A Sensitive Stylistic Approach to Identify Fake News on Social Networking

Nicollas R. de Oliveira, Dianne S. V. Medeiros and Diogo M. F. Mattos, *Member, IEEE*

***Abstract*—Human inefficiency to distinguish between true and false facts poses fake news as a threat to logical truth, which deteriorates democracy, journalism, and credibility in governmental institutions. In this letter, we propose a computational-stylistic analysis based on natural language processing, efficiently applying machine learning algorithms to detect fake news in texts extracted from social media. The analysis considers news from Twitter, from which approximately 33,000 tweets were collected, assorted between real and proven false. In assessing the quality of detection, 86% accuracy, and 94% precision stand out even employing a dimensional reduction to one-sixth of the number of original features. Our approach introduces a minimum overhead, while it has the potential of providing a high confidence index on discriminating fake from real news.**

***Index Terms*—Fake News Detection, One-Class SVM.**

I. Introduction

THE potential for dissemination, acceptance, and destruction of fake news poses them as one of the greatest threats to the concept of logical truth. Since the popularization of the spread of fake news, there has been a growing joint effort by the academic community to research and develop approaches capable of analyzing, detecting and intervening in the performance of misleading content. Scientific evidence has already revealed the human vulnerability in distinguishing real from false facts, while the capacity for human differentiation reduces up to a random probability of approximately 54% correctness. Furthermore, the fight against fake news renders the social network and data consumption problems inseparable. By spreading malicious content, a user is wasting network and processing resources and undermining the credibility of the service provided. In turn, fake news hamper the Quality of Trust (QoT) applied to news distribution, that is, how much a user trusts in the content of a particular source [1]–[6].

In this letter, we characterize the recent phenomenon of fake news and propose a new combination of techniques and algorithms capable of saving time and cognitive effort when identifying false news on social networks. At the heart of the contribution is the presentation of a stylistic-computational analysis, based on Natural Language Processing (NLP), efficiently applying machine learning algorithms for detecting fake news in texts extracted from social media. The analysis considers news extracted from Twitter, from which approximately 33 thousand tweets were collected, assorted between real and proven false. In the end, the proposed approach's effectiveness is measured using information retrieval metrics, such as sensitivity, precision, and accuracy. Previous works that aim to identify fake content focus on linguistic characteristics, grammatical resource, word pattern, term count, and frequency of certain expressions [7]. Other approaches differ by applying network analysis, which considers other information instead of inspecting only the content. Such an approach generally performs well but depends on a large number of related meta-data [8]. Given these works, our approach emerges, introducing minimal overhead, while allowing the discrimination of false news with high precision.

Manuscript received February 15, 2020. This work was supported in part by CNPq, CAPES, RNP, FAPERJ and FAPESP.

Nicollas R. de Oliveira, Dianne S. V. Medeiros and Diogo M. F. Mattos are with the Graduate Program in Electrical and Telecommunications Engineering, at Universidade Federal Fluminense (UFF), Niterói, Brazil (e-mail: {nicollas_rodrigues, dianneshcerly, diogo_mattos}@id.uff.br).

II. The Branches of Fake News

Fake news can be divided into three aspects: (i) those of a purely fraudulent nature, whose intention is to deceive the reader by leading him to confusion; (ii) rumors, which are information with dubious truth but publicly accepted; (iii) and those with humorous character using sarcasm and irony to produce parodies and satires [9]. The growth in the spread of fake news is the result of the expansion of social networks that speed up the spread of rumors, satires, and wrong information. Social network users tend to rely on fake news because there is currently public disbelief concerning traditional media and because such news is often shared by friends or confirm prior knowledge. Moreover, it is hard to discriminate fake news while users are overwhelmed with misleading information that is received continuously [10].

