

Indonesian Journal of Computer Science

ISSN 2549-7286 (*online*) Jln. Khatib Sulaiman Dalam No. 1, Padang, Indonesia Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Fake News Detection Using Machine Learning

Pshko Rasul Mohammed Amin

Pshko.amin@uor.edu.krd University of Raparin

Article Information	Abstract	
Submitted : 27 Jul 2023 Reviewed: 1 Ags 2023 Accepted : 17 Ags 2023	Today one of the fattest environments for opinion exchanging is the internet. Individuals can share their data or post any news they want through social media platforms. these data sharing platforms do not verify the users or the data and information they post. so, some users can share fake or untrusted	
Keywords	data through these platforms on the Internet. Fake news is described as a propaganda tool against any Individual, society, government, or political	
Fake News, Machine Learning, Social Data, Classifiers	party. Humans cannot detect and understand all the fake news or data through these online platforms. In this study, the challenge has been defined through a Machine learning concept. machine learning classification was applied to the dataset. Finally, a comparison of the working of these classifiers is presented along with the results, Decision Tree and Support Vector Machine classification models are performing well.	

A. Introduction

The world is enhancing concurrently. There are undoubtedly many benefits to living in a digital age, but there are also drawbacks. Different problems exist in this digital age. Fake news is among them. False news is simple to distribute. Spreading fake news to damage a person's reputation either a person or a business. A political party or other entity may be the target of the propaganda. There are numerous online channels where one might propagate false information. Among them are Facebook, Twitter, etc. Artificial intelligence's machine learning component aids in creating systems that can learn and carry out various tasks [1].

Various machine learning algorithms are available. B. Supervised, Unsupervised, and Reinforcement Learning Methods. To train your algorithm, you first need to use a data set called the training data set. Many tasks can be performed with these algorithms after training. Machine learning is used in various industries to perform various tasks. Machine learning algorithms are often used to make predictions and find hidden things. Users benefit from online platforms because they can quickly access news. The issue is that this offers online criminals the chance to use these sites to distribute false information. An individual or society may be harmed by this news. Readers read the news and start believing it before it is verified. Identifying fake news is not easy and presents a significant challenge [2]. If fake news isn't debunked soon, people will spread it and everyone will start believing it. Fake news can influence people, groups, and political parties. Fake news during the 2016 US elections influenced public opinion and judgment [3]. Various scholars are trying to identify fake news. Machine learning can help here. Researchers use various algorithms to identify fake news. Researchers [4] argue that fake news is difficult to identify.

The article's author [6] explored ways to make the ANN algorithm more efficient to yield better results. The results are similarly excellent compared to using an evolving genetic algorithm to select the most accurate nonlinear function parameters appropriate for each trait. Preeti Nair and Indu Kashyap in their paper [7] standardize the data input to the classifier and improve the performance of the algorithm by introducing resampling and interquartile range (IQR) techniques in the preprocessing step.

In a study [8], the authors developed a model that uses user social site visits and headline data to detect fake news. In an article, Nagashri and J. Sangeetha found TFIDF to be a good text preprocessing technique for fake news identification [9]. They used counting vector techniques and various machine learning concepts and evaluated them with accuracy, precision, recall, and F1 score.

In their paper [10], the authors describe the relationship between words and the environment in which they appear in the text, and distinguish this relationship between real text (negative case) and invented text (positive case). We explored how it can be used to Models such as Count Vectorize were used to convert character-based text to numeric representations. Next, we considered which model was best at recognizing whether the content was authentic. In a study [11], Shlok Gilda used probabilistic context-free grammar (PCFG). Machine learning classification techniques are used to identify fake news. His accuracy rating was 77.2%. A study by Julio C.S. Reis et al. This study explored several ideas for identifying fake news. Message parameters are, for example, the title of the message and its source. Additionally, the model is built using the title and source as building blocks, and her SVM classifier is used for classification. The Lustum algorithm is used for forecasting. And the author reported that the accuracy of his algorithm is 87%. However, a careful examination of this structure reveals many instances of incorrect comparisons. The impossibility of identifying fake news based on headlines and sources alone seems to be an obstacle to proper evaluation[7].

