

# MUCIC at CheckThat! 2021: FaDo-Fake News Detection and Domain Identification using Transformers Ensembling

Fazlourrahman Balouchzahi<sup>1</sup>, Hosahalli Lakshmaiah Shashirekha<sup>2</sup> and Grigori Sidorov<sup>1</sup>

<sup>1</sup>Center for Computing Research, Instituto Politécnico Nacional, CDMX, Mexico

<sup>2</sup>Department of Computer Science, Mangalore University, Mangalore, India

## Abstract

Since the beginning of Covid-19 era in November 2019, the patient growth curve is closely accompanied by the growth of fake news. Therefore, developing tools and models for the detection of fake news from real ones in various domains have become more significant than the earlier days. To address the detection of fake news, in this paper, we, team MUCIC, describe the models submitted to 'Fake News Detection', a shared task organized by CLEF-2021-CheckThat! Lab. This shared task contains two sub-tasks namely; Fake News Detection of News Articles (Subtask 3A) and Topical Domain Classification of News Articles (Subtask 3B) and both are multi-class text classification tasks. The proposed models have been developed by fine-tuning the three transformer-based language models namely; Roberta, Distilbert, and BERT from HuggingFace using training data and then ensembling them as estimators with majority voting. The proposed models performances evaluated through the evaluation script provided by organizers obtained F1-scores of 0.5309 and 0.8550 for Subtask 3A and Subtask 3B respectively.

## Keywords

Fake News Detection, Domain Identification, Transformers, BERT, Roberta, Distilbert

## 1. Introduction

Anonymity is a significant attribute of the cyber world [1] which also provides ample opportunities to mislead and manipulate peoples' thoughts and damnify social trust [2]. Ease of access, low cost, and swift broadcasting of the information in social media and network is exerting negative influence on sharing fake news in various domains [3]. Fake news detection in different domains has received much attention especially after the outbreak of Covid-19 and its effects on the entire world.

The nature of texts in social media is highly unstructured and noisy, and texts may belong to various domains as people comment or share messages about various topics. Therefore, identifying the domain of texts in which news are disseminating in social media is very important

---

CLEF 2021 – Conference and Labs of the Evaluation Forum, September 21–24, 2021, Bucharest, Romania

✉ frs1\_b@yahoo.com (F. Balouchzahi); hlsrekha@gmail.com (H. L. Shashirekha); sidorov@cic.ipn.mx (G. Sidorov)

🌐 <https://mangaloreuniversity.ac.in/dr-h-l-shashirekha> (H. L. Shashirekha); <http://www.cic.ipn.mx/~sidorov/> (G. Sidorov)

🆔 0000-0003-1937-3475 (F. Balouchzahi)

© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

and might help in designing the models/tools to detect fake news. Fake news detection and domain identification tasks could be modeled as binary Text Classification (TC) problems if there are only two classes, otherwise they will be modeled as multi-class TC problem.

Hitherto many models based on Machine Learning (ML) and Deep Learning (DL) are experimented by researchers for TC in general. Of late Transfer Learning (TL) and transformers-based models are also getting attention due to their efficient performances in various Natural Language Processing (NLP) and TC tasks [4, 5].

Many DL and Neural Network (NN) based models such as CNN, LSTM, BiLSTM, etc. are considered as best models for many NLP and TC tasks. But, the introduction of transformers have changed the game and since 2017 transformer<sup>1</sup>-based models are beating DL models in many NLP related applications<sup>2</sup>. Transformers are novel architectures with self-attention mechanism that are primarily used for NLP. Transformer-based models are able to handle long-range dependencies and usually are used for sequence-to-sequence NLP tasks<sup>3</sup>.

NLP researchers are challenged by Conference and Labs of the Evaluation Forum (CLEF) 2021<sup>4</sup> CheckThat! Lab on Detecting Check-Worthy Claims, Previously Fact-Checked Claims, and Fake News<sup>5</sup> [6]. Fake News Detection (Task 3)<sup>6</sup> is one of the three tasks evaluated by the CheckThat! Lab and this task has two subtasks. Figure 1 illustrates the graphical representation of subtasks and details of the subtasks are as follows:

- **Subtask 3A - Multi-Class Fake News Detection of News Articles (English):** is a multi-class TC task that accepts a given English text and classifies it into one of four pre-defined categories namely, 'False', 'Partially False', 'True', and 'Other'. Table 1 gives the description of labels in Subtask 3A;
- **Subtask 3B - Topical Domain Classification of News Articles:** is also a multi-class TC task that further classifies a given fake news into one of six categories representing six different domains namely, 'Health', 'Climate', 'Economy', 'Crime', 'Elections' and 'Education'. It is worth to note that classification results of Subtask 3A are incommunicable to Subtask 3B and only known fake news provided by the organizers are used for Subtask 3B.

