

**PERINGKASAN ARTIKEL *ONLINE* BERBASIS ALGORITMA *TEXTRANK*  
DAN *SIMILARITY***

**SKRIPSI**

Oleh :  
**ILMI RASYIDAH FIRDAUS**  
**NIM. 19650115**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**PERINGKASAN ARTIKEL *ONLINE* BERBASIS ALGORITMA  
*TEXTRANK DAN SIMILARITY***

**SKRIPSI**

**Oleh :**  
**ILMI RASYIDAH FIRDAUS**  
**NIM. 19650115**

Diajukan kepada:  
Fakultas Sains dan Teknologi Universitas Islam Negeri (UIN) Malang  
Untuk memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Komputer (S. Kom)

**PROGRAM STUDI TEKNIK INFORMATIKA**  
**FAKULTAS SAINS DAN TEKNOLOGI**  
**UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM**  
**MALANG**  
**2023**

## HALAMAN PERSETUJUAN

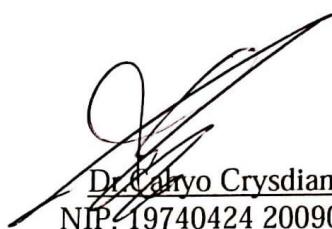
### PERINGKASAN ARTIKEL *ONLINE BERBASIS ALGORITMA TEXTRANK DAN SIMILARITY*

### SKRIPSI

Oleh :  
**ILMI RASYIDAH FIRDAUS**  
**NIM. 19650115**

Telah Diperiksa dan Disetujui untuk Diuji:  
Tanggal: 09 Juni 2023

Pembimbing I,



Dr. Cahyo Crysdiyan MCS  
NIP. 19740424 200901 1 008

Pembimbing II,



Okta Qomaruddin Aziz, S.Si., M.Kom  
NIP. 19911019 201903 1 013

Mengetahui,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Fachru Kurniawan, M.MT, IPM  
NIP. 19771020 200912 1 001

## HALAMAN PENGESAHAN

### PERINGKASAN ARTIKEL *ONLINE BERBASIS ALGORITMA TEXTRANK DAN SIMILARITY*

#### SKRIPSI

Oleh :  
**ILMI RASYIDAH FIRDAUS**  
NIM. 19650115

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Sarjana Komputer ( S.Kom )  
Tanggal: 26 Juni 2023

#### Susunan Dewan Penguji

Ketua Penguji : Dr. M. Amin Hariyadi, M.T (  )  
NIP. 19670018 200501 1 001

Anggota Penguji I : Zainal Abidin, M.Kom (  )  
NIP. 19760613 200501 1 004

Anggota Penguji II : Dr. Cahyo Crysdiyan MCS (  )  
NIP. 19740424 200901 1 008

Anggota Penguji III : Okta Qomaruddin Aziz, S.Si., M.Kom (  )  
NIP. 19911019 201903 1 013

Mengetahui dan Mengesahkan,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Fachru'l Kurniawan, M.MT, IPM  
NIP. 19771020 200912 1 001

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Ilmi Rasyidah Firdaus  
NIM : 19650115  
Fakultas / Jurusan : Sains dan Teknologi / Teknik Informatika  
Judul Skripsi : Peringkasan Artikel *Online* Berbasis Algoritma *TextRank* dan *Similarity*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 26 Juni 2023  
Yang membuat pernyataan,



Ilmi Rasyidah Firdaus  
NIM. 19650115

## **HALAMAN MOTTO**

*... ~ Bagaimanapun hasilnya nanti, hargailah setiap proses yang telah dilalui,  
banggalah pada diri sendiri atas perjuangan untuk terus menjadi yang lebih baik,  
bukan yang terbaik ~ ...*

*~ Jangan terlalu merendahkan diri sendiri, karena hanya kamu yang dapat  
menilai dirimu sendiri dengan baik ~*

## **HALAMAN PERSEMBAHAN**

**الْحَمْدُ لِلّٰهِ رَبِّ الْعَالَمِينَ**

Dengan mengucapkan syukur dan terima kasih, penulis mempersembahkan skripsi ini untuk orang tua, keluarga, dosen, para sahabat dan semua pihak-pihak yang telah memberikan saran, motivasi, doa, serta dukungan sehingga penelitian ini selesai pada waktunya.

## **KATA PENGANTAR**

*Assalamu'alaikum Wr.Wb.*

Puji syukur penulis panjatkan kepada Allah SWT yang senantiasa memberikan rahmat serta kesehatan, sehingga penulis mampu menyelesaikan penelitian ini dengan baik. Penulis menyampaikan ucapan terima kasih kepada semua pihak yang pernah terlibat langsung maupun tidak langsung dalam menyelesaikan penelitian ini, bukan hanya karena usaha keras dari penulis sendiri, akan tetapi karena adanya dukungan dari berbagai pihak. Oleh karena itu penulis berterima kasih kepada:

1. Prof. Dr. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Sri Hariani, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan M.MT., IPM selaku Ketua Prodi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Cahyo Crysian, MCS selaku Dosen Pembimbing 1 yang telah membimbing serta memberikan arahan serta motivasi dalam penulisan skripsi dari awal hingga akhir.
5. Okta Qomaruddin Aziz, S.Si., M.Kom selaku Dosen Pembimbing 2 yang telah memberikan bimbingan, arahan serta bantuan dalam terwujudnya karya tulis skripsi ini dari awal hingga akhir.

6. Dr. Ir. M. Amin Hariyadi. M.T selaku penguji I dan Zainal Abidin M.Kom selaku penguji II yang telah meluangkan waktunya untuk menguji dan dengan sabar memberi arahan dan saran dalam menyelesaikan skripsi ini.
7. Segenap civitas akademik Program Studi Teknik Informatika, dan seluruh dosen yang telah memberikan ilmu serta arahan semasa kuliah.
8. Kedua orang tua penulis, Bapak Imam Bukhori dan Ibu Titi Belistini, serta keluarga besar yang telah memberikan banyak dukungan, doa serta selalu menjadi semangat sehingga penulis mampu menyelesaikan masa studi hingga mencapai gelar sarjana.
9. Teman-teman yang telah memberikan bantuan yang sangat besar dalam pengumpulan dataset penelitian saya, serta yang menjadi teman diskusi dan teman belajar.

Skripsi yang telah ditulis ini masih jauh dari kata sempurna, oleh karena itu penulis sangat menghargai dan senang jika terdapat kritik dan saran yang diberikan. Semoga skripsi ini dapat memberikan manfaat.

*Wassalamu alaikum, Wr. Wb.*

Malang, 26 Juni 2023

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>ii</b>
<b>HALAMAN PERSETUJUAN .....</b>	<b>iii</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>iv</b>
<b>PERNYATAAN KEASLIAN TULISAN .....</b>	<b>v</b>
<b>HALAMAN MOTTO .....</b>	<b>vi</b>
<b>HALAMAN PERSEMBAHAN .....</b>	<b>vii</b>
<b>KATA PENGANTAR.....</b>	<b>viii</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR TABEL .....</b>	<b>xii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>ABSTRAK .....</b>	<b>xiv</b>
<b>ABSTRACT .....</b>	<b>xv</b>
<b>مستخلص البحث.....</b>	<b>xvi</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Pernyataan Masalah .....	4
1.3 Tujuan Penelitian .....	5
1.4 Batasan Masalah .....	5
1.5 Manfaat Penelitian .....	5
<b>BAB II STUDI PUSTAKA .....</b>	<b>7</b>
2.1 <i>Automatic Text Summarization</i> .....	7
2.2 <i>TextRank</i> .....	11
2.3 <i>Cosine Similarity</i> .....	14
2.4 <i>Jaccard Similarity</i> .....	15
<b>BAB III DESAIN DAN IMPLEMENTASI .....</b>	<b>17</b>
3.1 Data Penelitian .....	17
3.2 Desain Sistem.....	17
3.3 <i>Scraping Data</i> .....	18
3.4 <i>Preprocessing</i> .....	19
3.4.1 <i>Splitting Sentences</i> .....	20
3.4.2 <i>Cleaning</i> .....	20
3.4.3 <i>Case Folding</i> .....	22
3.4.4 <i>Stemming</i> .....	22
3.4.5 <i>Tokenizing</i> .....	25
3.4.6 Hapus <i>Stopword</i> .....	25
3.4.7 Hapus Item Kosong pada <i>List</i> .....	27
3.5 Perhitungan <i>Similarity</i> Kalimat.....	27
3.5.1 <i>Similarity</i> Berbasis Word Overlapping.....	27
3.5.2 <i>Jaccard Similarity</i> .....	28
3.5.3 <i>Cosine Similarity</i> .....	30
3.6 Matriks <i>Similarity</i> antar Kalimat.....	34
3.7 Iterasi <i>TextRank</i> .....	36
3.8 Pembuatan Ringkasan .....	39

<b>BAB IV UJI COBA DAN PEMBAHASAN.....</b>	<b>40</b>
4.1 Skenario Uji Coba.....	40
4.2 Data Pengujian .....	42
4.3 Hasil Uji Coba.....	47
4.4 Pembahasan.....	54
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>83</b>
5.1 Kesimpulan .....	83
5.2 Saran .....	84

**DAFTAR PUSTAKA**

**LAMPIRAN**

## DAFTAR TABEL

Tabel 3.1 Contoh Data <i>Splitting Sentence</i> .....	20
Tabel 3.2 Contoh Data <i>Cleaning</i> .....	21
Tabel 3.3 Contoh Data <i>Case Folding</i> .....	22
Tabel 3.4 Contoh Proses <i>Stemming</i> Kata .....	23
Tabel 3.5 Contoh Data <i>Tokenizing</i> .....	25
Tabel 3.6 Contoh Data Hapus <i>Stopword</i> .....	26
Tabel 3.7 Contoh Perhitungan Komponen BoW <i>Cosine Similarity</i> .....	31
Tabel 3.8 Contoh Perhitungan TF-IDF .....	33
Tabel 3.9 Contoh Perhitungan Komponen TF-IDF <i>Cosine Similarity</i> .....	34
Tabel 3.10 Hasil Perhitungan <i>Similarity Word Overlapping</i> .....	36
Tabel 4.1 Data Artikel Uji Coba .....	43
Tabel 4.2 Hasil Pemberian Skor Artikel-1 oleh Annotator.....	44
Tabel 4.3 Perangkingan Skor Total Kalimat Artikel-1 .....	46
Tabel 4.4 Matrix <i>Similarity word overlapping</i> Artikel-1.....	47
Tabel 4.5 Matrix <i>Jaccard Similarity</i> Artikel-1 .....	48
Tabel 4.6 Matrix BoW <i>cosine Similarity</i> Artikel-1.....	48
Tabel 4.7 Matrix TFIDF <i>cosine Similarity</i> Artikel-1 .....	49
Tabel 4.8 Hasil Skor Ground Truth dan Skor Final <i>TextRank</i> .....	50
Tabel 4.9 Hasil Ringkasan Manual dan Prediksi pada Artikel-1 .....	51
Tabel 4.10 Hasil Confusion Matrix pada Artikel-1 .....	51
Tabel 4.11 Hasil ROUGE-1 pada Artikel-1 .....	53
Tabel 4.12 Rata-rata Hasil <i>Accuracy</i> dan <i>F-Score</i> ( <i>Confusion Matrix</i> ).....	54
Tabel 4.13 Rata-rata Hasil <i>Precision</i> , <i>Recall</i> dan <i>F-Score</i> (ROUGE-1) .....	63

## DAFTAR GAMBAR

Gambar 3.1 Desain Sistem.....	18
Gambar 3.2 <i>Flowchart Preprocessing</i> .....	19
Gambar 3.3 <i>Flowchart Cleaning</i> .....	21
Gambar 3.4 <i>Flowchart Stemming</i> .....	24
Gambar 3.5 <i>Flowchart Hapus Stopword</i> .....	26
Gambar 3.6 <i>Flowchart Matrix Dokumen TF-IDF</i> .....	32
Gambar 3.7 <i>Flowchart Matrix Similarity</i> .....	35
Gambar 3.8 <i>Flowchart Iterasi TextRank</i> .....	38
Gambar 4.1 Diagram Rata-rata Hasil <i>Accuracy (Confusion Matrix)</i> .....	55
Gambar 4.2 Diagram Rata-rata Hasil ( <i>Confusion Matrix</i> ).....	59
Gambar 4.3 Diagram Rata-rata Hasil <i>Precision (ROUGE-1)</i> .....	64
Gambar 4.4 Diagram Rata-rata Hasil <i>Recall (ROUGE-1)</i> .....	68
Gambar 4.5 Diagram Rata-rata Hasil (ROUGE-1).....	72

## ABSTRAK

Firdaus, Ilmi Rasyidah. 2023. **Peringkasan Artikel Online Berbasis Algoritma TextRank dan Similarity**. Skripsi. Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (1) Dr. Cahyo Crysian MCS (2) Okta Qomaruddin Aziz, S.Si., M.Kom

**Kata Kunci :** peringkasan teks, similarity, textRank

Informasi yang beredar di internet mengalami pertumbuhan yang cukup pesat. Hal ini memberikan dampak yang signifikan, termasuk meningkatnya kesulitan dalam mencari informasi yang relevan, serta timbulnya overload informasi. Maka untuk mengatasi permasalahan tersebut, dapat dibuat suatu ringkasan teks. Ringkasan teks mampu mengurangi informasi berlebihan dengan cara mengambil intisarinya. Sehingga dengan membaca ringkasan teks dapat membantu pembaca mendapatkan informasi yang relevan dengan cepat. Sistem peringkasan teks dapat dibangun dengan memanfaatkan metode yang memiliki kemampuan untuk mengekstraksi kalimat seperti metode *TextRank*. Proses peringkasan pada *TextRank* memanfaatkan konsep *similarity* untuk mengetahui keterkaitan antar kalimat. Pada penelitian ini akan dilakukan perbandingan hasil peringkasan teks menggunakan algoritma *TextRank* dengan metode *similarity* berbasis *word overlapping*, *Jaccard similarity*, *BOW cosine similarity* dan *TF-IDF cosine similarity*. Berdasarkan hasil uji coba yang telah dilakukan, rata-rata nilai *f-score* terbaik dicapai oleh metode *TextRank* dengan *similarity word overlapping* pada semua kategori *compression rate*. Pada ringkasan 50%, secara berurutan nilai *f-score* dan *accuracy* pada *confusion matrix* sebesar 0,705 dan 0,706, sedangkan nilai *precision*, *recall* dan *f-score* pada ROUGE-1 sebesar 0,82, 0,793 dan 0,772. Pada ringkasan 40%, secara berurutan nilai *f-score* dan *accuracy* pada *confusion matrix* sebesar 0,66 dan 0,728, sedangkan nilai *precision*, *recall* dan *f-score* pada ROUGE-1 sebesar 0,82, 0,793 dan 0,772. Pada ringkasan 30%, secara berurutan nilai *f-score* dan *accuracy* pada *confusion matrix* sebesar 0,588 dan 0,754, sedangkan nilai *precision*, *recall* dan *f-score* pada ROUGE-1 sebesar 0,742, 0,706 dan 0,721. Pada ringkasan 20%, secara berurutan nilai *f-score* dan *accuracy* pada *confusion matrix* sebesar 0,512 dan 0,805, sedangkan nilai *precision*, *recall* dan *f-score* pada ROUGE-1 sebesar 0,665, 0,657 dan 0,656. Pada ringkasan 10%, secara berurutan nilai *f-score* dan *accuracy* pada *confusion matrix* sebesar 0,355 dan 0,871, sedangkan nilai *precision*, *recall* dan *f-score* pada ROUGE-1 sebesar 0,540, 0,546 dan 0,535. Hasil penelitian ini menyatakan bahwa kalimat yang diambil menjadi ringkasan berdasarkan metode *TextRank*, merupakan kalimat dengan keterkaitan yang tinggi dengan kalimat lain.

## ABSTRACT

Firdaus, Ilmi Rasyidah. 2023. **Summarization in Online Articles Based on TextRank and Similarity Algorithm.** Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim Islamic State University Malang. Supervisor : (1) Dr. Cahyo Crysday MCS (2) Okta Qomaruddin Aziz, S.Si., M.Kom

Information on the internet has advanced rapidly. It brings a significant impact, including the difficulty in finding relevant information and the occurrence of overloaded information. To deal with the situation, we can make a text summary to decrease overloaded information by grabbing the text essence. By reading the summary, the reader can acquire relevant information quickly. A text summarization system can be built using sentence extraction methods, like the Textrank method. The Textrank uses the similarity concept to find out the correlation among sentences. The researcher compares the text summarization result using the Textrank algorithm and similarity method based on word overlapping, Jaccard similarity, BOW cosine similarity, and TF-IDF cosine similarity. The try-out result shows that the highest f-score mean is achieved using the Textrank method with similar word overlapping on each category of compression rate. For 50% summary, the f-score and accuracy of the confusion matrix are 0.705 and 0.706, respectively. Meanwhile, the precision, recall, and f-score on ROUGE-1 are 0.82, 0.793, and 0.772. For the 40% summary, the f-score and accuracy of the confusion matrix are 0.66 and 0.728. Meanwhile, the precision, recall, and f-score on ROUGE-1 are 0.82, 0.793, and 0.772. For the 30% summary, the f-score and accuracy of the confusion matrix are 0.588 and 0.754, respectively. Meanwhile, the precision, recall, and f-score on ROUGE-1 are 0.742, 0.706, and 0.721, respectively. For the 20% summary, the f-score and accuracy of the confusion matrix are 0.512 and 0.805. Meanwhile, the precision, recall, and f-score on ROUGE-1 are 0.665, 0.657, and 0.656, respectively. For the 10% summary, the f-score and accuracy of the confusion matrix are 0.355 and 0.871, respectively. Meanwhile, the precision, recall, and f-score of the ROUGE-1 are 0.540, 0.546, and 0.535. The research result shows that the sentences taken for the summary based on the Textrank method are highly correlated with other sentences.

**Keywords :** *similarity, text summarization, textRank*

## مستخلص البحث

فروض، علمي رشيدة ٢٠٢٣ . "تلخيص المقالات عبر الإنترت بناء على خوارزميات ترتيب النص والتشابه". البحث الجامعي. قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج، المشرف الأول: د. جهيو كريسيديان، الماجستير. المشرف الثاني: أوكتا قمر الدين عزيز، الماجستير.

الكلمات الرئيسية: تلخيص النص، التشابه، ترتيب النص.

تشهد المعلومات المتداولة على الإنترت نموا سريعا. وهذا له تأثير كبير، بما في ذلك زيادة صعوبة العثور على المعلومات ذات الصلة، فضلا عن ظهور فائض من المعلومات. لذلك للتغلب على هذه المشاكل، يمكن عمل ملخص نصي. تلخيص النص قادر على تقليل المعلومات الزائدة عن الحاجة عن طريقأخذ جوهرها. لذا فإن قراءة ملخص للنص يمكن أن يساعد القراء في الحصول على المعلومات ذات الصلة بسرعة. يمكن بناء أنظمة تلخيص النص باستخدام طريقة لديها القدرة على استخراج الجمل مثل طريقة ترتيب النص. تستخدم عملية التلخيص في ترتيب النص (*textrank*) مفهوم التشابه لتحديد العلاقة بين الجمل. في هذا البحث، سيتم إجراء مقارنة لنتائج تلخيص النص باستخدام خوارزمية ترتيب النص والتشابه القائمة على تداخل الكلمات (*Jaccard similarity*)، وتشابه جاكارد (*word overlapping*)، وتشابه جيب التمام (*BOW*)، وتشابه جيب التامام (*TF-IDF*). بناء على نتائج التجارب التي أجريت، تم تحقيق أفضل متوسط قيمة الفاء القياسي بواسطة طريقة ترتيب النص مع التشابه بداخل كلمات في جميع فئات معدل الضغط. في ملخص ٥٥٪، كانت قيم درجة الفاء القياسي والدقة على مصفوفة الخلط بالترتيب ٠٠٠٧٥٥ و ٠٠٠٧٦٠، بينما كانت قيم الضبط والاستدعاء ودرجة الفاء القياسي على ROUGE-1 هي ٠٠٠٨٢، ٠٠٠٧٩٣ و ٠٠٠٧٧٢ . في ملخص ٤٠٪، كانت قيم درجة الفاء القياسي والدقة على مصفوفة الخلط ٠٠٠٦٦٥ و ٠٠٠٦٥٧، بينما كانت قيم الضبط والاستدعاء ودرجة الفاء القياسي على ROUGE-1 هي ٠٠٠٧٤٢، ٠٠٠٧٢١ و ٠٠٠٧٢٠ . بينما كانت قيم درجة الفاء القياسي والدقة على مصفوفة الخلط هي ٠٠٠٥١٢ و ٠٠٠٥٥٧، بينما كانت قيم الدقة والاستدعاء ودرجة الفاء القياسي على ROUGE1 هي ٠٠٠٣٥٥ و ٠٠٠٣٥٥، بينما كانت قيم الضبط والاستدعاء ودرجة الفاء القياسي على ROUGE-1 هي ٠٠٠٥٤٦، ٠٠٠٥٤٠ و ٠٠٠٥٣٥ . بينما كانت قيم الضبط والاستدعاء ودرجة الفاء القياسي على طريقة ترتيب النص، هي جمل ذات صلة عالية بالجمل الأخرى.

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Artikel merupakan karya tulis yang memuat informasi mengenai suatu topik, dipublikasikan melalui media cetak maupun *online* untuk dikomunikasikan kepada publik sebagai sarana edukasi dan hiburan (Effendi *et al.*, 2020). Informasi dalam artikel dapat berbentuk fakta maupun opini yang ditulis secara sistematis berdasarkan permasalahan yang diangkat. Topik dalam suatu artikel dapat berisi suatu pembahasan tentang politik, pemerintah, gaya hidup, ekonomi, bisnis, teknologi, kesehatan dan masih banyak lagi. Saat ini, artikel *online* lebih diminati masyarakat dikarenakan kemudahan dalam mengaksesnya. Artikel berbahasa Indonesia dapat diakses secara bebas melalui portal artikel atau berita *online*. Membaca artikel dapat memberi manfaat yang besar jika pembaca dapat mengambil informasi penting dalam artikel.

Diperlukan suatu pemahaman yang baik menggunakan akal dan pikiran untuk memperoleh informasi penting dari suatu teks artikel. Akal dan pikiran merupakan keistimewaan yang diberikan oleh Allah untuk digunakan dengan baik oleh manusia. Orang yang memiliki akal akan dapat mengambil suatu pelajaran sebagaimana firman Allah SWT pada surah Al-Baqarah ayat 269 :

يُؤْتَى الْحِكْمَةَ مَنْ يَشَاءُ وَمَنْ يُؤْتَ الْحِكْمَةَ فَقَدْ أُوتِيَ حَيْرًا كَثِيرًا وَمَا يَذَّكُرُ إِلَّا أُولُوا الْأَلْبَابِ

*“Dia memberikan hikmah kepada siapa saja yang Dia kehendaki. Barang siapa yang diberi hikmah, sesungguhnya dia telah diberi kebaikan yang banyak. Dan tidak ada yang dapat mengambil pelajaran kecuali orang-orang yang mempunyai akal sehat.”* (Al-Baqarah ayat 269)

Berdasarkan tafsir Jalalain, يُوتى الحكمة memiliki maksud bahwa ilmu yang berguna dapat menolong manusia untuk bekerja dan berkarya. Manusia akan diberi hikmah kebahagiaan ketika mereka memiliki ilmu yang berguna. Ilmu dapat diperoleh dari mana saja, termasuk dari suatu artikel. Artikel memuat informasi penting yang dapat menambah wawasan seseorang yang berguna bagi kehidupan.

Salah satu cara yang dapat dilakukan untuk mengambil informasi penting suatu artikel adalah dengan membaca ringkasannya (Saputra et al., 2017). Ringkasan mengandung intisari yang memberikan gambaran umum suatu teks, disajikan secara singkat dengan tetap mempertahankan urutan isi teks. Suatu artikel dapat berisi tulisan yang panjang maupun pendek tergantung dengan jumlah informasi yang disajikan. Semakin panjang ukuran artikel maka semakin banyak informasi yang terkandung di dalamnya, sehingga diperlukan waktu lebih lama untuk membaca dan memahami isi artikel. Dibutuhkan suatu ringkasan untuk menghemat waktu serta tenaga saat membaca dan memahami suatu teks artikel. Pembuatan ringkasan teks dapat mengatasi permasalahan *information overload* pada artikel yang memiliki banyak informasi yang harus cepat dianalisis oleh pembaca (Budiharto & Suhartono, 2014). Jadi, ringkasan teks memberi kemudahan dalam pengambilan informasi penting secara lebih cepat sehingga mampu mempersingkat waktu dalam proses pemahaman isi dokumen.

Peringkasan teks dapat dilakukan dengan dua cara, yaitu secara manual atau secara otomatis. Peringkasan text otomatis merupakan salah satu cabang penting dalam *Natural Language Processing* yang bertujuan untuk memadatkan dokumen yang memiliki ukuran teks panjang menjadi informasi yang singkat dan mudah

dipahami (Joshi *et al.*, 2019). Sistem peringkasan otomatis ini memiliki input berupa teks yang panjang dan output berupa kalimat ringkasan singkat yang berisi informasi penting dokumen.

Terdapat beberapa pendekatan peringkasan teks yaitu secara abstraktif, ekstraktif dan hybrid (Kryściński *et al.*, 2019). Pada penelitian ini, jenis peringkasan teks yang akan dibuat yaitu peringkasan ekstraktif. Hal ini dikarenakan jenis peringkasan ekstraktif memiliki proses yang terbilang cukup mudah dibandingkan pendekatan peringkasan teks yang lain. Peringkasan ekstraktif hanya mengambil sebagian kalimat dari dokumen yang dinilai sebagai representasi dari pokok pembahasan dokumen tanpa adanya parafrase terhadap kalimat yang diringkas. Tujuan utama peringkasan ekstraktif adalah untuk memperoleh informasi yang dinilai paling relevan dari dokumen dan juga untuk menghilangkan redundansi serta kalimat yang kurang penting dalam suatu dokumen.

Sistem peringkasan teks dapat dibangun dengan memanfaatkan metode yang memiliki kemampuan untuk mengekstraksi kalimat-kalimat mana yang akan diambil sebagai output ringkasan. Salah satu metode yang cocok untuk ekstraksi kalimat adalah metode *TextRank*. *TextRank* merupakan metode yang cukup populer digunakan dalam peringkasan teks secara ekstraktif karena memiliki kinerja yang cukup baik. Sampai saat ini metode *TextRank* masih menjadi salah satu pilihan untuk penelitian pada sistem peringkasan teks. *TextRank* memiliki beberapa kelebihan diantaranya tidak memerlukan data *training* sehingga mudah beradaptasi dengan bahasa lain, selain itu *TextRank* dapat menghasilkan hasil ekstraksi yang

sangat pendek (bisa hanya satu kalimat) karena menghasilkan suatu ranking kalimat (Mihalcea & Tarau, 2004).

Proses peringkasan pada *TextRank* memanfaatkan algoritma *similarity* untuk mengetahui hubungan kemiripan antar kalimat. Pada program awal *TextRank*, metode *similarity* kalimat yang digunakan yaitu metode berbasis *content* atau *word overlapping*. Namun, terdapat penelitian lainnya yang menggunakan metode pengukuran kemiripan kalimat lain seperti *cosine similarity* dan *Jaccard similarity*. Mempertimbangkan beberapa penerapan metode kemiripan kalimat yang digunakan dalam proses peringkasan *TextRank*, maka pada penelitian ini akan dilakukan pengukuran dan perbandingan antara beberapa metode kemiripan kalimat. Metode yang dibandingkan yaitu metode *similarity* berbasis *word overlapping*, metode *Jaccard similarity* dan metode *cosine similarity*. Metode *Jaccard similarity* dipilih karena perhitungannya yang sederhana dan memiliki notasi perhitungan yang hampir sama dengan metode berbasis *word overlapping*. Metode *cosine similarity* dipilih karena banyak penelitian peringkasan dengan *TextRank* yang menggunakan *cosine similarity* sebagai metode untuk menghitung kemiripan antar kalimat.

## 1.2 Pernyataan Masalah

- a. Seberapa baik hasil *precision*, *recall*, *f-score* dan *accuracy* ringkasan teks menggunakan algoritma *TextRank* dengan metode *similarity* berbasis *word overlapping*, *Jaccard similarity*, *BOW cosine similarity* dan *TF-IDF cosine similarity*?

- b. Apa saja faktor-faktor yang mempengaruhi hasil *precision*, *recall*, *f-score* dan *accuracy* sistem peringkasan teks berdasarkan metode yang digunakan?

### **1.3 Tujuan Penelitian**

- a. Mengukur dan membandingkan hasil *precision*, *recall*, *f-score* dan *accuracy* algoritma *TextRank* menggunakan metode *similarity* berbasis *word overlapping*, *Jaccard similarity*, *BOW cosine similarity* dan *TF-IDF cosine similarity*.
- b. Menganalisis faktor-faktor yang mempengaruhi hasil *precision*, *recall*, *f-score* dan *accuracy* sistem peringkasan teks berdasarkan metode yang digunakan.

### **1.4 Batasan Masalah**

- a. Data yang digunakan adalah dokumen artikel berbahasa Indonesia yang tersedia di *website artikel online* yang dikumpulkan pada 25 sampai 30 April 2023.
- b. Fokus penelitian adalah pada peringkasan ekstraktif dokumen tunggal.

### **1.5 Manfaat Penelitian**

- a. Manfaat untuk lembaga akademik, sistem peringkasan teks dapat membantu memproses banyak dokumen untuk memperoleh informasi penting yang berkaitan dengan riset studi.
- b. Manfaat untuk instansi perpustakaan, dengan adanya sistem peringkasan teks dapat memudahkan pengguna perpustakaan untuk menentukan apakah suatu dokumen memiliki isi yang selaras dengan kebutuhan mereka secara cepat.
- c. Manfaat untuk media massa (surat kabar, majalah dan situs berita), sistem peringkasan teks dapat digunakan untuk memproses informasi dari berbagai

sumber sebagai referensi penulisan artikel, selain itu dapat digunakan dalam meringkas berita untuk menjelaskan poin-poin penting.

- d. Manfaat untuk pembaca artikel yang memiliki keterbatasan waktu dan memiliki banyak kesibukan, sistem peringkasan teks dapat menjadi solusi dengan menyajikan informasi singkat yang merepresentasikan seluruh isi artikel.
- e. Metode yang digunakan dapat dijadikan referensi untuk penelitian selanjutnya.

## **BAB II**

### **STUDI PUSTAKA**

#### **2.1 Automatic Text Summarization**

*Text summarization* merupakan teknik untuk menghasilkan tulisan yang lebih pendek dari sebuah teks panjang. Tulisan pendek yang dihasilkan harus mengandung informasi yang mengandung intisari dari teks tersebut. Teknik ini bertujuan untuk membantu pemahaman teks dan memungkinkan seseorang untuk memperoleh informasi yang dibutuhkan secara lebih cepat dan efisien (Pabbi & Sindhu, 2021). Gagasan mengenai *Automatic Text Summarization* (ATS) pertama kali diusulkan oleh Luhn (1958) pada penelitian pembuatan abstrak secara otomatis pada artikel majalah dan paper teknikal. Pembuatan ringkasan secara otomatis dilakukan oleh program komputer dengan memangkas sebagian teks sehingga didapatkan satu ringkasan yang menyajikan informasi terpenting sebagai representasi dari seluruh isi dokumen. Ringkasan dokumen harus mengandung informasi yang mencangkup keseluruhan isi dokumen dan memberikan redundansi rendah serta keterbacaan tinggi (Zong *et al.*, 2021).

El-Kassas *et al.* (2021) melakukan survey mengenai *automatic text summarization* (ATS) menghasilkan suatu tinjauan mengenai berbagai aspek ATS seperti klasifikasi, aplikasi dan kategorisasi ATS. Pada aspek klasifikasi, sistem ATS diklasifikasikan berdasarkan beberapa kriteria. Klasifikasi berdasarkan ukuran input dibagi menjadi dua kategori yaitu *single-document* dan *multi-document*. Klasifikasi berdasarkan pendekatan teknik *summarization* dibagi menjadi tiga kategori yaitu *extractive*, *abstractive* dan *hybrid*. Klasifikasi berdasarkan output

ringkasan dibagi menjadi dua yaitu *generic* dan *query-based*. Klasifikasi berdasarkan algoritma ringkasan dibagi menjadi dua yaitu *supervised* dan *unsupervised*. Sistem ATS dapat diterapkan pada beberapa aplikasi seperti untuk ringkasan berita, ringkasan email, ringkasan opini, ringkasan buku, ringkasan novel dan ringkasan artikel ilmiah. Pada peringkasan secara *extractive*, terdapat metode yang dapat digunakan seperti metode berbasis *statistical*, *concept*, *topic*, *graph*, *clustering*, *machine-learning*, *semantic* dan *deep-learning*.

Victor et al. (2019) melakukan penelitian mengenai peringkasan otomatis dari *speech documents recorded* sebagai sinyal audio. Teknik peringkasan yang digunakan yaitu peringkasan ekstraktif. Penelitian tersebut menggunakan enam metode *text summarization* yaitu *Luhn*, *TextRank*, *Lexrank*, *LSA*, *KLSum* dan *SumBasic*. Metode kemiripan kalimat yang digunakan dalam metode *TextRank* yaitu pengukuran kemiripan yang bergantung pada banyaknya kata yang tumpang tindih (*word overlapping*). Dataset yang digunakan yaitu kumpulan data artikel berita berbahasa Inggris dari dataset OWIDSum, DUC2001 dan ICSI-Corpus. Metode evaluasi yang digunakan yaitu ROUGE metrik. Hasil penelitian menunjukkan, pada dataset OWIDSum, metode mencapai nilai ROUGE rata-rata yang lebih baik dibandingkan dengan metode lain, yaitu R-1 0.39015, R-2 0.26119, R-L 0.35759, R-W 0.18348, R-S 0.18787, R-SU 0.19145.

Shah & Jivani (2017) melakukan penelitian mengenai *automatic text summarization* dengan kategori *multi-document extractive text summarization* pada data artikel dari Wikipedia berbahasa Inggris. Metode peringkasan teks yang digunakan yaitu *LexRank*, *TextRank*, *LSA* dan *SumBasic*. Metode kemiripan antar

dokumen menggunakan *cosine similarity*. Metode evaluasi yang digunakan yaitu ROUGE metrik. Hasil penelitian menunjukkan bahwa metode *LexRank* dan *TextRank* menunjukkan kinerja yang lebih baik dibandingkan dengan metode LSA dan *SumBasic*.

Gulden et al. (2019) melakukan penelitian mengenai *automatic text summarization* dengan kategori *single-document extractive text summarization*. Dataset yang digunakan yaitu kumpulan data dari deskripsi dan ringkasan singkat dari *trials* yang terdaftar di clinicaltrials.gov. Metode peringkasan teks yang digunakan yaitu *LexRank*, *TextRank*, LSA, Luhn, *SumBasic*, and KLSum. Kinerja terbaik untuk dicapai oleh metode *TextRank* dan skor ROUGE dengan nilai ROUGE-1 F1-score 0.3531, ROUGE-2 F1-score 0.1723, dan ROUGE-L F1-score 0.3003.

Rani & Bidhan (2021) melakukan penelitian mengenai *automatic text summarization* dengan kategori *single-document extractive text summarization*. Terdapat tiga dataset yang digunakan sebagai perbandingan yaitu dataset yang berasal dari penelitian-penelitian sebelumnya yaitu data review dokumen, artikel berita, dan informasi hukum. Metode yang digunakan yaitu *TextRank*, TF-IDF, *Latent Dirichlet Allocation* (LDA). Hasil penelitian menunjukkan metode *TextRank* memiliki kinerja lebih baik dibandingkan dengan pendekatan TF-IDF dan LDA.

Chouigui et al. (2021) melakukan penelitian mengenai *automatic text summarization* dengan kategori *single-document extractive text summarization* pada data korpus berita *multi-source* bahasa Arab. Terdapat dua jenis pendekatan metode yang digunakan yaitu pendekatan berbasis *graph* dan berbasis *statistical*.

Algoritma pendekatan berbasis *graph* yaitu algoritma *LexRank* dan *TextRank*. Algoritma pendekatan berbasis *statistical* yaitu algoritma *Luhn* dan LSA. Evaluasi kinerja sistem peringkas ini dilakukan dengan menghitung metrik ROUGE-1 dan BLEU-1. Hasil penelitian menunjukkan ringkasan menggunakan algoritma *LexRank* mencapai hasil terbaik dengan nilai BLEU-1 sebesar 0,67 dan nilai ROUGE-1 mencapai 0,9.

Elbarougy *et al.* (2020) melakukan penelitian mengenai *automatic text summarization* dengan kategori *single-document extractive text summarization* pada dokumen bahasa Arab. Pendekatan berbasis *graph* diusulkan dengan dilakukan modifikasi pada algoritma PageRank. Modifikasi PageRank dilakukan dengan memberi skor awal pada tiap node, skor dihitung dengan menghitung banyaknya *noun* (kata benda) dalam kalimat. Semakin besar *noun*, maka semakin besar pula informasi yang terkandung dalam kalimat. *Cosine similarity* dengan fitur TF-IDF digunakan untuk menghitung kesamaan antar kalimat. Sistem diuji dengan menetapkan jumlah iterasi yang berbeda untuk menemukan hasil yang terbaik. Hasil ringkasan bergantung pada *compression ratio* dengan mempertimbangkan penghapusan redundansi. Metode yang diusulkan memberikan hasil yang lebih baik ketika jumlah iterasi adalah 10000 dengan nilai *f-measure* yang didapat sebesar 67,68.

Nawaz *et al.* (2020) melakukan penelitian mengenai *automatic text summarization* dengan kategori *single-document extractive text summarization* pada dokumen berita bahasa Urdu. Penelitian tersebut memiliki tujuan untuk memecahkan masalah peringkasan teks Urdu dan untuk memberikan wawasan

tentang berbagai praktik peringkasan teks. Terdapat dua pendekatan perhitungan bobot yang digunakan yaitu pendekatan bobot lokal dan pendekatan bobot global. Penelitian menunjukkan bahwa pendekatan berbasis pembobotan lokal menghasilkan ringkasan ekstraktif yang lebih baik. Nilai *f-score* pada metode bobot kalimat sebesar 80% sedangkan nilai *f-score* pada metode pembobotan *term-frequency* sebesar 76%.

Firman A *et al.* (2022) melakukan penelitian mengenai *automatic text summarization* dengan kategori *single-document extractive text summarization* terhadap dokumen portal berita olahraga berbahasa Indonesia. Metode yang digunakan yaitu metode pembobotan TF-IDF dan *Maximum Marginal Relevance* (MMR). MMR diimplementasikan dalam perhitungan nilai relevansi antara nilai kemiripan antara kalimat dengan dokumen, dan kemiripan antara query dengan kalimat. Nilai kemiripan didapatkan dengan *cosine similarity*. Evaluasi sistem dengan menghitung nilai *precision*, *recall* dan *f-measure* yang menghasilkan nilai secara berurutan sebesar 57,7%, 48,5% dan 50,3%.

## 2.2 TextRank

*TextRank* merupakan algoritma perangkingan berbasis graph dalam pemrosesan teks yang diusulkan oleh Mihalcea & Tarau (2004) sebagai metode *unsupervised learning* untuk ekstraksi kalimat dan kata kunci. Metode *TextRank* merupakan algoritma perangkuman berbasis graf yang dibangun berdasarkan algoritma *PageRank* (Zhou *et al.*, 2022). Graph pada *TextRank*, terdiri dari *vertex* dan *edge*, dimana *vertex* merepresentasikan kalimat pada dokumen sedangkan *edge* merepresentasikan hubungan kemiripan antar kalimat. Konsep utama dari

perangkingan berbasis graph adalah pemberian voting atau rekomendasi ketika suatu *vertex* terhubung ke *vertex* lain. Semakin tinggi jumlah vote yang diberikan pada suatu *vertex*, maka semakin tinggi pula skor *vertex* tersebut. Pada konsep *text summarization*, algoritma *TextRank* diterapkan untuk ekstraksi kalimat untuk mendapatkan kalimat-kalimat penting dalam dokumen. *TextRank* bekerja dalam dua tahap yaitu mengeksplorasi kesamaan antara dua kalimat dalam dokumen dan menghitung kepentingan kalimat (P. Verma et al., 2019).

Mallick *et al.* (2018) mengusulkan metode modifikasi *TextRank* pada penelitian mengenai *single-document summarization*. Metode *TextRank* diusulkan dengan membangun *graph* yang terdiri dari kalimat sebagai *node* dan *similarity* antar kalimat sebagai bobot pada *edge*. Terdapat modifikasi menggunakan *frequency-cosine similarity* untuk memberi bobot berbeda pada kata-kata yang berbeda dalam kalimat. Bobot didapatkan dengan menghitung nilai *tf* dan *isf* tiap kalimat, kemudian dilakukan penghitungan *isf-modified-cosine similarity* untuk mengukur kemiripan antara tiap pasangan kalimat yang digunakan sebagai bobot pada *edge*. Selanjutnya dilakukan perhitungan skor *TextRank* dengan rumus *TextRank* yang telah dimodifikasi. Evaluasi kinerja modifikasi *TextRank* pada penelitian tersebut menunjukkan nilai ROUGE yang lebih tinggi dari penggunaan metode *TextRank* dengan TF-IDF.

Fakhrezi *et al.* (2021) melakukan penelitian tentang *automatic text summarization* dengan metode *TextRank* pada pengembangan ensiklopedia kosakata Al-Quran. Metode pada penelitian tersebut menggabungkan seluruh makna kata kemudian diringkas dengan *TextRank* untuk mendapat hasil ringkasan

berdasarkan seluruh makna kata. Input sistem berupa *query* yang ingin dicari dan interpretasi dari al-Misbah yang digunakan untuk makna *query*. Kemudian dilakukan pemilihan dokumen yang relevan dengan *query* dan meringkas dokumen yang dipilih dengan metode *TextRank*. Evaluasi sistem dilakukan dengan membandingkan hasil ringkasan sistem dengan ringkasan yang dibuat oleh manusia secara manual. Penerapan metode *TextRank* pada penelitian tersebut menghasilkan nilai rata-rata *f-Score* 0,6173.

Yang & Zhang (2019) melakukan penelitian tentang *automatic text summarization* dengan metode *TextRank* pada paper penelitian. Untuk menemukan hubungan antar kalimat, digunakan metode *Jaccard Distance* dengan menghitung persentase kata yang *overlapping* dalam dua kalimat. Algoritme berbasis *TextRank* lebih fleksibel daripada abstrak resmi, karena pengguna dapat menentukan jumlah kalimat yang akan ditampilkan dalam ringkasan, yaitu jumlah informasi.

Cheng & Guo (2022) mengusulkan peringkasan teks pada aplikasi *WeChat* dengan metode *TextRank* yang di improvisasi yang memperhitungkan *user demand* dan fitur kalimat. Improvisasi dilakukan dengan menggunakan korpus eksternal dan fitur kalimat. Metode *Word2Vec* digunakan untuk memberi nilai vektor pada kalimat. Fitur kalimat yang digunakan yaitu dengan mempertimbangkan permintaan *user*, lokasi informasi pada kalimat, kesamaan antar kalimat dan kesamaan antara kalimat dengan judul. Metode improvisasi *TextRank* yang diusulkan menghasilkan ringkasan teks yang lebih baik dalam memenuhi permintaan *user*.

### 2.3 Cosine Similarity

*Cosine similarity* merupakan ukuran kemiripan antara dua vektor pada *inner product space* dan sering digunakan dalam analisis teks (Han et al., 2012). *Cosine similarity* digunakan karena kemampuannya untuk mengukur kesamaan antara vektor dengan cepat dan efektif, serta mampu menangani data yang berdimensi tinggi. *Cosine similarity* juga mudah dikombinasikan dengan berbagai feature extraction.

