

# Text Mining To Detect Plagiarism In E-Learning System Using Rabin Karp Algorithm

DITA BAITU RAHMAWATI<sup>1</sup>, MUHAMMAD LUTHFI IRFANI<sup>2</sup>, RITA BR PURBA<sup>3</sup>, INDRA RANGGADARA<sup>4</sup>

<sup>1,2,3,4</sup> Faculty of Computer Science, Mercu Buana University, Jakarta, Indonesia

**Abstract-** *Plagiarism is copying someone else's work, by someone who makes the work as his work. Plagiarism can be considered a criminal act for stealing the works of others. The most common plagiarism acts are found in many academic fields, such as in the E-Learning system. E-Learning is a learning system that utilizes information technology in the process of learning and teaching, from accessing modules, filling the forum, and finish quizzes. This study aims to analyze plagiarism when filling out E-Learning forums. We apply the strategic approach of the Rabin-Karp algorithm, as a tool to detect plagiarism. The data we use in this study is processed using the text mining method. The process of detecting plagiarism in E-Learning systems uses the white insensitivity, K-Gram, Hash, Fingerprint, and Dice Similarity Coefficient processes. The results of plagiarism detection are obtained from the percentage of the similarity of answers in the E-Learning forum.*

**Indexed Terms-** *Algorithm Rabin Karp, Data Mining, E-learning, Plagiarism, and Similarity.*

## I. INTRODUCTION

Plagiarism is a general problem that occurs in various aspects of life. There are various forms of plagiarism in the academic field, easily find at school and universities. Plagiarism at universities can be a demotivating factor for the lecturers and students. If plagiarism is not addressed properly, plagiarists will get benefits that they are not actually entitled to, from high scores on assignments, credits from the others, while they didn't do sufficient effort during the working process.

The most frequent plagiarism act that being done in the academic field, is submitting the assignment without citing any reference, or claim others work as

their own works. Many factors encourage plagiarists into committing acts of plagiarism, from fear of failing in certain subjects, low grades in difficult classes, low awareness of copyright and intellectual property, lack of confidence in submitting their own work.

Plagiarism act will result in consequences and disadvantages, from reduced creativity and innovation of students, ruined reputation in both academic and professional field, violation of copyright and intellectual property law.

Studies from previous research show that there are several factors that influence the performance of Rabin-Karp algorithm, from: the amount of content in the file lengthens running time, stemming and preprocessing makes the processing time tend to be longer, but the similarity accuracy is higher, the given modulo value affects time of the process, and the smaller K-Gram will result in better accuracy of similarity values [1].

Observations of plagiarism behavior in practice reveal a number of commonly found methods for illegitimate text usage, which are characterized below [2]:

1. Copy & Paste (c & p) plagiarism specifies the act of taking over text verbatim from another author.
2. Disguised plagiarism subsumes practices intended to mask copied segments. Four different masking techniques have been identified. These are:
  - Shake & Paste (s & p) plagiarism is characterized by copying and merging sentences or paragraphs from different sources with slight adjustments necessary for forming a coherent text.
  - Expansive plagiarism refers to the insertion of additional text into or in addition to copied segments.
  - Contractive plagiarism describes the summary or trimming of copied material.

- Mosaic plagiarism encompasses the merge of text segments from different sources and obfuscating the plagiarism by changing word order, substituting words with synonyms or entering/deleting filling words.

In Mercu Buana University's E-Learning system, students are assigned to fill forum as a form of understanding the subjects given and as a student obligation during learning process. When filling in the E-Learning forum, students are required to analyze the problems, and provide opinions/responses along with solutions. However, during the process, it is often found that students commit act of plagiarism, by doing copy and paste the answer of other student's answers. There are other acts of plagiarism also being done, from using sources from others without proper citing, *paraphrasing text* or proposing ideas without citing the source.

The plagiarism issue can be treated from two perspectives, prevention and detection. As Schleimer, Wilkerson, and Aiken (2003) state, both can be combined to effectively reduce it. While copy detection methods can only help after the plagiarism has been committed, prevention methods can and should educate and encourage people not to do it, further decreasing its level [3].

