Fake or real? The novel approach to detecting online disinformation based on multi ML classifiers

Martyna Tarczewska, Anna Marciniak[0000−0002−9900−3951], and Agata Giełczyk[0000−0002−5630−7461]

University of Science and Technology, Bydgoszcz, Poland
agata.gielczyk@utp.edu.pl

Abstract. Background: the machine learning (ML) techniques have been implemented in numerous applications and domains, including healthcare, security, entertainment, and sports. This paper presents how ML can be used for detecting fake news. The problem of online disinformation has recently become one of the most challenging issues of computer science. Methods: in this research, a fake news detection method based on multi classifiers (CNN, XGBoost, Random Forest, Naive Bayes, SVM) has been developed. In the proposed method, two classifiers cooperate; consequently, they obtain better results. Realistic, publicly available data was used in order to train and test the classifiers, Results: in the article, several experiments were presented; they differ in the implemented classifiers, and some improved parameters. Promising results (accuracy = 0.95, precision = 0.99, recall = 0.91, and F1-score = 0.95) were reported. Conclusion: the presented research proves that machine learning is a promising approach to fake news detection.

Keywords: Fake news · Online disinformation · Machine learning

1 Introduction

According to the Collins dictionary, fake news can be defined as 'false, often sensational, information disseminated under the guise of news reporting'[1]. Despite the fact that fake news existed for many years, its impact has recently increased. This trend can be easily observed, e.g., by means of the Google Trends tool[2]. It shows that the phrase 'fake news' has rapidly become more popular since November 2016. Traditionally, fake news was known as rumors or propaganda, mostly used in order to make political or economic gains. The main goal of creating fake news has remained unchanged. However, currently it can spread more easily thanks to the popularity of social networks. The current pandemic reality has led to a serious outbreak of misinformation. It can be very dangerous in social, health-care and political aspects, like in the case of the fake news concerning the COVID-19 pandemic and its connection with the 5G transmission [2], etc.

[1] https://www.collinsdictionary.com/dictionary/english/fake-news
[2] https://trends.google.com/trends/explore?date=today%205-y&q=fake%20news

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77970-2_2

Several subtypes of fake news can be listed [12]:

- rumor - an item of circulating information the veracity status of which is yet to be verified at the time of posting;
- hoax - a deliberately fabricated falsehood made to masquerade as truth;
- click-bait - a piece of low-quality journalism which is intended to attract traffic and monetize via advertising revenue;
- disinformation - fake or inaccurate information which is intentionally false and spread deliberately;
- misinformation - fake or inaccurate information which is spread unintentionally;
- fake news - a news article that is intentionally and verifiable false.

Moreover, it is possible to find some pieces of information which can be classified as satire. Unlike subtler forms of deception, satire may feature more obvious cues that reveal its disassociation from the truth. In fact, satire is meant to be recognized as a joke, at least by some readers.

Due to the variety of fake news types, the methods used in order to classify it are also diverse. The typology of fake news detection approaches is presented in Fig. 1. One of the most popular approaches to detecting fake news is NLP, consisting in analyzing the text of the news/tweet/post [5]. In such approaches, pattern recognition systems are trained in order to discover lexical [10], linguistic [8], psycholinguistic [23], syntactic [4] and semantic [9] features.

The general concept behind the authors' reputation system is to evaluate the source of the information - it can be a publisher, a www address or an IP address. In such an approach, some websites or information providers (e.g., CNN or BBC) can be assumed to be reliable. A sample system focusing on the author's credibility was presented in [20].

Another approach is to implement network analysis, which refers to the network and graph theory. In this approach, the relations between the news' author and the user who reposts or shares it are discovered, as presented in [19].

Since images have become a dominant and powerful communication channel, the last but not least group of methods used in order to detect the fake news is based on image analysis. The ML-based approach to fake news detection by recognizing image forgery is presented in [11] and in [14].

The remainder of the paper is structured as follows: Section 2 presents the current state of the art. In Section 3, the proposed solution is described in detail. Section 4 contains and discusses the obtained results. Section 5 provides threats to validity, and conclusions.