Rubin *et al.* argue that in the construction of any database, known as *corpus*, which is composed of fake news, the following must be observed. (i) Considering both false and true instances allows any predictive method applied to the database to consider characteristic patterns of each type of news. (ii) Information should preferably be in textual format. (iii) The homogeneity of the news in terms of size and (iv) writing style must be considered, avoiding very different instances whenever possible. Equally, there is a concern with (v) the influence of the manner of news delivery in the reader's inclination towards believing in its content. Also, (vi) the acquisition of news from the same time interval is a crucial factor. (vii) It is also advisable to attend to some pragmatic aspects, such as copyright costs, availability, ease of obtaining, and writer's privacy. One should not neglect the (viii) language and (ix) culture to which the collected data belong. The translation may imply ambiguities or misinterpretations, negatively affecting the efficiency of detection processes [9], [11].

III. The Proposed Approach

The identification of fake news can be carried out manually, e.g., by professionals in journalism, being the most commonly used approach. Nevertheless, the focus of this proposal is on automatic identification based computational methods. Within this automatic approaches, there are four distinct action fronts:

(i) the automatic proof of logical statements through facts already known; (ii) the analysis of news spread on social networks; (iii) the analysis of the profile of users who share the news; or (iv) the natural language processing for knowledge extraction in a stylistic-computational approach [2]. The focus on the stylistic-computational approach based on natural language processing justifies the fact that the users' consumption of data on social networks is restricted to information that reaches the end-user. The end-user does not access content dissemination statistics nor reputation models. Our proposed approach first implements a sequence of preparation tasks, including extracting data from Twitter, textual treatment with NLP and reducing complexity. Subsequently, three methodologies for classifying the veracity of the news are tested and compared against each other.

A. Data Collection Process and Database

Faced with negative consequences caused by an eventual erroneous data collection, such as the particularization of the analysis or the obtaining of dissonant results, the nine conditions suggested by Rubin *et al.*, presented in Section II, are adopted as a guideline for the formation of a corpus of fake news [9]. Therefore, the database's composition includes both real and fake news, collected from specific Twitter accounts. In order to obtain this information, a script was developed in `Python` using *Twitter's API*. Accessing this API, using developer credentials, allows the continuous extraction of textual content from tweets from any open profile on the social network. However, besides the time limitations also faced by Barreto *et al.*, related to the maximum number of requests per time window of 15 minutes [12], there is also a limitation on the number of historical tweets that can be collected. Obtaining tweets is restricted to a period of up to a maximum of two months. Due to data recovery restrictions, one solution is to diversify sources for searching for real news, collecting tweets from other journalistic sources. The choice of news media profiles as a source of real content is based on the premise that these profiles are less likely to share the content of dubious origin than individual user accounts. Similarly, the collection of tweets proven to be false is also promoted, previously verified by journalists, and made available by the profile "Boatos.org". The database counts 33.000 tweets.

B. Data Representation and Dimensional Reduction

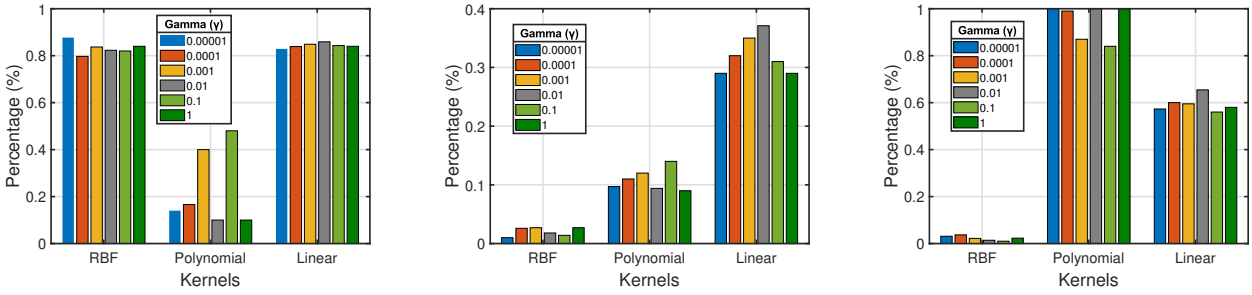
Cleaning and shaping data in the *corpus* are the initial procedures in any information extraction process and are performed using the following natural language processing techniques: tokenization, punctuation and special character removal, elimination of stop words, spelling correction, recognition of named entities and stemming. Following the order above, each sentence in the *corpus* is first submitted to tokenization, responsible for transforming a contiguous phrase into a token list and thus allowing individual manipulation. Each token is seen as an instance of a string [13]. Subsequently, various spelling features such as punctuation, and special characters, are removed from each token. Next, we eliminate the stop words, considered the most frequent words, such as connectors, articles, and pronouns. This particular task derived from the principle that the higher the frequency of a word in the *corpus*, the less relevant information the word has. In the specific case of tweets, it is equally important to remove specific words from Twitter, such as eventual hashtags, referrals, or links for sharing. In the next step, spell checking occurs by comparing the token with its closest correspondent in the dictionary [14]. Another necessary procedure is the recognition of named entities, mainly proper names, and their subsequent removal. To this end, we need to calculate the Levenshtein distance, the minimum number of operations required to transform a name in the database into another name in a dictionary. Stemming reduces inflected or derived words to their radicals, eliminating possible variants or plurals. Once these techniques are performed, each sentence abandons its purely contiguous textual form to be expressed as a list of the remaining words [15].

