

CIC at CheckThat! 2022: Multi-class and Cross-lingual Fake News Detection

Muhammad Arif¹, Atnafu Lambebo Tonja¹, Iqra Ameer¹, Olga Kolesnikova¹, Alexander Gelbukh¹, Grigori Sidorov¹ and Abdul Gafar Manuel Meque¹

¹Instituto Politécnico Nacional (IPN), Centro de Investigación en Computación (CIC), Av. Juan de Dios Batiz, s/n, 07320, Mexico City, Mexico

Abstract

Nowadays, social media is one widely used platform to access information. Fake news on social media and various other media is widely spreading. It is a matter of serious concern due to its ability to cause a lot of social and national damage with destructive impacts. Therefore, detecting misleading news is critical to detect automatically. Fake news detection software has been used in a variety of fields, such as social media, health, political news, etc. This paper presents the Instituto Politécnico Nacional (Mexico) at CheckThat! 2022. In this paper, we discuss using different algorithms for the multiclass and cross-lingual fake news detection task. We achieved a macro F1-score of 28.60% for a mono-lingual task in English (task 3a) using RoBERTa pre-trained model and 17.21% for a cross-lingual task for English and German (task 3b) using Bi-LSTM deep learning algorithm.

Keywords

Fake news detection, Cross-lingual classification, Multi-class detection, Fake news detection for low resource languages, Transfer learning

1. Introduction

Fake news refers to falsified news or propaganda that is disseminated through traditional media platforms such as print and television, as well as non-traditional media platforms such as social media [1]. The primary objective of disseminating such information is to deceive readers, harm a company's reputation, or profit from sensationalism. It is widely regarded as one of the most serious threats to democracy, free speech, and social order [2]. Fake news is rapidly being disseminated through social media platforms such as Twitter and Facebook, according to [3]. These platforms provide a venue for the general public to express their thoughts and opinions in an unfiltered and uncensored manner. Compared to conventional method views from media publishers' platforms, some news pieces hosted or shared on social media sites receive more views. According to researchers who researched the speed with which fake news spreads on Twitter, tweets containing misleading information reach individuals six times faster than factual

CLEF 2022: Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy

✉ mariff2021@cic.ipn.mx (M. Arif); atnafu.lambebo@wsu.edu.et (A. L. Tonja); iqra@nlp.cic.ipn.com (I. Ameer); kolesolga@gmail.com (O. Kolesnikova); gelbukh@cic.ipn.mx (A. Gelbukh); sidorov@cic.ipn.mx (G. Sidorov); gafar1_meque@cic.ipn.mx (A. G. M. Meque)

🌐 <https://atnafuatx.github.io/> (A. L. Tonja)

🆔 0000-0001-6141-0204 (M. Arif); 0000-0002-3501-5136 (A. L. Tonja); 0000-0002-1134-9713 (I. Ameer)



© 2022 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)

tweets [2]. The negative consequences of false news appear to be unavoidable, ranging from making people believe Hillary Clinton had an alien baby to affecting the 2016 US presidential elections [3]. A few facts about fake news in the United States are as follows:

- 62% of Americans get their news from social media [4]
- Fake news had a higher Facebook share than legitimate news [5]

Another interesting, however, a sad example is a false statement claiming that "alcohol is a cure for COVID-19" caused multiple deaths and hospitalizations in Iran, according to [6]. This shows how helpless we are against fake news in some critical situations and how severe the consequences can be if we ignore them. The first step in dealing with fake news is recognizing and distinguishing it from real news.

Detecting fake information on social media poses numerous new and challenging research problems. Although fake news itself isn't a new problem—nations or groups have been using the news media to execute propaganda or affect operations for centuries—the rise of web-generated news on social media makes fake news a more powerful force that challenges traditional journalistic norms. There are numerous traits of this hassle that make it uniquely challenging for automatic detection [7]. In this paper, we propose a methodology to train the model that detects whether an article is authentic or fake based on its words, phrases, sources, and titles. It will apply supervised machine learning algorithms on an annotated (labeled) dataset and for automatic fake news detection.