2 Related work

Amongst the approaches to detecting fake news, convolutional neural networks (CNN), support vector machine (SVM), random forest (RF) and the XGBoost classifier are currently the most commonly used ones.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77970-2_2

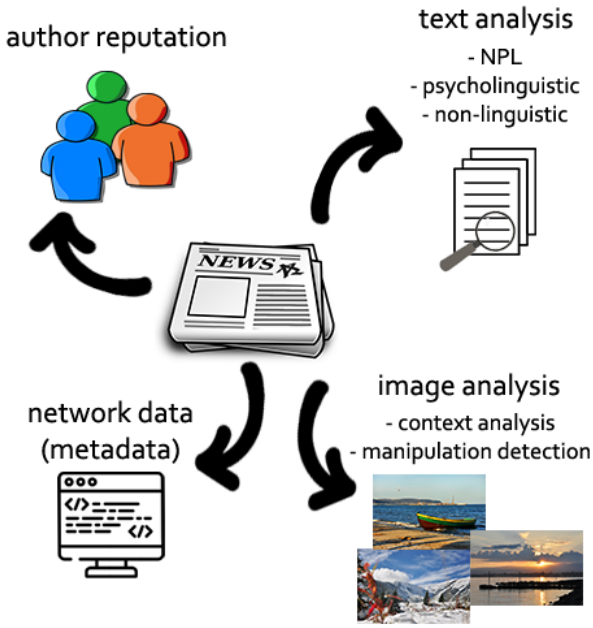

Fig. 1. The typology of fake news detection approaches [5]

The authors in [18] developed a system for fake news detection with the use of supervised learning methods. The authors compared several algorithms, including k-Nearest Neighbors (kNN), Naive Bayes (NB), Random Forests (RF), Support Vector Machine with RBF kernel (SVM), and XGBoost (XGB). In this case, the best results were obtained with RF and XGB - AUC 0.85 and 0.86, respectively; F1-score 0.81 for both models.

The authors in [1] tested various machine learning methods on four different datasets. Altogether, 13 methods were tested, including CNN, LSTM and XGB. Among those three methods, XGB achieved the best accuracy (over 89% each time). Among the other tested methods, the best results were obtained by RF and linear SVM classifiers, with accuracy over 91% (in 3 out of 4 datasets) and over 90% (in 3 out of 4 datasets), respectively.

The authors in [22] developed a novel, hybrid CNN to integrate metadata with text. Authors compared two approaches: text-only models (including SVM, logistic regression, Bi-LSTM and CNN) and test and meta-data hybrid models (hybrid CNN). The better results were obtained with the hybrid CNN approach, both in the test and in the validation dataset.

In [7], the authors proposed a deep neural network approach, where CNN and LSTM models were used. Both single models and their combinations have been tested and compared with previously developed models, including SVM and the model described above. However, the authors' approach did not result in better accuracy (97.84%) than the previously published models [24].

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77970-2_2

The authors in [13] studied fake news detection with different degrees of fakeness by integrating several sources. The authors proposed a Multi-source Multi-class Fake news Detection framework (MMFD), with automated feature extraction (performed with the use of CNN and LSTM), incorporated multi-source fusion and fakeness discrimination. Moreover, the authors compared MMFD to SVM, RF, kNN and Wang method (described above [22]). In each case, the MMFD achieved better accuracy.

The authors in [17] developed a system for automatic fake news detection; they focused on the preparation of new data for further analysis, which consists of the evaluation of the linguistic features, creating a machine learning model and making a comparison to human performance. Linear SVM classifier and five-fold cross-validation were used to create the fake news detector and R, caret and e1071 packages were used to conduct machine learning classification. The results achieved by the models were comparable to those achieved by humans.

The authors in [21] focused on fake news detection using deep learning architecture. In this model, the authors included both the CNN and LSTM neural networks and combined them with principle component analysis (PCA) and Chi-Square. By using this approach, the authors achieved the fake news detection accuracy of 97.8%.

In [15], the approach based on kNN was developed. The authors used the dataset which has been collected from Buzz Feed News organization and which is commonly used in scientific methods. It contains Facebook posts. Using the proposed approach, it was possible to obtain accuracy reaching 79%.

