

Text Summarization in Multi Document Using Genetic Algorithm

Nirwana Hendrastuty*¹, Azhari²

¹Master Program of Computer Science, FMIPA UGM, Yogyakarta, Indonesia

²Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia

e-mail: *nirwanahendrastuty@mail.ugm.ac.id, arison@ugm.ac.id

Abstrak

Peringkasan teks otomatis merupakan representasi dari sebuah dokumen yang berisi intisari atau fokus utama dari dokumen tersebut. Peringkasan teks secara otomatis dilakukan menggunakan metode ekstraksi. Metode ekstraksi melakukan peringkasan dengan menyalin teks yang dianggap paling penting atau paling informatif dari teks sumber menjadi ringkasan [1]. Dokumen dibedakan menjadi dua jenis yaitu single dokumen dan multi dokumen. Multi dokumen merupakan masukan yang berasal dari banyak dokumen dari satu sumber atau lebih yang memiliki lebih dari satu ide pokok.

Penelitian ini bertujuan melakukan peringkasan teks menggunakan Algoritma Genetika dengan memperhatikan ekstraksi fitur teks di setiap kromosom. Ekstraksi fitur yang digunakan yaitu posisi kalimat, kata kunci positif, kata kunci negatif, kemiripan antar kalimat, kalimat yang mengandung kata entitas, kalimat yang mengandung angka, panjang kalimat, koneksi antar kalimat, jumlah koneksi antar kalimat. Jumlah kromosom yang digunakan adalah setengah dari jumlah aduan masyarakat. Data yang digunakan yaitu data pengaduan masyarakat terhadap pemerintah DIY dari bulan Februari 2018 sampai Juli 2020. Data di dapatkan dari website e-lapor DIY. Dari hasil pengujian di dapatkan rata-rata nilai Precision 1, Recall 0.71, dan nilai f-measure sebesar 0.79.

Kata kunci— *Peringkasan Teks Otomatis, Ekstraksi Fitur, Algoritma Genetika, Pemerintah DIY.*

Abstract

Automatic text summarization is a representation of a document that contains the essence or main focus of the document. Text summarization is automatically performed using the extraction method. The extraction method summarizes by copying the text that is considered the most important or most informative from the source text into a summary [1]. Documents can be divided into two types, namely single documents and multi documents. Multi document is input that comes from many documents from one or more sources that have more than one main idea.

This study aims to summarize the text using a Genetic Algorithm by paying attention to the extraction of text features on each chromosome. The feature extraction used is sentence position, positive keywords, negative keywords, similarity between sentences, sentences containing entity words, sentences containing numbers, sentence length, connections between sentences, the number of connections between sentences. The number of chromosomes used is half of the number of public complaints. The data used is data on public complaints against the DIY government from February 2018 to July 2020. The data is obtained from the e-lapor DIY website. From the test results, the average value of Precision 1, Recall is 0.71, and f-measure value is 0.79.

Keywords— *Automatic Text Sumarization, Feature Extraction , DIY government, Genetic Algorithm.*

1. INTRODUCTION

The government is essentially formed to serve the community and create conditions that allow each member [2]. Currently the government does not only provide services and receive complaints from the public face-to-face, but many websites have been developed by the government. The more the website develops, the more reports accumulate and it is necessary to summarize the complaints reported by the public.

Summary is a representation of a document that contains the essence or main focus of the document. Documents can be divided into two types, namely single documents and multi documents. Multi document is input that comes from many documents from one or more sources that have one main idea, which can also have different main ideas. To make it easier for readers Summarization of a text or several texts can be done automatically. Automatic text summarization is a technique of making a summary of a text automatically which maintains the important points of the original document by utilizing an application running on a computer [1].

Automatic text summarization has two approaches, namely the extraction method and the abstraction method. The extraction method summarizes by copying the text that is considered the most important or most informative from the source text into a summary [1]. The copied word or sentence can be in the form of a main clause, main sentence, or main paragraph. Extractive methods used to maintain the original words that exist in the actual report in order to minimize the difference in meaning is derived.

One method that has been developed to perform automatic text summarization is genetic algorithms. [3], [4] have performed automatic text summarization in a single document in Indonesian using the genetic algorithm method. Based on the explanation above, this study proposes a Genetic Algorithm which is expected to automatically summarize text in multiple documents using the extraction method.

