# Detection of Truthful, Semi-Truthful, False and Other News with Arbitrary Topics Using BERT-Based Models

Elena Shushkevich
Technological University Dublin
Dublin, Ireland
elena.n.shushkevich@gmail.com

John Cardiff
Technological University Dublin
Dublin, Ireland
john.cardiff@tudublin.ie

Anna Boldyreva
FRUCT Association, Finland
Helsinki, Finland
anna.boldyreva@phystech.edu

*Abstract*—**Easy and uncontrolled access to the Internet provokes the wide propagation of false information, which freely circulates in the Internet. Researchers usually solve the problem of fake news detection (FND) in the framework of a known topic and binary classification. In this paper we study possibilities of BERT-based models to detect fake news in news flow with unknown topics and four categories: true, semi-true, false and other. The object of consideration is the dataset CheckThat! Lab proposed for the conference CLEF-2022. The subjects of consideration are the models SBERT, RoBERTa, and mBERT. To improve the quality of classification we use two methods: the addition of a known dataset (LIAR), and the combination of several classes (true + semi-true, false + semi-true). The results outperform the existing achievements, although the state-of-the-art in the FND area is still far from practical applications.**

## I. INTRODUCTION AND PROBLEM SETTINGS

The easy access and up-to-date content of the news portals and social media have made them more and more popular, attracting a huge number of users. Unfortunately, with the increase in traffic, the number of fake news is increasing too. Such a circumstance makes the task of fake news detection one the most important in the development of modern human communications.

A wide range of fake news themes exists, such as politics, Covid-19, and ecology - and fake news for each theme have their own peculiar properties. Nowadays the fake news detection problem is usually posed as *the binary classification of fake news of a specific topic into fake or real news*. The results of such classification can provide very good results, via modern models, especially involving transformers.

However, in real life fake news are usually not related to a single specific topic, for example in a single article we could see fake news about both politics and Covid-19. Moreover, when we speak about real cases of fake news, we are often faced with news which contain both fake and true information. This fact could be explained by the assumption that it is easy to believe in a news which looks more realistic (and contains a piece of truth), and the creators of fake news may be using this trick intentionally or not.

Because of this, we concentrate our efforts on the problem of fake news detection in social media where we have to deal with the nonspecific fake news and try to implement multi-class classification of fake news. *Such conditions - an absence of a*

*specific topic of fake news and a multiclassing of fake news - bring us as close as possible to the situation we have in the real world.*

We have two main goals for our work:

● We will study the possibilities of multi-class fake news detection, where we indicate false, true, partially false, and other classes of news. We will use state-of-the-art transformers for this purpose;
● We will propose novel classifiers based on the classes' combination to improve the quality of fake news detection.

## II. RELATED WORK

The problem of fake news detection is wide and includes a lot of important questions and approaches which could be interesting for researchers. In this Section, we discuss the existing techniques for fake news detection, and the works of other researchers, which show high results in the area of fake news detection. Our work is devoted to multi-class fake news detection, but to understand how to implement multi-class classification in the best way, firstly we need to understand how to work with fake news classification in general. To facilitate comparison and analysis, we have classified the approaches for fake news detection into three main groups: Classical Machine Learning, Neural Networks, and other approaches. In addition, we will observe the works devoted to the multi-class classification of fake news.

### A. Classical Machine Learning Approach for Detection of Fake News

The most commonly used algorithms for fake news detection are classical machine learning algorithms. They demonstrate good results when we deal with binary classification.

The authors [1] analyzed fake news connected with COVID pandemic. They collected a large dataset using materials of 150 users from different social media including Twitter, email, mobile, Whatsapp and Facebook for 4 months from March 2020 to June 2020. Traditional K-Nearest Neighbor provided the quality of results 0.79 F1-score and 0.91 F1-score in March and June respectively.

In [2,3] the researchers aimed to detect fake news connected with Covid-19 on a small dataset of 1000 fake and real news. The researchers compared Logistic Regression, Support Vector

Machine, Gradient Boosting and Random Forest. The winners proved to be Support Vector Machine and Random Forest c with their 69% micro-F1 score. Although this result is worse than the previous one, it may be useful for those who deal with limited datasets.