Rabin-Karp algorithm as a basis for testing in this study is used to detect plagiarism by students in Mercu Buana University E-Learning system. The method of plagiarism detection using Rabin-Karp algorithm is easier to implement than the manual method. The advantage of Rabin-Karp algorithm compared to other string-matching algorithms is the capability to find many string patterns.

Scheimer, Wilkerson, and Aiken (2013) presented in their work that, if the detection method can only be done after a plagiarism action is committed, the prevention method can be done by educating and encouraging students to avoid plagiarism, which is expected to reduce the level of plagiarism. Nevertheless, prevention methods surely require active participation from all of the society members and academic environment.

According to the site: [www.plagiarism.com](http://www.plagiarism.com), in 2016, nearly 80% of students in all degree did academic

fraud at least once during their time as students. In the same year, The Psychological Scientific Journal reported that as many as 36% of undergraduate students did plagiarism in their academic writing. From the data in Education Week in 2017, 74% of students commit serious academic fraud [4].

In the circular letter of the Director General of Higher Education (Dirjen Dikti) dated January 4, 2012 addressed to the State Higher Education Leadership and Kopertis Coordinator, there is a list of names of state and private universities included in the list of cases of violations of academic norms in the form of plagiarism. The universities included in the list are 21 PTNs and 7 PTSs.

## II. THE PROPOSED METHOD & ALGORITHM

Several method and algorithm we proposed in this paper, including:

### a. Data Mining

Data Mining is a data collection process that is processed by various methods. Another term for data mining is knowledge-discovery in the database (KDD). The purpose of data mining is to utilize the data and process it to obtain new and useful information [5].

Data mining is the process of discovering interesting patterns and knowledge from large amounts of data. The data sources can include databases, data warehouses, websites, other information repositories, or data that are streamed into the system dynamically. Data mining processes according to Jiawei Han et. al as follow [6]:

1. Data Cleaning (to remove noise and inconsistent data)
2. Data integration (where multiple data sources may be combined)
3. Data selection (where data relevant to the analysis task are retrieved from the database)
4. Data transformation (where data are transformed and consolidated into forms appropriate for mining by performing summary or aggregation operations)

5. Data mining (an essential process where intelligent methods are applied to extract data patterns)
6. Pattern evaluation (to identify the truly interesting patterns representing knowledge based on interestingness measures)
7. Knowledge presentation (where visualization and knowledge representation techniques are used to present mined knowledge to users).

#### b. Text Mining

Text analytics (text mining) refers to techniques that extract information from textual data. It involves imposing structure upon text so that relevant information can be extracted from it [7].

Text mining is a technique used to deal with classification, clustering, information extraction and information retrieval. Basically, the work process of text mining many adopted from research Data Mining but the difference is the pattern used by the text mining is taken from the set of natural languages that are not structured in Data Mining patterns taken from structured databases [8].

Text mining is a variation of data mining that seeks to find interesting patterns from a large number of textual data sets. Text mining can be interpreted as the discovery of new information that was not previously known by computers, automatically extracting information from different sources. The key to this process is combining information that was successfully extracted from various sources. Text mining is defined as data in the form of text which is usually the source of data obtained from documents, with the aim of finding words that can represent the contents of the document which can later be analyzed by the relationship between documents [9].

Text mining usually marks by analyzing large amount of data in a form of text stored in the electronic file. This electronic file is usually in the form of several documents that will be processed, but certainly, these documents are not in a structured form. It takes a mechanism to mine the texts in the document collection so that information is more valuable and structured. The mechanism is divided into several stages (pre-processing phase).

There are several processes carried out in text mining to obtain hidden information, including the initial processing of text (text preprocessing), text transformation (text transformation), feature selection (feature selection) and discovery pattern (pattern discovery) [10].