Even if accurately standardized, each sentence is not mathematically operable, as it is still composed of word radicals and not measurable values. To obtain a numerical representation, we use the vector space model. This model defines that texts can be interpreted as a vector space of words, in which each word can be represented in different patterns. Among the possibilities of vector representation, it is worth mentioning that in this proposal, we have adopted representation by frequency *tf-idf*, a statistical measure that indicates the importance of a word in a sentence in relation to the database [16].

An important point to be understood is that the dimension of the vector is linked to the remaining number of distinct words in the entire database since several of them are removed during the process. The words kept in the sentence carry meaning, revealing themselves to be essential for understanding the central idea of the text. However, when using an extensive database, it is inevitable to deal with very long word vectors. In particular, immediately after the vectorization process, 13,716 distinct words are identified in the database. Thus, it is convenient to apply some dimensionality reduction technique – Principal Component Analysis (PCA), Latent Semantic Analysis (LSA), Multidimensional Scale, or FastMap – since these techniques can find less complex vector representations and respect the main characteristics of the original representations [17]. However, due to the sparse nature of the feature vectors, the LSA is chosen. Unlike the PCA, the LSA technique does not centralize the data before calculating the singular value decomposition [18]. By preserving about 70% of the original characteristics' variance, the LSA technique reduced the number of characteristics to 2,000, corresponding to a reduction of 85.4% of the memory space.

C. Fake News Detection Methodologies

Our proposal use three different methodologies to classify the legitimacy of news. The first two methodologies implement different combinations of machine learning algorithms, both for unsupervised clustering and classification, to predict the type of news from the training only on the real news. In parallel, the third methodology expands the detection proposal for a statistical scenario, based on the hypothesis that real and fake news have different probability distributions when



(a) Accuracy. (b) Precision. (c) Sensitivity.

Fig. 1. Results obtained by applying the LSA together with the one-class SVM. The best accuracy is found using the linear core function and $\gamma = 0{,}01$.

considering the module of their representation in the vector space of frequency of words.

1) The Reduction Methodology with Training: The use of supervised algorithms for detecting fake news depends on an extensive database containing both true and fake news. However, this imposes the limitation of having a labeled base with real and fake news. One strategy to overcome the limitation in the number of fake news for classifier training is to learn a single class, such as that based on the one-class support vector machine method (SVM). The one-class SVM is an unsupervised learning algorithm that derives a decision hyperplane for detecting anomalies. New data are classified as similar or different from the training set. In contrast to typical SVM implementations, the one-class takes into account a set of training samples for a single class. Any new sample that does not fit the decision surface defined by the training set is considered an instance of a new class and, therefore, a sample of fake news [19], [20]. The learning process using one-class SVM relies on the use of the linear kernel function.

The presented results split the database into 90% of the samples for training and 10% for the test. It is worth mentioning that the use of one-class SVM as a classifier model imposes a singular requirement: the training step must be performed using only the real news instances, the single class. Figure 1 shows the execution of the single class classifier for different core functions and varying the γ coefficient that determines how far distant samples influence the SVM hyperplane calculation. It is worth mentioning that the best accuracy and precision for the fake news class are obtained using the linear core function and $\gamma = 0.01$.