Sitikhu et al. (2019) melakukan penelitian dengan membandingkan tiga metode yang tidak hanya berfokus pada teks, namun juga penggabungan fitur vektor dan perhitungan *semantic similarity*. Dataset yang digunakan pada penelitian adalah dataset artikel berita *AG's news topic classification*. Terdapat tiga metode yang digunakan yaitu *cosine similarity* dengan TF-IDF, *cosine similarity* dengan *Word2Vec* dan *soft cosine similarity* dengan *Word2Vec*. Hasil akurasi paling besar dicapai oleh metode *cosine similarity* dengan TF-IDF yaitu 76,8%. Pengukuran dengan *word2vec* memberikan hasil yang lebih rendah karena banyak kata yang tidak tersedia dalam kosakata *word2vec* sehingga kata tersebut ditetapkan sebagai nilai nol.

Ubaidillah et al. (2020) menggunakan *Cross Latent Semantic Analysis* (CLSA) dan *cosine similarity* untuk melakukan peringkasan dokumen berita *online*. Penelitian tersebut menggunakan pembobotan kata TF-IDF untuk membuat matriks. Matriks TF-IDF direduksi supaya menjadi lebih sedikit dengan *Singular Value Decomposition* (SVD). Metode CLSA digunakan untuk menentukan hasil ringkasan berdasarkan skor tertinggi. *Cosine similarity* digunakan untuk mengukur

kesamaan antara judul artikel dengan kalimat isi berita. Hasil akurasi ringkasan yang didapat mencapai rata-rata *f-measure*, *precision* dan *recall* berturut-turut sebesar 58%, 0,62% dan 57%.

Sintia *et al.* (2021) melakukan penelitian dengan *cosine similarity* dengan pembobotan TF-IDF dalam sistem pencarian kodefikasi pada aplikasi SiPaDa. Tujuan dari penelitian tersebut adalah untuk meningkatkan akurasi kodifikasi barang. *Cosine similarity* digunakan untuk menghitung kesamaan antara kata kunci pencarian dengan keyword barang. Berdasarkan hasil penelitian, penerapan TF-IDF dan *cosine similarity* berhasil meningkatkan akurasi pencarian kodifikasi barang.

#### **2.4 Jaccard Similarity**

*Jaccard similarity* merupakan metode untuk melakukan perhitungan kemiripan kalimat yang memanfaatkan irisan dan gabungan elemen kata dari kedua himpunan kalimat. Verma *et al.* (2017) menggunakan *Jaccard similarity* berdasarkan clustering untuk mendapatkan polaritas item pada data review user. *Jaccard* mengukur kesamaan antara kumpulan dokumen dalam jumlah terbatas. Definisi ini merujuk kepada perhitungan dengan rasio ukuran irisan dengan ukuran gabungan pada himpunan yang diberikan.

Arifin *et al.* (2018) menggunakan *Jaccard similarity* dengan menginterpolasi pola-pola penting menggunakan sinonim kata untuk mendapatkan hasil perbandingan produk dengan memanfaatkan data opini publik. *Jaccard similarity* digunakan untuk melakukan identifikasi fitur serupa dengan membandingkan kemiripan karakter. Hasil penelitian tersebut menunjukkan bahwa strategi yang diusulkan dapat meningkatkan nilai akurasi hingga mencapai 90%.

Manalu *et al.* (2019) menggunakan *Jaccard similarity* untuk memngembangkan sistem *Document similarity Detector* (DSD). Pengembangan dilakukan pada perhitungan kemiripan dokumen berdasarkan *fitness value*. Penelitian tersebut juga menggunakan algoritma Genetic untuk mengeliminasi beberapa *keyword* untuk mendapatkan *potential keyword* dari dokumen.

## **BAB III**

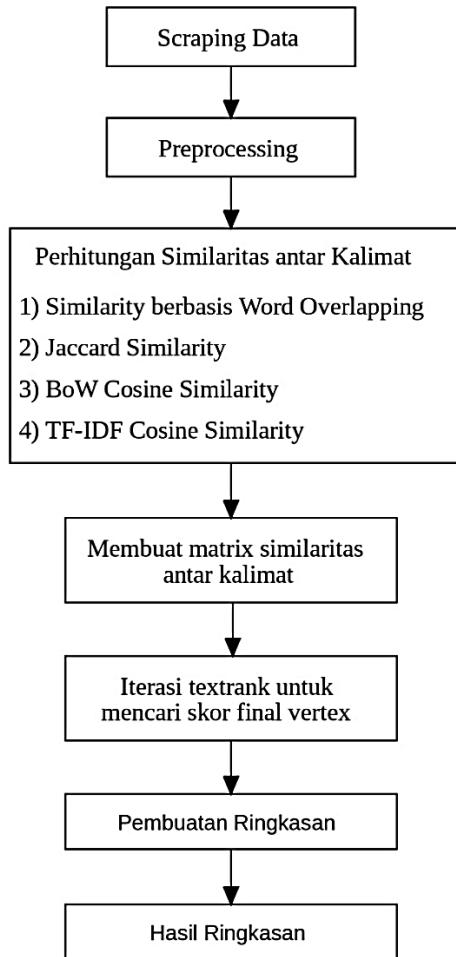
### **DESAIN DAN IMPLEMENTASI**

#### **3.1 Data Penelitian**

Data yang digunakan pada penelitian ini bersumber dari *website* artikel berbahasa Indonesia yang dapat diakses secara bebas, diantaranya: detik.com, cnnindonesia.com, kumparan.com, idntimes.com dan beautynesia.id. Data artikel yang dikumpulkan berjumlah 80 artikel yang dikumpulkan pada 25 April 2023 hingga 30 April 2023. Data artikel yang telah dikumpulkan, diringkas dengan dua cara yaitu secara otomatis oleh sistem dan secara manual oleh annotator (manusia). Data ringkasan manual atau sering disebut ringkasan referensi ini dimanfaatkan sebagai *ground truth* yang nantinya akan digunakan pada saat evaluasi sistem.

#### **3.2 Desain Sistem**

Tahapan penelitian dimulai dengan pengumpulan data berupa teks artikel melalui proses *scraping data*. Kemudian dilakukan *preprocessing* membersihkan data sehingga menghasilkan data berupa *list* kalimat yang tersusun dari token kata dasar. Selanjutnya masuk ke dalam proses inti peringkasan text yang dimulai dengan perhitungan *similarity* antar kalimat. Berdasarkan nilai *similarity* tersebut, maka dibuat *matrix similarity*. Selanjutnya dilakukan pencarian skor tiap kalimat berdasarkan iterasi menggunakan rumus *TextRank*. Kemudian dilakukan pemeringkatan skor kalimat dan dibuat suatu ringkasan. Gambar 3.1 merupakan desain sistem peringkasan text dengan metode *TextRank* berbasis teknik *similarity*.



Gambar 3.1 Desain Sistem

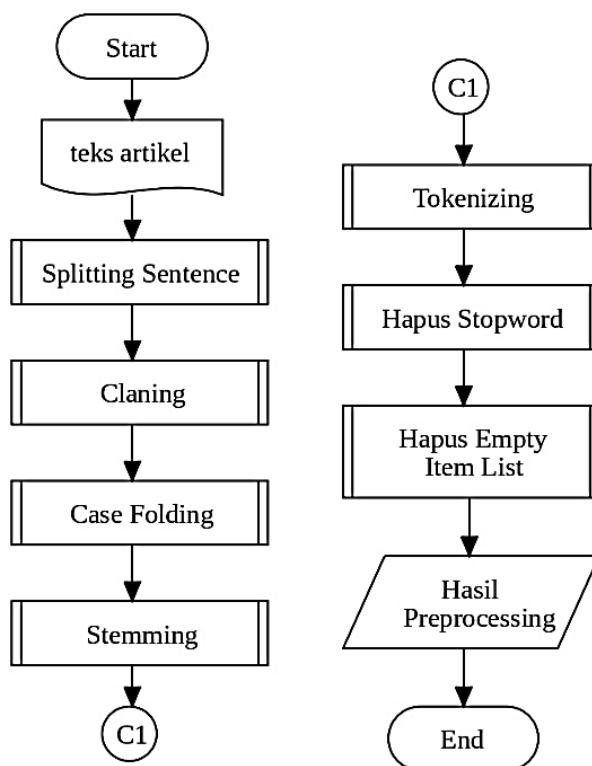
### 3.3 Scraping Data

Pengambilan data pada penelitian dapat dilakukan dengan dua cara yaitu menyalin isi artikel secara langsung dari halaman *website* dan juga dilakukan menggunakan teknik *scraping* pada halaman *website* untuk mempercepat proses pengambilan data. *Web scraping* dilakukan dengan mengambil bagian tertentu dari halaman *website* dan menyimpannya ke dalam *database* atau file sistem untuk selanjutnya akan dianalisis (Zhao, 2017). Pada penelitian ini, *web scraping* dilakukan dengan memanfaatkan *library BeautifulSoup* yang merupakan salah satu

*library python* dengan kemampuan menavigasi struktur HTML, mengekstrak data yang dibutuhkan.

### 3.4 Preprocessing

*Preprocessing* dokumen bertujuan untuk membuat dokumen berupa teks mentah atau teks kotor menjadi teks yang bersih dan siap digunakan pada sistem peringkasan teks. Tahapan *preprocessing* data pada penelitian ini dimulai dari proses memecah dokumen menjadi kalimat tunggal yang disimpan dalam bentuk *list* melalui proses *splitting sentences*. Kemudian dilakukan penghapusan karakter yang tidak penting, *case folding*, *stemming*, *tokenizing*, penghapusan *stopword* dan tahap terakhir yaitu penghapusan item kosong pada *list*. Gambar 3.2. menunjukkan alur proses *preprocessing* data yang dilakukan pada penelitian.



Gambar 3.2 Flowchart Preprocessing

### 3.4.1 *Splitting Sentences*

Pada tahap ini dokumen artikel akan dipecah menjadi kalimat-kalimat yang kemudian disimpan dalam *list* dengan memanfaatkan notasi *regular expression* (*regex*), fungsi *replace* dan fungsi *split* pada *python*. Fungsi *split* dapat memecah text berdasarkan *delimiter* yang telah ditentukan. *Delimiter* yang digunakan untuk memisahkan kalimat yaitu titik (.), tanda tanya (?) dan tanda seru (!). Namun terdapat beberapa kondisi ketika *delimiter* bukan merupakan pemisah antar kalimat, seperti tanda titik pada akronim, pemisah angka, *website*, singkatan nama orang dan singkatan gelar. Data input proses *splitting sentences* berupa data teks artikel utuh yang didapatkan dari proses *scraping data*. Sedangkan outputnya berupa *list* yang terdiri dari item kalimat pada teks artikel.

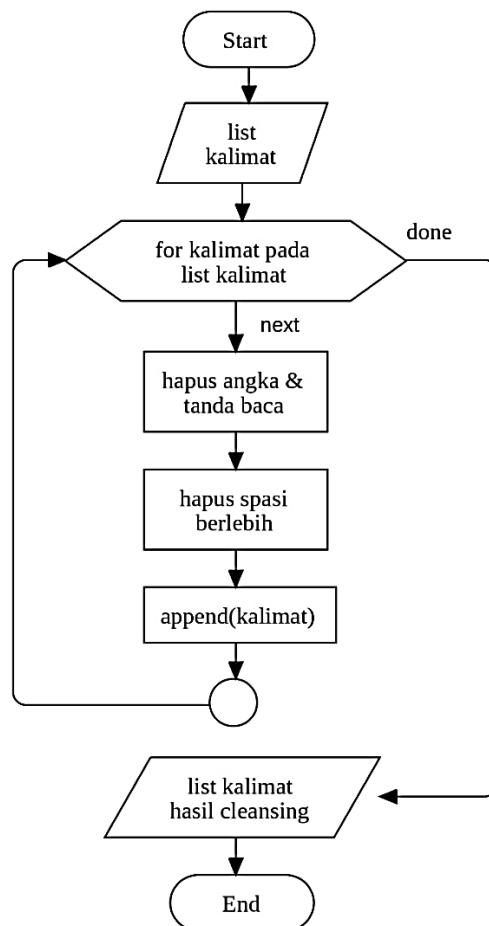
Tabel 3.1 Contoh Data *Splitting Sentence*

<b>Sebelum <i>splitting sentences</i></b>	Indonesia adalah salah satu negara dengan kekayaan alam dan budaya yang sangat beragam. Banyak destinasi wisata yang dapat dikunjungi di berbagai provinsi di Indonesia. Beberapa destinasi wisata terkenal di Indonesia antara lain Bali, Lombok, Yogyakarta, dan Raja Ampat. Bali adalah salah satu destinasi wisata terpopuler di Indonesia. Pulau ini terkenal dengan keindahan pantai dan budaya yang kaya. Selain pantai, Bali juga memiliki tempat wisata lain seperti Pura Besakih, Gunung Batur, dan Ubud yang terkenal dengan seni dan kerajinan. Lombok juga merupakan salah satu destinasi wisata yang terkenal dengan keindahan pantainya.....
<b>Sesudah <i>splitting sentences</i></b>	[‘Indonesia adalah salah satu negara dengan kekayaan alam dan budaya yang sangat beragam.’, ‘Banyak destinasi wisata yang dapat dikunjungi di berbagai provinsi di Indonesia.’, ‘Beberapa destinasi wisata terkenal di Indonesia antara lain Bali, Lombok, Yogyakarta, dan Raja Ampat.’, ‘Bali adalah salah satu destinasi wisata terpopuler di Indonesia.’, ‘Pulau ini terkenal dengan keindahan pantai dan budaya yang kaya.’, ‘Selain pantai, Bali juga memiliki tempat wisata lain seperti Pura Besakih, Gunung Batur, dan Ubud yang terkenal dengan seni dan kerajinan.’, ‘Lombok juga merupakan salah satu destinasi wisata yang terkenal dengan keindahan pantainya.’, .....]

### 3.4.2 *Cleaning*

Pada tahap ini *cleaning* dilakukan untuk penghapusan *noise* pada kalimat seperti angka, tanda baca, serta spasi yang berlebih. *Noise* merupakan karakter yang

tidak diperlukan pada proses analisis teks. Program *cleansing* dibuat dengan memanfaatkan notasi *regular expression (regex)*. Gambar 3.3. merupakan *flowchart* dari proses *cleaning*.



Gambar 3.3 Flowchart Cleaning

Tabel 3.2 merupakan contoh data sebelum dan sesudah dilakukan proses *cleaning*. Data input merupakan *list* kalimat, sedangkan data output merupakan *list* kalimat dengan data yang sudah dilakukan penghapusan angka dan tanda baca.

Tabel 3.2 Contoh Data *Cleaning*

<b>Sebelum <i>cleansing</i></b>	['Beberapa destinasi wisata terkenal di Indonesia antara lain Bali, Lombok, Yogyakarta, dan Raja Ampat.', 'Raja Ampat terletak di Papua dan memiliki lebih dari 1.500 spesies ikan dan terumbu karang yang indah.']}
<b>Sesudah <i>cleansing</i></b>	['Beberapa destinasi wisata terkenal di Indonesia antara lain Bali Lombok Yogyakarta dan Raja Ampat', 'Raja Ampat terletak di Papua dan memiliki lebih dari spesies ikan dan terumbu karang yang indah']

### 3.4.3 Case Folding

*Case folding* dilakukan untuk mengubah huruf besar yang terdapat pada text menjadi huruf kecil dengan memanfaatkan fungsi *lower* pada python. *Case folding* membuat suatu teks menjadi lebih konsisten dengan menyeragamkan format teks, hal ini dapat mengurangi kesalahan akibat penggunaan huruf besar dan kecil. Tabel 3.3 merupakan contoh data sebelum dan sesudah dilakukan proses *case folding*.

Tabel 3.3 Contoh Data *Case Folding*

<b>Sebelum case folding</b>	['Indonesia adalah salah satu negara dengan kekayaan alam dan budaya yang sangat beragam', 'Banyak destinasi wisata yang dapat dikunjungi di berbagai provinsi di Indonesia']
<b>Sesudah case folding</b>	['indonesia adalah salah satu negara dengan kekayaan alam dan budaya yang sangat beragam', 'banyak destinasi wisata yang dapat dikunjungi di berbagai provinsi di indonesia']

### 3.4.4 Stemming

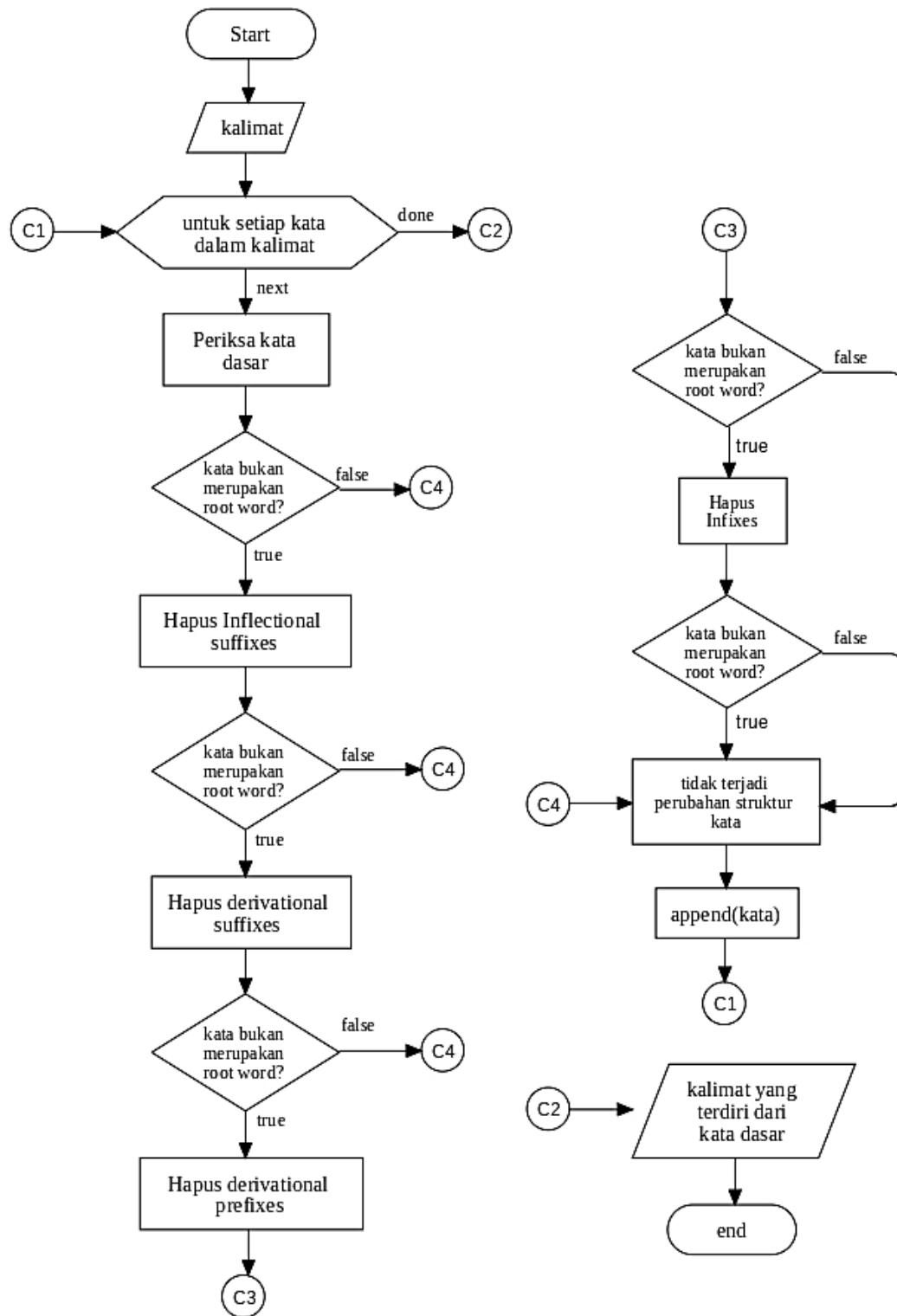
*Stemming* merupakan proses untuk penghapusan imbuhan pada suatu kata untuk mendapatkan kata dasarnya. Algoritma *stemming* yang digunakan dalam penelitian ini yaitu *stemming* menggunakan Sastrawi Bahasa Indonesia. Algoritma ini dikembangkan oleh Adriani et al., (2007) dengan menggunakan aturan Bahasa Indonesia. Proses *stemming* dimulai dari memeriksa tiap kata, apakah kata tersebut merupakan *root word dictionary*. Jika kata tersebut terdapat pada *root word dictionary*, maka kata tersebut merupakan kata dasar sehingga proses *stemming* berhenti. Jika kata tersebut tidak terdapat pada kamus kata dasar, maka dilakukan penghapusan imbuhan pada kata tersebut. Terdapat beberapa imbuhan dalam aturan Bahasa Indonesia, dalam proses penghapusan imbuhan dengan *stemming*, imbuhan yang dihapus pertama kali yaitu *inflectional suffixes* yang terdiri dari *particles suffixes* berupa {-kah, -lah, -tah, -pun} dan *inflectional possessive pronouns suffixes*

berupa -ku, -mu atau -nya. Selama kata yang diinputkan belum terdapat pada *root word dictionary*, maka dilakukan penghapusan untuk jenis imbuhan yang lain yaitu *derivational suffixes* (-i, -kan, -an) dan *derivational prefixes* (be-, di-, ke-, me-, pe-, se-, te-). Terdapat 33 aturan lain dalam proses *stemming* ketika suatu kata memiliki *prefix disambiguation* pada penelitian oleh Adriani et al., (2007). Selain itu, terdapat sisipan yang disebut dengan *infixes* yang berupa -el-, -em-, -er-, dan -in-. *Infixes* adalah jenis afiks yang dimasukkan ke dalam kata dasar atau *root word*. Berbeda dengan *prefixes* yang ditambahkan di awal kata, dan *suffixes* yang ditambahkan di akhir kata, *infixes* diletakkan di dalam kata itu sendiri. Proses *stemming* berhenti saat kata terdapat pada kamus kata dasar. Tabel 3.4 berisi contoh kata sebelum dan sesudah dilakukan proses *stemming*. Data input kata merupakan kata yang terdapat imbuhan. Data output kata merupakan kata yang sudah dihapus imbuhanya sehingga menjadi kata dasar.

Tabel 3.4 Contoh Proses *Stemming* Kata

Kata	Affixes	Pengganti	Hasil
membuat	me- ( <i>prefix</i> )	-	buat
dikunjungi	di- ( <i>prefix</i> ) -i ( <i>suffix</i> )	-	kunjung
keindahan	ke- ( <i>prefix</i> ) -an ( <i>suffix</i> )	-	indah
menyukai	men- ( <i>prefix</i> ) -i ( <i>suffix</i> )	s	suka
menawarkan	men- ( <i>prefix</i> ) -kan ( <i>suffix</i> )	t	tawar
makanan	-an ( <i>suffix</i> )	-	makan
mengunyah	-meng ( <i>prefix</i> )	k	kunyah
telapak	-el- ( <i>infixes</i> )	-	tapak
gemetar	-em- ( <i>infixes</i> )	-	getar
gerigi	-er- ( <i>infixes</i> )	-	gigi

Gambar 3.4 merupakan *flowchart* proses *stemming* menurut penelitian yang dilakukan oleh Adriani et al., (2007).



Gambar 3.4 Flowchart Stemming

### 3.4.5 Tokenizing

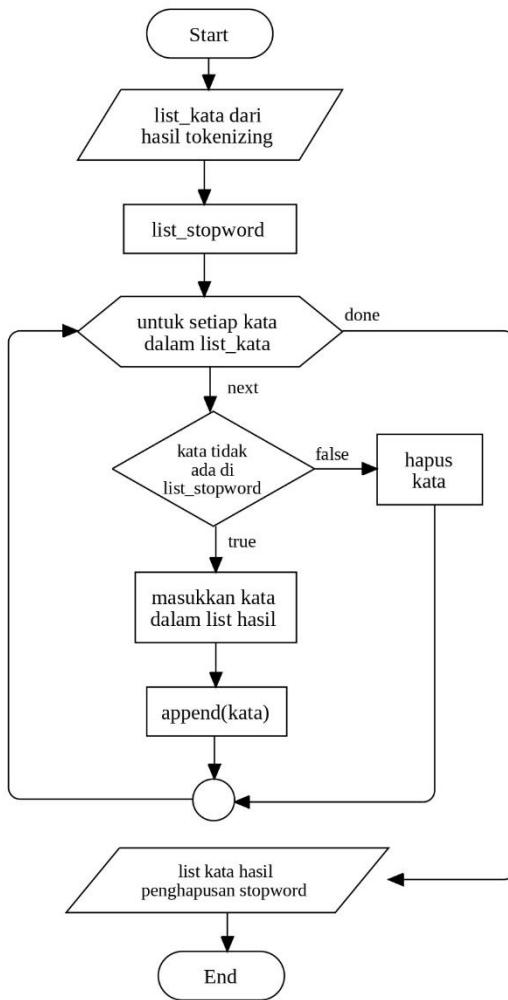
*Tokenizing* dilakukan untuk memecah kalimat menjadi *list* token kata. Tahap ini dilakukan untuk mempermudah pemrosesan teks. *Tokenizing* dapat dilakukan dengan memanfaatkan notasi *regular expression (regex)* dan fungsi *split* untuk memecah kalimat berdasarkan karakter spasi sehingga dihasilkan *list* token kata. Tabel 3.5 merupakan contoh data sebelum dan sesudah dilakukannya *tokenizing*.

Tabel 3.5 Contoh Data *Tokenizing*

<b>Sebelum <i>tokenizing</i></b>	['indonesia adalah salah satu negara dengan kaya alam dan budaya yang sangat ragam', 'banyak destinasi wisata yang dapat kunjung di bagai provinsi di indonesia']
<b>Sesudah <i>tokenizing</i></b>	[['indonesia', 'adalah', 'salah', 'satu', 'negara', 'dengan', 'kaya', 'alam', 'dan', 'budaya', 'yang', 'sangat', 'agam'], ['banyak', 'destinasi', 'wisata', 'yang', 'dapat', 'kunjung', 'di', 'bagai', 'provinsi', 'di', 'indonesia']]

### 3.4.6 Hapus Stopword

*Stopword* merupakan kata yang dinilai kurang memiliki makna dalam suatu teks sehingga tidak banyak berkontribusi dalam proses analisis teks. Penghapusan *stopword* dilakukan untuk meningkatkan kualitas pada hasil analisis teks. Pada tahap ini, *list stopword* didapatkan dari corpus *stopword* Bahasa Indonesia yang disediakan oleh NLTK. Setiap kata dalam kalimat akan dilakukan pemeriksaan. Jika kata terdapat pada *list stopword*, maka kata tersebut dihapus. Jika kata tidak terdapat pada *list stopword* maka kata dimasukkan ke dalam *list* hasil penghapusan *stopword*. Gambar 3.5 merupakan *flowchart* dari proses penghapusan *stopword*.



Gambar 3.5 Flowchart Hapus Stopword

Tabel 3.6 merupakan contoh data sebelum dan sesudah dilakukan penghapusan *stopword*. Data input merupakan *list* kalimat yang terdiri dari token seluruh kata dalam kalimat. Data output merupakan *list* kalimat yang terdiri dari token kata tanpa *stopword*.

Tabel 3.6 Contoh Data Hapus Stopword

<b>Sebelum Hapus Stopword</b>	[['indonesia', 'adalah', 'salah', 'satu', 'negara', 'dengan', 'kaya', 'alam', 'dan', 'budaya', 'yang', 'sangat', 'ragam'], ['banyak', 'destinasi', 'wisata', 'yang', 'dapat', 'kunjung', 'di', 'bagai', 'provinsi', 'di', 'indonesia']]
<b>Setelah Hapus Stopword</b>	[['indonesia', 'salah', 'negara', 'kaya', 'alam', 'budaya', 'ragam'], ['destinasi', 'wisata', 'kunjung', 'provinsi', 'indonesia']]

### 3.4.7 Hapus Item Kosong pada List

*Empty item* merupakan item kosong pada *list* kalimat, dimana semua token kata pada item kalimat tersebut terhapus saat proses *cleaning* dan penghapusan *stopword*. *Empty item* muncul karena adanya kalimat yang hanya terdiri dari *stopword* ataupun hanya berupa angka atau karakter sehingga kalimat tersebut dinilai terlalu penting. *Empty item* dapat menyebabkan *error* saat proses analisis teks, sehingga harus dihapus.

## 3.5 Perhitungan *Similarity* Kalimat

Perhitungan *similarity* dilakukan untuk mengetahui hubungan antar kalimat. Nilai *similarity* ini, nantinya akan digunakan sebagai bobot *edge* dalam pembuatan graf *TextRank*. Pada sistem orisinal *TextRank*, kemiripan antar kalimat diukur berdasarkan *overlapping* kata antara dua kalimat. Perhitungan *similarity* kalimat juga dapat dihitung menggunakan metode lain seperti *cosine similarity* dan *Jaccard similarity*.

### 3.5.1 Similarity Berbasis Word Overlapping

Teknik *similarity* yang digunakan pada sistem orisinal *TextRank* yang diusulkan oleh Mihalcea & Tarau (2004) yang dihitung dengan membagi *content overlap* antara dua kalimat dan panjang tiap kalimat. Panjang kalimat sangat mempengaruhi hasil perhitungan, semakin banyak kata yang *overlapping* maka hasil kemiripannya akan semakin besar. Kemiripan dari kalimat-A dan kalimat-B dapat didefinisikan dengan persamaan 3.1:

$$\text{Similarity}(A, B) = \frac{|\{w_k | w_k \in A \& w_k \in B\}|}{\log(|A|) + \log(|B|)} \quad (3.1)$$

di mana  $\text{Similarity}(A, B)$  merupakan skor *similarity* antara kalimat-A dengan kalimat-B. A merupakan himpunan dari n kata unik yang muncul pada kalimat-A dan B merupakan himpunan dari n kata unik yang muncul pada kalimat-B. Himpunan  $w_k$  merupakan himpunan kata dimana  $w_k$  merupakan elemen atau anggota dari himpunan A dan  $w_k$  juga merupakan elemen dari himpunan B. Perhitungan nilai *similarity* antara kalimat A dan B dilakukan dengan menghitung kardinalitas himpunan  $w_k$ , kemudian membaginya dengan penjumlahan nilai logaritma kardinalitas himpunan A dan B.

Contoh perhitungan *similarity* kalimat berdasarkan *word overlapping* untuk kalimat “Banyak destinasi wisata yang dapat dikunjungi di berbagai provinsi yang ada di Indonesia.” dan “Beberapa destinasi wisata terkenal di Indonesia antara lain Bali, Lombok, Yogyakarta, dan Raja Ampat.” adalah :

$$A = \{ 'destinasi', 'wisata', 'kunjung', 'provinsi', 'indonesia' \}, |A| = 5$$

$$B = \{ 'destinasi', 'wisata', 'kenal', 'indonesia', 'bali', 'lombok', 'yogyakarta', 'raja', 'ampat' \}, |B| = 9$$

$$\{w_k | w_k \in A \& w_k \in B\} = \{ 'destinasi', 'wisata', 'indonesia' \}$$

$$\text{Similarity}(A, B) = \frac{|\{w_k | w_k \in A \& w_k \in B\}|}{\log(|A|) + \log(|B|)} = \frac{3}{\log(5) + \log(9)} = 1.81465$$

### 3.5.2 Jaccard Similarity

*Jaccard similarity* memiliki cara perhitungan yang hampir sama dengan perhitungan *similarity* berdasarkan *word overlapping*. Kedua metode ini sama-

sama menghitung banyaknya kata yang tumpang tindih pada dua himpunan kalimat. Pembeda dari kedua metode ini adalah pada pembaginya. *Jaccard similarity* memiliki nilai antara 0 dan 1, dimana 0 menandakan tidak ada kemiripan pada kedua elemen himpunan, dan 1 menandakan kedua elemen himpunan identik. *Jaccard similarity* memiliki rumus banyaknya elemen pada irisan kedua himpunan dibagi banyaknya elemen pada gabungan antara kedua himpunan. Persamaan 3.2 merupakan rumus dari *Jaccard similarity* :

$$\text{Similarity}(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cup B|} \quad (3.2)$$

di mana A dan B merupakan dua himpunan kalimat, dan  $|A|$  dan  $|B|$  secara berturut-turut merupakan banyaknya elemen himpunan A dan himpunan B.  $|A \cap B|$  merupakan banyaknya elemen yang termasuk ke dalam irisan antara kedua himpunan.  $|A \cup B|$  merupakan banyaknya elemen dari gabungan (*union*) antara kedua himpunan.

Contoh perhitungan *similarity* kalimat berdasarkan *Jaccard similarity* untuk kalimat “Banyak destinasi wisata yang dapat dikunjungi di berbagai provinsi yang ada di Indonesia.” dan “Beberapa destinasi wisata terkenal di Indonesia antara lain Bali, Lombok, Yogyakarta, dan Raja Ampat.” adalah :

$A = \{ \text{'destinasi'}, \text{'wisata'}, \text{'kunjung'}, \text{'provinsi'}, \text{'indonesia'} \}$ ,  $B = \{ \text{'destinasi'}, \text{'wisata'}, \text{'kenal'}, \text{'indonesia'}, \text{'bali'}, \text{'lombok'}, \text{'yogyakarta'}, \text{'raja'}, \text{'ampat'} \}$ ,  $|A| = 5$ ,  $|B| = 9$   
 $A \cap B = \{ \text{'destinasi'}, \text{'wisata'}, \text{'indonesia'} \}$ ,  $|A \cap B| = 3$ ,  $|A \cup B| = 11$

$$\text{Similarity}(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{3}{11} = 0,2728$$

### 3.5.3 Cosine Similarity

*Cosine similarity* dihitung berdasarkan nilai kosinus sudut antara dua vektor. Pada saat melakukan perhitungan kemiripan dengan *cosine similarity*, tiap kalimat harus diubah ke bentuk vektor terlebih dahulu. Kalimat dapat diubah menjadi vektor dengan memanfaatkan *feature extraction* seperti *Bag of Word* (BoW) dan *Term Frequency Inverse Document Frequency* (TF-IDF). *Cosine similarity* memiliki rentang nilai antara 0 dan 1, hasil 0 menandakan kedua kalimat tidak memiliki hubungan kemiripan, sedangkan hasil 1 menunjukkan kedua kalimat memiliki hubungan kemiripan yang identik, sehingga nilai *cosine similarity* lebih tinggi menunjukkan bahwa kedua vektor memiliki lebih banyak kata yang sama. Persamaan *cosine similarity* dari kalimat-A dan kalimat-B dapat didefinisikan pada persamaan 3.3 :

$$\text{Similarity}(A, B) = \cos \theta = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (3.3)$$

di mana A merupakan vektor representasi kalimat 1 dan B merupakan vektor representasi kalimat 2,  $\theta$  merupakan *angle* atau sudut yang terbentuk antara vektor A dan vektor B. Dimensi vektor A dan vektor B adalah sama, maka sebelum memasuki perhitungan *cosine similarity*, tahap pertama yang harus dilakukan yaitu membuat himpunan kata dalam dokumen, panjang himpunan kata ini menjadi nilai panjang dimensi dari kedua vektor. *Cosine similarity* dilakukan dengan menghitung nilai *dot product* A dan B, kemudian dibagi dengan nilai perkalian magnitude (panjang vektor) A dan B.

Teknik BoW mengubah teks menjadi himpunan kata-kata unik, kemudian dilakukan perhitungan frekuensi kemunculan tiap kata unik tersebut pada kalimat. Frekuensi kemunculan kata tersebut dibuat menjadi bentuk vektor dengan panjang dimensi vektor sama dengan banyaknya elemen dalam himpunan kata-kata unik. Nilai pada elemen vektor merepresentasikan jumlah kemunculan kata dalam kalimat.

Contoh perhitungan *similarity* kalimat berdasarkan *cosine similarity* dengan menggunakan *bag of word* untuk kalimat “Banyak destinasi wisata yang dapat dikunjungi di berbagai provinsi yang ada di Indonesia.” dan “Beberapa destinasi wisata terkenal di Indonesia antara lain Bali, Lombok, Yogyakarta, dan Raja Ampat”.

Tabel 3.7 Contoh Perhitungan Komponen BoW *Cosine Similarity*

<b>Set Kata</b>	<b>A</b>	<b>B</b>	<b>A<sup>2</sup></b>	<b>B<sup>2</sup></b>	<b>AB</b>
destinasi	1	1	1	1	1
wisata	1	1	1	1	1
kunjung	1	0	1	0	0
provinsi	1	0	1	0	0
indonesia	1	1	1	1	1
kenal	0	1	0	1	0
bali	0	1	0	1	0
lombok	0	1	0	1	0
yogyakarta	0	1	0	1	0
raja	0	1	0	1	0
ampat	0	1	0	1	0
<b>Jumlah</b>			<b>5</b>	<b>9</b>	<b>3</b>

Berdasarkan perhitungan pada Tabel 3.7 maka dapat dihitung besar kemiripan antara vektor A dan vektor B :

$$\text{Similarity}(A, B) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} = \frac{3}{\sqrt{5}\sqrt{9}} = 0.4472$$

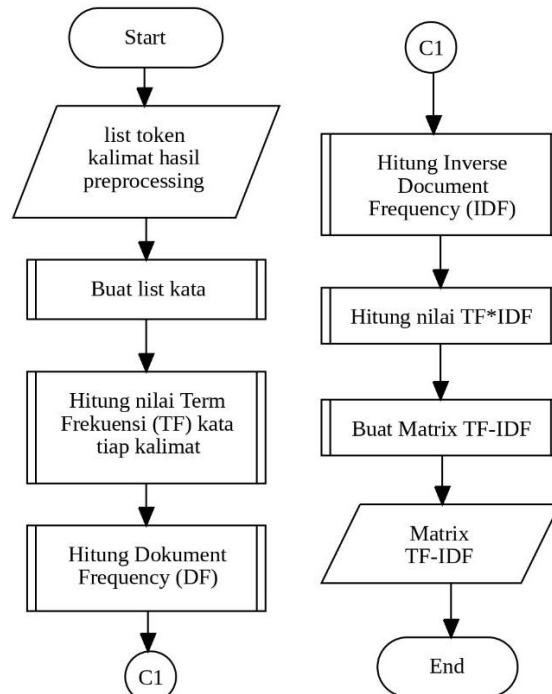
Teknik TF-IDF merupakan pengembangan dari teknik BoW untuk mengubah kalimat menjadi vektor. Teknik ini sering digunakan untuk mengurangi

bobot kata-kata yang sering muncul dalam teks dan meningkatkan bobot kata-kata yang lebih jarang muncul, sehingga menghasilkan representasi teks yang lebih akurat. Pembuatan vektor TF-IDF dimulai dengan membuat himpunan kata yang berisi daftar seluruh kata pada dokumen. Kemudian menghitung *Term Frekuensi* (TF) dan *Inverse Document Frequency* (IDF) pada masing-masing kata dalam kalimat. Skor TF-IDF dihitung melalui perkalian antara nilai TF dan IDF. Nilai TF dihitung dengan membagi banyaknya kata  $i$  pada kalimat  $k$  dengan banyaknya kata dalam kalimat  $k$ .

$$IDF_{(i,d)} = 1 + \log \left( \frac{D}{DF_i} \right) \quad (3.4)$$

$$TF - IDF_{(i,k,d)} = TF_{(i,k)} * IDF_{(i,d)} \quad (3.5)$$

Di mana D merupakan banyaknya kalimat yang terdapat pada dokumen, DF merupakan dokumen frekuensi yaitu banyaknya kalimat yang mengandung kata i.



Gambar 3.6 Flowchart Matriks Dokumen TF-IDF

Setelah mendapatkan skor TF-IDF tiap kata dalam kalimat, selanjutnya yaitu membuat matrik dokumen TF-IDF. Tiap baris matrix menggambarkan vektor kalimat pada dokumen. Gambar 3.6 merupakan *flowchart* pembuatan matrik dokumen TF-IDF.

Contoh perhitungan *similarity* kalimat berdasarkan *cosine similarity* dengan menggunakan TF-IDF untuk kalimat “Banyak destinasi wisata yang dapat dikunjungi di berbagai provinsi yang ada di Indonesia.” dan “Beberapa destinasi wisata terkenal di Indonesia antara lain Bali, Lombok, Yogyakarta, dan Raja Ampat.” jika dimisalkan dalam satu dokumen hanya terdapat dua kalimat tersebut adalah:

$A = ['destinasi', 'wisata', 'kunjung', 'provinsi', 'indonesia']$ ,  $B = ['destinasi', 'wisata', 'kenal', 'indonesia', 'bali', 'lombok', 'yogyakarta', 'raja', 'ampat']$ .

Banyaknya kata pada kalimat A adalah 5, banyaknya kata pada kalimat B adalah 9, banyaknya kalimat pada dokumen adalah 2. Tabel 3.8 menunjukkan perhitungan TF-IDF pada dokumen D yang terdiri dari kalimat A dan kalimat B.

Tabel 3.8 Contoh Perhitungan TF-IDF

Set Kata	TF		DF	D/DF	IDF	TF-IDF	
	A	B				A	B
destinasi	0,2	0,11	2	1	1	0,2	0,11
wisata	0,2	0,11	2	1	1	0,2	0,11
kunjung	0,2	0	1	2	1,3	0,26	0
provinsi	0,2	0	1	2	1,3	0,26	0
indonesia	0,2	0,11	2	1	1	0,2	0,11
kenal	0	0,11	1	2	1,3	0	0,14
bali	0	0,11	1	2	1,3	0	0,14
lombok	0	0,11	1	2	1,3	0	0,14
yogyakarta	0	0,11	1	2	1,3	0	0,14
raja	0	0,11	1	2	1,3	0	0,14
ampat	0	0,11	1	2	1,3	0	0,14

Tabel 3.9 Contoh Perhitungan Komponen TF-IDF *Cosine Similarity*

<b>Set Kata</b>	<b>A</b>	<b>B</b>	<b>A<sup>2</sup></b>	<b>B<sup>2</sup></b>	<b>AB</b>
destinasi	0,2	0,11	0,04	0,012	0,00048
wisata	0,2	0,11	0,04	0,012	0,00048
kunjung	0,26	0	0,06	0	0
provinsi	0,26	0	0,06	0	0
indonesia	0,2	0,11	0,04	0,012	0,00048
kenal	0	0,14	0	0,019	0
bali	0	0,14	0	0,019	0
lombok	0	0,14	0	0,019	0
yogyakarta	0	0,14	0	0,019	0
raja	0	0,14	0	0,019	0
ampat	0	0,14	0	0,019	0
<b>Jumlah</b>			<b>0,24</b>	<b>0,15</b>	<b>0,0144</b>

Berdasarkan perhitungan pada Tabel 3.9 maka dapat dihitung besar kemiripan antara vektor A dan vektor B :

$$\text{Similarity}(A, B) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} = \frac{0,0144}{\sqrt{0,24} \sqrt{0,15}} = 0,07589$$

### 3.6 Matriks *Similarity* antar Kalimat

Matriks *similarity* digunakan untuk merepresentasikan hubungan antar kalimat. Elemen matriks terdiri dari nilai *similarity* dari seluruh kombinasi kalimat dalam dokumen. Nilai *similarity* dihitung menggunakan salah satu metode *similarity* yang telah didefinisikan sebelumnya. Ukuran (*ordo*) dari matriks *similarity* ini adalah  $n \times n$ , di mana  $n$  merupakan banyaknya kalimat dalam text.