We used three models: Passive Aggressive Classifier (PAC), a machine learning algorithm; Bi-LSTM, a deep learning algorithm; and RoBERTa a pre-trained language model from the BERT family for fine-tuning. Then, according to confusion matrix results, feature selection methods are applied to experiment and choose the best-fit features to obtain the highest precision. We propose to create the model using different classification algorithms. The product model will test the unseen data, the results will be plotted, and accordingly, the product will be a model that detects and classifies fake articles and can be used and integrated with any system for future use.[8].

This paper discusses multi-class and cross-lingual fake news detection methods for the shared task at CheckThat!2022. The paper is organized as follows: section 2 describes the past work related to this study, section 3 gives an overview of the dataset statistics, section 4 explains the methodology adopted in this study including used algorithms, section 5 emphasizes on the experimental results and description. Finally, section 6 concludes the paper and sheds some light on possible future work.

2. Related work

Faking a piece of news has been part of all eras of technology in the form of yellow journalism. However, since the advent of social media, the impact of the harm has grown many folds. It has hence been one of the most challenging problems for researchers to solve for the last decade as it is very difficult to distinguish fake text from real text [9, 10, 11]. Fake news detection approaches, by and large, fall into two classes relying upon whether they use (1) news content or (2) social settings [7]. Theoretical fake news studies have examined the classification of fake

news in the form of misinformation, disinformation, hysteria, falsehood, propaganda, clickbait, and conspiracy theories. The last decade has evidenced considerable advances in research on fake news that has had a real-life impact.

With the emergence of larger datasets, the research on fake news detection got into force when numerous studies arose in 2017. Among the first, Wang [12] introduced LIAR - a relatively large dataset of 12,000 truth-checked and multi-class labeled short news statements from the political area. To this novel data, he applied various algorithms and deep learning architectures such as Support Vector Machines(SVM), Logistic Regression, Bi-LSTM, and Convolutional Neural Network (CNN).

The domain of automated fake news detection and fake news classification across languages was not studied in depth in previous research. Though [13] trained their models with news texts from one language to classify them in another, their approach relied on prior machine translation as a pre-processing step. The lack of cross-language classification studies in fake news detection was due to the absence of appropriate data. Yet, recently[14] presented a multilingual dataset that offers a possibility to study transfer learning in fake news classification. Utilizing the embedding of XLM-R without fine-tuning the model, the authors reported accuracy of 82% for transfer classification from Italian to French. As their overall results show room for improvement, this study aims to extend the proposed model further to gain valuable insights from transfer learning.

3. Dataset

In the experimentation phase, we used the dataset for task 3 from shared task organizers [15, 16]. The dataset contains train, validation, and test sets for English and test data for the German language. The training data includes about 1,300 articles in English with the respective labels such as true, partially false, false, or other [17]. The features are as follows:

- ID is a unique identifier to identify each instance uniquely.
- Text is the content of the news.
- Title is the headlines of the news.
- Label is the class assigned to the instance as true, partially false, false, or other.

The "partially false" label is associated with articles that contain partially true and partially false information but cannot be considered 100% true. While the "other" label is assigned to articles that cannot be categorized as true, false, or partially false due to the lack of evidence about their claims.

Figure 1 shows training and validation dataset statistics and the imbalance between the four classes in both the training and validation dataset. The training dataset is slightly imbalanced 46% of the texts labeled as false, 28% were labelled as partially false, 17% were labelled as true, and 9% were labelled as other. This shows that approximately half of the training dataset was labelled as false, and approximately one-third was labelled as partially false.