The article [3] presented a hybrid architecture is which is based on Bidirectional LSTM and Convolutional Neural Network. Using both types of classifiers enabled to incorporate news content and information concerning the user profile as well. The proposed hybrid architecture performed better than individual architecture and it gave overall accuracy of 42.2%. The experiments were conducted using the Liar dataset. Authors pointed that the similarity of classes in the dataset (pants-fire, false, barely-true were claimed to be almost same) was the biggest challenge in this research.

Authors in [16] also proposed a novel hybrid method. Their model combined the Convolutional Neural Network and the Recurrent Neural Network for fake news classification. It was successfully validated on two fake news datasets (ISO and FA-KES), achieving detection results that was significantly better than other non-hybrid baseline methods. One of the key points of the proposed method was the pre-processing, namely Word2Vec provided by Google and GloVe, pre-trained word embedding.

The interesting solution was introduced in [25], where the explainable fake news detection tool was presented. In this approach the XGBoost classifier was implemented in order to detect the online disinformation. The usability of the proposed solution was demonstrated on a real-world dataset crawled from PolitiFact, where thousands of verified political news have been collected.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77970-2_2

3 Presented approach

3.1 Dataset and pre-processing

In this research, a publicly available dataset was used, which can be downloaded from the Kaggle website (https://www.kaggle.com/c/fake-news/data). A single row in the dataset contains the following elements: id - unique id for a news article, title - the title of a news article, author - the author of the news article, text - the text of the article and label - that marks the article as potentially unreliable. The label equal to 1 shows that the article is fake news, whereas 0 means that it is reliable news. The initial pipeline of the proposed method is presented in Fig. 2. First of all, the pre-processing needs to be performed. Thus, the article body is converted to the lower case, and the stop words which are commonly used words (such as 'the', 'a', 'an', 'in'), and do not significantly impact of the whole text's sense, are deleted.

The dataset contains over 20k labeled rows. The dataset is well balanced - half of the articles were marked as fake, and the other half as real. During the experiments, the dataset was divided into the training set (80%) and the testing set (20%). Thus, the training data was obtained, which also was balanced - 2k fake news and 2k reliable news. the experiments were performed using the 5-fold cross validation.

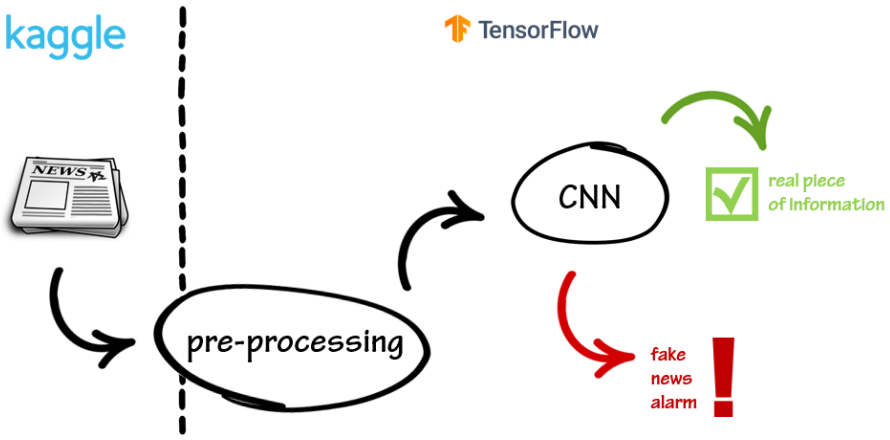

Fig. 2. The pipeline of the CNN-based method - the starting point for further experiments

3.2 Machine learning

All the experiments were performed using the Keras API that works with Tensorflow. These tools enabled using the machine learning methods. As mentioned in

