International Journal of Computer Science and Information Technology (IJISIT)
Volume 1, No. 2 (2024) pp. 73-80
e-ISSN 3063-4156 | p-ISSN 3063-4172
DOI: 10.55123/ijisit.v1i2.5

Analysis of Hate Speech in 2024 Elections on Social Media Platforms Using Natural Language Processing (NLP) Methods

Richardo Johan Tanujaya[1], Farhan Mohammed[2], Gabriela Callista Halim[3], Cindy Thalia[4], Shane Michael Colyn[5]
[1,2,3,4,5]Mahasiswa Informatika, Universitas Pelita Harapan, Medan, Indonesia
Email: [1]03082220020@student.uph.edu, [2]03082220012@student.uph.edu, [3]03082220005@student.uph.edu, [4]03082220003@student.uph.edu, [5]03082220024@student.uph.edu

Article Info

Article history:
Received April 19, 2024
Revised October 02, 2024
Accepted December 25, 2024

Keywords:

Hate Speech Detection
NLP
Lexicon
Naive Bayes
Bi - LSTM

ABSTRACT

The surge in digital activity during the 2024 General Election in Indonesia has triggered the spread of hate speech on social media, potentially disrupting democratic stability. This study aims to analyze the effectiveness of hate speech detection on the X platform using Natural Language Processing (NLP) methods. The technical approach combines Regular Expressions for data pre-processing and a Lexicon-Based method to identify offensive words based on a predefined dictionary. Data collection was conducted using the Uninvolved Conversation Observation Technique (TSBLC) with a total of 100 election-related comment samples. Test results indicate that this rule-based program is capable of detecting the presence of hate speech within the data samples. However, comparative analysis reveals that the lexicon method yields lower accuracy compared to advanced Machine Learning models, such as Naive Bayes (93%) and Bi-LSTM (96.9%), utilized in previous studies. In conclusion, while the lexicon approach offers structured basic detection, the integration of machine learning models is highly recommended to enhance accuracy and achieve deeper contextual understanding in future research.

Corresponding Author:

Richardo Johan Tanujaya
Departement of Informatics, Faculty of Computer Science
Universitas Pelita Harapan Medan
20112, Medan, Indonesia
Email: 03082220020@student.uph.edu

1. INTRODUCTION

Public elections as the highest point of modern democracy have become, and always will be, a very dynamic moment, reflected in the intensity of information circulating on social media[1]. In an era where technology is becoming more dominant, so does the role of social media in creating public opinion. Unfortunately, a phenomenon known as ”hate speech” is becoming increasingly common on many social media platforms. The 2024 Indonesian elections has become the next stage for the contest of democracy, and one of its potential challenges is the spreading of hate speech on social media which could affect the elections[2].

Hate speech on social media is speech that attacks, insults, or belittles an individual or group. The purpose of hate speech is to spread hatred, provoke conflict, or degrade the dignity of certain individuals or groups[3]. Hate speech can be transmitted through many platforms, a few of which are Facebook, Instagram, and Whatsapp[4].

As an example, on X, formerly known as Twitter, there are countless examples of comments containing hate speech. We have observed a number of these hate speech comments regarding the 2024 elections.

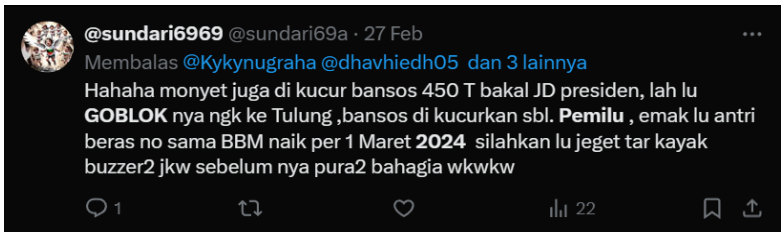

Figure 1. Hate Speech

Figure 2. Hate Speech

The picture above contains the word "*Goblok*" which could be defined as "something very stupid"[5]. The second picture contains the word "*K****l*" which references the male genital organ[5]. Both of these examples clearly contain words that are not ethical and could hurt somebody.

According to Abaido, at least 40.5% of users in social media have spread hate speech in some form[6]. Analysing hate speech related to the 2024 election issue on social media is crucial to understand the impact and expansion of this phenomenon. With a large proportion of the population being exposed to information on social media, controlling hate speech must be strict and efficient.