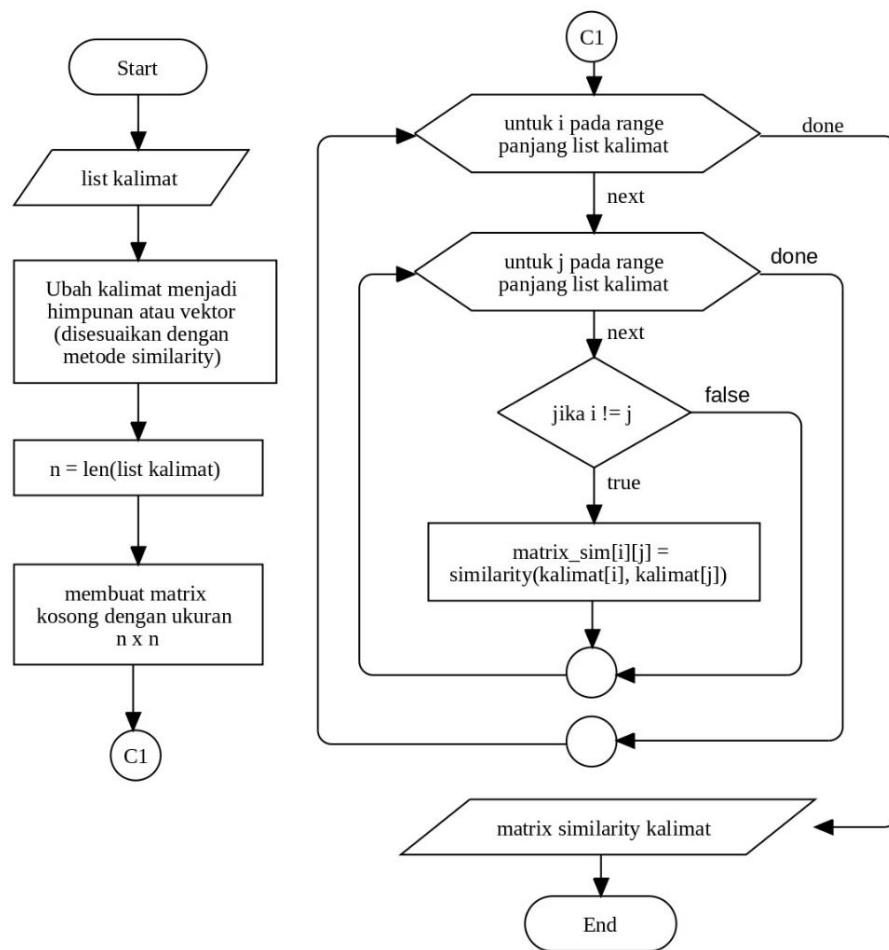
Persamaan 3.6 merupakan persamaan matriks *similarity* antar kalimat.

$$S = \begin{bmatrix} S_{0,0} & S_{0,1} & \cdots & S_{0,n} \\ S_{1,0} & S_{1,1} & \cdots & S_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,0} & S_{n,1} & \cdots & S_{n,n} \end{bmatrix} \quad (3.6)$$

Matriks  $S$  merupakan matriks *similarity* dengan elemen matriks berupa nilai *similarity*  $s_{i,j}$  dimana  $i$  adalah kalimat dalam dokumen pada *index* ke- $i$  dengan nilai

$i = 0, 1, 2, \dots, n$  dan  $j$  adalah kalimat dalam dokumen pada *index* ke- $j$  dengan  $j = 0, 1, 2, \dots, n$ . Sehingga elemen  $s_{0,1}$  merupakan nilai *similarity* antara kalimat pada *index*-0 dengan kalimat pada *index*-1.

Untuk *similarity* dengan *word overlapping* dan *Jaccard similarity*, maka *list* kalimat harus diubah ke bentuk himpunan yang terdiri dari item kata. Sedangkan untuk perhitungan *similarity* dengan *cosine similarity*, *list* kalimat harus diubah menjadi vektor kalimat. Gambar 3.7 merupakan *flowchart* sistem pembuatan matriks *similarity* kalimat.



Gambar 3.7 *Flowchart Matrix Similarity*

Tabel 3.10 Hasil Perhitungan *Similarity Word Overlapping*

Kalimat	<b>K0</b>	<b>K1</b>	<b>K2</b>	<b>K3</b>	<b>K4</b>	<b>K5</b>	<b>K6</b>	<b>K7</b>	<b>K8</b>	<b>K9</b>	<b>K10</b>	<b>K11</b>
<b>K0</b>	0	0,6	0,5	2,5	0	0,9	0	1,5	0	1,0	0,5	3,7
<b>K1</b>	0,6	0	1,8	1,7	0,5	1,1	0,6	1,1	0	1,1	0	1,2
<b>K2</b>	0,5	1,8	0	2,5	1,4	1,9	1,0	1,9	1,0	2,5	0,9	0,5
<b>K3</b>	2,5	1,7	2,5	0	1,8	2,7	1,5	2,3	0,5	1,9	0,4	2,0
<b>K4</b>	0	0,5	1,4	1,8	0	1,3	1,4	1,3	0,5	0,9	0,4	0
<b>K5</b>	0,9	1,1	1,9	2,7	1,3	0	1,4	1,3	0,4	1,8	0,4	0,9
<b>K6</b>	0	0,6	1,0	1,5	1,4	1,4	0	0,4	0	0,5	0,4	0
<b>K7</b>	1,5	1,1	1,9	2,3	1,3	1,3	0,4	0	1,0	0,9	0,4	1,4
<b>K8</b>	0	0	1,0	0,5	0,5	0,4	0	1,0	0	1,0	0	0
<b>K9</b>	1,0	1,1	2,5	1,9	0,9	1,8	0,5	0,9	1,0	0	1,4	1,0
<b>K10</b>	0,5	0	0,9	0,4	0,4	0,4	0,4	0,4	0	1,4	0	0,4
<b>K11</b>	3,7	1,2	0,5	2,0	0	0,9	0	1,4	0	1,0	0,4	0

Tabel 3.10 menunjukkan hasil perhitungan *similarity* terhadap seluruh kalimat yang telah di proses melalui tahap preprocessing. Terdapat 12 kalimat dalam teks dengan *index* 0 sampai 11, sehingga *matrix similarity* yang akan dihasilkan berkuran 12x12. Tiap *cell* pada tabel merepresentasikan item pada matriks dimana *cell* tersebut berisi nilai *similarity* antara kalimat pada kolom *Ki* dan baris *Kj*. Nilai *similarity* untuk suatu kalimat dengan kalimat itu sendiri tidak dihitung karena tidak diperlukan dalam sistem namun pada matriks *similarity* akan diinisialisasi dengan nilai 0.

### 3.7 Iterasi *TextRank*

Iterasi *TextRank* dilakukan untuk menentukan skor final tiap kalimat. Perhitungan skor *textrank* dilakukan secara rekursif dengan rumus yang berbasis graf. Rumus perhitungan skor *TextRank* merupakan pengembangan dari rumus *Pagerank*. Berdasarkan matriks yang telah dibuat sebelumnya, maka perhitungan skor kalimat pada iterasi *TextRank* adalah sebagai persamaan 3.7:

$$WS(V_i) = (1 - d) + d \times \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WV(K_j) \quad (3.7)$$

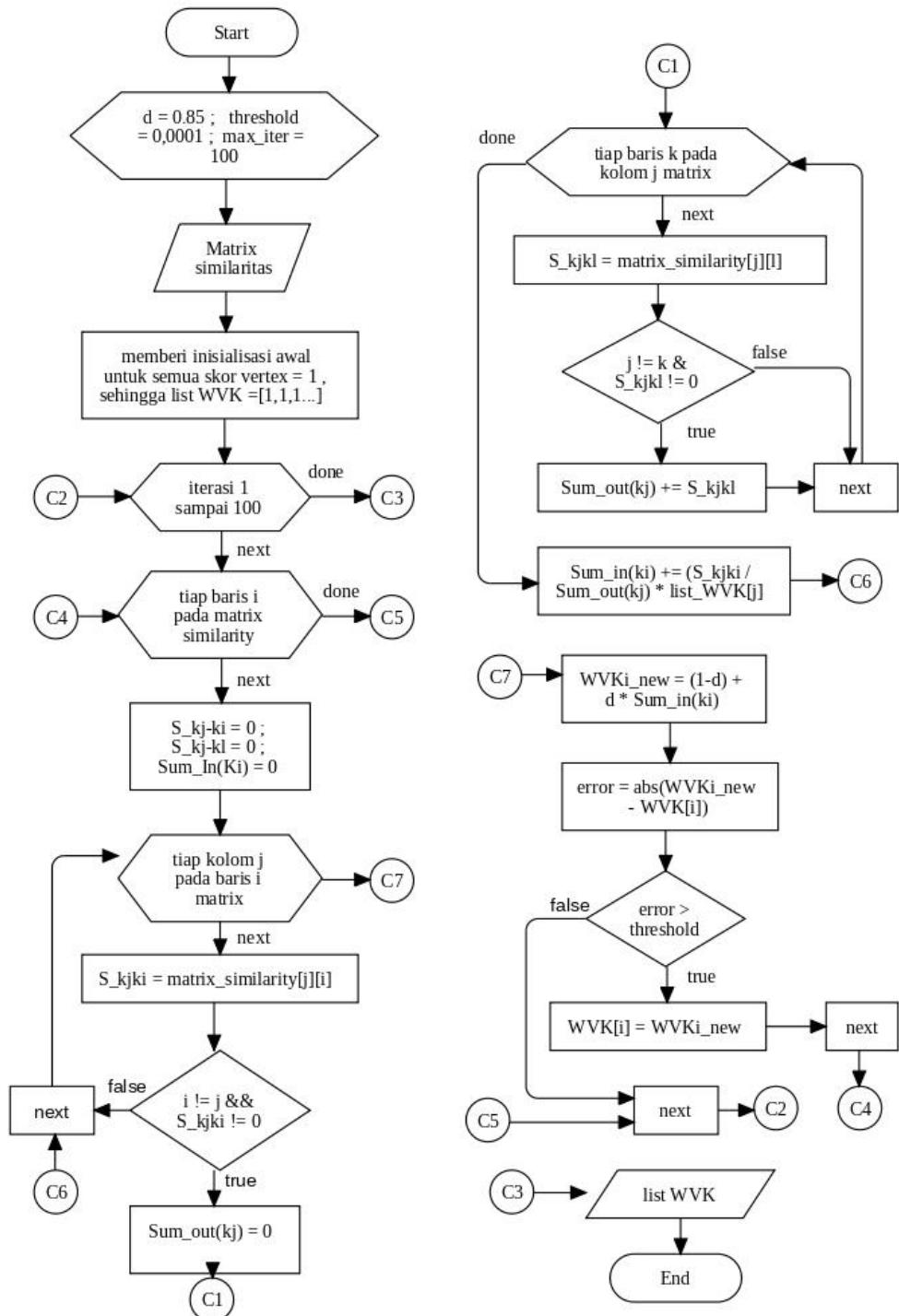
di mana  $WS(V_i)$  merupakan skor kalimat  $V_i$ ,  $d$  merupakan *damping factor* yang memiliki rentang nilai antara 0 sampai 1, dan  $w_{ji}$  merupakan bobot yang berupa nilai kemiripan antara kalimat  $j$  dan kalimat  $i$ . Rumus *tetxrank* ini merupakan rumus yang berbasis pada graf, dimana kalimat direpresentasikan sebagai *vertex* sedangkan nilai kemiripan antar dua kalimat direpresentasikan sebagai *edge*. Fungsi  $In(V_i)$  merupakan suatu fungsi yang menghasilkan *output* yang terdiri dari elemen *vertex* yang mengarah ke *vertex*  $V_i$  (*predecessor*). Sedangkan  $Out(V_j)$  merupakan himpunan yang terdiri dari elemen *vertex* dimana  $V_j$  mengarah ke *vertex* tersebut (*successor*). Jika dalam matriks yang telah dibuat sebelumnya, maka  $In(V_i)$  merupakan fungsi yang menghasilkan *output* berupa kalimat  $V_j$  untuk  $j = 1, 2, 3..n$ , dengan  $V_j \neq V_i$  dan nilai *similarity* antara  $V_j$  dengan  $V_i$  tidak sama dengan 0.

Iterasi *TextRank* dilakukan hingga hasil perhitungan skor tiap *vertex* mencapai konvergensi. Konvergensi dicapai ketika nilai *error rate* setiap *vertex* berada di bawah toleransi. *Error rate* dapat dihitung berdasarkan selisih antara skor *vertex* pada dua iterasi berturut-turut. Persamaan untuk menghitung nilai *error rate* dapat didefinisikan sebagai persamaan 3.8:

$$\text{error rate} = WS^{a+1}(V_i) - WS^a(V_i) \quad (3.8)$$

di mana  $WS^a(V_i)$  merupakan skor *vertex* pada iterasi ke  $a$ , dan  $WS^{a+1}(V_i)$  merupakan pada iterasi selanjutnya. Inisialisasi skor awal *vertex*  $WS(V_i)$  dapat ditentukan secara random, disini skor awal *vertex* diinisialisasi dengan nilai 1. Inisialisasi awal ini tidak mempengaruhi skor final *vertex* dalam iterasi *TextRank*,

namun hanya mempengaruhi jumlah iterasi. Gambar 3.8 merupakan *flowchart* sistem untuk iterasi menggunakan *TextRank*.



Gambar 3.8 Flowchart Iterasi *TextRank*

### 3.8 Pembuatan Ringkasan

Langkah pembuatan ringkasan dimulai dengan mengurutkan skor final *vertex* dalam *descending order* sehingga terbentuk ranking pada kalimat dalam dokumen. Pembentukan ringkasan dilakukan dengan mengelompokkan kalimat berdasarkan n kalimat dengan peringkat awal (*top-n*). Banyaknya kalimat yang diambil (n) dihitung berdasarkan *compression rate* yang telah ditentukan. *Compression rate* adalah perbandingan antara ukuran teks ringkasan dengan ukuran teks asli sebelum diringkas. *Compression rate* menggambarkan banyaknya teks ekstraksi atau dipertahankan sebagai ringkasan teks dengan menggunakan algoritma *TextRank*. *Compression rate* biasanya diukur dalam bentuk persentase. Persamaan 3.9 adalah persamaan untuk menghitung n panjang kalimat ringkasan:

$$nRingkasan = CR * N \quad (3.9)$$

di mana *nRingkasan* merupakan banyaknya kalimat yang diambil menjadi ringkasan, CR merupakan *compression rate*, sedangkan N merupakan banyaknya kalimat pada dokumen asli sebelum diringkas. Setelah didapatkan n kalimat dengan skor tertinggi, maka dilakukan pembentukan ringkasan teks dengan menggabungkan kalimat berdasarkan n kalimat yang diambil. Penggabungan ringkasan tidak mengubah urutan atau susunan awal dari teks dokumen.

## **BAB IV**

### **UJI COBA DAN PEMBAHASAN**

#### **4.1 Skenario Uji Coba**

Uji coba dilakukan pada data teks artikel *online* berbahasa Indonesia dengan jumlah sebanyak 80 artikel. Kemudian dilakukan *preprocessing* teks untuk membersihkan data yang berupa teks mentah menjadi data yang siap dianalisis pada proses peringkasan teks. Tiap artikel akan diolah menggunakan metode *TextRank* dengan empat metode *similarity* kalimat yang berbeda, yaitu :

- a) *TextRank* dengan *similarity word overlapping*,
- b) *TextRank* dengan *Jaccard similarity*,
- c) *TextRank* dengan *cosine similarity* menggunakan vektor BoW, dan
- d) *TextRank* dengan *cosine similarity* menggunakan vektor TF-IDF.

Selanjutnya dibentuk matriks *similarity* antar kalimat pada masing-masing metode *similarity* dan pemberian skor kalimat melalui iterasi *TextRank* berdasarkan matriks *similarity* yang telah dibuat. Skor kalimat *TextRank*, diurutkan secara *descending order* dan diambil n kalimat berdasarkan urutan skor tertinggi (*top-n*) untuk diambil menjadi *output* ringkasan prediksi. Hal ini sesuai dengan penelitian yang dilakukan oleh Mihalcea & Tarau (2004) yang menyatakan bahwa hasil ringkasan diambil berdasarkan kalimat dengan skor *TextRank* tertinggi. Namun, pada penelitian ini akan dilakukan perbandingan hasil ringkasan berdasarkan skor *TextRank* tertinggi (*top-n*) dengan hasil ringkasan berdasarkan skor *TextRank* terendah (*bottom-n*), hal ini dilakukan untuk memperkuat argumen pengambilan ringkasan teks berdasarkan skor *TextRank* tertinggi.

Suatu ringkasan tidak boleh memiliki panjang lebih dari 50% dari keseluruhan isi dokumen asli (Algulihev & Aligulihev, 2009). Maka pada penelitian ini, digunakan nilai *compression rate* sebesar 10%, 20%, 30%, 40%, dan 50%, nilai ini diambil berdasarkan penelitian yang dilakukan oleh Andriani & Furqon (2019) dan Verma et al. (2019). Alasan penggunaan *compression rate* yang berbeda ini adalah untuk mempelajari perilaku kinerja dari teknik yang berbeda.

Hasil ringkasan prediksi akan dievaluasi dengan cara dibandingkan dengan ringkasan referensi (*ground truth*) yang dibuat secara manual oleh manusia. Terdapat dua teknik evaluasi yang digunakan, yaitu menggunakan *confusion matrix* dan ROUGE matrix. *Confusion matrix* untuk mengevaluasi kinerja model berbasis kalimat antara hasil output prediksi ringkasan dengan ringkasan *ground truth*-nya. Parameter yang digunakan sebagai ukuran kinerja sistem dengan *confusion matrix* ini yaitu *recall*, *precision*, *accuracy* dan *f-score*.

$$Precision = \frac{TP}{TP + FP} \quad (4.1)$$

$$Recall = \frac{TP}{TP + FN} \quad (4.2)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.3)$$

$$Precision = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4.5)$$

Di mana TP merupakan banyaknya kalimat pada ringkasan prediksi yang memiliki nilai error kurang dari nilai error minimum terhadap kalimat pada *ground truth*. TN merupakan banyaknya kalimat yang tidak terdapat pada ringkasan prediksi dan tidak pula terdapat pada *ground truth*. FP merupakan banyaknya kalimat yang

terdapat pada ringkasan prediksi namun tidak terdapat pada *ground truth*. FN merupakan banyaknya kalimat yang terdapat pada *ground truth* namun tidak terdapat pada ringkasan prediksi.

Teknik evaluasi kedua yang digunakan yaitu ROUGE *scoring* yang diperkenalkan oleh Lim (2004). Jenis metode ROUGE yang digunakan dalam penelitian ini yaitu ROUGE-1 untuk menghitung skor *recall*, *precision*, dan *F-score* berbasis unigram. Pada dasarnya ROUGE-1 digunakan untuk mengetahui jumlah token kata yang saling tumpang tindih antara ringkasan prediksi dan ringkasan referensi (Allahyari *et al.*, 2017).

$$Recall = \frac{Overlaps\ kata\ antara\ Ringkasan\ Referensi\ &\ Prediksi}{Banyaknya\ kata\ pada\ Ringkasan\ Referensi} \quad (4.6)$$

$$Precision = \frac{Overlaps\ kata\ antara\ Ringkasan\ Referensi\ &\ Prediksi}{Banyaknya\ kata\ pada\ Ringkasan\ Prediksi} \quad (4.7)$$

$$F - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4.8)$$

## 4.2 Data Pengujian

Data pengujian pada penelitian ini sebanyak 80 data artikel yang dikumpulkan dari portal artikel *online* : detik.com, cnnindonesia.com, kumparan.com, idntimes.com dan beautynesia.id. Semua artikel dan ringkasan manualnya, dikumpulkan dan disimpan dalam bentuk *excel*, kemudian tiap dokumen akan diproses satu per satu untuk mendapatkan hasil ringkasan sistem dan juga hasil skor kinerja sistem. Tabel 4.1 menunjukkan seluruh judul dari artikel yang telah dikumpulkan.

Tabel 4.1 Data Artikel Uji Coba

No.	Judul Artikel	Statistik		
		Kalimat	Kata	Unik
1	Kenapa Lengan T-rex Pendek tapi Jadi Predator Ganas?	20	331	186
2	Mahasiswa Kerap Overthinking? Ini 4 Tips Menjaga Kesehatan Mental ala Pakar Itra	26	450	232
3	Kenapa Kita Sering Overthinking pada Malam Hari? Ini Alasannya	37	565	265
4	7 Cara Atasi Overthinking yang Kerap Muncul di Malam Hari, Sudah Lakukan?	41	528	234
5	Tidak Tidur Semalaman Bikin Otak Jadi Tua, Begini Studinya	29	657	277
6	6 Cara Memperbaiki Jadwal Tidur yang Berantakan	29	447	189
7	Bisakah ChatGPT Menggeser Peran Guru di Ruang Kelas?	47	966	373
8	Cuci Muka Pakai Air Dingin atau Air Hangat, Mana yang Lebih Sehat untuk Kulit	35	509	223
...	...	...	...	...
79	Sering Ngeblank dan Susah Fokus, Kenali Gejala Brain Fog dan Cara Mengatasinya	36	506	258
80	3 Cara Memilih Sunscreen yang Tepat untuk Cuaca Terik agar Kulit Terlindungi	21	380	195

Masing-masing artikel dicari ringkasan manual atau *ground truth*-nya terlebih dahulu untuk digunakan dalam proses evaluasi sistem. Ringkasan *ground truth* dibuat dengan cara pemberian skor pada setiap kalimat dalam artikel. Skor diberikan untuk melihat seberapa penting kalimat tersebut untuk dipilih sebagai ringkasan. Pemberian skor antara 0 - 3 untuk setiap kalimat, skor 0 untuk kalimat yang benar tidak penting dan tidak cocok untuk diekstraksi menjadi ringkasan, skor 1 untuk kalimat yang tidak terlalu penting namun sedikit berhubungan dengan seluruh paragraf, skor 2 untuk kalimat yang cukup penting diekstraksi menjadi ringkasan, skor 3 untuk kalimat yang sangat penting untuk diekstraksi menjadi ringkasan (mencangkup intisari artikel dan memiliki keterkaitan tinggi dengan kalimat lain). Setiap kalimat pada satu artikel akan diberi skor oleh lima orang *human annotator*. Skor yang diberikan oleh masing-masing *annotator*, dikumpulkan melalui *form online* kemudian hasil skor akan dijumlah dan dihitung

rata-ratanya kemudian diambil kalimat sebagai ringkasan referensi berdasarkan nilai rata-rata tertinggi. Tabel 4.2 menunjukkan hasil pemberian skor pada artikel pertama oleh lima orang annotator : A1, A2, A3, A4 dan A5.

Tabel 4.2 Hasil Pemberian Skor Artikel-1 oleh Annotator

Id K	Kalimat	Skor Annotator					Total	Rata rata
		A1	A2	A3	A4	A5		
0	Tyrannosaurus rex dikenal sebagai dinosaurus karnivora ganas tetapi berlengan pendek.	3	2	0	3	1	9	1,8
1	Tidak hanya T-rex, kerabatnya di kelompok theropoda atau dinosaurus berjari kaki tiga juga punya lengan dan tangan pendek dibandingkan dengan besar kepala dan badannya.	3	3	2	3	3	14	2,4
2	Kenapa, ya?	0	0	0	0	0	0	0
3	Hipotesis Evolusi Lengan.	0	0	1	0	0	1	0,2
4	Penelitian Kevin Padian di jurnal Acta Paleontologica Polonica menunjukkan, lengan T-rex berevolusi jadi kecil agar tidak tergigit oleh T-rex lain saat makan bersama dan merobek mangsa.	3	3	1	3	3	13	2,6
5	Peneliti mendapati, T-rex melahap mangsa secara berkelompok, seperti ditulisnya dalam Why Tyrannosaurid Forelimbs were So Short: An Integrative Hypothesis.	2	1	0	0	0	3	0,6
6	Hipotesis Evolusi Kepala.	0	0	1	0	0	1	0,2
7	Sementara itu, ahli biologi John Hutchinson dari Royal Veterinary College, University of London memperkirakan bahwa bukan tungkai depan T-rex yang berevolusi jadi pendek, melainkan kaki belakang dan kepalanya.	3	2	1	3	2	11	2,8
8	Hutchinson, yang mempelajari gerak pada hewan darat besar punah maupun yang masih hidup, berpendapat bahwa makin besar seekor hewan, makin kecil kaki depannya dan makin besar kepalanya.	2	1	0	1	0	4	0,8
9	"(T-rex khususnya) mengembangkan gigitan mematikan yang mampu menghancurkan tulang dari evolusi kepala.	2	0	2	0	0	4	0,8
10	Jadi, dinosaurus ini benar-benar berfokus mengembangkan kemampuan fisik di bagian kepala dan mengurangi bagian lengannya," kata Hutchinson, dikutip dari Live Science.	3	2	0	2	1	8	1,6
11	Menurut Hutchinson, hewan predator ganas ini berevolusi di bagian tertentu fisiknya sehingga kemampuannya terspesialisasi dalam menggigit mangsa.	2	1	0	2	1	6	1,2
12	Sementara itu, hewan yang memakan tumbuhan di darat dan bukan di pohon tinggi cenderung punya evolusi lebih merata secara fisik, tidak terspesialisasi.	1	0	0	0	0	1	0,2

Id K	Kalimat	Skor Annotator					Total	Rata rata
		A1	A2	A3	A4	A5		
13	Ia menjelaskan, lengan T-rex kemudian menjadi terlalu pendek untuk membantunya berburu dan menangkap mangsa.	3	3	1	2	3	12	2,4
14	Karena itu, dinosaurus ini menggunakan giginya sebagai pengganti lengan.	3	0	3	0	1	7	1,4
15	Praktik T-rex ini menurut Hutchinson senada dengan cara berburu komodo modern ( <i>Varanus komodoensis</i> ).	1	0	0	0	0	1	0,2
16	Bagian kaki belakang membantu hewan-hewan ini menstabilkan pergerakannya.	1	0	1	0	0	2	0,4
17	Ia menggarisbawahi, peneliti tetap belum mengetahui penuh fungsi lengan T-rex.	1	2	0	1	2	6	1,2
18	Bisa jadi, T-rex dan dinosaurus theropoda lainnya sudah menggunakan lengannya untuk hal penting sehari-hari semasa hidupnya.	3	3	0	2	2	10	2
19	Untuk mengetahuinya, butuh lebih banyak penelitian dari fosil-fosil yang terawetkan dengan baik.	2	1	0	1	1	5	1

Kolom ‘id K’ pada Tabel 4.2 merupakan *index* yang mewakili kalimat dalam dokumen. Skor yang diberikan oleh kelima annotator dijumlah, kemudian dilakukan perangkingan kalimat berdasarkan nilai rata-rata tertinggi. Tabel 4.3 menunjukkan urutan kalimat berdasarkan total skor tertinggi oleh annotator. Berdasarkan urutan kalimat pada tabel 4.3, maka akan dibentuk suatu ringkasan berdasarkan *compression rate*-nya. Pada hasil ringkasan 10%, kalimat yang diekstraksi terdiri dari 2 kalimat yaitu K[1] dan K[4], dengan K[x] merupakan kalimat pada *index* ke-x. Pada hasil ringkasan 20%, kalimat yang diekstraksi terdiri dari 4 kalimat yaitu K[1], K[4], K[7] dan K[13]. Pada hasil ringkasan 30%, kalimat yang diekstraksi terdiri dari 6 kalimat yaitu K[0], K[1], K[4], K[7], K[13] dan K[18]. Pada hasil ringkasan 40%, kalimat yang diekstraksi terdiri dari 8 kalimat yaitu K[0], K[1], K[4], K[7], K[10], K[13], K[14] dan K[18]. Pada hasil ringkasan 50%, kalimat yang diekstraksi terdiri dari 10 kalimat yaitu K[0], K[1], K[4], K[7], K[10], K[11], K[13], K[14], K[17] dan K[18].

Tabel 4.3 Perangkingan Skor Total Kalimat Artikel-1

No	Id K	Kalimat	Rata-rata
1	1	Tidak hanya T-rex, kerabatnya di kelompok theropoda atau dinosaurus berjari kaki tiga juga punya lengan dan tangan pendek dibandingkan dengan besar kepala dan badannya.	2,8
2	4	Penelitian Kevin Padian di jurnal Acta Paleontologica Polonica menunjukkan, lengan T-rex berevolusi jadi kecil agar tidak tergigit oleh T-rex lain saat makan bersama dan merobek mangsa.	2,6
3	13	Ia menjelaskan, lengan T-rex kemudian menjadi terlalu pendek untuk membantunya berburu dan menangkap mangsa.	2,4
4	7	Sementara itu, ahli biologi John Hutchinson dari Royal Veterinary College, University of London memperkirakan bahwa bukan tungkai depan T-rex yang berevolusi jadi pendek, melainkan kaki belakang dan kepalamanya.	2,4
5	18	Bisa jadi, T-rex dan dinosaurus theropoda lainnya sudah menggunakan lengannya untuk hal penting sehari-hari semasa hidupnya.	2
6	0	Tyrannosaurus rex dikenal sebagai dinosaurus karnivora ganas tetapi berlengan pendek.	1,8
7	10	Jadi, dinosaurus ini benar-benar berfokus mengembangkan kemampuan fisik di bagian kepala dan mengurangi bagian lengannya," kata Hutchinson, dikutip dari Live Science.	1,6
8	14	Karena itu, dinosaurus ini menggunakan giginya sebagai pengganti lengan.	1,4
9	11	Menurut Hutchinson, hewan predator ganas ini berevolusi di bagian tertentu fisiknya sehingga kemampuannya terspesialisasi dalam menggigit mangsa.	1,2
10	17	Ia menggarisbawahi, peneliti tetap belum mengetahui penuh fungsi lengan T-rex.	1,2
11	19	Untuk mengetahuinya, butuh lebih banyak penelitian dari fosil-fosil yang terawetkan dengan baik.	1
12	8	Hutchinson, yang mempelajari gerak pada hewan darat besar punah maupun yang masih hidup, berpendapat bahwa makin besar seekor hewan, makin kecil kaki depannya dan makin besar kepalamanya.	0,8
13	9	"(T-rex khususnya) mengembangkan gigitan mematikan yang mampu menghancurkan tulang dari evolusi kepala.	0,8
14	5	Peneliti mendapati, T-rex melahap mangsa secara berkelompok, seperti ditulisnya dalam Why Tyrannosaurid Forelimbs were So Short: An Integrative Hypothesis.	0,6
15	16	Bagian kaki belakang membantu hewan-hewan ini menstabilkan pergerakannya.	0,4
16	3	Hipotesis Evolusi Lengan.	0,2
17	6	Hipotesis Evolusi Kepala.	0,2
18	12	Sementara itu, hewan yang memakan tumbuhan di darat dan bukan di pohon tinggi cenderung punya evolusi lebih merata secara fisik, tidak terspesialisasi.	0,2
19	15	Praktik T-rex ini menurut Hutchinson senada dengan cara berburu komodo modern ( <i>Varanus komodoensis</i> ).	0,2
20	2	Kenapa, ya?	0

### 4.3 Hasil Uji Coba

Uji coba dilakukan berdasarkan langkah-langkah yang telah dijelaskan sebelumnya pada 80 artikel yang dikumpulkan. Satu per satu artikel diringkas dengan sistem yang telah dibangun sehingga terdapat 80 kali uji coba. Uji coba dimulai dari input artikel, *preprocessing*, perhitungan *similarity*, perhitungan skor *TextRank*, penentuan ringkasan dan yang terakhir yaitu melakukan evaluasi dengan *confusion matrix* dan *rouge matrix*. Proses *preprocessing* artikel menghasilkan suatu *list* kata dasar tanpa *stopword*. Dari hasil *preprocessing* ini akan dibuat matriks *similarity*. Perhitungan *similarity* dilakukan berdasarkan kombinasi antara kalimat satu dengan yang lain menggunakan metode *word overlapping*, *Jaccard similarity*, *BoW cosine similarity* dan *TFIDF cosine similarity*.

Artikel pertama memiliki 20 kalimat yang berada pada index 0 sampai 19 yang ditunjukkan pada kolom ‘id K’ pada Tabel 4.2. Terdapat empat hasil matriks *similarity* pada artikel pertama yang ditunjukkan pada Tabel 4.4 hingga Tabel 4.7 berdasarkan teknik *similarity* kalimat yang digunakan. Header tabel pada baris dan kolom merepresentasikan index kalimat pada artikel-1. Sementara *cell* tabel merepresentasikan hasil *similarity* kedua kalimat. *similarity* antara suatu kalimat dengan kalimat itu sendiri tidak dihitung sehingga diinisialisasi dengan nilai 0.

Tabel 4.4 Matrix Similarity word overlapping Artikel-1

	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	...	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>
<b>0</b>	0	1,951	0	0,725	0,962	...	0,525	0	1,144	1,784	0
<b>1</b>	1,952	0	0	0,616	1,292	...	0,932	0,542	1,507	2,598	0
<b>2</b>	0	0	0	0	0	...	0	0	0	0	0
<b>3</b>	0,725	0,616	0	0	1,21	...	0	0	0,756	0,797	0
<b>4</b>	0,962	1,292	0	1,21	0	...	0,919	0	1,979	1,535	0,562
...	...	...	...	...	...	...	...	...	...	...	...
<b>16</b>	0	0,542	0	0	0	...	0	0	0	0	0
<b>17</b>	1,144	1,507	0	0,756	1,979	...	1,084	0	0	1,848	0,691
<b>18</b>	1,784	2,598	0	0,797	1,535	...	1,125	0	1,848	0	0
<b>19</b>	0	0	0	0	0,562	...	0	0	0,691	0	0

Tabel 4.4 merupakan matriks *similarity* dengan menggunakan *word overlapping*. Hasil yang diberikan menunjukkan rentang yang cukup besar karena metode ini dapat memberikan nilai lebih dari 1 antara dua kalimat jika kedua kalimat memiliki ukuran cukup panjang dan memiliki kemiripan tinggi. Sehingga pada metode ini, ukuran suatu kalimat sangat berpengaruh terhadap hasil *similarity*.

Tabel 4.5 Matrix *Jaccard Similarity* Artikel-1

	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	...	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>
<b>0</b>	0	0,222	0	0,1	0,095	...	0,059	0	0,154	0,273	0
<b>1</b>	0,222	0	0	0,063	0,115	...	0,091	0,056	0,167	0,333	0
<b>2</b>	0	0	0	0	0	...	0	0	0	0	0
<b>3</b>	0,1	0,063	0	0	0,125	...	0	0	0,111	0,125	0
<b>4</b>	0,095	0,115	0	0,125	0	...	0,087	0	0,222	0,167	0,056
...	...	...	...	...	...	...	...	...	...	...	...
<b>15</b>	0,059	0,091	0	0	0,087	...	0	0	0,133	0,143	0
<b>16</b>	0	0,056	0	0	0	...	0	0	0	0	0
<b>17</b>	0,154	0,167	0	0,111	0,222	...	0,133	0	0	0,3	0,1
<b>18</b>	0,273	0,333	0	0,125	0,167	...	0,143	0	0,3	0	0
<b>19</b>	0	0	0	0	0,056	...	0	0	0,1	0	0

Tabel 4.5 merupakan matriks *similarity* dengan menggunakan *Jaccard similarity*. Hasil yang diberikan menunjukkan rentang antara 0 sampai 1. Hasil perhitungan *similarity* antar kalimat memberikan nilai yang cukup rendah, lebih rendah dibanding menggunakan metode *word overlapping*. Hal ini dikarenakan metode *Jaccard similarity* menggunakan irisan dan gabungan antara kedua himpunan kalimat.

Tabel 4.6 Matrix BoW cosine Similarity Artikel-1

	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	...	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>
<b>0</b>	0	0,378	0	0,204	0,231	...	0,112	0	0,267	0,433	0
<b>1</b>	0,378	0	0	0,154	0,292	...	0,169	0,094	0,303	0,546	0
<b>2</b>	0	0	0	0	0	...	0	0	0	0	0
<b>3</b>	0,204	0,154	0	0	0,252	...	0	0	0,218	0,236	0
<b>4</b>	0,231	0,292	0	0,252	0	...	0,276	0	0,495	0,445	0,082
...	...	...	...	...	...	...	...	...	...	...	...
<b>15</b>	0,112	0,169	0	0	0,276	...	0	0	0,239	0,258	0
<b>16</b>	0	0,094	0	0	0	...	0	0	0	0	0
<b>17</b>	0,267	0,303	0	0,218	0,495	...	0,239	0	0	0,463	0,143
<b>18</b>	0,433	0,546	0	0,236	0,445	...	0,258	0	0,463	0	0
<b>19</b>	0	0	0	0	0,082	...	0	0	0,143	0	0

Tabel 4.6 merupakan matriks *similarity* dengan menggunakan *bag of word cosine similarity*. Hasil yang diberikan menunjukkan rentang antara 0 sampai 1. Hasil perhitungan *similarity* antar kalimat memberikan nilai yang cukup rendah, lebih rendah dibandingkan dengan *word overlapping* namun lebih tinggi dibandingkan *Jaccard similarity*. *Cosine similarity* memanfaatkan vektor *bag of word* dalam perhitungan similaritas antar kalimat.

Tabel 4.7 Matrix TFIDF *cosine Similarity* Artikel-1

	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	...	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>
<b>0</b>	0	0,182	0	0,086	0,077	...	0,03	0	0,087	0,221	0
<b>1</b>	0,182	0	0	0,065	0,1	...	0,048	0,067	0,1	0,368	0
<b>2</b>	0	0	0	0	0	...	0	0	0	0	0
<b>3</b>	0,086	0,065	0	0	0,131	...	0	0	0,094	0,124	0
<b>4</b>	0,077	0,1	0	0,131	0	...	0,084	0	0,209	0,191	0,051
...	...	...	...	...	...	...	...	...	...	...	...
<b>15</b>	0,03	0,048	0	0	0,084	...	0	0	0,07	0,092	0
<b>16</b>	0	0,067	0	0	0	...	0	0	0	0	0
<b>17</b>	0,087	0,1	0	0,094	0,209	...	0,07	0	0	0,192	0,084
<b>18</b>	0,221	0,368	0	0,124	0,191	...	0,092	0	0,192	0	0
<b>19</b>	0	0	0	0	0,051	...	0	0	0,084	0	0

Tabel 4.7 merupakan matriks *similarity* dengan menggunakan TFIDF *cosine similarity*. Hasil yang diberikan menunjukkan rentang antara 0 sampai 1. Hasil perhitungan *similarity* antar kalimat memberikan nilai yang paling rendah dari ketiga metode lainnya. Dibandingkan dengan menggunakan vektor BoW, perhitungan *cosine similarity* dengan vektor TFIDF, memberikan nilai yang lebih rendah karena TFIDF mengurangi bobot kata yang sering muncul dan meningkatkan bobot kata yang jarang muncul.

Setelah dibuat matriks *similarity*, maka selanjutnya dapat dihitung skor tiap kalimat menggunakan rumus iterasi *TextRank*. Tabel 4.8 merupakan hasil dari pemberian skor kalimat berdasarkan id kalimat pada tabel 4.2 dan juga matriks *similarity* yang telah dibuat sebelumnya.

Tabel 4.8 Hasil Skor *Ground Truth* dan Skor Final *TextRank*

id K	Skor <i>Ground Truth</i>	Skor Final <i>TextRank</i>			
		WordOverlap	Jaccard	Bow Cosine	TFIDF Cosine
0	1,8	1.067296	1.126059	1.079152	0.994881
1	2,4	1.540227	1.393062	1.353433	1.291049
2	0	0.15	0.15	0.15	0.15
3	0,2	0.927122	1.143393	1.115126	1.309474
4	2,6	1.494194	1.291939	1.472454	1.331763
5	0,6	0.915003	0.759716	0.780477	0.646584
6	0,2	0.749075	0.895669	0.873547	1.150202
7	2,8	1.365241	1.101612	1.13322	0.963185
8	0,8	0.78556	0.758697	0.822926	0.990055
9	0,8	1.233676	1.182947	1.174611	1.03345
10	1,6	1.02428	0.990609	0.974465	0.992557
11	1,2	0.981722	0.982779	0.986316	1.266411
12	0,2	0.698365	0.712677	0.733889	0.855377
13	2,4	1.350661	1.424645	1.337315	1.325326
14	1,4	0.68518	0.800038	0.7808	0.701507
15	0,2	0.884966	0.837182	0.840596	0.667188
16	0,4	0.44144	0.456098	0.570048	0.725126
17	1,2	1.23079	1.331531	1.273614	1.015948
18	2	1.351906	1.538773	1.448259	1.4717
19	1	0.273296	0.272574	0.249753	0.268219

Pada Tabel 4.8, skor *ground truth* merupakan total skor yang diperoleh dari annotator. Sedangkan skor final *TextRank* merupakan hasil perhitungan menggunakan iterasi *TextRank* pada masing-masing metode *similarity* yang digunakan. Berdasarkan skor final yang terdapat pada Tabel 4.8 maka dapat dibentuk suatu ringkasan yang diambil dengan mengurutkan skor final kemudian mengambil  $n$  kalimat dengan skor tertinggi. Tabel 4.9 menunjukkan hasil pengambilan kalimat ringkasan pada artikel pertama. Pada Tabel 4.9,  $K[x]$  merupakan kalimat pada *index* ke- $x$ , dan  $x$  merupakan id kalimat yang dapat dilihat pada kolom “id K” di Tabel 4.8.  $CR$  merupakan *compression rate*. Tabel 4.9 menunjukkan pengambilan ringkasan berdasarkan *compression rate* dengan metode *similarity* yang digunakan dalam *TextRank* berdasarkan ranking paling tinggi.

Tabel 4.9 Hasil Ringkasan Manual dan Prediksi pada Artikel-1

CR	Kalimat pada index-x, K[x]				
	Ground Truth	Word Overlap	Jaccard	Bow Cosine	TFIDF Cosine
50%	K[0], K[1], K[4], K[7], K[10], K[11], K[13], K[14], K[17], K[18]	K[0], K[1], K[4], K[7], K[9], K[10], K[11], K[13], K[17], K[18]	K[0], K[1], K[3], K[4], K[7], K[9], K[10], K[13], K[17], K[18]	K[0], K[1], K[3], K[4], K[7], K[9], K[11], K[13], K[17], K[18]	K[0], K[1], K[3], K[4], K[6], K[9], K[11], K[13], K[17], K[18]
40%	K[0], K[1], K[4], K[7], K[10], K[13], K[14], K[18]	K[0], K[1], K[4], K[7], K[9], K[13], K[17], K[18]	K[0], K[1], K[3], K[4], K[9], K[13], K[17], K[18]	K[1], K[3], K[4], K[7], K[9], K[13], K[17], K[18]	K[1], K[3], K[4], K[6], K[9], K[11], K[13], K[18]
30%	K[0], K[1], K[4], K[7], K[13], K[18]	K[1], K[4], K[7], K[9], K[13], K[18]	K[1], K[4], K[9], K[13], K[17], K[18]	K[1], K[4], K[9], K[13], K[17], K[18]	K[1], K[3], K[4], K[11], K[13], K[18]
20%	K[1], K[4], K[7], K[13]	K[1], K[4], K[7], K[18]	K[1], K[13], K[17], K[18]	K[1], K[4], K[13], K[18]	K[3], K[4], K[13], K[18]
10%	K[4], K[7]	K[1], K[4]	K[13], K[18]	K[4], K[18]	K[4], K[18]

Berdasarkan hasil ringkasan yang terdapat pada Tabel 4.9 maka dapat dihitung nilai kinerja sistem dengan *confusion matrix* dan juga ROUGE-1. Tabel 4.10 merupakan hasil perhitungan nilai *precision*, *recall*, *accuracy* dan *f-score* menggunakan *confusion matrix*.