In [4] the authors compared the same algorithms on the large dataset connected with Covid-19 and taken from Facebook, Twitter and other social media platforms. The best result - 0.93 F1-score - was achieved with the SVM model.

*B. Neural Networks for Detection of Fake News*

Researchers use a lot of interesting linguistic models, the most important of which is BERT [5], which stands for Bidirectional Encoder Representations from Transformers. Such models demonstrated high results in various natural language processing tasks including text classification [6, 7, 8].

DistilBERT model is a modification of BERT having reduced size and providing acceleration 60% [9]. CT-BERT (Covid-Twitter-BERT) is a trained variant of DistilBERT, which showed very good results of fake news detection on Twitter messages [10].

RoBERTa is a robust variant of BERT, which needs larger datasets and longer time for its training [11]. Researchers used the RoBERTa-base variant implementation with cosine similarity computed by averaging over the token vectors to obtain contextual word embeddings. They tuned this 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. This study is described in [12].

Hierarchical Attention Networks (HAN) is based on LSTM (Long Short-Term Memory) architecture. It includes four sequential levels – word encoder, word-level attention, sentence encoder and sentence-level attention [13]. It was successfully used for political fact-checking [14].

The authors of [15] used well-known CNN (Convolutional Neural Network) to detect fake news about Covid-19 in the LIAR dataset. With binary classification they could obtain 0.46 accuracy.

In [16] the authors created an ensemble of linguistic models XLNet, RoBERTa, XLM-RoBERTa, DeBERTa, ERNIE 2.0, and ELECTRA for the task of fake news detection. Their news were connected with several given topics. Furthermore, the authors implemented Heuristic Post-Processing, which takes Soft-voting prediction vectors into account. Thanks to the careful preprocessing step, such an ensemble, and the soft-voting through prediction vectors instead of the hard-voting approach, the authors could achieve an F1-score of 0.98.

To deal with a problem of fake news detection, the authors of [17] implemented a combination of topical distributions from Latent Dirichlet Allocation (LDA) with contextualized representations from XLNet. Their dataset included real and fake information about Covid-19. For the implementation, the authors used the Transformers library maintained by the researchers and engineers at Hugging Face [18], which provides the PyTorch interface for XLNet. Such a complex model allowed researchers to achieve 0.97 F1-score on the test dataset.

The problem of fake news detection was a main focus of the competition named Constraint@AAAI2021 - COVID19 Fake News Detection. The challenge contained the same task for two languages: English and Hindi. The participants needed to create a system for binary classification of fake news and real news. The English dataset contained 10700 messages, 5100 messages were the real one and 5600 news were the false one. Real news were collected from reliable sources. It is WHO (World Health Organization) and CDC (Centers for Disease Control and Prevention). Fake news were collected from social media such as Facebook posts, Twitter tweets, Instagram posts, etc. The dataset contains 37,503 unique words.

166 teams took part in the challenge with the English dataset, and 114 of them overcame the baseline of 93% F1-score. The best results proved to be very close to each other and all were higher than 98% F1-score.

The winner of the challenge g2tmn team [19] achieved 98.69% F1-score using the ensemble of three pretrained CT-BERT models. The second result 98.65% F1-score was achieved by the saradhix team [20]. During the research, the authors used several classical machine learning methods such as Naive Bayes, etc. and also several Transformer models. The third result with 98.60% F1-score was obtained by the xiangyangli team [21]. In this case, the authors also used Text Transformers for their research. Additionally, they used a Pseudo Label Algorithm to do data augmentation

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

●     The most successful models were created as ensembles of Text Transformers,

●     The most important step was to fine-tune such Transformers,

●     The preprocessing steps, which are usually very important for the classical ML models, had no significant role here.

*C. Other features for Detection of Fake News*

Both classical machine learning models and neural networks play an important role in solving the problem of fake news detection, but we may point out many other features and approaches, which are popular in natural language processing, and which could be helpful for fake news detection. We mean here first of all the task of text classification. For example, we can mention algorithms using n-grams of words or characters [22], GloVe [23], an unsupervised learning algorithm for obtaining vector representations for words, Fast-text [24], 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 technique ща regularization that uses noise for the labels [25]; adversarial training [26], a novel regularization method for classifiers to improve model robustness for small, approximately worst case perturbations; and tax2vec [27], a Semantic space vectorization

algorithm.