The Natural Language Processing (NLP) method is one of the most effective techniques for extracting, understanding, and analysing widely distributed text on social media platforms[7]. In this context, NLP opens up opportunities to identify unique patterns[8].

This research aims to present an in-depth analysis of Hate Speech found on social media platforms using the Natural Language Processing approach. By focusing on natural language analysis, this study is expected to uncover not only the types of hate speech disseminated but also the underlying sentiments and potential impacts on public perception[9].

Previously, research on reducing hate speech on social media has been conducted using smart contracts[10]. Previously, research on reducing hate speech on social media has been conducted using smart contracts. The smart contract is a method that has its connections to blockchain, where it establishes agreements between two clients[11]. However, smart contracts are not effective when applied to more public social media platforms such as Instagram or Twitter. In the research, they could only practise it within a group chat application[12].

Through the use of NLP technology, this research aims to detect hate speech or hate speech. The ultimate goal is to provide deep insights to stakeholders, including the government, electoral institutions, and the general public, to design effective mitigation strategies [13].

By integrating NLP technology to detect hate speech, this research is expected to make a significant contribution to understanding and addressing issues of misinformation that arise during the 2024 election process on social media platforms.

2. METHOD

This research employs a combination of Natural Language Processing (NLP) and Lexicon Base methods to detect hate speech in Indonesian language. NLP method is utilized for cleaning text data from noise and irrelevant characters, such as links and symbols. Regular Expression (RE) is employed to match specific text patterns, like links and non-alphanumeric characters[14].

Once the data is cleaned, Lexicon Base is utilized to identify words containing hate speech. Lexicon Base is a dictionary containing words and phrases categorized as hate speech based on certain criteria, such as SARA (Ethnicity, Religion, Race, and Inter-group)[15]. The method assigns values to each comment enabling us to ascertain whether the comment contains hate speech.

2.1 Program Workflow

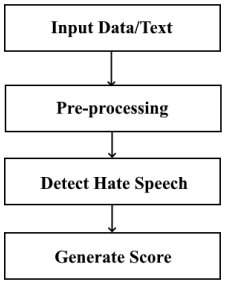

Figure 3. Workflow Diagram[16]

Pre-processing is a crucial aspect of NLP as it involves the process of "cleaning" the input so that the program can read it[17]. For pre-processing, regular expressions can be employed due to their flexibility in detecting a pattern[18].

2.2 Regular Expression

This research employs a method called Regular Expression. This method is one of the techniques in NLP that has the capability to search not only for specific strings, but also conducts searches based on patterns within those strings[19].

In the program, Regular Expression will be utilized twice to process the input text.

The first one is to remove links from the input:

$$text = re.sub(r'http\S+', '', text)$$

The pattern used here, 'http\S+', matches text starting with "http" followed by one or more non-space characters. When a link is found, it is replaced with an empty string, essentially means that the link is removed from the text[20].

The second one is to remove non-letter or non-word characters:

$$text = re.sub(r'[^\w\s]', '', text)$$

'[^\w\s]' is a pattern that matches any character that is not a word character (\w) or a space character (\s) because (^) is a negation symbol[20]. This means that if there are numbers or integers in our input, they will be removed and ignored.

2.3 Lexicon Base

The main part of the program is the Lexicon Base where it will be used to detect hate speech. This method involves creating a dictionary, including opinions [21].Words that was found in this dictionary will be used to classify words containing hate speech within a string.

$$self.hate\_speech\_lexicon = ["g **** k", "b ** o", "b *** h", "t *** l", "b ***** t", "j **** k", "a * u", "a **** g", "k **** l", "k ** l", "gemoy", "samul", "wowo", "bahlul", "rezim", "ponakan paman", "anakk h *** m", ,"pemilu curang"]$$

The lexicon list can expand during the testing process due to the variety of hate speech expressions on social media.

$$for\ term\ in\ self.hate\_speech\_lexicon:$$
$$if\ term.lower()\ in\ text.lower():$$
$$hate\_speech\_score\ +=\ 1$$

With Lexicon, the inputted text can be scanned for its content to determine whether it contains hate speech or not[22].