Many tools and algorithms have been introduced by researchers for TC in general. However, an algorithm which performs well for a particular dataset may not give the same performance for other datasets [3]. Therefore, it is not logical to claim that an algorithm or a model leads to the same performance for all the datasets. As a result, inspired by H. L. Shashirekha et al. [3] and Gundapu S et al. [7], we, team MUCIC, utilized available at HuggingFace<sup>7</sup> to develop transformer-based models, namely, BERT, Distilbert, and Roberta, extended each of these models with a step of fine-tuning the respective Language Model (LM) and finally ensembled the extended models with majority voting for fake news detection and domain identification.

---

<sup>1</sup><https://pypi.org/project/transformers/>

<sup>2</sup>[https://yale-lily.github.io/public/matt\\_f2018.pdf](https://yale-lily.github.io/public/matt_f2018.pdf)

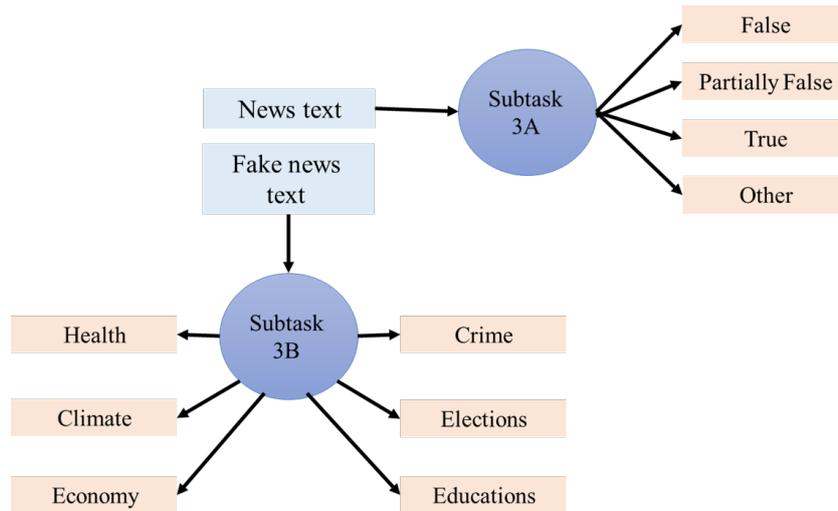
<sup>3</sup><https://towardsdatascience.com/transformers-89034557de14>

<sup>4</sup><http://clef2021.clef-initiative.eu/>

<sup>5</sup><https://sites.google.com/view/clef2021-checkthat/home>

<sup>6</sup><https://sites.google.com/view/clef2021-checkthat/tasks/task-3-fake-news-detection>

<sup>7</sup><https://huggingface.co/>



**Figure 1:** Graphical representation of Fake News Detection subtasks

**Table 1**

Classes in Subtask 3A

| Category        | Description   |
|-----------------|---|
| False           | The main content of given text is fake.   |
| Partially false | Main claim in given text might be True but also contain false or misleading information, not surely true and not certainly false. |
| True            | The given text includes contents that are clearly apparent or capable of being logically proved.                                  |
| Other           | Texts with no enough evidence to categorize as of one of earlier categories.  |

The rest of paper is organized as follows: Section 2 highlights the recent literature and related works followed by the proposed Ensemble of transformer-based model for fake news detection and domain identification in Section 3. While Section 4 describes the experiments and results, Section 5 concludes the paper with future scope.

