A Highly Accurate Internet-Based Fake Information Detection Tool for Indonesian Twitter

Rizal Arifin*[1], Gus Nanang Syaifuddiin[2], Desriyanti[1], Zulkham Umar Rosyidin[1], Ghulam Asrofi Buntoro[1]
E-mail: rarifin@umpo.ac.id, gus.nanang@pnm.ac.id, yunandes@gmail.com, zzhulqam@gmail.com, ghulam@umpo.ac.id
*Corresponding author
[1]Faculty of Engineering, Universitas Muhammadiyah Ponorogo, Indonesia
[2]Department of Information Technology, Politeknik Negeri Madiun, Indonesia

Keywords: fake information, identification, Indonesian Twitter, machine learning, application

Received: September 22, 2022

The dissemination of fake information through social media has several harmful effects on the social life of a nation. Indonesia has been afflicted by the dissemination of erroneous information regarding the negative health consequences of vaccination, resulting in widespread unwillingness to undergo immunization. Therefore, it is necessary to combat such misleading information. We developed a web application using machine learning technologies to identify bogus information flowing on Indonesian Twitter. A Passive-Aggressive Classifier and n-gram tokenization are used to handle data. The application test results indicate that the detection accuracy, precision, and recall for 1-3 grams of tokenization are higher than 90%. In addition, the black box approach yields reliable findings for all application functionalities.

Povzetek: Opisana je metoda strojnega učenja za zaznavanje lažnih novic za Indonezijo za Twitter.

1 Introduction

With the advancement of technology, it is possible to access all types of information using a variety of Internet-connected gadgets. Recently, besides the COVID-19 pandemic, the globe is also facing an infodemic. Pandemics have proven to be fertile ground for the dissemination of hoaxes, particularly through social media. Indonesia has been subjected to false assaults perpetrated by unscrupulous groups. Researchers have found that joint efforts are required to combat the spread of false news to keep people informed. Infodemic resilience is the capacity of a system to rebound from a significant change [1]. There is a figure that can be used as data for sentiment analysis, and a great deal of data on social media can be used to determine public or consumer sentiment toward particular items, and there are several more uses of data from social media. Using machine learning techniques, a sentiment analysis of the 2019 Indonesian presidential election was performed [2]. The findings may be used to illustrate the attitudes of the Indonesian people regarding the 2019 Indonesian presidential candidate, particularly among Twitter users.

However, the rise of information technology has a detrimental effect on society if individuals are less able to sift and choose information. This is because in cyberspace or on the Internet, anyone may assume any identity and wear either a mask of good or a mask of evil, or even both masks simultaneously. As a result, many sorts of irresponsibly anonymous, misleading, or fake information may swiftly spread over the Internet, particularly on social media platforms, such as Twitter, Facebook, and Instagram, which are not reviewed by editors or subject matter experts. This can lead people to rapidly accept false information or hoaxes without filters, which might harm the nation's unity and integrity [3]–[5]. In a hoax, false information is packed and presented in such a way that it seems genuine. During the COVID-19 pandemic, we received information regarding the virus, particularly COVID-19, for over two years. Some of that material is accurate and useful, but some false information has been widely disseminated [6]–[8]. This produces concern and dread among the Indonesian people and the international community.

The propagation of fake information and hoaxes in Indonesia has been on the rise, with several examples being propagated each minute. Every big event in Indonesia, from the Indonesian presidential election campaign to natural disasters, can be used by unresponsible parties to promote false information and perpetuate widespread hoaxes. Indonesia is an ethnically and racially diverse nation with a significant number of people and cultures. The propagation of false information may harm the nation's unity and integrity [9]–[12]. To combat the spread of urban legends, considerable efforts are required from all stakeholders, particularly the authorities, and specifically the government. The Ministry of Communication and Information of the Republic of Indonesia blocks sites that disseminate hoaxes, and it educates the public on the hazards of fake news in print media and on television, the Internet, and social media [11].