In data collection, we utilized *Teknik Simak Bebas Libat Cakap (TSBLC)*, an observation method to gather data. *TSBLC* is a technique used in data collection by taking screenshots of existing comments and also by reading and transcribing comments[23].

3. RESULTS AND DISCUSSION

The data sampling in this study was conducted using the *Teknik Simak Bebas Libas Cakap (TSBLC)* method, which is related to the 2024 Election on X application. The data collection process involved typing keywords into the search column of X application, after which we began gathering specific data samples classified as containing hate speech according to the provided keywords. The data that was collected using the *TSBLC* method was obtained through screenshots of comments related to the keywords. After collecting the data, the we began testing on the following link.
Link: https://bit.ly/DataSample_Kelompok3

For example, it can be seen in the 41th sample data, where in that data there is a hate speech that is noticed from the searched keywords.

Figure 4. The 41th Data Sample

Out of 100 collected data that consist of comments containing hate speech towards candidates, political parties, or election organizers, it can be inferred from the gathered data that social media platforms like X can serve as dangerous channels for spreading hate speech. The most prevalent types of hate speech that was found, are insults and defamation. This could pose a threat to democracy and stability in Indonesia.

3.1. Using Python Algorithms

We utilized Python for analyzing hate speech within comments on X platform social media. The algorithms employed included regular expressions and lexicon-based techniques. The combination of these algorithms proved effective in detecting words that can incite hatred. We defined rules to remove all characters that are not letters or spaces.

```python
import re
class HateSpeechDetector:
    def __init__(self):
        self.hate_speech_lexicon = ["goblok", "bego", "bodoh", "tolol", "bangsat", "jancok", "asu", "anjing", "kontol", "kntl",
                                    "gemoy", "samsul", "wowo", "bahlul", "rezim", "ponakan paman", "anak haram", "pemilu curang" ]
        self.hate_speech_score_threshold = 1  # Adjust as needed
    def preprocess_text(self, text):
        # Remove punctuation, URLs, emojis, and special characters
        text = re.sub(r'http\S+', '', text)
        text = re.sub(r'[^\w\s]', '', text)
        # Convert text to lowercase
        print(text)
        return text.lower()
    def detect_hate_speech(self, input_text):
        text = self.preprocess_text(input_text)
        hate_speech_score = 0
        # Lexicon-based detection (case-insensitive)
        for term in self.hate_speech_lexicon:
            # Count the number of occurrences of the term in the text
            count = text.lower().count(term.lower())
            # Increment the hate_speech_score by the count
            hate_speech_score += count
        print(hate_speech_score)
        # Dynamic thresholding
        if hate_speech_score >= self.hate_speech_score_threshold:
            return True
        else:
            return False
# Example usage:
detector = HateSpeechDetector()
while True:
    input_text = input("Enter text to check for hate speech (or type 'exit' to quit): ")
    if input_text.lower() == 'exit':
        break
    if detector.detect_hate_speech(input_text):
        print("Hate speech detected.")
    else:
        print("No hate speech detected.")
```

Figure 5. Snip Program

In the program, we used the "re" library for text processing using regular expressions. Then, we created a function named pre-process text to conduct pre-processing on the text according to the flow diagram in Figure 3. In this function, the process involves removing links, punctuation, and symbols. Additionally, the text will be converted to lowercase.

In the program, there is a function called detect_hate_speech used to detect hate speech in texts. This function has two attributes: hate_speech_lexicon and hate_speech_score. The hate_speech_lexicon attribute contains words or phrases that we consider to be hate speech.

The hate_speech_score attribute functions to tally each hate speech word or phrase that appears. The program uses a while loop to continuously prompt for text input. If the entered text is "exit", the loop will terminate. The entered text is then checked, and the processed results will be displayed.

The steps for using this algorithm are as follows: first, the program will prompt the user to enter any text to detect hate speech. Then, the text will be pre-processed using the preprocess_text function. Next, the algorithm will proceed to detect words or phrases from the hate_speech_lexicon. Each word that indicated as hate speech will be assigned a score based on its occurrence in the text. The accumulated score will be compared against a predefined threshold (hate_speech_score_threshold). If the score is greater than or equal to the specified threshold, the result will be true, indicating the presence of hate speech in the input text. Conversely, if the score is less than the threshold, the result will be false, indicating the absence of hate speech in the input text.