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77970-2_2

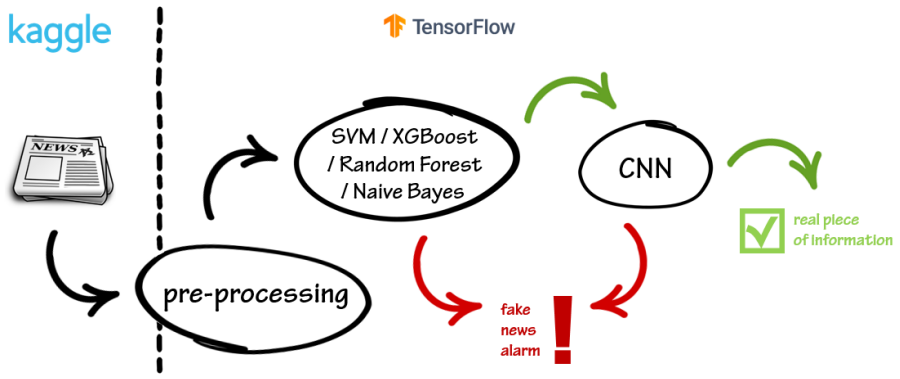

Fig. 3. The pipeline of the proposed method

the state-of-the-art review, there are several ML methods that are widely implemented in fake news detection. In order to perform this research, the following ones were selected: Convolutional Neural Network (CNN), eXtreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Naive Bayes (NB) and Random Forest (RF).

First of all, the CNN training and testing were run with the default parameters of the network. This kind of approach would be the starting point for the next algorithms; its general pipeline is presented in Fig. 2. The next step was to improve the parameters of the CNN so that it could give higher accuracy. Thus, 128 convolutional layers with activation type=relu were added, the dropout as modified to 0.2 and 10 dense layers with activation type=relu were added. The proposed improvements were performed according to the state-of-the-art review and the authors' experience. The CNN with improved parameters is further called the boosted CNN.

The next step of the proposed method was to implement a number of methods: XGBoost, SVM, NB, and RF as a single classifier in place of CNN.

The last part of the research was to add the additional classifier which would initially scan the articles. This approach is presented in Fig. 3. The first classifier verifies the article. If the article receives the label 'fake', it is finally marked as the false piece of information (red arrow in Fig. 3). Otherwise, the next classifier analyzes the article (green arrow in Fig. 3). The decision of the second classifier is final and the article gets the label fake or reliable. The additional classifiers were again: XGBoost, SVM, NB and RF, whereas the final decision was made by the boosted CNN.

4 Results and discussion

Since the fake news detection problem can be understood as a binary classification, confusion matrices were used in order to evaluate and compare the ML-based methods. Four measures were defined as follows:

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77970-2_2

– TP - true positives - fake news classified as fake news;
– FP - false positives - fake news classified as reliable pieces of information;
– FN - false negatives - real news classified as fake news;
– TN - true negatives - real news classified as reliable pieces of information.

Each model in this research was evaluated using Accuracy (Eq. 1), Precision (Eq. 2), Recall (Eq. 3) and F1-score (Eq. 4), which use the above mentioned measures TP, FP, FN and TN.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{4}$$

Table 1. Obtained results

Classifier	Accuracy	Precision	Recall	F1-score
CNN (default)	0.8889	0.8696	0.9006	0.8846
CNN (boosted)	0.9213	0.9150	0.9248	0.9194
XGBoost	0.8992	0.8778	0.9135	0.8953
SVM	0.6311	0.6096	0.6276	0.6184
Naive Bayes	0.5810	0.4801	0.5893	0.5291
Random Forest	0.7853	0.7112	0.8269	0.7647
CNN + XGBoost	**0.9487**	**0.9941**	**0.9117**	**0.9511**
CNN + SVM	0.8328	0.9736	0.7603	0.8538
CNN + Bayes	0.7921	0.9565	0.7205	0.8219
CNN + Random Forest	0.9458	0.9941	0.9068	0.9485

The obtained results are presented in Table 1. As seen in it, by modifying selected parameter of CNN it was possible to improve the results. When it comes to the comparison of the single classifiers (CNN excluded), the most promising results were obtained by XGBoost (Acc=90%, Prec=88%, Rec=91% and F1=90%). Each classifier used with CNN gave the improved results. It is also remarkable that the combination CNN + Random Forest is very promising, even though Random Forest used as a single classifier was not impressive (Acc=79%, Prec=71%, Rec=83% and F1=76%). Nevertheless, the highest values of Accuracy, Precision, Recall and F1-score were achieved by connecting CNN with XGBoost, namely Acc = 0.95, Prec = 0.99, Rec = 0.91 and F1 = 0.95. This result is the most encouraging and thus, marked in bold in the Table 1. The obtained results are also presented in a visual way as the confusion

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77970-2_2

matrices in Fig. 4. The selected experiments results have been presented there: default CNN, boosted CNN, XGBoost, CNN+XGBoost, RF and CNN+RF. The results' improvement is especially visible between RF and CNN+RF (the third row), where the number of FP was decreased significantly.