Researchers in the area of information technology, particularly artificial intelligence, have attempted to use a variety of techniques to identify social media hoaxes reliably. Other researchers, including Hasan et al., created an artificial neural network-based technique for identifying possible centrifugal pump problems [13] and characterizing the human sleep state [14]. In addition, Hasan et al. [15] created a novel technique for segmentation-based texture fractal analysis (SFTA), which is said to be more accurate than traditional SFTA. The most precise approach is the N-gram method. These methods function by tokenizing sentences according to length N, so researchers can estimate how long N is in n-grams to obtain the highest level of precision [16]–[18]. The phrase frequency-inverse document frequency is the most popular weighting approach employed by artificial intelligence researchers, notably in natural language processing (NLP) sentence extraction (TF-IDF). The TF-IDF algorithm determines the frequency of a word's occurrence in a sentence and compares it to the inverse of the data. The TF-IDF technique also assesses how often a word occurs in a phrase. The greater the frequency, the lower the weight value, indicating that the word—typically a conjunction (and, which, in, will, with, etc.)—is unimportant in a sentence [19], [20]. Passive-aggressive algorithms are large-scale learning algorithms that are commonly used in applications involving huge amounts of data. They do not need the same learning rate as Perception. However, unlike Perception, they contain regularization settings. This algorithm excels at identifying false information on social media platforms, such as Twitter and WhatsApp, where fresh data are posted every second [21]–[23]. Very recently, Natural Language Processing (NLP) methods were used to examine tweets on the Covid-19 vaccine. According to reports, the Support Vector Machine (SVM) classifier is the best match for the dataset with an accuracy of 84.32 percent. This study illustrates how Twitter data and machine learning techniques may be used to analyze the developing public discourse and attitudes on the Covid-19 vaccine deployment campaign [24]. In more recent study, Hutama and Suhartono used the pre-trained transformer multilingual model (XLM-R and mBERT) in conjunction with a BERTopic model as a topic distribution model to categorize Indonesian fake news. They obtained an accuracy value of 90.51% [25].

Table 1 provides some instances of findings from past studies on the identification of fake information. The accuracy of the acquired detection findings varies between 78.6% and 92.0% [17], [18], [24], [26], [27]. The results of the n-gram tokenization approach and machine learning are more precise than those of the Naive Bayes method, according to these data. In this work, we used the n-gram approach and machine learning based on the findings of a prior investigation.

Table 1: Some previous related works.

Year	Data	Technique	Accuracy
2017 (ref. [17])	published literature	n-gram and machine learning	92.0%
2017 (ref. [27])	Indonesian hoax news	Naïve Bayes	78.6%
2018 (ref. [18])	online news articles	n-gram and machine learning	92.0%
2020 (ref. [26])	Indonesian fake news	Naïve Bayes	87.0%
2022 (ref. [24])	Covid-19 vaccine-related tweets	n-gram and machine learning	84.3%

Although there has been a significant amount of research in the field of artificial intelligence, particularly to detect false information or hoaxes, there is still a need for research that continues to increase the accuracy and effectiveness of hoax detection to reduce and prevent the spread of hoaxes in the community. The 1-2 gram combination model has been shown to be more than 90% accurate when the proportion of training data to test data is 70:30 [28]. In this research, we construct an Internet-based application system by modifying n-grams and re-sorting the datasets used in the training process. The aim is to develop an easy-to-use method for detecting hoaxes in Indonesian tweets. This program may therefore be used by social media users in Indonesia, particularly Twitter users, to determine the veracity of the information they receive, thus reducing the negative effects of fake news.

2 Methods

The waterfall approach is employed for application design, while black box testing is used for system testing. This approach is selected because the application design process is executed sequentially.

The waterfall approach facilitates departmentalization and control, which is a significant benefit. To reduce the likelihood of mistakes, the model creation process is broken down into sequential steps. The waterfall method has five stages: requirements analysis and definition, system and software design, implementation and unit testing, integration and system testing, and operations and maintenance.

2.1 Requirements analysis and definition

This step entails collecting requirements at the system level, business strategy level, requirements data level, and business unit level, for instance, at the system level.

2.2 System and software design

This phase of the process comprises the design of the management workflow and the implementation system. This phase is also known as the programming or implementation phase. During this phase, the application design created using a programming language is converted into an executable application. The outcomes of the design are expressed in strings or lines of computer code that are comprehensible.

2.3 Implementation and unit testing

In this programming step, the creation of software is separated into discrete modules that are integrated in the

subsequent phase. In addition, testing and verification of the functionality of the created modules are performed at this step to see whether they fulfill the specified requirements. Confusion matrix testing is conducted during unit testing to anticipate performance.