3.2. Data Processing

A total of 100 samples of data have been successfully collected. With this sample datas, testing will be conducted using the Python program that has been developed. After going through the processing stages, each sample datas will produce a different output.

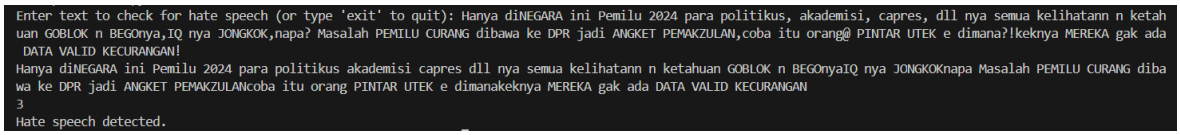

Figure 6. Program's Output

The number of detected hate speech instances in each sample data is annotated in the output results after processing. Here is a summary table of the output related to the detection of hate speech and the number of hate speech instances contained in each sample data.

Table 1. Results of Data Processing
Link : https://bit.ly/DataProcessed_Kelompok3

No.	Text	The amount of hate speech	Hate speech detection status
1.	geplak kepala koruptor, tolol. bukan geplak mereka yang lagi have fun dan encourage generasinya buat ikut pemilu. lo hidup beneran musingin orang yang lagi asik sama hiburannya sendiri? segabut itu lo? mo gemoy fandom kek, mo indogov fandom kek, terserah dong orang seneng apa.	1	Detected
2.	Hanya diNEGARA ini diPEMILU 2024 para POLITIKUS, AKADEMISI, CAPRES dll nya semua kelihatan n ketahuan GOBLOK n BEGOnya, IQ nya JONGKOK, napa? Masalah PEMILU CURANG dibawa ke DPR jadi ANGKET PEMAKZULAN, coba itu orang2 PINTAR UTEK e dimana?!keknya MEREKA gak ada DATA VALID KECURANGAN!	3	Detected
3.	Gw pikir pemilu 2019 adlh yg paling kacau dgn 800an petugas meninggal dunia. Ternyata pemilu 2024 lebih goblok Ig dengan segala ketololan kpu	2	Detected
4.	Lain kali si abah tahun 2029 jangan di ajak, pada ribet pendukungnya, goblok goblok di pliara	3	Detected
5.	Kenapa ada indikasi kecurangan pada pemilu 2024? 1. Ada UU yg di rekayasa oleh pamannya di MK tapi semuanya pada pura pura goblok, karena mestinya pembuat UU & bisa merubah UU adalah DPR. 2. Imbas dari rekayasa UU di MK itu ketua KPU terbukti pelanggaran etik. 3. 1 rangkaian.	1	Detected
6.	Hahaha monyet juga di kucur bansos 450 T bakal JD presiden, lah lu GOBLOK nya ngk ke Tulung, bansos di kucurkan sbl. Pemilu, emak lu antri beras no sama BBM naik per 1 Maret 2024 silahkan lu jeget tar kayak buzzer2 jkw sebelum nya pura2 bahagia wkwkw	1	Detected

7.	Sama goblok dan tololnya. Hasil pemilu tanggal 13 Febuary 2024 itu sudah ada. Sudah dikunci pake Quick Count secara CURANG modifikasi rumus Algoritma. Pemilu tggl 14 February 2024 hanya selebrasi dan pesta coblos Penipuan terbesar. Pengkhianatan terparah terhdp Rakyat. Paham??	2	Detected
8.	Simple, yg kontra dgn hasil pemilu 2024 ini jumlahnya pasti lebih besar drpd dgn yg pro. Makanya yg kontra teruslah berteriak, suarakan terus keadilan! JGN TERUS2AN JADI BANGSA YG GOBLOK!	1	Detected
9.	Usut sampai ke akar akarnya kecurangan pemilu 2024,,,, pake nyalahin kamera,,,goblok	1	Detected
10.	lah yang Goblok lu itu, Lu gak sungkem sama senior Yaman kadrun Babe Haikal Dan si botak HTI Achmad Dhani Dari awal pemilu 2024, imam jumbo gak pernah keliatan Sejak kapan surya Paloh bermesraan dengan imam jumbo ngakak 02 emang Goblok*	2	Detected
…	…	…	…
100.	Bocils jaman skrg semakin mengerikan girls. Yang mau punya anak hati2 aja. Sekarang zaman semakin rusak semakin KONTOL	1	Detected

Our Python program aims to detect the amount of hate speech that contained within the sample data with a certain level of accuracy. After conducting the test, the sample data overall contains varying amounts of hate speech. The statistical analysis results from testing the 100 samples against the program are displayed through the following diagram.

