

# Fake Article Detection Application with Rabin Karp Algorithm

Mawaddah Harahap<sup>1)\*</sup>, Rico Marcel<sup>2)</sup>, Dwi Yunita Milatrisna<sup>3)</sup>, Sri Ajeng Kuswulandari<sup>4)</sup>

<sup>1)2)3)4)5)</sup> Universitas Prima Indonesia, Sumatera Utara, Indonesia

<sup>1)</sup>[mawaddah@unprimdn.ac.id](mailto:mawaddah@unprimdn.ac.id), <sup>2)</sup>[marcelrico7@gmail.com](mailto:marcelrico7@gmail.com), <sup>3)</sup>[dwiynmt@gmail.com](mailto:dwiynmt@gmail.com), <sup>4)</sup>[kuswulandarisri@gmail.com](mailto:kuswulandarisri@gmail.com)

**Submitted** : Nov 28, 2022 | **Accepted** : Dec 16, 2022 | **Published** : Jan 6, 2023

**Abstract:** Currently, fake information/articles are increasingly widespread. Surveys show that people receive fakes more than once a day. The most used channel for spreading fakes is social media. The fake phenomenon in Indonesia raises doubts about the information received and confuses the public. Artificial intelligence is an intelligent machine (computer), where a machine is trained in advance (machine learning) to solve certain problems. This artificial intelligence is expected to work faster and more accurately. There are several methods that can be implemented, one of which is using the Rabin-Karp algorithm. Where the reason for choosing the Rabin karp algorithm is that there is a process that can filter existing writing so that it is suitable for the problems discussed, namely article writing and an algorithm that can work at high speed so that users do not need a long time to receive predictive results for articles.

**Keywords:** Article detection, Fake article, Hoax, classification, Rabin Karp

## INTRODUCTION

The mass media must be seen as an institution that is free of values and conveys reality as it is. The media has the power to construct reality as it is in society. Because of this, the media must be balanced in reporting every event that occurs in society (Kaila & Bait, 2021). Currently, the dissemination of information is mainly done through network media. The ease of delivering and mediating information in the network to the public makes information or articles not filterable properly. No editor is responsible for the dissemination of information disseminated by online media, because anyone who has the rights to trade online media data can disseminate information. Large amounts of anonymous information allow hoaxes to spread quickly in online media. False information/articles are becoming more and more common these days. Surveys show that people receive fakes more than once a day. The most used channel for spreading fake news is social media. Counterfeiting in Indonesia raises doubts about the information received and confuses the public (Rahmadhany, Safitri & Irwansyah, 2021).

Social media is a very vulnerable platform and is often used as a place to spread fake articles. The large number of active users can even be said to be social media enthusiasts in Indonesia, which makes it very easy for fake spreaders to carry out their actions (Masi, Lestahulu & Bandjar, 2022). Fake articles are very difficult to identify, because they are well structured and equipped with convincing fake evidence. These fake articles can easily influence readers and believe and do what the creator of the fake article wants (Batoebara, Suyani & Nuraflah, 2020). The spread was so fast and wide, of course it was very dangerous and caused misunderstanding and uproar in society. Even if the government does not intervene and handle it in a timely manner, it will threaten the unity and integrity of the country. It seems very difficult to identify these fake articles manually. Therefore, an artificial intelligence (AI) approach is needed. Artificial intelligence is an intelligent machine (computer) where the machine is pre-trained (machine learning) to solve certain problems. This AI promises to work faster and more accurately (Bahri & Wajhillah, 2020). There are several methods that can be implemented, one of which is using the Rabin-Karp algorithm. The Rabin-Karp algorithm is one of the string matching algorithms that can be used to measure the level of text similarity (Aldian & Mubarak, 2022). In computer science, predicting how likely an article is fake can be done using language modeling. One of the modeling languages is the Rabin Karp algorithm. The reason for choosing the Rabin Karp algorithm is to have a process for filtering existing posts to match the problem being discussed, namely writing articles and an algorithm that can work at high speed so that users don't take long to receive predictive results. article. Rabin Karp was able to solve the problems that exist in LSA, which are related to vector dimensions, because Rabin Karp works directly by matching text using the hash as a parameter to measure the level of text similarity (Bahri & Wajhillah, 2020).