## 2. Related Work

Alongside the outbreak of Covid-19, fake news detection task has gained more and more attention due to critical situation the entire globe is facing. Many researchers have developed many models to combat fake news and avoid its spread thereby. Few of the latest and related works are briefly discussed in this section. Gundapu et al. [7] have proposed several models based on ML, DL and TL approaches for the task of fake news detection in English language in which a given news item is categorized as either ‘fake’ or ‘real’. They conducted experiments on the dataset provided by ConstraintAI’21<sup>8</sup> shared task organizers and their ensemble of transformer-based models outperformed the rest of the models with an F1-score of 0.9855. The

<sup>8</sup><https://constraint-shared-task-2021.github.io/>

dataset as described in detail by Patwa et al. [8] consists of 10,700 texts from various social media platforms such as Instagram, Twitter, etc. Their proposed transformer-based ensemble model architecture consists of preprocessing and ensemble model construction. Data is preprocessed by converting emoji and hashtags to text, removing punctuation, digits, non-ASCII characters, stop words and stemming. Three transformer-based models namely, BERT, ALBERT, and XLNet are ensembled in such way that the average of all softmax values from estimators gives probability of the final predicted label.

A multilingual cross-domain dataset containing 5,182 fact-checked news articles related to Covid-19 is developed by Shahi et al. [9] by collecting articles from 92 fact checking websites shared during the first five months of 2020. The collected articles include texts in 40 languages, categorized into 11 classes and then filtered into two categories namely, 'False' and 'Other'. 4,132 texts which belong to 'False' category contain false information and the remaining belongs to 'Other' category. Authors used Bert-based model as a baseline and obtained an average F1-score of 0.76 on identifying fake news for the developed dataset.

Another work towards fake news detection in Covid-19 domain carried out by Paka et. al [10] includes developing a COVID-19 (Twitter) Fake news (CTF) dataset. CTF is a large-scale text dataset in Covid-19 domain collected from Twitter containing 21.85M unlabelled tweets along with 45.26K labeled tweets; out of which 18.55K are labeled as genuine and remaining 26.71K as fake. Data collection and annotation procedure has been done in four stages namely, Segregating COVID-19 related tweets, Collecting COVID-19 supporting statements, Filtering genuine and fake tweets, and Human annotation. The authenticity of the CTF dataset is guaranteed by fact checking websites such as PolitiFact, Snopes, TruthOrFiction, etc. and certain health organizations. The authors proposed a semi-supervised DL model based on Neural Attention model called Cross-SEAN which leverages huge unlabelled data to improve its performance and obtained F1-score of 0.95 on CTF dataset.

Random Forest (RF), k-Nearest Neighbor (kNN), Logistic Regression (LR) and Support Vector Machine (SVM), with TF and TF-IDF as features and also DL models, namely, Convolutional Neural Network (CNN), Long Short Term Memory (LSTM), and Gated Recurrent Network (GRU), utilizing Glove word embedding have been surveyed by Jiang et al. [11] for fake news detection. Further, the authors proposed a stacking approach using the above mentioned classifiers and evaluated that model on two fake news datasets namely, ISOT<sup>9</sup> and KDnugget [12] and obtained accuracies of 99.94% and 96.05% respectively. However, in terms of individual performance of each estimator, LR classifier with an accuracy of 92.82% and RF with an accuracy of 99.87% both fed with TF-IDF as features outperformed other individual classifiers for both the datasets.

Despite the lack of availability of annotated dataset for fake news detection in low resource languages, Forum for Information Retrieval Evaluation (FIRE) 2020<sup>10</sup> called for UrduFake<sup>11</sup>, a shared task on fake news detection in Urdu language and provided a dataset consisting of 500 real and 400 fake news articles in Urdu [13]. Balouchzahi et al. [14] proposed various models based on ML, DL, TL, and hybrid approaches for UrduFake. ML model is a Voting Classifier (VC) with three estimators namely, Multinomial Naïve Bayes (MNB), Multilayer Perceptron (MLP),

---

<sup>9</sup><https://www.uvic.ca/engineering/ece/isot/datasets/fake-news/index.php>

<sup>10</sup><http://fire.irsir.res.in/fire/2020/home>

<sup>11</sup><https://www.urdufake2020.cicling.org/home>

and LR, had been fed with set of char and word ngrams features. While DL model has been developed based on a multi-channel CNN and Skipgram word embedding, an implementation of Universal Language Model Fine-Tuning (ULMFiT) is based on TL approach. Further, all models are ensembled as an hybrid model based on majority voting. ML VC outperformed the rest of the models with an average F1-score of 0.7894.