There are some interesting works describing the impact of fake news about Covid-19, for example [28], which is devoted to researching the social impact of fake news in a context of health information in social media. The authors analyzed materials from three social media platforms: Twitter, Facebook and Reddit. The authors' research questions concentrated on ways that social media messages focused on fake health information, and how interactions based on health evidence with social impacts helped overpower fake health information.

As a result, the authors revealed that misinformation and fake news is higher in Twitter (19%) than in Facebook (4%) or Reddit (7%). They also revealed emotional characteristics messages, namely, messages focused on false health information are mostly aggressive, while messages based on evidence of social impact prove to be more peaceful and respectful.

The authors [29, 30] studied emotional reactions, and their role in fake news detection on the Twitter platform and Politifact dataset (available at https://www.mpi-inf.mpg.de/dl-cred-analysis/). Fake news can reflect completely different topics.

The authors [31] determined the contribution of different topics in dynamics to the publications of well-known media. Here the multi-class definition of topics can be considered as a consequence of reliable and unreliable (real and fake) information in publications.

The approach taken in [32] could help create a dataset for fake news detection, choosing articles for it using key phrases, the advantage of which is that there is no need for a priori information.

*D. Multi-class Classification of Fake News*

As we mentioned previously, in the real world the data could contain not just fake information and true information, but both fake and true information in one news. In this case, we speak about multi-class classification - a state when we have two hypothetical poles – completely fake news and completely real news, and all other news between these two poles, which contain true and fake information in different proportions.

Here we firstly mention the paper [33]. The authors classified political materials from the LIAR dataset in 5 classes: false, barely-true, half-true, mostly-true, true. The training set and test set contains 10200 and 1200 texts respectively. With LSTM the authors obtained accuracy of 0.42.

The authors of [34] presented the fake news classification of 6 classes: pants-fire, false, barely-true, half-true, mostly-true, and true messages. For the experiments, they used SVM, Logistic Regression, Bi-LSTM, CNN, and Hybrid CNNs, and they obtained the highest accuracy of 0.27 using Hybrid CNNs.

Multi-class fake news detection was one of the central topics of the conference CLEF-2022. The competition was organized in the framework of challenge Shared Task 3 - CheckThat! Lab [35]. The aim of the task was to classify news articles in English and German into 4 classes: true, partially true, false, or other, where partially true articles were a mixture of true and false

information, and 'other' class of classification had articles which cannot be categorized as true, false, or partially false due to lack of evidence about its claims. The proposed dataset included several given topics.

The winners of the challenge on English data achieved 0.34 macro-averaged F1-score, where researchers used a BERT-base-uncased model. The authors also conducted experiments with RoBERTa, but the BERT model obtained higher results [36]. The second-best result in the experiments on English dataset was 0.33 macro-averaged F1-score, where researchers created an ensemble of a Funnel Transformer and a Feed Forward Neural Network [37]

For the English-German cross lingual task, the highest result was achieved with 0.29 macro-averaged F1-score, where researchers used BERT-large model [38]. The second-best result was obtained with 0.23 macro-averaged F1-score. Researchers conducted experiments with different transformers and the best one proved to be mDeBERTa model [39].

The problem of multi-class classification, the classification for more than two (fake news and real news) classes is not researched well and, as we mentioned above, even complex and well-trained models show quite low results on the task of multi classification in comparison with binary classification of fake news.

III. DATASETS

In this Section, we describe the datasets we use for the experiments on the multiclass fake news classification.

*A. Basic dataset for CheckThat-2022*

As the main dataset for the experiments, we use the CheckThat-2022 Task 3 dataset [35]. We described this challenge in the previous Section and mentioned the best results that were achieved on the dataset. For the work, we use only the English part of the dataset.

The data for the dataset was collected from 2010 to 2022. The dataset has 4 labels: true, partially true, false, and other. While a 'fake' group of messages contain articles with the untrue main claim, and a 'true' group of messages contains articles with the primary elements of the main claim are demonstrably true, the 'partially true' articles contain information, which can't be accepted as completely true. The 'other' group of messages contains articles that cannot be categorized as neither 'true', 'false', nor 'partially false' due to a lack of evidence about its claims. This category includes articles in dispute and unproven articles. The Table I presents the contents of the dataset.