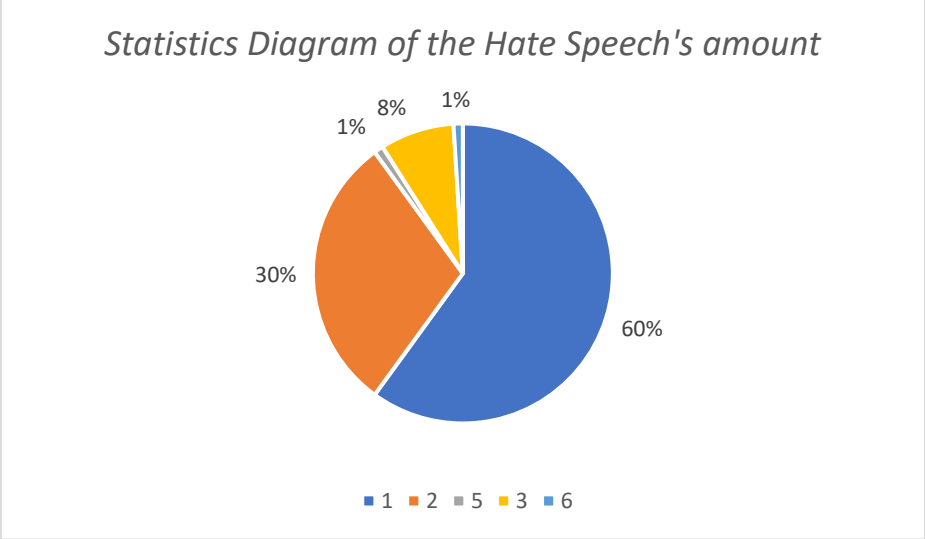

Figure 7. Percentage of the Detected Hate Speech's amount

The symbols, numbers, and colors in the diagram above serve as indicators of the detected amount of hate speech. The dark blue color (number 1) represents the variable for detecting one instance of hate speech, indicating that there are sixty samples containing one hate speech word each. This also applies to the other four color indicators representing the percentage of samples containing hate speech words to a certain variable extent.

We also utilized datasets from two different journals to compare the accuracy of the developed program. We then inputted samples from these datasets into the Python program that was created.

Table 2. Comparison of the accuracy of Processed data

No.	Authors	Method	Amount of data	Accuracy (%)
1	(Richardo et al.,2024)	Natural Language Processing	13,169	59.8
2	(Elisabeth et al.,2020)	Naive Bayes (NB)	13,169	93

From Table 2, we can see that Elisabeth utilized a different method compared to ours and although they used the same dataset they were able to acquire a much higher accuracy than ours, 93% [24].

Table 3. Comparison of the accuracy of Processed data

No.	Authors	Method	Amount of data	Accuracy (%)
1	(Richardo et al.,2024)	Natural Language Processing	713	69
2	(Isnain et al., 2020)	Bidirectional Long Short Term Method (Bi - LSTM)	713	96.9

We also compared our model with another journal. As seen in Table 3, which also used a different method compared to us, Bi-LSTM, which achieved a high accuracy with 96.9% [25].

4. CONCLUSION

In conclusion, the analysis of hate speech in the 2024 election on social media platforms, utilizing the *Teknik Simak Bebas Libas Cakap (TSBLC)* method alongside Python-based natural language processing techniques, revealed significant insights into the prevalence and nature of hate speech online. However, our study underscores the limitations of rule-based approaches in comparison to more sophisticated machine learning methods. Despite the meticulous design of algorithms employing regular expressions and lexicon-based techniques, our program demonstrated a lower accuracy in detecting hate speech when compared to machine learning models. This disparity became evident when benchmarking our results against those achieved by Elisabeth and another journal, both of which employed different methodologies, yet achieved notably higher accuracy rates of 93% and 96.9% respectively.

