Multi-Label Fake News Detection using Multi-layered Supervised Learning

Tayyaba Rasool
National University of Sciences and Technology
Islamabad 44000, Pakistan
tayyaba.rasool93@gmail.com

Wasi Haider Butt
National University of Sciences and Technology
Islamabad 44000, Pakistan
wasi@ceme.nust.edu.pk

Arslan Shaukat
National University of Sciences and Technology
Islamabad 44000, Pakistan
arslan.shaukat@ce.ceme.edu.pk

M. Usman Akram
National University of Sciences and Technology
Islamabad 44000, Pakistan
usman.akram@ceme.nust.edu.pk

ABSTRACT

Rapid spreading of misinformation is a growing worldwide concern as it has the capacity to greatly influence individual reputation and societal behavior. The consequences of unchecked spreading of misinformation can not only vary from political to financial but also effect global opinion for a long time. Thus, detecting fake news is important but challenging as the ability to accurately categorize certain information as true or fake is limited even in human. Moreover, fake news are a blend of correct news and false information making accurate classification even more confusing. In this paper, we propose a novel method of multilevel multiclass fake news detection based on relabeling of the dataset and learning iteratively. The proposed method outperforms the benchmark and our experiments indicate that profile of the source of information contributes the most in fake news detection.

CCS Concepts

• **CCS →Computing methodologies → Artificial intelligence**

Keywords

Fake News; data mining; SVM; Decision tree; Misinformation.

1. INTRODUCTION

Information transfer has become very rapid today, thanks to the internet and everything built around it. With faster information exchange and lack of control over the content that is communicated over the internet, comes faster misinformation exchange. Misinformation can spread rapidly over wide areas and is being perceive as a global challenge because disinformation has the potential to be used effectively as a weapon in modern conflicts. This is because online spread of misinformation has the power to shape public opinions and sentiments.

One type of misinformation is fake news, consisting of false information that is disseminated as factually accurate to mislead the audience using traditional or social media. Fake news on any media can effect a society or individual in various ways. It can have severe consequences in terms of financial and political decision making as well as destroying an individual's social standing. Numerous examples of sharing of unverified news by individuals and organizations can be observed in the recent past. For example, after the Las Vegas shooting false information about the gunman began to circulate on social media. A picture of comedian Sam Hyde was repeatedly shared as being that of the gunman[1].

Despite newer manifestations, fake news is by no means a new problem. Different entities use news media for propaganda and attempt to influence other entities. The faster information exchange makes fake news more powerful, especially since US elections 2016. The open nature of the internet makes it easier to create and disseminate fake news causing it to spread faster and deeper [1]. Therefore, detecting and countering fake news is essential for a free and fair global society.

The process of fake news detection is aimed at predicting the probability of a news being intentionally fake [2]. It is a challenging task primarily for two reasons. Firstly, the challenge is due to the fact that the human ability to correctly classify true and fake news is quite unimpressive. Human beings, by rough comparison, can achieve only 50-63% success rate in identifying fake news [3]. Secondly, fake news usually include a mixture of correct news and misinformation. It can easily confuse the audience and capture their attention without them noticing the fabricated information. Several other tasks related to fake news detection exist including rumor detection [4] and spam detection [5].

Categorization of detection algorithms can be determined by feature extraction of news. Fake news detection techniques mainly rely on the content of a news story and the social context associated with it [6]. News content describes the information related to the piece of news including attributes such as the headline, source, detail and visual content. On the other hand, social context feature can be derived from the social engagement of news. Social context features may be related to the audience, network of news propagation and reaction to the news.

As mentioned previously, a news item may not necessarily be completely true or fabricated. In practice, real news may be distorted or corrupted with an inaccurate or exaggerated account of events related to the news story. A multiclass approach is needed to address the fake news detection problem. In this paper,

ICCAE 2019, February 23–25, 2019, Perth, WN, Australia

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6287-0/19/02…$15.00

https://doi.org/10.1145/3313991.3314008

[1] https://www.vanityfair.com/news/2018/01/the-6-fakest-fake-news-stories-of-2017

we propose a solution to the multiclass fake news detection problem using multilayer supervised learning.