Tools and modules for the analysis of fake news on social media are not only bounded to detect the category of a given text but also to identify the spreaders of such false information to find out the intention beyond sharing fake news. In this direction, PAN at Conference and Labs of the Evaluation Forum (CLEF) 2020 had called for shared task to identify the spreader of fake news in Spanish and English. Dataset provided by task organizers consists of 100 tweets per user (news spreaders) for 300 users. To tackle this task Shashirekha et al. [2, 3] proposed two classifiers namely, ULMFiT based on TL approach and ensemble of ML classifiers as a voting classifier. The ULMFiT model initially uses unlabelled data collection from Wikipedia to train the universal LMs for both English and Spanish languages. As texts from Wikipedia are from general domain, there is every chance of missing valuable words and features related to fake news. Hence, the LMs obtained by training on Wikipedia are fine-tuned using the training data provided by PAN to make LMs more specific for the task of fake news spreader detection. The Fast.ai<sup>12</sup> library is used to develop LMs and classifiers. The proposed ML VC is built with three estimators, namely, two linear SVM and LR classifiers. After preprocessing the texts by removing stopwords and punctuation, converting emoji to text and performing lemmatization, features are extracted. Unigram TFIDF, N-gram TF combined with Doc2vec features are scaled using MaxAbsScaler and used to train the VC. Chi-square test, Mutual Information, and F-test are used to select the important features. The final results reported by PAN illustrate that ULMFiT and ML VC obtained average accuracies of 0.63 and 0.70 respectively.

### 3. Methodology

The performance and effectiveness of ensembling models based on majority voting of ML classifiers have already been proved for many tasks. Inspired by Gundapu S et al. [7], ensembling fine tuned transformer-based models are experimented in this work for fake news detection and domain identification. Transformer-based models generally include two steps of pre-training LM and then fine-tuning the model for the new task. Usually as pre-trained models are publicly available at HuggingFace it is only required to fine-tune LM with a labeled dataset for the target task. The structure of the proposed model consists of the following four steps:

1. Pre-training: pre-trained models available at HuggingFace are used;
2. Fine-tuning LMs with texts from training set to make LM more domain specific for the intended task;
3. Training the models obtained in step 2 for classification (each model separately);
4. Ensembling for prediction of labels based on majority voting.

As domain of texts used for pre-training a LM might be different from the training set, initially the individual LMs for transformer-based models namely, BERT [15], Distilbert [16], and

---

<sup>12</sup><https://www.fast.ai/>

**Table 2**  
Configuration of transformer-based models

| Model      | Type                    | Max Length | Batch size | Learning rate | epochs |
|------------|-------------------------|------------|------------|---------------|--------|
| Roberta    | roberta-base            |            |            |               |        |
| Distilbert | distilbert-base-uncased | 512        | 16         | 4e-5          | 5      |
| BERT       | bert-base-uncased       |            |            |               |        |

Roberta [17] are fine-tuned using the training data and then these fine-tuned models are used as estimators in a VC. The main objectives of each estimator are given below:

- **BERT** is a pre-trained model that employs Masked Language Modeling (MLM) in a self-supervised fashion. In other words, only raw texts are used to pre-train the model without manual annotation. MLM concept along with Next Sentence Prediction (NSP) enables the model to learn deep representation of a language to extract more efficient features from texts for downstream tasks.
- **Distilbert** is a lightweight BERT model. It follows the objectives of BERT model with distillation loss and returns the same probabilities as BERT and it also utilizes Cosine embedding concept to generate hidden states as close as BERT model.
- **Roberta** is an optimized BERT model re-trained with improved training methodology, more data and hardware resources<sup>13</sup>. Roberta without NSP concept is similar to BERT and employs dynamic masking results in changing the masked token during the training epochs.

Fast-bert<sup>14</sup> library with configuration given in Table 2 is used to fine-tune the LMs in (step 2). Except for model and type, the same configuration is used for all the three LMs and fine-tuned only for 5 epochs due to resource (RAM and GPU) constraint. Fast-bert associated with HuggingFace enables a very simple manner to train and evaluate the transformer-based models. Each individual transformer-based model is trained for 20 epochs.

As per the general idea of the proposed model illustrated in Figure 2, preprocessing the texts has been skipped as the models performed better without preprocessing. The same architecture is followed to construct models for both the subtasks.