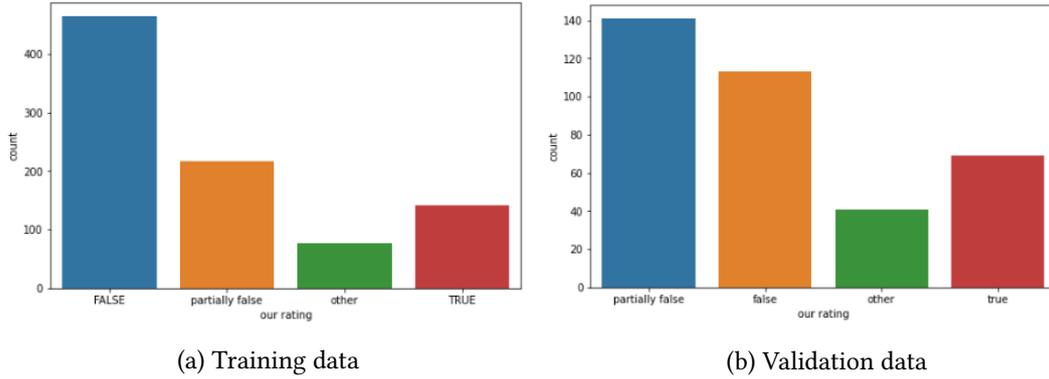


Figure 1: Label distribution of training and validation dataset

4. Methodology

For fake news detection, we used three models: Passive Aggressive Classifier (PAC), a machine learning algorithm; Bi-LSTM, a deep learning algorithm; and RoBERTa [18] a pre-trained language model from the BERT family for fine-tuning. For task 3a, we used the English training dataset that contains four labels to train the model with three different algorithms. We used the English test dataset to test the models' performance. For task 3b, we used the English dataset to train the models and then tested them on the German test dataset to evaluate cross-lingual performance. F1-score was used to measure the models' efficiency.

4.1. Data pre-processing

Data pre-processing is one of the important steps in natural language processing (NLP) tasks. We performed the following data pre-processing in order to prepare the data for training. We performed several pre-processing steps, including removing unwanted characters, stop word removal, as well as converting labels to integers.

4.2. Algorithms

In this section, we discuss algorithms used in this paper: passive aggressive classifier, a machine learning algorithm, Bi-LSTM, a deep learning algorithm, and RoBERTa, a pre-trained language model.

4.2.1. Passive Aggressive Classifier

Passive aggressive classification (PAC) is an online learning classifier that is used in cases where there is a need to keep a 24×7 check on the data like news, social media, etc. [19]. PAC is a noteworthy classifier among the online learning algorithms. The classification function is updated if there is a misclassification in newly seen data or if a predetermined margin is not exceeded by its classification score [20]. The input to PAC is a matrix of TF-IDF features. Thus, a model is formed while trained on the training data and then applied to the test set to evaluate

its performance. We used the following parameters in training fake news detection using PAC, tfidf_vectorizer to transform text into vector, max_df=0.7, C=0.5, max_iter=50, random_state=5.

4.2.2. Bi-LSTM

Deep learning models are widely used for linguistic modeling. Typical deep learning models such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) can detect complex patterns in textual data. Long Short-Term Memory (LSTM) is a tree-structured recurrent neural network used to analyze variable-length sequential data [21]. Bi-directional LSTM(Bi-LSTM) allows looking at a particular sequence both from front to back as well as from back to front. Long short-term memory (LSTM) is a structure that learns how much of the previous network state to apply when input data is received. By using the hidden state, information can flow in both directions. At each time step, the outputs of the two LSTMs are merged. This BiLSTM model helps remove the barriers of traditional RNNs. The BiLSTM provides excellent accuracy and makes the context much easier to understand. Table 1 shows the parameters used for the Bi-LSTM model for our experiment.