1. Text Pre-processing is the process of preparing text into data that will undergo further processing. In pre-processing do text cleaning for each sentence.
2. Text Transformation is a transformation process that aims to get the expected document representation. Text transformation changes words into their basic form and reduces the word dimensions in the document.
3. Feature Selection is the selection of word features is the stage of the dimensional reduction process in the transformation process. In addition to deleting words that are considered not descriptive (stop words), not all words in the document have important meanings. Therefore, to reduce the dimensions of the document, the selection of words that are considered important and have a close relationship with the contents of the document is done. The basic idea of feature selection is to delete words that appear in a document too little or too much.
4. Pattern Discovery is an important stage to find patterns or knowledge of the entire text. The initial input of the text mining process is text data and produces pattern output as a result of interpretation. If the output of the pattern discovery is not suitable for the application, then the evaluation is continued by iterating to one or several previous stages.

#### c. Rabin-Karp Algorithm

Rabin-Karp algorithm was first introduced by Michael O. Rabin and Richard M. Karp in 1987. This algorithm uses array tables and Hashing methods in the operation. This hashing method is used primarily to increase search speed by increasing equality testing in the text. The characteristics of this algorithm are [11]:

1. Using the hashing function.
2. The preprocessing phase uses the time complexity  $O(m)$ .
3. Or the search phase the complexity is  $O(mn)$ .
4. Time needed:  $O(m + n)$ .

In fact, the hash function stores the form of a string in another form, namely enumeration so that a certain string will have its own enumeration value (unique). Because a string only has an enumeration value, this is what the Rabin-Karp algorithm uses to speed up the search for strings in hash tables.

Using a method like this, there will be a leak in the search in a long text, because in the long text there will be the same string numbering even though the target string is different. So there is a need for further verification of the contents of the string. This can actually consume time longer if it occurs in a long substring. But a good hash function guarantees that this lack is rare, so the average search time using this method is relatively good.

This algorithm works well in most practical cases but will result in relatively long computing time for certain cases, such as searching for words consisting of 10,000 characters 'a' followed by a character 'b' in a word consisting of 10,000,000 characters 'a'. In this case, this algorithm will operate with its worst computational time, namely  $O(nm)$ .

Rabin-Karp algorithm accelerates the similarity testing for patterns of substrings in text using the hash function. Rabin-Karp algorithm is a string-matching algorithm that will use the hash function as a comparison between the search string ( $m$ ) with substring in text ( $n$ ). The hashing function provides a simple method for avoiding quadratic comparisons of character numbers in many cases or situations. Instead of checking each position of the text when pattern matching occurs, it is more efficient to carry out checks only if the text being processed has similarities to the pattern. To check the similarity between these two words the hash function is used (Fernando, 2009). If 2 strings are identical, the hash value will be the same, so string search can be derived by calculating the hash value of the pattern, and then searching for a pattern with the same hash value in the input data [12]. If the hash values are the same, a comparison will be made of the characters. If the results of the two are not the same, then the substring will shift to the right. Shift is carried out as much as  $(n-m)$  times. The efficient calculation of hash values when shifting will affect the performance of this algorithm.

#### d. Similarity

Similarity is the level of similarity between the two identified patterns. In the Rabin-Karp algorithm, the method used to calculate the level of similarity between the two patterns is Dice's Similarity Coefficients. The concept of similarity has become a very important issue in almost every field of science. There are three types of techniques that were built to determine the document similarity value [13].

##### q) Distance-based similarity measure

Distance-based similarity measure measures the level of similarity between two objects in terms of geometric distance from the variables included in the two objects. These Distance-based similarity methods include Minkowski Distance, Manhattan / City block distance, Euclidean distance, Jaccard Distance, Dice's Coefficient, Cosine similarity, Levenshtein Distance, Hamming Distance, and Soundex distance.

##### r) Feature-based similarity measure

Feature-based similarity measure performs a calculation of the degree of similarity by representing an object in the form of compared features. Feature-based similarity measure is widely used in classifying or pattern matching for images and text.

##### s) Probabilistic-based similarity measure

Probabilistic-based similarity measure calculates the degree of similarity between two objects by representing two sets of objects compared to probability. Kullback Leibler Distance and Posterior Probability are included in this method.