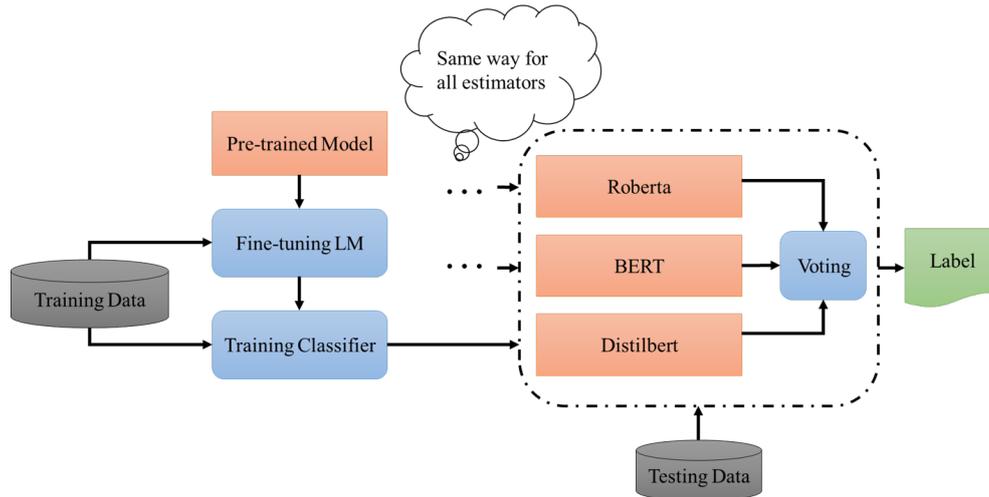
## 4. Experimental Results

For any supervised ML task such as TC, annotated data sets are essential to train the model and enable machine to quickly and clearly understand the input patterns [18]. Therefore, datasets provided by the CheckThat! Lab are used and the data collections steps are detailed by Shahi et al. [19].

Due to mistakes while generating prediction files for the task, the results of the proposed models reported in the leaderboard are much less than the actual performances of the systems.

<sup>13</sup><https://towardsdatascience.com/bert-roberta-distilbert-xlnet-which-one-to-use-3d5ab82ba5f8>

<sup>14</sup><https://github.com/utterworks/fast-bert>



**Figure 2:** Proposed Ensemble of Transformer-based models

Therefore, along with the results reported by task organizers, the non-official results, i.e., the actual performances of system evaluated through evaluation script<sup>15</sup> provided by the organizer on re-generated prediction files are also included in this paper. The re-generated and re-evaluated submissions for both the subtasks can be found in our GitHub page<sup>16</sup>.

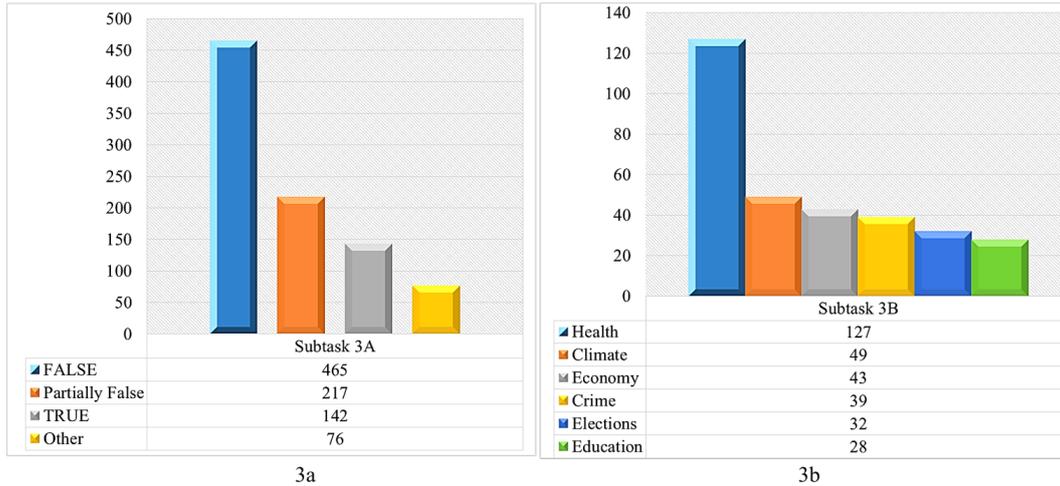
The training set for Subtask 3A consists of 900 texts in four categories namely, ‘False’, ‘Partially False’, ‘True’, and ‘Other’. Description of labels is given in Table 1 and label distributions over training sets are given in Figure 3a. It can be observed that the dataset is highly imbalanced and as expected this has affected the performance of the proposed model in a negative way. As per the results reported in leaderboard, for the test set consisting of 364 texts, the proposed model obtained an F1-score of 0.2334 which is far lower than the expectation due to earlier mentioned reason of mistakes in prediction files. However, non-officially the model obtained 0.5034 F1-score on re-generated prediction files.