2. LITERATURE REVIEW

Fake news detection is the process of classifying news by estimating the measure of certainty. A review of existing literature indicates that it is typically achieved through feature extraction and the features can be mainly divided into two categories: news context and social context.

News context features describe the meta-information related to a piece of news that includes the source of news, a short title and the details of the story. Different features can be extracted from these content. Typically, the news content will mostly be linguistic-based [7-20] or visual based [20, 21].

Linguistic approaches analyze the language pattern to extract cues of a deceptive message, following the belief that the language of truth differs from fabricated information [22, 23]. For example, fake news may contain some contradiction with facts or authentic resources [24]. Alternatively, a fake news item may contain more sentiment, negative emotion, less self-oriented pronouns, etc. [25] or a different grammatical structure [26].

Visual-based features are extracted from visual elements. Images were classified based on user-level and tweet-level features on twitter using a classification framework [27]. In [21], the authors propose several visual and statistical features to characterize different image distribution patterns while in [20] a combination of both visual as well as textual features of news item were used for fake news identification.

Social context features are user-level social engagements of news on media platforms. These features fall in three broad categories: user-based [13, 16, 21, 28-30], based on generated posts [14-16, 29-32] and network-based [13, 21, 33-36].

User based features are extracted by analyzing the attributes of users who interacted with the news [28, 37]. Post-based features emphasis on extracting useful information to estimate the validity of news from relevant sources such as topic, credibility etc. Network-based features are derived by constructing network of users who are involved in the circulation of news on media.

A review of existing literature indicates that the fake news identification problem is often considered as a binary classification problem. However, most news cannot be explicitly classified as absolutely true or absolutely false. As argued earlier, they are generally a combination of misinformation and true news. Therefore, a multi-class classification approach may be more realistic. Multiclass classification was used for stance detection where the authors' opinion or perception about a news story is automatically detected. The classification in this case predicts labels related to whether the author 'agrees', 'disagrees', 'discusses' a piece or the author's opinion is 'unrelated' [12, 18, 38]. Moreover, most of existing related work has been targeted at identifying false information in the domain of social media. Third, classification using the speaker's profile is in early phase of research and we believe the speaker's profile can be effectively used for fake news detection. In this research, we propose a novel multilevel approach for multi-label classification of fake news.

3. SOLUTION

We propose a novel approach for fake news classification based on a three step process consisting of feature extraction, relabeling and learning. Two of these steps are performed iteratively at multiple levels (See Figure 1). Let $N = \{n_1, n_2, \ldots, n_n\}$ be a set of news items, L_0 be the set of original labels and $M = \{m_1, m_2, \ldots, m_n\}$ be the set of trained models.

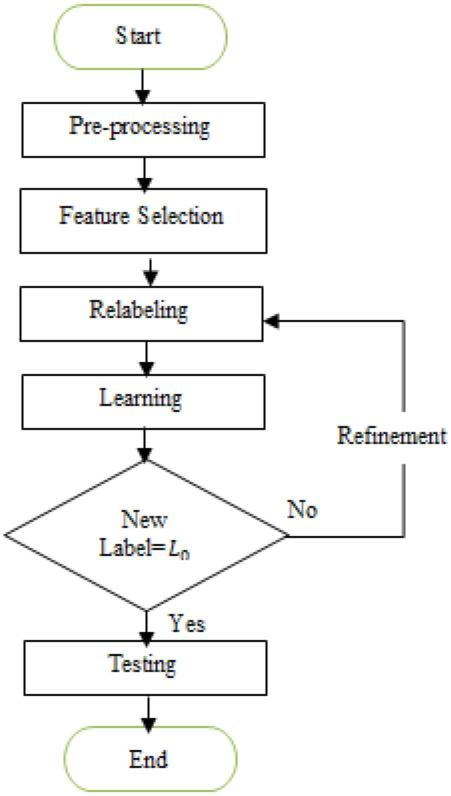

Figure 1. Multi-Label Classification Algorithm

In our approach, the data N is pre-processed manually before it can be used for training a classifier. This is achieved by first cleaning the records manually in N. Moreover, the training data N is converted to numeric values.

We perform feature selection to exclude certain irrelevant features that may increase complexity and degrade the performance and/or accuracy of the algorithm.

