Reliable or not? An automated classification of webpages about early childhood vaccination using supervised machine learning

Meppelink, C.S.; Hendriks, H.; Trilling, D.; van Weert, J.C.M.; Shao, A.; Smit, E.S.

DOI
10.1016/j.pec.2020.11.013

Publication date
2021

Document Version
Final published version

Published in
Patient Education and Counseling

Patient Education and Counseling 104 (2021) 1460–1466

Contents lists available at ScienceDirect

Patient Education and Counseling

journal homepage: www.elsevier.com/locate/pateducou

Reliable or not? An automated classification of webpages about early childhood vaccination using supervised machine learning

Corine S. Meppelink[a,*], Hanneke Hendriks[a], Damian Trilling[a], Julia C.M. van Weert[a], Anqi Shao[a,b], Eline S. Smit[a]

[a] Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, the Netherlands
[b] Life Sciences Communication, University of Wisconsin-Madison, United States

ARTICLE INFO

Article history:
Received 1 May 2020
Received in revised form 2 October 2020
Accepted 10 November 2020

Keywords:
Supervised machine learning
Consumer health information
Vaccination
Misinformation
Reliability

ABSTRACT

Objective: To investigate the applicability of supervised machine learning (SML) to classify health-related webpages as 'reliable' or 'unreliable' in an automated way.
Methods: We collected the textual content of 468 different Dutch webpages about early childhood vaccination. Webpages were manually coded as 'reliable' or 'unreliable' based on their alignment with evidence-based vaccination guidelines. Four SML models were trained on part of the data, whereas the remaining data was used for model testing.
Results: All models appeared to be successful in the automated identification of unreliable (F1 scores: 0.54–0.86) and reliable information (F1 scores: 0.82–0.91). Typical words for unreliable information are 'dr', 'immune system', and 'vaccine damage', whereas 'measles', 'child', and 'immunization rate', were frequent in reliable information. Our best performing model was also successful in terms of out-of-sample prediction, tested on a dataset about HPV vaccination.
Conclusion: Automated classification of online content in terms of reliability, using basic classifiers, performs well and is particularly useful to identify reliable information.
Practice implications: The classifiers can be used as a starting point to develop more complex classifiers, but also warning tools which can help people evaluate the content they encounter online.

1. Introduction

When people in the Netherlands turn to the Internet for health information, they can visit some major health websites which provide information on many different general health related topics (e.g., thuisarts.nl) or specific health related topics, such as vaccines (e.g., rivm.nl/vaccinaties). The content of these websites is based on medical guidelines and checked by medical professionals. Aside from these websites, however, there are many websites that are beyond medical control, and present information that is not always correct [1]. Since reliable and unreliable information can have the same appearance (e.g., a professionally looking website), it is difficult to determine which information is reliable and which is not [2]. As a vast majority of the Dutch population (69%) uses the Internet as a health information source [3], this may cause problems, especially for people with lower levels of health literacy [4]. It would therefore be beneficial if information-seekers could be aided in their judgement of the information they encounter. This study aims to provide input for such aid, by investigating whether the reliability of health websites can be determined in an automated way using supervised machine learning.

Our study focusses on early childhood vaccinations, since stories claiming that vaccines are harmful are widespread online [5–8]. This omnipresence of misinformation coincides with decreasing immunization rates among newborns in many Western countries [9–11]. Many parents consult the Internet for vaccine-related information before making a vaccination decision [12]. However, research has shown that particularly this medium can fuel vaccine hesitancy among parents by disseminating rumors and myths about vaccines [13]. Yet, it should be noted that the group of parents that radically rejects vaccinations (as indicated by very negative attitudes towards vaccination) is relatively small [14,15], while the group of parents that is simply unsure is much bigger. As especially these parents could be influenced by online misinformation, possibly leading to uninformed vaccination decisions, it is important to aid them in distinguishing reliable from unreliable information.

* Corresponding author at: Amsterdam School of Communication Research, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV, Amsterdam, P.O. Box 15793, 1001, NG, Amsterdam, the Netherlands.
E-mail address: c.s.meppelink@uva.nl (C.S. Meppelink).

https://doi.org/10.1016/j.pec.2020.11.013

To evaluate online health information, there are multiple measures in use by the scientific community. For information quality, the Health on the Net (HON) criteria [16] and DISCERN criteria [17] are frequently used. Furthermore, for vaccine-related pages, the Online Vaccination Information Quality Codebook has also been applied [18]. These measures consist of a checklist that health information professionals and consumers can use to develop and evaluate online medical information. To qualify a website as 'high quality', the explicit aim of the information should be mentioned for example, medical credentials of the authors should be stated, or the information provided should be balanced. Although these measures are well-developed and widely used in academia, there are two associated problems.

