for fake news detection about Covid-19, which are very helpful for the problem of the detection and classification of such news.

---

[3] https://www.elastic.co/

[4] https://www.healthnewsreview.org/

The large number of fake news on Covid-19 can be viewed as a manifestation of information wars. Models of information wars have already become the subject of consideration of specialists in mathematical modelling, see for example [42, 43]. In future, it would be possible to consider such models to predict the dynamics of fake news and real news of Covid-19.

Finally, we should not forget about models and methods for fake news detection, which were created previously, such as methods that take into account the peculiarities of the styles of texts [44]. Such methods would be useful in case of Covid-19 fake news too, with the assumption that such news are part of fake news in general.

References

1. **Hunt, E. (2016).** What is fake news? How to spot it and what you can do to stop it. The Guardian. Retrieved from https://www.theguardian.com /media/2016/dec/18/what-is-fake-news-pizzagate.

2. **Choraś, M., Demestichas, K., Giełczyk, A., Herrero, Á., Ksieniewicz, P., Remoundou, K., Urda, D., Wozniak, M.(2020).** Advanced Machine Learning techniques for fake news (online disinformation) detection: A systematic mapping study. Applied Soft Computing. 101. 107050. DOI: 10.1016/j.asoc.2020.107050.

3. **Zhou, X., Zafarani, R. (2020).** A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys. 53. DOI: 10.1145/3395046.

4. **Tasnim, S., Hossain, M., Mazumder, H. (2020).** Impact of rumors and misinformation on COVID-19 in social media. J Prev Med Public Health. 53(3):171–174. DOI: 10.3961/jpmph.20.094.

5. **Bandyopadhyay, S., Dutta, S. (2020).** Analysis of fake news in social medias for four months during lockdown in COVID-19. DOI: 10.20944/preprints 202006.0243.v1.

6. **Shushkevich, E. & Cardiff, J. (2021).** Detecting fake news about Covid-19 on small datasets with machine learning algorithms. Proceedings of the 30th Conference of Open Innovations Association FRUCT, pp. 253–258.

7. **Shushkevich, E.,M. Alexandrov, M., Cardiff, J. (2021).** Detecting fake news about Covid-19 using classifiers from Scikit-learn. International Workshop on Inductive Modeling IWIM'2021, 5 pp.

8. **Patwa, P., Sharma, S., Pykl, S., Guptha, V., Kumari, G., Akhtar, M., Ekbal, A., Das, A., Chakraborty, T. (2021).** Fighting an infodemic: COVID-19 fake news dataset. arXiv:2011.03327.

9. **Devlin J., Chang, M., Lee, K., Toutanova, K. (2018).** Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

10. **Sanh, V., Debut, L., Chaumond, J., Wolf, T. (2019).** Distilbert, a distilled version of Bert: Smaller, faster, cheaper and lighter. CoRR 1910.01108.

11. **Muller, M., Salathe, M., Kummervold, P. E.: (2020).** COVID-Twitter-BERT: A natural language processing model to analyse COVID-19 content on Twitter. arXiv preprint arXiv:2005.07503.

12. **Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V. (2019).** Roberta: A robustly optimized Bert pretraining approach. arXiv preprint arXiv:1907.11692.

13. **Clark, K., Luong, M.T., Le, Q., Manning, C. (2020).** Electra: Pre-training text encoders as discriminators rather than generators. arXiv preprintarXiv: 2003.10555.

14. **Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., Soricut, R., Carvalho, M. (2019).** Albert: A lite Bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.

15. **Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. & Le, Q. (2019).** Xlnet: Generalized autoregressive pretraining for language understanding. In Advances in neural information processing systems, pp. 5753–5763.

16. **Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E. (2016).** Hierarchical attention networks for document classification. Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. pp. 1480–1489.

17. **Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L. (2018).** Deep contextualized word representations. arXiv preprint arXiv:1802.05365.

18. **Hossain, T., Logan, R., Ugarte, A., Matsubara, Y., Young, S., & Singh, S. (2020).** Detecting COVID-19 misinformation on social media. Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020. DOI:10.18653/v1/2020.nlpcovid19-2.1.

19. **Dipta, S., Basak, A., Dutta, S. (2021).** A heuristic-driven ensemble framework for COVID-19 fake news detection. In Combating Online Hostile Posts

in Regional Languages during Emergency Situation pp. 164–176.

20. **Hancock, J., Markowitz, D. (2014).** Linguistic traces of a scientific fraud: The case of Diederik Stapel. PLoS One 9, no. 8.

21. **Gautam, A., Venktesh, V., Masud, S. (2021).** Fake news detection system using XLNet model with topic distributions: CONSTRAINT@AAAI2021 Shared Task, 2101.11425, arXiv, cs.CL.