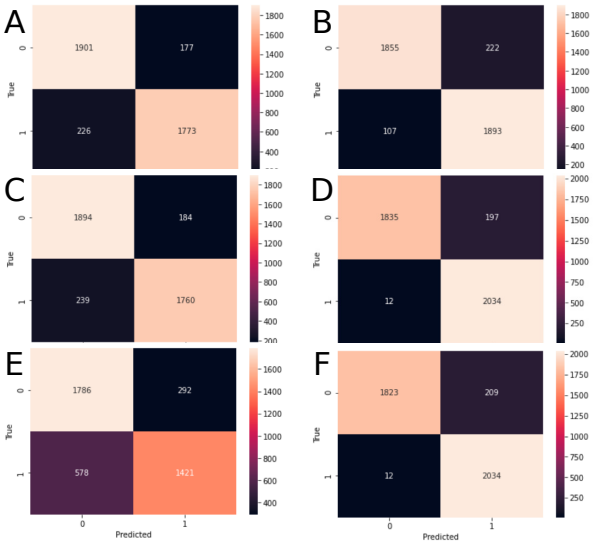

Fig. 4. Confusion matrices for the selected experiments - A: CNN default, B: CNN boosted, C: XGBoost, D: CNN + XGBoost, E: RF and F: CNN + RF

5 Conclusions

In this paper, an efficient and accurate ML-based approach to the fake news detection has been presented. The obtained results were promising, as seen in Table 2. Consequently, the approach enables obtaining results that are similar to the results of the current state-of-the-art approaches. However, it is essential to mention that the fake news detecting methods are hardly comparable due to the variety of the fake news datasets.

The proposed solution may be extended in the future, e.g., by implementing another type of classifier or introducing some more pre-processing methods. Other possible extension that could be done in the nearest future are rebuilding the pipeline of the proposed solution and adding the block of explanation. This kind of approach could give both the fake/real assessment and the explanation why the algorithm decided in such a way.

It is also remarkable, as proposed in [6], that automatic fake news detection tools should be designed to augment human judgement, not to replace it. The human aspect would be especially helpful in recognizing satire and jokes.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77970-2_2

Table 2. Obtained results in related works and in the proposed method

Reference	Dataset	Method	Result
De Sarkar et al. [7]	LIAR	CNN	Accuracy=98%, Precision=93%, Recall=84%, F1-score=88%
Reis et al. [18]	BuzzFace	XGBoost	AUC=86%, F1=81%
Ahmad et al. [1]	ISOT Fake News Dataset	XGBoost	Accuracy=98%, Precision=99%, Recall=99%, F1-score=99%
Tarczewska et al.	**Kaggle**	**CNN+XGBoost**	**Accuracy=95%, Precision=99%, Recall=91%, F1-score=95%**

References

1. Ahmad, I., Yousaf, M., Yousaf, S., Ahmad, M.O.: Fake news detection using machine learning ensemble methods. Complexity **2020** (2020)
2. Ahmed, W., Vidal-Alaball, J., Downing, J., López Seguí, F.: Covid-19 and the 5g conspiracy theory: Social network analysis of twitter data. J. Med Internet Res **22**(5), e19458 (May 2020). https://doi.org/10.2196/19458
3. Balwant, M.K.: Bidirectional lstm based on pos tags and cnn architecture for fake news detection. In: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). pp. 1–6. IEEE (2019)
4. Capistrano, J.L.C., Suarez, J.J.P., Naval Jr, P.C.: Salsa: Detection of cybertrolls using sentiment, aggression, lexical and syntactic analysis of tweets. In: Proceedings of the 9th International Conference on Web Intelligence, Mining and Semantics. pp. 1–6 (2019)
5. Choraś, M., Demestichas, K., Giełczyk, A., Herrero, Á., Ksieniewicz, P., Remoundou, K., Urda, D., Woźniak, M.: Advanced machine learning techniques for fake news (online disinformation) detection: A systematic mapping study. Applied Soft Computing p. 107050 (2020)
6. Conroy, N.K., Rubin, V.L., Chen, Y.: Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology **52**(1), 1–4 (2015)
7. De Sarkar, S., Yang, F., Mukherjee, A.: Attending sentences to detect satirical fake news. In: Proceedings of the 27th International Conference on Computational Linguistics. pp. 3371–3380 (2018)
8. Dey, A., Rafi, R.Z., Parash, S.H., Arko, S.K., Chakrabarty, A.: Fake news pattern recognition using linguistic analysis. In: 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR). pp. 305–309. IEEE (2018)
9. Gaglani, J., Gandhi, Y., Gogate, S., Halbe, A.: Unsupervised whatsapp fake news detection using semantic search. In: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS). pp. 285–289. IEEE (2020)