2. METHODS

2.1 System Architecture

At this stage an analysis of the information obtained will be carried out. The system architecture to be built has five parts, including data collection, preprocessing, feature extraction, automatic text summarization and finally evaluated. The architecture to be designed in this study is shown in Figure 1.

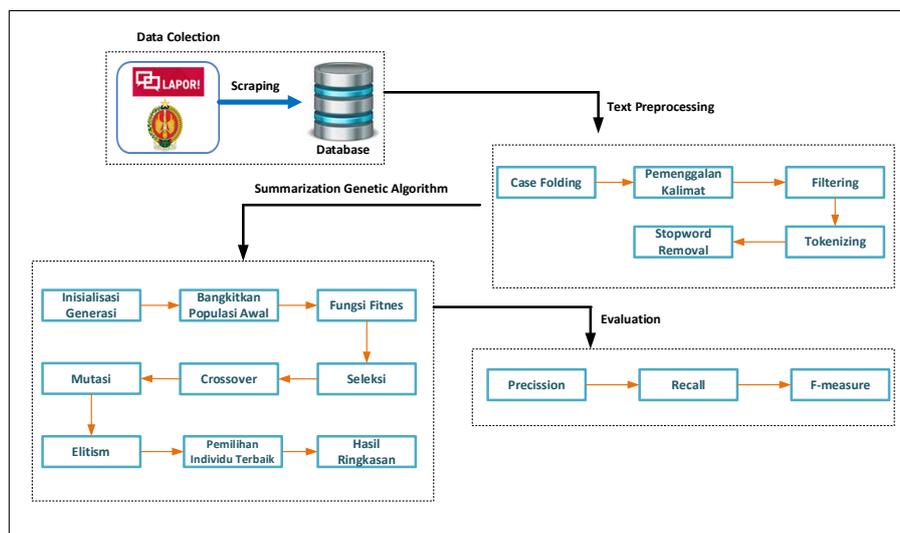


Figure 1 System Architecture

2.2 Data Description

The data in this study used data obtained from public complaint scraping on the DIY e-report website from February 2018 to July 2020. The number of public complaint data used was 1000 community complaints.

2.3 Preprocessing

The preprocessing stage is carried out to manage data on the results of public complaints in order to get data that is cleaner than noise to facilitate further processing. The preprocessing stage can be seen in the following [5]:

Table 1 Preprocessing

Preprocessing	Review
	Mohon bantuan agar traffic light di per4an Mirota swalayan di jl godean di aktifkan kembali.. karena sangat semrawut kondisinya.
Case Folding	mohon bantuan agar traffic light di per4an mirota swalayan di jl godean di aktifkan kembali.. karena sangat semrawut kondisinya.
Pemenggalan Kalimat	mohon bantuan agar traffic light di per4an mirota swalayan di jl godean di aktifkan kembali. karena sangat semrawut kondisinya
Filtering	mohon bantuan agar traffic light di per4an mirota swalayan di jl godean di aktifkan kembali karena sangat semrawut kondisinya
Tokenizing	[mohon, bantuan, agar, traffic, light, di, per4an, mirota, swalayan, di, jl, godean, di, aktifkan, Kembali, karena, sangat, semrawut, kondisinya]
Stopword Removal	[mohon, bantuan, agar, traffic, light, per4an, mirota, swalayan, jl, godean, aktifkan, Kembali, karena, sangat, semrawut, kondisinya]

2.3.1 Case Folding

At this stage, the uniformity of words in a complaint becomes a lower case or all letters in a word become all lowercase.

2.3.2 Sentencing

At this stage, sentences are divided into each community complaint using punctuation marks (.), exclamation (!), Ask (?).

2.3.3 Filtering

This stage is used to remove special characters that are often found in public complaints such as hashtags (#). In addition, it will also remove punctuation marks (.,?!, Etc.), and other characters (\$,% , * , etc.) and scientific symbols ({,}, [,], <,> , etc.).

2.3.4 Tokenizing

This stage performs the process of separating words in one sentence into tokens, where each word in one sentence is separated by a space.

2.3.5 Stopword Removal

This stage serves to remove words that have no influence (which, and, or, to, from, etc.).