In the relabeling step, we simplify the multiclass label problem by relabeling records. Initially, multiple class labels with similar properties are considered as a single class label. For example, for a set of original labels $L_0 = \{1,2,3,4,5\}$, the first three label i.e. 1, 2 and 3 are can be considered as a single label. On the other hand, 4 and 5 show higher rating so they are high rating labels.

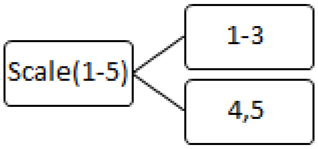

Figure 2. Conversion of multiclass to binary class

Consequently, we have reduced the multiclass labels into 2 class labels (See Figure 2). From here on, we solve the multiclass problem with binary class solutions.

After relabeling process, we train classifier on the new class labels assigned in the previous step. The learning algorithm determines patterns in the training data that enable it to map input data attributes to the new class labels. The resulting ML model $m_i \in M$ captures these relationships.

We define the process of refinement as the process of relabeling and (re-)learning iteratively. Refinement is performed for each binary class individually which means that labels in each class are refined such that they move one hierarchical level closer to the original labels L_0. Consider the previous example, the class label *low* is relabeled again as *low* (1, 2) and *high* (3). This relabeled data is then used to train a new model $m_j \in M$.

Label refinement repeats itself until new labels converge to the original labels. The result is multiple trained models $m = m_{1;}\, m_{1;} \ldots m_n$ that will be used for classification. For the previous example, different trained models are represented by dotted lines (See Figure 3).

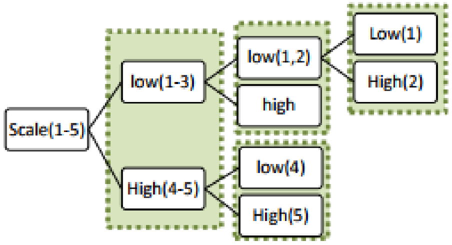

Figure 3. Example of Multiple trained models

4. EXPERIMENTATION AND RESULTS

The proposed solution has been evaluated using the LIAR dataset [39]. The dataset contains 12,836 news statements from various speakers in POLITIFACT.COM [2]. The dataset contains text of news, topic, and some features related to speaker's profile that are speaker name, job title, state, party affiliation, location of speech and credit history. The credit history consist of historical records of incorrect statements for each speaker. There are six labels; *pants-fire, false, barely-true, half-true, mostly-true,* and *true*. We have manually clean the dataset, thus the remaining 10235 records have been used for experimentation. The dataset was then codified into numeric dataset manually. The main reasons of manual codification are the spelling mistakes and use of acronyms in the dataset. For example, fb is used for Facebook in some records.

For feature selection, forward feature selection technique is used. It was performed using Weka 3.8 and 5 features were extracted that were *barely-true* count, *false* count, *half-true* count, *mostly-true* count and *pants-fire* count. It can be seen that all the selected features belong to the speaker's credit history.

As already discussed, the dataset has six labels, two of them belong to the false category and four belong to the true category. Initially, the dataset was relabeled in such a way that class labels *false* and *pants-fire* are considered false. Similarly, *barely-true, half-true, mostly-true* and *true* are considered as true.

As discussed in the proposed methodology, the classifier is then trained on new labels. At first, the machine is trained for *all trues* (*barely-true, half-true, mostly-true, true)* and *all false* (*false, pants-fire*) class labels.

After m_1 is trained for true and false class labels, they are considered two different paths. The training dataset is relabeled to new labels one step closer to the original such that a true is classified further as either {*barely-true, half-true}* or *{mostly-true, true}* (see Figure 4).

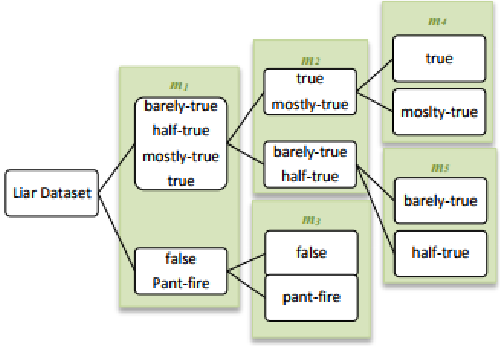

Figure 4. Trained Models on LIAR dataset

4.1 Experimentation

The algorithm is tested using different classification algorithms that are SVM and decision tree. In order to illustrate the working of the proposed approach classification using SVM is presented however, the procedure for other classification is similar. Three evaluation methods are tested i.e. hold-out method, test set evaluation and cross validation.