### III. RESEARCH METHOD

The research methodology is the use of a scientific method by researchers in collecting research data. And has a connection with procedures in conducting research and technical research. A research method approach is used to solve problems in the research process.

Stipulating research methods is very important because it can help direct researchers in collecting, processing and analyze research data [14]. To facilitate the conduct of research related to the problem under study, the writers approached using the case study method.

Using this approach, the data collected can be adjusted to the actual situation and compared to the supporting

theory. Thus, it can provide a fairly clear picture and can draw conclusions from the object under study.

The research method used is the Rabin Karp Algorithm. The Rabin Karp algorithm was discovered by Michael O. Rabin and Richard M. Karp. The Rabin Karp algorithm is a word search algorithm that searches for a pattern in the form of a substring in a text using hashing. This theory is rarely used to search for a single word, but it is quite important and very effective when used for plural word searches.

Steps to detect plagiarism in the Mercu Buana University e-learning system using the Rabin Karp Algorithm method include:

1. Eliminate punctuation and change to the source text and the word before searching process into words without capital letters.
2. Divide the text into gram-grams which are determined by the k-gram value.
3. Find hash values with the rolling hash function of each gram formed.
4. Perform an iteration of index  $i = 0$  to  $i = n - m + 1$  to compare the hash value of the word that need to be found with the hash value of the source text at index  $i$  to  $i + m - 1$ . If the hash value is the same, then it will be further examined whether the search word is the same as the text part from the source index  $i$  to  $i + m - 1$ . If it's the same, a suitable word has been found. If not, then proceed by comparing the hash value of the word searched with the source text hash value in the next index.
5. Determine the e-learning forum answers similarity with the Dice's Similarity Coefficient equation.

In string matching, this algorithm has the complexity  $O(n + m)$  to find substrings with length  $m$  in strings with length  $n$ , but this algorithm has the worst case with complexity  $O(nm)$ . This algorithm uses the hash function as a comparison between the string pattern that is searched with substring in the text, so the execution speed of this algorithm is also determined by the hash function used.

For example, a string pattern with a length of  $m$  characters will be searched in text with length  $n$  characters. If the substring and string pattern have the same hash results, then each character will be checked further. However, if the results of the two are not the

same, then the substring will be shifted to the right. The shift can be carried out up to  $(n - m)$  times. The efficient calculation of hash values when shifting will affect the performance of this algorithm.

Mathematical equation:

$$ts + 1 = (d(ts - T[s + 1]h) + T[s + m + 1]) \bmod q$$

Remarks:

$ts$  = decimal value with the length  $m$  of the substring  $T[s + 1 .. s + m]$ , for  $s = 0, 1, \dots, n - m$

$ts + 1$  = the next decimal value calculated from  $ts$  = radix decimal (base number 10)

$h = d^{m-1}$

$n$  = text length

$m$  = pattern length

$q$  = modulo value

The reduction with  $T[s + 1] * h$  is to eliminate the high-order digit from  $ts$ , multiply the result by 10 to shift one digit to the left, and add a low-order digit with  $T[s + m + 1]$ .

#### IV. RESULT AND ANALYSIS

At this stage, the process of testing plagiarism is discussed using the Rabin-Karp algorithm. Data processing used for testing plagiarism using data mining. Data mining processes according to Jiawei Han et. al as follow [19]:

##### 1. Data Cleaning

This stage in the detection of plagiarism in the Mercu Buana University's e-learning forum uses the Rabin Karp algorithm applied to eliminate punctuation and change to the searched source texts and words are words without capital letters.

##### 2. Data integration

In the process of plagiarism detection of the Mercu Buana University e-learning forum using the Rabin Karp algorithm, this stage is used to combine data obtained from the student forum answers.

##### 3. Data selection

In the process of plagiarism detection of the Mercu Buana University e-learning forum using the Rabin Karp algorithm, this stage is used to select student forum answer data which will be used as sampling for the plagiarism detection process.