Table 1

LSTM parameters used in model training

Parameters used in Bi-LSTM	
Parameters	Values
hidden_units	128
embed_units	100
Dropout	0.2
learning_rate	0.0001
optimizer	adam
batch_size	32
loss	binary cross entropy
num_itr	50
activation	sigmoid
Total params	3,717,745

After pre-processing the text we used tokenizer function from Keras to tokenize the input and added padding to tokenized input. We converted tokenized input after padding to tensors using convert to tensor. To build BiLSTM model we used the keras library package. Our BiLSTM model contains embedding layer which is composed of vocab size of 34,172, embed units of 100, input length of 5,840, dropout layers, a fully connected layer with 256 neurons, binary cross entropy as loss function, adam as optimizer and relu activation. The model is trained using batch size of 32, 50 number of iteration and 0.0001 learning rate.

4.2.3. RoBERTa

RoBERTa is a transformer model pre-trained on a large corpus of English data in a self-supervised fashion. RoBERTa is an optimized BERT model re-trained with improved training methodology, more data, and hardware resources proposed by [18]. RoBERTa without the next sentence prediction concept is similar to BERT and employs dynamic masking, which results in changing

the masked token during the training epochs. RoBERTa is trained with dynamic masking , full-sentences without NSP (Next-Sentence Prediction) loss, large mini-batches and a larger byte-level BPE (Byte-Pair Encoding). Furthermore, RoBERTa is pre-trained on more data, longer sequences and with bigger batch sizes. Table 2 shows parameters used to train the RoBERTa model.

Table 2
RoBERTa parameters used in model training

Parameters used in RoBERTa	
Parameters	Values
hidden_units	128
Dropout	0.1
learning_rate	0.0001
optimizer	adam
batch_size	32
num_itr	20
activation	softmax
Total params	125,798,148

The RoBERTa base model used for this paper was fine-tuned on a given fake news dataset. We used a "roberta-base" model from the Hugging Face library, which is already pre-trained. For both tasks we added batch normalization layer to speed up training, to make learning easier and a fully-connected output layer with a SoftMax function so that a probabilistic output of all labels for fake news detection would be produced. For both tasks we fine tuned the model using 20 number of iteration. For Task 3a, the model is fine-tuned and tested in English dataset while for Task 3b we used English dataset for fine tuning and tested the model with German test data without explicitly training in German dataset.

5. Result and Discussion

Table 3 presents the Accuracy, Precision, Recall, and F1- scores obtained by applying the deep learning and transfer learning methods on the dataset provided by the organizers of the CheckThat! 2022 fake news detection shared task at CLEF 2022. In this table, "Task" refers to the monolingual task in English (Task 3a) and the cross-lingual task for English and German (Task 3b). The "Model" refers to deep learning and transfer learning-based models applied in this study.

As shown in Table 3 for Task 3a (multi-class fake news classification), fine-tuning RoBERTa model on the English dataset gives better results than using Bi-LSTM and PAC. This shows that languages with fewer resources can benefit from using pre-trained models. For Task 3b, cross-lingual fake news classification Bi-LSTM gives better results than RoBERTa and passive classifier by only testing the model on the German test dataset. This shows that monolingual pre-trained language models have lower performance on new languages if a large amount of dataset does not efficiently fine-tune them. Figure 2 and 3 shows the training loss, validation loss, as well as train and validation accuracy of the Bi-LSTM algorithm for task 3b. It can be

Task	Model	Accuracy	Precision	Recall	F1-score
Task 3a	PAC	0.51	0.29	0.25	0.2
	Bi-LSTM	0.52	0.13	0.25	0.17
	RoBERTa	0.47	0.36	0.34	0.29
Task 3b	PAC	-	-	-	-
	Bi-LSTM	0.61	0.37	0.39	0.35
	RoBERTa	0.28	0.13	0.26	0.17

Table 3
Results for Task 3a and Task 3b using three different models

observed that the validation loss is greater than the training loss after 20 epochs, and train accuracy is very high, but the validation accuracy is low and constant as we increase the number of epochs. This shows that the model is not predicting well on the validation data. It may indicate that the model is underfitting and depends on the size of the datasets. The latter is typical for deep learning algorithms.

