

Analyzing radicalism sentiments in Indonesian da'wah content on website da'wah through text mining techniques

Aulia Aziza^{1,2,3}, Risqiatul Hasanah¹, Juairiah², Munsyi³

¹Islamic Communication and Broadcasting, Faculty of Da'wah and Communication Sciences, UIN Antasari, Banjarmasin, Indonesia

²Islamic Library Science, Faculty of Tarbiyah and Teaching, UIN Antasari, Banjarmasin, Indonesia

³Department of Information Technology, Faculty of Da'wah and Communication Sciences, UIN Antasari, Banjarmasin, Indonesia

Article Info

Article history:

Received Aug 2, 2024

Revised Dec 2, 2024

Accepted Dec 15, 2024

Keywords:

Classification

K-nearest neighbors

K-NN algorithm

Radical content

Text mining

Website da'wah

ABSTRACT

This study investigates the classification of radical content in Indonesian Da'wah websites using text mining techniques. A content search engine application, developed with PHP, processes queries by comparing results against a database of keywords, classifying content into four categories: red, yellow, green, and white. Manual labeling based on data from the Ministry of Communication and Informatics yielded 126 labeled articles, forming the dataset for classification. The K-nearest neighbors (K-NN) algorithm, with an optimal k value of 7, achieved a classification accuracy of 66.37%, demonstrating its reliability compared to manual methods. The "White" class showed the highest precision and recall. System testing revealed efficient performance, with 0.704 seconds per classification task and 884,656 bytes of memory usage. Future enhancements include incorporating synonym identification for Indonesian keywords and exploring machine learning algorithms such as Naive Bayes and neural networks to improve accuracy. This research highlights the potential for text mining in identifying online radical content while emphasizing the need for system adaptability.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Munsyi

Department of Information Technology, Faculty of Da'wah and Communication Sciences

UIN Antasari, Banjarmasin, Indonesia

Email: munsyi@uin-antasari.ac.id

1. INTRODUCTION

The swift development of internet technology has brought about a global society that transcends geographic boundaries. This change has markedly boosted online interactions, particularly on website da'wah platforms like Website Da'wah [1], [2]. While the extensive use of website da'wah has numerous benefits, it also has the potential to steer individuals toward actions that deviate from societal norms. In addition to the many advantages of information technology, there has been an increase in crimes facilitated by these digital tools [3], [4]. This problem has captured the attention of researchers in fields such as psychology, social science, and computer science, who seek to create methods to detect and prevent these activities. Radical content that incites violence, spreads hatred, and promotes anti-nationalism is a major concern. According to Indonesia's National Counterterrorism Agency (BNPT) [5], radical content encourages violence in the name of religion, incites suicide, and advocates jihad against those labeled as infidels. In Indonesia, religious radicalism is deeply connected to the internal dynamics and challenges faced by Indonesian Muslims, both domestically and internationally [6], [7]. There is significant internal polarization among the various ethnic and cultural groups within the Muslim community. In the face of pressures such as secularization, capitalism, and cultural denigration, website da'wah has become a potent tool for the spread of radical ideas. This dissemination significantly affects Muslim youth at various levels, including their discourse, thought

processes, and actions [8]. Online platforms facilitate the spread of radical ideologies, with recruitment for radical Islamic groups often taking place through website da'wah. Documented cases show that website da'wah has been used to organize radical activities, including armed attacks and the use of explosives [3].

This situation is particularly concerning given that 87% of Indonesia's population is Muslim [9]. Indonesia's low media literacy ranking, placed 60th out of 61 countries according to the CCSU World Ranking of Nations, further exacerbates the issue [10]. The spread of radical Islamic ideas through da'wah websites presents a significant threat to the integrity and stability of the Indonesian nation and the Unitary State of the Republic of Indonesia (NKRI). Therefore, it is crucial to improve media literacy skills among Muslim youth, especially Islamic activist students, to counter the influence of radical Islamic content disseminated through da'wah website networks. The convergence of website da'wah and the rise of radicalism has created a complex landscape for policymakers, educators, and researchers in Indonesia. The extensive use of platforms like Website Da'wah has not only amplified the reach of radical content but has also facilitated the recruitment and radicalization of vulnerable individuals. This phenomenon is particularly pronounced in Indonesia, where the large Muslim population and low media literacy levels create fertile ground for the spread of extremist ideologies [11]. The Indonesian government and various non-governmental organizations have recognized the threat posed by online radicalism and have initiated several countermeasures. However, the rapid evolution of website da'wah technologies and the sophisticated methods employed by radical groups necessitate continuous adaptation and improvement of these strategies. Efforts to counter radical content on website da'wah must be multi-faceted, involving not only technological solutions but also educational and community-based approaches [12]. Enhancing media literacy is a critical component of these efforts. By equipping young Muslims, particularly those involved in Islamic activism, with the skills to critically analyze and interpret online content, it is possible to reduce the influence of radical ideologies. Educational programs that focus on critical thinking, digital literacy, and the responsible use of website da'wah are essential in building resilience against radicalization [11], [13].