Furthermore, our comparative analysis with these studies suggests that while rule-based approaches provide a structured means of identifying hate speech, they may lack the nuanced understanding and adaptability of machine learning models. The success of Elisabeth's method, for instance, indicates the effectiveness of machine learning algorithms in discerning complex patterns and contexts within textual data, thereby achieving superior accuracy rates. Likewise, the journal's approach further highlights the potential of machine learning techniques to outperform rule-based systems in hate speech detection tasks. Thus, while our study contributes valuable insights into the landscape of hate speech during the 2024 election, it underscores the need for further exploration and adoption of machine learning methodologies to enhance the accuracy and efficacy of hate speech detection on social media platforms.

ACKNOWLEDGEMENTS

We, the writitng team, would like to thank Universitas Pelita Harapan for providing the opportunity to conduct this research and complete our assignment from the scientific writing course. The author would also like to thank the lecturers who have guided the writing of this journal from start to finish.

REFERENCES

[1] R. M. Widayat, A. Nurmandi, Y. Rosilawati, Z. Qodir, S. Usman, and T. Baharuddin, "2019 Election Campaign Model in Indonesia Using Social Media," *Webology*, vol. 19, no. 1, pp. 5216–5235, Jan. 2022, doi: 10.14704/web/v19i1/web19351.

[2] F. E. Siregar, "The Role of the Elections Supervisory Agency to Contend Hoax and Hate Speech in the Course of 2019 Indonesian General Election," *Padjadjaran Jurnal Ilmu Hukum*, vol. 7, no. 2, pp. 158–180, 2020, doi: 10.22304/pjih.v7n2.a2.

[3] M. M. Nasution, J. Izar, and I. H. Afifah, "https://ejournal.unida-aceh.ac.id/index.php/jetli AN ANALYSIS OF HATE SPEECH AGAINST K-POP IDOLS AND THEIR FANS ON INSTAGRAM AND TWITTER FROM THE PERSPECTIVE OF PRAGMATICS 1*." [Online]. Available: https://ejournal.unida-aceh.ac.id/index.php/jetli

[4] O. K. Lekik, S. Palinggi, and I. C. Ranteallo, "The Descriptive Analysis of Hoax Spread through Social Media in Indonesia Media Perspective," Scitepress, Sep. 2020, pp. 276–286. doi: 10.5220/0009441402760286.

[5] F. Deni, D. dan Muhammad, and Ikhwan M. Said, "JENIS UJARAN KEBENCIAN (HATE SPEECH) DALAM KOLOM KOMENTAR INSTAGRAM JOKOWI PADA MASA PPKM: ANALISIS LINGUISTIK FORENSIK," *Jurnal Indonesia Sosial Teknologi*, vol. 3, no. 5, pp. 574–585, May 2022, doi: 10.36418/jist.v3i5.422.

[6] G. M. Abaido, "Cyberbullying on social media platforms among university students in the United Arab Emirates," *Int J Adolesc Youth*, vol. 25, no. 1, pp. 407–420, Dec. 2020, doi: 10.1080/02673843.2019.1669059.

[7] H. Jiang, Y. Hua, D. Beeferman, and D. Roy, "Annotating the Tweebank Corpus on Named Entity Recognition and Building NLP Models for Social Media Analysis," Jan. 2022, [Online]. Available: http://arxiv.org/abs/2201.07281

[8] T. Fontes, F. Murcos, E. Carneiro, J. Ribeiro, and R. J. F. Rossetti, "Leveraging Social Media as a Source of Mobility Intelligence: An NLP-Based Approach," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 4, pp. 663–681, 2023, doi: 10.1109/OJITS.2023.3308210.

[9] T. T. A. Putri, S. Sriadhi, R. D. Sari, R. Rahmadani, and H. D. Hutahaean, "A comparison of classification algorithms for hate speech detection," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, May 2020. doi: 10.1088/1757-899X/830/3/032006.