The first problem refers to the fact that information seekers infrequently use these formal criteria to assess the quality of online information, and use criteria unrelated to the quality instead, such as a professional design [19,20]. In response to this problem, researchers investigated whether this evaluation process could be automated. Boyer and Dolamic [21] used supervised learning to automatically classify medical webpages based on the HON-criteria. Their study, based on 27 websites, showed mixed results. Sondhi, Vydiswaran, and Zhia [22] also applied supervised learning to automatically classify webpages based on the HON-criteria. With a sample of 360 websites, they were able to reach prediction accuracies of over 80%. Kinkead, Allam, and Krauthammer [23] reached the same levels of accuracy (over 80%) with their automated version of the DISCERN instrument. The fact that these authors succeeded in the automated evaluation of the quality of medical websites is promising, and could be of great value to information seekers when integrated into an intervention suitable for implementation in the information seeking process, such as a traffic light system [24].

The second problem that is associated with quality instruments, is that they do not evaluate the content of the webpages, or assess whether the information is scientifically or medically correct. Even if a website fulfills the criteria of balanced information and mentions the medical credentials of the authors, the content can still be inaccurate. Indeed, it has been shown that quality criteria for medical webpages and accuracy of the information presented do not necessarily correlate [25,26].

Therefore, our study focuses on the automated classification of the content of vaccine-related webpages. In our study, we consider information about early childhood vaccines that advises in line with the guidelines provided by the Dutch National Institute for Public Health and the Environment (RIVM) to be reliable information, whereas information that deviates from these guidelines is seen as unreliable. The RIVM is a governmental organization which is responsible for the Dutch immunization program. The decision to incorporate a vaccine in the immunization program is made by the Dutch Ministry of Public Health, based on advice of the Health Council of the Netherlands and scientific research [27]. We therefore consider the RIVM as an expert organization in this field.

The aim of our study is to investigate whether we can classify textual information on Dutch webpages about early childhood vaccination as reliable or unreliable in an automated way. Our study builds on research focusing on automated text analysis, a technique that has been developed and widely applied in computer science, and is rapidly growing in the social sciences domain [28,29]. While using many different tools and techniques (e.g. [30,31]), one of the major advantages of automated text analysis is its scalability, as large numbers of texts can be processed at the same time. Furthermore, once a classifier has been trained successfully and works well, it could be applied to new websites, or other topics or domains.

2. Methods

2.1. Data collection

Data were collected by means of a systematic online search for webpages that discuss early childhood vaccines, using the search engine Google.nl on a cookie-deleted computer. All search terms (e.g., 'vaccinations safe', 'vaccinations unsafe') that were used are presented in Table 1. The list of unique webpages that were identified can be found at https://github.com/corinemeppelink/vaccmisinfo.

Data retrieval took place between April and July 2018. The textual content of 476 webpages was retrieved, and the search stopped when no new webpages could be found. To be included, webpages had to address early childhood vaccinations, whereas webpages about other vaccines (e.g., flu or travel vaccines) were excluded. After duplicate removal, the textual content of 468 unique webpages was manually coded by coder LS. Webpages were coded as 'reliable' if their content advised in line with the RIVM, i.e., recommended to vaccinate children against early childhood diseases, whereas content deviating from those guidelines was coded as 'unreliable'. For 26 pages (5.6%) this could not be determined, for example if no advice was given or if the text advocated contradictory perspectives. These pages were coded as 'unclear' and they were not used in the analyses. After the manual annotation, 259 texts were classified as reliable (58.6%), and 183 (41.4%) as unreliable. To ascertain reliability of the data, a random sample of 59 cases (13.3%) was coded by a different coder (LW), resulting in high inter-coder reliability (*kappa* = .966).

Table 1
Search terms used in data collection and corresponding number of webpages.

Original search term (Dutch)	Translation search term (English)	Number of webpages found
'Vaccinaties veilig'	'Vaccinations safe'	129
'Vaccinaties onveilig'	'Vaccinations unsafe'	22
'Vaccinaties slecht'	'Vaccinations bad'	83
'Vaccinaties goed'	'Vaccination good'	47
'Vaccinaties gevaarlijk'	'Vaccinations dangerous'	31
'Vaccinaties ongevaarlijk'	'Vaccinations not dangerous'	5
'Tegenstanders vaccinaties'	'Opponents of vaccinations'	40
'Voorstanders vaccinaties'	'Supporters of vaccinations'	31
'Argumenten voor vaccinaties'	'Arguments in favor of vaccinations'	17
'Argumenten tegen vaccinaties'	'Arguments against vaccinations'	16
'Waarom vaccineren'	'Why vaccinate'	7
'Waarom niet vaccineren'	'Why not vaccinate'	18
'Kinderen vaccinaties'	'Vaccinations children'	30

2.2. From text to numbers

Before doing any statistical modeling, we needed to transform each text into a set of vector of numbers. We compared two approaches: a so-called count vectorizer and a so-called tf·idf vectorizer. A count vectorizer simply represents a text by simply counting the frequency of all words in it. In other words: Each possible word becomes an independent variable, and its value is how often it occurs in the specific document. Compared to such a simple count vectorizer, a. tf·idf vectorizer (which stands for term frequency times inverse document frequency), additionally weighs the word counts by the number of documents it occurs in at least once. This means that uncommon words get a higher weight. The reason behind this is that one could assume that these words are more helpful in determining to which category a text belongs. Both vectorizers were provided by scikit-learn [32]. We used the standard settings of these vectorizers.