4.1.1 Holdout Testing using SVM

For testing, the dataset is divided into training and test datasets in the ratio of 70:30. Thus, of the total 10235 records, 7165 were used for training and 3071 records were used for testing purposes. The detailed statistics of holdout testing are as follows:

1. After training the m_1 produces 85.57% accurate results for *all trues* and *all false* class labels. Among these 448 out of 811 records belong to the *false* label while 2180 of 2260 belong to the *true* category.
2. The accuracy of m_2 is 80.59% which means 1757 out of 2180 records were classified correctly. Among these 1099 records belong to {*mostly-true, true}* class label and 1081 records belong to {*barely-true, half-true}* class label.
3. The model m_3 produces an accuracy of 89.7% which means 402 out of the 448 records were predicted correctly. 308 *false* records and 94 *pants-fire* records are predicted correctly.
4. The accuracy of m_4 is 80.7% that are 877 records out of 1099 records are accurately classified. 584 of these records belong to *mostly-true* and 303 belong to the *true* class label.
5. Similarly, m_5 produces 85.8% accuracy, which means 870 records are correctly predicted. Among these 307 records are *barely-true* and 440 records are *half-true.*

Collectively, 2036 records are correctly predicted out of 3071 records, hence the accuracy of the proposed solution with hold-out testing method is 66.29%.

4.1.2 Testing on Test Set using SVM

The complete dataset of size 10235 is used for training and a separately given test dataset is used for testing which consists of another 1265 records.

1. After training the m_1 produces 78.8% accurate results for *all trues* and *all false* class labels. Among these, 130 records were *false* while 867 records were *true.*

[2] https://www.politifact.com/

2. Accuracy of m_2 is 65.7%, 292 of these records belong to {*mostly-true, true}* class label and 278 records belongs to {*barely-true, half-true}* class label.
3. The model m_3 has the accuracy of 82.30%. Among these, 67 *false* records and 40 *pants-fire* records are predicted correctly.
4. The accuracy of m_4 is 68.15% that are 149 *mostly-true* records and 60 true records.
5. Similarly, m_5 produces 69.7% accuracy, which includes 75 *barely-true* records and 119 *half-true* records.

Collectively, 500 records are correctly predicted out of these 1256 records. Hence, the accuracy of proposed method on testing data is 39.5%.

4.1.3 Cross Validation using SVM

The complete dataset of size 10235 is used for k -fold classification where $k = 10$.

1. After training the m_1 produces 80% accurate results for *all trues* and *all false* class labels. Out of which, 1207 records were false and 6981 records were true.
2. Accuracy of m_2 is 69.08%, 2408 of these records belong to *mostly-true and true* class label and 2415 records belongs to *barely-true and half-true* class label.
3. The model m_3 has the accuracy of 82.02%, 687 *false* records and 303 *pants-fire* records are predicted correctly.
4. The accuracy of m_4 is 70.39% that are 1247 *mostly-true* records and 448 true records.
5. Similarly, m_5 produces 72.38% accuracy, which includes 657 *barely-true* records and 1091 *half-true* records

Collectively, 4433 records are correctly classified out of 10235 records. Therefore, the accuracy of the proposed method on test data is 43.3%.

4.2 Results

Decision tree and SVM classifier produces better and approximately similar results. The accuracy peaks of each method are shown in bold (See Table 1).

Table 1. Percentage Accuracy of Experiments

	Holdout method	**Test set method**	**Cross validation**
Single-level SVM	22.5	21.10	21.35
Single-level decision tree	40.6	38.6	39.35
Proposed solution with decision tree	60.4	**39.7**	**43.7**
Proposed solution with SVM	**66.29**	39.5	43.38

4.3 Comparison with Benchmark

The problem of fake news detection has been explored by Wang et. Al [39] where he classifies text as well as the corresponding metadata based on a hybrid Convolutional Neural Networks framework that integrates text and meta-data. The authors have only used the test dataset for evaluation of their approach and not considered the hold-out or cross validation methods. Their analysis shows that best results are obtained when using all attributes including text.