22. **Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T., Gugger, S., Rush, A. (2020).** Huggingface's transformers: State-of-the-art natural language processing. ArXiv.

23. **Martinc, M., Skrlj, B., Pollak, S. (2018).** Multilingual gender classification with multiview deep learning: Notebook for PAN at CLEF 2018. In: **Cappellato, L., Ferro, N., Nie, J., Soulier, L. editors** Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, Avignon, France, September 10-14, 2018. CEUR Workshop Proceedings, vol. 2125.

24. **Pennington, J., Socher, R., Manning, C.D. (2014).** Glove: Global vectors for word representation. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). pp. 1532–1543.

25. **Joulin, A., Grave, E., Bojanowski, P., Mikolov, T. (2016).** Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759.

26. **Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. (2016).** Inception-v4, inception-Resnet and the impact of residual connections on learning. arXiv preprint arXiv:1602.07261.

27. **Goodfellow, J., Shlens, J., Szegedy, K. (2014).** Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.

28. **Skrlj, B., Martinc, M., Kralj, J., Lavrac, N., Pollak, S. (2020).** tax2vec: Constructing interpretable features from taxonomies for short text classification. Computer Speech & Language p. 101104.

29. **Pulido, C., Ruiz-Eugenio, L., Redondo-Sama, G., Villarejo, B. (2020).** A New application of social impact in social media for overcoming fake news in health. Int. J. Environ. Res. Public Health 2020, 17, 2430.

30. **Oyebode, O., Ndulue, C., Mulchandani, D., Suruliraj, B., Adib, A., Orji, F., Milios, E., Matwin, S., Orji, R. (2020).** COVID-19 pandemic: Identifying key issues using social media and natural language processing. ArXiv, abs/2008.10022.

31. **Gomez-Adorno, H., Posadas-Duran, J.P., Enguix, G.B., Porto, C. (2021).** Overview of fakedes at Iberlef 2021: Fake news detection in Spanish shared task. Procesamiento del Lenguaje Natural 67, 223–231.

32. **Patwa, P., Bhardwaj, M., Guptha, V., Kumari, G., Sharma, S., PYKL, S., Das, A., Ekbal, A., Akhtar, M., Chakraborty, T. (2021).** Overview of constraint 2021 shared tasks: Detecting English covid-19 fake news and Hindi hostile posts. Proceedings of the First Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation (CONSTRAINT), Springer.

33. **Glazkova, A.,Glazkov, M., Trifonov, T. (2020).** g2tmn at Constraint@AAAI2021: Exploiting CT-BERT and ensembling learning for COVID-19 fake news detection. In Combating Online Hostile Posts in Regional Languages during Emergency Situation, pp.116–127.

34. **Python Package Index (PyPI),** emoji 0.6.0, (2021). Available at: https://pypi.org/project/tweet-emoji/.

35. **Raha, T., Indurthi, V., Upadhyaya, A., Kataria, J., Bommakanti, P., Keswani, V., Varma, V. (2021).** Identifying COVID-19 fake news in social media. arXiv.

36. **Li, X., Xia, Y., Long, X., Li, Z., Li, S.(2021).** Exploring text-transformers in AAAI 2021 shared task: Covid-19 fake news detection in English. Proceedings of the First Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation.

37. **Cui L., Lee D. (2020).** CoAID: COVID-19 Healthcare misinformation dataset. arXiv preprint arXiv:2006.00885.

38. **Shahi G. K., Nandini D. (2020).** FakeCovid–A multilingual cross-domain fact check news dataset for COVID-19. arXiv preprint arXiv:2006.11343.

39. **Zenodo (2021).** Available at: https://zenodo.org/record/3862989#.YE9CNF2mO3I.

40. **Alam, F., Dalvi, F., Shaar, S., Durrani, N., Mubarak, H., Nikolov, A., Martino, G., Abdelali, A., Sajjad, H., Darwish, K., Nakov, P. (2020).** Fighting the Covid-19 infodemic in social media: A holistic perspective and a call to arms. arXiv preprint arXiv:2007.07996.

41. **Kaggle (2021).** Available at: https://www.kaggle.com/gpreda/covid19-tweets.

42. **Petrov, A., Proncheva, O. (2018).** Modeling propaganda battle: Decision-making, homophily,

and echo chambers. Proc. AINL-2018, Springer, series CCIS, vol. 930.

43. **Proncheva, O. (2020).** Modeling position selection by individuals during informational warfare with a two-component agenda. J. Mathematical Models and Computer Simulations, vol. 12, No. 2, pp.154–163.

44. **Posadas-Duran, J.-P., H. Gomez-Adorno, Sidorov, G., Escobar, J.J.M. (2019).** Detection of fake news in a new corpus for the Spanish language. Journal of Intelligent & Fuzzy Systems, 36(5):4869–4876.