2.3. Training the machine

We compared two machine learning classifiers: a Naïve-Bayes classifier and a Logistic Regression classifier. A classifier is any machine learning model that does not aim to predict a continuous value, but rather tries to predict to which of multiple classes an item belongs – in our case, whether a given text belongs to the class 'reliable' or to the class 'unreliable'. We specifically chose to compare two very popular yet simple classifiers. The Naïve-Bayes algorithm is a typical probabilistic one based on Bayes' theorem. It assumes that all the features of a text (e.g. textual features in this study) are independent. Logistic Regression, in contrast, does not make this assumption and uses the prediction results of the linear regression model to approximate the occurrence ratio of the posterior probability. In particular, we used the Multinomial Naive Bayes classifier with the default settings as provided by scikit-learn and a logistic regression classifier with l2 regularization and the lbfgs solver, which are the defaults suggested by scikit learn. Because we considered two different classifiers and two different vectorizers, we estimated (trained) four different models. To do so, we used 5-fold cross-validation. This means that the total of 442 entries was grouped into 5 groups. We then estimated the models five times, where each group served one time as a test set to evaluate the performance on, and the remaining four groups were used to train the model. We then reported the mean performance, along with standard deviation and 90% confidence interval.

When evaluating our trained classifier, we needed to decide whether we wanted to evaluate the machine's performance to identify *unreliable* information or its performance to identify *reliable* information. This is not the same: a hypothetical classifier that is always right when it classifies some text as reliable (and hence would be good for applying some "stamp of approval" checkmark to texts) may still miss some reliable texts – and that, then, would mean that not everything it classifies as "unreliable" really is unreliable. We would not want to use this classifier then to delete or filter out everything deemed unreliable, even though it works perfectly the other way around. In short, for the evaluation (see next section), we needed to choose which of the two classes ("reliable" or "unreliable") we wanted to predict. Which option is preferred depends on the goal of an intervention for which it may be used: do we want to label unreliable information to warn people, or do we want to identify reliable information to recommend it to people? Because both scenarios are plausible, the models were evaluated in both ways.

2.4. Model evaluation

We evaluated the models based on two measures: precision and recall. While one could be tempted to just evaluate how often the model was "right" in predicting the class of a text in the test dataset (the so-called accuracy, which we report for the sake of completeness as well), this percentage of correct classifications can be misleading. If, for instance, we had 90 reliable and 10 unreliable texts, and our classifier would just always predict reliable, it would still be correct in 90% of the cases, which clearly is misleading. Precision and recall, instead, are based on the number of true positives (TP), false positives (FP), and false negatives (FN). Precision answers the question: "How many of the texts I classified as X really are X?"; more formally, we calculated: precision = TP/(TP + FP). Complementary, recall asks: "How many of all X-s did I classify as such?"; more formally, we calculated: recall = TP/(TP + FN).

2.5. Feature analysis

To get a better understanding *which* of the words differ between unreliable and reliable information, we produced a word shift graph using the Python package shifterator [33]. In this graph, the most characteristic words for unreliable and reliable information are presented, using the Kullback-Leibler divergence, a asymmetric measure of how two texts differ.

2.6. Out-of-sample prediction

After we determined which of our four models performed best, we tested whether it was also capable of predicting a different dataset – a much harder task. This so-called out-of-sample prediction allowed us to evaluate the generalizability of our model. To this end, we retrieved the textual content that was present on all available Dutch webpages on HPV vaccination in May 2019 (see Appendix A). Similar to the webpages on early childhood vaccination, the texts on these webpages were

Table 2
Average performance of the Naive Bayes and Logistic Regression classifiers when predicting unreliable information.

Classifier	metric	*M*	*SD*	*CI-lower*	*CI-upper*
Naïve Bayes, count vectorizer	F1	0.86	0.06	0.80	0.92
	recall	0.84	0.12	0.72	0.96
	precision	0.89	0.06	0.83	0.96
Naïve Bayes, tf·idf vectorizer	F1	0.54	0.16	0.37	0.71
	recall	0.39	0.17	0.22	0.57
	precision	0.98	0.04	0.93	1.02
Logistic regression, count vectorizer	F1	0.83	0.06	0.76	0.90
	recall	0.79	0.12	0.67	0.92
	precision	0.89	0.05	0.84	0.95
Logistic regression, tf·idf vectorizer	F1	0.81	0.11	0.69	0.92
	recall	0.72	0.15	0.56	0.88
	precision	0.96	0.04	0.92	1.01

Table 3
Average performance of the Naive Bayes and Logistic Regression classifiers when predicting reliable information.