Additionally, collaboration between government agencies, educational institutions, and technology companies is vital in developing effective counter-radicalization strategies. This includes monitoring and analyzing online content, identifying and removing radical material, and promoting positive and moderate narratives that counteract extremist messages. Research plays a crucial role in informing these efforts [14]. By utilizing text mining techniques and sentiment analysis, researchers can gain valuable insights into the nature and dynamics of radical content on website da'wah. This information can help in developing targeted interventions that address the specific needs and vulnerabilities of the Indonesian Muslim youth. Technological advancements, particularly the Internet, have become a primary need, significantly changing societal behavior globally [13]. The Internet has dissolved geographical boundaries and rapidly instigated significant social changes. However, this advancement also poses risks, facilitating negative activities and unlawful actions. The sophistication of this technology can lead individuals to act against prevailing social norms, creating a new global society without territorial limitations. Despite the numerous benefits of information technology, its dark side has seen a rise in crimes committed through technological means. One such crime is cyber terrorism, where the Internet is misused for terrorist activities. Cyber terrorism involves politically motivated attacks using technology, computer networks, and technical infrastructure to cause harm. Experts argue that cyber terrorism can be more dangerous than traditional terrorism [5], [15]. The rise of online radicalization, also known as cyber-terrorism, cyber-racism, or cyber hate, has become a significant concern for communities, governments, and law enforcement agencies worldwide [16].

Cyber racism is a type of cybercrime characterized by radical behavior, which varies across nations and is a topic of multidisciplinary research. Radical content, which incites violence, spreads hatred, and promotes anti-nationalism, is of particular concern. For BNPT, radical content includes advocating violence in the name of religion, interpreting jihad as suicide bombing, and denouncing others as infidels. The criteria for radicalism, according to BNPT, include:

- a) Advocating rapid change through violence in the name of religion.
- b) Denouncing others as infidels.
- c) Supporting, spreading, and inviting others to join ISIS/IS.
- d) Interpreting jihad narrowly.

Researchers face challenges in detecting radical online content due to the large volume of unstructured, user-generated content that changes dynamically. The secretive nature of radical content also complicates detection, making traditional web crawlers ineffective. Studies suggest using Text Mining and Web Scraping methods to classify radical content [17]. Research has shown that various Internet platforms can be easily misused for harmful purposes, forming hate groups, spreading extremist ideologies, and inciting violence. Developed a system to detect users accessing terrorist websites by monitoring and analyzing web content in real-time [18]. Another study by [19] proposed an efficient algorithm to destabilize terrorist networks by uncovering hidden hierarchies within these networks using social network analysis. Introduced a

semi-automatic approach for detecting radical content using the link-based bootstrap (LBB) algorithm, which identifies and expands authoritative URLs to capture related radical content. In conclusion, the intersection of website da'wah and radicalism presents a significant challenge for Indonesia. Addressing this issue requires a comprehensive approach that combines technological, educational, and community-based strategies. Enhancing media literacy, fostering collaboration, and leveraging research are key elements in the fight against online radicalism, ensuring the stability and integrity of Indonesian society.

2. RESEARCH METHOD

The methods and techniques used in this research explain the steps or process of developing an application for detecting radical content. This application helps users identify radical and non-radical content in Indonesia. The system design for this research is shown in Figure 1.

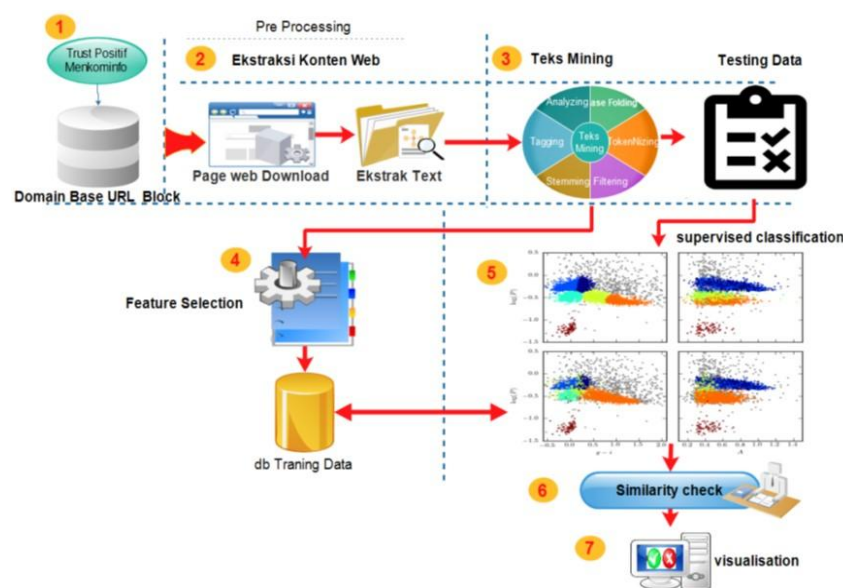


Figure 1. Radical content research system design

As shown in Figure 1, the system workflow starts by gathering news content data from the ministry of communication and information's approved internet sources. Next, web content extraction removes HTML tags, leaving only the text. Then, text mining identifies keywords through several stages: case folding, tokenization, filtering, stemming, tagging, and analyzing, which result in a set of keywords [20], [21]. These keywords are stored in a database and used for feature selection to reduce data dimensionality. The classification process follows, determining which keywords correspond to class labels provided by the ministry of communication and information.