2) The Matrix Transformation Methodology: After the dimensional reduction, the D matrix holds all the samples in the database and accounts for a number of columns j equal to 2,000, equivalent to the number of features remaining in each sample. Assuming the existence of a transformation matrix T, which dimensions are $j \times i$, such that i is much smaller than j, we achieve the projection of the matrix D over T, resulting in a matrix S even more compact. In practice, obtaining the transformed matrix consists of:

$$D_{k \times i} \times T_{i \times j} = S_{k \times j}. \quad (1)$$

A possible transformation matrix is formed by the centroids that best describe the data in the original data before feature reduction, M matrix. The k-means algorithm is a heuristic capable of partitioning data into k clusters by minimizing the sum of the squares of the distances in each group. Although robust, fast, and easy to understand, k-means, as well as other grouping techniques, *k-medoids* and Expectation-Maximization Algorithm, are subject to the disadvantage of indeterminacy as to the appropriate number k groups. In order to circumvent this indeterminacy, two methods, *Elbow* and *Silhouette*, are used to previously analyze the conformity of the data to different amounts of clusters and, thus, obtain a suitable result for the data. Although both are intended to solve the same problem, each method has different calculations and criteria for identifying k. In particular, *Elbow* measures the clusters' compactness by establishing a relationship between the number of groups and their influence on the total variation of data within the group. Graphically, the best value of k is identified by looking at the point at which the curve gain declines sharply, remaining approximately constant after that. Similarly, the *Silhouette* method measures the quality of a cluster. The ideal number of k groups is one that maximizes the average silhouette over a range of possible values for k.

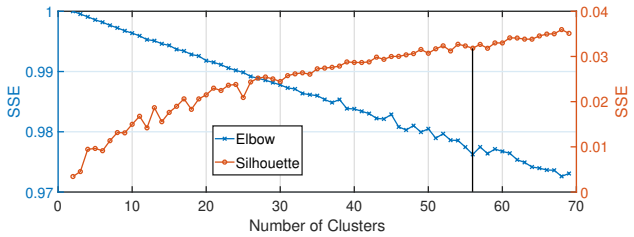


Fig. 2. Methods for determining the best number of clusters k (*clusters*). For both methods, the normalized mean square error values are shown. For the values tested for k, 56 clusters proved to be a local minimum for *Elbow* and a local maximum for *Silhouette*.

In practice, finding the appropriate number of clusters implies analyzing the Figure 2, obtained for the news set used in this letter, simultaneously. The goal is to identify which value of k minimizes data dispersion in each group, Elbow's curve while maximizing the distance between data grouped in different groups, Silhouette's curve. Under these circumstances, the best way to apply *k-means* to the textual data used is to set k equal to 56, as it is a local minimum in Elbow's curve and a local maximum in Silhouette's curve. Although not the optimal global value, $k = 56$ allows defining 56 centroids, forming the $T_{i \times j}$ matrix that compacts the 2,000 characteristics provided by the LSA in a set of 56 artificial dimensions. On the $S_{k \times j}$ matrix, we apply the one-class SVM classifier. The approach of using the transformation for the vector space defined by the k centroids aims at concentrating

the data because even after the application of the LSA, the D matrix is still sparse.

3) The Radial Limit Methodology: The radial limit approach starts from the D matrix, with 2,000 features, resulting from the LSA's dimensional reduction over the M matrix, and implements the Equation 2. Equation 2 defines that the sum of all values x_{ij}^2, belonging to a sample, line i of the matrix D, results in a random variable R_i^2, which can be interpreted as the square of the radius of a hypersphere that contains the real news. Expanding this logic to the whole matrix, but grouping by type of news, false and true, a set of random variables specific to each type of news is obtained.

$$R_i^2 = x_{i0}^2 + x_{i1}^2 + x_{i2}^2 + ... + x_{ik}^2 \quad (2)$$

Figure 3 shows the probability density function, and we assume the hypothesis of inequality between the averages regarding the sets of random variables of false and legitimate news. The proof of this assumption, refuting the null hypothesis, with a 95% confidence interval, arises through the application of Welch's t-test, an adaptation of Student's t-test for sample sets with different variance or sizes. Exploring this statistical relationship between the news sets and the fact that Student's t distribution approaches the normal distribution for high degrees of freedom, we develop a probabilistic news classifier based on the samples of the random variable R^2.