Classifier	Metric	*M*	*SD*	*CI-lower*	*CI-upper*
Naïve Bayes, count vectorizer	F1	0.91	0.03	0.88	0.94
	recall	0.92	0.05	0.87	0.98
	precision	0.90	0.07	0.83	0.97
Naïve Bayes, tf·idf vectorizer	F1	0.82	0.04	0.78	0.87
	recall	0.99	0.02	0.98	1.01
	precision	0.70	0.06	0.64	0.77
Logistic regression, count vectorizer	F1	0.89	0.03	0.87	0.92
	recall	0.93	0.05	0.88	0.98
	precision	0.87	0.06	0.81	0.93
Logistic regression, tf·idf vectorizer	F1	0.90	0.04	0.86	0.94
	recall	0.98	0.03	0.95	1.01
	precision	0.84	0.07	0.76	0.91

annotated (by coder NV) based on whether their content advocated a perspective which is consistent with the Dutch RIVM or not. This resulted in a sample of 198 unique webpages, of which 132 (66.7%) were coded as reliable and 66 (33.3%) as unreliable. To ascertain reliability of the coding, a second coder (JA) coded 44 of the HPV texts (22%) showing good inter coder reliability (*kappa* = .952). Next, we used the best model trained on the dataset on early childhood vaccines, to predict this HPV vaccine dataset.

3. Results

3.1. Identifying unreliable information

First, we tested how well our models were able to detect unreliable information. Therefore, two different reliability indicators are reported; precision and recall (see method section for a detailed explanation). All scores are presented in Table 2. Starting with the aspect of precision, we see that both tf·idf models (0.98; 0,96) outperform the count models (0.89; 0.89), but all scores are quite good. Based on these scores, it can be concluded that for all models, at least 89% of the texts that are classified as unreliable are actually unreliable. For the tf·idf models this percentage reaches almost perfect prediction (96–98%). With respect to the aspect of recall, the Naïve Bayes/count model preforms best (scoring 0.84). This means that 84% of the unreliable texts were classified as such. The other models score lower in recall, particularly the Naïve Bayes/tf·idf model (0.39). Also, if we look at the harmonic mean of precision and recall, the so-called F1 score, we see that – except the Naïve Bayes Model with tf·idf vectorizer, all models are reliable and able to identify unreliable information in an automated way.

3.2. Identifying reliable information

We also trained and tested our models to identify reliable information. All precision, recall, and F1 scores for each model are presented in Table 3. For reliable information, precision scores show that particularly the count models perform well (Naïve Bayes; 0.90, Logistic Regression; 0.87). To elaborate, around 9 in 10 of all texts that were classified as reliable by those models are indeed reliable. Recall scores are the highest for both tf·idf classifiers (0.99 and 0.98). This means that nearly all reliable texts from our data set are correctly classified as reliable. Based on the high F1 scores (> 0.82), we can conclude that our models are well able to identify reliable information.

3.3. Information features

Fig. 1 shows the most characteristic words of unreliable (left) and reliable (right) information. Results show that typical words for unreliable information are 'immune system (immuunsysteem)'

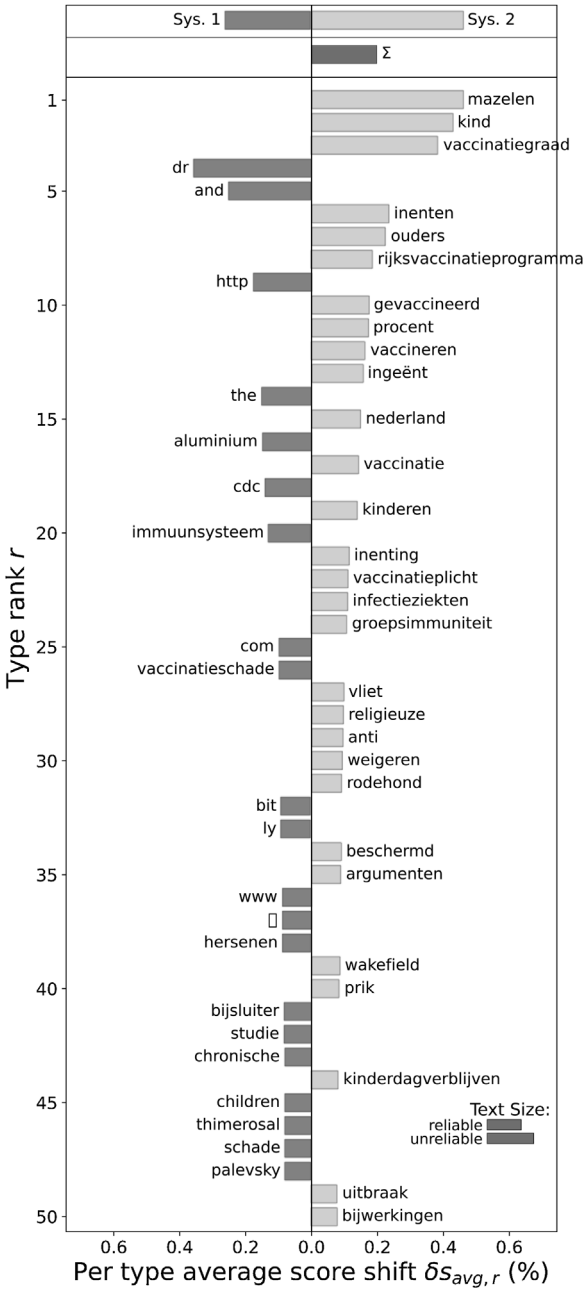

Fig. 1. A word shift graph showing the most characteristic words for unreliable (left) and reliable (right) information, measured using the Kullback-Leibler divergence (KLD).