Tabel 4.10 Hasil *Confusion Matrix* pada Artikel-1

Metode	CR	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
Word Overlap	10%	20	2	1	17	1	1	0,5	0,5	0,9	0,5
	20%	20	4	3	15	1	1	<b>0,75</b>	<b>0,75</b>	<b>0,9</b>	<b>0,75</b>
	30%	20	6	5	13	1	1	<b>0,833</b>	<b>0,833</b>	<b>0,9</b>	<b>0,833</b>
	40%	20	8	6	10	2	2	<b>0,75</b>	<b>0,75</b>	<b>0,8</b>	<b>0,75</b>
	50%	20	10	9	9	1	1	<b>0,9</b>	<b>0,9</b>	<b>0,9</b>	<b>0,9</b>
Jaccard similarity	10%	20	2	0	16	2	2	0	0	0,8	0
	20%	20	4	2	14	2	2	0,5	0,5	0,8	0,5
	30%	20	6	4	12	2	2	0,667	0,667	0,8	0,667
	40%	20	8	5	9	3	3	0,625	0,625	0,7	0,625
	50%	20	10	8	8	2	2	0,8	0,8	0,8	0,8
BoW cosine similarity	10%	20	2	1	17	1	1	0,5	0,5	0,9	0,5
	20%	20	4	3	15	1	1	<b>0,75</b>	<b>0,75</b>	<b>0,9</b>	<b>0,75</b>
	30%	20	6	4	12	2	2	0,667	0,667	0,8	0,667
	40%	20	8	5	9	3	3	0,625	0,625	0,7	0,625
	50%	20	10	8	8	2	2	0,8	0,8	0,8	0,8
TFIDF cosine similarity	10%	20	2	1	17	1	1	0,5	0,5	0,9	0,5
	20%	20	4	2	14	2	2	0,5	0,5	0,8	0,5
	30%	20	6	4	12	2	2	0,667	0,667	0,8	0,667
	40%	20	8	4	8	4	4	0,5	0,5	0,6	0,5
	50%	20	10	7	7	3	3	0,7	0,7	0,7	0,7

Pada Tabel 4.10, CR merupakan *compression rate*, A merupakan banyaknya kalimat pada dokumen asli, R merupakan banyaknya kalimat pada ringkasan referensi atau ringkasan prediksi. Hasil uji coba artikel pertama menunjukkan pada kategori *compression rate* 50%, nilai *precision*, *recall*, dan *accuracy* tertinggi dicapai oleh metode *TextRank* dengan *word overlapping* yaitu sebesar 0,9 pada semua parameter penilaian. Pada kategori *compression rate* 40%, kinerja terbaik dicapai oleh metode *word overlapping* yaitu dengan nilai *f-score* 0,75 dan *accuracy* sebesar 0,8. Pada kategori *compression rate* 30%, kinerja terbaik dicapai oleh metode *word overlapping* yaitu dengan nilai *f-score* 0,833 dan *accuracy* sebesar 0,9. Pada kategori *compression rate* 20%, kinerja terbaik dicapai oleh metode *word overlapping* yaitu dengan nilai *f-score* 0,75 dan *accuracy* sebesar 0,9. Pada kategori *compression rate* 10%, kinerja terbaik dicapai oleh metode tiga metode yaitu *word overlapping*, BoW dan TFIDF *cosine similarity*, yaitu dengan nilai *f-score* 0,5 dan *accuracy* sebesar 0,9. Nilai *precision*, *recall* dan *f-score* menunjukkan hasil yang sama pada perhitungan dengan *confusion matrix*, karena FP dan FN memiliki nilai yang selalu sama karena kalimat pada *ground truth* sama dengan kalimat pada ringkasan prediksi.

Tabel 4.11 merupakan hasil perhitungan nilai *precision*, *recall* dan *f-score* menggunakan *rouge matrix*. Pada Tabel 4.10, CR merupakan *compression rate*, A merupakan banyaknya kalimat pada dokumen artikel, R merupakan banyaknya kalimat pada ringkasan referensi atau ringkasan prediksi. GT merupakan banyaknya kata yang berada pada ringkasan *ground truth*, RP merupakan banyaknya kata yang

berada pada ringkasan prediksi, dan  $O$  merupakan banyaknya kata yang saling tumpang tindih antara  $GT$  dan  $RP$ .

Tabel 4.11 Hasil ROUGE-1 pada Artikel-1

Metode	CR	A	GT	RP	O	Precision	Recall	F-Score
Word Overlap	10%	331	59	55	35	0,63636	0,59322	0,61404
	20%	331	101	104	90	<b>0,86538</b>	<b>0,89109</b>	<b>0,87805</b>
	30%	331	131	135	122	<b>0,9037</b>	<b>0,9313</b>	<b>0,91729</b>
	40%	331	164	158	137	<b>0,86709</b>	<b>0,83537</b>	<b>0,85093</b>
	50%	331	194	199	185	<b>0,92965</b>	<b>0,95361</b>	<b>0,94148</b>
Jaccard similarity	10%	331	59	35	12	0,34286	0,20339	0,25532
	20%	331	101	73	51	0,69863	0,50495	0,58621
	30%	331	131	117	97	0,82906	0,74046	0,78226
	40%	331	164	132	111	0,84091	0,67683	0,75
	50%	331	194	185	169	0,91351	0,87113	0,89182
BoW cosine similarity	10%	331	59	48	34	<b>0,70833</b>	<b>0,57627</b>	<b>0,63551</b>
	20%	331	101	90	76	0,84444	0,75248	0,79581
	30%	331	131	117	97	0,82906	0,74046	0,78226
	40%	331	164	151	128	0,84768	0,78049	0,8127
	50%	331	194	180	167	0,92778	0,86082	0,89305
TFIDF cosine similarity	10%	331	59	48	34	<b>0,70833</b>	<b>0,57627</b>	<b>0,63551</b>
	20%	331	101	68	54	0,79412	0,53465	0,63905
	30%	331	131	112	95	0,84821	0,72519	0,78189
	40%	331	164	131	107	0,81679	0,65244	0,72542
	50%	331	194	154	141	0,91558	0,7268	0,81034

Hasil uji coba artikel pertama pada kategori *compression rate* 50%, nilai *precision*, *recall* dan *f-score* terbaik dicapai oleh metode *word overlapping* secara berturut-turut sebesar 0,9296; 0,9536 dan 0,9415. Pada kategori *compression rate* 40%, nilai *precision*, *recall* dan *f-score* terbaik dicapai oleh metode *word overlapping* secara berturut-turut sebesar 0,86709; 0,83537 dan 0,85093. Pada kategori *compression rate* 30%, nilai *precision*, *recall* dan *f-score* terbaik dicapai oleh metode *word overlapping* secara berturut-turut sebesar 0,9037; 0,9313 dan 0,9173. Pada kategori *compression rate* 20%, nilai *precision*, *recall* dan *f-score* terbaik dicapai oleh metode *word overlapping* secara berturut-turut sebesar 0,86538; 0,89109 dan 0,87805. Pada kategori *compression rate* 10%, nilai

*precision, recall* dan *f-score* terbaik dicapai oleh metode BoW dan TFIDF *cosine similarity* secara berturut-turut sebesar 0,70833; 0,57627 dan 0,63551.

#### 4.4 Pembahasan

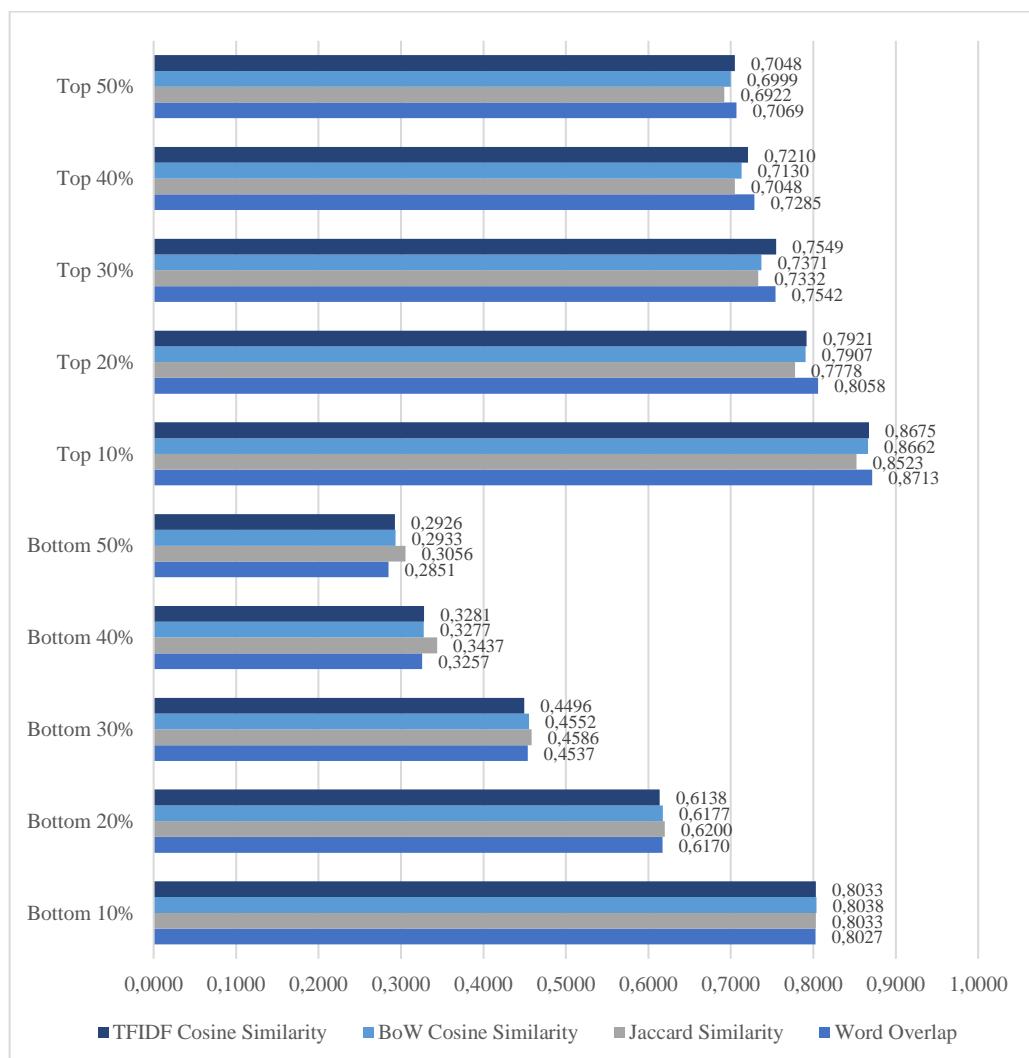
Berdasarkan skenario pengujian yang telah dilakukan, maka didapatkan hasil uji coba dengan *confusion matrix* pada tabel 4.12 dan juga uji coba dengan ROUGE-1 pada tabel 4.13. Hasil uji coba terdiri dari kombinasi empat metode *similarity* dengan lima kategori *compression rate*, serta terdapat pengujian terhadap pengambilan ringkasan berdasarkan skor *TextRank* tertinggi (*top-n*) dan skor terendah (*bottom-n*).

Tabel 4.12 Rata-rata Hasil *Accuracy* dan *F-Score* (*Confusion Matrix*)

<b>Metode</b>	<b>CR</b>	<b>Top-n</b>		<b>Bottom-n</b>	
		<b>Accuracy</b>	<b>F-Score</b>	<b>Accuracy</b>	<b>F-Score</b>
Word Overlap	10%	<b>0,871334</b>	<b>0,355551</b>	0,8027	0,012292
	20%	<b>0,805846</b>	<b>0,512157</b>	0,617005	0,040045
	30%	0,754198	0,58869	0,453711	0,085054
	40%	<b>0,728508</b>	<b>0,660242</b>	0,325724	0,158133
	50%	<b>0,706933</b>	<b>0,705678</b>	0,285075	0,282235
Jaccard similarity	10%	0,852291	0,259836	0,803309	0,015417
	20%	<b>0,777755</b>	0,44241	0,619992	0,04808
	30%	0,733179	0,553419	0,458577	0,093864
	40%	0,704837	0,631102	0,34373	0,180293
	50%	0,692217	0,691002	0,305582	0,302793
BoW cosine similarity	10%	0,866234	0,325744	0,803841	0,017917
	20%	<b>0,790721</b>	0,475703	0,617652	0,041295
	30%	0,73708	0,559108	0,455237	0,087901
	40%	0,713001	0,641023	0,327684	0,159932
	50%	0,69991	0,698309	0,293323	0,290973
TFIDF cosine similarity	10%	0,867544	0,332009	0,803309	0,015417
	20%	0,792058	0,478397	0,613807	0,031662
	30%	<b>0,754914</b>	<b>0,589666</b>	0,44964	0,078116
	40%	0,721012	0,651339	0,328105	0,160306
	50%	0,704827	0,703206	0,292591	0,290095

Tabel 4.12 berisi seluruh nilai rata-rata hasil uji coba sistem pada parameter *accuracy* dan *f-score* menggunakan *confusion matrix*. Terdapat empat metode *similarity* yang digunakan yaitu *word overlap*, *Jaccard similarity*, *BoW cosine*

*similarity* dan TFIDF *cosine similarity*. Masing-masing metode dicari ringkasan sistemnya berdasarkan lima *compression rate* yang ditentukan yaitu 10%, 20%, 30%, 40%, dan 50%. Kemudian dilakukan perbandingan antara pengambilan ringkasan berdasarkan skor *TextRank* tertinggi (*top-n ranking*) dan berdasarkan skor *TextRank* terendah (*bottom-n ranking*). Sehingga terdapat 40 hasil ringkasan berdasarkan kombinasi ketiga kategori yang telah disebutkan. Gambar 4.1 dan Gambar 4.2 merupakan diagram visualisasi dari perhitungan rata-rata nilai *accuracy* dan *f-score* pada *confusion matrix*.



Gambar 4.1 Diagram Rata-rata Hasil Accuracy (*Confusion Matrix*)

Diagram pada Gambar 4.1 menunjukkan rata-rata hasil *accuracy* terbaik pada kategori *top-n* ringkasan (ringkasan diambil berdasarkan skor *TextRank* tertinggi). Nilai akurasi yang dihasilkan mengalami kenaikan yang cukup konsisten ketika nilai *compression rate*-nya semakin menurun (dari 50% ke 10%). Nilai akurasi menunjukkan hasil yang lebih baik ketika nilai *compression rate* rendah ini dikarenakan saat *compression rate* rendah, maka kalimat yang dipertahankan menjadi ringkasan akan semakin sedikit dan kalimat yang dibuang semakin banyak. Kalimat yang dibuang ini sebagian besar merupakan TN (True Negatif) yang memiliki pengaruh besar dalam perhitungan akurasi. Semakin besar nilai TN dan TP, maka semakin besar pula nilai akurasi yang dihasilkan.

Pada ringkasan dengan *compression rate* 50%, rata-rata hasil akurasi dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,706, TFIDF *cosine similarity* sebesar 0,704, BoW *cosine similarity* sebesar 0,699 dan *Jaccard similarity* sebesar 0,692. Nilai akurasi dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Akurasi dari keempat metode menunjukkan perbedaan nilai yang relatif kecil, dibuktikan dengan kecilnya jarak antar nilai akurasi tiap metode dengan rata-rata akurasi keempat metode. Rata-rata dari keempat metode menunjukkan nilai 0,70 sedangkan nilai standar deviasinya sebesar 0,006. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Sehingga penggunaan metode yang berbeda tidak terlalu berpengaruh terhadap hasil akurasi pada *compression rate* 50%. Namun berdasarkan data artikel yang digunakan, metode yang memberikan

nilai akurasi terbaik pada *compression rate* 50% adalah metode *TextRank* dengan *similarity word overlapping*.

Pada ringkasan dengan *compression rate* 40%, rata-rata hasil akurasi dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,728, TFIDF *cosine similarity* sebesar 0,721, BoW *cosine similarity* sebesar 0,713 dan *Jaccard similarity* sebesar 0,704. Nilai akurasi dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata dari keempat metode menunjukkan nilai 0,716 sedangkan nilai standar deviasinya sebesar 0,01. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai akurasi terbaik pada *compression rate* 40% adalah metode *TextRank* dengan *similarity word overlapping*.

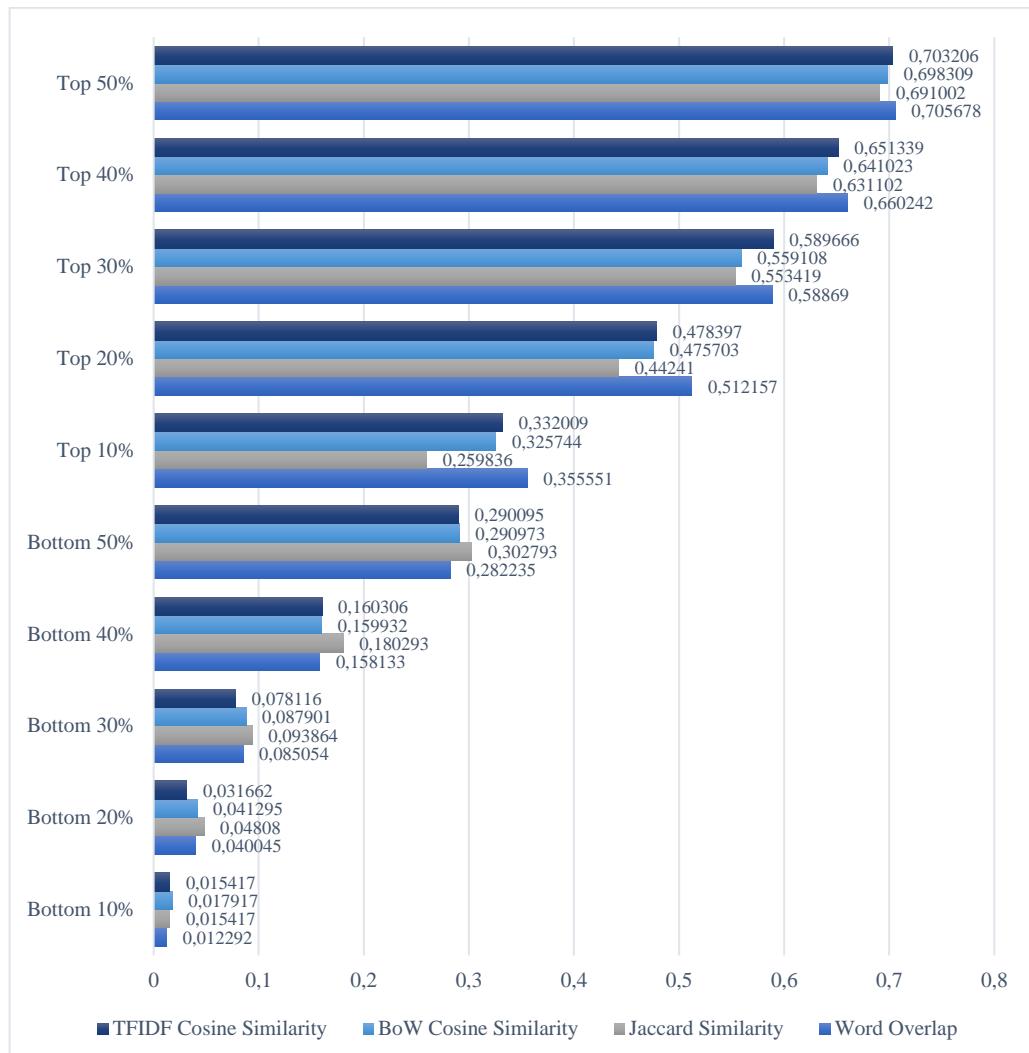
Pada ringkasan dengan *compression rate* 30%, rata-rata hasil akurasi dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* dan TFIDF *cosine similarity* yang memiliki nilai sama sebesar 0,754, BoW *cosine similarity* sebesar 0,737 dan *Jaccard similarity* sebesar 0,733. Nilai akurasi dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata dari keempat metode menunjukkan nilai 0,744 sedangkan nilai standar deviasinya sebesar 0,011. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang

memberikan nilai akurasi terbaik pada *compression rate* 30% adalah metode *TextRank* dengan *similarity word overlapping* dan *TFIDF cosine similarity*.

Pada ringkasan dengan *compression rate* 20%, rata-rata hasil akurasi dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,805, *TFIDF cosine similarity* sebesar 0,792, *BoW cosine similarity* sebesar 0,790 dan *Jaccard similarity* sebesar 0,777. Nilai akurasi dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata dari keempat metode menunjukkan nilai 0,791 sedangkan nilai standar deviasinya sebesar 0,01. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai akurasi terbaik pada *compression rate* 20% adalah metode *TextRank* dengan *similarity word overlapping*.

Pada ringkasan dengan *compression rate* 10%, rata-rata hasil akurasi dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,871, *TFIDF cosine similarity* sebesar 0,867, *BoW cosine similarity* sebesar 0,866 dan *Jaccard similarity* sebesar 0,852. Nilai akurasi dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata dari keempat metode menunjukkan nilai 0,864 sedangkan nilai standar deviasinya sebesar 0,008. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai

akurasi terbaik pada *compression rate* 10% adalah metode *TextRank* dengan *similarity word overlapping*.



Gambar 4.2 Diagram Rata-rata Hasil (*Confusion Matrix*)

Diagram pada Gambar 4.2 menunjukkan rata-rata hasil terbaik pada kategori *top-n* ringkasan (ringkasan diambil berdasarkan skor *TextRank* tertinggi). Nilai *f-score* yang dihasilkan mengalami penurunan ketika *compression rate*-nya menurun (dari 50% ke 10%). Ketika nilai *compression rate* rendah, maka kalimat yang dipertahankan menjadi ringkasan semakin sedikit, sehingga peluang satu

kalimat pada ringkasan prediksi menjadi TP juga semakin kecil. Nilai TP yang kecil ini mempengaruhi perhitungan *precision* dan *recall*, yang juga sangat berpengaruh terhadap nilai *f-score*. Sedangkan ketika nilai *compression rate* tinggi, maka kalimat yang dipertahankan menjadi ringkasan semakin banyak, sehingga peluang satu kalimat pada ringkasan menjadi TP juga semakin tinggi. Nilai TP yang tinggi membuat hasil perhitungan *f-score* menjadi lebih tinggi.

Pada ringkasan dengan *compression rate* 50%, rata-rata hasil *f-score* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,705, TFIDF *cosine similarity* sebesar 0,703, BoW *cosine similarity* sebesar 0,698 dan *Jaccard similarity* sebesar 0,691. Nilai *f-score* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Nilai *f-score* dari keempat metode menunjukkan perbedaan nilai yang relatif kecil, dibuktikan dengan kecilnya jarak antar nilai *f-score* tiap metode dengan nilai rata-rata *f-score* dari keempat metode. Rata-rata *f-score* dari keempat metode menunjukkan nilai 0,699 sedangkan nilai standar deviasinya sebesar 0,006. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Sehingga penggunaan metode yang berbeda tidak terlalu berpengaruh terhadap hasil *f-score* pada *compression rate* 50%. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *f-score* terbaik pada *compression rate* 50% adalah metode *TextRank* dengan *similarity word overlapping*.

Pada ringkasan dengan *compression rate* 40%, rata-rata hasil *f-score* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank*

dengan *similarity word overlapping* sebesar 0,66, TFIDF *cosine similarity* sebesar 0,651, BoW *cosine similarity* sebesar 0,641 dan *Jaccard similarity* sebesar 0,631. Nilai *f-score* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata *f-score* dari keempat metode menunjukkan nilai 0,645 sedangkan nilai standar deviasinya sebesar 0,01. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *f-score* terbaik pada *compression rate* 40% adalah metode *TextRank* dengan *similarity word overlapping*.

Pada ringkasan dengan *compression rate* 30%, rata-rata hasil *f-score* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan TFIDF *cosine similarity* sebesar 0,689, *word overlapping* sebesar 0,688, BoW *cosine similarity* sebesar 0,559 dan *Jaccard similarity* sebesar 0,553. Nilai *f-score* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata *f-score* dari keempat metode menunjukkan nilai 0,572 sedangkan nilai standar deviasinya sebesar 0,01. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *f-score* terbaik pada *compression rate* 30% adalah metode *TextRank* dengan TFIDF *cosine similarity*.

Pada ringkasan dengan *compression rate* 20%, rata-rata hasil *f-score* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,512, TFIDF *cosine similarity* sebesar

0,478, BoW *cosine similarity* sebesar 0,475 dan *Jaccard similarity* sebesar 0,442. Nilai *f-score* dari keempat metode menunjukkan hasil yang cukup dalam melakukan peringkasan teks. Rata-rata *f-score* dari keempat metode menunjukkan nilai 0,477 sedangkan nilai standar deviasinya sebesar 0,02. Nilai standar deviasi yang cukup kecil ini menunjukkan perbedaan signifikan yang sangat kecil diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *f-score* terbaik pada *compression rate* 20% adalah metode *TextRank* dengan *similarity word overlapping*.

Pada ringkasan dengan *compression rate* 10%, rata-rata hasil *f-score* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,355, TFIDF *cosine similarity* sebesar 0,332, BoW *cosine similarity* sebesar 0,325 dan *Jaccard similarity* sebesar 0,259. Nilai *f-score* dari keempat metode menunjukkan hasil yang tidak cukup baik dalam melakukan peringkasan teks. Rata-rata *f-score* dari keempat metode menunjukkan nilai 0,318 sedangkan nilai standar deviasinya sebesar 0,04. Nilai standar deviasi yang cukup kecil ini menunjukkan perbedaan signifikan yang sangat kecil diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *f-score* terbaik pada *compression rate* 10% adalah metode *TextRank* dengan *similarity word overlapping*.

Tabel 4.13 berisi seluruh nilai rata-rata hasil uji coba sistem pada parameter *precision*, *recall* dan *f-score* menggunakan ROUGE *matrix*. Terdapat empat metode *similarity* yang digunakan, masing-masing metode dicari ringkasan sistemnya berdasarkan lima *compression rate* yang ditentukan yaitu 10%, 20%, 30%, 40%,

dan 50%. Kemudian dilakukan perbandingan antara pengambilan ringkasan berdasarkan skor *TextRank* tertinggi (*top-n ranking*) dan berdasarkan skor *TextRank* terendah (*bottom-n ranking*).

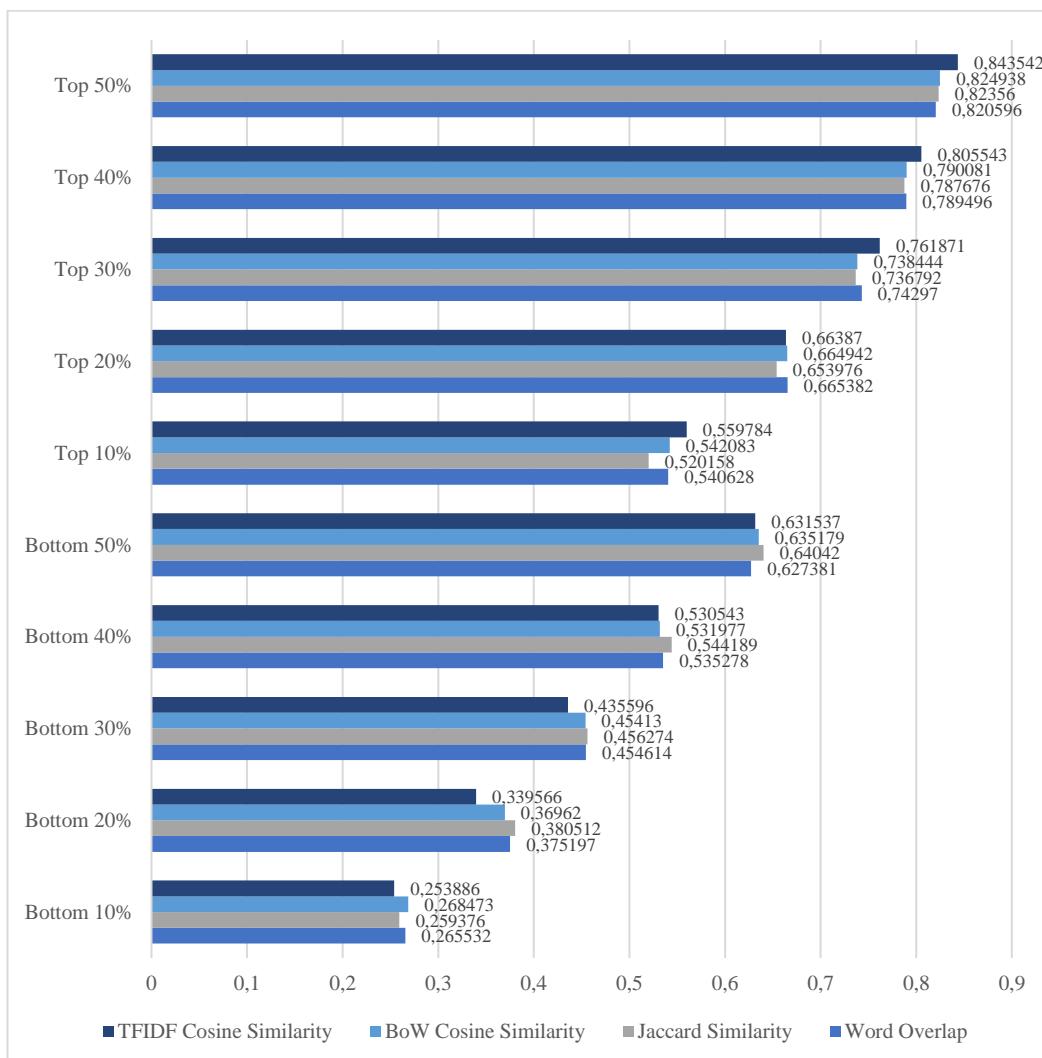
Tabel 4.13 Rata-rata Hasil *Precision*, *Recall* dan *F-Score* (ROUGE-1)

<b>Metode</b>	<b>CR</b>	<b>Top-n</b>			<b>Bottom-n</b>		
		<b>Precision</b>	<b>Recall</b>	<b>F-Score</b>	<b>Precision</b>	<b>Recall</b>	<b>F-Score</b>
Word Overlap	10%	0,540628	<b>0,546316</b>	<b>0,535732</b>	0,265532	0,168335	0,196058
	20%	<b>0,665382</b>	<b>0,657805</b>	<b>0,656862</b>	0,375197	0,2438	0,288559
	30%	0,74297	<b>0,70651</b>	<b>0,721871</b>	0,454614	0,309057	0,36185
	40%	0,789496	<b>0,759321</b>	<b>0,772427</b>	0,535278	0,389974	0,445803
	50%	0,820596	<b>0,793333</b>	<b>0,805585</b>	0,627381	0,482587	0,541672
Jaccard similarity	10%	0,520158	0,440405	0,466398	0,259376	0,174236	0,197703
	20%	0,653976	0,563944	0,600196	0,380512	0,266937	0,305617
	30%	0,736792	0,650752	0,687901	0,456274	0,334082	0,379441
	40%	0,787676	0,707797	0,743319	0,544189	0,427143	0,473141
	50%	0,82356	0,754345	0,785813	0,64042	0,525153	0,573325
BoW cosine similarity	10%	0,542083	0,504017	0,513604	0,268473	0,176187	0,201587
	20%	0,664942	0,611371	0,631611	0,36962	0,251545	0,293105
	30%	0,738444	0,671159	0,70025	0,45413	0,326858	0,374097
	40%	0,790081	0,732264	0,758052	0,531977	0,405875	0,455271
	50%	0,824938	0,773695	0,797181	0,635179	0,50675	0,56004
TFIDF cosine similarity	10%	<b>0,559784</b>	0,508584	0,522804	0,253886	0,177298	0,198741
	20%	0,66387	0,610925	0,630851	0,339566	0,242309	0,277021
	30%	<b>0,761871</b>	0,684783	0,71799	0,435596	0,3251	0,367127
	40%	<b>0,805543</b>	0,735032	0,766482	0,530543	0,416767	0,462898
	50%	<b>0,843542</b>	0,769151	0,80288	0,631537	0,52082	0,567464

Rata-rata *precision* menunjukkan nilai yang lebih tinggi dibandingkan dengan rata-rata *recall*. Hal ini disebabkan karena sebagian besar ringkasan referensi pada data artikel memiliki jumlah kata yang lebih banyak dibandingkan dengan jumlah kata pada ringkasan prediksinya. Gambar 4.3, 4.4 dan Gambar 4.5 merupakan diagram visualisasi dari perhitungan rata-rata nilai *precision*, *recall* dan *f-score* menggunakan ROUGE-1.

Diagram pada Gambar 4.3 menunjukkan rata-rata hasil *precision* terbaik pada kategori *top-n* ringkasan (ringkasan diambil berdasarkan skor *TextRank* tertinggi). Namun pada kategori *bottom* dengan *compression rate* 50%, nilai

*precision* menunjukkan nilai yang lebih tinggi dibandingkan dengan nilai *precision* pada kategori *top* 50%. Hal ini dapat terjadi karena nilai TP pada *bottom* 50% memiliki jumlah yang lebih besar dibandingkan dengan nilai TP pada *top* 10%, hal ini juga dipengaruhi oleh banyaknya kalimat yang diambil menjadi ringkasan pada *compression rate* 10% tidak sebanyak ringkasan pada *compression rate* 50%.



Gambar 4.3 Diagram Rata-rata Hasil *Precision* (ROUGE-1)

Nilai *precision* yang dihasilkan mengalami penurunan ketika *compression rate*-nya menurun (dari 50% ke 10%). Besar kecilnya nilai *precision* dipengaruhi

oleh banyaknya kata pada ringkasan prediksi dan banyaknya kata yang tumpang tindih dengan ringkasan referensi. Saat *compression rate* rendah, maka kalimat yang dipertahankan menjadi ringkasan semakin sedikit. Hal ini menyebabkan semakin kecilnya peluang suatu kata pada ringkasan prediksi saling tumpang tindih dengan kata yang terdapat pada ringkasan referensi. Nilai tumpang tindih kata yang kecil ini mempengaruhi perhitungan. Sedangkan ketika nilai *compression rate* tinggi, maka kalimat yang dipertahankan menjadi ringkasan semakin banyak. Sehingga peluang suatu kata pada ringkasan prediksi saling tumpang tindih dengan kata yang terdapat pada ringkasan referensi semakin besar. Besar kecilnya nilai tumpang tindih antar kata (*overlaps*) sangat mempengaruhi perhitungan *precision*, *recall* dan *f-score*. Nilai *precision* yang tinggi, berarti ada banyak kata pada ringkasan referensi yang terdapat pada ringkasan prediksi.

Pada ringkasan dengan *compression rate* 50%, rata-rata hasil *precision* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan TFIDF *cosine similarity* sebesar 0,843, BoW *cosine similarity* sebesar 0,824, *Jaccard similarity* sebesar 0,824, dan *word overlapping* sebesar 0,820. Nilai *precision* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Nilai *precision* dari keempat metode menunjukkan perbedaan nilai yang relatif kecil, dibuktikan dengan kecilnya jarak antar nilai *precision* tiap metode dengan nilai rata-rata *precision* dari keempat metode. Rata-rata *precision* dari keempat metode menunjukkan nilai 0,828, sedangkan nilai standar deviasinya sebesar 0,01. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Sehingga penggunaan

metode yang berbeda tidak terlalu berpengaruh terhadap hasil *precision* pada *compression rate* 50%. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *precision* terbaik pada *compression rate* 50% adalah metode *TextRank* dengan TFIDF *cosine similarity*.

Pada ringkasan dengan *compression rate* 40%, rata-rata hasil *precision* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan TFIDF *cosine similarity* sebesar 0,805, BoW *cosine similarity* sebesar 0,79, *similarity word overlapping* sebesar 0,789 dan *Jaccard similarity* sebesar 0,787. Nilai *precision* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata *precision* dari keempat metode menunjukkan nilai 0,793 sedangkan nilai standar deviasinya sebesar 0,008. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *f-score* terbaik pada *compression rate* 40% adalah metode *TextRank* dengan TFIDF *cosine similarity*.

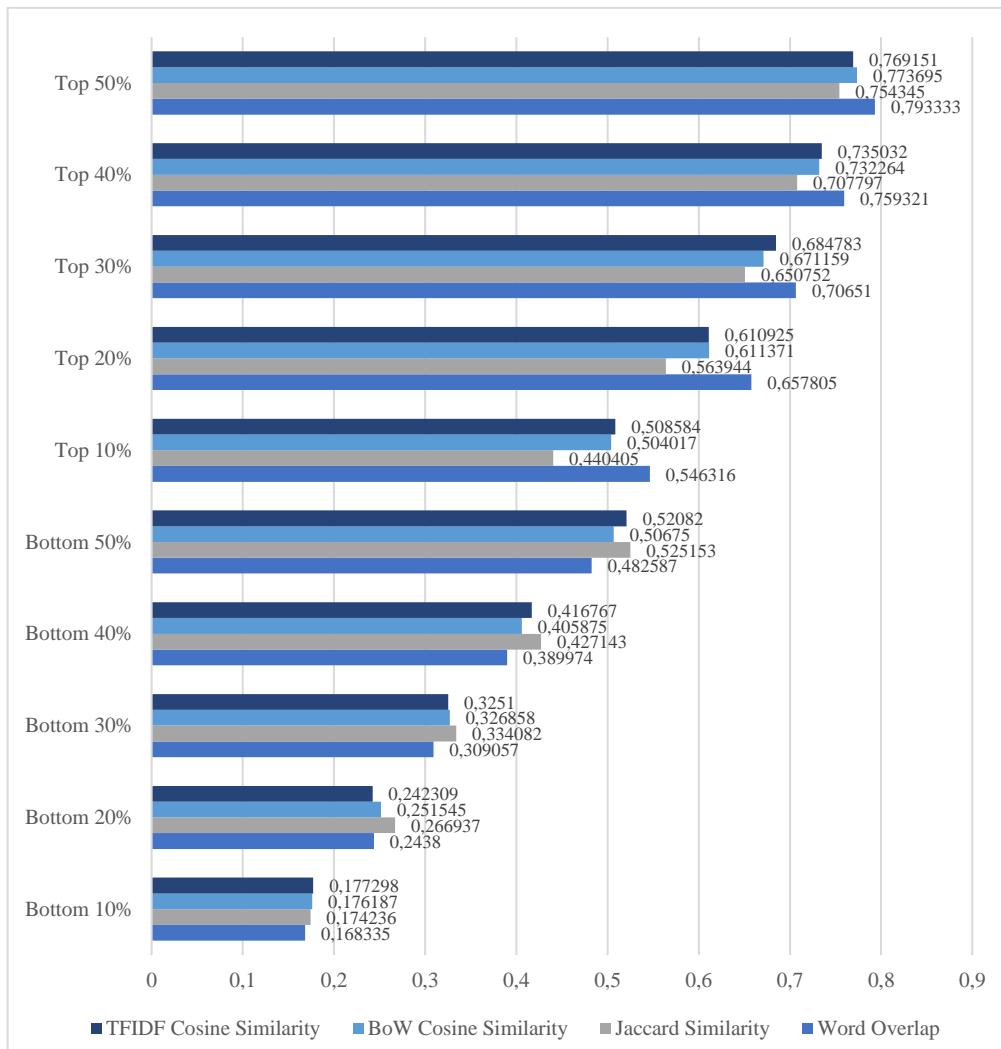
Pada ringkasan dengan *compression rate* 30%, rata-rata hasil *precision* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan TFIDF *cosine similarity* sebesar 0,761, *similarity word overlapping* sebesar 0,742, BoW *cosine similarity* sebesar 0,738 dan *Jaccard similarity* sebesar 0,787. Nilai *precision* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata *precision* dari keempat metode menunjukkan nilai 0,745 sedangkan nilai standar deviasinya sebesar 0,01. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara

keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *f-score* terbaik pada *compression rate* 30% adalah metode *TextRank* dengan TFIDF *cosine similarity*.

Pada ringkasan dengan *compression rate* 20%, rata-rata hasil *precision* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,665, BoW *cosine similarity* sebesar 0,664, TFIDF *cosine similarity* sebesar 0,663 dan *Jaccard similarity* sebesar 0,653. Nilai *precision* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata *precision* dari keempat metode menunjukkan nilai 0,745 sedangkan nilai standar deviasinya sebesar 0,01. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *f-score* terbaik pada *compression rate* 20% adalah metode *TextRank* dengan *similarity word overlapping*.

Pada ringkasan dengan *compression rate* 10%, rata-rata hasil *precision* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan TFIDF *cosine similarity* sebesar 0,559, BoW *cosine similarity* sebesar 0,542, *similarity word overlapping* sebesar 0,549 dan *Jaccard similarity* sebesar 0,520. Nilai *precision* dari keempat metode menunjukkan hasil yang tidak terlalu baik dalam melakukan peringkasan teks. Rata-rata *precision* dari keempat metode menunjukkan nilai 0,662 sedangkan nilai standar deviasinya sebesar 0,01. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang

memberikan nilai *f-score* terbaik pada *compression rate* 10% adalah metode *TextRank* dengan TFIDF cosine similarity.



Gambar 4.4 Diagram Rata-rata Hasil *Recall* (ROUGE-1)

Diagram *recall* pada Gambar 4.4 menunjukkan grafik yang hampir sama seperti diagram *precision*. Hal ini dikarenakan perhitungan *recall* juga dipengaruhi oleh kata yang tumpang tindih antara ringkasan referensi dan prediksi. Namun, perhitungan *recall* didasarkan pada jumlah kata pada ringkasan referensinya. Saat *compression rate* rendah, maka kalimat yang dipertahankan menjadi ringkasan

semakin sedikit, sehingga menyebabkan kecilnya peluang kata pada ringkasan prediksi saling tumpang tindih dengan kata yang terdapat pada ringkasan referensi.

Pada ringkasan dengan *compression rate* 50%, rata-rata hasil *recall* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,793, BoW *cosine similarity* sebesar 0,773, TFIDF *cosine similarity* sebesar 0,769 dan *Jaccard similarity* sebesar 0,754. Nilai *recall* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Nilai *recall* dari keempat metode menunjukkan perbedaan nilai yang relatif kecil, dibuktikan dengan kecilnya jarak antar nilai *recall score* tiap metode dengan nilai rata-rata *recall* dari keempat metode. Rata-rata *recall* dari keempat metode menunjukkan nilai 0,772 sedangkan nilai standar deviasinya sebesar 0,01 Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Sehingga penggunaan metode yang berbeda tidak terlalu berpengaruh terhadap hasil *recall* pada *compression rate* 50%. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *recall* terbaik pada *compression rate* 50% adalah metode *TextRank* dengan *similarity word overlapping*.

Pada ringkasan dengan *compression rate* 40%, rata-rata hasil *recall* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,759, TFIDF *cosine similarity* sebesar 0,735, BoW *cosine similarity* sebesar 0,732 dan *Jaccard similarity* sebesar 0,707. Nilai *recall* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata *recall* dari keempat metode menunjukkan

nilai 0,733 sedangkan nilai standar deviasinya sebesar 0,02. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *recall* terbaik pada *compression rate* 40% adalah metode *TextRank* dengan *similarity word overlapping*.

Pada ringkasan dengan *compression rate* 30%, rata-rata hasil *recall* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,706, TFIDF *cosine similarity* sebesar 0,684, BoW *cosine similarity* sebesar 0,671 dan *Jaccard similarity* sebesar 0,650. Nilai *recall* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata *recall* dari keempat metode menunjukkan nilai 0,678 sedangkan nilai standar deviasinya sebesar 0,02. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *recall* terbaik pada *compression rate* 30% adalah metode *TextRank* dengan *similarity word overlapping*.