Confusion matrix is a technique often used to calculate the precision of data mining concepts. The matrix uses four terms to indicate the outcomes of the classification process: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). True Negative (TN) is the number of negative data that are successfully identified, while False Positive (FP) is the number of negative data that are incorrectly recognized as positive (see Table 2). True Positive (TP) is the number of data that are appropriately identified. False Negative (FN) is the inverse of True Positive (TP), indicating that the data are positive but are identified as negative.

Table 2: Confusion matrix

		Values	
		True	**False**
Prediction	**True**	True Positive (correct result)	True Negative (unexpected result)
	False	False Negative (missing result)	True Negative (correct absence of result)

2.4 Integration and system testing

Software testing serves an essential purpose in its development by detecting faults (defects) produced by disparities between predicted and actual outcomes [29]. The objective of system testing is to guarantee that all processes conform to the requirements. At this point, testing is performed. Black box testing is a software testing technique that focuses on functionality, particularly the input and output of the program and whether they agree with expectations.

2.5 Operations and maintenance

At this level, the manager/administrator takes action to support the system's functioning.

3 Results and discussion

At the stage of requirements analysis and definition, the authors study system requirements to determine the to-be-built process flow and system data. The system model is developed using a tweeter dataset of 4617 tweets. These are raw data that have not been preprocessed, as seen in Figure 1. The data are accessible in csv format with the following columns: subjects, keywords, tweets, photos, URLs, and labels. The data are characterized as fake, unclassified, and valid. Figure 1 shows that 3042 records are identified as hoaxes, 730 are valid, and 845 are not classified.

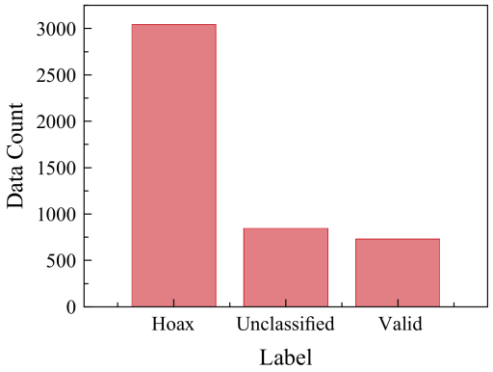

Figure 1: Number of data by label

Next is the system design and software design phase. Figure 2 shows the process and flow of the constructed system. The procedure has two broad stages: (*i*) creating a model for classification, which starts with separating the dataset into training data and testing data, and (*ii*) evaluating the model's performance. The raw data are then preprocessed to prepare for the feature retrieval procedure. In the preprocessing phase, the cases are folded and stemmed.

The next step, feature extraction, is performed using tokenization and n-grams to extract word characteristics from the data. The extracted feature data are used in the classification model training procedure. In addition (ii) the data detection procedure employs a classification model to identify user-entered texts. The system architecture is shown in Figure 3.

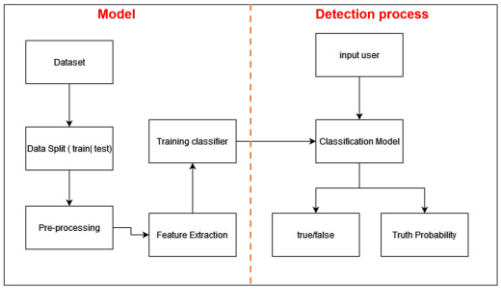

Figure 2: Detection system

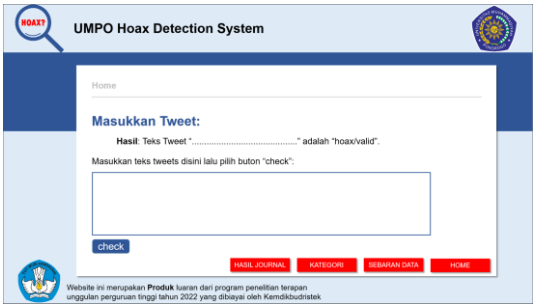

Figure 3: User interface design

Implementation and unit testing of the system come next. The system is implemented using the computer language Python. Figure 4 shows an example of the

implementation of the system. Using the diagram in Figure 2, the system process is developed.

Figure 5 shows the results of assessing the performance of fake information identification using a combination of Passive-Aggressive Classifier and *n*-gram. In the testing phase, 70% of the data are training data, and 30% are testing data. The *n*-gram tokenization is performed using n values from 1 to 3 and their permutations.