and 'vaccine damage (vaccinatieschade)'. These texts are also characterized by science-related terminology and qualifications such as 'dr (dr)' and 'study (studie)'. Also 'http' and 'www' are typical for this category, indicating that webpages classified as unreliable generally link or refer to other webpages, which is less common on webpages that present reliable information. We also found that reliable information is characterized by a different discourse. Here, a more abstract viewpoint seems to be taken, referring to the vaccination effects on the population level. Typical words are for example 'immunization rate (vaccinatiegraad)', 'immunize (inenten)' and 'National Immunization Program (rijksvaccinatieprogramma)'. References to vaccine preventable diseases such as 'measles (mazelen)' and 'rubella (rode hond)' are also common.

3.4. Out-of-sample prediction

Out of all classifiers we compared, the Naïve Bayes classifier with count vectorizer had the most consistent overall performance. This can be illustrated with the so-called macro F1 value, the average between the F1 scores for both classes. It is 0.88 (*SD* = .04, *CI* = [.84–.93]) for the Naive Bayes classifier with count vectorizer, followed by the logistic regression with count vectorizer (macro F1 = .86, *SD* = .04, *CI* = [.82–.91]), followed by the logistic regression with tf·idf vectorizer (macro F1 = .85, *SD* = .07, *CI* = [.78–.93]) and, as clearly worst performing combination, the Naive Bayes with tf·idf vectorizer (macro F1 = .68, *SD* = .10, *CI* = .57–.79]).

We therefore used the Naïve Bayes classifier with count vectorizer for the out-of-sample prediction task. This is a test in which we evaluate its performance, on a different data set than it was trained on, in our case a dataset consisting of texts retrieved from webpages about HPV vaccination (see method section). The results show that the recall score for the identification of reliable information is particularly high (0.93), indicating that our classifier was also successful in the identification of reliable information about HPV vaccination, although it was trained on texts about early-childhood vaccines. Regarding the identification of unreliable information, the recall score is considerably lower (0.59). Therefore, our model is better used to classify reliable information compared to unreliable information. This is also reflected in the F1 scores. If the overall goal is to identify reliable information, our classifier performs well. Table 4 shows all the scores.

4. Discussion and conclusion

4.1. Discussion

The aim of this study was to investigate whether it is possible to classify web texts about early childhood vaccination as reliable or unreliable in an automated way, which turned out to be possible. We used very basic classifiers and vectorizers which already generated these significant results. The fact that these relatively simple models reliably distinguish both types of information, in multiple contexts, is promising. If the scientific community develops this further, by applying more complex models such as neural networks [23], the accuracy of the predictions could be even further improved. It is therefore important that the models are tested on many more text samples and various topics. Ultimately, the aim would be to develop a model that is able to classify texts about new diseases as well (e.g., the recent outbreak of the corona virus).

Table 4
Average performance of the Naïve Bayes/count vectorizer model in terms of classifying texts about HPV vaccination.

Type of information	Metric	Value
unreliable	F1	0.68
unreliable	Recall	0.59
unreliable	Precision	0.81
reliable	F1	0.87
reliable	Recall	0.93
reliable	Precision	0.82
	Accuracy	0.81

The finding that our models performed better in identifying reliable information than unreliable information - particularly in terms of out-of-sample prediction - is possibly associated with varying levels of homogeneity. Reliable websites seem to share a common structure characterized by use of terminology (immunization rate, National Immunization Program) and generic terms (vaccinate, infectious diseases) that apply to vaccines in general. As a result, the models trained on early childhood vaccines also perform well on HPV, and probably others types of vaccinations. In contrast, the arguments and ideas presented on webpages advocating against vaccination might be quite diverse (e.g., potential side effects, conspiracy theories) making the classifiers overfit the context they are trained on and less suitable for identifying unreliable information in a different context. In our study, we used a bag-of-words representation, where each feature was a word. Future research could add more features, such as average sentence length, reading ease and the use of specific grammatical categories, which may improve the robustness of the classifier, especially in terms of out of sample prediction.