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Based on the frequent rapid spread of fake articles that have occurred recently, there are lots of ordinary people who immediately forward the fake articles as a sign of validating an article first, so that with an application that can provide fake or non-false values for an article it is really needed by ordinary people today, so that researchers are interested in doing research, as well as reducing the negative impact caused by these fake articles.

### LITERATURE REVIEW

In accordance with the title Strategy for Countering False Information on Social Media by the Cyber Crime Unit in Makassar City by Tirta Raharja 2020, the data analysis technique used in this study is descriptive and qualitative. The steps for processing the data are as follows. The data from the observations summarize the potentials and problems. The data obtained from the interviews were in the form of recorded interviews, narrated and edited to produce good and correct grammar. The narratives and results of the interviews were analyzed based on existing theories and concepts, followed by drawing conclusions. The findings indicate that the Makassar City Cybercrime Unit's strategy to combat social media disinformation is one of its kind. Conduct outreach at schools, campuses according to the network membership program b. Establish cooperation with various stakeholders c. Direct supervision of online social media, especially online social media on behalf of network patrols (Raharja, 2020).

According to Haidar Ihzaulhaq's 2021 paper entitled Application of Hashing in the Rabin-Karp Algorithm to determine Keywords in a Text, the application of hashing to the Rabin-Karp algorithm can be used to do things related to words in a text, where one of them is keyword search. With the definition of a keyword, a word that occurs frequently and is unique in the text, a program can be made to search for it. Using the Rabin-Karp algorithm also makes keyword searches faster compared to other algorithms such as brute force (Ihzaulhaq, 2021)

Aldian, Mubarak 2022 entitled "Implementation of the Rabin-Karp Algorithm to Detect Plagiarism in Web-Based Text Document Files" states that according to designing and building an application using the Rabin-Karp algorithm to find textual forms that are tested for the percentage of similarity. Based on the results of testing the original documents and documents tested, among the results of testing 10 text documents using the Rabin-Karp algorithm, the highest accuracy rate is 47.58%. And the smallest accuracy is 19.28% (Aldian & Mubarak, 2022).

### METHOD

At this stage, the manufacturing process includes designing a flowchart to find out the flow of the program from start to finish, designing a database to store data for evacuation sites and lists of missing persons, designing program interfaces. The work procedures for applications designed include the following:

1. Processing Datasets

The initial dataset given was 65 datasets taken from the website <http://turnbackhoax.id/>, these datasets can be added via the dataset menu, and simply fill in the written articles in the richtextbox provided and press the save button to save data, the edit button to change the writing articles that have been saved, the delete button to delete the existing writing articles.

2. *Look for text similarities*

In the process of searching for text similarity, programs designed using the karp method. Where the process steps that occur include: Select a text file with the .txt extension and automatically the text in the .txt file will appear in the application; Retrieval of datasets that have been stored in the database; Make all writing into uppercase so that articles that are in upper and lower case have no effect; Repeating the length of the article sentence to be processed; Doing iterative recalculation of the length of the article sentence to be processed; Checking the similarity of the written article database in the database with the previously selected writing articles

### RESULT

Hoax datasets are datasets that have been confirmed as articles that are considered fake/hoax articles. The hoax dataset used in this study is in the appendix.