Web mining applies data mining techniques to analyze web data and extract meaningful information. It aims to transform large amounts of web data into actionable insights. In the context of this system, web mining is key for processing and identifying patterns in the extracted text data. Web mining can be categorized into three types: content mining, structure mining, and usage mining. Content mining extracts valuable data from web documents using techniques like NLP, IR, and ML. Structure mining analyzes hyperlink relationships to improve page ranking and navigation. Usage mining focuses on analyzing web logs to understand user behavior and improve personalization and adaptability.

The web mining process includes data collection, preprocessing, pattern discovery, evaluation, and deployment [22]. Data sources include web pages, logs, and user profiles, with preprocessing involving data cleaning and transformation [23]. After applying mining algorithms to discover patterns, these insights are evaluated and used in real-world applications. Despite its potential, web mining faces challenges such as inconsistent, noisy data, the need for scalable techniques, privacy concerns, and the dynamic nature of the web [24]. Nevertheless, it remains a powerful tool for enhancing web services and user experiences [25]. Text mining, applying data mining to text data, extracts patterns from unstructured text by converting it into a structured format suitable for analysis [26].

A. Key preprocessing steps:

– Case folding

Definition: case folding is the process of converting all characters in a text document to a standard case, usually lowercase. This step ensures consistency in text processing since text documents may not use capital letters consistently.

– Tokenization

Tokenization is the process of dividing the input text into individual words or tokens, while removing punctuation, numbers, and non-alphabetic characters, which serve as delimiters and do not add value in text processing. For example, the sentence “Text mining, also known as text analytics, is valuable.” would be tokenized into [“text”, “mining”, “also”, “known”, “as”, “text”, “analytics”, “is”, “valuable”].

– Filtering

Filtering involves selecting significant words from the tokens created during tokenization by eliminating common words, known as stop words, that do not provide meaningful content, such as “is”, “the”, and “and”. For instance, from the tokens [“text”, “mining”, “also”, “known”, “as”, “text”, “analytics”, “is”, “valuable”], the stop words “also”, “as”, and “is” would be removed.

– Stemming

Stemming refers to the process of reducing words to their root form to standardize variations of words, improving the efficiency of text mining. This step converts inflected or derived words into their base form. Common stemming algorithms include the Porter Stemmer for English and Arifin-Setiono Stemmer for Indonesian. For example, the words “running”, “runs”, and “ran” would be reduced to the root word “run”.

– Tagging

Tagging is the step of assigning labels or categories to words, usually after stemming, to identify their part of speech or other semantic properties. This helps to further clarify the role of each word within the context of the text.

– Analyzing

Definition: analyzing determines the relationships between words across documents. The simplest algorithm for scoring in text mining is term frequency-inverse document frequency (TF-IDF). Example: calculating the TF-IDF scores for words in a document helps identify the importance of each word relative to the document corpus.

B. Classification

Definition: classification is a data mining technique used to solve real-world problems by learning patterns from historical data to categorize new instances into predefined groups or classes. Methodology: the typical two-step process includes ‘model development/training’ followed by ‘model testing/deployment’. For example, algorithms like K-nearest neighbors (K-NN), decision trees, and others are used to classify data. The K-NN algorithm classifies new data based on its similarity to labeled data, usually measured by Euclidean distance. Through text mining, these preprocessing and classification steps convert unstructured text data into structured, valuable insights that can be applied in areas such as business intelligence, customer sentiment analysis, and more.

In this research, the research and development (R&D) method is used, which is a systematic approach aimed at improving or developing new products through rigorous testing to ensure their reliability. The study presents theories and concepts relevant to the problem and scope of the discussion, drawing on references from diverse fields such as politics, sociology, culture, economics, and information technology. The research explores fundamental and radical definitions from these perspectives, along with sentiment analysis and text mining theories. Furthermore, the concept of fundamentalism, often associated with revivalist religious groups, is discussed, along with radicalism, both of which can involve violence as a means of achieving objectives rooted in religious beliefs. The paper reviews various understandings from multiple viewpoints to provide a comprehensive framework for the study shown in Table 1.

Radicalism often arises as a reaction to injustice and the unequal distribution of power in society. Politically, it represents the use of state power by individuals or groups to achieve specific objectives. Sociologically, conflict theory suggests that radicalism emerges from the concentration of power among certain groups, who seek to maintain their dominance. Anthropologically, radicalism is seen as a social system developed by groups to defend against perceived threats. Economically, radical acts are often viewed as resistance by marginalized groups against the ruling class and state, seen as failures in providing prosperity. In religious contexts, radicalism is often fueled by narrow interpretations of religious teachings, such as the misapplication of jihad to justify violence. In information technology, radical content typically involves promoting violence, hate, and anti-nationalism. From these perspectives, radical features are identified to detect and analyze radical content.

In Figure 2, the system evaluates the similarity between the data and the keywords in the database using the Euclidean distance method. The keywords are visualized based on how closely the data matches the labels in the database. The goal of this process is to develop a model for identifying radical content. This process involves several stages: starting with a literature review, gathering data on radical content, preprocessing through text mining, manually labeling the content into classes, and selecting relevant features. The final result is a model that categorizes radical content according to the characteristics and definitions of radicalism in Indonesia.