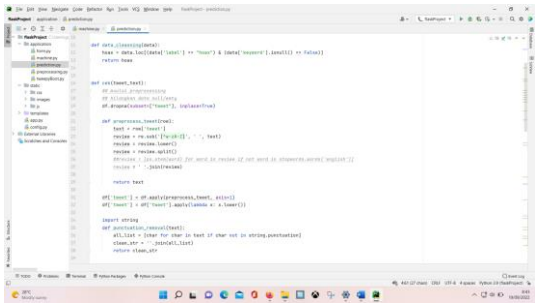

Figure 4: Example of the implementation part of a system written in the Python programming language

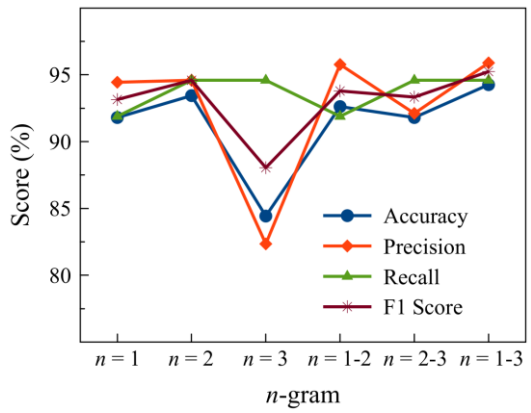

Figure 5: Results of the fake information detection system

Table 3. Comparison of accuracy with findings from recent related works

Works	Data	Technique	Accuracy
(ref. [17])	published literature	n-gram and machine learning	92.0%
(ref. [27])	Indonesian hoax news	Naïve Bayes	78.6%
(ref. [18])	online news articles	n-gram and machine learning	92.0%
(ref. [26])	Indonesian fake news	Naïve Bayes	87.0%
(ref. [24])	Covid-19 vaccine-related tweets	n-gram and machine learning	84.3%
This work	Tweets from Indonesian Twitter	n-gram and machine learning	94.3%

The findings in Figure 5 demonstrate that the combination of Passive-Aggressive Classifier and 1-3-grams generates superior performance. Compared to other combinations, the accuracy, precision, and F1 scores for this combination were the greatest at 94.26%, 95.89%, and 95.22%, respectively. In contrast, the recall score for the combination of the Passive-Aggressive Classifier with 2-grams, 3-grams, 2-3-grams, and 1-3-grams is 94.59%. We constructed an Internet-based fake information detection program using machine learning and a mix of the Passive-Aggressive Classifier and 1-3-gram algorithms based on the acquired findings. From Table 3, it is evident that our system's accuracy and precision outperformed those reported by Zaman et al. [26] and Pratiwi et al. [27], both of which use the Naïve Bayes Classifier method. Using n-gram tokenization and machine learning, we achieve a better degree of accuracy than a number of earlier research that used the same methodology [17], [18], [24].

Next, system testing is conducted using the black box technique. The results of system testing are shown in Table 4. The results demonstrate that all the application's features and menus are operating and working as planned. This indicates that the program is ready for use by the end user. Figure 6 demonstrates how application test results may be shown on the "topik" and "hasil journal" menus. In this test, both may display the "topik" page for the hoax category based on tweet data and the "hasil journal" page, which provides a comparison table for the method's findings.

Figure 6: Results of the fake information detection system

Table 4: The outcomes of system testing using the black box technique.

Test Case	Input	Expected result	Actual result	Description
go to the home menu view	click the home menu	displays a home page containing input text and a check button for hoax detection	as expected	Valid
input tweet sentence data based on category topics in the input text	click the check button	displays the results of hoax detection, namely hoax or valid	as expected	Valid
go to sebaran data menu view	click the sebaran data button	displays a sebaran data page containing a bar chart of the number of valid and hoax data	as expected	Valid
go to topik menu view	click the topic button	displays the hoax category in topik page from tweet data	as expected	Valid
go to hasil journal menu view	click the hasil journal button	displays the results of the journal that contains a comparison table for the results of the method used	as expected	Valid

4 Conclusion

Using machine learning, we built an Internet-based program to identify fraudulent information or hoax information on Indonesian Twitter. Data are handled using a combination of Passive-Aggressive Classifier and n-gram tokenization. The results of the application tests reveal that the detection accuracy, precision, and recall for 1-3 grams of tokenization are more than 90%. In addition, black box testing of the hoax detection program shows that all the features and menus operate and function as intended. The "topic" and "journal results" menu options may show, among other things, a "topic" page for the hoax category based on Twitter data and a comparison table of performance-testing results for the combination of the Passive-Aggressive Classifier and n-gram algorithms. From these data, it can be stated that the program for detecting hoaxes is ready for user deployment. Attempts are made to increase the accuracy of the application model for hoax detection using more datasets and other machine learning methods.