2.4 Feature Extraction

Feature extraction is an important process for converting an unstructured textual format into a structured one. The feature extraction used in this study is the same as the feature extraction that has been done by [6].

a. Sentence Position

The sentence position can be calculated using equation (1). Assuming that the sentence in each paragraph is the most important sentence.

$$Score_{f_1}(s) = \frac{X}{N} \quad (1)$$

b. Positive Keywords (f2)

Calculating positive keywords can be done using equation (2). Positive keywords are the words that appear the most in a sentence.

$$Score_{f_2}(s) = \frac{\text{Number of } i^{\text{th}} \text{ keywords in sentence}}{\text{Number of keywords in sentence}} \quad (2)$$

c. Negative Keywords (f3)

Negative keywords can be calculated using equation (3). Negative keywords are the keywords that appear the least in the sentence.

$$Score_{f_3}(s) = \frac{\text{the least number of keywords in the sentence}}{\text{Number of keywords in sentence}} \quad (3)$$

d. Similarities Between Sentences (f4)

The similarity between sentences can be calculated using equation (4). The similarity between sentences is to look for the same word that is on the first chromosome with the word that is on the other chromosome.

$$Score_{f_4}(s) = \frac{\text{keywords in the sentence} \cap \text{keywords between sentence}}{\text{keywords in the sentence} \cup \text{keywords between sentence}} \quad (4)$$

e. Sentence contains Entity (f5)

Sentences containing the word entity can be calculated using equation (5). Sentences containing entities are sentences that contain meanings such as the name of the island, the name of the person, the institution, the place, and so on.

$$Score_{f_5}(s) = \frac{\text{entity name in the sentence}}{\text{sentence length}} \quad (5)$$

f. Sentence contains number (f6)

Sentences containing numbers can be seen in equation (6). Sentences containing numbers are usually considered important.

$$Score_{f_6}(s) = \frac{\text{data that contains numbers in sentences}}{\text{sentence length}} \quad (6)$$

g. Sentence length (f7)

The sentence length can be calculated using equation (7). The length of the sentence aims to find out how long the sentence.

$$Score_{f_7}(s) = \frac{\text{number of words in the sentence}}{\text{unique words in the document}} \quad (7)$$

h. Connection Between Sentences (f8)

The connection between sentences can be calculated using equation (8). The connection between sentences is the number of sentences that have the same word as another sentence.

$$Score_{f_8}(s) = \text{koneksi antar kalimat} \quad (8)$$

i. Number of Connections Between Sentences (f9)

The number of connections between sentences can be calculated using equation (9). The connection between sentences is the number of sentences that have the same word as another sentence.

$$Score_{f_9}(s) = \sum_{k=0}^n \text{connection between sentences} \quad (9)$$

2.5 Genetic Algorithm

Genetic algorithm is a branch of evolutionary algorithms methods adaptively used to solve a search value in an optimization problem on the mechanism of natural selection and the creatures living, in which the algorithm is to follow the principle of natural selection and "whoever is strong, he survived (*survive*) ", By imitating this theory of evolution, genetic algorithms can be used to find solutions to problems that exist in the real world [7].

2.5.1 Generation Initialization

Generation Initialization (Iteration) is carried out to determine how many iterations will be done to get the best individual. In this study, the repetition was done 10, 100, 200, 500 times to get the best individual.

2.5.2 Population Initialization

Population initialization is the stage where the *chromosome string* randomization will be carried out with multiples of 10 populations (popsize) that have been determined at the beginning and the number of chromosomes in each individual is $n/2$ where n is the number of chromosomes in the public complaint data. Examples of individual representations can be seen in Table 2 below:

Table 2 Representation of individuals

P_i	K1	K2	K3	K4	Kn
-------	----	----	----	----	-------	----

After randomization string is completed the next stage is taking the results of the calculation extract features (F1-F9) which has been carried out for each chromosome of each individual [6].

2.5.3 Fitness Function

The fitness function is carried out to determine whether or not the existing solution is in an individual, each individual in the population must have a comparative value (fitness). Through this comparison value, the best solution will be obtained. In this study, each individual will be calculated the fitness value using equation (10) [8].

$$Fitness = \sum_{i=1}^n W_i * (f1_i + \dots + f10_i) \quad (10)$$

2.5.4 Selection

The selection process is a stage for selecting the best individual to be used as a parent for marriage or *crossover*. The selection method used is the *roulette wheel*. On the *roulette wheel* each individual will be represented as a collection of elements in the draw wheel. Each individual is calculated the probability value. This probability value is the value of how likely an individual is to be selected. The higher the probability value of an individual, the more likely that individual is selected as the best individual.