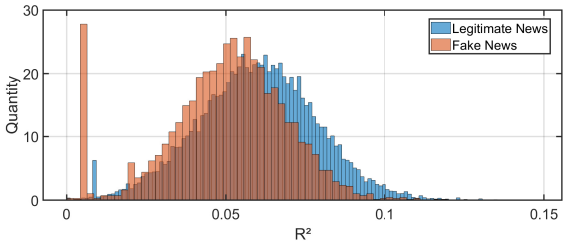


Fig. 3. Statistical behavior of the entire *corpus* divided according to type. There is a noticeable vertical shift in the amounts by R^2 as well as in the accumulated probabilities depending on the news.

IV. Results Evaluation

The detection quality assessment process for each proposed methodology is based on the calculation of information retrieval metrics such as accuracy, precision, and sensitivity. All algorithms and computational techniques were performed on a personal computer with an Intel Core i7 4770 processor, with 24 GB of RAM and 1 TB hard disk. Figure 4 depicts the best results obtained in each methodology, specifying their performance in each metric.

In the results of the reduction methodology with training, the proposal demonstrates a more homogeneous performance among the metrics, standing out mainly for the high accuracy and more significant percentage of sensitivity among the three methodologies. In the results referring to the matrix transformation methodology using the linear core function, there is a clear predominance in the ability to classify news as being fake. On the other hand, it has low percentages of accuracy and sensitivity, which are possibly the result of the loss of features, important in the discrimination of news, imposed by two levels of dimensional reduction - LSA and *k-means*. Analogously, the radial limit methodology has the same precise character in

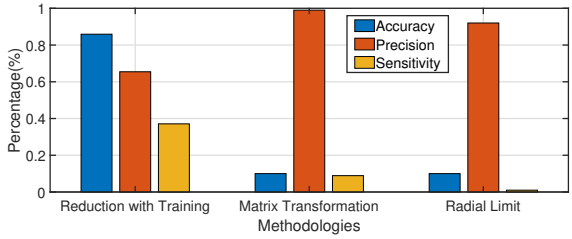


Fig. 4. A comparison of the methodologies reveals a different but slightly complementary behavior in the levels of information retrieval.

identifying fake news, although there is a depreciation in its sensitivity levels. As an additional assessment, we check the Receiver Operating Characteristic Curve, the ROC, for each methodology expressed in Figure 5. The ROC represents the ratio between the true positive rate and the true negative rate, for various thresholds. The curve graphically describes the performance of a classification model. Briefly, the larger the area under the curve, the better the model's performance. The result obtained by the matrix transformation, an area under the curve of 0.82, derivates from the *k-means* clustering that substantially reduced the dimension of the data. The clustering procedure allows the one-class SVM to define a hypersurface more adjusted to the data. The result of the radial limit is based on the hypothesis test performed *a priori*, which shows that the average of the sum of the square values of the frequencies between true and fake news is different. The radial limit shows good results because fake news tends to have a higher frequency of use of words, Figure 3, while real news tends to vary words, using more vocabulary. As the difference in distributions is small, the classifier based on this distribution has an average result, with an area under the curve of 0.63.

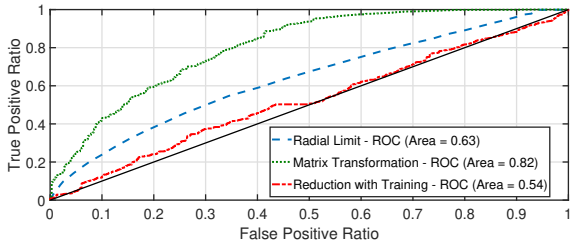


Fig. 5. The ROC curves reflect the performance of a binary classifier. The matrix transformation presents the best performance.