Table 1. Various aspect for reviews definition of radicalism

Various aspects	Definition
Politic aspect	Radicalism in politics is linked to the pursuit of state power, where power is used to achieve personal goals, with the state's authority differing from that of other organizations.
Sociology aspect	Radicalism arises from the unequal distribution of power, where dominant groups use their influence to maintain control, often through radical means to sustain their power.
Culture aspect	Radicalism is seen as a social construct developed by groups to defend against perceived threats, often using retaliation or demands for justice in response to perceived wrongs.
Economy aspect	Radicalism or terrorism is viewed as resistance by marginalized groups against capitalist elites and the state, a response to social inequality caused by exploitation.
Religion aspect	Radicalism emerges from a distorted interpretation of religious teachings, where concepts like jihad are misused to justify violent actions for achieving certain goals.
Information technology aspect	In information technology, radical content often involves promoting violence, hatred, and anti-nationalism, using religion to justify violent acts like suicide bombings.

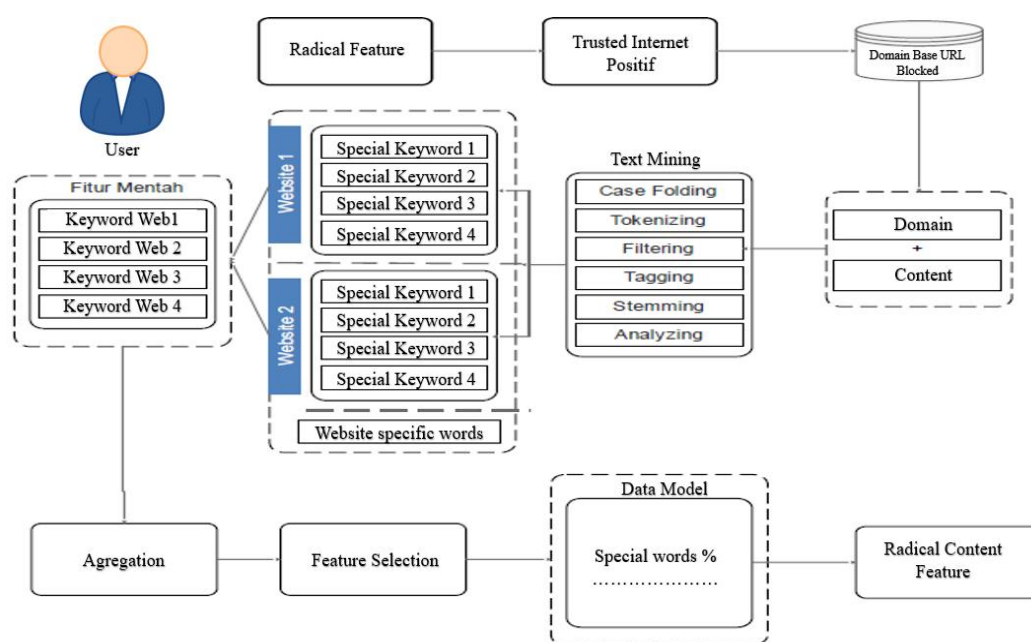


Figure 2. System learning diagram

The system first collects the data, followed by text mining. The purpose of text mining is to identify the words that represent the content and to analyze and connect them. Text mining applies data mining techniques to uncover patterns in text and extract valuable information. Due to the unstructured nature of text data, text mining requires preprocessing steps to organize the data effectively. These steps include five stages: tokenizing, filtering, stemming, tagging, and analyzing, which are sequential and interconnected. After these steps, four class labels are assigned: green, yellow, red, and white. The red label denotes content that has significant influence, such as material that incites readers to mobilize, raise funds, or oppose the government. Each label is defined based on the ministry of communication and information technology's criteria for radical content.

The literature review in this study involved identifying radical and non-radical content, examining global radicalism research, understanding the characteristics of radical content in different countries, and analyzing how radical features can be identified specifically for Indonesia. Methods used in other countries to

categorize radical content were explored and partially adapted for the system. Data collection involved interviewing experts on radical Islam and using search indexes for radical content, extremist websites, terrorism, and issues related to ethnicity, religion, and class. Examples included content with hate speech, insults, and websites of religious organizations like MUI, NU, Hidayatullah, Hizbut Tahrir, PGI, MAGABUDI, and PHDI. Additionally, the study included domain-based content deemed harmful by the government, identified by Ministry of Communication and Informatics (Menkominfo's) trusted links. A total of 33 content pieces related to ethnicity, religion, and class, along with 33 radical content examples, were collected.

3. RESULTS AND DISCUSSION

We implemented the system as a content search engine application using the PHP programming language. This choice was made because PHP is open source and can be accessed by any operating system, allowing anyone to develop or build applications with it. The process starts with creating a search engine form to simplify the user's search for news-related content. The user inputs a query, which could be a news topic or a sentence. This query undergoes text mining and is then compared with keywords stored in the database. The system checks the similarity between the queried words and the database keywords. Based on this, the system assigns news labels according to the user's search. In this classification process, documents that have been processed up to the text frequency calculation stage will be classified based on predefined classes (red, yellow, green, white). The determination of class labels is done manually using human judgment. The process involves reviewing all the learning content obtained from the positive trust data crawling by the ministry of communication and informatics. A total of 126 news articles were obtained and then sorted based on the specified labels. After that, text mining is performed, and the content is converted into document vectors and aggregated. For class labels, there are four class labels assigned: green, yellow, red, and white. The red label is symbolized as content that has a very significant impact, such as enabling readers to mobilize and raise funds, or provoking opposition against the government. Below are the descriptions of each class label for radical content obtained from the Menkominfo show in Table 2.