Machine learning was used to identify fake news. Researchers [5] found that the prevalence of fake news increases over time. This is why it is important to recognize fake news. Machine learning algorithms have been developed for this purpose. Once machine learning algorithms are trained, they will be able to detect fake news on their own. Through this literature review, many research questions will be resolved. This literature review demonstrates the value of machine learning in identifying fake news. The application also covers machine learning for detecting fake news. Describes machine learning techniques used to identify fake The this document is organized follows: news. rest of as a literature review, methodology, results, and discussion, the conclusions were drawn.

B. Research Method

The In this study, we used Python programming language as the main functional programming language. Python provides many useful libraries for data processing and machine learning. In the sections below the explanation about the proposed system has been provided.

a) In the first step we have to import the required libraries for the system to be functional, the below table shows us the libraries and their purpose of using.

Libraries	Functionality
Pandas	Data structures
NumPy	Array, Linear Algebra
seaborn	Plot graphs
matplotlib	Integrating charts
sklearn	Classification models
Train_test_split ()	Prediction algorithms
Accuracy_sccore	Classification accuracy
	measurement
Classification_report	Assess the accuracy of
	classification
re	Regular expression test
string	Examining unstructured text

Table 1. Python L	ibraries
-------------------	----------

As shown in the above table, these libraries have been used in our fake news detection system. Pandas are a library that contains data structure algorithms. Numpy is also a useful Python library that contains an array and linear algebra data structures, seaborn is used to plot graphs. Matplotlib is used to integrate charts. Sklearn is the most useful Python library which contains machine learning classification models. The train test is used to make predictions regarding classification models. The accuracy score library was used to measure the accuracy of classification models in this study, re is the library used to determine whether the given string is a regular expression or not.

b) Importing the required datasets, for this system we have collected the data online from different social media platforms. For this study, we have two collections of data placed in. excel file, one of the files is the true data or trusted data, and the other file is the fake data. The labels of the dataset are (ID, title, text, subject, and date). The following figures below show the structure of the data.

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'	Donald Trump just couldn t wish all Americans	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian	House Intelligence Committee Chairman Devin Nu	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke	On Friday, it was revealed that former Milwauk	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name	On Christmas day, Donald Trump announced that	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur	Pope Francis used his annual Christmas Day mes	News	December 25, 2017

Figure 1.Fake news data

	title	text	subject	date
0	As U.S. budget fight looms, Republicans flip t	WASHINGTON (Reuters) - The head of a conservat	politicsNews	December 31, 2017
1	U.S. military to accept transgender recruits o	WASHINGTON (Reuters) - Transgender people will	politicsNews	December 29, 2017
2	Senior U.S. Republican senator: 'Let Mr. Muell	WASHINGTON (Reuters) - The special counsel inv	politicsNews	December 31, 2017
3	FBI Russia probe helped by Australian diplomat	WASHINGTON (Reuters) - Trump campaign adviser	politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor	SEATTLE/WASHINGTON (Reuters) - President Donal	politicsNews	December 29, 2017

Figure 2. True news data

The next steps are assigning classes to each dataset, and checking the number of rows and columns in each dataset. Then we have to merge both datasets and perform some text processing algorithms, which they call preprocessing techniques by doing this we can eliminate or clean up unnecessary data in both datasets. Finally, defining training and testing data and splitting them. The models of classification we have used in this study are Logistic Regression, DecisionTreeClassifier-Nearest Neighbor, Support Vector Machine, and Naïve Bayes classifiers.

C. Result and Discussion

In this section, after implementing the machine learning algorithms, and using different classification models, the results are described below. Firstly, while processing the dataset to detect fake news for the initial step it is important to distinguish the type of the news according to different categories. The following table and figure below show the type of news and the frequency of the news according to different categories.

1	Government News	1570
2	Middle-east	778
3	News	9050
4	US_News	783
5	left-news	4459
6	politics	6841
7	politicsNews	11272
8	World news	10145

Tab	le	2.	News	Categories
Iup	IC.	4.	110 113	Gategories



Figure 3. News Categories

Secondly, after the most frequent word counter according to different categories, the determination of the most frequent word in fake and real news was conducted. The following table and figure below show the dataset's count of fake and real news words after preprocessing steps.