Table 2. Comparison with Benchmark

Model	**Features**	**Test accuracy %**
Hybrid CNN Wang et. al [39]	Text + speakers profile	27.4
Single level SVM	Speakers profile	21.10
Proposed solution with SVM	Credit history	39.5
Proposed solution with Decision tree	Credit history	39.7

5. Conclusion

In this paper we propose a multi-level fake news detection method based on supervised learning. Input dataset consists of multi-labels and speakers profile (name, designation, party affiliation, credit history etc.). Multi-level models are trained using machine learning techniques which are tested with SVM and decision tree classifiers. Evaluations have been made using hold-out, test dataset and cross validation approaches. The results are compared with benchmark techniques and the results demonstrate higher accuracy in comparison. The proposed method outperforms the benchmark with an accuracy of 39.5%. Our experiments indicate that the profile of the source of information contributes significantly in fake news detection. The proposed method can also be extended to classify text and metadata together. Another direction of future research is to use ensembles instead of simple classifiers.

6. REFERENCES

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

[2] Chen, Y., Conroy, N. J. and Rubin, V. L. News in an online world: The need for an “automatic crap detector”. *Proceedings of the Association for Information Science and Technology*, 52, 1 (2015), 1-4.

[3] Rubin, V. L. Deception detection and rumor debunking for social media. *The SAGE Handbook of Social Media Research Methods* (2017), 342-363.

[4] Jin, Z., Cao, J., Jiang, Y.-G. and Zhang, Y. *News credibility evaluation on microblog with a hierarchical propagation model*. IEEE, 2014.

[5] Shen, H., Ma, F., Zhang, X., Zong, L., Liu, X. and Liang, W. Discovering social spammers from multiple views. *Neurocomputing*, 225 (2017), 49-57.

[6] Shu, K., Sliva, A., Wang, S., Tang, J. and Liu, H. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19, 1 (2017), 22-36.

[7] Singhania, S., Fernandez, N. and Rao, S. *3HAN: A Deep Neural Network for Fake News Detection*. Springer International Publishing, 2017.

[8] Dewang, R. K. and Singh, A. K. Identification of Fake Reviews Using New Set of Lexical and Syntactic Features. In *Proceedings of the Proceedings of the Sixth International Conference on Computer and Communication Technology 2015* (Allahabad, India, 2015). ACM.

[9] Granik, M. and Mesyura, V. *Fake news detection using naive Bayes classifier*. IEEE, 2017.

[10] Thu, P. P. and New, N. *Implementation of emotional features on satire detection*. IEEE, 2017.

[11] Ahmed, H., Traore, I. and Saad, S. *Detection of online fake news using N-gram analysis and machine learning techniques*. Springer, 2017.

[12] Thorne, J., Chen, M., Myrianthous, G., Pu, J., Wang, X. and Vlachos, A. *Fake news stance detection using stacked ensemble of classifiers*. 2017.

[13] Vosoughi, S., Mohsenvand, M. N. and Roy, D. Rumor gauge: Predicting the veracity of rumors on Twitter. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 11, 4 (2017), 50.

[14] Ruchansky, N., Seo, S. and Liu, Y. *Csi: A hybrid deep model for fake news detection*. ACM, 2017.

[15] Shin, J., Jian, L., Driscoll, K. and Bar, F. The diffusion of misinformation on social media: Temporal pattern, message, and source. *Computers in Human Behavior*, 83 (2018), 278-287.

[16] Boididou, C., Papadopoulos, S., Apostolidis, L. and Kompatsiaris, Y. Learning to Detect Misleading Content on Twitter. In *Proceedings of the Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval* (Bucharest, Romania, 2017). ACM.

[17] Boididou, C., Papadopoulos, S., Zampoglou, M., Apostolidis, L., Papadopoulou, O. and Kompatsiaris, Y. Detection and visualization of misleading content on Twitter. *International Journal of Multimedia Information Retrieval*, 7, 1 (2018), 71-86.