TABLE I. STATISTICS OF THE CHECKTHAT-2022 DATASET

| **Dataset**: CheckThat-2022 | |
| --- | --- |
| Number of False messages | 893 (48% of all messages) |
| Number of True messages | 421 (22% of all messages) |
| Number of Partially false messages | 414 (22% of all messages) |
| Number of Other messages | 148 (8% of all messages) |
| Total number of messages | 1,876 |
| Average length of a message | 731 words |

The training dataset includes 900 messages, the development

dataset consists of 364 messages, and the test dataset includes 612 messages.

*B. LIAR dataset*

With the idea that the increase in the number of the train messages could improve the results of the classification, we add the LIAR [34] dataset to the training dataset of CheckThat-2022.

The LIAR dataset includes more than 12,000 short messages collected from POLITIFACT.COM between 2007 and 2016, and all messages are labeled for six groups: pants-fire, false, barely-true, half-true, mostly-true, and true. As our goal was to implement a three-class (because of the absence of the 'other' class) classification for fake and non-fake messages, we re-labeled these labels: pants-fire and false messages as fake messages, barely-true, half-true and mostly-true messages as partially false messages, and true messages as real messages. The statistics of the LIAR dataset are presented in Table II.

TABLE II. STATISTICS OF THE LIAR DATASET

| Dataset: LIAR | |
|---|---|
| Number of Pants-fire messages | 1,047 |
| Number of False messages | 2,506 |
| Number of Barely-true messages | 2,102 |
| Number of Half-true messages | 2,627 |
| Number of Mostly-true messages | 2,454 |
| Number of True messages | 2,052 |
| Total number of messages | 12,791 |
| Average length of a message | 18 words |

After relabeling, we have 2,052 true messages, 3,553 false messages, and 7,183 partially false messages. The below Fig.1 illustrates the percentage ratio of different messages in the ChechThat-2022 and the LIAR datasets.
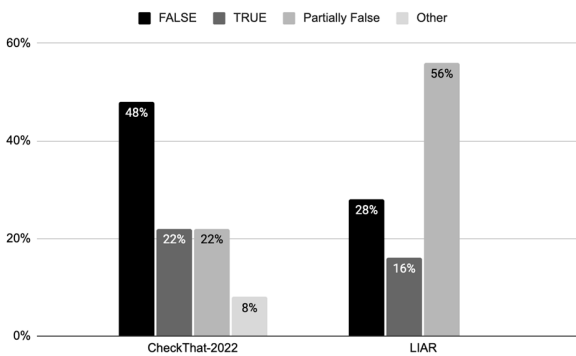


Fig. 1. Percentage ratio of different messages in the CheckThat-2022 and the LIAR datasets

## IV. TESTING BERT-BASED MODELS

*A. BERT-based models*

We conducted the experiments on the CheckThat-2022 dataset using the modern transformers models, including:

● The mBERT [5] model, which supports 104 languages, and is 12-layer, 768-hidden, 12-heads, 110M parameters - the transformer model which was pre-trained on the top 104 languages in Wikipedia, using a masked language

modeling (MLM) objective. As we showed in our previous work, the BERT model provides really high results on fake news detection task in social media, and mBERT could be used not only for English, but also for multilingual experiments. In this work, we apply it for English dataset only, but with an idea to apply it to the multilingual corpora in future. For the experiments, we used the batch size of 8, 512-token input, and 0.5 dropout.

● XLM-RoBERTa [40] is a generic cross lingual sentence encoder that obtains state-of-the-art results on many cross-lingual understanding (XLU) benchmarks. It is trained on filtered CommonCrawl data in 100 languages. This model is also very prospective for our case and could be used for multilingual classification as well. We declared the batch size of 8, 128-token input, and 0.5 dropout.

● SBERT [41] is a modification of the pretrained BERT including siamese structure and triplet network structure to derive semantically meaningful sentence embeddings that can be compared using cosine-similarity. Because of this, SBERT is faster than BERT/RoBERTa, but still provides quality results at the BERT level. SBERT also supports more than 100 languages and could be used for multilingual experiments. We used the batch size of 8, 128-token input, and 0.3 dropout.