V. Conclusion

This letter presented a stylistic-computational analysis, based on natural language processing, efficiently applying unsupervised learning algorithms, such as one-class SVM, in detecting fake news in texts extracted from social media. We propose to apply to original data both dimensionality reduction technique, through latent semantic analysis (LSA), and data compaction through our proposed methodologies. After a reduction of more than 85% in the number of characteristics, three different news classification methodologies were implemented – two employing cascading or unique configurations of learning algorithms and the other statistically evaluating the difference between the types of news. In the process of assessing the quality of the detection of the methodologies, an accuracy of 86% and a precision of 94% stands out.

REFERENCES

[1] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[2] X. Zhou and R. Zafarani, "Fake news: A survey of research, detection methods, and opportunities," *arXiv preprint arXiv:1812.00315*, 2018.

[3] W. Y. Wang, ""Liar, liar pants on fire": A new benchmark dataset for fake news detection," in *Annual Meeting of the Association for Computational Linguistics - ACL 2017*, 2017.

[4] V. L. Rubin, "On deception and deception detection: Content analysis of computer-mediated stated beliefs," in *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem-Volume 47*. American Society for Information Science, 2010, p. 32.

[5] V. Rubin, N. Conroy, Y. Chen, and S. Cornwell, "Fake news or truth? using satirical cues to detect potentially misleading news," in *Proceedings of the second workshop on computational approaches to deception detection*, 2016, pp. 7–17.

[6] G. Liu, Y. Wang, and M. A. Orgun, "Quality of trust for social trust path selection in complex social networks," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems, 2010, pp. 1575–1576.

[7] Y. Chen, N. J. Conroy, and V. L. Rubin, "Misleading online content: Recognizing clickbait as false news," in *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*, 2015, pp. 15–19.

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

[9] V. L. Rubin, Y. Chen, and N. J. Conroy, "Deception detection for news: three types of fakes," in *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*. American Society for Information Science, 2015, p. 83.

[10] M. Hardalov, I. Koychev, and P. Nakov, "In search of credible news," in *International Conference on Artificial Intelligence: Methodology, Systems, and Applications*. Springer, 2016, pp. 172–180.

[11] V. Rubin, "Pragmatic and cultural considerations for deception detection in asian languages," *ACM Transactions on Asian Language Information Processing*, vol. 13, 06 2014.

[12] H. F. Barreto, M. E. M. Campista, and L. H. M. Costa, "Spammers no twitter: Quando contatos deixam de ser bem-vindos," in *Workshop de Redes P2P, Dinâmicas, Sociais e Orientadas a Conteúdo (Wp2p+ 2014) - SBRC 2014*, vol. 1, 2014, pp. 23–36.

[13] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, "The stanford corenlp natural language processing toolkit," in *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, 2014, pp. 55–60.

[14] N. R. de Oliveira, L. H. Reis, N. C. Fernandes, C. A. M. Bastos, D. S. V. de Medeiros, and D. M. F. Mattos, "Natural language processing characterization of recurring calls in public security services," in *Proceedings of the 2020 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 2020, pp. 1009–1013.

[15] R. Oshikawa, J. Qian, and W. Y. Wang, "A survey on natural language processing for fake news detection," in *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC)*, 2020, pp. 6086–6093.

[16] F. Raposo, R. Ribeiro, and D. M. de Matos, "On the application of generic summarization algorithms to music," *IEEE Signal Processing Letters*, vol. 22, no. 1, pp. 26–30, Jan 2015.

[17] Q. Du and J. E. Fowler, "Low-complexity principal component analysis for hyperspectral image compression," *The International Journal of High Performance Computing Applications*, vol. 22, no. 4, pp. 438–448, 2008.

[18] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM review*, vol. 53, no. 2, pp. 217–288, 2011.

[19] R. Perdisci, G. Gu, and W. Lee, "Using an ensemble of one-class svm classifiers to harden payload-based anomaly detection systems," in *Sixth International Conference on Data Mining*, 2006, pp. 488–498.

[20] S. Gaonkar, S. Itagi, R. Chalippatt, A. Gaonkar, S. Aswale, and P. Shetgaonkar, "Detection of online fake news : A survey," in *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*, 2019, pp. 1–6.