The way this method works is as follows:

1. Calculated the fitness value of each individual (f_i , where i is the 1st to the n th individual).
2. Calculated the total fitness of all individuals.
3. Calculated the probability of each individual.
4. From this probability, the ration of each individual is calculated at numbers 1 to 100.
5. Generated random numbers between 1 and 100.
6. From the random numbers generated, it is determined which individuals will be selected in the selection process.

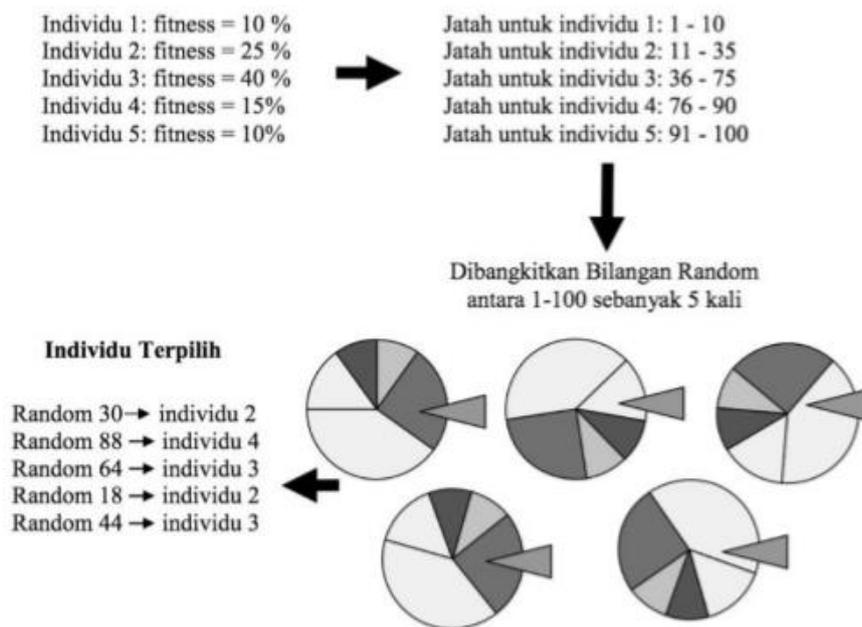


Figure 2 Illustration of selection using the roulette wheel

2.5.5 Crossover

Crossover or cross breeding is a stage that will cross two individuals in a population, to get two new individuals. In this implementation, the cross-breeding method used is *single point crossover*. For *single crossover points*, the chromosomes to be crossover are selected randomly.

2.5.6 Mutations

Mutations are changes in genes in an individual who are born in a new generation. At this stage been some new people as much as the value of mutation probability her. then one of the genes randomly selected from the mutated individual will be changed. The selected gene has a sentence weight will be added with a random value between -1 to 1.

2.5.7 Elitism

Elitism is carried out using the *continuous update* method, namely by adding a new generation to the old population to become individuals. Then do the ranking according to the *fitness* value of each individual. After ranking, the individuals who survive are only the top individuals as many as the population that was initiated at the beginning. The results of elitism will be inputted in the next iteration process.

2.5.8 Selection of the Best Individual

Selection of the best individual is done by sorting each individual based on the highest fitness value. The individual with the highest fitness value is the best individual among other individuals in the same population. The best individual will be used as the result of the *summary*.