##### 4. Data transformation

In the process of plagiarism detection of the Mercu Buana University e-learning forum using the Rabin

Karp algorithm, this stage is used for the data transformation process, by changing the data by dividing the text into gram-grams which are determined by the k-gram value.

5. Data mining

In the process of plagiarism detection of the Mercu Buana University e-learning forum using the Rabin Karp algorithm, this stage is used to identify data patterns. The process carried out to identify data patterns is to search for hash values with the rolling hash function of each gram formed.

6. Pattern evaluation

In the process of plagiarism detection of the Mercu Buana University e-learning forum using the Rabin Karp algorithm, this stage is applied to the Dice's Similarity Coefficient equation because this pattern is easily understood and presented to the user.

7. Knowledge presentation

In the process of plagiarism detection of the Mercu Buana University e-learning forum using the Rabin Karp algorithm, in this stage visualization techniques are displayed in the form of a percentage of the Dice's Similarity Coefficient equation.

For the plagiarism detection test in the e-learning learning system, Mercu Buana University uses the Rabin Karp Algorithm Application. This application is website based on application. In this application, there are several testing criteria based on the Rabin Karp Algorithm for plagiarism detection.

There are five main components used for the plagiarism detection process using the Rabin Karp Application:

- White Insensitivity

The function of white insensitivity is to convert the text in documents into lowercase letters and delete punctuation.

- K-Gram

The k-gram method is used to extract letters from a number of characters from a word that is continuously read from the source text to the end of the document. K-gram is a series of terms with length K. Most of which are used as terms are words. K-gram is a method that is applied to the generation of words or characters. This k-grams method is used to extract letters from a number of characters from a word that is continuously

read from the source text to the end of the document [13].

- Hashing

Hashing is transforming a string into a fixed-length unique value that serves as the string marker. The function to generate this value is called the hash function, while the resulting value is called the hash value

- Fingerprint

Fingerprinting is a method used to detect the accuracy of copies between documents, both all the text contained in documents or only a portion of the text. The working principle of this fingerprinting method is to use the hashing technique. The hashing technique is a function that converts every 9 strings into numbers [15]. For example, Rabin-Karp, Winnowing and Manbe.

- Dice Similarity Coefficient

The Dice Similarity Coefficient is a calculation of similarity by calculating the KG-Gram used in the two documents tested. The similarity value can be calculated by the formula:

$$S = 2C / A + B$$

Where S is the value, C is the same number of K-Gram, and A and B are respectively the number of K-Gram of each string tested. The result is a circle from 0 to 100 where the number is the percentage of the similarity of the 2 strings tested.

- Result of Manual Testing

Teks 1	Teks 2
plagiarisme	plagiarisme
bidang	tingkat
akademik	universitas

The text used for manual testing is in Indonesian.

K-Gram : 5

Basis : 3

Teks 1	Hash	Teks 2	Hash
plagi	13275	plagi	13275
lagia	12706	lagia	12706
agiar	11988	agiar	11988
giari	12498	giari	12498
iaris	12580	iaris	12580

arsim	12394	arsim	12394
risme	13532	risme	13532
ismeb	12992	ismet	13010
smebi	13566	smeti	13620
mebid	12853	metin	13025
ebida	12169	eting	12691
bidan	12074	tingk	13637
idang	12511	ingka	12820
danga	12115	ngkat	13061
angak	12152	gkatu	12570
ngaka	12982	katun	12791
gakad	12316	atuni	12477
akade	12020	tuniv	13978
kadem	12598	unive	13847
ademi	11898	niver	13224
demik	12230	ivers	13057
		versi	13761
		ersit	12725
		rsita	13729
		sitas	13600

Rabin Karp Algorithm Formula:

$$P = \frac{2 * IH}{THA + THB} * 100$$

P : Plagiarism percentage

IH : Similar hash

THA : Total Hash in Text 1

THB : Total Hash in Text 2

Output :

$$P = \frac{2 * 7}{21 + 25} * 100$$

$$P = \frac{14}{46} * 100\%$$

$$P = 0,3043478261 * 100\%$$

$$P = 30,434\%$$

• Result of Systematic Testing

The plagiarism detection test uses a web-based Rabin Karp Algorithm Application. The application can be accessed through this following website: <https://program.arfianhidayat.com/rabinkarp>. For this test using k-gram value = 4. In this paper, the researches use text in Indonesian to perform manual testing.