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77970-2_2

10. Giglou, H.B., Razmara, J., Rahgouy, M., Sanaei, M.: Lsaconet: A combination of lexical and conceptual features for analysis of fake news spreaders on twitter. In: CLEF (2020)
11. Gragnaniello, D., Marra, F., Poggi, G., Verdoliva, L.: Analysis of adversarial attacks against cnn-based image forgery detectors. In: 2018 26th European Signal Processing Conference (EUSIPCO). pp. 967–971. IEEE (2018)
12. Guo, B., Ding, Y., Sun, Y., Ma, S., Li, K., Yu, Z.: The mass, fake news, and cognition security. Frontiers of Computer Science **15**(3), 1–13 (2021)
13. Karimi, H., Roy, P., Saba-Sadiya, S., Tang, J.: Multi-source multi-class fake news detection. In: Proceedings of the 27th International Conference on Computational Linguistics. pp. 1546–1557 (2018)
14. Kasban, H., Nassar, S.: An efficient approach for forgery detection in digital images using hilbert–huang transform. Applied Soft Computing **97**, 106728 (2020)
15. Kesarwani, A., Chauhan, S.S., Nair, A.R.: Fake news detection on social media using k-nearest neighbor classifier. In: 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE). pp. 1–4. IEEE (2020)
16. Nasir, J.A., Khan, O.S., Varlamis, I.: Fake news detection: A hybrid cnn-rnn based deep learning approach. International Journal of Information Management Data Insights **1**(1), 100007 (2021)
17. Pérez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R.: Automatic detection of fake news. arXiv preprint arXiv:1708.07104 (2017)
18. Reis, J.C., Correia, A., Murai, F., Veloso, A., Benevenuto, F.: Supervised learning for fake news detection. IEEE Intelligent Systems **34**(2), 76–81 (2019)
19. Shu, K., Bernard, H.R., Liu, H.: Studying fake news via network analysis: detection and mitigation. In: Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining, pp. 43–65. Springer (2019)
20. Sitaula, N., Mohan, C.K., Grygiel, J., Zhou, X., Zafarani, R.: Credibility-based fake news detection. In: Disinformation, Misinformation, and Fake News in Social Media, pp. 163–182. Springer (2020)
21. Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A., Choi, G.S., On, B.W.: Fake news stance detection using deep learning architecture (cnn-lstm). IEEE Access **8**, 156695–156706 (2020)
22. Wang, W.Y.: "liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648 (2017)
23. Wawer, A., Wojdyga, G., Sarzyńska-Wawer, J.: Fact checking or psycholinguistics: How to distinguish fake and true claims? In: Proceedings of the Second Workshop on Fact Extraction and VERification (FEVER). pp. 7–12 (2019)
24. Yang, F., Mukherjee, A., Dragut, E.: Satirical news detection and analysis using attention mechanism and linguistic features. arXiv preprint arXiv:1709.01189 (2017)
25. Yang, F., Pentyala, S.K., Mohseni, S., Du, M., Yuan, H., Linder, R., Ragan, E.D., Ji, S., Hu, X.: Xfake: explainable fake news detector with visualizations. In: The World Wide Web Conference. pp. 3600–3604 (2019)

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77970-2_2