3. RESULTS AND DISCUSSION

3.1 Result of Population Generation

The results of population generation are done by taking random sentences (random) on documents per week. Selected sentences will be stored in an *array*. The results of the generation array are shown in Figure 3.

```

New population:
[[ 5 23 24  2 15  3 21 26 12  7  6  1  4]
 [15 19 13 16 20 25 21  5 23 22  8  6 17]
 [18 10 25  2 24  6 12 21  7 16 23  9 20]
 [26  0  4 19  7 10 12 21 20 25 22  6  3]
 [ 2  3 19 10  4  7 21 13 15  0 26  6 24]
 [10 23  1 19 12  3 16  0  7 21 14 18 15]
 [20  2 17 24  5 22 19 18 23  8 14  9 21]
 [ 7 13  9 22  3  4  1 20  0 12  5  8 24]
 [ 7 19 21 20 17 25 23  6 12  2  5  8 26]
 [ 8 11 12 14  6 19 18  7 26  5 16 10 24]]

```

Figure 3 Result of New Population

3.2 Fitness Results

Each individual who has been formed, his fitness value will be calculated. The fitness value is obtained by multiplying the sentence weight by the sentence feature. The fitness results can be seen in Figure 4.

```

Fitness
[525126.08545039 525126.08545039 525126.08545039 525126.08545039
525126.08545039 525126.08545039 525126.08545039 525126.08545039
525126.08545039 525126.08545039 525126.08545039 525126.08545039
525126.08545039 525126.08545039 525126.08545039 525126.08545039
525126.08545039 525126.08545039 525126.08545039 525126.08545039]

```

Figure 4 Fitness Results

3.3 Best Individual Results

Individuals who have the highest fitness value are considered to be the best individuals. The best individual results can be seen in Figure 5.

```

Best Result: 525126.085450387
Individu terbaik: [ 4 18  2 12 19  4 13 12  4  1]

```

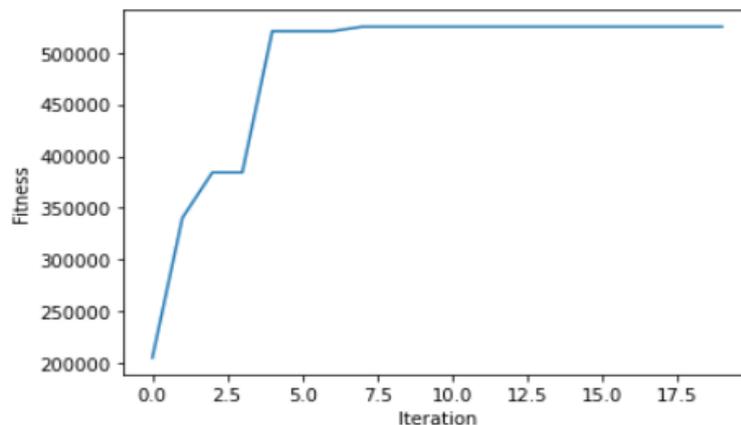


Figure 5 Best Individual Results

The graph in Figure 7 is the result of the *fitness* of each generation. The *fitness* value starts to stabilize in the second iteration because each generation of selected individuals is an individual who has a *fitness* value of 525126.085450387. Data original can be seen in Figure 6 and best individual results who have been returned to the initial sentences can be seen in Figure 7.

Mohon beberapa lampu di Jalan Ring Road Utara atau Jalan Padjadjaran, segera diperbaiki karena mati. Jalan berlubang lampu merah perempatan gejayan dari arah timur .terima kasih.Trans jogja bukan pada jalurnya dan ambil jalur paling kiri.membahayakan pengguna motor. Foto ini saya ambil tanggal 10 Desember 2017. Ada ranting palem yang menyangkut di kabel (entah pln / telpon) cukup berbahaya karen mudah terhembus angin dan lalulintas sangat padat, pernah saya posting di ICJ. Sampai hari ini saya buat laporan, masih menggantung di sana. Padahal lokasi dekat dengan Pos Polisi n imigrasi bahkan bandara. Mohon segera diamankan. Mohon jln Suryowijayan sampai jln KH Wahid Hasyim aspalnya diperbaiki,diratakan krn semenjak ada penambalan, jalan menjadi tidak rata,membahayakan pengemudi kendaraan bermotor, sepeda dll yg melintasi jalan raya tersebut. Tks. Jalan disebelah utara kandang menjangkan kalau hujan airnya selalu menggenang cukup dalam karena open drainnya mungkin buntu. Dan sekarang jalan tersebut rusak parah. Mohon perhatian pihak berwenang. Lampu penerangan jalan arah timur pasar sekarsuli hingga perempatan kidsfun sangat minim. Padahal kondisi lalu lintas sangat ramai. Juga banyak terdapat persimpangan jalan yang rawan akan kecelakaan. Mohon untuk dapat ditindak lanjuti. Lampu lalulintas di perempatan dari arah timur warnanya error.saat lampu itu harusnya sudah berwarna hijau tetapi warna merah juga tetap menyala.sehingga membingungkan para pengguna jalan. Tolong dong dinas yg terkait. Pengamen disepanjang jln malioboro sekarang sudah tidak sopan lagi. Mereka preman, minta maksa2, orang lagi jalan kaki aja dimintain duwit. Tolong malioboro bebaskan dari preman yg berkedok ngamen. Biar jogja nyaman pagi. Tolong dong dinas yg terkait. Pengamen disepanjang jln malioboro sekarang sudah tidak sopan lagi. Mereka preman, minta maksa2, orang lagi jalan kaki aja dimintain duwit. Tolong malioboro bebaskan dari preman yg berkedok ngamen. Biar jogja nyaman pagi.