A subset of fake news from the dataset of Subtask 3A is used for Subtask 3B. The dataset for Subtask 3B which includes 318 texts distributed into 6 categories is also imbalanced and the distribution of labels is shown in Figure 3b. Based on results reported in leaderboard, for the test set consisting of 137 texts, the proposed model obtained an F1-score of 0.1450 which is far less than the expectation but actual performance obtained non-officially is 0.8550 F1-score on re-generated prediction files.

Comparison of results both official (in leaderboard) and non-official, between the proposed models and top models in the shared task are given in Table 3. It is very much clear that there is a huge gap between the results reported officially and the results obtained non-officially. Considering actual performances of the proposed model once again the effectiveness of ensembling classifiers to utilize the strength of single models has been proved.

<sup>15</sup>[https://gitlab.com/checkthat\\_lab/clef2021-checkthat-lab/-/tree/master/task3/evaluation](https://gitlab.com/checkthat_lab/clef2021-checkthat-lab/-/tree/master/task3/evaluation)

<sup>16</sup>[https://github.com/fazlfrs/CheckThat-\\_Task3\\_Submissions](https://github.com/fazlfrs/CheckThat-_Task3_Submissions)



**Figure 3:** Label distribution over training sets for subtasks

**Table 3**

Comparison of performances of the top models and proposed models in the shared task

| Team/ participant name                | F1-score   |            |
|---------------------------------------|------------|------------|
|                                       | Subtask 3A | Subtask 3B |
| sushmakumari                          | 0.8376     | 0.8552     |
| <b>MUCIC(non-official)</b>            | 0.5039     | 0.8550     |
| kannanrrk                             | 0.5034     | 0.8178     |
| jmartinez595                          | 0.4680     | –          |
| <b>MUCIC(official in leaderboard)</b> | 0.2334     | 0.1450     |

## 5. Conclusion and Future Work

In this paper, we, team, MUCIC, have presented the description of the proposed Ensemble of transformer-based VC models for Fake News Detection, a shared task (Task 3) in CLEF-2021-CheckThat! Lab. Three transformer-based models namely, Roberta, Distilbert, and BERT are double fine-tuned (once on respective LM and then down streamed for respective TC task) and ensembled as VC that predicts the label of a given text by majority voting. The proposed models achieved low F1-scores of 0.233 and 0.145 for Subtask 3A: Multi-Class Fake News Detection of News Articles and Subtask 3B: Topical Domain Classification of News Articles respectively, against our expectations due to our mistakes in submission files. However, the actual performances of the systems show very competitive F1-scores of 0.5309 and 0.8550 for Subtask 3A and Subtask 3B respectively, on re-generated prediction files. Improving the performances of the proposed models by addressing the problems followed by exploring ML and DL approaches with various feature sets will be the future work.

## 6. Acknowledgment

Team MUCIC sincerely appreciates the efforts, guidance and support of the shared task organizers and reviewers for the valuable comments and suggestions.