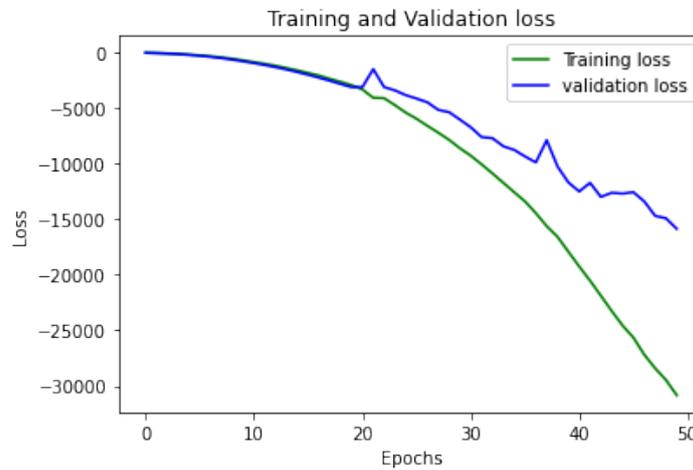


Figure 2: BiLSTM train loss vs validation loss for Task 3b

Figure 4 and 5 displays training loss, validation loss, and train and validation accuracy of the RoBERTa algorithm for task 3b. It is seen that the model’s both training loss and validation loss decrease and stabilize at a specific point. This indicates an optimal fit. Training and Validation accuracy also demonstrate how the model performs on a new dataset. Our results demonstrate that using pre-trained language models for low-resourced languages can give better results than deep learning algorithms.

6. Conclusion and Future Work

This paper discussed some classification models for detecting multi-class and cross-lingual fake news. RoBERTa pre-trained model performed well for the monolingual task compared

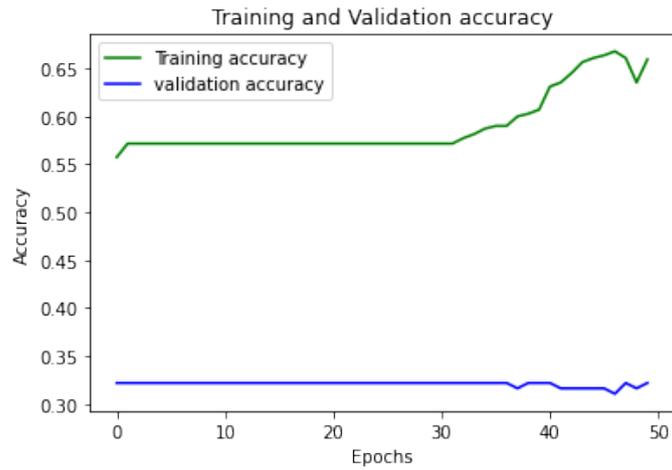


Figure 3: BiLSTM train accuracy vs validation accuracy for Task 3b



Figure 4: RoBERTa train and validation loss for Task 3b

to other Bi-LSTM and PAC algorithms we applied in our experiments. We also observed that for cross-lingual fake news detection, Bi-LSTM performed well compared to RoBERTa. In the future, we will explore how increasing the amount of data will influence the performance of pre-trained models. We also suggest multilingual pre-trained models may improve cross-lingual fake news classification.

Acknowledgments

The work was done with partial support from the Mexican Government through the grant A1S-47854 of CONACYT, Mexico, grants 20220852, 20220859, and 20221627 of the Secretaría de