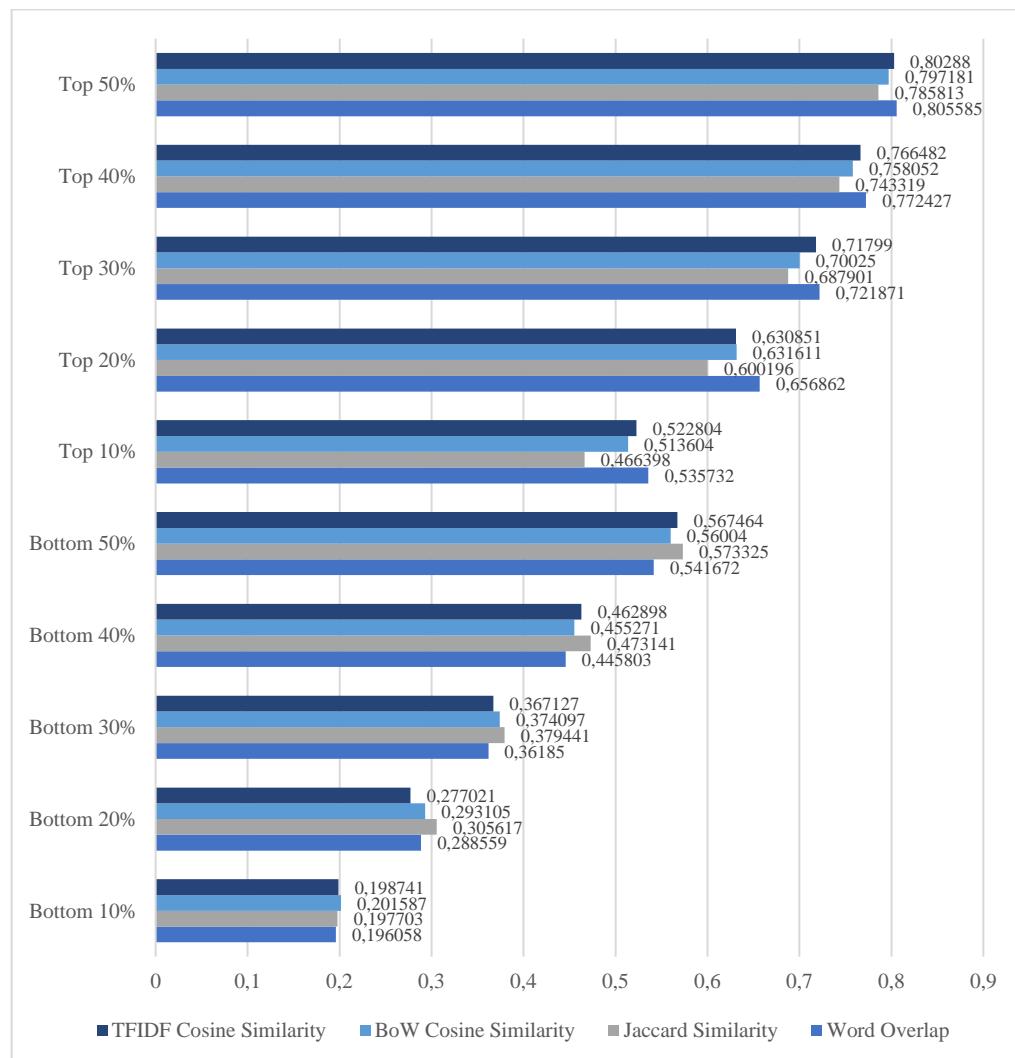
Pada ringkasan dengan *compression rate* 30%, rata-rata hasil *recall* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,706, TFIDF *cosine similarity* sebesar 0,684, BoW *cosine similarity* sebesar 0,671 dan *Jaccard similarity* sebesar 0,650. Nilai *recall* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata *recall* dari keempat metode menunjukkan nilai 0,678 sedangkan nilai standar deviasinya sebesar 0,02. Nilai standar deviasi

yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *recall* terbaik pada *compression rate* 30% adalah metode *TextRank* dengan *similarity word overlapping*.

Pada ringkasan dengan *compression rate* 20%, rata-rata hasil *recall* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,657, BoW *cosine similarity* sebesar 0,61, TFIDF *cosine similarity* sebesar 0,661 dan *Jaccard similarity* sebesar 0,563. Nilai *recall* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata *recall* dari keempat metode menunjukkan nilai 0,611 sedangkan nilai standar deviasinya sebesar 0,03. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *recall* terbaik pada *compression rate* 30% adalah metode *TextRank* dengan *similarity word overlapping*.

Pada ringkasan dengan *compression rate* 10%, rata-rata hasil *recall* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,546, TFIDF *cosine similarity* sebesar 0,508, BoW *cosine similarity* sebesar 0,504 dan *Jaccard similarity* sebesar 0,44. Nilai *recall* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata *recall* dari keempat metode menunjukkan nilai 0,499 sedangkan nilai standar deviasinya sebesar 0,04. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat

metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *recall* terbaik pada *compression rate* 10% adalah metode *TextRank* dengan *similarity word overlapping*.



Gambar 4.5 Diagram Rata-rata Hasil (ROUGE-1)

Diagram pada Gambar 4.5 menunjukkan grafik yang hampir sama seperti diagram *precision*. Hal ini dikarenakan perhitungan dipengaruhi oleh nilai *precision* dan *recall*. Nilai *f-score* yang dihasilkan mengalami penurunan ketika

*compression rate*-nya menurun (dari 50% ke 10%). Nilai *f-score* juga lebih tinggi ketika ringkasan yang diambil berada dalam kategori *top-n ranking*.

Pada ringkasan dengan *compression rate* 50%, rata-rata hasil *f-score* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,805, TFIDF *cosine similarity* sebesar 0,802, BoW *cosine similarity* sebesar 0,797 dan *Jaccard similarity* sebesar 0,785. Nilai *f-score* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Nilai *f-score* dari keempat metode menunjukkan perbedaan nilai yang relatif kecil, dibuktikan dengan kecilnya jarak antar nilai *f-score* tiap metode dengan nilai rata-rata *f-score* dari keempat metode. Rata-rata *f-score* dari keempat metode menunjukkan nilai 0,797 sedangkan nilai standar deviasinya sebesar 0,008. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Sehingga penggunaan metode yang berbeda tidak terlalu berpengaruh terhadap hasil *f-score* pada *compression rate* 50%. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *f-score* terbaik pada *compression rate* 50% adalah metode *TextRank* dengan *similarity word overlapping*.

Pada ringkasan dengan *compression rate* 40%, rata-rata hasil *f-score* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,772, TFIDF *cosine similarity* sebesar 0,766, BoW *cosine similarity* sebesar 0,758 dan *Jaccard similarity* sebesar 0,743. Nilai *f-score* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata *f-score* dari keempat metode menunjukkan

nilai 0,76 sedangkan nilai standar deviasinya sebesar 0,01. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *f-score* terbaik pada *compression rate* 40% adalah metode *TextRank* dengan *similarity word overlapping*.

Pada ringkasan dengan *compression rate* 30%, rata-rata hasil *f-score* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,721, TFIDF *cosine similarity* sebesar 0,717, BoW *cosine similarity* sebesar 0,70 dan *Jaccard similarity* sebesar 0,687. Nilai *f-score* dari keempat metode menunjukkan hasil yang cukup baik dalam melakukan peringkasan teks. Rata-rata *f-score* dari keempat metode menunjukkan nilai 0,70 sedangkan nilai standar deviasinya sebesar 0,01. Nilai standar deviasi yang kecil ini menunjukkan tidak ada perbedaan signifikan diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *f-score* terbaik pada *compression rate* 30% adalah metode *TextRank* dengan *similarity word overlapping*.

Pada ringkasan dengan *compression rate* 20%, rata-rata hasil *f-score* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,656, BoW *cosine similarity* sebesar 0,631, TFIDF *cosine similarity* sebesar 0,630 dan *Jaccard similarity* sebesar 0,60. Nilai *f-score* dari keempat metode menunjukkan hasil yang cukup dalam melakukan peringkasan teks. Rata-rata *f-score* dari keempat metode menunjukkan nilai 0,629 sedangkan nilai standar deviasinya sebesar 0,02. Nilai standar deviasi

yang cukup kecil ini menunjukkan perbedaan signifikan yang sangat kecil diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *f-score* terbaik pada *compression rate* 20% adalah metode *TextRank* dengan *similarity word overlapping*.

Pada ringkasan dengan *compression rate* 10%, rata-rata hasil *f-score* dari yang tertinggi ke yang terendah secara berurutan dicapai oleh metode *TextRank* dengan *similarity word overlapping* sebesar 0,535, TFIDF *cosine similarity* sebesar 0,522, BoW *cosine similarity* sebesar 0,513 dan *Jaccard similarity* sebesar 0,466. Nilai *f-score* dari keempat metode menunjukkan hasil yang tidak cukup baik dalam melakukan peringkasan teks. Rata-rata *f-score* dari keempat metode menunjukkan nilai 0,5 sedangkan nilai standar deviasinya sebesar 0,03. Nilai standar deviasi yang cukup kecil ini menunjukkan perbedaan signifikan yang sangat kecil diantara keempat metode. Namun berdasarkan data artikel yang digunakan, metode yang memberikan nilai *f-score* terbaik pada *compression rate* 10% adalah metode *TextRank* dengan *similarity word overlapping*.

Berdasarkan penelitian yang dilakukan oleh Mihalcea & Tarau (2004), ringkasan *TextRank* diambil berdasarkan skor kalimat tertinggi (*top-n ranking*). Ringkasan text dengan mengambil skor *TextRank* tertinggi terbukti menghasilkan ringkasan yang lebih baik dibandingkan dengan mengambil kalimat dengan skor *TextRank* paling kecil. Hal ini dapat dilihat dari hasil uji coba sistem menggunakan *confusion matrix* pada Tabel 4.12 dan uji coba menggunakan ROUGE-1 pada Tabel 4.13. Semua parameter menunjukkan hasil yang lebih tinggi pada kategori peringkasan dengan mengambil *score* *TextRank* tertinggi. Berdasarkan

perbandingan pengujian dengan dua kategori *top-n* dan *bottom-n*, terdapat perbedaan jarak nilai yang cukup besar antara kedua kategori. Hal ini dikarenakan kalimat yang benar diekstraksi menjadi ringkasan (sesuai dengan *ground truth*) sebagian besar berada pada peringkat atas *TextRank (top-n)*.

Data yang digunakan dalam penelitian ini merupakan data artikel *online*. Sehingga struktur penulisan artikel disesuaikan dengan kebutuhan dan jenis artikel yang digunakan. Sebagian besar artikel memiliki struktur yang terdiri dari judul, pendahuluan (*lead*), *body* artikel dan penutup (Jho, 2023). Judul menggambarkan isi artikel dan mengandung kata kunci yang relevan. Pendahuluan terdiri dari kumpulan kalimat yang membahas latar belakang dari topik yang diangkat secara umum. *Body* artikel merupakan inti artikel yang terdiri dari sub-bagian dan diikuti kalimat-kalimat argumen yang mendukung sub-bagian tersebut. Sub-bagian disajikan dalam bentuk kumpulan kalimat dengan suatu kalimat utama yang menjadi pembahasan dari sub-bagian tersebut. Penutup berisi kesimpulan dari pemaparan isi artikel. Kesimpulan dapat berisi pandangan penulis atau rekomendasi yang sesuai.

Kalimat yang terdapat dalam suatu artikel harus memiliki keterkaitan satu sama lain, yaitu keterkaitan yang menunjuk pada topik atau judul artikel yang dibahas. Keterkaitan antar kalimat dalam artikel merupakan salah satu aspek yang penting untuk memastikan alur dari artikel secara jelas. Kalimat dengan *similarity* antar kalimat yang tinggi, berarti kalimat tersebut saling terkait dengan kalimat lain. Kalimat ini cocok dipertahankan sebagai ringkasan karena kalimat tersebut mencangkap intisari dari kalimat-kalimat lainnya. Kalimat dengan *similarity* antar

kalimat yang rendah, tidak cocok diambil menjadi ringkasan karena tidak memiliki keterkaitan dengan kalimat lain. Biasanya dalam artikel, kalimat tersebut berbentuk kalimat yang terlalu umum, tidak relevan dan tidak memberi informasi khusus mengenai topik yang dibahas, selain itu bisa juga berbentuk informasi tambahan yang tidak terkait konteks artikel. Maka berdasarkan pernyataan tersebut, dapat dibuktikan bahwa kalimat dengan skor *similarity* tinggi merupakan kalimat yang harus dipertahankan menjadi ringkasan sistem.

Berdasarkan analisis dari tiap data artikel yang digunakan, Ringkasan dengan *compression rate* 10% tidak dapat mengambil semua poin penting yang terdapat pada dokumen. Suatu ringkasan harus memuat seluruh poin penting dalam dokumen, sedangkan ketika menggunakan *compression rate* 10%, beberapa poin penting terlewatkan. Hal ini menyebabkan kurangnya kelengkapan informasi suatu ringkasan. Banyaknya informasi penting yang terdapat dalam suatu dokumen dapat dilihat dari keberagaman informasi yang disajikan dalam sub-bagian artikel. Contohnya pada artikel-6 terdapat 6 sub-bagian penting yang harus ada dalam ringkasan. Banyaknya kalimat pada artikel-6 yaitu 29, sehingga kalimat yang akan diekstraksi menjadi ringkasan dengan *compression rate* 10% berjumlah 3 kalimat saja, padahal informasi penting yang harus ada dalam ringkasan berjumlah 6 informasi penting. Hal ini juga terjadi pada data artikel lain dimana hasil ringkasan 10% tidak menunjukkan kelengkapan informasi teks. Maka berdasarkan penelitian ini, ringkasan dengan *compression rate* 10% kurang baik dalam melakukan peringkasan teks. Sedangkan untuk hasil ringkasan dengan *compression rate* 20%

hingga 50%, masih mencangkup semua informasi penting dalam teks artikel, sehingga menunjukkan hasil ringkasan yang lebih baik.

Beberapa faktor yang mempengaruhi hasil kinerja sistem yaitu persebaran kata dan keberagaman informasi yang terkandung dalam teks. Persebaran kata, mengacu pada seberapa sering suatu kata muncul dalam teks dokumen yang akan diringkas. Kata yang sering muncul cenderung dianggap lebih penting dalam teks. Jika terdapat kata-kata yang tersebar secara merata dan seragam dalam teks, algoritma *TextRank* dapat mengalami kesulitan dalam mengidentifikasi kata-kata kunci yang lebih penting. Sebaliknya, jika ada kata-kata yang sangat dominan atau muncul secara berlebihan dalam teks, algoritma *TextRank* akan cenderung memberikan skor yang terlalu tinggi pada kalimat yang mengandung kata-kata tersebut. Selain itu, persebaran kata juga dapat mempengaruhi kualitas ringkasan dalam hal kelengkapan informasi. Jika kata-kata yang penting untuk memahami konteks atau inti teks tersebar secara merata, hasil ringkasan mungkin tidak dapat menangkap informasi yang penting atau keseluruhan makna teks dengan baik.

Faktor kedua yang dapat mempengaruhi hasil kinerja sistem yaitu keberagaman informasi yang terkandung dalam teks. Keberagaman informasi dapat dilihat dari jumlah poin penting yang dibahas dalam teks artikel. Pada *TextRank*, informasi yang muncul secara berulang atau memiliki keterkaitan yang kuat dianggap lebih penting. Jika teks asli memiliki keberagaman informasi tinggi yang terdiri dari banyak poin penting yang tersebar di berbagai bagian teks, maka beberapa poin penting bisa jadi tidak mendapatkan skor yang cukup dalam proses ringkasan. Algoritma *TextRank* tidak mampu mengenali semua poin penting

tersebut atau memberikan skor yang optimal pada masing-masing poin penting. Sebaliknya, jika informasi penting muncul secara berlebihan dalam satu bagian teks, algoritma *TextRank* memberikan skor yang lebih tinggi pada informasi tersebut.

Adanya sistem peringkasan teks secara otomatis ini dapat memberikan manfaat dengan cara memudahkan orang lain mengambil informasi penting pada suatu teks dokumen panjang. Ringkasan yang dihasilkan sistem juga dapat mengatasi kesulitan orang lain ketika membutuhkan informasi penting secara cepat. Manfaat yang dihasilkan oleh ringkasan sistem ini sesuai dengan hadist riwayat Thabranī :

أَحَبُّ النَّاسِ إِلَى اللَّهِ تَعَالَى أَنْفَعُهُمْ لِلنَّاسِ ، وَأَحَبُّ الْأَعْمَالِ إِلَى اللَّهِ تَعَالَى سُرُورٌ تُدْخِلُهُ عَلَى مُسْتَلِيمٍ ، أَوْ تَكْشِفُ عَنْهُ كُبْرَيَّةً ، أَوْ تَضْطَبِي عَنْهُ دِينًا ، أَوْ تَطْرُدُ عَنْهُ بُنُوغاً ، وَلَا يَنْأِي شَيْءٌ مَعَ أَخٍ فِي حَاجَةٍ أَحَبُّ إِلَيَّ مِنْ أَنْ أَعْتَكُنَّ فِي هَذَا الْمَسْجِدِ بِعْنَيِ مَسْجِدَ الْمَدِينَةِ شَهْرًا

*“Manusia yang paling dicintai oleh Allah adalah yang paling bermanfaat untuk manusia. Dan amalan yang paling dicintai oleh Allah adalah kegembiraan yang engkau masukan ke hati seorang mukmin, atau engkau hilangkan salah satu kesusahannya, atau engkau membayarkan hutangnya, atau engkau hilangkan kelaparannya. Dan aku berjalan bersama saudaraku untuk memenuhi kebutuhannya itu lebih aku cintai daripada ber-i’tikaf di masjid Nabawi selama sebulan lamanya. Dan siapa yang menahan marahnya maka Allah akan tutupi auratnya. Barangsiapa yang menahan marahnya padahal ia bisa menumpahkannya, maka Allah akan penuhi hatinya dengan keridhaan di hari kiamat. Dan barangsiapa berjalan bersama saudaranya sampai ia memenuhi kebutuhannya, maka Allah akan mengokohkan kedua kakinya di hari ketika banyak kaki-kaki terpeleset ke api neraka”* (HR. Ath Thabranī 6/139, dishahihkan Al Albani dalam Silsilah Ash Shahihah 2/575).

Berdasarkan hadist diatas, maka salah satu amalan yang disukai Allah SWT adalah mempermudah urusan orang lain, hal ini sesuai dengan konsep *Hablim Minannas* (hubungan dengan manusia). Adanya sistem peringkasan teks dapat

mempermudah urusan orang lain dalam memperoleh informasi secara cepat dan mudah. Ringkasan berisi intisari suatu teks, sehingga dengan membaca ringkasan, orang lain akan lebih cepat menangkap informasi penting yang terdapat dalam teks. Membaca sendiri merupakan suatu hal yang diperintahkan oleh Allah sesuai dengan Al-Qur'an surah Al-Alaq ayat 1-5 :

إِقْرَأْ بِاسْمِ رَبِّكَ الَّذِي حَلَقَ ۚ ۱ٖ خَلَقَ الْإِنْسَانَ مِنْ عَلَقٍ ۲ٗ إِقْرَأْ وَرِثْكَ الْأَكْرَمُ ۳ٗ الَّذِي عَلَمَ بِالْقَلْمَنِ ۴ٗ عَلَمَ الْإِنْسَانَ مَا  
لَمْ يَعْلَمْ ۵ٗ

*“1) Bacalah dengan (menyebut) nama Tuhanmu yang menciptakan, 2) Dia telah menciptakan manusia dari segumpal darah. 3) Bacalah, dan Tuhanmulah Yang Mahamulia, 4) Yang mengajar (manusia) dengan pena. 5) Dia mengajarkan manusia apa yang tidak diketahuinya.”*

Berdasarkan tafsir Jalalain, {إِقْرَأْ} memiliki maksud untuk mulai membaca dan memulainya dengan cara menyebut nama Rabb yang menciptakan seluruh makhluk. Kemudian {عَلَمَ الْإِنْسَانَ} memiliki maksud mengajarkan kepada manusia {مَا لَمْ يَعْلَمْ} suatu hal yang tidak diketahui oleh manusia. Kata {إِقْرَأْ} pada ayat tersebut memiliki berbagai makna diantaranya membaca, menyampaikan, menelaah, mendalami, meneliti, dan lain sebagainya. Ayat tersebut juga menganjurkan manusia untuk mencari ilmu sebanyak mungkin. Salah satu cara mencari ilmu pada ayat tersebut adalah dengan membaca. Salah satu sumber bacaan yang paling populer saat ini adalah artikel *online*. Dengan membaca artikel *online*, seseorang dapat mengambil informasi berupa ilmu pengetahuan yang bermanfaat. Adanya sistem peringkasan teks dapat membantu manusia dalam mengambil informasi karena isi ringkasan teks merupakan informasi yang penting yang berguna bagi

masyarakat. Orang yang berilmu akan diangkat derajadnya oleh Allah sebagaimana firman-Nya dalam Al-Qur'an surah Al-Mujadilah ayat 11:

يَأَيُّهَا الْمُلْكَيْنَ إِذَا قِيلَ لَكُمْ تَفَسَّحُوا فِي الْمَجَlisِ فَأَفْسِحُوا يَفْسِحَ اللَّهُ لَكُمْ وَإِذَا قِيلَ أَنْشُرُوا فَانْشُرُوا يَرْفَعُ اللَّهُ الْمُلْكَيْنَ إِذَا مَنَعُوا إِنْكُمْ وَالْمُلْكَيْنَ أُوتُوا الْعِلْمَ دَرَجَتٌ وَاللَّهُ إِنَّمَا تَعْمَلُونَ حَسِيرٌ

*"Wahai orang-orang yang beriman! Apabila dikatakan kepadamu, "Berilah kelapangan di dalam majelis-majelis," maka lapangkanlah, niscaya Allah akan memberi kelapangan untukmu. Dan apabila dikatakan, "Berdirilah kamu," maka berdirilah, niscaya Allah akan mengangkat (derajat) orang-orang yang beriman di antaramu dan orang-orang yang diberi ilmu beberapa derajat. Dan Allah Maha teliti apa yang kamu kerjakan." (Al-Mujadilah ayat 11)*

Berdasarkan tafsir Jalalain, Allah akan meninggikan derajat orang-orang yang diberi ilmu pengetahuan di surga nanti. Dengan ayat tersebut, pada saat mencari ilmu pengetahuan, manusia akan selalu mengingat bahwa Allah akan meninggikan derajat orang yang berilmu. Allah juga mencintai orang-orang yang memberi kemudahan atas kesusahan orang lain. Hal ini sesuai dengan tujuan peringkasan teks yaitu untuk memudahkan seseorang dalam mengambil informasi (pengetahuan) penting dalam suatu teks.

Konsep algoritma *textrank*, yaitu pengambilan ringkasan berdasarkan kalimat dengan keterkaitan (*similarity*) yang tinggi ini sesuai dengan konsep hukum *qiyas* dalam islam. *Qiyas* menurut ahli bahasa artinya menyamakan, mengukur atau membandingkan. *Qiyas* merupakan menetapkan suatu hukum atas permasalahan baru yang belum diketahui hukumnya dengan cara melihat kesamaan dalam '*illat*' hukum dari permasalahan yang sudah jelas *nash* (Tihami, 1995). Analogi *qiyas* seperti pada hukum minuman yang membuat mabuk yang terdapat pada Hadist Riwayat Muslim, no 2003:

كُلُّ مُسْكِرٍ حَمْرٌ وَكُلُّ مُسْكِرٍ حَرَامٌ وَمَنْ شَرِبَ الْحَمْرَ فِي الدُّنْيَا فَمَاتَ وَهُوَ يُدْمِنُهَا لَمْ يَئُنْ مَمْ يَشْرُبُهَا فِي الْآخِرَةِ

*“Segala sesuatu yang memabukkan itu khamar. Segala sesuatu yang memabukkan itu haram. Siapa saja meminum khamar di dunia lalu ia meninggal dunia dalam keadaan kecanduan dan tidak bertaubat, maka ia tidak akan meminum khamar (yang penuh nikmat) di akhirat.”* (HR. Muslim, no. 2003)

Pada hadist diatas, dinyatakan bahwa setiap yang memabukkan adalah *khamr*, sehingga hukumnya sama dengan *khamr* yaitu haram. Maka berdasarkan hasil penelitian, dapat diambil pembelajaran bahwa kalimat yang diambil menjadi ringkasan merupakan kalimat yang memiliki *similarity* yang tinggi dengan kalimat lain. Hal ini untuk memastikan alur dari dokumen dengan jelas. Penetapan kalimat dengan *similarity* tertinggi ini juga sesuai dengan konsep penetapan hukum *qiyas*, yaitu dengan melihat adanya kesamaan atas dua peristiwa.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan hasil implementasi serta analisis hasil uji coba sistem, maka dapat diambil kesimpulan bahwa metode *TextRank* dapat diterapkan dalam peringkasan ekstraktif pada artikel *online*. Pada uji coba berdasarkan metode *similarity* yang digunakan, didapatkan nilai *f-score* tertinggi dicapai oleh metode *similarity word overlapping*. Namun jika dibandingkan dengan metode *similarity* lain, maka didapatkan standar deviasi yang sangat kecil berdasarkan nilai *f-score* keempat metode. Maka dapat dikatakan bahwa penggunaan metode *similarity* yang berbeda dalam penelitian ini tidak menunjukkan perbedaan yang signifikan. Pada uji coba berdasarkan nilai *compression rate*, maka didapatkan hasil bahwa semakin tinggi nilai *compression rate* maka nilai *f-score* juga semakin tinggi. Sehingga nilai *f-score* tertinggi dicapai ketika *compression rate* sebesar 50%. Pada uji coba pengambilan ringkasan berdasarkan posisi ranking skor *TextRank*. Maka didapatkan hasil bahwa pada posisi *top-n* ranking memberikan nilai *f-score* yang lebih tinggi dibandingkan pada posisi *bottom-n* ranking. Maka jika ditarik kesimpulan, berdasarkan hasil uji coba yang telah dilakukan, rata-rata nilai *f-score* terbaik dicapai oleh metode *TextRank* dengan *similarity word overlapping* pada *compression rate* 50% di kategori posisi *top-n* ranking. Nilai *f-score* pada *confission matrix* sebesar 0,705 sedangkan nilai *f-score* dengan ROUGE-1 sebesar 0,805.

## 5.2 Saran

Peringkasan artikel menggunakan algoritma *TextRank* ini masih memiliki beberapa kekurangan. Maka untuk menghasilkan kinerja yang lebih maksimal, terdapat saran untuk penelitian di masa depan adalah:

- a. Menambahkan fitur *query expansion* untuk meningkatkan kinerja algoritma *similarity*.
- b. Menggunakan ekstraksi fitur *word embedding* untuk membangun vektor kalimat untuk metode *cosine similarity*.
- c. Menggunakan metode *similarity* lain yang seperti *soft cosine similarity*, *word mover distance* dll.
- d. Menggunakan data berupa dokumen yang memiliki ukuran lebih panjang.

## DAFTAR PUSTAKA

- Adriani, M., Asian, J., Nazief, B., Tahaghoghi, S. M. M., & Williams, H. E. (2007). Stemming Indonesian: A confix-stripping approach. *ACM Transactions on Asian Language Information Processing*, 6(4), 1–33. <https://doi.org/10.1145/1316457.1316459>
- Alguliyev, R., & Aliguliyev, R. (2009). Evolutionary Algorithm for Extractive Text Summarization. *Intelligent Information Management*, 01(02), 128–138. <https://doi.org/10.4236/iim.2009.12019>
- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., D., E., B., J., & Kochut, K. (2017). Text Summarization Techniques: A Brief Survey. *International Journal of Advanced Computer Science and Applications*, 8(10), 319–321. <https://doi.org/10.14569/ijacsa.2017.081052>
- Andriani, D., & Tanzil Furqon, M. (2019). Peringkasan Teks Otomatis Pada Artikel Berita Hiburan Berbahasa Indonesia Menggunakan Metode BM25. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(3), 2548–2964. <http://j-ptiik.ub.ac.id>
- Arifin, A. Z., Qomariah, D. U. N., Riduwan, M., Haniefardy, A., Azhar, Y., Sholikah, R. W., & Navastara, D. A. (2018). Automatic comparison of products based on opinion features using synonym and jaccard similarity. *Proceedings of the 3rd International Conference on Informatics and Computing, ICIC 2018*, 1–6. <https://doi.org/10.1109/IAC.2018.8780458>
- Cheng, Z., & Guo, S. (2022). Automatic Text Summarization for Public Health WeChat Official Accounts Platform Base on Improved TextRank. *Journal of Environmental and Public Health*, 2022, 1–9. <https://doi.org/10.1155/2022/1166989>
- Chouigui, A., Ben Khiroun, O., & Elayeb, B. (2021). An Arabic Multi-source News Corpus: Experimenting on Single-document Extractive Summarization. *Arabian Journal for Science and Engineering*, 46(4), 3925–3938. <https://doi.org/10.1007/s13369-020-05258-z>
- El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2021). Automatic text summarization: A comprehensive survey. *Expert Systems with Applications*, 165, 113679. <https://doi.org/10.1016/j.eswa.2020.113679>
- Elbarougy, R., Behery, G., & El Khatib, A. (2020). Extractive Arabic Text Summarization Using Modified PageRank Algorithm. *Egyptian Informatics Journal*, 21(2), 73–81. <https://doi.org/10.1016/j.eij.2019.11.001>
- Fakhrezi, M. F., Bijaksana, M. A., & Huda, A. F. (2021). Implementation of

Automatic Text Summarization with TextRank Method in the Development of Al-Qur'an Vocabulary Encyclopedia. *Procedia Computer Science*, 179(2020), 391–398. <https://doi.org/10.1016/j.procs.2021.01.021>

Firman A, D., Fahrur Rozi, I., & Kusumaning Putri, I. (2022). Peringkasan Teks Otomatis pada Portal Berita Olahraga menggunakan metode Maximum Marginal Relevance. *Jurnal Informatika Polinema*, 8(3), 21–30. <https://doi.org/10.33795/jip.v8i3.519>

Gulden, C., Kirchner, M., Schüttler, C., Hinderer, M., Kampf, M., Prokosch, H. U., & Toddenroth, D. (2019). Extractive summarization of clinical trial descriptions. *International Journal of Medical Informatics*, 129(May), 114–121. <https://doi.org/10.1016/j.ijmedinf.2019.05.019>

Han, J., Kamber, M., & Pei, J. (2012). Getting to Know Your Data. In *Data Mining*. <https://doi.org/10.1016/b978-0-12-381479-1.00002-2>

Luhn, H. P. (1958). The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, 2(2), 159–165. <https://doi.org/10.1147/rd.22.0159>

Mallick, C., Das, A. K., Dutta, M., Das, A. K., & Sarkar, A. (2018). Graph-based text summarization using modified TextRank. In *Advances in Intelligent Systems and Computing* (Vol. 758). Springer Singapore. [https://doi.org/10.1007/978-981-13-0514-6\\_14](https://doi.org/10.1007/978-981-13-0514-6_14)

Manalu, D. R., Rajagukguk, E., Siringoringo, R., Siahaan, D. K., & Sihombing, P. (2019). The Development of Document Similarity Detector by Jaccard Formulation. *2019 International Conference of Computer Science and Information Technology, ICoSNIKOM 2019*. <https://doi.org/10.1109/ICoSNIKOM48755.2019.9111494>

Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing order into texts. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, EMNLP 2004 - A Meeting of SIGDAT, a Special Interest Group of the ACL Held in Conjunction with ACL 2004*, 85, 404–411.

Nawaz, A., Bakhtyar, M., Baber, J., Ullah, I., Noor, W., & Basit, A. (2020). Extractive Text Summarization Models for Urdu Language. *Information Processing and Management*, 57(6), 102383. <https://doi.org/10.1016/j.ipm.2020.102383>

Pabbi, K., & Sindhu, C. (2021). Opinion Summarisation using Bi-Directional Long-Short Term Memory. *2021 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2021*, 256–259. <https://doi.org/10.1109/WiSPNET51692.2021.9419412>

- Rani, U., & Bidhan, K. (2021). Comparative Assessment of Extractive Summarization: TextRank, TF-IDF and LDA. *Journal of Scientific Research*, 65(01), 304–311. <https://doi.org/10.37398/jsr.2021.650140>
- Shah, C., & Jivani, A. G. (2017). Multi-document summarization: Study on existing techniques. In *International Conference on Advances in Computing, Communication and Informatics, September*.
- Sintia, S., Defit, S., & Nurcahyo, G. W. (2021). Product Codification Accuracy With Cosine Similarity And Weighted Term Frequency And Inverse Document Frequency (TF-IDF). *Journal of Applied Engineering and Technological Science (JAETS)*, 2(2), 62–69. <https://doi.org/10.37385/jaets.v2i2.210>
- Sitikhu, P., Pahi, K., Thapa, P., & Shakya, S. (2019). A Comparison of Semantic Similarity Methods for Maximum Human Interpretability. *International Conference on Artificial Intelligence for Transforming Business and Society, AITB 2019*. <https://doi.org/10.1109/AITB48515.2019.8947433>
- Tihami. (1995). *Al-qiyas dan al-dalil*.
- Ubaidillah, M. A., Dwidasmara, I. B. G., & Muliantara, A. (2020). Peringkas Tekst Otomatis Berita Online Menggunakan Metode Cross Latent Semantic Analysis & Cosine Similarity. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, 9(1), 105. <https://doi.org/10.24843/jlk.2020.v09.i01.p11>
- Verma, P. K., Agarwal, S., & Khan, M. A. (2017). Opinion mining considering roman words using Jaccard similarity algorithm based on clustering. *8th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2017*, 5–8. <https://doi.org/10.1109/ICCCNT.2017.8204041>
- Verma, P., Pal, S., & Om, H. (2019). A comparative analysis on Hindi and English extractive text summarization. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 18(3). <https://doi.org/10.1145/3308754>
- Victor, D. M., Eduardo, F. F., Biswas, R., Alegre, E., & Fernández-Robles, L. (2019). Application of Extractive Text Summarization Algorithms to Speech-to-Text Media. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11734 LNAI(December 2020), 540–550. [https://doi.org/10.1007/978-3-030-29859-3\\_46](https://doi.org/10.1007/978-3-030-29859-3_46)
- Yang, D., & Zhang, A. N. (2019). Performing literature review using text mining, Part III: Summarizing articles using TextRank. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 3186–3190. <https://doi.org/10.1109/BigData.2018.8622408>

- Zhao, B. (2017). Encyclopedia of Big Data. In *Encyclopedia of Big Data* (Issue May, pp. 1–3). <https://doi.org/10.1007/978-3-319-32001-4>
- Zhou, N., Shi, W., Liang, R., & Zhong, N. (2022). TextRank Keyword Extraction Algorithm Using Word Vector Clustering Based on Rough Data-Deduction. *Computational Intelligence and Neuroscience*, 2022. <https://doi.org/10.1155/2022/5649994>
- Zong, C., Xia, R., & Zhang, J. (2021). Text Data Mining. In *Tsinghua University Press*. [https://doi.org/10.1007/978-981-16-0100-2\\_5](https://doi.org/10.1007/978-981-16-0100-2_5)

## **LAMPIRAN**

## Lampiran 1 :

Data judul dari teks artikel yang digunakan

No.	Judul Artikel	Statistik		
		Kalimat	Kata	KataUnik
1	Kenapa Lengan T-rex Pendek tapi Jadi Predator Ganas?	20	331	186
2	Mahasiswa Kerap Overthinking? Ini 4 Tips Menjaga Kesehatan Mental ala Pakar Itera	26	450	232
3	Kenapa Kita Sering Overthinking pada Malam Hari? Ini Alasannya	37	565	265
4	7 Cara Atasi Overthinking yang Kerap Muncul di Malam Hari, Sudah Lakukan?	41	528	234
5	Tidak Tidur Semalaman Bikin Otak Jadi Tua, Begini Studinya	29	657	277
6	6 Cara Memperbaiki Jadwal Tidur yang Berantakan	29	447	189
7	Bisakah ChatGPT Menggeser Peran Guru di Ruang Kelas?	47	966	373
8	Cuci Muka Pakai Air Dingin atau Air Hangat, Mana yang Lebih Sehat untuk Kulit	35	509	223
9	5 Manfaat Minum Air Hangat, Turunkan Berat Badan hingga Buang Racun dalam Tubuh	31	386	180
10	Parenting Itu Penting?	20	368	175
11	Moms, Ini Cara Bantu Anak Agar Berani Hadapi Rasa Malunya	39	528	235
12	5 Tanda Anak Dapat Kekerasan dari Orang Terdekat	29	513	198
13	5 Manfaat Mengurangi Konsumsi Gula, Bikin Kulit Glowing	38	644	225
14	6 Bahan Alami yang Bisa Bantu Detoksifikasi Ginjal Kamu	36	534	272
15	Naik Motor Sambil Merokok Bisa Didenda, Ini Aturan Hukumnya	20	285	130
16	Benarkah Rokok Elektrik Lebih Aman dari Rokok Konvensional?	41	657	319
17	10 Rokok Termahal di Dunia, Perokok Ketengen Minggir Dulu!	68	799	273
18	Tips Memulai Karier sebagai Investor Pemula, yuk Mulai!	64	1029	448
19	Tips Keuangan untuk Mempersiapkan Pernikahan, Poin Pertama Sering Skip	28	340	184
20	6 Cara Mengurangi Kecemasan Secara Alami, Jangan Langsung Minum Obat	49	695	326
21	Benarkah Nyeri Haid Setara dengan Serangan Jantung?	45	574	265
22	Seperti Emily in Paris, Kenapa Perempuan Potong Rambut saat Stres soal Cinta?	28	532	263
23	Terimakasih atau Terima Kasih, Manakah yang Benar?	23	321	145
24	Pengertian Law Of Attraction dan Cara Menerapkannya secara Positif	25	373	169
25	Filosofi Serigala sebagai Hewan yang Memegang Teguh Prinsip	31	366	196
26	3 Filosofi Padi yang Bisa Diterapkan pada Kehidupan	28	329	165
27	5 Filosofi Garam untuk Menghadapi Kerasnya Kehidupan	28	383	117

No.	Judul Artikel	Statistik		
		Kalimat	Kata	KataUnik
28	Filosofi Gitar, Alat Musik yang Ternyata Penuh Makna	30	359	182
29	Cara Bersosialisasi dengan Baik dalam Kehidupan Masyarakat	20	352	159
30	4 Manfaat Afirmasi Positif untuk Diri Sendiri, Bikin Pikiran Lebih Positif	32	409	189
31	Punya Pasangan dengan Love Language Words of Affirmation? Lakukan 3 Hal Ini	33	592	264
32	Survei Ungkap Situationship Jadi Tren di Kalangan Perempuan Gen Z Selama 2022	28	468	214
33	5 Fakta tentang Kegagalan, Semua Momen Ada Waktunya!	40	481	199
34	5 Cara Dewasa Menyikapi Kegagalan, Jangan Putus Asa!	34	331	183
35	Bukan Arabika, Inilah Pilihan Kopi Terbaik untuk Begadang Kata Ahli	24	364	189
36	Wajib Tahu, Inilah Tips Menikmati Kopi yang Baik untuk Kesehatan	37	436	209
37	4 Cara Mengatasi Susah Tidur demi Kesehatan	25	320	142
38	Feminisme Sebagai Perolehan dalam Kesamaan Hak Asasi Manusia	24	350	184
39	Sering Ngeblank dan Susah Fokus, Kenali Gejala Brain Fog dan Cara Mengatasinya	36	506	258
40	3 Cara Memilih Sunscreen yang Tepat untuk Cuaca Terik agar Kulit Terlindungi	21	380	195
41	Selain Dimilai Kurang Cerdas, Ini Dampak Negatif Tidak Mampu Mengontrol Emosi dengan Baik, Simak!	32	417	177
42	Tanpa Sadar, 3 Kalimat 'Toksik' Ini Sering Diucapkan Orang Cerdas! Duh, Apa Saja?	35	571	295
43	Terlalu Perfeksionis, Ini 5 Tanda Kamu Orang yang Mudah Insecure!	28	376	169
44	Toxic Masculinity, Pemicu Ketidaksetaraan Gender?	24	666	274
45	5 Tips yang Bisa Diterapkan saat Kamu Terjebak dalam Kesepian	32	432	188
46	5 Alasan Tegas pada Orang Egois Itu Perlu, demi Harga Diri	43	454	211
47	5 Tips agar Anak Mau Mengakui Kesalahan Tanpa Memarahinya	37	472	199
48	5 Tips Membangun Hubungan Tipe Low Maintenance	37	576	221
49	Penembakan AS Menggilir, Kenapa Biden Tak Tetapkan Darurat Senjata Api?	36	556	269
50	Buntut Penembakan Massal di AS, Biden Dukung UU Larangan Senjata Serbu	20	405	212
51	Harusnya Ekspor Tembaga Disetop Juni 2023, Kenapa Jokowi Masih Izinkan Freeport?	37	554	287
52	Pemerintah Mau Tambah Kepemilikan Saham di Freeport 10 Persen	31	518	243
53	Investor Percaya Indonesia, Penanaman Modal Asing Tumbuh 20,2 Persen	17	393	171
54	Kenapa Sering Mengantuk saat Puasa? Ternyata Ini Penyebabnya	37	394	189
55	Saat Bersin, Seberapa Jauh Kuman Infeksius Bisa Menyebar?	49	931	383

No.	Judul Artikel	Statistik		
		Kalimat	Kata	KataUnik
56	Apakah Bisa Bersin saat Tidur? Ini Fakta Sebenarnya	34	390	186
57	Mengenali Tanda-tanda Depresi pada Anak-anak, Jangan Diabaikan	38	637	209
58	Apa yang Terjadi pada Tubuh saat Kamu Dibius Total?	75	1103	378
59	5 Tips Menyikapi Orang Toksik Tanpa Konflik Berkepanjangan	56	652	318
60	Taklukkan Orangtua Egois dengan 5 Sikap Cerdas Kamu	31	387	173
61	Perbedaan Makna Traveling dengan Holiday, Kamu Sudah Tahu?	31	395	194
62	5 Alasan Kenapa Banyak Orang Gak Senang Jadi Dewasa	46	666	285
63	Jangan Jadikan Wikipedia sebagai Sumber Tulisanmu, Ini Penjelasannya	49	759	294
64	5 Hal yang Bikin Permintaan Maafmu Terlihat Murah, Segera Hindari	38	465	215
65	Mengenal Apa Itu Quality Time dalam Love Language	30	432	163
66	Perkawinan Sedarah Ancam Kepunahan Badak Jawa di Ujung Kulon	19	359	189
67	Hati-hati Tertipu QRIS Palsu Kotak Amal, Cek 2 Kali Bagian Ini Sebelum Transfer	17	278	161
68	Waspada, Serangan Ransomware Tumbuh Signifikan	11	225	120
69	Fenomena Geologi: Afrika Bakal Terbelah Jadi Dua, Ciptakan Samudra Baru	24	434	220
70	Berapa Banyak Air Liur yang Kamu Hasilkan Seumur Hidup?	29	494	254
71	Elon Musk Ngaku Beli Twitter Karena 'Terpaksa'	14	219	118
72	Berapa Banyak Pelukan yang Dibutuhkan Seorang Anak dalam Sehari?	22	385	194
73	Generasi Z: Melek Intelektual, Merem Tata Krama	38	874	328
74	Pentingnya Netralitas Tubuh untuk Mengubah Pola Pikir Citra Diri	32	523	246
75	Gampang Terdistraksi Saat Bekerja? Lakukan 5 Hal Ini untuk Tingkatkan Konsentrasi!	42	512	259
76	Bukan Memenangkan Perdebatan, Ini 3 Cara Terbaik untuk Mengubah Pikiran Orang Lain	21	370	184
77	Alasan yang Mendorong Seseorang Jadi Oversharing, Kamu Pernah Mengalaminya Juga?	26	335	173
78	Waspada, Ini 6 Tanda Kamu adalah People Pleaser dalam Hubungan Asmara	25	478	217
79	5 Kebiasaan Buruk yang Wajib Dihindari Demi Menyehatkan Mental, Kebiasaan Perfeksionis Termasuk!	34	530	234
80	Sering Rebahan Bisa Membuat Kamu Jadi Kurang Cerdas, Ini 5 Dampak Buruk Lainnya!	27	328	163