Table 2. Class label for radical content

No.	Classification	Description
1.	Red	Calls for specific actions, mobilization of funds, and people
2.	Yellow	Calls for a certain attitude, provocation
3.	Green	Spreading false news about certain groups
4.	White	Normal content

The decision-making process in this method involves evaluating test data by comparing it to all training data using the Euclidean distance to measure their proximity. These distances are ranked from smallest to largest, and a voting system is applied based on the ranked distances. The number of votes, denoted by k in the KNN algorithm, is usually set to an odd number such as 1, 3, 5, or 7 to avoid ties. The classification of the test data is determined by the majority vote from the nearest neighbors. Each test data point is assigned a label through the KNN algorithm. The system then searches for data by using keywords from the database and calculates the Euclidean distance for each word. The final result is a label (red, yellow, green, or white). The algorithm's accuracy is evaluated by using a radical dataset, divided into various decision-making attributes.

The system was developed with a web scraping application, and the study includes a flowchart for the tests conducted. Experimental parameters include the query results, classification accuracy (measured using the confusion matrix), and performance in terms of processing time. The experiments are divided into three main tests: querying the system, evaluating classification accuracy, and assessing system performance in processing time and memory usage. The retrieval of news content from da'wah websites, such as Website Da'wah, was carried out through scripts. For this experiment, however, the news content was directly sourced from the scrapped database of Website Da'wah content.

Data testing is conducted directly by the system, beginning with text mining to extract significant words from the news content. This process is similar to the training procedure described earlier. However, unlike the training process, which includes a feature selection step after text mining, the testing process does not. Instead, the tested data is stored in the database alongside the training data for classification. The system tests the search data with the keywords stored in the database, calculates the Euclidean distance for each word, and classifies the data into one of the labels (red, yellow, green, or white). The results are visualized, as shown in Figure 3.

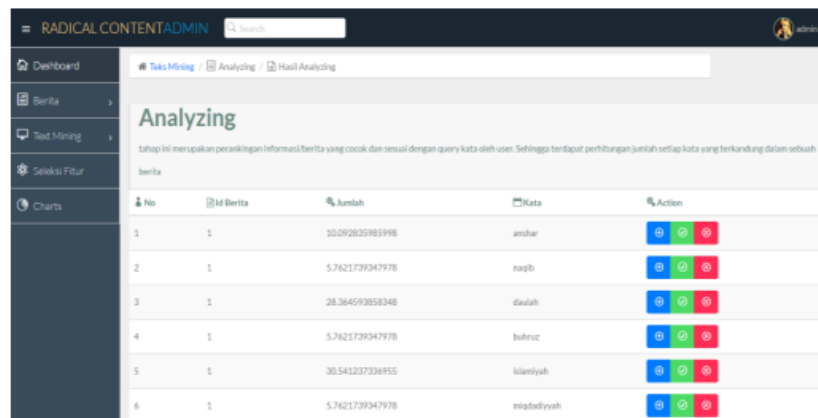


Figure 3. Analyzing system result

The system interface is designed to be simple to facilitate user navigation. Users only need to enter the sentence or content of the news to be text-mined. The system will then display a table of important words resulting from the text mining process. In this system testing, a sample of data was taken randomly from the database, constituting 10% of the dataset. The selected content was then entered into the news input form, resulting in the following outcomes:

Table 3 shows the results of the query testing experiment conducted on the system. Ten news articles were randomly selected from the content database and then input into the system using the news input form and Figure 4 show the extraction content from the website da'wah.

Table 3. Result of query testing experiment

No	Test content	Label	Result label
1	Content 1	Red	White
2	Content 2	Red	Yellow
3	Content 3	Yellow	White
4	Content 4	Yellow	White
5	Content 5	White	White
6	Content 6	Yellow	Yellow
7	Content 7	White	White
8	Content 8	Green	Green
9	Content 9	Green	Green
10	Content 10	White	White

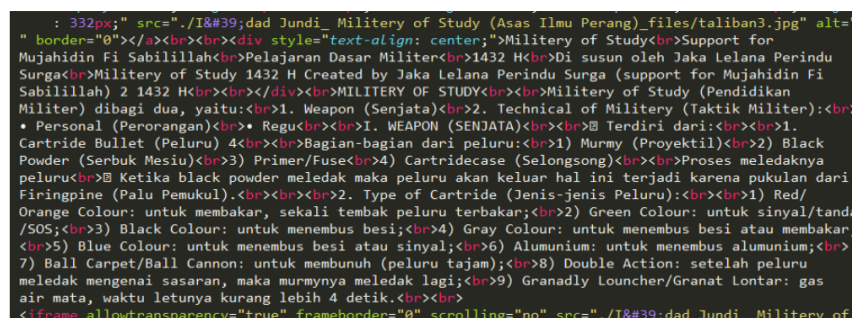


Figure 4. Extraction content result

This application measures the algorithm's accuracy in correctly classifying documents into the four specified categories. The accuracy calculation is detailed below.