For the experiments with all models, we used a learning rate of 1e-7. Using these models, we conducted the experiments on the datasets we described earlier: the English part of the CheckThat-2022 dataset, and on the combination of the English part of the CheckThat-2022 dataset and the LIAR dataset (where we used the LIAR dataset for the training purpose only, and the test dataset included only test CheckThat-2022 dataset).

*B. Experiments with basic datasets*

We conducted experiments with all three transformers: mBERT, SBERT, and XLM-RoBERTa of the CheckThat-2022 dataset, and the results of experiments are presented in Table III.

TABLE III. RESULTS OF EXPERIMENTS ON THE CHECKTHAT-2022 DATASET

| Model | Accuracy | Macro F1-score |
|---|---|---|
| mBERT | 0.40 | 0.22 |
| XLM-RoBERTa | 0.29 | 0.18 |
| SBERT | 0.30 | 0.07 |

As the highest results were obtained using the mBERT model, we rely on this model as the most prospective one, and it is important to clarify the results we have obtained using mBERT in detail. The full results of classification using mBERT are presented in Table IV.

TABLE IV. RESULTS OF EXPERIMENTS WITH MBERT MODEL

| Class | Precision | Recall | Macro F1-score |
|---|---|---|---|
| False | 0.51 | 0.71 | 0.60 |
| True | 0.36 | 0.04 | 0.07 |
| Partially false | 0.10 | 0.21 | 0.13 |
| Other | 0.08 | 0.06 | 0.07 |

As it is shown in Table 4, the results of the classification by classes are very heterogeneous, and the classifier predicts false messages with high results in comparison with predictions of other classes and with the overall results. Our future experiments are aimed to improve these classification results.

*C. Experiments with extended dataset*

As we mentioned previously, the idea of the explanation of the CheckThat-2022 dataset for the results improvement looks promising, and we added the LIAR dataset for the train CheckThat-2022 dataset. The results of experiments on the expanded dataset using mBERT, XLM-RoBERTa, and SBERT are presented in Table V.

TABLE V. RESULTS OF EXPERIMENTS ON THE CHECKTHAT-2022 + LIAR DATASET

| Model | Accuracy | Macro F1-score |
|---|---|---|
| mBERT | 0.40 | 0.25 |
| XLM-RoBERTa | 0.33 | 0.25 |
| SBERT | 0.44 | 0.25 |

Based on the results of the expanded dataset which are higher than the results on the single CheckThat-2022 dataset, we could conclude that our assumption where the increase of the training data will improve the result of classification is correct for this task of multi-class fake news classification.

## V. TESTING MBERT ON COMBINED CLASSES

*A. Combined classes*

As we mentioned above, the state-of-the-art results of multi-class fake news classification on the CheckThat-2022 dataset are low (less than 0.40 macro F1-score), so the most important target is to improve the performance on this dataset. We propose the approach for multi-class classification which is based on the idea of a combination of different classes with the aim to find the intersection of them.

In the first step of our approach, we combine the false and partially false in one joint false class. The idea is to collect all messages with fake news in one class, and to implement the new three-classes classification for joint false, true, and other news.

The same way, we combine true and partially false classes in one joint true class, and implement the three-classes classification for joint true, false, and other news.

In the second step of our approach, we choose all the messages that were labeled as joint false and joint true in the first step of the classification, and find the intersection of these two classes. According to our idea, the messages that fell into the intersection area belong to a partially false class. The graphic illustration of our idea is presented in Fig. 2.
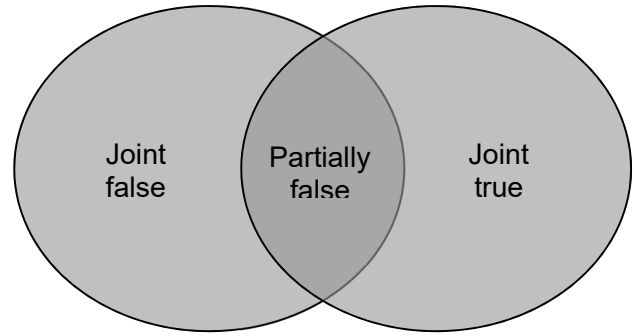


Fig. 2. Search for messages belonging to the partially false class

We believe that this two-steps approach could allow us to improve the quality of partially false messages detection, and, in case of success, could be implemented for the other classes of implementation too.