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

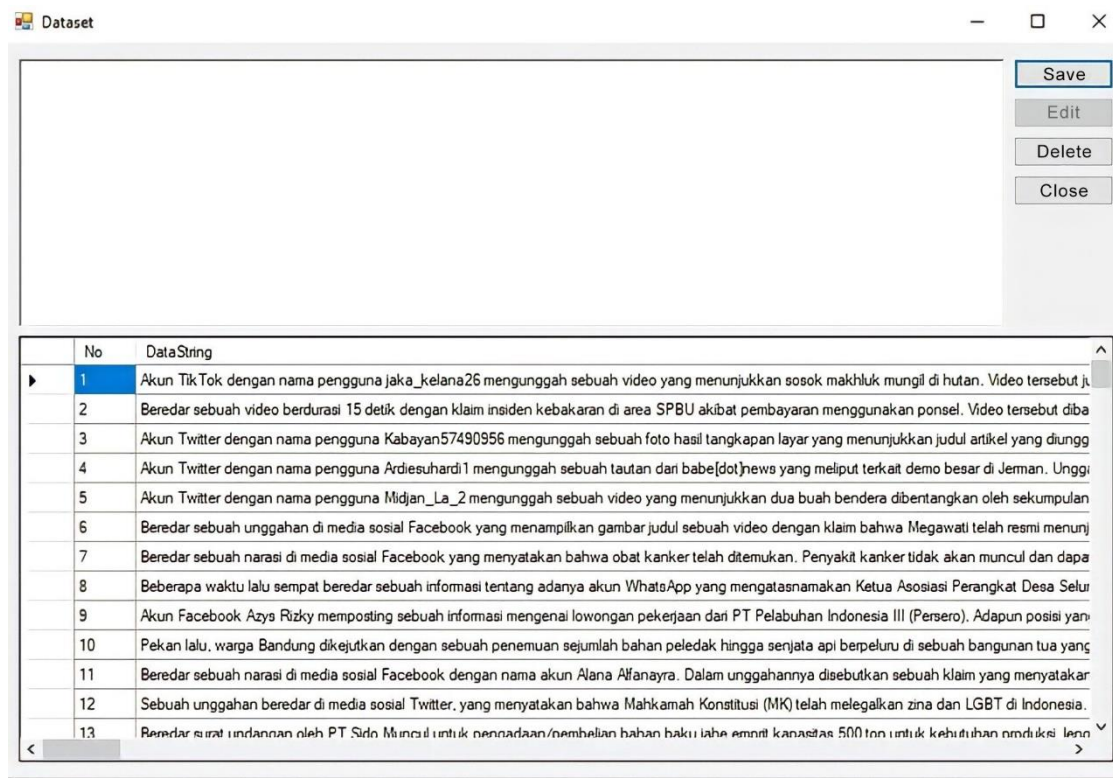


Figure 1. Program dataset display

In the program dataset display, users can add fake news by inputting data in the textbox provided and pressing the save button to save the news that has been inputted. After that, users can also change news by pressing the datagrid at the bottom, and pressing the edit button after the user changes the news. there, the user can also delete news data that has been stored by pressing datagrid then pressing the delete button then the selected data will be deleted, and will not be used for calculations using rabin karp.

The application display when performing the calculation process using the Rabin Karp method is shown in the following figure:

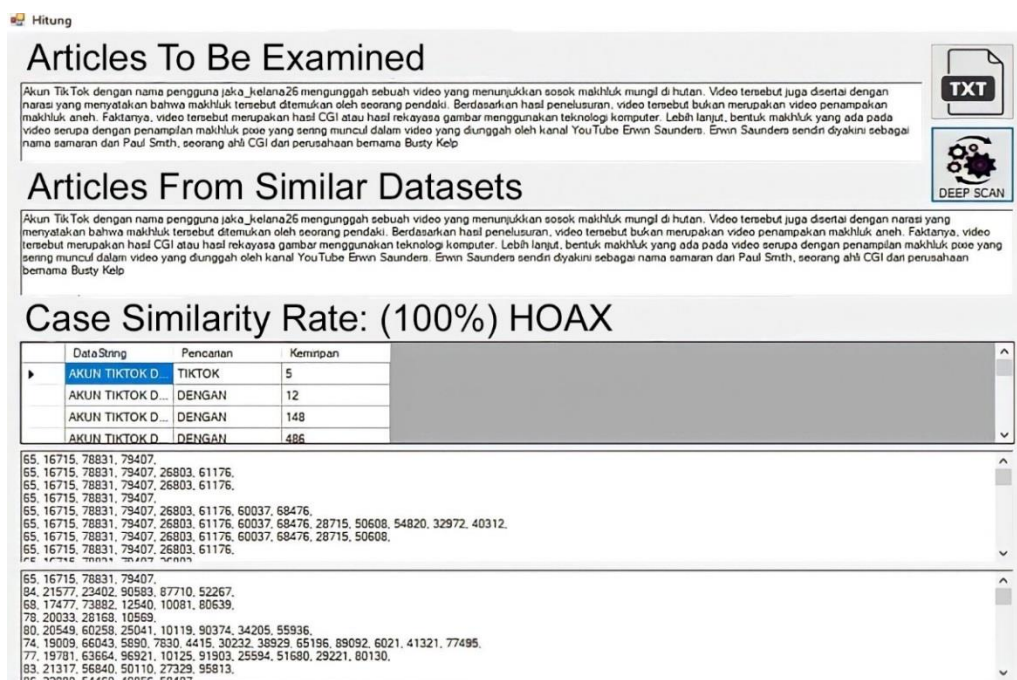


Figure 2. The Calculation Process Using Rabin Karp Display

\*name of corresponding author



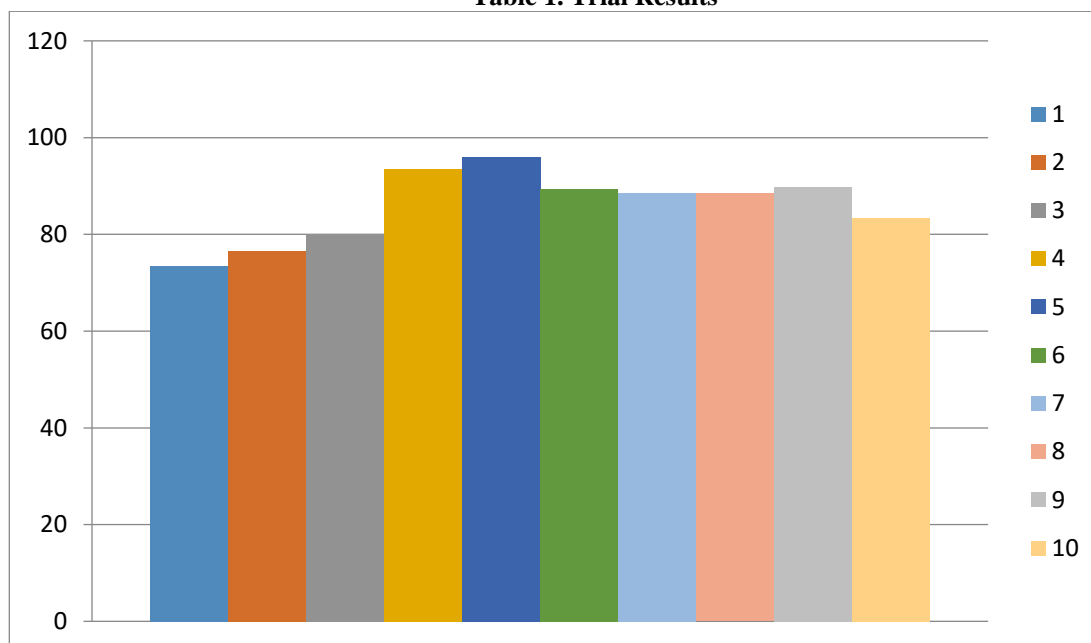
This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

The calculation process consists of several steps, namely:

1. Selection of a text file with the extension .txt where in the textfile there are sentences that will be calculated using rabin karp against the scattered hoax dataset.
2. After the news appears from the textfile previously selected by the user, the user can use the process button to start calculating the fake news dataset against the selected file.
3. Each word in the previously selected news will be compared to the hoax dataset using the rabin karp formula in the application, for example in the sentence "A TikTok account with the username jaka\_kelana26 uploaded a video showing a tiny creature in the forest." Each word starts from "Account", "Tiktok", "with", "name", "user", "jaka\_kelana26", "upload", "a", "video", "which", "shows", "figure", "creature", "tiny", "in", and "jungle" will be compared to the dummy article dataset.
4. After the calculation, datasets that have a similarity value will appear, where the iteration of the dataset will be recalculated to record the level of similarity.