## Lampiran 2 :

Hasil Confusion Matrix - Textrank dengan *similarity word overlapping*, dengan  
Compression Rate 10%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	2	1	17	1	1	0,5	0,5	0,9	0,5
2	26	3	0	20	3	3	0	0	0,769231	0
3	37	4	1	30	3	3	0,25	0,25	0,837838	0,25
4	41	4	1	34	3	3	0,25	0,25	0,853659	0,25
5	29	3	2	25	1	1	0,666667	0,666667	0,931034	0,666667
6	29	3	1	24	2	2	0,333333	0,333333	0,862069	0,333333
7	47	5	0	37	5	5	0	0	0,787234	0
8	35	4	3	30	1	1	0,75	0,75	0,942857	0,75
9	31	3	2	27	1	1	0,666667	0,666667	0,935484	0,666667
10	20	2	1	17	1	1	0,5	0,5	0,9	0,5
11	39	4	1	32	3	3	0,25	0,25	0,846154	0,25
12	29	3	1	24	2	2	0,333333	0,333333	0,862069	0,333333
13	38	4	1	31	3	3	0,25	0,25	0,842105	0,25
14	36	4	0	28	4	4	0	0	0,777778	0
15	20	2	2	18	0	0	1	1	1	1
16	41	4	2	35	2	2	0,5	0,5	0,902439	0,5
17	68	7	2	56	5	5	0,285714	0,285714	0,852941	0,285714
18	64	6	4	56	2	2	0,666667	0,666667	0,9375	0,666667
19	28	3	0	22	3	3	0	0	0,785714	0
20	49	5	2	41	3	3	0,4	0,4	0,877551	0,4
21	45	4	2	39	2	2	0,5	0,5	0,911111	0,5
22	28	3	1	23	2	2	0,333333	0,333333	0,857143	0,333333
23	23	2	0	19	2	2	0	0	0,826087	0
24	25	2	1	22	1	1	0,5	0,5	0,92	0,5
25	31	3	0	25	3	3	0	0	0,806452	0
26	28	3	0	22	3	3	0	0	0,785714	0
27	28	3	0	22	3	3	0	0	0,785714	0
28	30	3	2	26	1	1	0,666667	0,666667	0,933333	0,666667
29	20	2	0	16	2	2	0	0	0,8	0
30	32	3	1	27	2	2	0,333333	0,333333	0,875	0,333333
31	33	3	1	28	2	2	0,333333	0,333333	0,878788	0,333333
32	28	3	2	24	1	1	0,666667	0,666667	0,928571	0,666667
33	40	4	1	33	3	3	0,25	0,25	0,85	0,25
34	34	3	1	29	2	2	0,333333	0,333333	0,882353	0,333333
35	24	2	1	21	1	1	0,5	0,5	0,916667	0,5
36	37	4	2	31	2	2	0,5	0,5	0,891892	0,5
37	25	2	0	21	2	2	0	0	0,84	0
38	24	2	1	21	1	1	0,5	0,5	0,916667	0,5
39	36	4	4	32	0	0	1	1	1	1
40	21	2	0	17	2	2	0	0	0,809524	0
41	32	3	0	26	3	3	0	0	0,8125	0
42	35	4	1	28	3	3	0,25	0,25	0,828571	0,25
43	28	3	1	23	2	2	0,333333	0,333333	0,857143	0,333333
44	24	2	2	22	0	0	1	1	1	1
45	32	3	1	27	2	2	0,333333	0,333333	0,875	0,333333

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	4	0	35	4	4	0	0	0,813953	0
47	37	4	0	29	4	4	0	0	0,783784	0
48	37	4	2	31	2	2	0,5	0,5	0,891892	0,5
49	36	4	2	30	2	2	0,5	0,5	0,888889	0,5
50	20	2	0	16	2	2	0	0	0,8	0
51	37	4	2	31	2	2	0,5	0,5	0,891892	0,5
52	31	3	1	26	2	2	0,333333	0,333333	0,870968	0,333333
53	17	2	2	15	0	0	1	1	1	1
54	37	4	4	33	0	0	1	1	1	1
55	49	5	2	41	3	3	0,4	0,4	0,877551	0,4
56	34	3	1	29	2	2	0,333333	0,333333	0,882353	0,333333
57	38	4	3	33	1	1	0,75	0,75	0,947368	0,75
58	75	8	1	60	7	7	0,125	0,125	0,813333	0,125
59	56	6	1	45	5	5	0,166667	0,166667	0,821429	0,166667
60	31	3	2	27	1	1	0,666667	0,666667	0,935484	0,666667
61	31	3	1	26	2	2	0,333333	0,333333	0,870968	0,333333
62	46	5	0	36	5	5	0	0	0,782609	0
63	49	5	2	41	3	3	0,4	0,4	0,877551	0,4
64	38	4	1	31	3	3	0,25	0,25	0,842105	0,25
65	30	3	2	26	1	1	0,666667	0,666667	0,933333	0,666667
66	19	2	1	16	1	1	0,5	0,5	0,894737	0,5
67	17	2	0	13	2	2	0	0	0,764706	0
68	11	1	1	10	0	0	1	1	1	1
69	24	2	1	21	1	1	0,5	0,5	0,916667	0,5
70	29	3	2	25	1	1	0,666667	0,666667	0,931034	0,666667
71	14	1	0	12	1	1	0	0	0,857143	0
72	22	2	1	19	1	1	0,5	0,5	0,909091	0,5
73	38	4	1	31	3	3	0,25	0,25	0,842105	0,25
74	32	3	1	27	2	2	0,333333	0,333333	0,875	0,333333
75	42	4	2	36	2	2	0,5	0,5	0,904762	0,5
76	21	2	0	17	2	2	0	0	0,809524	0
77	26	3	0	20	3	3	0	0	0,769231	0
78	25	2	0	21	2	2	0	0	0,84	0
79	34	3	0	28	3	3	0	0	0,823529	0
80	27	3	1	22	2	2	0,333333	0,333333	0,851852	0,333333
Rata-rata							0,355551	0,355551	0,871334	0,355551

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

### Lampiran 3 :

Hasil Confusion Matrix - Textrank dengan *similarity word overlapping*, dengan  
Compression Rate 20%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	4	3	15	1	1	0,75	0,75	0,9	0,75
2	26	5	2	18	3	3	0,4	0,4	0,769231	0,4
3	37	7	2	25	5	5	0,285714	0,285714	0,72973	0,285714
4	41	8	2	27	6	6	0,25	0,25	0,707317	0,25
5	29	6	2	19	4	4	0,333333	0,333333	0,724138	0,333333
6	29	6	4	21	2	2	0,666667	0,666667	0,862069	0,666667
7	47	9	2	31	7	7	0,222222	0,222222	0,702128	0,222222
8	35	7	5	26	2	2	0,714286	0,714286	0,885714	0,714286
9	31	6	4	23	2	2	0,666667	0,666667	0,870968	0,666667
10	20	4	2	14	2	2	0,5	0,5	0,8	0,5
11	39	8	4	27	4	4	0,5	0,5	0,794872	0,5
12	29	6	3	20	3	3	0,5	0,5	0,793103	0,5
13	38	8	5	27	3	3	0,625	0,625	0,842105	0,625
14	36	7	4	26	3	3	0,571429	0,571429	0,833333	0,571429
15	20	4	3	15	1	1	0,75	0,75	0,9	0,75
16	41	8	6	31	2	2	0,75	0,75	0,902439	0,75
17	68	14	10	50	4	4	0,714286	0,714286	0,882353	0,714286
18	64	13	9	47	4	4	0,692308	0,692308	0,875	0,692308
19	28	6	3	19	3	3	0,5	0,5	0,785714	0,5
20	49	10	7	36	3	3	0,7	0,7	0,877551	0,7
21	45	9	6	33	3	3	0,666667	0,666667	0,866667	0,666667
22	28	6	4	20	2	2	0,666667	0,666667	0,857143	0,666667
23	23	5	3	16	2	2	0,6	0,6	0,826087	0,6
24	25	5	2	17	3	3	0,4	0,4	0,76	0,4
25	31	6	3	22	3	3	0,5	0,5	0,806452	0,5
26	28	6	3	19	3	3	0,5	0,5	0,785714	0,5
27	28	6	2	18	4	4	0,333333	0,333333	0,714286	0,333333
28	30	6	2	20	4	4	0,333333	0,333333	0,733333	0,333333
29	20	4	1	13	3	3	0,25	0,25	0,7	0,25
30	32	6	5	25	1	1	0,833333	0,833333	0,9375	0,833333
31	33	7	4	23	3	3	0,571429	0,571429	0,818182	0,571429
32	28	6	4	20	2	2	0,666667	0,666667	0,857143	0,666667
33	40	8	5	29	3	3	0,625	0,625	0,85	0,625
34	34	7	2	22	5	5	0,285714	0,285714	0,705882	0,285714
35	24	5	3	17	2	2	0,6	0,6	0,833333	0,6
36	37	7	4	27	3	3	0,571429	0,571429	0,837838	0,571429
37	25	5	4	19	1	1	0,8	0,8	0,92	0,8
38	24	5	3	17	2	2	0,6	0,6	0,833333	0,6
39	36	7	6	28	1	1	0,857143	0,857143	0,944444	0,857143
40	21	4	2	15	2	2	0,5	0,5	0,809524	0,5
41	32	6	4	24	2	2	0,666667	0,666667	0,875	0,666667
42	35	7	4	25	3	3	0,571429	0,571429	0,828571	0,571429
43	28	6	2	18	4	4	0,333333	0,333333	0,714286	0,333333
44	24	5	4	18	1	1	0,8	0,8	0,916667	0,8
45	32	6	3	23	3	3	0,5	0,5	0,8125	0,5

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	9	3	28	6	6	0,333333	0,333333	0,72093	0,333333
47	37	7	1	24	6	6	0,142857	0,142857	0,675676	0,142857
48	37	7	4	27	3	3	0,571429	0,571429	0,837838	0,571429
49	36	7	4	26	3	3	0,571429	0,571429	0,833333	0,571429
50	20	4	2	14	2	2	0,5	0,5	0,8	0,5
51	37	7	4	27	3	3	0,571429	0,571429	0,837838	0,571429
52	31	6	5	24	1	1	0,833333	0,833333	0,935484	0,833333
53	17	3	2	13	1	1	0,666667	0,666667	0,882353	0,666667
54	37	7	5	28	2	2	0,714286	0,714286	0,891892	0,714286
55	49	10	6	35	4	4	0,6	0,6	0,836735	0,6
56	34	7	3	23	4	4	0,428571	0,428571	0,764706	0,428571
57	38	8	7	29	1	1	0,875	0,875	0,947368	0,875
58	75	15	6	51	9	9	0,4	0,4	0,76	0,4
59	56	11	3	37	8	8	0,272727	0,272727	0,714286	0,272727
60	31	6	2	21	4	4	0,333333	0,333333	0,741935	0,333333
61	31	6	2	21	4	4	0,333333	0,333333	0,741935	0,333333
62	46	9	2	30	7	7	0,222222	0,222222	0,695652	0,222222
63	49	10	4	33	6	6	0,4	0,4	0,755102	0,4
64	38	8	3	25	5	5	0,375	0,375	0,736842	0,375
65	30	6	2	20	4	4	0,333333	0,333333	0,733333	0,333333
66	19	4	3	14	1	1	0,75	0,75	0,894737	0,75
67	17	3	0	11	3	3	0	0	0,647059	0
68	11	2	1	8	1	1	0,5	0,5	0,818182	0,5
69	24	5	2	16	3	3	0,4	0,4	0,75	0,4
70	29	6	4	21	2	2	0,666667	0,666667	0,862069	0,666667
71	14	3	1	9	2	2	0,333333	0,333333	0,714286	0,333333
72	22	4	2	16	2	2	0,5	0,5	0,818182	0,5
73	38	8	4	26	4	4	0,5	0,5	0,789474	0,5
74	32	6	4	24	2	2	0,666667	0,666667	0,875	0,666667
75	42	8	3	29	5	5	0,375	0,375	0,761905	0,375
76	21	4	1	14	3	3	0,25	0,25	0,714286	0,25
77	26	5	2	18	3	3	0,4	0,4	0,769231	0,4
78	25	5	2	17	3	3	0,4	0,4	0,76	0,4
79	34	7	3	23	4	4	0,428571	0,428571	0,764706	0,428571
80	27	5	1	18	4	4	0,2	0,2	0,703704	0,2
Rata-rata						0,512157	0,512157	0,805846	0,512157	

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

#### Lampiran 4 :

Hasil Confusion Matrix - Textrank dengan *similarity word overlapping*, dengan  
Compression Rate 30%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	6	5	13	1	1	0,833333	0,833333	0,9	0,833333
2	26	8	5	15	3	3	0,625	0,625	0,769231	0,625
3	37	11	6	21	5	5	0,545455	0,545455	0,72973	0,545455
4	41	12	3	20	9	9	0,25	0,25	0,560976	0,25
5	29	9	5	16	4	4	0,555556	0,555556	0,724138	0,555556
6	29	9	4	15	5	5	0,444444	0,444444	0,655172	0,444444
7	47	14	6	25	8	8	0,428571	0,428571	0,659574	0,428571
8	35	10	6	21	4	4	0,6	0,6	0,771429	0,6
9	31	9	6	19	3	3	0,666667	0,666667	0,806452	0,666667
10	20	6	5	13	1	1	0,833333	0,833333	0,9	0,833333
11	39	12	6	21	6	6	0,5	0,5	0,692308	0,5
12	29	9	5	16	4	4	0,555556	0,555556	0,724138	0,555556
13	38	11	7	23	4	4	0,636364	0,636364	0,789474	0,636364
14	36	11	8	22	3	3	0,727273	0,727273	0,833333	0,727273
15	20	6	3	11	3	3	0,5	0,5	0,7	0,5
16	41	12	9	26	3	3	0,75	0,75	0,853659	0,75
17	68	20	14	42	6	6	0,7	0,7	0,823529	0,7
18	64	19	13	39	6	6	0,684211	0,684211	0,8125	0,684211
19	28	8	5	17	3	3	0,625	0,625	0,785714	0,625
20	49	15	8	27	7	7	0,533333	0,533333	0,714286	0,533333
21	45	14	8	25	6	6	0,571429	0,571429	0,733333	0,571429
22	28	8	5	17	3	3	0,625	0,625	0,785714	0,625
23	23	7	5	14	2	2	0,714286	0,714286	0,826087	0,714286
24	25	8	2	11	6	6	0,25	0,25	0,52	0,25
25	31	9	6	19	3	3	0,666667	0,666667	0,806452	0,666667
26	28	8	5	17	3	3	0,625	0,625	0,785714	0,625
27	28	8	4	16	4	4	0,5	0,5	0,714286	0,5
28	30	9	4	16	5	5	0,444444	0,444444	0,666667	0,444444
29	20	6	2	10	4	4	0,333333	0,333333	0,6	0,333333
30	32	10	6	18	4	4	0,6	0,6	0,75	0,6
31	33	10	8	21	2	2	0,8	0,8	0,878788	0,8
32	28	8	5	17	3	3	0,625	0,625	0,785714	0,625
33	40	12	6	22	6	6	0,5	0,5	0,7	0,5
34	34	10	4	18	6	6	0,4	0,4	0,647059	0,4
35	24	7	5	15	2	2	0,714286	0,714286	0,833333	0,714286
36	37	11	8	23	3	3	0,727273	0,727273	0,837838	0,727273
37	25	8	5	14	3	3	0,625	0,625	0,76	0,625
38	24	7	5	15	2	2	0,714286	0,714286	0,833333	0,714286
39	36	11	7	21	4	4	0,636364	0,636364	0,777778	0,636364
40	21	6	5	14	1	1	0,833333	0,833333	0,904762	0,833333
41	32	10	7	19	3	3	0,7	0,7	0,8125	0,7
42	35	10	4	19	6	6	0,4	0,4	0,657143	0,4
43	28	8	3	15	5	5	0,375	0,375	0,642857	0,375
44	24	7	6	16	1	1	0,857143	0,857143	0,916667	0,857143
45	32	10	8	20	2	2	0,8	0,8	0,875	0,8

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	13	6	23	7	7	0,461538	0,461538	0,674419	0,461538
47	37	11	3	18	8	8	0,272727	0,272727	0,567568	0,272727
48	37	11	6	21	5	5	0,545455	0,545455	0,72973	0,545455
49	36	11	8	22	3	3	0,727273	0,727273	0,833333	0,727273
50	20	6	3	11	3	3	0,5	0,5	0,7	0,5
51	37	11	6	21	5	5	0,545455	0,545455	0,72973	0,545455
52	31	9	7	20	2	2	0,777778	0,777778	0,870968	0,777778
53	17	5	3	10	2	2	0,6	0,6	0,764706	0,6
54	37	11	6	21	5	5	0,545455	0,545455	0,72973	0,545455
55	49	15	9	28	6	6	0,6	0,6	0,755102	0,6
56	34	10	4	18	6	6	0,4	0,4	0,647059	0,4
57	38	11	9	25	2	2	0,818182	0,818182	0,894737	0,818182
58	75	22	13	44	9	9	0,590909	0,590909	0,76	0,590909
59	56	17	8	30	9	9	0,470588	0,470588	0,678571	0,470588
60	31	9	6	19	3	3	0,666667	0,666667	0,806452	0,666667
61	31	9	6	19	3	3	0,666667	0,666667	0,806452	0,666667
62	46	14	7	25	7	7	0,5	0,5	0,695652	0,5
63	49	15	9	28	6	6	0,6	0,6	0,755102	0,6
64	38	11	5	21	6	6	0,454545	0,454545	0,684211	0,454545
65	30	9	6	18	3	3	0,666667	0,666667	0,8	0,666667
66	19	6	5	12	1	1	0,833333	0,833333	0,894737	0,833333
67	17	5	2	9	3	3	0,4	0,4	0,647059	0,4
68	11	3	2	7	1	1	0,666667	0,666667	0,818182	0,666667
69	24	7	6	16	1	1	0,857143	0,857143	0,916667	0,857143
70	29	9	6	17	3	3	0,666667	0,666667	0,793103	0,666667
71	14	4	1	7	3	3	0,25	0,25	0,571429	0,25
72	22	7	3	11	4	4	0,428571	0,428571	0,636364	0,428571
73	38	11	8	24	3	3	0,727273	0,727273	0,842105	0,727273
74	32	10	6	18	4	4	0,6	0,6	0,75	0,6
75	42	13	8	24	5	5	0,615385	0,615385	0,761905	0,615385
76	21	6	2	11	4	4	0,333333	0,333333	0,619048	0,333333
77	26	8	6	16	2	2	0,75	0,75	0,846154	0,75
78	25	8	3	12	5	5	0,375	0,375	0,6	0,375
79	34	10	4	18	6	6	0,4	0,4	0,647059	0,4
80	27	8	6	17	2	2	0,75	0,75	0,851852	0,75
Rata-rata						0,58869	0,58869	0,754198	0,58869	

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 5 :

Hasil Confusion Matrix - Textrank dengan *similarity word overlapping*, dengan  
Compression Rate 40%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	8	6	10	2	2	0,75	0,75	0,8	0,75
2	26	10	8	14	2	2	0,8	0,8	0,846154	0,8
3	37	15	11	18	4	4	0,733333	0,733333	0,783784	0,733333
4	41	16	5	14	11	11	0,3125	0,3125	0,463415	0,3125
5	29	12	10	15	2	2	0,833333	0,833333	0,862069	0,833333
6	29	12	8	13	4	4	0,666667	0,666667	0,724138	0,666667
7	47	19	11	20	8	8	0,578947	0,578947	0,659574	0,578947
8	35	14	10	17	4	4	0,714286	0,714286	0,771429	0,714286
9	31	12	8	15	4	4	0,666667	0,666667	0,741935	0,666667
10	20	8	6	10	2	2	0,75	0,75	0,8	0,75
11	39	16	9	16	7	7	0,5625	0,5625	0,641026	0,5625
12	29	12	10	15	2	2	0,833333	0,833333	0,862069	0,833333
13	38	15	11	19	4	4	0,733333	0,733333	0,789474	0,733333
14	36	14	10	18	4	4	0,714286	0,714286	0,777778	0,714286
15	20	8	4	8	4	4	0,5	0,5	0,6	0,5
16	41	16	12	21	4	4	0,75	0,75	0,804878	0,75
17	68	27	21	35	6	6	0,777778	0,777778	0,823529	0,777778
18	64	26	19	31	7	7	0,730769	0,730769	0,78125	0,730769
19	28	11	8	14	3	3	0,727273	0,727273	0,785714	0,727273
20	49	20	14	23	6	6	0,7	0,7	0,755102	0,7
21	45	18	11	20	7	7	0,611111	0,611111	0,688889	0,611111
22	28	11	8	14	3	3	0,727273	0,727273	0,785714	0,727273
23	23	9	6	11	3	3	0,666667	0,666667	0,73913	0,666667
24	25	10	3	8	7	7	0,3	0,3	0,44	0,3
25	31	12	7	14	5	5	0,583333	0,583333	0,677419	0,583333
26	28	11	9	15	2	2	0,818182	0,818182	0,857143	0,818182
27	28	11	5	11	6	6	0,454545	0,454545	0,571429	0,454545
28	30	12	7	13	5	5	0,583333	0,583333	0,666667	0,583333
29	20	8	5	9	3	3	0,625	0,625	0,7	0,625
30	32	13	10	16	3	3	0,769231	0,769231	0,8125	0,769231
31	33	13	10	17	3	3	0,769231	0,769231	0,818182	0,769231
32	28	11	8	14	3	3	0,727273	0,727273	0,785714	0,727273
33	40	16	11	19	5	5	0,6875	0,6875	0,75	0,6875
34	34	14	10	16	4	4	0,714286	0,714286	0,764706	0,714286
35	24	10	7	11	3	3	0,7	0,7	0,75	0,7
36	37	15	9	16	6	6	0,6	0,6	0,675676	0,6
37	25	10	6	11	4	4	0,6	0,6	0,68	0,6
38	24	10	6	10	4	4	0,6	0,6	0,666667	0,6
39	36	14	8	16	6	6	0,571429	0,571429	0,666667	0,571429
40	21	8	5	10	3	3	0,625	0,625	0,714286	0,625
41	32	13	11	17	2	2	0,846154	0,846154	0,875	0,846154
42	35	14	8	15	6	6	0,571429	0,571429	0,657143	0,571429
43	28	11	4	10	7	7	0,363636	0,363636	0,5	0,363636
44	24	10	7	11	3	3	0,7	0,7	0,75	0,7
45	32	13	10	16	3	3	0,769231	0,769231	0,8125	0,769231

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	17	10	19	7	7	0,588235	0,588235	0,674419	0,588235
47	37	15	9	16	6	6	0,6	0,6	0,675676	0,6
48	37	15	9	16	6	6	0,6	0,6	0,675676	0,6
49	36	14	10	18	4	4	0,714286	0,714286	0,777778	0,714286
50	20	8	5	9	3	3	0,625	0,625	0,7	0,625
51	37	15	10	17	5	5	0,666667	0,666667	0,72973	0,666667
52	31	12	9	16	3	3	0,75	0,75	0,806452	0,75
53	17	7	6	9	1	1	0,857143	0,857143	0,882353	0,857143
54	37	15	11	18	4	4	0,733333	0,733333	0,783784	0,733333
55	49	20	15	24	5	5	0,75	0,75	0,795918	0,75
56	34	14	9	15	5	5	0,642857	0,642857	0,705882	0,642857
57	38	15	13	21	2	2	0,866667	0,866667	0,894737	0,866667
58	75	30	22	37	8	8	0,733333	0,733333	0,786667	0,733333
59	56	22	12	24	10	10	0,545455	0,545455	0,642857	0,545455
60	31	12	7	14	5	5	0,583333	0,583333	0,677419	0,583333
61	31	12	8	15	4	4	0,666667	0,666667	0,741935	0,666667
62	46	18	10	20	8	8	0,555556	0,555556	0,652174	0,555556
63	49	20	12	21	8	8	0,6	0,6	0,673469	0,6
64	38	15	9	17	6	6	0,6	0,6	0,684211	0,6
65	30	12	8	14	4	4	0,666667	0,666667	0,733333	0,666667
66	19	8	5	8	3	3	0,625	0,625	0,684211	0,625
67	17	7	4	7	3	3	0,571429	0,571429	0,647059	0,571429
68	11	4	2	5	2	2	0,5	0,5	0,636364	0,5
69	24	10	9	13	1	1	0,9	0,9	0,916667	0,9
70	29	12	9	14	3	3	0,75	0,75	0,793103	0,75
71	14	6	4	6	2	2	0,666667	0,666667	0,714286	0,666667
72	22	9	5	9	4	4	0,555556	0,555556	0,636364	0,555556
73	38	15	10	18	5	5	0,666667	0,666667	0,736842	0,666667
74	32	13	8	14	5	5	0,615385	0,615385	0,6875	0,615385
75	42	17	12	20	5	5	0,705882	0,705882	0,761905	0,705882
76	21	8	4	9	4	4	0,5	0,5	0,619048	0,5
77	26	10	7	13	3	3	0,7	0,7	0,769231	0,7
78	25	10	6	11	4	4	0,6	0,6	0,68	0,6
79	34	14	8	14	6	6	0,571429	0,571429	0,647059	0,571429
80	27	11	8	13	3	3	0,727273	0,727273	0,777778	0,727273
Rata-rata							0,660242	0,660242	0,728508	0,660242

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 6 :

Hasil Confusion Matrix - Textrank dengan *similarity word overlapping*, dengan  
Compression Rate 50%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	10	9	9	1	1	0,9	0,9	0,9	0,9
2	26	13	10	10	3	3	0,769231	0,769231	0,769231	0,769231
3	37	18	13	14	5	5	0,722222	0,722222	0,72973	0,722222
4	41	20	10	11	10	10	0,5	0,5	0,512195	0,5
5	29	14	11	12	3	3	0,785714	0,785714	0,793103	0,785714
6	29	14	8	9	6	6	0,571429	0,571429	0,586207	0,571429
7	47	24	17	16	7	7	0,708333	0,708333	0,702128	0,708333
8	35	18	14	13	4	4	0,777778	0,777778	0,771429	0,777778
9	31	16	10	9	6	6	0,625	0,625	0,612903	0,625
10	20	10	8	8	2	2	0,8	0,8	0,8	0,8
11	39	20	13	12	7	7	0,65	0,65	0,641026	0,65
12	29	14	11	12	3	3	0,785714	0,785714	0,793103	0,785714
13	38	19	13	13	6	6	0,684211	0,684211	0,684211	0,684211
14	36	18	16	16	2	2	0,888889	0,888889	0,888889	0,888889
15	20	10	6	6	4	4	0,6	0,6	0,6	0,6
16	41	20	16	17	4	4	0,8	0,8	0,804878	0,8
17	68	34	28	28	6	6	0,823529	0,823529	0,823529	0,823529
18	64	32	23	23	9	9	0,71875	0,71875	0,71875	0,71875
19	28	14	10	10	4	4	0,714286	0,714286	0,714286	0,714286
20	49	24	19	20	5	5	0,791667	0,791667	0,795918	0,791667
21	45	22	15	16	7	7	0,681818	0,681818	0,688889	0,681818
22	28	14	11	11	3	3	0,785714	0,785714	0,785714	0,785714
23	23	12	8	7	4	4	0,666667	0,666667	0,652174	0,666667
24	25	12	6	7	6	6	0,5	0,5	0,52	0,5
25	31	16	11	10	5	5	0,6875	0,6875	0,677419	0,6875
26	28	14	12	12	2	2	0,857143	0,857143	0,857143	0,857143
27	28	14	8	8	6	6	0,571429	0,571429	0,571429	0,571429
28	30	15	11	11	4	4	0,733333	0,733333	0,733333	0,733333
29	20	10	6	6	4	4	0,6	0,6	0,6	0,6
30	32	16	14	14	2	2	0,875	0,875	0,875	0,875
31	33	16	11	12	5	5	0,6875	0,6875	0,69697	0,6875
32	28	14	10	10	4	4	0,714286	0,714286	0,714286	0,714286
33	40	20	13	13	7	7	0,65	0,65	0,65	0,65
34	34	17	11	11	6	6	0,647059	0,647059	0,647059	0,647059
35	24	12	9	9	3	3	0,75	0,75	0,75	0,75
36	37	18	12	13	6	6	0,666667	0,666667	0,675676	0,666667
37	25	12	8	9	4	4	0,666667	0,666667	0,68	0,666667
38	24	12	7	7	5	5	0,583333	0,583333	0,583333	0,583333
39	36	18	12	12	6	6	0,666667	0,666667	0,666667	0,666667
40	21	10	6	7	4	4	0,6	0,6	0,619048	0,6
41	32	16	12	12	4	4	0,75	0,75	0,75	0,75
42	35	18	11	10	7	7	0,611111	0,611111	0,6	0,611111
43	28	14	5	5	9	9	0,357143	0,357143	0,357143	0,357143
44	24	12	9	9	3	3	0,75	0,75	0,75	0,75
45	32	16	12	12	4	4	0,75	0,75	0,75	0,75

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	22	13	12	9	9	0,590909	0,590909	0,581395	0,590909
47	37	18	13	14	5	5	0,722222	0,722222	0,72973	0,722222
48	37	18	13	14	5	5	0,722222	0,722222	0,72973	0,722222
49	36	18	13	13	5	5	0,722222	0,722222	0,722222	0,722222
50	20	10	8	8	2	2	0,8	0,8	0,8	0,8
51	37	18	13	14	5	5	0,722222	0,722222	0,72973	0,722222
52	31	16	14	13	2	2	0,875	0,875	0,870968	0,875
53	17	8	6	7	2	2	0,75	0,75	0,764706	0,75
54	37	18	12	13	6	6	0,666667	0,666667	0,675676	0,666667
55	49	24	15	16	9	9	0,625	0,625	0,632653	0,625
56	34	17	11	11	6	6	0,647059	0,647059	0,647059	0,647059
57	38	19	16	16	3	3	0,842105	0,842105	0,842105	0,842105
58	75	38	28	27	10	10	0,736842	0,736842	0,733333	0,736842
59	56	28	16	16	12	12	0,571429	0,571429	0,571429	0,571429
60	31	16	12	11	4	4	0,75	0,75	0,741935	0,75
61	31	16	13	12	3	3	0,8125	0,8125	0,806452	0,8125
62	46	23	16	16	7	7	0,695652	0,695652	0,695652	0,695652
63	49	24	15	16	9	9	0,625	0,625	0,632653	0,625
64	38	19	12	12	7	7	0,631579	0,631579	0,631579	0,631579
65	30	15	10	10	5	5	0,666667	0,666667	0,666667	0,666667
66	19	10	7	6	3	3	0,7	0,7	0,684211	0,7
67	17	8	5	6	3	3	0,625	0,625	0,647059	0,625
68	11	6	4	3	2	2	0,666667	0,666667	0,636364	0,666667
69	24	12	10	10	2	2	0,833333	0,833333	0,833333	0,833333
70	29	14	11	12	3	3	0,785714	0,785714	0,793103	0,785714
71	14	7	5	5	2	2	0,714286	0,714286	0,714286	0,714286
72	22	11	6	6	5	5	0,545455	0,545455	0,545455	0,545455
73	38	19	14	14	5	5	0,736842	0,736842	0,736842	0,736842
74	32	16	12	12	4	4	0,75	0,75	0,75	0,75
75	42	21	17	17	4	4	0,809524	0,809524	0,809524	0,809524
76	21	10	7	8	3	3	0,7	0,7	0,714286	0,7
77	26	13	10	10	3	3	0,769231	0,769231	0,769231	0,769231
78	25	12	8	9	4	4	0,666667	0,666667	0,68	0,666667
79	34	17	13	13	4	4	0,764706	0,764706	0,764706	0,764706
80	27	14	11	10	3	3	0,785714	0,785714	0,777778	0,785714
Rata-rata						0,705678	0,705678	0,706933	0,705678	

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 7 :

Hasil Confusion Matrix - Textrank dengan Jaccard *similarity*, dengan

Compression Rate 10%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	2	0	16	2	2	0	0	0,8	0
2	26	3	1	21	2	2	0,333333	0,333333	0,846154	0,333333
3	37	4	0	29	4	4	0	0	0,783784	0
4	41	4	1	34	3	3	0,25	0,25	0,853659	0,25
5	29	3	1	24	2	2	0,333333	0,333333	0,862069	0,333333
6	29	3	1	24	2	2	0,333333	0,333333	0,862069	0,333333
7	47	5	0	37	5	5	0	0	0,787234	0
8	35	4	2	29	2	2	0,5	0,5	0,885714	0,5
9	31	3	2	27	1	1	0,666667	0,666667	0,935484	0,666667
10	20	2	0	16	2	2	0	0	0,8	0
11	39	4	0	31	4	4	0	0	0,794872	0
12	29	3	0	23	3	3	0	0	0,793103	0
13	38	4	1	31	3	3	0,25	0,25	0,842105	0,25
14	36	4	0	28	4	4	0	0	0,777778	0
15	20	2	0	16	2	2	0	0	0,8	0
16	41	4	1	34	3	3	0,25	0,25	0,853659	0,25
17	68	7	3	57	4	4	0,428571	0,428571	0,882353	0,428571
18	64	6	3	55	3	3	0,5	0,5	0,90625	0,5
19	28	3	0	22	3	3	0	0	0,785714	0
20	49	5	1	40	4	4	0,2	0,2	0,836735	0,2
21	45	4	1	38	3	3	0,25	0,25	0,866667	0,25
22	28	3	1	23	2	2	0,333333	0,333333	0,857143	0,333333
23	23	2	0	19	2	2	0	0	0,826087	0
24	25	2	0	21	2	2	0	0	0,84	0
25	31	3	0	25	3	3	0	0	0,806452	0
26	28	3	0	22	3	3	0	0	0,785714	0
27	28	3	1	23	2	2	0,333333	0,333333	0,857143	0,333333
28	30	3	1	25	2	2	0,333333	0,333333	0,866667	0,333333
29	20	2	0	16	2	2	0	0	0,8	0
30	32	3	2	28	1	1	0,666667	0,666667	0,9375	0,666667
31	33	3	1	28	2	2	0,333333	0,333333	0,878788	0,333333
32	28	3	1	23	2	2	0,333333	0,333333	0,857143	0,333333
33	40	4	1	33	3	3	0,25	0,25	0,85	0,25
34	34	3	1	29	2	2	0,333333	0,333333	0,882353	0,333333
35	24	2	0	20	2	2	0	0	0,833333	0
36	37	4	3	32	1	1	0,75	0,75	0,945946	0,75
37	25	2	0	21	2	2	0	0	0,84	0
38	24	2	0	20	2	2	0	0	0,833333	0
39	36	4	1	29	3	3	0,25	0,25	0,833333	0,25
40	21	2	0	17	2	2	0	0	0,809524	0
41	32	3	0	26	3	3	0	0	0,8125	0
42	35	4	1	28	3	3	0,25	0,25	0,828571	0,25
43	28	3	0	22	3	3	0	0	0,785714	0
44	24	2	2	22	0	0	1	1	1	1
45	32	3	1	27	2	2	0,333333	0,333333	0,875	0,333333

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	4	1	36	3	3	0,25	0,25	0,860465	0,25
47	37	4	0	29	4	4	0	0	0,783784	0
48	37	4	1	30	3	3	0,25	0,25	0,837838	0,25
49	36	4	1	29	3	3	0,25	0,25	0,833333	0,25
50	20	2	0	16	2	2	0	0	0,8	0
51	37	4	2	31	2	2	0,5	0,5	0,891892	0,5
52	31	3	0	25	3	3	0	0	0,806452	0
53	17	2	2	15	0	0	1	1	1	1
54	37	4	1	30	3	3	0,25	0,25	0,837838	0,25
55	49	5	2	41	3	3	0,4	0,4	0,877551	0,4
56	34	3	1	29	2	2	0,333333	0,333333	0,882353	0,333333
57	38	4	3	33	1	1	0,75	0,75	0,947368	0,75
58	75	8	1	60	7	7	0,125	0,125	0,813333	0,125
59	56	6	0	44	6	6	0	0	0,785714	0
60	31	3	2	27	1	1	0,666667	0,666667	0,935484	0,666667
61	31	3	1	26	2	2	0,333333	0,333333	0,870968	0,333333
62	46	5	1	37	4	4	0,2	0,2	0,826087	0,2
63	49	5	3	42	2	2	0,6	0,6	0,918367	0,6
64	38	4	1	31	3	3	0,25	0,25	0,842105	0,25
65	30	3	1	25	2	2	0,333333	0,333333	0,866667	0,333333
66	19	2	1	16	1	1	0,5	0,5	0,894737	0,5
67	17	2	0	13	2	2	0	0	0,764706	0
68	11	1	1	10	0	0	1	1	1	1
69	24	2	1	21	1	1	0,5	0,5	0,916667	0,5
70	29	3	1	24	2	2	0,333333	0,333333	0,862069	0,333333
71	14	1	0	12	1	1	0	0	0,857143	0
72	22	2	1	19	1	1	0,5	0,5	0,909091	0,5
73	38	4	1	31	3	3	0,25	0,25	0,842105	0,25
74	32	3	1	27	2	2	0,333333	0,333333	0,875	0,333333
75	42	4	3	37	1	1	0,75	0,75	0,952381	0,75
76	21	2	0	17	2	2	0	0	0,809524	0
77	26	3	0	20	3	3	0	0	0,769231	0
78	25	2	0	21	2	2	0	0	0,84	0
79	34	3	0	28	3	3	0	0	0,823529	0
80	27	3	1	22	2	2	0,333333	0,333333	0,851852	0,333333
Rata-rata							0,259836	0,259836	0,852291	0,259836

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 8 :

Hasil Confusion Matrix - Textrank dengan Jaccard *similarity*, dengan

Compression Rate 20%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	4	2	14	2	2	0,5	0,5	0,8	0,5
2	26	5	2	18	3	3	0,4	0,4	0,769231	0,4
3	37	7	3	26	4	4	0,428571	0,428571	0,783784	0,428571
4	41	8	2	27	6	6	0,25	0,25	0,707317	0,25
5	29	6	2	19	4	4	0,333333	0,333333	0,724138	0,333333
6	29	6	2	19	4	4	0,333333	0,333333	0,724138	0,333333
7	47	9	2	31	7	7	0,222222	0,222222	0,702128	0,222222
8	35	7	4	25	3	3	0,571429	0,571429	0,828571	0,571429
9	31	6	2	21	4	4	0,333333	0,333333	0,741935	0,333333
10	20	4	2	14	2	2	0,5	0,5	0,8	0,5
11	39	8	5	28	3	3	0,625	0,625	0,846154	0,625
12	29	6	1	18	5	5	0,166667	0,166667	0,655172	0,166667
13	38	8	4	26	4	4	0,5	0,5	0,789474	0,5
14	36	7	3	25	4	4	0,428571	0,428571	0,777778	0,428571
15	20	4	1	13	3	3	0,25	0,25	0,7	0,25
16	41	8	5	30	3	3	0,625	0,625	0,853659	0,625
17	68	14	11	51	3	3	0,785714	0,785714	0,911765	0,785714
18	64	13	9	47	4	4	0,692308	0,692308	0,875	0,692308
19	28	6	4	20	2	2	0,666667	0,666667	0,857143	0,666667
20	49	10	5	34	5	5	0,5	0,5	0,795918	0,5
21	45	9	6	33	3	3	0,666667	0,666667	0,866667	0,666667
22	28	6	4	20	2	2	0,666667	0,666667	0,857143	0,666667
23	23	5	1	14	4	4	0,2	0,2	0,652174	0,2
24	25	5	2	17	3	3	0,4	0,4	0,76	0,4
25	31	6	2	21	4	4	0,333333	0,333333	0,741935	0,333333
26	28	6	2	18	4	4	0,333333	0,333333	0,714286	0,333333
27	28	6	2	18	4	4	0,333333	0,333333	0,714286	0,333333
28	30	6	2	20	4	4	0,333333	0,333333	0,733333	0,333333
29	20	4	1	13	3	3	0,25	0,25	0,7	0,25
30	32	6	4	24	2	2	0,666667	0,666667	0,875	0,666667
31	33	7	4	23	3	3	0,571429	0,571429	0,818182	0,571429
32	28	6	4	20	2	2	0,666667	0,666667	0,857143	0,666667
33	40	8	5	29	3	3	0,625	0,625	0,85	0,625
34	34	7	2	22	5	5	0,285714	0,285714	0,705882	0,285714
35	24	5	3	17	2	2	0,6	0,6	0,833333	0,6
36	37	7	4	27	3	3	0,571429	0,571429	0,837838	0,571429
37	25	5	3	18	2	2	0,6	0,6	0,84	0,6
38	24	5	2	16	3	3	0,4	0,4	0,75	0,4
39	36	7	3	25	4	4	0,428571	0,428571	0,777778	0,428571
40	21	4	2	15	2	2	0,5	0,5	0,809524	0,5
41	32	6	2	22	4	4	0,333333	0,333333	0,75	0,333333
42	35	7	3	24	4	4	0,428571	0,428571	0,771429	0,428571
43	28	6	2	18	4	4	0,333333	0,333333	0,714286	0,333333
44	24	5	4	18	1	1	0,8	0,8	0,916667	0,8
45	32	6	4	24	2	2	0,666667	0,666667	0,875	0,666667

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	9	3	28	6	6	0,333333	0,333333	0,72093	0,333333
47	37	7	1	24	6	6	0,142857	0,142857	0,675676	0,142857
48	37	7	4	27	3	3	0,571429	0,571429	0,837838	0,571429
49	36	7	3	25	4	4	0,428571	0,428571	0,777778	0,428571
50	20	4	0	12	4	4	0	0	0,6	0
51	37	7	3	26	4	4	0,428571	0,428571	0,783784	0,428571
52	31	6	4	23	2	2	0,666667	0,666667	0,870968	0,666667
53	17	3	2	13	1	1	0,666667	0,666667	0,882353	0,666667
54	37	7	3	26	4	4	0,428571	0,428571	0,783784	0,428571
55	49	10	2	31	8	8	0,2	0,2	0,673469	0,2
56	34	7	2	22	5	5	0,285714	0,285714	0,705882	0,285714
57	38	8	6	28	2	2	0,75	0,75	0,894737	0,75
58	75	15	8	53	7	7	0,533333	0,533333	0,813333	0,533333
59	56	11	2	36	9	9	0,181818	0,181818	0,678571	0,181818
60	31	6	2	21	4	4	0,333333	0,333333	0,741935	0,333333
61	31	6	1	20	5	5	0,166667	0,166667	0,677419	0,166667
62	46	9	3	31	6	6	0,333333	0,333333	0,73913	0,333333
63	49	10	5	34	5	5	0,5	0,5	0,795918	0,5
64	38	8	3	25	5	5	0,375	0,375	0,736842	0,375
65	30	6	2	20	4	4	0,333333	0,333333	0,733333	0,333333
66	19	4	3	14	1	1	0,75	0,75	0,894737	0,75
67	17	3	0	11	3	3	0	0	0,647059	0
68	11	2	1	8	1	1	0,5	0,5	0,818182	0,5
69	24	5	3	17	2	2	0,6	0,6	0,833333	0,6
70	29	6	4	21	2	2	0,666667	0,666667	0,862069	0,666667
71	14	3	0	8	3	3	0	0	0,571429	0
72	22	4	2	16	2	2	0,5	0,5	0,818182	0,5
73	38	8	5	27	3	3	0,625	0,625	0,842105	0,625
74	32	6	3	23	3	3	0,5	0,5	0,8125	0,5
75	42	8	4	30	4	4	0,5	0,5	0,809524	0,5
76	21	4	2	15	2	2	0,5	0,5	0,809524	0,5
77	26	5	2	18	3	3	0,4	0,4	0,769231	0,4
78	25	5	2	17	3	3	0,4	0,4	0,76	0,4
79	34	7	2	22	5	5	0,285714	0,285714	0,705882	0,285714
80	27	5	2	19	3	3	0,4	0,4	0,777778	0,4
Rata-rata						0,44241	0,44241	0,777755	0,44241	

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 9 :