*B. Experiments with combined classes*

We implemented the experiments in two steps. Firstly, we combined the false and partially false messages in one class (joint false), and implemented the 3-classes classification for the true, joint false, and other classes.

We also combined the true and partially false (partially true) messages of the CheckThat-2022 dataset in one class (joint true), and implemented the 3-classes classification for the false, joint true, and other classes. The results of classification for the joint false and the joint true classes are presented in Table VI.

TABLE VI. RESULTS FOR THE JOINT FALSE AND JOINT TRUE CLASSES OF CLASSIFICATION

| Class | Precision | Recall | Macro F1-score | Number of messages in the class |
|---|---|---|---|---|
| Joint false | 0.61 | 0.97 | 0.75 | 371 |
| Joint true | 0.47 | 0.38 | 0.42 | 266 |

In the second and final step, we merged the same messages from the joint true and the joint false classes with the assumption that the area of intersection of these two classes consists of partially false (partially true) messages. The results of classification for partially false messages on the intersection are presented in Table VII.

TABLE VII. RESULTS FOR THE PARTIALLY FALSE CLASS OF CLASSIFICATION

| Class | Precision | Recall | Macro F1-score | Number of messages in the class |
|---|---|---|---|---|
| Partially false | 0.11 | 1.00 | 0.21 | 33 |

The macro F1-score for partially false class on the intersection of the joint true and the joint false classes is 0.21, while the macro F1-score for partially false class on the full CheckThat-2022 dataset is 0.13, therefore we can conclude that our idea is correct and we can improve the results of partially false messages detection by the detection of intersections of the combined classes.

The graphical illustration of the basic classification results using mBERT model, and the classification results using mBERT model with combined classes are presented in Fig. 3
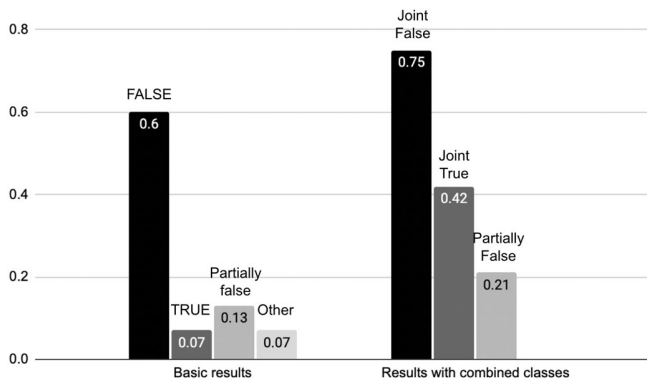


Fig. 3. Basic results of classification using mBERT model and results with the combined classes (Macro F1-score).

## VI. CONCLUSION

### A. Results

In this paper we presented the most common and effective approaches for fake news detection and classification, including not only binary classification for fake and real messages, but also multi-class classification of fake news.

We conducted experiments on the CheckThat-2022 dataset with three transformers: mBERT, SBERT, XLM-RoBERTa, and the best results were obtained with mBERT, namely 0.22 macro F1-score, while the SOTA results on the dataset provides a 0.34 macro F1-score. Despite the fact that our result is lower than SOTA, we should mention that, firstly, all the results on the dataset are low, and, secondly, all the results obtained on the dataset are relatively close to each other. This could be explained by the unbalanced nature of the dataset where the size of the biggest class (fake) is six times bigger than the size of the smallest class (other).

To improve the performance of our models, we expanded the CheckThat-2022 dataset by the LIAR dataset and showed that the increase in the number of training messages improves the results of multi-class classification. The important moment here is that the average length of the message in the LIAR dataset is much shorter that the average length of the message in the CheckThat-2022 dataset (18 words vs 731 words), and it means that short and long messages are interchangeable in the context of classification using transformers.

Also, to improve classification we implemented the new two-step procedure for partially false news detection. This approach includes the relabelling of false, true, and partially true classes at the first step with the idea to isolate the partially true news in the intersection of classes at the second step. This approach allowed us to improve the quality of partially false news detection from a 0.13 macro F1-score to a 0.21 macro F1-score. Although such a value is still far from the practical use, it is 1.5 times better than that we have before and simultaneously it shows one of the ways to improve all the results.