The feature analysis that we conducted seems to confirm what is known about the discourses used on websites advocating pro or against vaccination. Vaccine-critical webpages frequently address the poisonous ingredients of vaccines and use scientific terminology, ironically making these words indicators of unreliable information. The finding that especially webpages classified as unreliable link to other webpages should be taken seriously. Once people start reading through these webpages, and they are guided from one unreliable webpage to another, this will probably not result in a balanced perspective. Furthermore, webpages classified as reliable seem to present the topic from a public health perspective, whereas vaccine critical webpages tend to focus on the individual perspective. This points to the problem that all parents are facing, but vaccine hesitant parents in particular: they have to weigh the benefits of vaccination on a population level against the perceived risks for their own child. Therefore, it would be good if information supporting vaccinations would discuss individual benefits as well.

Our study has some limitations. First, our sample is fairly small compared to traditional machine learning studies focusing on, for example, Twitter data [34,35]. This is due to the fact that the number of webpages addressing early childhood and HPV vaccination in Dutch is limited. The number of webpages included in our study is however comparable to the number of cases included in other machine learning studies that classified health websites [22,23]. Furthermore, even though the data on early childhood vaccines were collected in 2018 and things may change fast online, a random check (conducted in August 2020) of 10% the webpages showed that over 80% of the webpages were still available and 86% presented the same message as two years ago. A final limitation lies in our conceptualization of reliable versus unreliable misinformation. Whereas the accuracy of medical information can always be debated and there will always be people that hold different opinions, we believe that standards and guidelines developed on the basis of scientific and medical evidence are the best reliability indicators. Since the guidelines applied by the Dutch RIVM are based on advice provided by the Health Council of the Netherlands, a medical authority that bases its advice on scientific research [36], we argued that this would be

the best indication available to judge a webpage's reliability. However, it should be noted that a classification in just two categories is very strict. Especially in the case of longer texts, multiple perspectives can be addressed, and different arguments can be discussed. Therefore, future research could explore a more nuanced classification.

4.2. Conclusion

Considering the vast amount of incorrect online information, and the fact that many parents turn to the Internet to inform themselves about vaccines, automated classification of online content can be helpful to guide people towards reliable information. Using supervised machine learning, we successfully trained and tested multiple classifiers on texts that were retrieved from existing webpages. Furthermore, the robustness of the most optimal classifier was tested on texts retrieved from webpages about HPV vaccination. In both contexts, the classifier performed well, especially in terms of the identification of reliable information.

4.3. Practice implications

When collecting the webpages for our study, we made an effort to include as much Dutch webpages as possible. It is encouraging to see that for both topics more webpages with reliable information were identified than webpages presenting unreliable information. Furthermore, despite the fact that in the Netherlands HPV vaccination uptake is much lower than early childhood vaccination (around 53% versus 93% [37]), there appeared to be less unreliable information available about HPV. This means that the high amount of anti-HPV information on social media that has been associated with lower vaccine uptake [38] is not so much reflected in websites. Efforts on correcting misinformation on websites should therefore mainly focus on early childhood vaccines.

Since many people find it difficult to determine if online information can be trusted, our classifiers could be used as a basis to develop online labeling tools. Labels can inform information-seekers about the reliability of the webpages they encounter. Since the results of our out-of-sample prediction were especially reliable with respect to identifying reliable information, it is recommended to label reliable information, not unreliable information. Labeling unreliable information as such makes people aware of the fact that the content of the message should be treated with caution, but also has some disadvantages. It could for example cause an implied truth effect, meaning that incorrect information without a label is considered validated, and thus seen as more accurate [39]. Also, labeling vaccine critical information as unreliable information because it goes against the advice of health authorities could give vaccine critical parents the impression that health authorities do not take their arguments seriously. Solely labeling reliable information therefore seems to be preferred over also labeling unreliable information.

Funding

Funding was provided by the Amsterdam School of Communication Research, which had no role in the study design, data collection, analysis, interpretation, or writing of the report.

CRediT authorship contribution statement

Corine S. Meppelink: Conceptualization, Methodology, Resources, Writing - original draft. **Hanneke Hendriks:** Conceptualization, Methodology, Resources, Writing - review & editing. **Damian Trilling:** Methodology, Software, Formal analysis, Data curation, Writing - original draft. **Julia C.M. van Weert:** Writing - review & editing, Funding acquisition. **Anqi Shao:** Methodology, Software, Data curation. **Eline S. Smit:** Conceptualization, Methodology, Resources, Writing - review & editing.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

The authors want to thank Lea Schwerdt, Leonie Westerbeek, Jacobien Adam, and Nynke Visscher for their assistance with the coding. We also thank the reviewers for their valuable comments.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.pec.2020.11.013.