Figure 6 Original Data

di lampu merah simpanh apmd sebelah utara ada 2 pengamen anak di malam hari. biasanya menunggu di lampu merah sebelah timur. mohon untuk dinas terkait pelayanan puskesmas sanden terutama pada pelayanan pasien ugd, beberapa kali saya berbincang soal pelayanan pada seorang pasien ternyata hasil juga sama pelayanan pasien ugd tidak langsung ditangani padahal ada juga yang dibentak oleh salah satu karyawan puskesmas. trmkasih. jalan ini menjadi salah satu jalan utama mahasiswa menuju ugm. kondisi jalan sudah rusak berat tetapi sepertinya belum ada tindakan lanjut dari pemerintah sekitar. tentunya kami mengharapkan adanya perbaikan segera sehingga tidak perlu ada korban akibat rusaknya jalan ini. saya tadi sampai jalan di pinggir yang mana lalulintas sangat padat ditambah adanya kendaraan yang parkir di sepanjang jalan tersebut yang jelas sekali sudah ada tanda larangan parkir dan sangat berbahaya bagi kami pejalan kaki.

Figure 7 Best individual result that have been returned to the original sentence

3.4 Results of Summarization for Every Category

The summary results for each category are obtained from the best individual summary results. After the best individual summary is obtained, the next step is to categorize it manually. The summary results for each category can be seen in Figure 8.

```

Kategori Pengamen :
di lampu merah simpanh apmd sebelah utara ada 2 pengamen anak di
malam hari

Kategori Pelayanan :
mohon untuk dinas terkait pelayanan puskesmas sanden terutama pada
pelayanan pasien ugd, beberapa kali saya berbincang soal pelayanan
pada seorang pasien ternyata hasil juga sama pelayanan pasien ugd
tidak langsung ditangani padahal ada juga yang dibentak oleh salah
satu karyawan puskesmas

Kategori Jalan :
jalan ini menjadi salah satu jalan utama mahasiswa menuju ugm

```

Figure 8 Results of Summarization for Every Category

3.5 Testing

This test method will discuss the testing phase of the system being built. Testing is done by measuring system performance which includes *precision*, *recall*, and *f-measure* testing against the summary results of public complaints made by genetic algorithms. From the results of this test, it will be known how many iterations and the number of populations that produce the best performance that produces *precision*, *recall*, and *f-measure* between the test results.

The public complaint data used for this test was taken from the DIY e-report website from February 2018 to July 2020 using 297 community complaints. Not all parameter combinations can be done due to limited time and resources, so the parameters tested in this study used 20 individuals and 20 iterations. The test results can be seen in Table 3.

Table 3 Testing Results

Eks	Samplng	Jumlah Data	Jumlah Kategori	Precision	Recall	F-measure
1	Minggu ke 14 Tahun 2018	3	1	1	0,5	0,6
2	Minggu ke 36 Tahun 2018	4	1	1	0,6	0,8
3	Minggu ke 11 Tahun 2018	12	4	1	0,8	0,8
4	Minggu ke 3 Tahun 2019	4	1	1	0,2	0,4
5	Minggu ke 11 Tahun 2019	9	3	1	0,7	0,8
6	Minggu ke 7 Tahun 2019	13	3	1	0,6	0,7
7	Minggu ke 12 Tahun 2020	2	1	1	1	1
8	Minggu ke 14 Tahun 2020	3	1	1	1	1
9	Minggu ke 13 Tahun 2020	18	1	1	1	1

Experiments carried out at this stage used different data sampling and were taken randomly from 2018 to 2020.