Table 3. Real & Fake news word frequency			
Real	21417		
Fake	23481		



Figure 4. Real & Fake news word frequency

Finally, using machine learning classification models, the assessment of the trained data according to different classification models has been provided. The following table below shows the accuracy of each model on the trained dataset to detect fake news.

Tab	le 4. Classification Models & Accu	iracy
ID	Model	Accuracy
1	Naive Bayes	95.68%
2	Logistic regression	99.14%
3	Decision Tree	99.7%
4	Random Forest	99.18%
5	SVM	99.71%



By looking at the table above, we can see that Decision Tree and Support Vector Machine classification models are performing well. The following figure below shows the confusion matrix of all classification models and their accuracy.

Figure 5. Confusion Matrix for Classification Models

D. Conclusion

Fake news can be defined as untrusted knowledge spreading across different online platforms. This fake news can be against an individual or company or a political party. Today through advancements in technology and specifically in the field of artificial intelligence we can address this challenge. In this paper, the challenge has been described and by using machine learning a valuable approach to detecting fake news and applying different machine learning classification models we assessed the accuracy of each model. And we can see that Decision Tree and Support Vector Machine classification models are performing well.

E. References

- [1] Donepudi, P. K. (2019). Automation and Machine Learning in Transforming the Financial Industry. Asian Business Review, 9(3), 129-138. https://doi.org/10.18034/abr.v9i3.494.
- [2] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective.ACM SIGKDD Explorations Newsletter, 19(1), 22-36. https://doi.org/10.1145/3137597.3137600H. Poor, An Introduction to Signal Detection and Estimation. New York: Springer-Verlag, 1985, ch. 4.
- [3] Dewey, C. (2016). Facebook has repeatedly trended fake news since firing its human editors. The Washington Post, Oct. 12, 2016.
- [4] Wang, W. Y. (2017). "Liar, liar pants on fire": A new benchmark dataset for fake news detection. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2(Short Papers), 422–426. http://dx.doi.org/10.18653/v1/P17-2067
- [5] Zhou, X., Zafarani, R., Shu, K., & Liu, H.(2019).Fake News: Fundamental theories, detection strategies, and challenges. InWSDM 2019 -Proceedings of the 12th ACM International Conference on Web Search and Data Mining(pp. 836-837). (WSDM 2019 -Proceedings of the 12th ACM International Conference on Web Search and Data Mining). Association for Computing Machinery, Inc.https://doi.org/10.1145/3289600.3291382
- [6] F. Sanei, A. Harifi and S. Golzari, "Improving the precision of KNN classifier using nonlinear weighting method based on the spline interpolation," 2017 7th International Conference on Computer and Knowledge Engineering (ICCKE), 2017, pp. 289-292.
- [7] P. Nair and I. Kashyap, "Hybrid Pre-processing Technique for Handling Imbalanced Data and Detecting Outliers for KNN Classifier," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019, pp. 460-464.
- [8] Kesarwani, A., Chauhan, S.S., Nair, A.R. and Verma, G., 2021. Supervised Machine Learning Algorithms for Fake News Detection. InAdvances in Communication and Computational Technology (pp. 767-778). Springer, Singapore.
- [9] A. Kesarwani, S. S. Chauhan and A. R. Nair, "Fake News Detection on Social Media using K- Nearest Neighbor Classifier," 2020 International Conference on Advances in Computing andCommunication Engineering (ICACCE), 2020, pp. 1-4,10.1109/ICACCE49060.2020.9154997.
- [10] Nagashri, K. and Sangeetha, J., 2021. Fake News Detection Using Passive-AggressiveClassifierand Other Machine Learning Algorithms. InAdvances in Computing and Network Communications (pp. 221-233). Springer, Singapore.
- [11] Vijayaraghavan, S., Wang, Y., Guo, Z., Voong, J., Xu, W., Nasseri, A., Cai, J., Li, L., Vuong, K. and Wadhwa, E., 2020. Fake news detection with different models. arXiv preprint arXiv:2003.04978.