Acknowledgements

This study was funded by a PTUPT 2022 grant from the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia [contract number 073/E5/P6.02.00.PT/2022].

References

[1] V. L. Muzykant, M. A. Muqsith, R. R. Pratomo, and V. Barabash, "Fake News on COVID-19 in Indonesia," in *Pandemic Communication and Resilience. Risk, Systems and Decisions*, D. M. Berube, Ed. Cham: Springer, 2021, pp. 363–378.

[2] G. A. Buntoro, R. Arifin, G. N. Syaifuddiin, A. Selamat, O. Krejcar, and H. Fujita, "The implementation of the machine learning algorithm for the sentiment analysis of Indonesia's 2019 presidential election," *IIUM Eng. J.*, vol. 22, no. 1, pp. 78–92, 2021, doi: https://doi.org/10.31436/iiumej.v22i1.1532.

[3] T. Buchanan, "Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation," *PLoS One*, vol. 15, no. 10, p. e0239666, 2020, doi: https://doi.org/10.1371/journal.pone.0239666.

[4] M. Celliers and M. Hattingh, "A Systematic Review on Fake News Themes Reported in Literature," in *Responsible Design, Implementation and Use of Information and Communication Technology*, 2020, pp. 223–234, doi: https://doi.org/10.1007/978-3-030-45002-1_19.

[5] T. Khan, A. Michalas, and A. Akhunzada, "Fake news outbreak 2021: Can we stop the viral spread?," *J. Netw. Comput. Appl.*, vol. 190, p. 103112, 2021, doi: https://doi.org/10.1016/j.jnca.2021.103112.

[6] A. Alasmari, A. Addawood, M. Nouh, W. Rayes, and A. Al-Wabil, "A Retrospective Analysis of the COVID-19 Infodemic in Saudi Arabia," *Futur. Internet*, vol. 13, no. 10, p. 254, 2021, doi: https://doi.org/10.3390/fi13100254.

[7] M. Montesi, "Understanding fake news during the Covid-19 health crisis from the perspective of information behaviour: The case of Spain," *J. Librariansh. Inf. Sci.*, vol. 53, no. 3, pp. 454–465, 2020, doi: https://doi.org/10.1177/0961000620949653.

[8] S. van der Linden, J. Roozenbeek, and J. Compton, "Inoculating Against Fake News About COVID-19," *Front. Psychol.*, vol. 11, p. 566790, 2020, doi: https://doi.org/10.3389/fpsyg.2020.566790.

[9] K. Lutfiyah, "Hoax and Fake News During Covid-19: Is the Law Effective in Overcoming It?," *Indones. J. Int'l Clin. Leg. Educ.*, vol. 2, no. 3, pp. 345–360, 2020, doi: https://doi.org/10.15294/ijicle.v2i3.38422.

[10] N. M. Nasir, B. Baequni, and M. I. Nurmansyah, "Misinformation Related to Covid-19 in

Indonesia," *J. Adm. Kesehat. Indones.*, vol. 8, no. 1, pp. 51–59, 2020, doi: http://dx.doi.org/10.20473/jaki.v8i0.2020.51-59.

[11] M. Rasidin, D. Witro, B. Yanti, R. Purwaningsih, and W. Nurasih, "The Role of Government in Preventing The Spread if Hoax Related The 2019 Elections in Social Media," *Diakom*, vol. 3, no. 2, pp. 127–3, 2020, doi: https://doi.org/10.17933/diakom.v3i2.76.

[12] Y. I. Ferdiawan, P. A. D. Nurjanah, E. P. Krisdyan, A. Hidayatullah, H. J. M. Sirait, and N. A. Rakhmawati, "HOAX Impact to Community Through Social Media Indonesia," *Cakrawala*, vol. 19, no. 1, pp. 121–124, 2019, doi: https://doi.org/10.31294/jc.v19i1.4452.

[13] M. J. Hasan, A. Rai, Z. Ahmad, and J.-M. Kim, "A Fault Diagnosis Framework for Centrifugal Pumps by Scalogram-Based Imaging and Deep Learning," *IEEE Access*, vol. 9, pp. 58052–58066, 2021, doi: https://doi.org/10.1109/CEEICT.2016.7873115.