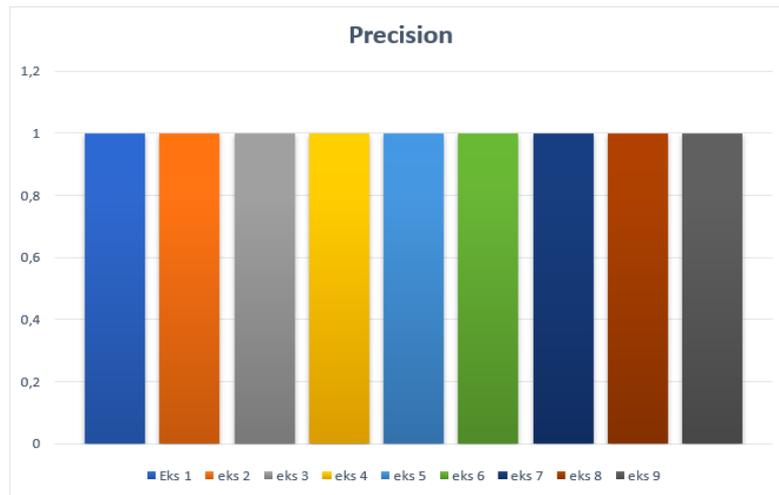


Figure 9 Graph of Precision Values

The experimental results in Figure 9 have the same precision value. This is because the categorization is still done manually and the precision value is obtained from the slice between the system summary results and the data that has been labeled, so that the resulting precision reaches a perfect value.

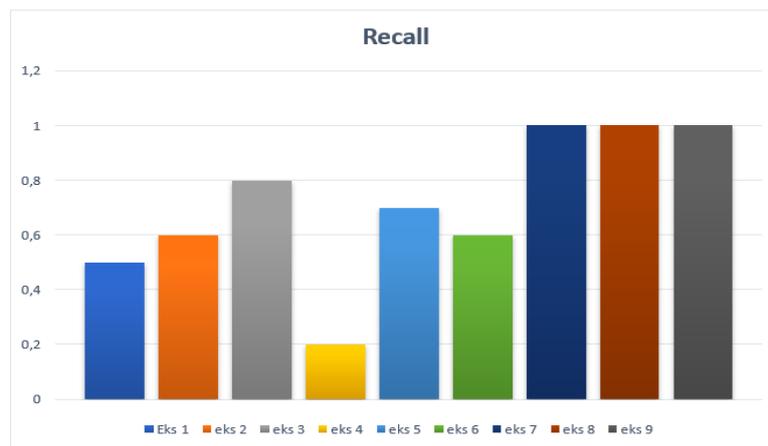


Figure 10 Graph of the Recall Value

The experimental results in Figure 10 have different recall values. The lowest recall value is in experiment four with a recall value of 0.2. This is because the labeling process is still done manually so that when summarizing according to the category, there are many public complaints that the label cannot be detected. The highest recall values were in experiments seven, eight and nine because in the experiment there was only one category so that the labeling was definitely detected.