## References

- [1] F. Balouchzahi, H. L. Shashirekha, Las for HASOC - learning approaches for hate speech and offensive content identification, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2020 - Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020, volume 2826 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 145–151. URL: <http://ceur-ws.org/Vol-2826/T2-6.pdf>.
- [2] H. L. Shashirekha, F. Balouchzahi, Ulmfit for twitter fake news spreader profiling, in: L. Cappellato, C. Eickhoff, N. Ferro, A. Névéol (Eds.), Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, September 22-25, 2020, volume 2696 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020. URL: [http://ceur-ws.org/Vol-2696/paper\\_126.pdf](http://ceur-ws.org/Vol-2696/paper_126.pdf).
- [3] H. L. Shashirekha, M. D. Anusha, N. S. Prakash, Ensemble model for profiling fake news spreaders on twitter, in: L. Cappellato, C. Eickhoff, N. Ferro, A. Névéol (Eds.), Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, September 22-25, 2020, volume 2696 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020. URL: [http://ceur-ws.org/Vol-2696/paper\\_136.pdf](http://ceur-ws.org/Vol-2696/paper_136.pdf).
- [4] S. Gao, M. Alawad, M. T. Young, J. Gounley, N. Schaefferkoetter, H.-J. Yoon, X.-C. Wu, E. B. Durbin, J. Doherty, A. Stroup, et al., Limitations of transformers on clinical text classification, *IEEE journal of biomedical and health informatics* (2021).
- [5] S. Prabhu, M. Mohamed, H. Misra, Multi-class text classification using bert-based active learning, *arXiv preprint arXiv:2104.14289* (2021).
- [6] P. Nakov, G. D. S. Martino, T. Elsayed, A. Barrón-Cedeño, R. Míguez, S. Shaar, F. Alam, F. Haouari, M. Hasanain, N. Babulkov, A. Nikolov, G. K. Shahi, J. M. Struß, T. Mandl, The CLEF-2021 checkthat! lab on detecting check-worthy claims, previously fact-checked claims, and fake news, in: D. Hiemstra, M. Moens, J. Mothe, R. Perego, M. Potthast, F. Sebastiani (Eds.), *Advances in Information Retrieval - 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28 - April 1, 2021, Proceedings, Part II*, volume 12657 of *Lecture Notes in Computer Science*, Springer, 2021, pp. 639–649. URL: [https://doi.org/10.1007/978-3-030-72240-1\\_75](https://doi.org/10.1007/978-3-030-72240-1_75). doi:10.1007/978-3-030-72240-1\_75.
- [7] S. Gundapu, R. Mamid, Transformer based automatic covid-19 fake news detection system, *arXiv preprint arXiv:2101.00180* (2021).
- [8] P. Patwa, S. Sharma, S. PYKL, V. Guptha, G. Kumari, M. S. Akhtar, A. Ekbal, A. Das, T. Chakraborty, Fighting an infodemic: Covid-19 fake news dataset, *arXiv preprint arXiv:2011.03327* (2020).
- [9] G. K. Shahi, D. Nandini, FakeCovid – a multilingual cross-domain fact check news dataset for covid-19, in: *Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media*, 2020. URL: [http://workshop-proceedings.icwsm.org/pdf/2020\\_14.pdf](http://workshop-proceedings.icwsm.org/pdf/2020_14.pdf).

- [10] W. S. Paka, R. Bansal, A. Kaushik, S. Sengupta, T. Chakraborty, Cross-sean: A cross-stitch semi-supervised neural attention model for covid-19 fake news detection, *Applied Soft Computing* 107 (2021) 107393.
- [11] T. Jiang, J. P. Li, A. U. Haq, A. Saboor, A. Ali, A novel stacking approach for accurate detection of fake news, *IEEE Access* 9 (2021) 22626–22639.
- [12] P. H. A. Faustini, T. F. Covões, Fake news detection in multiple platforms and languages, *Expert Systems with Applications* 158 (2020) 113503.
- [13] M. Amjad, G. Sidorov, A. Zhila, Data augmentation using machine translation for fake news detection in the urdu language, in: *Proceedings of The 12th Language Resources and Evaluation Conference*, 2020, pp. 2537–2542.
- [14] F. Balouchzahi, H. L. Shashirekha, Learning models for urdu fake news detection, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), *Working Notes of FIRE 2020 - Forum for Information Retrieval Evaluation*, Hyderabad, India, December 16-20, 2020, volume 2826 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 474–479. URL: <http://ceur-ws.org/Vol-2826/T3-7.pdf>.
- [15] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805* (2018).
- [16] V. Sanh, L. Debut, J. Chaumond, T. Wolf, Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter, *arXiv preprint arXiv:1910.01108* (2019).
- [17] 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, *arXiv preprint arXiv:1907.11692* (2019).
- [18] G. K. Shahi, Amused: An annotation framework of multi-modal social media data, *arXiv preprint arXiv:2010.00502* (2020).
- [19] G. K. Shahi, J. M. Struß, T. Mandl, Overview of the CLEF-2021 CheckThat! lab task 3 on fake news detection, in: *Working Notes of CLEF 2021—Conference and Labs of the Evaluation Forum*, CLEF '2021, Bucharest, Romania (online), 2021.