References

[1] G. Eysenbach, J. Powell, O. Kuss, et al., Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review, JAMA 287 (20) (2002) 2691–2700, doi:http://dx.doi.org/10.1001/jama.287.20.2691.
[2] A.J. Flanagin, M.J. Metzger, Perceptions of internet information credibility, J. Mass Commun. Q. 77 (3) (2000) 515–540, doi:http://dx.doi.org/10.1177/107769900007700304.
[3] Statistics Netherlands, Internet; toegang, gebruik en faciliteiten, (2020) [Internet; Access, Use, and Facilities]. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83429NED/table?fromstatweb. Accessed February.
[4] N. Diviani, B. van den Putte, S. Giani, J.C.M. van Weert, Low health literacy and evaluation of online health information: a systematic review of the literature, J. Med. Internet Res. 17 (5) (2015) e112, doi:http://dx.doi.org/10.2196/jmir.4018.
[5] P. Davies, S. Chapman, J. Leask, Anti-vaccination activists on the world wide web, Arch. Dis. Child. 87 (1) (2002) 22–25, doi:http://dx.doi.org/10.1136/adc.87.1.22.
[6] R.K. Zimmerman, R.M. Wolfe, D.E. Fox, et al., Vaccine criticism on the world wide web, J. Med. Internet Res. 7 (2) (2005) e17, doi:http://dx.doi.org/10.2196/jmir.7.2.e17.
[7] M. Dredze, D.A. Broniatowski, M.C. Smith, et al., Understanding vaccine refusal: why we need social media now, Am. J. Prev. Med. 50 (4) (2016) 550–552, doi:http://dx.doi.org/10.1016/j.amepre.2015.10.002.
[8] A. Kata, A. Postmodern Pandora's box: anti-vaccination misinformation on the internet, Vaccine 28 (7) (2010) 1709–1716, doi:http://dx.doi.org/10.1016/j.vaccine.2009.12.022.
[9] M.B. Gilkey, A. McRee, B.E. Magnus, et al., Vaccination confidence and parental refusal/delay of early childhood vaccines, PLoS One 11 (7) (2016)e0159087, doi:http://dx.doi.org/10.1371/journal.pone.0159087.
[10] E.A. van Lier, J.L.E. Geraedts, P.J. Oomen, et al., Vaccinatiegraad en jaarverslag Rijksvaccinatieprogramma Nederland 2016. [Immunisation Coverage and Annual Report National Immunisation Programme in the Netherlands 2016], (2017), doi:http://dx.doi.org/10.21945/RIVM-2017-0010 2017-0010.
[11] S.B. Omer, D.A. Salmon, W.A. Orenstein, et al., Vaccine refusal, mandatory immunization, and the risks of vaccine-preventable diseases, N. Engl. J. Med. 360 (19) (2009) 1981–1988, doi:http://dx.doi.org/10.1056/NEJMsa0806477.
[12] I.A. Harmsen, G.G. Doorman, L. Mollema, et al., Parental information-seeking behaviour in childhood vaccinations, BMC Public Health 13 (1) (2013) 1219, doi:http://dx.doi.org/10.1186/1471-2458-13-1219.
[13] E. Dubé, C. Laberge, M. Guay, P. Bramadat, R. Roy, J.A. Bettinger, Vaccine hesitancy: an overview, Hum. Vaccin. Immunother. 9 (8) (2013) 1763–1773, doi:http://dx.doi.org/10.4161/hv.24657.
[14] B.A. Lehmann, H.E. de Melker, D.R. Timmermans, L. Mollema, Informed decision making in the context of childhood immunization, Patient Educ. Couns. 100 (12) (2017) 2339–2345, doi:http://dx.doi.org/10.1016/j.pec.2017.06.015.
[15] C.S. Meppelink, E.G. Smit, N. Diviani, J.C.M. van Weert, Health literacy and online health information processing: unraveling the underlying mechanisms, J. Health Commun. 21 (sup2) (2016) 109–120, doi:http://dx.doi.org/10.1080/10810730.2019.1583701.
[16] Health on the Net, Health on the Net Foundation, (2020) https://www.hon.ch/. Accessed February.
[17] D. Charnock, S. Shepperd, G. Needham, R. Gann, DISCERN: an instrument for judging the quality of written consumer health information on treatment choices, J. Epidemiol. Commun. Health 53 (2) (1999) 105–111, doi:http://dx.doi.org/10.1136/jech.53.2.105.