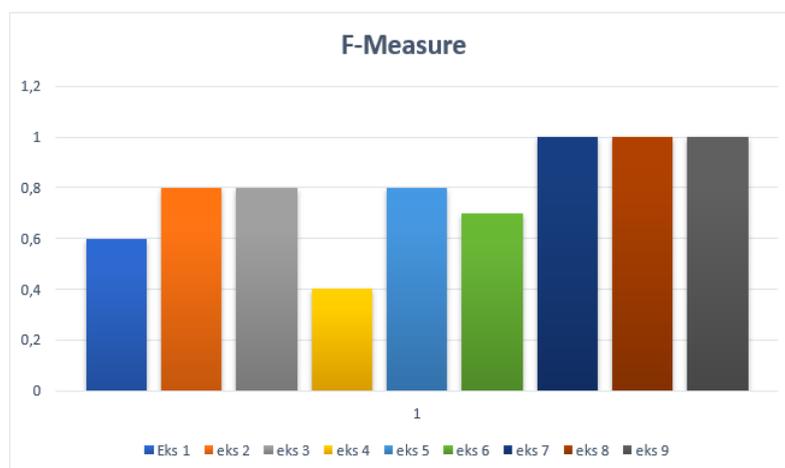


Figure 11 Graph of F-Measure Value

Figure 11 shows the f-measure value of the experiment that was done previously. The f-measure value is determined by the precision value and also the recall value. If the precision value is high and the recall value is high, the resulting f-measure value will be high too, and vice versa, if the precision value and recall value are low, the resulting f-measure value will be low. Because in this study the resulting precision value is the same and the lowest recall value is in the fourth experiment, the lowest f-measure value is in experiment four with a f-measure value of 0.4 and the highest f-measure value is in experiments seven, eight and nine with f-measure value 1.

4. CONCLUSIONS

Based on the test results obtained, the conclusions that can be drawn are as follows:

1. The value of *fitness* are produced at each generation is not always stable. It mentioned in caused due to the process of selection using the method of *the roulette wheel*, so that the individual who has the value of *fitness* highs could just not be selected even though the probability that possess high.
2. Complaints are often lodged by the public in the year 2018 and 2019 most of which relate to the street and the lights either light path or traffic light, because the facility that are used every day. Year 2020 complaint that most many complaints by the public is a category covid. It is because at the beginning of the year 2020 covid began to enter into Indonesia, especially in Yogyakarta area. Covid is a big problem for the community and also the government.
3. From the results of the test algorithm is genetic, a summary of which is produced by the system is already significant and in accordance with the contents of the document. The value of *f-measure* is the best that is generated by the system by using 20 individual and 20 iterations exist in the experiment seven, eight and nine with the value of f-measure amounted to 1. The cause of primary value *f-measure* is perfect is the process of labeling categories were conducted by researchers still at doing it manually.

5. SUGGESTIONS

Based on the research conducted, the suggestions that can be implemented for further research are as follows:

1. In further research, other feature extraction methods can be used so that they can be compared.
2. In the study further can use the classification using *Naïve Bayes* to categorize the results *summary*.
3. In further research, we can add another method to increase the *f-measure value*.
4. In further research, it is hoped that the system can be developed to summarize in other languages, not only Indonesian.

BIBLIOGRAPHY

- [1] T. Zhu and K. Li, "The similarity measure based on LDA for automatic summarization," *Procedia Engineering*, vol. 29, pp. 2944–2949, 2012, doi: 10.1016/j.proeng.2012.01.419.
- [2] R. Rasyid, *Desentralisasi dalam Menjunjung Pembangunan Daerah dalam Pembangunan Administrasi di Indonesia*. Jakarta, 1998.
- [3] P. Rukmana, V. R. ApriliRukmana, P., Aprilia, V.R., SuhaRukmana, P., ApriliRukmana, P., Aprilia, V.R., Suhartono, D. & Wongso, R., 2014, Summarizing Text for Indonesian Language by Using Latent Dirichlet Allocation and Genetic Algorithm, , , August, 20–21.a, V.R., Suharton, D. Suhartono, and R. Wongso, "Rukmana, P., ApriliRukmana, P., Aprilia, V.R., SuhaRukmana, P., ApriliRukmana, P., Aprilia, V.R., Suhartono, D. & Wongso, R., 2014, Summarizing Text for Indonesian Language by Using Latent Dirichlet Allocation and Genetic Igorithm, , , August, 20–21.a, V.," no. August, pp. 20–21, 2014.
- [4] M. A. Fatah and F. Ren, "Automatic Text Summarization," *Open Science Index 13, 2008*, vol. 2, pp. 90–93, 2008.
- [5] V. Srividhya and R. Anitha, "Evaluating Preprocessing Techniques in Text Categorization," no. 2010, p. 3, 2010.
- [6] F. Z. Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," in *M.Sc. Thesis, Appendix D*, vol. pp, 2003, pp. 39–46.
- [7] M. Negnevitsky, "Artificial Intelligence: A Guide to Intelligent Systems," *Second ed. Addison-Wesley*, Jan. 2005.
- [8] J. Pardede, D. Rosmala, and P. J. Artaguna, "Implementasi Algoritma Genetika Pada Aplikasi Peringkat Dokumen Berita Bahasa Indonesia," p. 8, 2018.