$$Accuracy (\%) = \frac{\sum \text{Correct corrections}}{\sum \text{test data}} * 100 \quad (1)$$

Based on the equation, the accuracy achieved is 69.05%. This was tested using cross-validation to verify the correctness of the manual calculations. The classification results were then analyzed to determine the accuracy using the confusion matrix. The confusion matrix calculation focuses on true positive items or variables, utilizing the (2).

$$\begin{aligned} \text{True Positif (TP)} &= \frac{TP}{TP+FN} \\ \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} = \frac{TP+TN}{N} \end{aligned} \quad (2)$$

The test results help determine the optimal k value for maximum success. In this study, k values from 1 to 30 were tested. However, not all k values yielded significant results, so only the most effective k values were selected, it is evident that the optimal and stable k value is achieved at k = 7 and beyond, indicating that k = 7 provides the best accuracy. The graph below in Figure 5 illustrates the optimal k value identified in this research.

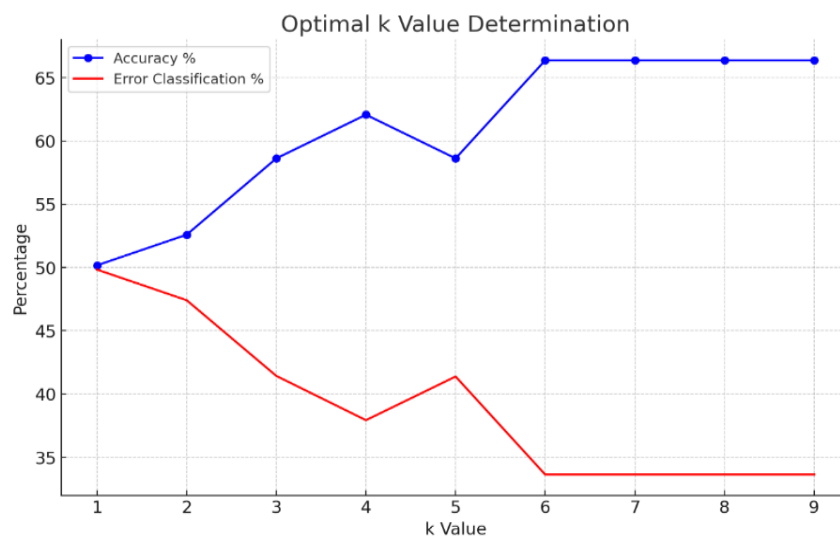


Figure 5. Radical k-optimal

The result shows that the optimal result is obtained with a k value of 6. This is due to the large spread of distance values between different classes. The number of k determines how many data points have the closest distance, which is then used to calculate the mode for predicting the most common classification class. After determining the performance of the classification, we then measure the precision and recall values for each class to understand the closeness of the relationships between the classes. Below are the precision results for each class show in Table 4.

Table 4. Measurement the precision classification

Class	K						
	1	2	3	4	5	6	
Red	0.133	0.214	0.143	0.143	0.233	0.457	0.457
Yellow	0.385	0.538	0.583	0.471	0.538	0.571	0.571
Green	0.089	0.078	0.089	0.128	0.178	0.210	0.210
White	0.632	0.701	0.690	0.690	0.632	0.741	0.741

When compared to the values in the tables above, it can be seen that the “White” class has the following precision values: k1 = 0.689, k2 = 0.824, k3 = 0.797, k4 = 0.784, k5 = 0.683, k6 = 0.811, and k7 = 0.811, with an average increase of 0.771. The highest value is obtained at k2, compared to all classes. The “White” class has the highest precision and recall values across all k values.

4. CONCLUSION

This study demonstrates the effectiveness of text mining techniques in classifying radical content in Indonesian Da'wah on websites using a PHP-based content search engine. The system processes queries by comparing text with keywords stored in a database, using 126 labeled articles from the Ministry of Communication and Informatics. The K-NN algorithm, with an optimal k of 7, achieved a 66.37% accuracy rate. The results highlighted the “White” class’s high precision and recall. Performance testing showed a classification time of 0.704 seconds and memory usage of 884,656 bytes, indicating efficiency with room for improvement. Future enhancements include adding synonym identification for Indonesian keywords, involving experts in theme classification, and combining K-NN with other machine learning algorithms to improve accuracy. Expanding the system to mobile platforms and proxy servers could increase its versatility, while integrating it with educational initiatives can strengthen resilience against online radicalization.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the ustadz of Islamic boarding schools in Banjarmasin, the National Counter-Terrorism Agency of Indonesia, and the Department of Communication and Informatics of South Kalimantan for their valuable assistance and insights during the course of this research. Their contributions have significantly enriched the study, and their support is greatly appreciated.