[18] G. Sak, N. Diviani, A. Allam, P.J. Schulz, Comparing the quality of pro-and anti-vaccination online information: a content analysis of vaccination-related webpages, BMC Public Health 16 (1) (2016) 38, doi:http://dx.doi.org/10.1186/s12889-016-2722-9.
[19] Y. Sun, Y. Zhang, J. Gwizdka, C.B. Trace, Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators, J. Med. Internet Res. 21 (5) (2019)e12522, doi:http://dx.doi.org/10.2196/12522.
[20] N. Diviani, B. van den Putte B, C.S. Meppelink, J.C.M. van Weert, Exploring the role of health literacy in the evaluation of online health information: insights from a mixed-methods study, Patient Educ. Couns. 99 (2016) 1017–1025, doi:http://dx.doi.org/10.1016/j.pec.2016.01.007.
[21] C. Boyer, L. Dolamic, Automated detection of HONcode website conformity compared to manual detection: an evaluation, J. Med. Internet Res. 17 (6) (2015) e135, doi:http://dx.doi.org/10.2196/jmir.3831.
[22] P. Sondhi, V.V. Vydiswaran, C. Zhai, Reliability prediction of webpages in the medical domain, European Conference on Information Retrieval, Springer, 2012.
[23] L. Kinkead, A. Allam, M. Krauthammer, AutoDiscern: rating the quality of online health information with hierarchical encoder attention-based neural networks, BMC Med. Inform. Decis. Mak. 20 (1) (2020) 1–13, doi:http://dx.doi.org/10.1186/s12911-020-01131-z.
[24] L. Westerbeek, H. Hendriks, E.S. Smit, C.S. Meppelink, Combatting online misinformation regarding vaccinations, Paper presented at the annual conference Etmaal van de Communicatiewetenschap, Amsterdam, The Netherlands, 2019.
[25] M. Frické, D. Fallis, M. Jones, G.M. Luszko, Consumer health information on the Internet about carpal tunnel syndrome: indicators of accuracy, Am. J. Med. 118 (2) (2005) 168–174, doi:http://dx.doi.org/10.1016/j.amjmed.2004.04.032.
[26] E.V. Bernstam, M.F. Walji, S. Sagaram, D. Sagaram, C.W. Johnson, F. Meric-Bernstam, Commonly cited website quality criteria are not effective at identifying inaccurate online information about breast cancer, Cancer 112 (6) (2008) 1206–1213.
[27] RIVM, Over het Rijksvaccinatieprogramma [About the Immunization Programme], (2020) . https://rijksvaccinatieprogramma.nl/over-het-programma.
[28] W. van Atteveldt, T. Peng, When communication meets computation: opportunities, challenges, and pitfalls in computational communication science, Commun. Methods Meas. 12 (2–3) (2018) 81–92, doi:http://dx.doi.org/10.1080/19312458.2018.1458084.
[29] G.C. Banks, H.M. Woznyj, R.S. Wesslen, R.L. Ross, A review of best practice recommendations for text analysis in R (and a user-friendly app), J. Bus. Psychol. 33 (4) (2018) 445–459, doi:http://dx.doi.org/10.1007/s10869-017-9528-3.
[30] T.C. Guetterman, T. Chang, M. DeJonckheere, T. Basu, E. Scruggs, V.V. Vydiswaran, Augmenting qualitative text analysis with natural language processing: methodological study, J. Med. Internet Res. 20 (6) (2018) e231, doi:http://dx.doi.org/10.2196/jmir.9702.
[31] J.W. Boumans, D. Trilling, Taking stock of the toolkit: an overview of relevant automated content analysis approaches and techniques for digital journalism scholars, Digit. J. 4 (1) (2016) 8–23, doi:http://dx.doi.org/10.1080/21670811.2015.1096598.
[32] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (Oct) (2011) 2825–2830.
[33] R.J. Gallagher, M.R. Frank, L. Mitchell, A.J. Schwartz, A.J. Reagan, C.M. Danforth, et al., Generalized Word Shift Graphs: a Method for Visualizing and Explaining Pairwise Comparisons Between Texts. arXiv Preprint arXiv:2008.02250, (2020) .
[34] B. Burscher, D. Odijk, R. Vliegenthart, et al., Teaching the computer to code frames in news: comparing two supervised machine learning approaches to frame analysis, Commun. Measures. 8 (3) (2014) 190–206, doi:http://dx.doi.org/10.1080/19312458.2014.937527.
[35] S.A. Vermeer, T. Araujo, S.F. Bernritter, et al., Seeing the wood for the trees: how machine learning can help firms in identifying relevant electronic word-of-mouth in social media, Int. J. Res. Mark. 36 (3) (2019) 492–508, doi:http://dx.doi.org/10.1016/j.ijresmar.2019.01.010.
[36] Gezondheidsraad, Commissie vaccinaties. [Committee Vaccines], (2020) https://www.gezondheidsraad.nl/over-ons/organisatie/vaste-commissies/vaccinaties. Accessed February.
[37] E. van Lier, L. Kamp, P. Oomen, H. Giesbers, J. van Vliet, I. Drijfhout, et al., Vaccinatiegraad en jaarverslag Rijksvaccinatieprogramma Nederland 2019 [Immunisation Coverage and Annual Report National Immunisation Programme in the Netherlands 2019]. Bilthoven, RIVM, 2020, doi:http://dx.doi.org/10.21945/RIVM-2020-0011 Report No.: 2020-0011.
[38] A.G. Dunn, D. Surian, J. Leask, A. Dey, K.D. Mandl, E. Coiera, Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States, Vaccine 35 (23) (2017) 3033–3040, doi:http://dx.doi.org/10.1016/j.vaccine.2017.04.060.
[39] G. Pennycook, T.D. Cannon, D.G. Rand, Prior exposure increases perceived accuracy of fake news, J Exp Psychol.: Gen. 147 (12) (2018) 1865, doi:http://dx.doi.org/10.1037/xge0000465.