[18] Zhang, Q., Yilmaz, E. and Liang, S. Ranking-based Method for News Stance Detection. In *Proceedings of the Companion Proceedings of the The Web Conference 2018* (Lyon, France, 2018). International World Wide Web Conferences Steering Committee.

[19] Bhatt, G., Sharma, A., Sharma, S., Nagpal, A., Raman, B. and Mittal, A. Combining Neural, Statistical and External Features for Fake News Stance Identification. In *Proceedings of the Companion Proceedings of the The Web Conference 2018* (Lyon, France, 2018). International World Wide Web Conferences Steering Committee.

[20] Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L. and Gao, J. EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. In *Proceedings of the Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery &#38; Data Mining* (London, United Kingdom, 2018). ACM,.

[21] Jin, Z., Cao, J., Zhang, Y., Zhou, J. and Tian, Q. Novel visual and statistical image features for microblogs news verification. *IEEE transactions on multimedia*, 19, 3 (2017), 598-608.

[22] Bachenko, J., Fitzpatrick, E. and Schonwetter, M. *Verification and implementation of language-based deception indicators in civil and criminal narratives*. Association for Computational Linguistics, 2008.

[23] Larcker, D. F. and Zakolyukina, A. A. Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, 50, 2 (2012), 495-540.

[24] Feng, V. W. and Hirst, G. *Detecting deceptive opinions with profile compatibility*. 2013.

[25] Long, Y., Lu, Q., Xiang, R., Li, M. and Huang, C.-R. *Fake news detection through multi-perspective speaker profiles*. 2017.

[26] Horne, B. D. and Adali, S. This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. *arXiv preprint arXiv:1703.09398* (2017).

[27] Gupta, A., Lamba, H., Kumaraguru, P. and Joshi, A. *Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy*. ACM, 2013.

[28] Yang, F., Liu, Y., Yu, X. and Yang, M. *Automatic detection of rumor on Sina Weibo*. ACM, 2012.

[29] Tschiatschek, S., Singla, A., Rodriguez, M. G., Merchant, A. and Krause, A. Fake News Detection in Social Networks via Crowd Signals. In *Proceedings of the Companion Proceedings of the The Web Conference 2018* (Lyon, France, 2018). International World Wide Web Conferences Steering Committee.

[30] Antoniadis, S., Litou, I. and Kalogeraki, V. *A model for identifying misinformation in online social networks*. Springer, 2015.

[31] Kim, J., Tabibian, B., Oh, A., Sch, B., #246, lkopf and Gomez-Rodriguez, M. Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation. In *Proceedings of the Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (Marina Del Rey, CA, USA, 2018). ACM.

[32] Shao, C., Ciampaglia, G. L., Flammini, A. and Menczer, F. Hoaxy: A Platform for Tracking Online Misinformation. In *Proceedings of the Proceedings of the 25th International Conference Companion on World Wide Web* (Canada, 2016). International World Wide Web Conferences Steering Committee.

[33] Jain, S., Sharma, V. and Kaushal, R. *Towards automated real-time detection of misinformation on Twitter*. IEEE, 2016.

[34] Zhao, J., Cao, N., Wen, Z., Song, Y., Lin, Y.-R. and Collins, C. # FluxFlow: Visual analysis of anomalous information spreading on social media. *IEEE transactions on visualization and computer graphics*, 20, 12 (2014), 1773-1782.

[35] Kumar, K. K. and Geethakumari, G. Detecting misinformation in online social networks using cognitive psychology. *Human-centric Computing and Information Sciences*, 4, 1 (2014), 14.

[36] Jang, S. M., Geng, T., Queenie Li, J.-Y., Xia, R., Huang, C.-T., Kim, H. and Tang, J. A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis. *Computers in Human Behavior*, 84 (2018/07/01/ 2018), 103-113.

[37] Castillo, C., Mendoza, M. and Poblete, B. *Information credibility on twitter*. ACM, 2011.

[38] Bourgonje, P., Schneider, J. M. and Rehm, G. *From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles*. 2017.

[39] Wang, W. Y. *" Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection*. 2017.