[10] F. Tchakounté, K. Amadou Calvin, A. A. A. Ari, and D. J. Fotsa Mbogne, "A smart contract logic to reduce hoax propagation across social media," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 3070–3078, Jun. 2022, doi: 10.1016/j.jksuci.2020.09.001.
[11] D. Sinha and S. Roy Chowdhury, "Blockchain-based smart contract for international business – a framework," *Journal of Global Operations and Strategic Sourcing*, vol. 14, no. 1, pp. 224–260, Mar. 2021, doi: 10.1108/JGOSS-06-2020-0031.
[12] Z. Zheng *et al.*, "An Overview on Smart Contracts: Challenges, Advances and Platforms," Dec. 2019, doi: 10.1016/j.future.2019.12.019.
[13] S. Kreps, "THE ROLE OF TECHNOLOGY IN ONLINE MISINFORMATION," 2020.
[14] S. Mitra, "Regular expressions: A detailed study for the understanding of their role and methods for efficient application Soham Mitra," *~ 71 ~ International Journal of Research in Circuits, Devices and Systems*, vol. 2, no. 2, pp. 71–76, 2021, [Online]. Available: www.circuitsjournal.com
[15] K. M. O. Nahar, A. Jaradat, M. S. Atoum, and F. Ibrahim, "Sentiment analysis and classification of arab jordanian facebook comments for jordanian telecom companies using lexicon-based approach and machine learning," *Jordanian Journal of Computers and Information Technology*, vol. 6, no. 3, pp. 247–262, Sep. 2020, doi: 10.5455/jjcit.71-1586289399.
[16] E. A. Abdelnabi, A. M. Maatuk, T. M. Abdelaziz, and S. M. Elakeili, "Generating UML Class Diagram using NLP Techniques and Heuristic Rules," in *Proceedings - STA 2020: 2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 277–282. doi: 10.1109/STA50679.2020.9329301.
[17] M. Kashina, I. D. Lenivtceva, and G. D. Kopanitsa, "Preprocessing of unstructured medical data: The impact of each preprocessing stage on classification," in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 284–290. doi: 10.1016/j.procs.2020.11.030.
[18] D. Gibney and S. V. Thankachan, "Text indexing for regular expression matching," *Algorithms*, vol. 14, no. 5, May 2021, doi: 10.3390/a14050133.
[19] F. A. T. Tobing and R. Nainggolan, "ANALISIS PERBANDINGAN PENGGUNAAN METODE BINARY SEARCH DENGAN REGULAR SEARCH EXPRESSION," *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, vol. 4, no. 2, pp. 168–172, Oct. 2021, doi: 10.46880/jmika.Vol4No2.pp168-172.
[20] P. Wang, C. Brown, J. A. Jennings, and K. T. Stolee, "An Empirical Study on Regular Expression Bugs," in *Proceedings - 2020 IEEE/ACM 17th International Conference on Mining Software Repositories, MSR 2020*, Association for Computing Machinery, Inc, Jun. 2020, pp. 103–113. doi: 10.1145/3379597.3387464.
[21] A. Syakur, "IMPLEMENTASI METODE LEXICON BASE UNTUK ANALISIS SENTIMEN KEBIJAKAN PEMERINTAH DALAM PENCEGAHAN PENYEBARAN VIRUS CORONA COVID-19 PADA TWITTER," *Jurnal Ilmiah Informatika Komputer*, vol. 26, no. 3, pp. 247–260, 2021, doi: 10.35760/ik.2021.v26i3.4720.
[22] A. Koufakou and J. Scott, "Lexicon-Enhancement of Embedding-based Approaches Towards the Detection of Abusive Language," 2020. [Online]. Available: https://www.kaggle.com/c/jigsaw-toxic-comment-classification-
[23] K. H. Prasetya, H. Subakti, and A. Musdolifah, "Pelanggaran Prinsip Kesantunan Berbahasa Peserta Didik terhadap Guru Sekolah Dasar," *Jurnal Basicedu*, vol. 6, no. 1, pp. 1019–1027, Jan. 2022, doi: 10.31004/basicedu.v6i1.2067.
[24] D. Elisabeth, I. Budi, and M. O. Ibrohim, "Hate Code Detection in Indonesian Tweets using Machine Learning Approach: A Dataset and Preliminary Study," in *2020 8th International Conference on Information and Communication Technology, ICoICT 2020*, Institute of Electrical and Electronics Engineers Inc., Jun. 2020. doi: 10.1109/ICoICT49345.2020.9166251.
[25] A. R. Isnain, A. Sihabuddin, and Y. Suyanto, "Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 14, no. 2, p. 169, Apr. 2020, doi: 10.22146/ijccs.51743.