Finally, with combined classes we achieved a relatively good result for the Joint False class with its 0.75 macro F1-score having in view unknown topics in the news under consideration. Such a result may be already interesting for practice.

### B. Future work

In the future, the most interesting and the most challenging task is to improve the results of multi-class classification of fake news. As we showed in this work, the possible way of improvement is to expand the dataset using external sources. As the database is unbalanced, the possible way to increase the performance is to balance the dataset, to collect the data for the classes with the lowest number of messages, such as other, partially false, and true classes. We demonstrated that adding even short messages to the train dataset helps for the performance improvement, and it would be interesting to study the case when the long news (the same length as the messages in the CheckThat-2022 dataset) are added to the dataset.

The second possible way for the performance improvement is to follow the approach we presented in this paper, when we implement the two-step multi-class classification with the aim to isolate the messages of one specific group. By grouping relevant classes it may be possible to implement our approach for the other classes of classification.

The idea of multi-class classification of fake news looks promising as it is close to real life, where we deal with not only purely true and purely false news, but also with a mix of them. The next step could be the splitting of the partially false news class into more specific classes, which could catch peculiarities of fake news and bring us closer to the goal of effective fake news detection.

## REFERENCES

[1] S. Bandyopadhyay, S. Dutta, "Analysis of Fake News In Social Medias for Four Months during Lockdown in COVID-19", 2020, Web: https://doi.org/(10.20944/preprints202006.0243.v1).

[2] E. Shushkevich, J. Cardiff, J., "Detecting fake news about Covid-19 on small datasets with machine learning algorithms. Proceedings of the 30th Conference of Open Innovations Association FRUCT", 2021, pp. 253–258.

[3] E. Shushkevich, M. Alexandrov, J. Cardiff, J., " Detecting fake news about Covid-19 using classifiers from Scikit-learn.", 2021, International Workshop on Inductive Modeling IWIM'2021, 5 pp.

[4] P. Patwa, S. Sharma, S. Pykl, V. Guptha, G. Kumari, M.S. Akhtar, and others ... & T. Chakraborty, "Fighting an infodemic: Covid-19 fake news dataset.", In Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1 (pp. 21-29). Springer International Publishing.

[5] J. Devlin, M. Chang, K. Lee, K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding.", 2018, arXiv preprint arXiv:1810.04805.

[6] L. Akhtyamova, "Named entity recognition in Spanish biomedical literature: Short review and BERT model", 2020, In 2020 26th Conference of Open Innovations Association (FRUCT) (pp. 1-7). IEEE.

[7] A. Adhikari, A. Ram, R. Tang, J. Lin, "Docbert: Bert for document classification", 2019, arXiv preprint arXiv:1904.08398.

[8] S. González-Carvajal, E. C. Garrido-Merchán, "Comparing BERT against traditional machine learning text classification", 2020, arXiv

preprint arXiv:2005.13012.

[9] V. Sanh, L. Debut, J. Chaumond, T. Wolf, "Distilbert, a distilled version of Bert: Smaller, faster, cheaper and lighter", 2019, CoRR 1910.01108.

[10] M. Muller, M. Salathe, P. Kummervold, "COVID-Twitter-BERT: A natural language processing model to analyse COVID-19 content on Twitter", 2020, arXiv preprint arXiv:2005.07503.

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

[12] T. Hossain, R. Logan, A. Ugarte, Y. Matsubara, S. Young, S. Singh, "Detecting COVID-19 misinformation on social media", 2020, Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020. DOI: 18653/v1/2020.nlpcovid19-2.1.

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

[14] H. Rashkin, E. Choi, J. Jang, S. Volkova, Y. Choi, "Truth of varying shades: Analyzing language in fake news and political fact-checking", 2017, In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2931–2937.

[15] S. D. Bhattacharjee, A. Talukder, B.V. Balantrapu, " Active learning based news veracity detection with feature weighting and deep-shallow fusion. In Big Data (Big Data)", 2017 IEEE International Conference on, pp. 556–565. IEEE.

[16] S.D. Das, A. Basak, S. Dutta, S., "A heuristic-driven ensemble framework for COVID-19 fake news detection", 2021, In International Workshop on Combating On line Ho st ile Posts in Regional Languages during Emergency Situation (pp. 164-176). Springer, Cham.