FUNDING INFORMATION

This research was funded by the Ministry of Religious Affairs of the Republic of Indonesia under the Research Grant Scheme with Proposal Number Detail 221180000058591.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Aulia Aziza				✓	✓			✓	✓	✓			✓	✓
Risqiatul Hasanah						✓			✓			✓		
Juairiah									✓	✓				
Munsiyi	✓	✓	✓				✓		✓		✓			

C : **C**onceptualization

M : **M**ethodology

So : **S**oftware

Va : **V**alidation

Fo : **F**ormal analysis

I : **I**nvestigation

R : **R**esources

D : **D**ata Curation

O : Writing - **O**riginal Draft

E : Writing - Review & **E**ding

Vi : **V**isualization

Su : **S**upervision

P : **P**roject administration

Fu : **F**unding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

The research related to human use has been complied with all the relevant national regulations and institutional policies in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee.

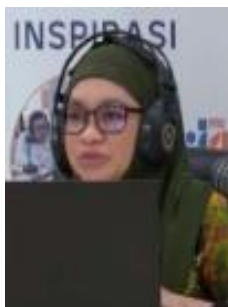
DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author, Munsiyi. The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

- [1] Y. Elovici *et al.*, "Content-based detection of terrorists browsing the web using an advanced terror detection system (ATDS)," *Lecture Notes in Computer Science*, vol. 3495, pp. 244–255, 2005, doi: 10.1007/11427995_20.
- [2] Y. Rahmatullah, "Radicalism, jihad and terror," *Al-Albab*, vol. 6, no. 2, pp. 151–166, 2017.
- [3] M. Fuad, "International publications on radicalism and terrorism in indonesia: a bibliometric assessment," *Wawasan: Jurnal Ilmiah Agama dan Sosial Budaya*, vol. 5, no. 1, pp. 96–107, 2020.
- [4] F. Wahyuni, "The role of citizens in preventing radicalism and terrorism in Indonesia," *Journal of Citizenship Studies*, vol. 4, no. 1, pp. 12–25, 2020.
- [5] H. Hasibuan, "The government's role in enforcing the law against perpetrators of terrorism," *Journal of Law and Sustainable Development*, vol. 12, no. 1, p. e3145, 2024, doi: 10.55908/sdgs.v12i1.3146.
- [6] D. Correa and A. Sureka, "Solutions to detect and analyze online radicalization : a survey," vol. 5, no. January, pp. 1–30, 2013.
- [7] A. Fauzi and M. Iqbal, "Meta-analysis: the root of terrorism from the perspective of islamic movements in Indonesia," *Journal of Islamic Studies*, vol. 8, no. 2, pp. 45–60, 2019.
- [8] A. Syaiful, "Radicalism on world wide web and propaganda strategy," *Journal of Communication Studies*, vol. 3, no. 1, pp. 77–89, 2015.
- [9] D. Solahudin and M. Fakhruroji, "Internet and islamic learning practices in Indonesia: Social media, religious populism, and religious authority," *Religions*, vol. 11, no. 1, pp. 1–12, 2020, doi: 10.3390/rel11010019.
- [10] Isabella and S. Nofrima, "Radicalism in the digital era: the role of digital literacy in preventing propaganda in Indonesia," *KnE Social Sciences*, vol. 2024, pp. 273–282, 2024, doi: 10.18502/kss.v9i18.16357.
- [11] B. Ganesh and J. Bright, "Countering extremists on social media: challenges for strategic communication and content moderation," *Policy and Internet*, vol. 12, no. 1, pp. 6–19, 2020, doi: 10.1002/poi3.236.
- [12] F. W. Githuthu, P. E. N. Gicheru, and R. D. N. Obiero, "The role of forgiveness in countering youth radicalization into violent extremism in Eastleigh, Kenya," *Journal of Sociology Psychology and Religious Studies*, vol. 4, no. 1, pp. 48–60, 2022, doi: 10.53819/81018102t2079.
- [13] M. R. A. Musa and Y. Sibaroni, "Radicalism speech detection in indonesia on twitter using backpropagation neural network method," *Journal of Computer System and Informatics (Josyc)*, vol. 3, no. 4, pp. 356–362, 2022, doi: 10.47065/josyc.v3i4.2146.
- [14] A. Subagyo, "The implementation of the pentahelix model for the terrorism deradicalization program in Indonesia," *Cogent Social Sciences*, vol. 7, no. 1, 2021, doi: 10.1080/23311886.2021.1964720.
- [15] A. Rizki, "Security threats of radicalism through social media amid COVID-19 Pandemic: Indonesia's perspective," *Connections the Quarterly Journal*, vol. 20, no. 3–4, pp. 95–106, 2021, doi: 10.11610/connections.20.3-4.05.
- [16] D. Kurnia, "The pentahelix strategy of the national counterterrorism agency in reducing the potential of radicalism in Indonesia," *International Journal of Humanities Education and Social Sciences (Ijhess)*, vol. 3, no. 2, 2023, doi: 10.55227/ijhess.v3i2.617.
- [17] Y. Patil and S. Patil, "Review of web crawlers with specification and working," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, no. 1, pp. 220–223, 2016, doi: 10.17148/IJARCCCE.2016.5152.
- [18] M. Moncrieff and P. Liénard, "Radicalization and violent extremism depend on envy; conspiracy ideation, sometimes," *Frontiers in Psychology*, vol. 14, 2023, doi: 10.3389/fpsyg.2023.1111354.
- [19] M. S. Muthuswamy, "Does sharia act as both a mediator and moderator in salafi radicalism?," 2021, doi: 10.31235/osf.io/nd4xq.
- [20] R. Irfan *et al.*, "A survey on text mining in social networks," *The Knowledge Engineering Review*, vol. 30, no. 2, pp. 157–170, 2015, doi: 10.1017/s0269888914000277.
- [21] M. Rifa'i, "Sentiment analysis using text mining techniques on social media using the support vector machine method case study seagames 2023 football final," *J. Of Artif. Intell. And Eng. Appl.*, vol. 3, no. 1, pp. 141–147, 2023, doi: 10.59934/jaiea.v3i1.274.
- [22] M. Q. Basit, "Web mining algorithms for social media analytic," *International Journal of Computing Programming and Database Management*, vol. 5, no. 1, pp. 52–54, 2024, doi: 10.33545/27076636.2024.v5.i1a.94.
- [23] J. Srivastava, R. Cooley, M. Deshpande, and P. Tan, "Web usage mining," *Acm Sigkdd Explorations Newsletter*, vol. 1, no. 2, pp. 12–23, 2000, doi: 10.1145/846183.846188.
- [24] K. Sukhija, "Web content mining equipped natural language processing for handling web data," *International Journal of Computer Applications Technology and Research*, vol. 4, no. 3, pp. 209–213, 2015, doi: 10.7753/ijcatr0403.1008.
- [25] H. Chen and M. Chau, "Web mining: machine learning for web applications," *Annual Review of Information Science and Technology*, vol. 38, no. 1, pp. 289–329, 2004, doi: 10.1002/aris.1440380107.
- [26] J. X. Koh and T. M. Liew, "How loneliness is talked about in social media during COVID-19 pandemic: text mining of 4,492 twitter feeds," *Journal of Psychiatric Research*, vol. 145, pp. 317–324, 2022, doi: 10.1016/j.jpsychires.2020.11.015.