Input Value:

Text 1 : Pada saat ini, perkembangan teknologi informasi dan komunikasi berkembang dengan sangat pesat.

Text 2 : Pada era globalisasi saat ini, dengan perkembangan teknologi yang sangat pesat berpengaruh serius terhadap nilai dan moral yang ada dalam masyarakat.

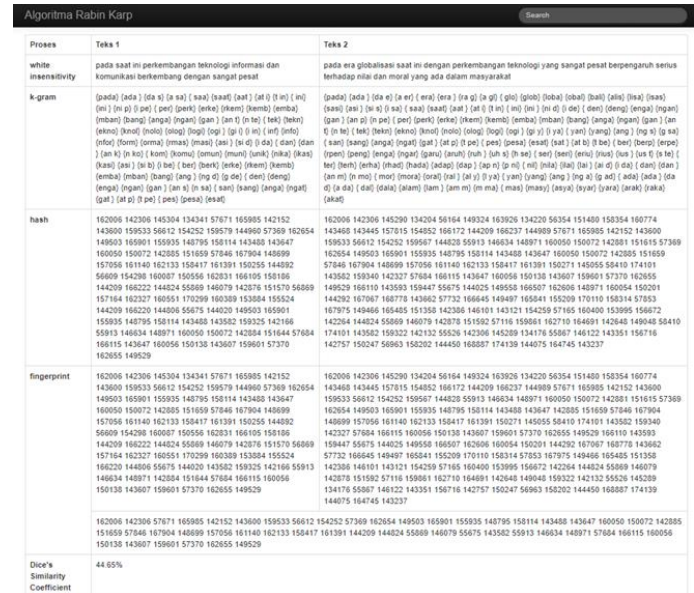


Fig.1. Plagiarism Testing Output

Testing is conducted using the web-based Rabin Karp Algorithm Application. Based on the test results with the test sample using text 1 and text 2, based on the application Rabin Karp made a comparison between text 1 and text 2 to determine the level of plagiarism. The comparison results in the form of percentage similarities. From the results of the test, it can be seen that the Coefficient of Dice Similarity is still below 50%, which is 44.65% so it is not considered plagiarism.

Apart from the limitations in this study, for further development that can be done is the integration of the university's e-learning system database with the detection of plagiarism using the Rabin-Karp Algorithm. For the system, it should be made more user-friendly by applying the form of Graphical User Interface (GUI) on the system. In addition, more varied testing data is needed, such as changing more sentence forms so that the influence of using hashing can be more accurate.

## V. CONCLUSION

Based on the results of the research and discussion presented in the previous chapter, some conclusions can be drawn as follows:

- a. Rabin Karp algorithm can be applied to the Mercu Buana University E-Learning system. The Rabin Karp algorithm is able to detect plagiarism in the answers to student forums. This algorithm is used to calculate plagiarism calculations based on the level of similarity of letters, numbers and symbols per character.
- b. Testing is conducted using the web-based Rabin Karp Algorithm Application. The test uses two samples, namely Text 1 (data tested) and Text 2 (comparative data). The process of detecting plagiarism in this application uses the Rabin Karp Algorithm. The detection testing process using the Rabin Karp Algorithm starts from determining the sampling of Text 1 and Text 2, then entering the gram value. Text 1 and Text 2 will be processed based on white insensitivity, Gram, hashing, and fingerprint. The final results obtained were in the form of a percentage of Dice Similarity Coefficient.
- c. Factors that affect the running time the speed of detecting plagiarism are the length of the text. K-Gram value also affects the level of similarity.
- d. Through the output results in the form of a percentage of Dice Similarity Coefficient, we can see the percentage of the forum's answers to determine whether or not there is a plagiarism action taken.