Hasil Confusion Matrix - Textrank dengan Jaccard *similarity*, dengan

Compression Rate 30%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	6	4	12	2	2	0,666667	0,666667	0,8	0,666667
2	26	8	5	15	3	3	0,625	0,625	0,769231	0,625
3	37	11	5	20	6	6	0,454545	0,454545	0,675676	0,454545
4	41	12	3	20	9	9	0,25	0,25	0,560976	0,25
5	29	9	3	14	6	6	0,333333	0,333333	0,586207	0,333333
6	29	9	4	15	5	5	0,444444	0,444444	0,655172	0,444444
7	47	14	6	25	8	8	0,428571	0,428571	0,659574	0,428571
8	35	10	6	21	4	4	0,6	0,6	0,771429	0,6
9	31	9	6	19	3	3	0,666667	0,666667	0,806452	0,666667
10	20	6	4	12	2	2	0,666667	0,666667	0,8	0,666667
11	39	12	6	21	6	6	0,5	0,5	0,692308	0,5
12	29	9	6	17	3	3	0,666667	0,666667	0,793103	0,666667
13	38	11	6	22	5	5	0,545455	0,545455	0,736842	0,545455
14	36	11	7	21	4	4	0,636364	0,636364	0,777778	0,636364
15	20	6	2	10	4	4	0,333333	0,333333	0,6	0,333333
16	41	12	8	25	4	4	0,666667	0,666667	0,804878	0,666667
17	68	20	14	42	6	6	0,7	0,7	0,823529	0,7
18	64	19	14	40	5	5	0,736842	0,736842	0,84375	0,736842
19	28	8	4	16	4	4	0,5	0,5	0,714286	0,5
20	49	15	8	27	7	7	0,533333	0,533333	0,714286	0,533333
21	45	14	8	25	6	6	0,571429	0,571429	0,733333	0,571429
22	28	8	6	18	2	2	0,75	0,75	0,857143	0,75
23	23	7	3	12	4	4	0,428571	0,428571	0,652174	0,428571
24	25	8	2	11	6	6	0,25	0,25	0,52	0,25
25	31	9	6	19	3	3	0,666667	0,666667	0,806452	0,666667
26	28	8	6	18	2	2	0,75	0,75	0,857143	0,75
27	28	8	5	17	3	3	0,625	0,625	0,785714	0,625
28	30	9	6	18	3	3	0,666667	0,666667	0,8	0,666667
29	20	6	1	9	5	5	0,166667	0,166667	0,5	0,166667
30	32	10	6	18	4	4	0,6	0,6	0,75	0,6
31	33	10	7	20	3	3	0,7	0,7	0,818182	0,7
32	28	8	5	17	3	3	0,625	0,625	0,785714	0,625
33	40	12	6	22	6	6	0,5	0,5	0,7	0,5
34	34	10	4	18	6	6	0,4	0,4	0,647059	0,4
35	24	7	4	14	3	3	0,571429	0,571429	0,75	0,571429
36	37	11	7	22	4	4	0,636364	0,636364	0,783784	0,636364
37	25	8	4	13	4	4	0,5	0,5	0,68	0,5
38	24	7	4	14	3	3	0,571429	0,571429	0,75	0,571429
39	36	11	7	21	4	4	0,636364	0,636364	0,777778	0,636364
40	21	6	3	12	3	3	0,5	0,5	0,714286	0,5
41	32	10	7	19	3	3	0,7	0,7	0,8125	0,7
42	35	10	4	19	6	6	0,4	0,4	0,657143	0,4
43	28	8	3	15	5	5	0,375	0,375	0,642857	0,375
44	24	7	5	15	2	2	0,714286	0,714286	0,833333	0,714286
45	32	10	8	20	2	2	0,8	0,8	0,875	0,8

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	13	7	24	6	6	0,538462	0,538462	0,72093	0,538462
47	37	11	4	19	7	7	0,363636	0,363636	0,621622	0,363636
48	37	11	6	21	5	5	0,545455	0,545455	0,72973	0,545455
49	36	11	7	21	4	4	0,636364	0,636364	0,777778	0,636364
50	20	6	2	10	4	4	0,333333	0,333333	0,6	0,333333
51	37	11	6	21	5	5	0,545455	0,545455	0,72973	0,545455
52	31	9	8	21	1	1	0,888889	0,888889	0,935484	0,888889
53	17	5	3	10	2	2	0,6	0,6	0,764706	0,6
54	37	11	7	22	4	4	0,636364	0,636364	0,783784	0,636364
55	49	15	9	28	6	6	0,6	0,6	0,755102	0,6
56	34	10	4	18	6	6	0,4	0,4	0,647059	0,4
57	38	11	9	25	2	2	0,818182	0,818182	0,894737	0,818182
58	75	22	12	43	10	10	0,545455	0,545455	0,733333	0,545455
59	56	17	7	29	10	10	0,411765	0,411765	0,642857	0,411765
60	31	9	6	19	3	3	0,666667	0,666667	0,806452	0,666667
61	31	9	5	18	4	4	0,555556	0,555556	0,741935	0,555556
62	46	14	7	25	7	7	0,5	0,5	0,695652	0,5
63	49	15	7	26	8	8	0,466667	0,466667	0,673469	0,466667
64	38	11	4	20	7	7	0,363636	0,363636	0,631579	0,363636
65	30	9	5	17	4	4	0,555556	0,555556	0,733333	0,555556
66	19	6	5	12	1	1	0,833333	0,833333	0,894737	0,833333
67	17	5	1	8	4	4	0,2	0,2	0,529412	0,2
68	11	3	2	7	1	1	0,666667	0,666667	0,818182	0,666667
69	24	7	5	15	2	2	0,714286	0,714286	0,833333	0,714286
70	29	9	6	17	3	3	0,666667	0,666667	0,793103	0,666667
71	14	4	1	7	3	3	0,25	0,25	0,571429	0,25
72	22	7	3	11	4	4	0,428571	0,428571	0,636364	0,428571
73	38	11	7	23	4	4	0,636364	0,636364	0,789474	0,636364
74	32	10	7	19	3	3	0,7	0,7	0,8125	0,7
75	42	13	7	23	6	6	0,538462	0,538462	0,714286	0,538462
76	21	6	2	11	4	4	0,333333	0,333333	0,619048	0,333333
77	26	8	5	15	3	3	0,625	0,625	0,769231	0,625
78	25	8	4	13	4	4	0,5	0,5	0,68	0,5
79	34	10	4	18	6	6	0,4	0,4	0,647059	0,4
80	27	8	6	17	2	2	0,75	0,75	0,851852	0,75
Rata-rata						0,553419	0,553419	0,733179	0,553419	

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 10 :

Hasil Confusion Matrix - Textrank dengan Jaccard *similarity*, dengan

Compression Rate 40%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	8	5	9	3	3	0,625	0,625	0,7	0,625
2	26	10	8	14	2	2	0,8	0,8	0,846154	0,8
3	37	15	10	17	5	5	0,666667	0,666667	0,72973	0,666667
4	41	16	6	15	10	10	0,375	0,375	0,512195	0,375
5	29	12	6	11	6	6	0,5	0,5	0,586207	0,5
6	29	12	6	11	6	6	0,5	0,5	0,586207	0,5
7	47	19	13	22	6	6	0,684211	0,684211	0,744681	0,684211
8	35	14	9	16	5	5	0,642857	0,642857	0,714286	0,642857
9	31	12	8	15	4	4	0,666667	0,666667	0,741935	0,666667
10	20	8	6	10	2	2	0,75	0,75	0,8	0,75
11	39	16	7	14	9	9	0,4375	0,4375	0,538462	0,4375
12	29	12	9	14	3	3	0,75	0,75	0,793103	0,75
13	38	15	11	19	4	4	0,733333	0,733333	0,789474	0,733333
14	36	14	10	18	4	4	0,714286	0,714286	0,777778	0,714286
15	20	8	3	7	5	5	0,375	0,375	0,5	0,375
16	41	16	11	20	5	5	0,6875	0,6875	0,756098	0,6875
17	68	27	18	32	9	9	0,666667	0,666667	0,735294	0,666667
18	64	26	18	30	8	8	0,692308	0,692308	0,75	0,692308
19	28	11	9	15	2	2	0,818182	0,818182	0,857143	0,818182
20	49	20	14	23	6	6	0,7	0,7	0,755102	0,7
21	45	18	11	20	7	7	0,611111	0,611111	0,688889	0,611111
22	28	11	8	14	3	3	0,727273	0,727273	0,785714	0,727273
23	23	9	5	10	4	4	0,555556	0,555556	0,652174	0,555556
24	25	10	3	8	7	7	0,3	0,3	0,44	0,3
25	31	12	8	15	4	4	0,666667	0,666667	0,741935	0,666667
26	28	11	8	14	3	3	0,727273	0,727273	0,785714	0,727273
27	28	11	8	14	3	3	0,727273	0,727273	0,785714	0,727273
28	30	12	7	13	5	5	0,583333	0,583333	0,666667	0,583333
29	20	8	4	8	4	4	0,5	0,5	0,6	0,5
30	32	13	10	16	3	3	0,769231	0,769231	0,8125	0,769231
31	33	13	9	16	4	4	0,692308	0,692308	0,757576	0,692308
32	28	11	8	14	3	3	0,727273	0,727273	0,785714	0,727273
33	40	16	10	18	6	6	0,625	0,625	0,7	0,625
34	34	14	9	15	5	5	0,642857	0,642857	0,705882	0,642857
35	24	10	7	11	3	3	0,7	0,7	0,75	0,7
36	37	15	9	16	6	6	0,6	0,6	0,675676	0,6
37	25	10	5	10	5	5	0,5	0,5	0,6	0,5
38	24	10	6	10	4	4	0,6	0,6	0,666667	0,6
39	36	14	7	15	7	7	0,5	0,5	0,611111	0,5
40	21	8	4	9	4	4	0,5	0,5	0,619048	0,5
41	32	13	11	17	2	2	0,846154	0,846154	0,875	0,846154
42	35	14	8	15	6	6	0,571429	0,571429	0,657143	0,571429
43	28	11	4	10	7	7	0,363636	0,363636	0,5	0,363636
44	24	10	6	10	4	4	0,6	0,6	0,666667	0,6
45	32	13	10	16	3	3	0,769231	0,769231	0,8125	0,769231

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	17	11	20	6	6	0,647059	0,647059	0,72093	0,647059
47	37	15	10	17	5	5	0,666667	0,666667	0,72973	0,666667
48	37	15	9	16	6	6	0,6	0,6	0,675676	0,6
49	36	14	9	17	5	5	0,642857	0,642857	0,722222	0,642857
50	20	8	5	9	3	3	0,625	0,625	0,7	0,625
51	37	15	10	17	5	5	0,666667	0,666667	0,72973	0,666667
52	31	12	9	16	3	3	0,75	0,75	0,806452	0,75
53	17	7	6	9	1	1	0,857143	0,857143	0,882353	0,857143
54	37	15	10	17	5	5	0,666667	0,666667	0,72973	0,666667
55	49	20	12	21	8	8	0,6	0,6	0,673469	0,6
56	34	14	9	15	5	5	0,642857	0,642857	0,705882	0,642857
57	38	15	11	19	4	4	0,733333	0,733333	0,789474	0,733333
58	75	30	19	34	11	11	0,633333	0,633333	0,706667	0,633333
59	56	22	11	23	11	11	0,5	0,5	0,607143	0,5
60	31	12	7	14	5	5	0,583333	0,583333	0,677419	0,583333
61	31	12	7	14	5	5	0,583333	0,583333	0,677419	0,583333
62	46	18	10	20	8	8	0,555556	0,555556	0,652174	0,555556
63	49	20	11	20	9	9	0,55	0,55	0,632653	0,55
64	38	15	9	17	6	6	0,6	0,6	0,684211	0,6
65	30	12	7	13	5	5	0,583333	0,583333	0,666667	0,583333
66	19	8	6	9	2	2	0,75	0,75	0,789474	0,75
67	17	7	4	7	3	3	0,571429	0,571429	0,647059	0,571429
68	11	4	2	5	2	2	0,5	0,5	0,636364	0,5
69	24	10	8	12	2	2	0,8	0,8	0,833333	0,8
70	29	12	9	14	3	3	0,75	0,75	0,793103	0,75
71	14	6	3	5	3	3	0,5	0,5	0,571429	0,5
72	22	9	6	10	3	3	0,666667	0,666667	0,727273	0,666667
73	38	15	9	17	6	6	0,6	0,6	0,684211	0,6
74	32	13	9	15	4	4	0,692308	0,692308	0,75	0,692308
75	42	17	11	19	6	6	0,647059	0,647059	0,714286	0,647059
76	21	8	5	10	3	3	0,625	0,625	0,714286	0,625
77	26	10	7	13	3	3	0,7	0,7	0,769231	0,7
78	25	10	6	11	4	4	0,6	0,6	0,68	0,6
79	34	14	8	14	6	6	0,571429	0,571429	0,647059	0,571429
80	27	11	7	12	4	4	0,636364	0,636364	0,703704	0,636364
Rata-rata						0,631102	0,631102	0,704837	0,631102	

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 11 :

Hasil Confusion Matrix - Textrank dengan Jaccard *similarity*, dengan

Compression Rate 50%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	10	8	8	2	2	0,8	0,8	0,8	0,8
2	26	13	11	11	2	2	0,846154	0,846154	0,846154	0,846154
3	37	18	13	14	5	5	0,722222	0,722222	0,72973	0,722222
4	41	20	11	12	9	9	0,55	0,55	0,560976	0,55
5	29	14	9	10	5	5	0,642857	0,642857	0,655172	0,642857
6	29	14	8	9	6	6	0,571429	0,571429	0,586207	0,571429
7	47	24	19	18	5	5	0,791667	0,791667	0,787234	0,791667
8	35	18	13	12	5	5	0,722222	0,722222	0,714286	0,722222
9	31	16	10	9	6	6	0,625	0,625	0,612903	0,625
10	20	10	7	7	3	3	0,7	0,7	0,7	0,7
11	39	20	13	12	7	7	0,65	0,65	0,641026	0,65
12	29	14	11	12	3	3	0,785714	0,785714	0,793103	0,785714
13	38	19	12	12	7	7	0,631579	0,631579	0,631579	0,631579
14	36	18	16	16	2	2	0,888889	0,888889	0,888889	0,888889
15	20	10	5	5	5	5	0,5	0,5	0,5	0,5
16	41	20	14	15	6	6	0,7	0,7	0,707317	0,7
17	68	34	28	28	6	6	0,823529	0,823529	0,823529	0,823529
18	64	32	21	21	11	11	0,65625	0,65625	0,65625	0,65625
19	28	14	10	10	4	4	0,714286	0,714286	0,714286	0,714286
20	49	24	18	19	6	6	0,75	0,75	0,755102	0,75
21	45	22	15	16	7	7	0,681818	0,681818	0,688889	0,681818
22	28	14	11	11	3	3	0,785714	0,785714	0,785714	0,785714
23	23	12	8	7	4	4	0,666667	0,666667	0,652174	0,666667
24	25	12	6	7	6	6	0,5	0,5	0,52	0,5
25	31	16	11	10	5	5	0,6875	0,6875	0,677419	0,6875
26	28	14	12	12	2	2	0,857143	0,857143	0,857143	0,857143
27	28	14	9	9	5	5	0,642857	0,642857	0,642857	0,642857
28	30	15	11	11	4	4	0,733333	0,733333	0,733333	0,733333
29	20	10	5	5	5	5	0,5	0,5	0,5	0,5
30	32	16	14	14	2	2	0,875	0,875	0,875	0,875
31	33	16	11	12	5	5	0,6875	0,6875	0,69697	0,6875
32	28	14	10	10	4	4	0,714286	0,714286	0,714286	0,714286
33	40	20	13	13	7	7	0,65	0,65	0,65	0,65
34	34	17	12	12	5	5	0,705882	0,705882	0,705882	0,705882
35	24	12	9	9	3	3	0,75	0,75	0,75	0,75
36	37	18	12	13	6	6	0,666667	0,666667	0,675676	0,666667
37	25	12	7	8	5	5	0,583333	0,583333	0,6	0,583333
38	24	12	7	7	5	5	0,583333	0,583333	0,583333	0,583333
39	36	18	11	11	7	7	0,611111	0,611111	0,611111	0,611111
40	21	10	7	8	3	3	0,7	0,7	0,714286	0,7
41	32	16	13	13	3	3	0,8125	0,8125	0,8125	0,8125
42	35	18	10	9	8	8	0,555556	0,555556	0,542857	0,555556
43	28	14	5	5	9	9	0,357143	0,357143	0,357143	0,357143
44	24	12	7	7	5	5	0,583333	0,583333	0,583333	0,583333
45	32	16	12	12	4	4	0,75	0,75	0,75	0,75

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	22	13	12	9	9	0,590909	0,590909	0,581395	0,590909
47	37	18	14	15	4	4	0,777778	0,777778	0,783784	0,777778
48	37	18	13	14	5	5	0,722222	0,722222	0,72973	0,722222
49	36	18	12	12	6	6	0,666667	0,666667	0,666667	0,666667
50	20	10	6	6	4	4	0,6	0,6	0,6	0,6
51	37	18	12	13	6	6	0,666667	0,666667	0,675676	0,666667
52	31	16	14	13	2	2	0,875	0,875	0,870968	0,875
53	17	8	6	7	2	2	0,75	0,75	0,764706	0,75
54	37	18	13	14	5	5	0,722222	0,722222	0,72973	0,722222
55	49	24	15	16	9	9	0,625	0,625	0,632653	0,625
56	34	17	11	11	6	6	0,647059	0,647059	0,647059	0,647059
57	38	19	16	16	3	3	0,842105	0,842105	0,842105	0,842105
58	75	38	28	27	10	10	0,736842	0,736842	0,733333	0,736842
59	56	28	16	16	12	12	0,571429	0,571429	0,571429	0,571429
60	31	16	12	11	4	4	0,75	0,75	0,741935	0,75
61	31	16	11	10	5	5	0,6875	0,6875	0,677419	0,6875
62	46	23	16	16	7	7	0,695652	0,695652	0,695652	0,695652
63	49	24	14	15	10	10	0,583333	0,583333	0,591837	0,583333
64	38	19	13	13	6	6	0,684211	0,684211	0,684211	0,684211
65	30	15	9	9	6	6	0,6	0,6	0,6	0,6
66	19	10	7	6	3	3	0,7	0,7	0,684211	0,7
67	17	8	5	6	3	3	0,625	0,625	0,647059	0,625
68	11	6	4	3	2	2	0,666667	0,666667	0,636364	0,666667
69	24	12	10	10	2	2	0,833333	0,833333	0,833333	0,833333
70	29	14	11	12	3	3	0,785714	0,785714	0,793103	0,785714
71	14	7	4	4	3	3	0,571429	0,571429	0,571429	0,571429
72	22	11	8	8	3	3	0,727273	0,727273	0,727273	0,727273
73	38	19	14	14	5	5	0,736842	0,736842	0,736842	0,736842
74	32	16	12	12	4	4	0,75	0,75	0,75	0,75
75	42	21	16	16	5	5	0,761905	0,761905	0,761905	0,761905
76	21	10	7	8	3	3	0,7	0,7	0,714286	0,7
77	26	13	10	10	3	3	0,769231	0,769231	0,769231	0,769231
78	25	12	8	9	4	4	0,666667	0,666667	0,68	0,666667
79	34	17	13	13	4	4	0,764706	0,764706	0,764706	0,764706
80	27	14	10	9	4	4	0,714286	0,714286	0,703704	0,714286
Rata-rata						0,691002	0,691002	0,692217	0,691002	

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 12 :

Hasil Confusion Matrix - Textrank dengan BoW Cosine *similarity*, dengan  
Compression Rate 10%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	2	1	17	1	1	0,5	0,5	0,9	0,5
2	26	3	0	20	3	3	0	0	0,769231	0
3	37	4	1	30	3	3	0,25	0,25	0,837838	0,25
4	41	4	1	34	3	3	0,25	0,25	0,853659	0,25
5	29	3	1	24	2	2	0,333333	0,333333	0,862069	0,333333
6	29	3	2	25	1	1	0,666667	0,666667	0,931034	0,666667
7	47	5	0	37	5	5	0	0	0,787234	0
8	35	4	2	29	2	2	0,5	0,5	0,885714	0,5
9	31	3	2	27	1	1	0,666667	0,666667	0,935484	0,666667
10	20	2	1	17	1	1	0,5	0,5	0,9	0,5
11	39	4	1	32	3	3	0,25	0,25	0,846154	0,25
12	29	3	0	23	3	3	0	0	0,793103	0
13	38	4	1	31	3	3	0,25	0,25	0,842105	0,25
14	36	4	1	29	3	3	0,25	0,25	0,833333	0,25
15	20	2	1	17	1	1	0,5	0,5	0,9	0,5
16	41	4	1	34	3	3	0,25	0,25	0,853659	0,25
17	68	7	1	55	6	6	0,142857	0,142857	0,823529	0,142857
18	64	6	4	56	2	2	0,666667	0,666667	0,9375	0,666667
19	28	3	0	22	3	3	0	0	0,785714	0
20	49	5	1	40	4	4	0,2	0,2	0,836735	0,2
21	45	4	1	38	3	3	0,25	0,25	0,866667	0,25
22	28	3	2	24	1	1	0,666667	0,666667	0,928571	0,666667
23	23	2	0	19	2	2	0	0	0,826087	0
24	25	2	1	22	1	1	0,5	0,5	0,92	0,5
25	31	3	0	25	3	3	0	0	0,806452	0
26	28	3	0	22	3	3	0	0	0,785714	0
27	28	3	1	23	2	2	0,333333	0,333333	0,857143	0,333333
28	30	3	2	26	1	1	0,666667	0,666667	0,933333	0,666667
29	20	2	0	16	2	2	0	0	0,8	0
30	32	3	1	27	2	2	0,333333	0,333333	0,875	0,333333
31	33	3	1	28	2	2	0,333333	0,333333	0,878788	0,333333
32	28	3	2	24	1	1	0,666667	0,666667	0,928571	0,666667
33	40	4	1	33	3	3	0,25	0,25	0,85	0,25
34	34	3	1	29	2	2	0,333333	0,333333	0,882353	0,333333
35	24	2	0	20	2	2	0	0	0,833333	0
36	37	4	3	32	1	1	0,75	0,75	0,945946	0,75
37	25	2	0	21	2	2	0	0	0,84	0
38	24	2	0	20	2	2	0	0	0,833333	0
39	36	4	3	31	1	1	0,75	0,75	0,944444	0,75
40	21	2	0	17	2	2	0	0	0,809524	0
41	32	3	1	27	2	2	0,333333	0,333333	0,875	0,333333
42	35	4	0	27	4	4	0	0	0,771429	0
43	28	3	1	23	2	2	0,333333	0,333333	0,857143	0,333333
44	24	2	1	21	1	1	0,5	0,5	0,916667	0,5
45	32	3	1	27	2	2	0,333333	0,333333	0,875	0,333333

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	4	1	36	3	3	0,25	0,25	0,860465	0,25
47	37	4	2	31	2	2	0,5	0,5	0,891892	0,5
48	37	4	3	32	1	1	0,75	0,75	0,945946	0,75
49	36	4	2	30	2	2	0,5	0,5	0,888889	0,5
50	20	2	0	16	2	2	0	0	0,8	0
51	37	4	2	31	2	2	0,5	0,5	0,891892	0,5
52	31	3	1	26	2	2	0,333333	0,333333	0,870968	0,333333
53	17	2	2	15	0	0	1	1	1	1
54	37	4	3	32	1	1	0,75	0,75	0,945946	0,75
55	49	5	2	41	3	3	0,4	0,4	0,877551	0,4
56	34	3	1	29	2	2	0,333333	0,333333	0,882353	0,333333
57	38	4	2	32	2	2	0,5	0,5	0,894737	0,5
58	75	8	2	61	6	6	0,25	0,25	0,84	0,25
59	56	6	1	45	5	5	0,166667	0,166667	0,821429	0,166667
60	31	3	1	26	2	2	0,333333	0,333333	0,870968	0,333333
61	31	3	1	26	2	2	0,333333	0,333333	0,870968	0,333333
62	46	5	0	36	5	5	0	0	0,782609	0
63	49	5	2	41	3	3	0,4	0,4	0,877551	0,4
64	38	4	1	31	3	3	0,25	0,25	0,842105	0,25
65	30	3	1	25	2	2	0,333333	0,333333	0,866667	0,333333
66	19	2	1	16	1	1	0,5	0,5	0,894737	0,5
67	17	2	0	13	2	2	0	0	0,764706	0
68	11	1	1	10	0	0	1	1	1	1
69	24	2	0	20	2	2	0	0	0,833333	0
70	29	3	2	25	1	1	0,666667	0,666667	0,931034	0,666667
71	14	1	0	12	1	1	0	0	0,857143	0
72	22	2	1	19	1	1	0,5	0,5	0,909091	0,5
73	38	4	2	32	2	2	0,5	0,5	0,894737	0,5
74	32	3	2	28	1	1	0,666667	0,666667	0,9375	0,666667
75	42	4	2	36	2	2	0,5	0,5	0,904762	0,5
76	21	2	0	17	2	2	0	0	0,809524	0
77	26	3	0	20	3	3	0	0	0,769231	0
78	25	2	0	21	2	2	0	0	0,84	0
79	34	3	0	28	3	3	0	0	0,823529	0
80	27	3	1	22	2	2	0,333333	0,333333	0,851852	0,333333
Rata-Rata						0,325744	0,325744	0,866234	0,325744	

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

### Lampiran 13 :

Hasil Confusion Matrix - Textrank dengan BoW Cosine *similarity*, dengan  
Compression Rate 20%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	4	3	15	1	1	0,75	0,75	0,9	0,75
2	26	5	3	19	2	2	0,6	0,6	0,846154	0,6
3	37	7	3	26	4	4	0,428571	0,428571	0,783784	0,428571
4	41	8	2	27	6	6	0,25	0,25	0,707317	0,25
5	29	6	2	19	4	4	0,333333	0,333333	0,724138	0,333333
6	29	6	2	19	4	4	0,333333	0,333333	0,724138	0,333333
7	47	9	2	31	7	7	0,222222	0,222222	0,702128	0,222222
8	35	7	4	25	3	3	0,571429	0,571429	0,828571	0,571429
9	31	6	2	21	4	4	0,333333	0,333333	0,741935	0,333333
10	20	4	2	14	2	2	0,5	0,5	0,8	0,5
11	39	8	5	28	3	3	0,625	0,625	0,846154	0,625
12	29	6	1	18	5	5	0,166667	0,166667	0,655172	0,166667
13	38	8	4	26	4	4	0,5	0,5	0,789474	0,5
14	36	7	4	26	3	3	0,571429	0,571429	0,833333	0,571429
15	20	4	2	14	2	2	0,5	0,5	0,8	0,5
16	41	8	6	31	2	2	0,75	0,75	0,902439	0,75
17	68	14	7	47	7	7	0,5	0,5	0,794118	0,5
18	64	13	10	48	3	3	0,769231	0,769231	0,90625	0,769231
19	28	6	3	19	3	3	0,5	0,5	0,785714	0,5
20	49	10	6	35	4	4	0,6	0,6	0,836735	0,6
21	45	9	6	33	3	3	0,666667	0,666667	0,866667	0,666667
22	28	6	3	19	3	3	0,5	0,5	0,785714	0,5
23	23	5	3	16	2	2	0,6	0,6	0,826087	0,6
24	25	5	2	17	3	3	0,4	0,4	0,76	0,4
25	31	6	2	21	4	4	0,333333	0,333333	0,741935	0,333333
26	28	6	2	18	4	4	0,333333	0,333333	0,714286	0,333333
27	28	6	2	18	4	4	0,333333	0,333333	0,714286	0,333333
28	30	6	2	20	4	4	0,333333	0,333333	0,733333	0,333333
29	20	4	1	13	3	3	0,25	0,25	0,7	0,25
30	32	6	4	24	2	2	0,666667	0,666667	0,875	0,666667
31	33	7	4	23	3	3	0,571429	0,571429	0,818182	0,571429
32	28	6	4	20	2	2	0,666667	0,666667	0,857143	0,666667
33	40	8	4	28	4	4	0,5	0,5	0,8	0,5
34	34	7	2	22	5	5	0,285714	0,285714	0,705882	0,285714
35	24	5	2	16	3	3	0,4	0,4	0,75	0,4
36	37	7	6	29	1	1	0,857143	0,857143	0,945946	0,857143
37	25	5	3	18	2	2	0,6	0,6	0,84	0,6
38	24	5	3	17	2	2	0,6	0,6	0,833333	0,6
39	36	7	4	26	3	3	0,571429	0,571429	0,833333	0,571429
40	21	4	2	15	2	2	0,5	0,5	0,809524	0,5
41	32	6	4	24	2	2	0,666667	0,666667	0,875	0,666667
42	35	7	3	24	4	4	0,428571	0,428571	0,771429	0,428571
43	28	6	2	18	4	4	0,333333	0,333333	0,714286	0,333333
44	24	5	4	18	1	1	0,8	0,8	0,916667	0,8
45	32	6	4	24	2	2	0,666667	0,666667	0,875	0,666667

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	9	3	28	6	6	0,333333	0,333333	0,72093	0,333333
47	37	7	2	25	5	5	0,285714	0,285714	0,72973	0,285714
48	37	7	3	26	4	4	0,428571	0,428571	0,783784	0,428571
49	36	7	4	26	3	3	0,571429	0,571429	0,833333	0,571429
50	20	4	2	14	2	2	0,5	0,5	0,8	0,5
51	37	7	3	26	4	4	0,428571	0,428571	0,783784	0,428571
52	31	6	3	22	3	3	0,5	0,5	0,806452	0,5
53	17	3	2	13	1	1	0,666667	0,666667	0,882353	0,666667
54	37	7	3	26	4	4	0,428571	0,428571	0,783784	0,428571
55	49	10	3	32	7	7	0,3	0,3	0,714286	0,3
56	34	7	3	23	4	4	0,428571	0,428571	0,764706	0,428571
57	38	8	6	28	2	2	0,75	0,75	0,894737	0,75
58	75	15	7	52	8	8	0,466667	0,466667	0,786667	0,466667
59	56	11	4	38	7	7	0,363636	0,363636	0,75	0,363636
60	31	6	3	22	3	3	0,5	0,5	0,806452	0,5
61	31	6	1	20	5	5	0,166667	0,166667	0,677419	0,166667
62	46	9	3	31	6	6	0,333333	0,333333	0,73913	0,333333
63	49	10	5	34	5	5	0,5	0,5	0,795918	0,5
64	38	8	4	26	4	4	0,5	0,5	0,789474	0,5
65	30	6	3	21	3	3	0,5	0,5	0,8	0,5
66	19	4	3	14	1	1	0,75	0,75	0,894737	0,75
67	17	3	1	12	2	2	0,333333	0,333333	0,764706	0,333333
68	11	2	1	8	1	1	0,5	0,5	0,818182	0,5
69	24	5	2	16	3	3	0,4	0,4	0,75	0,4
70	29	6	4	21	2	2	0,666667	0,666667	0,862069	0,666667
71	14	3	0	8	3	3	0	0	0,571429	0
72	22	4	2	16	2	2	0,5	0,5	0,818182	0,5
73	38	8	5	27	3	3	0,625	0,625	0,842105	0,625
74	32	6	3	23	3	3	0,5	0,5	0,8125	0,5
75	42	8	3	29	5	5	0,375	0,375	0,761905	0,375
76	21	4	2	15	2	2	0,5	0,5	0,809524	0,5
77	26	5	2	18	3	3	0,4	0,4	0,769231	0,4
78	25	5	2	17	3	3	0,4	0,4	0,76	0,4
79	34	7	2	22	5	5	0,285714	0,285714	0,705882	0,285714
80	27	5	1	18	4	4	0,2	0,2	0,703704	0,2
Rata-rata						0,475703	0,475703	0,790721	0,475703	

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 14 :

Hasil Confusion Matrix - Textrank dengan BoW Cosine *similarity*, dengan  
Compression Rate 30%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	6	4	12	2	2	0,666667	0,666667	0,8	0,666667
2	26	8	6	16	2	2	0,75	0,75	0,846154	0,75
3	37	11	6	21	5	5	0,545455	0,545455	0,72973	0,545455
4	41	12	3	20	9	9	0,25	0,25	0,560976	0,25
5	29	9	3	14	6	6	0,333333	0,333333	0,586207	0,333333
6	29	9	4	15	5	5	0,444444	0,444444	0,655172	0,444444
7	47	14	6	25	8	8	0,428571	0,428571	0,659574	0,428571
8	35	10	6	21	4	4	0,6	0,6	0,771429	0,6
9	31	9	5	18	4	4	0,555556	0,555556	0,741935	0,555556
10	20	6	5	13	1	1	0,833333	0,833333	0,9	0,833333
11	39	12	6	21	6	6	0,5	0,5	0,692308	0,5
12	29	9	5	16	4	4	0,555556	0,555556	0,724138	0,555556
13	38	11	7	23	4	4	0,636364	0,636364	0,789474	0,636364
14	36	11	7	21	4	4	0,636364	0,636364	0,777778	0,636364
15	20	6	3	11	3	3	0,5	0,5	0,7	0,5
16	41	12	9	26	3	3	0,75	0,75	0,853659	0,75
17	68	20	13	41	7	7	0,65	0,65	0,794118	0,65
18	64	19	13	39	6	6	0,684211	0,684211	0,8125	0,684211
19	28	8	4	16	4	4	0,5	0,5	0,714286	0,5
20	49	15	8	27	7	7	0,533333	0,533333	0,714286	0,533333
21	45	14	8	25	6	6	0,571429	0,571429	0,733333	0,571429
22	28	8	5	17	3	3	0,625	0,625	0,785714	0,625
23	23	7	4	13	3	3	0,571429	0,571429	0,73913	0,571429
24	25	8	2	11	6	6	0,25	0,25	0,52	0,25
25	31	9	5	18	4	4	0,555556	0,555556	0,741935	0,555556
26	28	8	4	16	4	4	0,5	0,5	0,714286	0,5
27	28	8	5	17	3	3	0,625	0,625	0,785714	0,625
28	30	9	4	16	5	5	0,444444	0,444444	0,666667	0,444444
29	20	6	2	10	4	4	0,333333	0,333333	0,6	0,333333
30	32	10	7	19	3	3	0,7	0,7	0,8125	0,7
31	33	10	8	21	2	2	0,8	0,8	0,878788	0,8
32	28	8	5	17	3	3	0,625	0,625	0,785714	0,625
33	40	12	5	21	7	7	0,416667	0,416667	0,65	0,416667
34	34	10	3	17	7	7	0,3	0,3	0,588235	0,3
35	24	7	4	14	3	3	0,571429	0,571429	0,75	0,571429
36	37	11	7	22	4	4	0,636364	0,636364	0,783784	0,636364
37	25	8	6	15	2	2	0,75	0,75	0,84	0,75
38	24	7	5	15	2	2	0,714286	0,714286	0,833333	0,714286
39	36	11	7	21	4	4	0,636364	0,636364	0,777778	0,636364
40	21	6	4	13	2	2	0,666667	0,666667	0,809524	0,666667
41	32	10	7	19	3	3	0,7	0,7	0,8125	0,7
42	35	10	3	18	7	7	0,3	0,3	0,6	0,3
43	28	8	3	15	5	5	0,375	0,375	0,642857	0,375
44	24	7	5	15	2	2	0,714286	0,714286	0,833333	0,714286
45	32	10	8	20	2	2	0,8	0,8	0,875	0,8

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	13	6	23	7	7	0,461538	0,461538	0,674419	0,461538
47	37	11	3	18	8	8	0,272727	0,272727	0,567568	0,272727
48	37	11	6	21	5	5	0,545455	0,545455	0,72973	0,545455
49	36	11	7	21	4	4	0,636364	0,636364	0,777778	0,636364
50	20	6	2	10	4	4	0,333333	0,333333	0,6	0,333333
51	37	11	6	21	5	5	0,545455	0,545455	0,72973	0,545455
52	31	9	6	19	3	3	0,666667	0,666667	0,806452	0,666667
53	17	5	4	11	1	1	0,8	0,8	0,882353	0,8
54	37	11	7	22	4	4	0,636364	0,636364	0,783784	0,636364
55	49	15	9	28	6	6	0,6	0,6	0,755102	0,6
56	34	10	4	18	6	6	0,4	0,4	0,647059	0,4
57	38	11	9	25	2	2	0,818182	0,818182	0,894737	0,818182
58	75	22	14	45	8	8	0,636364	0,636364	0,786667	0,636364
59	56	17	7	29	10	10	0,411765	0,411765	0,642857	0,411765
60	31	9	6	19	3	3	0,666667	0,666667	0,806452	0,666667
61	31	9	5	18	4	4	0,555556	0,555556	0,741935	0,555556
62	46	14	7	25	7	7	0,5	0,5	0,695652	0,5
63	49	15	8	27	7	7	0,533333	0,533333	0,714286	0,533333
64	38	11	4	20	7	7	0,363636	0,363636	0,631579	0,363636
65	30	9	6	18	3	3	0,666667	0,666667	0,8	0,666667
66	19	6	5	12	1	1	0,833333	0,833333	0,894737	0,833333
67	17	5	2	9	3	3	0,4	0,4	0,647059	0,4
68	11	3	1	6	2	2	0,333333	0,333333	0,636364	0,333333
69	24	7	5	15	2	2	0,714286	0,714286	0,833333	0,714286
70	29	9	6	17	3	3	0,666667	0,666667	0,793103	0,666667
71	14	4	1	7	3	3	0,25	0,25	0,571429	0,25
72	22	7	4	12	3	3	0,571429	0,571429	0,727273	0,571429
73	38	11	8	24	3	3	0,727273	0,727273	0,842105	0,727273
74	32	10	5	17	5	5	0,5	0,5	0,6875	0,5
75	42	13	7	23	6	6	0,538462	0,538462	0,714286	0,538462
76	21	6	2	11	4	4	0,333333	0,333333	0,619048	0,333333
77	26	8	5	15	3	3	0,625	0,625	0,769231	0,625
78	25	8	4	13	4	4	0,5	0,5	0,68	0,5
79	34	10	4	18	6	6	0,4	0,4	0,647059	0,4
80	27	8	6	17	2	2	0,75	0,75	0,851852	0,75
Rata-rata						0,559108	0,559108	0,73708	0,559108	

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 15 :

Hasil Confusion Matrix - Textrank dengan BoW Cosine *similarity*, dengan

Compression Rate 40%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	8	5	9	3	3	0,625	0,625	0,7	0,625
2	26	10	8	14	2	2	0,8	0,8	0,846154	0,8
3	37	15	10	17	5	5	0,666667	0,666667	0,72973	0,666667
4	41	16	5	14	11	11	0,3125	0,3125	0,463415	0,3125
5	29	12	7	12	5	5	0,583333	0,583333	0,655172	0,583333
6	29	12	7	12	5	5	0,583333	0,583333	0,655172	0,583333
7	47	19	11	20	8	8	0,578947	0,578947	0,659574	0,578947
8	35	14	10	17	4	4	0,714286	0,714286	0,771429	0,714286
9	31	12	8	15	4	4	0,666667	0,666667	0,741935	0,666667
10	20	8	6	10	2	2	0,75	0,75	0,8	0,75
11	39	16	7	14	9	9	0,4375	0,4375	0,538462	0,4375
12	29	12	10	15	2	2	0,833333	0,833333	0,862069	0,833333
13	38	15	11	19	4	4	0,733333	0,733333	0,789474	0,733333
14	36	14	10	18	4	4	0,714286	0,714286	0,777778	0,714286
15	20	8	3	7	5	5	0,375	0,375	0,5	0,375
16	41	16	11	20	5	5	0,6875	0,6875	0,756098	0,6875
17	68	27	21	35	6	6	0,777778	0,777778	0,823529	0,777778
18	64	26	18	30	8	8	0,692308	0,692308	0,75	0,692308
19	28	11	9	15	2	2	0,818182	0,818182	0,857143	0,818182
20	49	20	14	23	6	6	0,7	0,7	0,755102	0,7
21	45	18	11	20	7	7	0,611111	0,611111	0,688889	0,611111
22	28	11	8	14	3	3	0,727273	0,727273	0,785714	0,727273
23	23	9	5	10	4	4	0,555556	0,555556	0,652174	0,555556
24	25	10	3	8	7	7	0,3	0,3	0,44	0,3
25	31	12	8	15	4	4	0,666667	0,666667	0,741935	0,666667
26	28	11	8	14	3	3	0,727273	0,727273	0,785714	0,727273
27	28	11	7	13	4	4	0,636364	0,636364	0,714286	0,636364
28	30	12	7	13	5	5	0,583333	0,583333	0,666667	0,583333
29	20	8	5	9	3	3	0,625	0,625	0,7	0,625
30	32	13	10	16	3	3	0,769231	0,769231	0,8125	0,769231
31	33	13	9	16	4	4	0,692308	0,692308	0,757576	0,692308
32	28	11	7	13	4	4	0,636364	0,636364	0,714286	0,636364
33	40	16	11	19	5	5	0,6875	0,6875	0,75	0,6875
34	34	14	9	15	5	5	0,642857	0,642857	0,705882	0,642857
35	24	10	7	11	3	3	0,7	0,7	0,75	0,7
36	37	15	9	16	6	6	0,6	0,6	0,675676	0,6
37	25	10	7	12	3	3	0,7	0,7	0,76	0,7
38	24	10	7	11	3	3	0,7	0,7	0,75	0,7
39	36	14	7	15	7	7	0,5	0,5	0,611111	0,5
40	21	8	4	9	4	4	0,5	0,5	0,619048	0,5
41	32	13	11	17	2	2	0,846154	0,846154	0,875	0,846154
42	35	14	8	15	6	6	0,571429	0,571429	0,657143	0,571429
43	28	11	4	10	7	7	0,363636	0,363636	0,5	0,363636
44	24	10	7	11	3	3	0,7	0,7	0,75	0,7
45	32	13	9	15	4	4	0,692308	0,692308	0,75	0,692308

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	17	11	20	6	6	0,647059	0,647059	0,72093	0,647059
47	37	15	10	17	5	5	0,666667	0,666667	0,72973	0,666667
48	37	15	9	16	6	6	0,6	0,6	0,675676	0,6
49	36	14	9	17	5	5	0,642857	0,642857	0,722222	0,642857
50	20	8	5	9	3	3	0,625	0,625	0,7	0,625
51	37	15	9	16	6	6	0,6	0,6	0,675676	0,6
52	31	12	9	16	3	3	0,75	0,75	0,806452	0,75
53	17	7	5	8	2	2	0,714286	0,714286	0,764706	0,714286
54	37	15	11	18	4	4	0,733333	0,733333	0,783784	0,733333
55	49	20	13	22	7	7	0,65	0,65	0,714286	0,65
56	34	14	9	15	5	5	0,642857	0,642857	0,705882	0,642857
57	38	15	11	19	4	4	0,733333	0,733333	0,789474	0,733333
58	75	30	20	35	10	10	0,666667	0,666667	0,733333	0,666667
59	56	22	11	23	11	11	0,5	0,5	0,607143	0,5
60	31	12	7	14	5	5	0,583333	0,583333	0,677419	0,583333
61	31	12	8	15	4	4	0,666667	0,666667	0,741935	0,666667
62	46	18	10	20	8	8	0,555556	0,555556	0,652174	0,555556
63	49	20	11	20	9	9	0,55	0,55	0,632653	0,55
64	38	15	8	16	7	7	0,533333	0,533333	0,631579	0,533333
65	30	12	8	14	4	4	0,666667	0,666667	0,733333	0,666667
66	19	8	6	9	2	2	0,75	0,75	0,789474	0,75
67	17	7	4	7	3	3	0,571429	0,571429	0,647059	0,571429
68	11	4	2	5	2	2	0,5	0,5	0,636364	0,5
69	24	10	9	13	1	1	0,9	0,9	0,916667	0,9
70	29	12	9	14	3	3	0,75	0,75	0,793103	0,75
71	14	6	3	5	3	3	0,5	0,5	0,571429	0,5
72	22	9	7	11	2	2	0,777778	0,777778	0,818182	0,777778
73	38	15	10	18	5	5	0,666667	0,666667	0,736842	0,666667
74	32	13	8	14	5	5	0,615385	0,615385	0,6875	0,615385
75	42	17	12	20	5	5	0,705882	0,705882	0,761905	0,705882
76	21	8	5	10	3	3	0,625	0,625	0,714286	0,625
77	26	10	6	12	4	4	0,6	0,6	0,692308	0,6
78	25	10	6	11	4	4	0,6	0,6	0,68	0,6
79	34	14	8	14	6	6	0,571429	0,571429	0,647059	0,571429
80	27	11	7	12	4	4	0,636364	0,636364	0,703704	0,636364
Rata-rata						0,641023	0,641023	0,713001	0,641023	

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 16 :