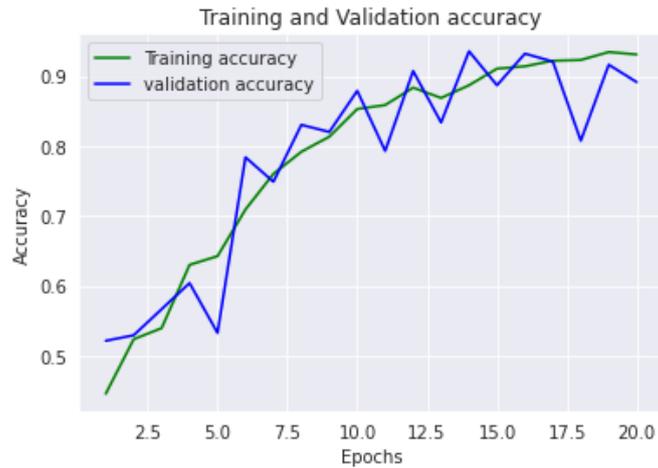


Figure 5: RoBERTa train and validation accuracy for Task 3b

Investigación y Posgrado of the Instituto Politécnico Nacional, Mexico. The authors thank the CONACYT for the computing resources brought to them through the Plataforma de Aprendizaje Profundo para Tecnologías del Lenguaje of the Laboratorio de Supercómputo of the INAOE, Mexico and acknowledge the support of Microsoft through the Microsoft Latin America PhD Award.

References

- [1] A. Thota, P. Tilak, S. Ahluwalia, N. Lohia, Fake news detection: a deep learning approach, *SMU Data Science Review* 1 (2018) 10.
- [2] K. Langin, Fake news spreads faster than true news on twitter—thanks to people, not bots, *Science magazine* (2018).
- [3] H. Allcott, M. Gentzkow, Social media and fake news in the 2016 election, *Journal of economic perspectives* 31 (2017) 211–36.
- [4] J. Gottfried, E. Shearer, News use across social media platforms 2016, <http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/> (2016).
- [5] C. Silverman, L. Alexander, How teens in the balkans are duping trump supporters with fake news. *buzzfeed*, 14 november, 2016.
- [6] N. Karimi, J. Gambrell, Hundreds die of poisoning in iran as fake news suggests methanol cure for virus, 2020.
- [7] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news detection on social media: A data mining perspective, *ACM SIGKDD explorations newsletter* 19 (2017) 22–36.
- [8] Z. Khanam, B. Alwasel, H. Sirafi, M. Rashid, Fake news detection using machine learning approaches, in: *IOP Conference Series: Materials Science and Engineering*, volume 1099, IOP Publishing, 2021, p. 012040.

- [9] V. L. Rubin, Y. Chen, N. K. Conroy, Deception detection for news: three types of fakes, *Proceedings of the Association for Information Science and Technology* 52 (2015) 1–4.
- [10] D. Fallis, A functional analysis of disinformation, *IConference 2014 Proceedings* (2014).
- [11] E. C. Tandoc Jr, Z. W. Lim, R. Ling, Defining “fake news” a typology of scholarly definitions, *Digital journalism* 6 (2018) 137–153.
- [12] W. Y. Wang, " liar, liar pants on fire": A new benchmark dataset for fake news detection, *arXiv preprint arXiv:1705.00648* (2017).
- [13] S. K. W. Chu, R. Xie, Y. Wang, Cross-language fake news detection, *Data and Information Management* 5 (2021) 100–109.
- [14] Y. Li, B. Jiang, K. Shu, H. Liu, Mm-covid: A multilingual and multimodal data repository for combating covid-19 disinformation, *arXiv preprint arXiv:2011.04088* (2020).
- [15] G. K. Shahi, J. M. Struß, T. Mandl, Overview of the clef-2021 checkthat! lab task 3 on fake news detection, *Working Notes of CLEF* (2021).
- [16] 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.
- [17] G. K. Shahi, A. Dirkson, T. A. Majchrzak, An exploratory study of covid-19 misinformation on twitter, *Online Social Networks and Media* 22 (2021) 100104.
- [18] 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).
- [19] K. Nagashri, J. Sangeetha, Fake news detection using passive-aggressive classifier and other machine learning algorithms, in: *Advances in Computing and Network Communications*, Springer, 2021, pp. 221–233.
- [20] J. Lu, P. Zhao, S. C. Hoi, Online passive-aggressive active learning, *Machine Learning* 103 (2016) 141–183.
- [21] P. Bahad, P. Saxena, R. Kamal, Fake news detection using bi-directional lstm-recurrent neural network, *Procedia Computer Science* 165 (2019) 74–82.