[14] M. J. Hasan, D. Shon, K. Im, H.-K. Choi, D.-S. Yoo, and J.-M. Kim, "Sleep State Classification Using Power Spectral Density and Residual Neural Network with Multichannel EEG Signals," *Appl. Sci.*, vol. 10, no. 21, p. 7639, 2020, doi: https://doi.org/10.3390/app10217639.

[15] M. J. Hasan, J. Uddin, and S. N. Pinku, "A novel modified SFTA approach for feature extraction," in *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, 2016, pp. 1–5, doi: https://doi.org/10.1109/CEEICT.2016.7873115.

[16] H. E. Wynne and Z. Z. Wint, "Content Based Fake News Detection Using N-Gram Models," in *Information Integration and Web-based Applications & Services*, 2019, pp. 669–673, doi: https://doi.org/10.1145/3366030.3366116.

[17] H. Ahmed, I. Traore, and S. Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques," in *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, 2017, pp. 127–138, doi: https://doi.org/10.1007/978-3-319-69155-8_9.

[18] H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," *Secur. Priv.*, vol. 1, p. e9, 2018, doi: https://doi.org/10.1002/spy2.9.

[19] J. Huang, "Detecting Fake News With Machine Learning," *J. Phys. Conf. Ser.*, vol. 1693, p. 012158, 2020, doi: https://doi.org/10.1088/1742-6596/1693/1/012158.

[20] M. J. Awan *et al.*, "Fake News Data Exploration and Analytics," *Electronics*, vol. 10, p. 2326, 2021, doi: https://doi.org/10.3390/electronics10192326.

[21] R. R. Mandical, N. Mamatha, N. Shivakumar, R. Monica, and A. N. Krishna, "Identification of Fake News Using Machine Learning," in *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, 2020, pp. 1–6, doi: https://doi.org/10.1109/CONECCT50063.2020.9198610.

[22] A. Chugh, Y. Arora, J. Singh, Shobhit, and Ronak, "Media Manipulation Detection System Using Passive Aggressive," *Int. J. Innov. Res. Comput. Sci. Technol.*, vol. 9, no. 3, pp. 48–52, 2021, doi: https://doi.org/10.21276/ijircst.2021.9.3.8.

[23] S. Gupta and P. Meel, "Fake News Detection Using Passive-Aggressive Classifier," in *Inventive Communication and Computational Technologies*, 2020, pp. 155–164, doi: https://doi.org/10.1007/978-981-15-7345-3_13.

[24] S. K. Akpatsa, H. Lei, X. Li, and V.-H. K. S. Obeng, "Evaluating Public Sentiments of Covid-19 Vaccine Tweets Using Machine Learning Techniques," *J. Inf. Syst. Telecommun.*, vol. 46, no. 1, pp. 69–75, 2022, doi: https://doi.org/10.31449/inf.v46i1.3483.

[25] L. B. Hutama and D. Suhartono, "Indonesian Hoax News Classification with Multilangual Transformer Model and Bertopic," *Informatica*, vol. 46, no. 8, pp. 81–90, 2022, doi: https://doi.org/10.31449/inf.v46i8.4336.

[26] B. Zaman, A. Justitia, K. N. Sani, and E. Purwanti, "An Indonesian Hoax News Detection System Using Reader Feedback and Naïve Bayes Algorithm," *Cybern. Inf. Technol.*, vol. 20, no. 1, pp. 82–94, 2020, doi: https://doi.org/10.2478/cait-2020-0006.

[27] I. Y. R. Pratiwi, R. A. Asmara, and F. Rahutomo, "Study of hoax news detection using naïve bayes classifier in Indonesian language," in *2017 11th International Conference on Information Communication Technology and System (ICTS)*, 2017, pp. 73–78, doi: https://doi.org/10.1109/ICTS.2017.8265649.

[28] G. N. Syaifuddiin *et al.*, "Hoax Identification of Indonesian Tweeters Using Ensemble Classifier," *J. Inf. Syst. Telecommun.*, vol. IN PRESS, 2022.

[29] B. P. Pratama, I. B. V. Ristianto, I. A. Prayogo, Nasrullah, and A. Saifudin, "Pengujian perangkat lunak sistem informasi penilaian mahasiswa dengan teknik boundary value analysis menggunakan metode black box testing," *J. Artif. Intell. Innov. Appl.*, vol. 1, no. 1, pp. 32–36, 2020.