BIOGRAPHIES OF AUTHORS



Aulia Aziza S. Ag, M.Si    is a Doctoral student in Gadjah Mada University. Born in Tanjung on November 23, she can be contacted at auliaaziza@uin-antasari.ac.id. Aulia has authored works such as "Empowerment of Historical Mosques in South Kalimantan" (2017), "Soul Meditation" (2018), "Transformation of the Dayak Maanyan Traditions" (2019), and "The Role of Female Ulama in Preaching in South Kalimantan" (2021). She completed her Bachelor's in Islamic Communication and Broadcasting (1991–1995) and a Master's in Sociology at UGM (1999–2002). Aulia has served as Secretary and Head of the KPI Study Program at IAIN Antasari, and Deputy Dean II at the Faculty of Da'wah and Communication Studies. Her organizational roles include Secretary of FORHATI Banjarmasin (2023–present) and Secretary of ALISA Khadijah ICMi South Kalimantan. She can be contacted at email: risqiatulhasanah@uin-antasari.ac.id.



Risqiatul Hasanah    Doctoral student at Antasari State Islamic University Banjarmasin. She earned a Master of Da'wah Studies (M. Sos) at UIN Sunan Ampel Surabaya. In the field of Da'wah and a Bachelor of Social Studies (S.Sos). Degree in Da'wah from the State Islamic College of Sorong, West Papua Indonesia. She is currently working as a Lecturer in the Islamic Communication and Broadcasting Study Program at Antasari State Islamic University Banjarmasin since January 2019. She has teaching experience at the Mambaul Hisan Islamic Boarding School, Kediri. She is a researcher specializing in the field of da'wah. Through his work, she aims to advance knowledge in the analysis of contemporary da'wah, thus providing a significant impact on the field of contemporary da'wah. She can be contacted at email: risqiatulhasanah@uin-antasari.ac.id.



Juairiah    is a faculty member at the Faculty of Tarbiyah and Education, specializing in Islamic Library and Information Sciences at the undergraduate level. She holds the academic title of S.Pd.I (Bachelor of Islamic Education) and M.Hum (Master of Humanities). Her National Identification Number (NIP) is 198406082019082001, and her NIDN/NITK/NUP is 2008068403. Juairiah is currently active in teaching but does not hold a teaching certificate (Akta Mengajar) or teaching permit. She has a professional role in the field of Islamic Library and Information, contributing to education and academic development within her faculty. She can be contacted at email: juairiah.efendi@gmail.com.



Munsyi    A doctoral student at Lambung Mangkurat University, he holds a Master of Applied Science (M.T.) from Politeknik Elektronika Negeri Surabaya and a Bachelor's degree (S.Kom.) in Informatics Engineering from the Islamic University of Kalimantan. Since March 2019, he has been a lecturer in the Department of Information Technology at Islamic State University Antasari Banjarmasin, with six years of teaching and research experience. His research specializes in spatial analysis, focusing on integrating remote sensing and IoT to address environmental and urban challenges. He uses satellite imagery and IoT sensor networks to monitor, assess, and analyze spatial data for environmental management and urban sustainability. His work aims to contribute to understanding climate change, biodiversity conservation, and sustainable urban planning through advanced geospatial technologies. He can be contacted at email: munsyi@uin-antasari.ac.id.