[17] A. Gautam, V. Venktesh, S. Masud, "Fake news detection system using XLNet model with topic distributions: CONSTRAINT@AAAI2021 Shared Task", 2021, 2101.11425, arXiv, cs.CL.

[18] T. Wolf, T. et al, "Transformers: State-of-the-art natural language processing", 2020, In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pp. 38–45.

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

[20] T. Raha, V. Indurthi, A. Upadhyaya, J. Kataria, P. Bommakanti, V. Keswani, V. Varma, V., "Identifying COVID-19 fake news in social media", 2021, arXiv preprint arXiv:2101.11954.

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

[22] M. Martinc, B. Skrlj, S. Pollak, "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.

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

[24] A. Joulin, E. Grave, P. Bojanowski, T. Mikolov, T., " Bag of tricks for efficient text classification", 2016, arXiv preprint arXiv:1607.01759.

[25] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, "Inception-v4, inception-Resnet and the impact of residual connections on learning", 2016, arXiv preprint arXiv:1602.07261.

[26] J. Goodfellow, J. Shlens, K. Szegedy, "Explaining and harnessing adversarial examples", 2014, arXiv preprint arXiv:1412.6572.

[27] B. Skrlj, M. Martinc, J. Kralj, N. Lavrac, S. Pollak, S., "tax2vec: Constructing interpretable features from taxonomies for short text classification", 2020, Computer Speech & Language p. 101104.

[28] C. M. Pulido, L. Ruiz-Eugenio, G. Redondo-Sama, B. Villarejo-Carballido, B., "A new application of social impact in social media for overcoming fake news in health.", 2020, International journal of environmental research and public health, 17(7), 2430.

[29] S. Vosoughi, D. Roy, S. Aral, S., "The spread of true and false news", 2018, online. science, 359(6380), 1146-1151.

[30] A. Giachanou, P. Rosso, and F. Crestani, "Leveraging Emotional Signals for Credibility Detection," in Proc. 42nd Int. ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR), Paris, France, July 21-25, 2019.

[31] A. Petrov and O. Proncheva, "Identifying the Topics of Russian Political Talk Shows," in Proc. Conf. Modeling and Analysis of Complex Systems and Processes (MACSPro), Oct. 22-24, pp. 79-86, 2020. [Online]. Available: http://ceur-ws.org/Vol-2795/short1.pdf

[32] S. Popova and G. Skitalinskaya, "Extended list of stop words: Does it work for keyphrase extraction from short texts?" in Proc. 12th Intern. Sci. Tech. Conf. Comp. Sci. and Inf. Tech. (CSIT), IEEE, 2017, pp. 401-404.

[33] Y. Long, Q., Lu, R. Xiang, M. Li, C.-R. Huang, "Fake news detection through multi-perspective speaker profiles", 2017, In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers), volume 2, pp. 252–256.

[34] W. Y. Wang, "'Liar, liar pants on fire': A new benchmark dataset for fake news detection," ArXiv, 2017, arXiv:1705.00648.

[35] J. Köhler et al., "Overview of the CLEF-2022 CheckThat! lab task 3 on fake news detection," Working Notes of CLEF, 2022.

[36] B. Taboubi, M. A. B. Nessir, and H. Haddad, "iCompass at CheckThat! 2022: Combining deep language models for fake news detection," Working Notes of CLEF, 2022.

[37] J. R. Martinez-Rico, J. Martinez-Romo, and L. Araujo, "NLP &IRUNED at CheckThat! 2022: Ensemble of classifiers for fake news detection," Working Notes of CLEF, 2022.

[38] H. N. Tran and U. Kruschwitz, "ur-iw-hnt at CheckThat! 2022: Cross-lingual text summarization for fake news detection," Working Notes of CLEF, 2022.

[39] J. Köhler et al., "Overview of the CLEF-2022 CheckThat! lab task 3 on fake news detection," Working Notes of CLEF, 2022.

[40] A. Conneau et al., "Unsupervised cross-lingual representation learning at scale," ArXiv, 2019, arXiv:1911.02116.

[41] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using siamese BERT-networks," ArXiv, 2019, arXiv:1908.10084.