Classifier for fake news detection and Topical Domain of News Articles

Kana Tsoplefack William[1]

[1]*University Duisburg-Essen, Germany*

Abstract

Digitization has resulted in a plethora of new methods to read articles or excerpts online using smartphones or tablets. Nowadays, everything is available online, and dealing with false information has grown increasingly. Online newspapers cannot check the veracity of all social media posts, and in order to combat the spread of fake news, machine learning algorithms might be beneficial in classifying articles based on labels provided by experts. This paper will present relevant algorithms and their outcomes.

Keywords

Fake News, Machine learning

1. Introduction

Today, spreading article content is as simple as clicking a link, and due to the rapid circulation of content, fake news articles follow the same pattern. With this expansion, it has become more vital to provide users with a concise overview of which class the entire article may belong to, whether such content is misleading, false, or true. To that end, many fact-checking websites are in high demand, with their already classified content increasing the amount of work to be done on each article with experts appropriate for their areas of expertise. Faced with this, machine learning algorithms may be useful in detecting fake news or identifying topics. First, article classification may be useful when there is a deficit of experts, or there are insufficient resources to hire them. Second, understanding the topic of articles may aid in the hiring process by providing companies with an estimate of how many specialists they will need to classify their articles. A basic description of applying machine learning to classify articles and the results gained will be discussed in this study.

2. Related Work

Many studies have been developed in the fields of classification using machine learning. In topics such as hate speech and fake news, [1] presents different classification methods used to detect hate speech using certain features extraction techniques: simple surface features such as Bag of Word (BoW) or Term Frequency-Inverse Document Frequency (TFIDF) that show the benefits of using character n-grams and words n-grams. [2] and [3] respectively describe how BoW and TFIDF operate. This study will just apply the logic of machine learning algorithms in articles classification.

3. Method

Many machine learning algorithms have been developed to assist humans in classification tasks, and their advances are becoming increasingly accurate. This paper discusses findings obtained by classifying articles for fake news detection and topic identification using two classifiers, Multinomial Naive Bayes and Random Forest.

Naive Bayes and Random Forest were chosen from among all machine learning algorithms. The following are the reasons why these algorithms were preferred to others for this paper: First, both Multinomial Naive Bayes and Random Forest [4] can predict non-binary outcomes. Second, Multinomial Naive Bayes uses term frequency, i.e. the number of times a specific phrase appears in a document. To normalise this term frequency, the raw term frequency is divided by the document length and the maximum likelihood based on the training data to estimate the conditional probability; the term frequency is used. Random Forest, on the other hand, builds many decision trees and then merges them for a more precise and consistent prediction.

4. Dataset

The dataset is downloaded from Zenodo [5]. Two datasets have been labelled: the first is aimed at fake news detection and contains 900 articles in the training set and 364 articles in the testing set. The training set contains labels such as false, partially false, true, and others when the article couldn't be categorised other mentioned early. The second section is for topic identification and includes a training set of 318 articles and a testing set of 137 articles. There are also many label values in the training set, such as crime, climate, economy, education, elections, and health. The dataset used in this paper was obtained through [6, 7, 8]. A preprocessing step is necessary using the NLK library [9]. The dataset was collected using the approach mentioned in [10, 11]. The four different categories of data are mentioned in [12, 13]. The cleaning process involves the removal of email addresses, hyperlinks, numerals, and special characters. Second, words will be lemmatised in order to determine their dictionary root.

4.1. Implementation

Most machine learning algorithms are simple to implement in many programming languages, and Python was used for this paper's task due to familiarity. The implementation of many

machine learning algorithms has become much easier with the help of the Scikit-learn library [14]. As features parameters, Scikit-learn default's function is used without any additional needs such as n_estimators (default value 100) for Random forest classifier.

5. Results and Discussion

5.1. Results

The above-mentioned machine learning algorithms were tested for fake news detection and identifying the topic of given articles. Two approaches were tested in order to obtain a more comprehensive view of classifiers and noises that could influence the results. As a result of these factors, two distinct outcomes were obtained, as shown in 1 and 2. The results in percent presented in the Table is the F1 score which is the weighted average of the precision and recall. Table 1 shows the results of the training dataset based on the labels provided, implying that the algorithms must determine the real label value. For example, if an expert classifies the article as "mostly false," the algorithm must return this value. Table 2, on the other hand, will put label values such as false, mostly false (true, mostly true) in the same category. For example, two articles A and B with labels true and mostly true will both obtain the value "true."

Table 1
Classification Results 1

Classifier	Multinomial Naive Bayes	Random Forest
accuracy	0.18	0.18

Table 2
Classification Results 2

Classifier	Multinomial Naive Bayes	Random Forest
accuracy	0.59	0.73

Table 3 displays F1 Scores for category classification, which aims to identify if an article belongs to a category such as "politic" or "health" using the algorithms mentioned in the preceding section.