## REFERENCES

- [1] Salmuasih and A. Sunyoto, -Implementasi Algoritma Rabin Karp untuk Pedeteksi Plagiat Dokumen Teks Menggunakan Konsep Similarity. Semin. Nas. Apl. Teknol. Inf., pp. F23–F28, 2013.
- [2] B. Gipp and N. Meuschke, -Citation Pattern Matching Algorithms for Citation-based Plagiarism Detection: Greedy Citation Tiling, Citation Chunking and Longest Common Citation Sequence Bela Gipp. Proc. 11th ACM Symp. Doc. Eng., pp. 249–258, 2011.
- [3] G. Oberreuter and J. D. Velásquez, -Text mining applied to plagiarism detection: The use of words for detecting deviations in the writing style. Expert Syst. Appl., vol. 40, no. 9, pp. 3756–3763, 2013.
- [4] S. Sukaesih, -Permasalahan Plagiarisme Dalam Penelitian Kualitatif Di Indonesia. J. Polit. Indones., vol. 3, no. 1, p. 210, 2018.
- [5] A. S. Fitriani, Fajrillah, and W. Novarika, -Implementation of Data Mining Using Naïve Bayes Classification Method To Predict Participation of Governor And Vocational Governor Selection in Jemirahan Village, Jabon District. The IJICS, vol. 3, no. 2, pp. 66, 2019.
- [6] Jiawei H., Micheline K., Jian P., Data Mining. 3rd Edition. Waltham, MA: Morgan Kaufmann. 2012: 6-8.
- [7] Z. Efendi and Mustakim, -Text Mining Classification Sebagai Rekomendasi Dosen Pembimbing Tugas Akhir Program Studi Sistem Informasi. Semin. Nas. Teknol. Informasi, Komun. dan Ind. 9, pp. 18–19, 2017.
- [8] Suhendra and I. Ranggadara, -Naive Bayes Algorithm with Chi Square and N-Gram Feature for Reviewing Laptop Product on Amazon Site. IRJCS, vol. 04, issue 12, pp. 29, 2017.
- [9] N. R. Apriliana and J. Nangi, -Implementasi Text Mining Klasifikasi Skripsi Menggunakan Metode Naïve Bayes Classifier. Semant. Vol.3, No.2, Jul-Des 2017, vol. 3, no. 2, pp. 329–343, 2017.
- [10] A. H. Purba and Z. Situmorang, -Analisis Perbandingan Algoritma Rabin-Karp Dan Levenshtein Distance Dalam Menghitung Kemiripan Teks. J. Tek. Inform. Unika St. Thomas, vol. 02, pp. 24–32, 2017.
- [11] A. A. Wiguna and I. Rizqa, -Mengetahui Tingkat Kemiripan Dari Source Code Pada Pemrograman Lisp.
- [12] D. Steveson, H. Agung, and F. Mulia, -RABIN KARP Plagiarisme Detection Applications For Tasks and Problems in School Using Rabin Karp Algorithm. J. Algoritma. Log. dan Komputasi, vol. 1, no. 1, pp. 12–17, 2018.
- [13] I. Widiastuti, C. Rahmad, and Y. Ariyanto, -Aplikasi pendeteksi kemiripan pada dokumen



menggunakan algoritma Rabin Karp. | J. Inform. Polinema, no. 2, pp. 13–18, 2015.

- [14] J. Priambodo, -Pendeteksian Plagiarisme Menggunakan Algoritma Rabin-Karp Dengan Metode Rolling Hash. | J. Inform. Univ. Pamulang, vol. 3, no. 1, pp. 39–45, 2018.
- [15] N. Alamsyah, -Perbandingan Algoritma Winnowing Dengan Algoritma Rabin Karp Untuk Mendeteksi Plagiarisme Pada Kemiripan Teks Judul Skripsi. | Technologia, vol. 8, no. 3, pp. 124–134, 2017.