The results of trials conducted by research using the Rabin Karp method for hoax datasets include:

**Table 1. Trial Results**



No.	ID Artikel	Similarity Value
1	1	73.4%
2	2	76.53%
3	3	80.12%
4	4	93.5%
5	5	95.93%
6	6	89.33%
7	7	88.48%
8	8	88.43%
9	9	89.83%
10	10	83.33%

### DISCUSSIONS

This research process can implement the validation of all existing functions. The rabin-karp algorithm has advantages in detecting the similarity of news articles. Where when doing user detection only enter the textfile that will be tested. And this process will filter unnecessary words or symbols so that it will display the results of the detection calculations. This is very useful for reducing errors in detecting fake articles to reduce the spread of wrong articles.

### CONCLUSION

Based on the results of the analysis and application that has been carried out in the previous chapter, it can be concluded that: A system has been created that can be used to detect similarities to a hoax document using the Rabin-Karp algorithm. Plagiarism can be done by changing some parts or even the whole by changing the words

\*name of corresponding author



with synonyms instead of those words. The Rabin-Karp algorithm that uses the Synonym Recognition method has a more accurate determination of similarity percentage values when compared to the Rabin-Karp algorithm without Synonym Recognition. Because even though the words in the test document are changed in such a way, they have the same meaning and will be detected by this system. For base values, not all numbers can be used. Because in some cases the wrong base value will result in the resulting hash value being the same as another hash value that has a different word. The test document is declared duplicate if the Similarity percentage value is above 50%.

#### REFERENCES

- Aldian & Mubarak (2022). Implementasi Algoritma Rabin-Karp Untuk Pendeteksian Plagiarisme Pada File Dokumen Berupa Text Berbasis Web. *Journal of Information System Research (JOSH)*. 3(3):150-154
- Bahri, S. & Wajhillah, R (2020). Optimalisasi Algoritma Rabin Karp menggunakan TF-IDF Dalam Pencocokan Text Pada Penilaian Ujian Essay Otomatis. *InfoTekJar*. 4(2): 292–295  
[Online]. Available: <https://jurnal.uisu.ac.id/index.php/infotekjar/article/view/2294>
- Batoebara, M. U., Suyani, E. & Nuraflah, E.S (2020). Literasi Media Dalam Menaggulangi Berita Hoaks. *Jurnal. Warta Edisi* 63,14(1): 34–41.  
[Online]. Available: <https://jurnal.dharmawangsa.ac.id/index.php/juwarta/article/view/541>
- Ihzaulhaq, H (2021). Penerapan Hashing dalam Algoritma Rabin-Karp untuk Menentukan Kata Kunci dalam Suatu Teks. Institut Teknologi Bandung
- Kalia, S. C., & Baits (2021). Analisis Penyebaran Berita Hoaks Pandemi Covid-19 Di Bondowoso Melalui Facebook. *Dr. Diss. Univ. Muhammadiyah Jember*, 20(1): 1–16
- Masi, W. F. L., Lestahulu, S., & Bandjar, A. (2022). Pola Penyebaran Hoax Melalui Media Sosial (Studi Kasus Penyebaran Informasi Wabah Covid-19 di Kota Ambon. 1(1): 30-37
- Raharja, T (2020). Strategi Penanggulangan Informasi Hoax Di Media Sosial Oleh Unit Cyber Crime Di Kota Makassar. Universitas Muhammadiyah. Makassar.
- Rahmadhany, A., Safitri, A. A., & Irwansyah (2021). Fenomena Penyebaran Hoax dan Hate Speech pada Media Sosial. *J. Teknol. Dan Sist. Inf. Bisnis*, 3(1): 30–43.

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.