Table 3
Category Classification

Classifier	Multinomial Naive Bayes	Random Forest
accuracy	0.81	0.64

6. Discussion

As mentioned in the preceding section, the classifier used here to detect fake news is insufficient when the label is a non-binary outcome, but it can be exploited when a group is established,

such as false and partially false in the same group with the value "false". This statement can be supported by the result 2. The task of topic detection yielded significantly better results with both classifiers, particularly Multinomial Naive Bayes, indicating a better outcome than the second, demonstrating that by adjusting its parameters, this classifier can perform better.

7. Conclusion

In this paper, the advantages of machine learning algorithms have been presented to categorize articles. Despite the lack of precision, these algorithms allow positive outcomes to be obtained in this paper. This work could help fight against the classification of articles and could help online newspapers in this challenge. Machine learning techniques with the support of innovative human insights are getting more effective in text labeling so that they are now being used to classify text and this case, articles. Other features, such as ngram [15, 16] combined with machine learning algorithms, have also demonstrated advantages in classification tasks. Alternative machine learning techniques, such as deep learning, may also be useful in fake news detection and topic identification.

References

[1] A. Schmidt, M. Wiegand, A survey on hate speech detection using natural language processing, in: Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media, Association for Computational Linguistics, Valencia, Spain, 2017, pp. 1–10. URL: https://www.aclweb.org/anthology/W17-1101. doi:10.18653/v1/W17-1101.

[2] Y. Zhang, R. Jin, Z.-H. Zhou, Understanding bag-of-words model: a statistical framework, International Journal of Machine Learning and Cybernetics 1 (2010) 43–52.

[3] T. Joachims, A probabilistic analysis of the rocchio algorithm with tfidf for text categorization, in: Proceedings of the Fourteenth International Conference on Machine Learning, ICML '97, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1997, p. 143–151.

[4] U. Parida, M. Nayak, A. K. Nayak, News text categorization using random forest and naïve bayes, in: 2021 1st Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology(ODICON), 2021, pp. 1–4. doi:10.1109/ODICON50556.2021.9428925.

[5] G. K. Shahi, J. M. Struß, T. Mandl, Task 3: Fake news detection at CLEF-2021 CheckThat!, 2021. URL: https://doi.org/10.5281/zenodo.4714517. doi:10.5281/zenodo.4714517.

[6] P. Nakov, G. Da San Martino, T. Elsayed, A. Barrón-Cedeño, R. Míguez, S. Shaar, F. Alam, F. Haouari, M. Hasanain, N. Babulkov, A. Nikolov, G. K. Shahi, J. M. Struß, T. Mandl, The CLEF-2021 CheckThat! lab on detecting check-worthy claims, previously fact-checked claims, and fake news, in: Proceedings of the 43rd European Conference on Information Retrieval, ECIR '21, Lucca, Italy, 2021, pp. 639–649. URL: https://link.springer.com/chapter/10.1007/978-3-030-72240-1_75.

[7] P. Nakov, G. Da San Martino, T. Elsayed, A. Barrón-Cedeño, R. Míguez, S. Shaar, F. Alam, F. Haouari, M. Hasanain, N. Babulkov, A. Nikolov, G. K. Shahi, J. M. Struß, T. Mandl, S. Modha, M. Kutlu, Y. S. Kartal, Overview of the CLEF-2021 CheckThat! lab on detecting check-worthy claims, previously fact-checked claims, and fake news", year = 2021, booktitle = "proceedings of the 12th international conference of the clef association: Information access evaluation meets multiliguality, multimodality, and visualization", series = CLEF '2021, address = Bucharest, Romania (online),, ????

[8] G. K. Shahi, J. M. Struß, T. Mandl, Overview of the CLEF-2021 CheckThat! lab task 3 on fake news detection, in: Working Notes of CLEF 2021—Conference and Labs of the Evaluation Forum, CLEF '2021, Bucharest, Romania (online), 2021.

[9] S. Bird, E. Klein, E. Loper, Natural language processing with Python: analyzing text with the natural language toolkit, " O'Reilly Media, Inc.", 2009.

[10] G. K. Shahi, Amused: An annotation framework of multi-modal social media data, 2020. arXiv:2010.00502.

[11] G. K. Shahi, D. Röchert, S. Stieglitz, Covid ct: Analysis and detection of different conspiracy theories on youtube in the context of covid-19 (????).

[12] G. K. Shahi, A. Dirkson, T. A. Majchrzak, An exploratory study of covid-19 misinformation on twitter, Online social networks and media (2021) 100104.

[13] G. K. Shahi, T. A. Majchrzak, Exploring the Spread of COVID-19 Misinformation on Twitter, Technical Report, EasyChair, 2021.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: Machine learning in python, Journal of machine learning research 12 (2011) 2825–2830.

[15] S. Malmasi, M. Zampieri, Detecting hate speech in social media, arXiv preprint arXiv:1712.06427 (2017).

[16] A. Gaydhani, V. Doma, S. Kendre, L. Bhagwat, Detecting hate speech and offensive language on twitter using machine learning: An n-gram and tfidf based approach, arXiv preprint arXiv:1809.08651 (2018).