Hasil Confusion Matrix - Textrank dengan BoW Cosine *similarity*, dengan  
Compression Rate 50%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	10	8	8	2	2	0,8	0,8	0,8	0,8
2	26	13	11	11	2	2	0,846154	0,846154	0,846154	0,846154
3	37	18	14	15	4	4	0,777778	0,777778	0,783784	0,777778
4	41	20	10	11	10	10	0,5	0,5	0,512195	0,5
5	29	14	9	10	5	5	0,642857	0,642857	0,655172	0,642857
6	29	14	8	9	6	6	0,571429	0,571429	0,586207	0,571429
7	47	24	18	17	6	6	0,75	0,75	0,744681	0,75
8	35	18	14	13	4	4	0,777778	0,777778	0,771429	0,777778
9	31	16	9	8	7	7	0,5625	0,5625	0,548387	0,5625
10	20	10	9	9	1	1	0,9	0,9	0,9	0,9
11	39	20	14	13	6	6	0,7	0,7	0,692308	0,7
12	29	14	10	11	4	4	0,714286	0,714286	0,724138	0,714286
13	38	19	13	13	6	6	0,684211	0,684211	0,684211	0,684211
14	36	18	16	16	2	2	0,888889	0,888889	0,888889	0,888889
15	20	10	5	5	5	5	0,5	0,5	0,5	0,5
16	41	20	14	15	6	6	0,7	0,7	0,707317	0,7
17	68	34	29	29	5	5	0,852941	0,852941	0,852941	0,852941
18	64	32	22	22	10	10	0,6875	0,6875	0,6875	0,6875
19	28	14	10	10	4	4	0,714286	0,714286	0,714286	0,714286
20	49	24	17	18	7	7	0,708333	0,708333	0,714286	0,708333
21	45	22	15	16	7	7	0,681818	0,681818	0,688889	0,681818
22	28	14	11	11	3	3	0,785714	0,785714	0,785714	0,785714
23	23	12	8	7	4	4	0,666667	0,666667	0,652174	0,666667
24	25	12	5	6	7	7	0,416667	0,416667	0,44	0,416667
25	31	16	11	10	5	5	0,6875	0,6875	0,677419	0,6875
26	28	14	12	12	2	2	0,857143	0,857143	0,857143	0,857143
27	28	14	10	10	4	4	0,714286	0,714286	0,714286	0,714286
28	30	15	11	11	4	4	0,733333	0,733333	0,733333	0,733333
29	20	10	6	6	4	4	0,6	0,6	0,6	0,6
30	32	16	14	14	2	2	0,875	0,875	0,875	0,875
31	33	16	10	11	6	6	0,625	0,625	0,636364	0,625
32	28	14	9	9	5	5	0,642857	0,642857	0,642857	0,642857
33	40	20	13	13	7	7	0,65	0,65	0,65	0,65
34	34	17	11	11	6	6	0,647059	0,647059	0,647059	0,647059
35	24	12	8	8	4	4	0,666667	0,666667	0,666667	0,666667
36	37	18	12	13	6	6	0,666667	0,666667	0,675676	0,666667
37	25	12	8	9	4	4	0,666667	0,666667	0,68	0,666667
38	24	12	7	7	5	5	0,583333	0,583333	0,583333	0,583333
39	36	18	11	11	7	7	0,611111	0,611111	0,611111	0,611111
40	21	10	6	7	4	4	0,6	0,6	0,619048	0,6
41	32	16	12	12	4	4	0,75	0,75	0,75	0,75
42	35	18	10	9	8	8	0,555556	0,555556	0,542857	0,555556
43	28	14	5	5	9	9	0,357143	0,357143	0,357143	0,357143
44	24	12	10	10	2	2	0,833333	0,833333	0,833333	0,833333
45	32	16	12	12	4	4	0,75	0,75	0,75	0,75

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	22	14	13	8	8	0,636364	0,636364	0,627907	0,636364
47	37	18	12	13	6	6	0,666667	0,666667	0,675676	0,666667
48	37	18	13	14	5	5	0,722222	0,722222	0,72973	0,722222
49	36	18	12	12	6	6	0,666667	0,666667	0,666667	0,666667
50	20	10	7	7	3	3	0,7	0,7	0,7	0,7
51	37	18	13	14	5	5	0,722222	0,722222	0,72973	0,722222
52	31	16	13	12	3	3	0,8125	0,8125	0,806452	0,8125
53	17	8	5	6	3	3	0,625	0,625	0,647059	0,625
54	37	18	13	14	5	5	0,722222	0,722222	0,72973	0,722222
55	49	24	15	16	9	9	0,625	0,625	0,632653	0,625
56	34	17	11	11	6	6	0,647059	0,647059	0,647059	0,647059
57	38	19	16	16	3	3	0,842105	0,842105	0,842105	0,842105
58	75	38	29	28	9	9	0,763158	0,763158	0,76	0,763158
59	56	28	17	17	11	11	0,607143	0,607143	0,607143	0,607143
60	31	16	12	11	4	4	0,75	0,75	0,741935	0,75
61	31	16	11	10	5	5	0,6875	0,6875	0,677419	0,6875
62	46	23	16	16	7	7	0,695652	0,695652	0,695652	0,695652
63	49	24	15	16	9	9	0,625	0,625	0,632653	0,625
64	38	19	12	12	7	7	0,631579	0,631579	0,631579	0,631579
65	30	15	10	10	5	5	0,666667	0,666667	0,666667	0,666667
66	19	10	7	6	3	3	0,7	0,7	0,684211	0,7
67	17	8	5	6	3	3	0,625	0,625	0,647059	0,625
68	11	6	5	4	1	1	0,833333	0,833333	0,818182	0,833333
69	24	12	10	10	2	2	0,833333	0,833333	0,833333	0,833333
70	29	14	11	12	3	3	0,785714	0,785714	0,793103	0,785714
71	14	7	4	4	3	3	0,571429	0,571429	0,571429	0,571429
72	22	11	9	9	2	2	0,818182	0,818182	0,818182	0,818182
73	38	19	15	15	4	4	0,789474	0,789474	0,789474	0,789474
74	32	16	12	12	4	4	0,75	0,75	0,75	0,75
75	42	21	17	17	4	4	0,809524	0,809524	0,809524	0,809524
76	21	10	8	9	2	2	0,8	0,8	0,809524	0,8
77	26	13	10	10	3	3	0,769231	0,769231	0,769231	0,769231
78	25	12	7	8	5	5	0,583333	0,583333	0,6	0,583333
79	34	17	13	13	4	4	0,764706	0,764706	0,764706	0,764706
80	27	14	10	9	4	4	0,714286	0,714286	0,703704	0,714286
Rata-rata						0,698309	0,698309	0,69991	0,698309	

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 17 :

Hasil Confusion Matrix - Textrank dengan TF-IDF Cosine *similarity*, dengan

Compression Rate 10%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	2	1	17	1	1	0,5	0,5	0,9	0,5
2	26	3	1	21	2	2	0,333333	0,333333	0,846154	0,333333
3	37	4	1	30	3	3	0,25	0,25	0,837838	0,25
4	41	4	0	33	4	4	0	0	0,804878	0
5	29	3	2	25	1	1	0,666667	0,666667	0,931034	0,666667
6	29	3	2	25	1	1	0,666667	0,666667	0,931034	0,666667
7	47	5	0	37	5	5	0	0	0,787234	0
8	35	4	1	28	3	3	0,25	0,25	0,828571	0,25
9	31	3	2	27	1	1	0,666667	0,666667	0,935484	0,666667
10	20	2	0	16	2	2	0	0	0,8	0
11	39	4	1	32	3	3	0,25	0,25	0,846154	0,25
12	29	3	0	23	3	3	0	0	0,793103	0
13	38	4	1	31	3	3	0,25	0,25	0,842105	0,25
14	36	4	0	28	4	4	0	0	0,777778	0
15	20	2	1	17	1	1	0,5	0,5	0,9	0,5
16	41	4	2	35	2	2	0,5	0,5	0,902439	0,5
17	68	7	2	56	5	5	0,285714	0,285714	0,852941	0,285714
18	64	6	4	56	2	2	0,666667	0,666667	0,9375	0,666667
19	28	3	0	22	3	3	0	0	0,785714	0
20	49	5	2	41	3	3	0,4	0,4	0,877551	0,4
21	45	4	1	38	3	3	0,25	0,25	0,866667	0,25
22	28	3	1	23	2	2	0,333333	0,333333	0,857143	0,333333
23	23	2	0	19	2	2	0	0	0,826087	0
24	25	2	0	21	2	2	0	0	0,84	0
25	31	3	0	25	3	3	0	0	0,806452	0
26	28	3	1	23	2	2	0,333333	0,333333	0,857143	0,333333
27	28	3	2	24	1	1	0,666667	0,666667	0,928571	0,666667
28	30	3	2	26	1	1	0,666667	0,666667	0,933333	0,666667
29	20	2	0	16	2	2	0	0	0,8	0
30	32	3	1	27	2	2	0,333333	0,333333	0,875	0,333333
31	33	3	1	28	2	2	0,333333	0,333333	0,878788	0,333333
32	28	3	2	24	1	1	0,666667	0,666667	0,928571	0,666667
33	40	4	2	34	2	2	0,5	0,5	0,9	0,5
34	34	3	1	29	2	2	0,333333	0,333333	0,882353	0,333333
35	24	2	0	20	2	2	0	0	0,833333	0
36	37	4	3	32	1	1	0,75	0,75	0,945946	0,75
37	25	2	0	21	2	2	0	0	0,84	0
38	24	2	0	20	2	2	0	0	0,833333	0
39	36	4	2	30	2	2	0,5	0,5	0,888889	0,5
40	21	2	0	17	2	2	0	0	0,809524	0
41	32	3	0	26	3	3	0	0	0,8125	0
42	35	4	1	28	3	3	0,25	0,25	0,828571	0,25
43	28	3	1	23	2	2	0,333333	0,333333	0,857143	0,333333
44	24	2	0	20	2	2	0	0	0,833333	0
45	32	3	1	27	2	2	0,333333	0,333333	0,875	0,333333

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	4	1	36	3	3	0,25	0,25	0,860465	0,25
47	37	4	1	30	3	3	0,25	0,25	0,837838	0,25
48	37	4	1	30	3	3	0,25	0,25	0,837838	0,25
49	36	4	2	30	2	2	0,5	0,5	0,888889	0,5
50	20	2	0	16	2	2	0	0	0,8	0
51	37	4	1	30	3	3	0,25	0,25	0,837838	0,25
52	31	3	1	26	2	2	0,333333	0,333333	0,870968	0,333333
53	17	2	1	14	1	1	0,5	0,5	0,882353	0,5
54	37	4	3	32	1	1	0,75	0,75	0,945946	0,75
55	49	5	1	40	4	4	0,2	0,2	0,836735	0,2
56	34	3	1	29	2	2	0,333333	0,333333	0,882353	0,333333
57	38	4	3	33	1	1	0,75	0,75	0,947368	0,75
58	75	8	1	60	7	7	0,125	0,125	0,813333	0,125
59	56	6	1	45	5	5	0,166667	0,166667	0,821429	0,166667
60	31	3	2	27	1	1	0,666667	0,666667	0,935484	0,666667
61	31	3	1	26	2	2	0,333333	0,333333	0,870968	0,333333
62	46	5	1	37	4	4	0,2	0,2	0,826087	0,2
63	49	5	3	42	2	2	0,6	0,6	0,918367	0,6
64	38	4	1	31	3	3	0,25	0,25	0,842105	0,25
65	30	3	2	26	1	1	0,666667	0,666667	0,933333	0,666667
66	19	2	1	16	1	1	0,5	0,5	0,894737	0,5
67	17	2	0	13	2	2	0	0	0,764706	0
68	11	1	1	10	0	0	1	1	1	1
69	24	2	0	20	2	2	0	0	0,833333	0
70	29	3	2	25	1	1	0,666667	0,666667	0,931034	0,666667
71	14	1	0	12	1	1	0	0	0,857143	0
72	22	2	1	19	1	1	0,5	0,5	0,909091	0,5
73	38	4	2	32	2	2	0,5	0,5	0,894737	0,5
74	32	3	2	28	1	1	0,666667	0,666667	0,9375	0,666667
75	42	4	3	37	1	1	0,75	0,75	0,952381	0,75
76	21	2	0	17	2	2	0	0	0,809524	0
77	26	3	1	21	2	2	0,333333	0,333333	0,846154	0,333333
78	25	2	1	22	1	1	0,5	0,5	0,92	0,5
79	34	3	1	29	2	2	0,333333	0,333333	0,882353	0,333333
80	27	3	2	23	1	1	0,666667	0,666667	0,925926	0,666667
Rata-rata							0,332009	0,332009	0,867544	0,332009

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 18 :

Hasil Confusion Matrix - Textrank dengan TF-IDF Cosine *similarity*, dengan

Compression Rate 20%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	4	2	14	2	2	0,5	0,5	0,8	0,5
2	26	5	3	19	2	2	0,6	0,6	0,846154	0,6
3	37	7	3	26	4	4	0,428571	0,428571	0,783784	0,428571
4	41	8	2	27	6	6	0,25	0,25	0,707317	0,25
5	29	6	2	19	4	4	0,333333	0,333333	0,724138	0,333333
6	29	6	3	20	3	3	0,5	0,5	0,793103	0,5
7	47	9	2	31	7	7	0,222222	0,222222	0,702128	0,222222
8	35	7	4	25	3	3	0,571429	0,571429	0,828571	0,571429
9	31	6	3	22	3	3	0,5	0,5	0,806452	0,5
10	20	4	2	14	2	2	0,5	0,5	0,8	0,5
11	39	8	4	27	4	4	0,5	0,5	0,794872	0,5
12	29	6	2	19	4	4	0,333333	0,333333	0,724138	0,333333
13	38	8	3	25	5	5	0,375	0,375	0,736842	0,375
14	36	7	3	25	4	4	0,428571	0,428571	0,777778	0,428571
15	20	4	2	14	2	2	0,5	0,5	0,8	0,5
16	41	8	6	31	2	2	0,75	0,75	0,902439	0,75
17	68	14	11	51	3	3	0,785714	0,785714	0,911765	0,785714
18	64	13	9	47	4	4	0,692308	0,692308	0,875	0,692308
19	28	6	2	18	4	4	0,333333	0,333333	0,714286	0,333333
20	49	10	5	34	5	5	0,5	0,5	0,795918	0,5
21	45	9	5	32	4	4	0,555556	0,555556	0,822222	0,555556
22	28	6	4	20	2	2	0,666667	0,666667	0,857143	0,666667
23	23	5	2	15	3	3	0,4	0,4	0,73913	0,4
24	25	5	2	17	3	3	0,4	0,4	0,76	0,4
25	31	6	3	22	3	3	0,5	0,5	0,806452	0,5
26	28	6	4	20	2	2	0,666667	0,666667	0,857143	0,666667
27	28	6	4	20	2	2	0,666667	0,666667	0,857143	0,666667
28	30	6	4	22	2	2	0,666667	0,666667	0,866667	0,666667
29	20	4	0	12	4	4	0	0	0,6	0
30	32	6	5	25	1	1	0,833333	0,833333	0,9375	0,833333
31	33	7	4	23	3	3	0,571429	0,571429	0,818182	0,571429
32	28	6	3	19	3	3	0,5	0,5	0,785714	0,5
33	40	8	4	28	4	4	0,5	0,5	0,8	0,5
34	34	7	2	22	5	5	0,285714	0,285714	0,705882	0,285714
35	24	5	3	17	2	2	0,6	0,6	0,833333	0,6
36	37	7	5	28	2	2	0,714286	0,714286	0,891892	0,714286
37	25	5	3	18	2	2	0,6	0,6	0,84	0,6
38	24	5	3	17	2	2	0,6	0,6	0,833333	0,6
39	36	7	4	26	3	3	0,571429	0,571429	0,833333	0,571429
40	21	4	2	15	2	2	0,5	0,5	0,809524	0,5
41	32	6	3	23	3	3	0,5	0,5	0,8125	0,5
42	35	7	3	24	4	4	0,428571	0,428571	0,771429	0,428571
43	28	6	2	18	4	4	0,333333	0,333333	0,714286	0,333333
44	24	5	4	18	1	1	0,8	0,8	0,916667	0,8
45	32	6	3	23	3	3	0,5	0,5	0,8125	0,5

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	9	4	29	5	5	0,444444	0,444444	0,767442	0,444444
47	37	7	2	25	5	5	0,285714	0,285714	0,72973	0,285714
48	37	7	4	27	3	3	0,571429	0,571429	0,837838	0,571429
49	36	7	3	25	4	4	0,428571	0,428571	0,777778	0,428571
50	20	4	2	14	2	2	0,5	0,5	0,8	0,5
51	37	7	3	26	4	4	0,428571	0,428571	0,783784	0,428571
52	31	6	3	22	3	3	0,5	0,5	0,806452	0,5
53	17	3	2	13	1	1	0,666667	0,666667	0,882353	0,666667
54	37	7	3	26	4	4	0,428571	0,428571	0,783784	0,428571
55	49	10	3	32	7	7	0,3	0,3	0,714286	0,3
56	34	7	3	23	4	4	0,428571	0,428571	0,764706	0,428571
57	38	8	6	28	2	2	0,75	0,75	0,894737	0,75
58	75	15	7	52	8	8	0,466667	0,466667	0,786667	0,466667
59	56	11	3	37	8	8	0,272727	0,272727	0,714286	0,272727
60	31	6	3	22	3	3	0,5	0,5	0,806452	0,5
61	31	6	1	20	5	5	0,166667	0,166667	0,677419	0,166667
62	46	9	3	31	6	6	0,333333	0,333333	0,73913	0,333333
63	49	10	5	34	5	5	0,5	0,5	0,795918	0,5
64	38	8	5	27	3	3	0,625	0,625	0,842105	0,625
65	30	6	2	20	4	4	0,333333	0,333333	0,733333	0,333333
66	19	4	3	14	1	1	0,75	0,75	0,894737	0,75
67	17	3	0	11	3	3	0	0	0,647059	0
68	11	2	1	8	1	1	0,5	0,5	0,818182	0,5
69	24	5	3	17	2	2	0,6	0,6	0,833333	0,6
70	29	6	3	20	3	3	0,5	0,5	0,793103	0,5
71	14	3	0	8	3	3	0	0	0,571429	0
72	22	4	2	16	2	2	0,5	0,5	0,818182	0,5
73	38	8	3	25	5	5	0,375	0,375	0,736842	0,375
74	32	6	4	24	2	2	0,666667	0,666667	0,875	0,666667
75	42	8	4	30	4	4	0,5	0,5	0,809524	0,5
76	21	4	2	15	2	2	0,5	0,5	0,809524	0,5
77	26	5	2	18	3	3	0,4	0,4	0,769231	0,4
78	25	5	2	17	3	3	0,4	0,4	0,76	0,4
79	34	7	2	22	5	5	0,285714	0,285714	0,705882	0,285714
80	27	5	2	19	3	3	0,4	0,4	0,777778	0,4
Rata-rata							0,478397	0,478397	0,792058	0,478397

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

### Lampiran 19 :

Hasil Confusion Matrix - Textrank dengan TF-IDF Cosine *similarity*, dengan

Compression Rate 30%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	6	4	12	2	2	0,666667	0,666667	0,8	0,666667
2	26	8	5	15	3	3	0,625	0,625	0,769231	0,625
3	37	11	5	20	6	6	0,454545	0,454545	0,675676	0,454545
4	41	12	4	21	8	8	0,333333	0,333333	0,609756	0,333333
5	29	9	4	15	5	5	0,444444	0,444444	0,655172	0,444444
6	29	9	5	16	4	4	0,555556	0,555556	0,724138	0,555556
7	47	14	8	27	6	6	0,571429	0,571429	0,744681	0,571429
8	35	10	6	21	4	4	0,6	0,6	0,771429	0,6
9	31	9	6	19	3	3	0,666667	0,666667	0,806452	0,666667
10	20	6	5	13	1	1	0,833333	0,833333	0,9	0,833333
11	39	12	5	20	7	7	0,416667	0,416667	0,641026	0,416667
12	29	9	7	18	2	2	0,777778	0,777778	0,862069	0,777778
13	38	11	7	23	4	4	0,636364	0,636364	0,789474	0,636364
14	36	11	8	22	3	3	0,727273	0,727273	0,833333	0,727273
15	20	6	3	11	3	3	0,5	0,5	0,7	0,5
16	41	12	9	26	3	3	0,75	0,75	0,853659	0,75
17	68	20	15	43	5	5	0,75	0,75	0,852941	0,75
18	64	19	14	40	5	5	0,736842	0,736842	0,84375	0,736842
19	28	8	4	16	4	4	0,5	0,5	0,714286	0,5
20	49	15	7	26	8	8	0,466667	0,466667	0,673469	0,466667
21	45	14	8	25	6	6	0,571429	0,571429	0,733333	0,571429
22	28	8	7	19	1	1	0,875	0,875	0,928571	0,875
23	23	7	4	13	3	3	0,571429	0,571429	0,73913	0,571429
24	25	8	2	11	6	6	0,25	0,25	0,52	0,25
25	31	9	7	20	2	2	0,777778	0,777778	0,870968	0,777778
26	28	8	6	18	2	2	0,75	0,75	0,857143	0,75
27	28	8	7	19	1	1	0,875	0,875	0,928571	0,875
28	30	9	6	18	3	3	0,666667	0,666667	0,8	0,666667
29	20	6	3	11	3	3	0,5	0,5	0,7	0,5
30	32	10	7	19	3	3	0,7	0,7	0,8125	0,7
31	33	10	6	19	4	4	0,6	0,6	0,757576	0,6
32	28	8	5	17	3	3	0,625	0,625	0,785714	0,625
33	40	12	7	23	5	5	0,583333	0,583333	0,75	0,583333
34	34	10	6	20	4	4	0,6	0,6	0,764706	0,6
35	24	7	5	15	2	2	0,714286	0,714286	0,833333	0,714286
36	37	11	8	23	3	3	0,727273	0,727273	0,837838	0,727273
37	25	8	5	14	3	3	0,625	0,625	0,76	0,625
38	24	7	5	15	2	2	0,714286	0,714286	0,833333	0,714286
39	36	11	8	22	3	3	0,727273	0,727273	0,833333	0,727273
40	21	6	4	13	2	2	0,666667	0,666667	0,809524	0,666667
41	32	10	7	19	3	3	0,7	0,7	0,8125	0,7
42	35	10	4	19	6	6	0,4	0,4	0,657143	0,4
43	28	8	4	16	4	4	0,5	0,5	0,714286	0,5
44	24	7	5	15	2	2	0,714286	0,714286	0,833333	0,714286
45	32	10	7	19	3	3	0,7	0,7	0,8125	0,7

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	13	7	24	6	6	0,538462	0,538462	0,72093	0,538462
47	37	11	4	19	7	7	0,363636	0,363636	0,621622	0,363636
48	37	11	6	21	5	5	0,545455	0,545455	0,72973	0,545455
49	36	11	7	21	4	4	0,636364	0,636364	0,777778	0,636364
50	20	6	2	10	4	4	0,333333	0,333333	0,6	0,333333
51	37	11	6	21	5	5	0,545455	0,545455	0,72973	0,545455
52	31	9	7	20	2	2	0,777778	0,777778	0,870968	0,777778
53	17	5	4	11	1	1	0,8	0,8	0,882353	0,8
54	37	11	5	20	6	6	0,454545	0,454545	0,675676	0,454545
55	49	15	8	27	7	7	0,533333	0,533333	0,714286	0,533333
56	34	10	4	18	6	6	0,4	0,4	0,647059	0,4
57	38	11	7	23	4	4	0,636364	0,636364	0,789474	0,636364
58	75	22	15	46	7	7	0,681818	0,681818	0,813333	0,681818
59	56	17	10	32	7	7	0,588235	0,588235	0,75	0,588235
60	31	9	7	20	2	2	0,777778	0,777778	0,870968	0,777778
61	31	9	5	18	4	4	0,555556	0,555556	0,741935	0,555556
62	46	14	7	25	7	7	0,5	0,5	0,695652	0,5
63	49	15	9	28	6	6	0,6	0,6	0,755102	0,6
64	38	11	6	22	5	5	0,545455	0,545455	0,736842	0,545455
65	30	9	4	16	5	5	0,444444	0,444444	0,666667	0,444444
66	19	6	4	11	2	2	0,666667	0,666667	0,789474	0,666667
67	17	5	1	8	4	4	0,2	0,2	0,529412	0,2
68	11	3	1	6	2	2	0,333333	0,333333	0,636364	0,333333
69	24	7	5	15	2	2	0,714286	0,714286	0,833333	0,714286
70	29	9	5	16	4	4	0,555556	0,555556	0,724138	0,555556
71	14	4	0	6	4	4	0	0	0,428571	0
72	22	7	4	12	3	3	0,571429	0,571429	0,727273	0,571429
73	38	11	7	23	4	4	0,636364	0,636364	0,789474	0,636364
74	32	10	6	18	4	4	0,6	0,6	0,75	0,6
75	42	13	8	24	5	5	0,615385	0,615385	0,761905	0,615385
76	21	6	3	12	3	3	0,5	0,5	0,714286	0,5
77	26	8	5	15	3	3	0,625	0,625	0,769231	0,625
78	25	8	5	14	3	3	0,625	0,625	0,76	0,625
79	34	10	5	19	5	5	0,5	0,5	0,705882	0,5
80	27	8	5	16	3	3	0,625	0,625	0,777778	0,625
Rata-rata						0,589666	0,589666	0,754914	0,589666	

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 20 :

Hasil Confusion Matrix - Textrank dengan TF-IDF Cosine *similarity*, dengan

Compression Rate 40%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	8	4	8	4	4	0,5	0,5	0,6	0,5
2	26	10	6	12	4	4	0,6	0,6	0,692308	0,6
3	37	15	9	16	6	6	0,6	0,6	0,675676	0,6
4	41	16	7	16	9	9	0,4375	0,4375	0,560976	0,4375
5	29	12	9	14	3	3	0,75	0,75	0,793103	0,75
6	29	12	8	13	4	4	0,666667	0,666667	0,724138	0,666667
7	47	19	14	23	5	5	0,736842	0,736842	0,787234	0,736842
8	35	14	9	16	5	5	0,642857	0,642857	0,714286	0,642857
9	31	12	8	15	4	4	0,666667	0,666667	0,741935	0,666667
10	20	8	6	10	2	2	0,75	0,75	0,8	0,75
11	39	16	7	14	9	9	0,4375	0,4375	0,538462	0,4375
12	29	12	9	14	3	3	0,75	0,75	0,793103	0,75
13	38	15	10	18	5	5	0,666667	0,666667	0,736842	0,666667
14	36	14	11	19	3	3	0,785714	0,785714	0,833333	0,785714
15	20	8	4	8	4	4	0,5	0,5	0,6	0,5
16	41	16	12	21	4	4	0,75	0,75	0,804878	0,75
17	68	27	23	37	4	4	0,851852	0,851852	0,882353	0,851852
18	64	26	17	29	9	9	0,653846	0,653846	0,71875	0,653846
19	28	11	9	15	2	2	0,818182	0,818182	0,857143	0,818182
20	49	20	14	23	6	6	0,7	0,7	0,755102	0,7
21	45	18	11	20	7	7	0,611111	0,611111	0,688889	0,611111
22	28	11	9	15	2	2	0,818182	0,818182	0,857143	0,818182
23	23	9	6	11	3	3	0,666667	0,666667	0,73913	0,666667
24	25	10	4	9	6	6	0,4	0,4	0,52	0,4
25	31	12	9	16	3	3	0,75	0,75	0,806452	0,75
26	28	11	8	14	3	3	0,727273	0,727273	0,785714	0,727273
27	28	11	9	15	2	2	0,818182	0,818182	0,857143	0,818182
28	30	12	7	13	5	5	0,583333	0,583333	0,666667	0,583333
29	20	8	6	10	2	2	0,75	0,75	0,8	0,75
30	32	13	10	16	3	3	0,769231	0,769231	0,8125	0,769231
31	33	13	7	14	6	6	0,538462	0,538462	0,636364	0,538462
32	28	11	8	14	3	3	0,727273	0,727273	0,785714	0,727273
33	40	16	11	19	5	5	0,6875	0,6875	0,75	0,6875
34	34	14	9	15	5	5	0,642857	0,642857	0,705882	0,642857
35	24	10	8	12	2	2	0,8	0,8	0,833333	0,8
36	37	15	9	16	6	6	0,6	0,6	0,675676	0,6
37	25	10	7	12	3	3	0,7	0,7	0,76	0,7
38	24	10	6	10	4	4	0,6	0,6	0,666667	0,6
39	36	14	9	17	5	5	0,642857	0,642857	0,722222	0,642857
40	21	8	6	11	2	2	0,75	0,75	0,809524	0,75
41	32	13	10	16	3	3	0,769231	0,769231	0,8125	0,769231
42	35	14	8	15	6	6	0,571429	0,571429	0,657143	0,571429
43	28	11	5	11	6	6	0,454545	0,454545	0,571429	0,454545
44	24	10	7	11	3	3	0,7	0,7	0,75	0,7
45	32	13	8	14	5	5	0,615385	0,615385	0,6875	0,615385

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	17	12	21	5	5	0,705882	0,705882	0,767442	0,705882
47	37	15	8	15	7	7	0,533333	0,533333	0,621622	0,533333
48	37	15	7	14	8	8	0,466667	0,466667	0,567568	0,466667
49	36	14	8	16	6	6	0,571429	0,571429	0,666667	0,571429
50	20	8	4	8	4	4	0,5	0,5	0,6	0,5
51	37	15	9	16	6	6	0,6	0,6	0,675676	0,6
52	31	12	9	16	3	3	0,75	0,75	0,806452	0,75
53	17	7	6	9	1	1	0,857143	0,857143	0,882353	0,857143
54	37	15	10	17	5	5	0,666667	0,666667	0,72973	0,666667
55	49	20	12	21	8	8	0,6	0,6	0,673469	0,6
56	34	14	9	15	5	5	0,642857	0,642857	0,705882	0,642857
57	38	15	10	18	5	5	0,666667	0,666667	0,736842	0,666667
58	75	30	20	35	10	10	0,666667	0,666667	0,733333	0,666667
59	56	22	12	24	10	10	0,545455	0,545455	0,642857	0,545455
60	31	12	8	15	4	4	0,666667	0,666667	0,741935	0,666667
61	31	12	7	14	5	5	0,583333	0,583333	0,677419	0,583333
62	46	18	10	20	8	8	0,555556	0,555556	0,652174	0,555556
63	49	20	11	20	9	9	0,55	0,55	0,632653	0,55
64	38	15	9	17	6	6	0,6	0,6	0,684211	0,6
65	30	12	7	13	5	5	0,583333	0,583333	0,666667	0,583333
66	19	8	6	9	2	2	0,75	0,75	0,789474	0,75
67	17	7	5	8	2	2	0,714286	0,714286	0,764706	0,714286
68	11	4	2	5	2	2	0,5	0,5	0,636364	0,5
69	24	10	8	12	2	2	0,8	0,8	0,833333	0,8
70	29	12	9	14	3	3	0,75	0,75	0,793103	0,75
71	14	6	3	5	3	3	0,5	0,5	0,571429	0,5
72	22	9	5	9	4	4	0,555556	0,555556	0,636364	0,555556
73	38	15	11	19	4	4	0,733333	0,733333	0,789474	0,733333
74	32	13	9	15	4	4	0,692308	0,692308	0,75	0,692308
75	42	17	11	19	6	6	0,647059	0,647059	0,714286	0,647059
76	21	8	5	10	3	3	0,625	0,625	0,714286	0,625
77	26	10	6	12	4	4	0,6	0,6	0,692308	0,6
78	25	10	6	11	4	4	0,6	0,6	0,68	0,6
79	34	14	9	15	5	5	0,642857	0,642857	0,705882	0,642857
80	27	11	8	13	3	3	0,727273	0,727273	0,777778	0,727273
Rata-rata							0,651339	0,651339	0,721012	0,651339

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 21 :

Hasil Confusion Matrix - Textrank dengan TF-IDF Cosine *similarity*, dengan

Compression Rate 50%

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
1	20	10	7	7	3	3	0,7	0,7	0,7	0,7
2	26	13	9	9	4	4	0,692308	0,692308	0,692308	0,692308
3	37	18	14	15	4	4	0,777778	0,777778	0,783784	0,777778
4	41	20	11	12	9	9	0,55	0,55	0,560976	0,55
5	29	14	11	12	3	3	0,785714	0,785714	0,793103	0,785714
6	29	14	10	11	4	4	0,714286	0,714286	0,724138	0,714286
7	47	24	18	17	6	6	0,75	0,75	0,744681	0,75
8	35	18	13	12	5	5	0,722222	0,722222	0,714286	0,722222
9	31	16	10	9	6	6	0,625	0,625	0,612903	0,625
10	20	10	8	8	2	2	0,8	0,8	0,8	0,8
11	39	20	12	11	8	8	0,6	0,6	0,589744	0,6
12	29	14	10	11	4	4	0,714286	0,714286	0,724138	0,714286
13	38	19	13	13	6	6	0,684211	0,684211	0,684211	0,684211
14	36	18	14	14	4	4	0,777778	0,777778	0,777778	0,777778
15	20	10	7	7	3	3	0,7	0,7	0,7	0,7
16	41	20	16	17	4	4	0,8	0,8	0,804878	0,8
17	68	34	28	28	6	6	0,823529	0,823529	0,823529	0,823529
18	64	32	24	24	8	8	0,75	0,75	0,75	0,75
19	28	14	10	10	4	4	0,714286	0,714286	0,714286	0,714286
20	49	24	16	17	8	8	0,666667	0,666667	0,673469	0,666667
21	45	22	16	17	6	6	0,727273	0,727273	0,733333	0,727273
22	28	14	11	11	3	3	0,785714	0,785714	0,785714	0,785714
23	23	12	8	7	4	4	0,666667	0,666667	0,652174	0,666667
24	25	12	5	6	7	7	0,416667	0,416667	0,44	0,416667
25	31	16	13	12	3	3	0,8125	0,8125	0,806452	0,8125
26	28	14	12	12	2	2	0,857143	0,857143	0,857143	0,857143
27	28	14	9	9	5	5	0,642857	0,642857	0,642857	0,642857
28	30	15	10	10	5	5	0,666667	0,666667	0,666667	0,666667
29	20	10	6	6	4	4	0,6	0,6	0,6	0,6
30	32	16	13	13	3	3	0,8125	0,8125	0,8125	0,8125
31	33	16	10	11	6	6	0,625	0,625	0,636364	0,625
32	28	14	11	11	3	3	0,785714	0,785714	0,785714	0,785714
33	40	20	15	15	5	5	0,75	0,75	0,75	0,75
34	34	17	11	11	6	6	0,647059	0,647059	0,647059	0,647059
35	24	12	9	9	3	3	0,75	0,75	0,75	0,75
36	37	18	13	14	5	5	0,722222	0,722222	0,72973	0,722222
37	25	12	8	9	4	4	0,666667	0,666667	0,68	0,666667
38	24	12	7	7	5	5	0,583333	0,583333	0,583333	0,583333
39	36	18	12	12	6	6	0,666667	0,666667	0,666667	0,666667
40	21	10	7	8	3	3	0,7	0,7	0,714286	0,7
41	32	16	13	13	3	3	0,8125	0,8125	0,8125	0,8125
42	35	18	11	10	7	7	0,611111	0,611111	0,6	0,611111
43	28	14	6	6	8	8	0,428571	0,428571	0,428571	0,428571
44	24	12	8	8	4	4	0,666667	0,666667	0,666667	0,666667
45	32	16	11	11	5	5	0,6875	0,6875	0,6875	0,6875

No	A	R	TP	TN	FP	FN	Precision	Recall	Accuracy	F-Score
46	43	22	14	13	8	8	0,636364	0,636364	0,627907	0,636364
47	37	18	11	12	7	7	0,611111	0,611111	0,621622	0,611111
48	37	18	12	13	6	6	0,666667	0,666667	0,675676	0,666667
49	36	18	12	12	6	6	0,666667	0,666667	0,666667	0,666667
50	20	10	7	7	3	3	0,7	0,7	0,7	0,7
51	37	18	11	12	7	7	0,611111	0,611111	0,621622	0,611111
52	31	16	14	13	2	2	0,875	0,875	0,870968	0,875
53	17	8	7	8	1	1	0,875	0,875	0,882353	0,875
54	37	18	11	12	7	7	0,611111	0,611111	0,621622	0,611111
55	49	24	15	16	9	9	0,625	0,625	0,632653	0,625
56	34	17	11	11	6	6	0,647059	0,647059	0,647059	0,647059
57	38	19	15	15	4	4	0,789474	0,789474	0,789474	0,789474
58	75	38	27	26	11	11	0,710526	0,710526	0,706667	0,710526
59	56	28	18	18	10	10	0,642857	0,642857	0,642857	0,642857
60	31	16	12	11	4	4	0,75	0,75	0,741935	0,75
61	31	16	10	9	6	6	0,625	0,625	0,612903	0,625
62	46	23	17	17	6	6	0,73913	0,73913	0,73913	0,73913
63	49	24	15	16	9	9	0,625	0,625	0,632653	0,625
64	38	19	14	14	5	5	0,736842	0,736842	0,736842	0,736842
65	30	15	9	9	6	6	0,6	0,6	0,6	0,6
66	19	10	8	7	2	2	0,8	0,8	0,789474	0,8
67	17	8	5	6	3	3	0,625	0,625	0,647059	0,625
68	11	6	6	5	0	0	1	1	1	1
69	24	12	11	11	1	1	0,916667	0,916667	0,916667	0,916667
70	29	14	11	12	3	3	0,785714	0,785714	0,793103	0,785714
71	14	7	3	3	4	4	0,428571	0,428571	0,428571	0,428571
72	22	11	6	6	5	5	0,545455	0,545455	0,545455	0,545455
73	38	19	17	17	2	2	0,894737	0,894737	0,894737	0,894737
74	32	16	13	13	3	3	0,8125	0,8125	0,8125	0,8125
75	42	21	15	15	6	6	0,714286	0,714286	0,714286	0,714286
76	21	10	7	8	3	3	0,7	0,7	0,714286	0,7
77	26	13	9	9	4	4	0,692308	0,692308	0,692308	0,692308
78	25	12	8	9	4	4	0,666667	0,666667	0,68	0,666667
79	34	17	12	12	5	5	0,705882	0,705882	0,705882	0,705882
80	27	14	11	10	3	3	0,785714	0,785714	0,777778	0,785714
Rata-rata						0,703206	0,703206	0,704827	0,703206	

Keterangan : A merupakan jumlah kalimat pada teks asli, R merupakan jumlah kalimat pada teks ringkasan, TP merupakan nilai True Positif, TN merupakan nilai True Negatif, FP merupakan nilai False Positif dan FN merupakan nilai False Negatif.

## Lampiran 22 :

Hasil ROUGE 1 - Textrank dengan *similarity word overlapping*, dengan

Compression Rate10%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	59	55	35	0,636364	0,59322	0,614035
2	450	43	75	17	0,226667	0,395349	0,288136
3	565	101	76	40	0,526316	0,39604	0,451977
4	528	70	51	28	0,54902	0,4	0,46281
5	657	92	81	66	0,814815	0,717391	0,763006
6	447	62	41	28	0,682927	0,451613	0,543689
7	966	38	156	20	0,128205	0,526316	0,206186
8	509	95	93	83	0,892473	0,873684	0,882979
9	386	43	46	34	0,73913	0,790698	0,764045
10	368	77	69	51	0,73913	0,662338	0,69863
11	528	68	57	34	0,596491	0,5	0,544
12	513	60	75	32	0,426667	0,533333	0,474074
13	644	72	85	44	0,517647	0,611111	0,56051
14	534	64	84	17	0,202381	0,265625	0,22973
15	285	47	47	47	1	1	1
16	657	88	81	59	0,728395	0,670455	0,698225
17	799	94	119	67	0,563025	0,712766	0,629108
18	1029	114	140	104	0,742857	0,912281	0,818898
19	340	18	54	7	0,12963	0,388889	0,194444
20	695	79	83	51	0,614458	0,64557	0,62963
21	574	50	52	34	0,653846	0,68	0,666667
22	532	63	74	35	0,472973	0,555556	0,510949
23	321	45	24	12	0,5	0,266667	0,347826
24	373	32	32	17	0,53125	0,53125	0,53125
25	366	33	33	11	0,333333	0,333333	0,333333
26	329	42	36	5	0,138889	0,119048	0,128205
27	383	37	56	18	0,321429	0,486486	0,387097
28	359	42	41	31	0,756098	0,738095	0,746988
29	352	34	48	10	0,208333	0,294118	0,243902
30	409	41	41	20	0,487805	0,487805	0,487805
31	592	64	62	40	0,645161	0,625	0,634921
32	468	62	56	38	0,678571	0,612903	0,644068
33	481	33	36	14	0,388889	0,424242	0,405797
34	331	30	32	15	0,46875	0,5	0,483871
35	364	30	36	22	0,611111	0,733333	0,666667
36	436	45	49	30	0,612245	0,666667	0,638298
37	320	28	25	9	0,36	0,321429	0,339623
38	350	37	35	27	0,771429	0,72973	0,75
39	506	77	77	77	1	1	1
40	380	44	33	10	0,30303	0,227273	0,25974
41	417	52	48	22	0,458333	0,423077	0,44
42	571	84	77	38	0,493506	0,452381	0,47205
43	376	61	38	24	0,631579	0,393443	0,484848
44	666	72	72	72	1	1	1
45	432	32	48	25	0,520833	0,78125	0,625

No	A	GT	RP	O	Precision	Recall	F-score
46	454	43	50	12	0,24	0,27907	0,258065
47	472	61	49	22	0,44898	0,360656	0,4
48	576	94	70	50	0,714286	0,531915	0,609756
49	556	67	66	43	0,651515	0,641791	0,646617
50	405	43	52	10	0,192308	0,232558	0,210526
51	554	94	64	35	0,546875	0,37234	0,443038
52	518	58	58	37	0,637931	0,637931	0,637931
53	393	39	39	39	1	1	1
54	394	57	57	57	1	1	1
55	931	88	95	38	0,4	0,431818	0,415301
56	390	45	39	24	0,615385	0,533333	0,571429
57	637	62	56	49	0,875	0,790323	0,830508
58	1103	152	131	65	0,496183	0,427632	0,459364
59	652	106	77	31	0,402597	0,292453	0,338798
60	387	30	34	20	0,588235	0,666667	0,625
61	395	58	56	26	0,464286	0,448276	0,45614
62	666	69	84	29	0,345238	0,42029	0,379085
63	759	58	84	44	0,52381	0,758621	0,619718
64	465	55	46	26	0,565217	0,472727	0,514851
65	432	70	67	51	0,761194	0,728571	0,744526
66	359	27	40	17	0,425	0,62963	0,507463
67	278	40	23	9	0,391304	0,225	0,285714
68	225	15	15	15	1	1	1
69	434	70	47	27	0,574468	0,385714	0,461538
70	494	58	64	50	0,78125	0,862069	0,819672
71	219	21	20	3	0,15	0,142857	0,146341
72	385	29	30	19	0,633333	0,655172	0,644068
73	874	84	73	43	0,589041	0,511905	0,547771
74	523	40	53	25	0,471698	0,625	0,537634
75	517	46	45	24	0,533333	0,521739	0,527473
76	370	32	55	10	0,181818	0,3125	0,229885
77	335	55	56	16	0,285714	0,290909	0,288288
78	478	50	54	14	0,259259	0,28	0,269231
79	530	60	68	17	0,25	0,283333	0,265625
80	328	44	51	23	0,45098	0,522727	0,484211
	Rata-rata				0,540628	0,546316	0,535732

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

### Lampiran 23 :

Hasil ROUGE 1 - Textrank dengan *similarity word overlapping*, dengan

Compression Rate 20%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	101	104	90	0,865385	0,891089	0,878049
2	450	91	119	63	0,529412	0,692308	0,6
3	565	139	131	64	0,48855	0,460432	0,474074
4	528	126	120	64	0,533333	0,507937	0,520325
5	657	152	160	89	0,55625	0,585526	0,570513
6	447	95	92	80	0,869565	0,842105	0,855615
7	966	111	235	89	0,378723	0,801802	0,514451
8	509	157	145	117	0,806897	0,745223	0,774834
9	386	100	103	76	0,737864	0,76	0,748768
10	368	112	99	78	0,787879	0,696429	0,739336
11	528	123	108	71	0,657407	0,577236	0,614719
12	513	148	129	91	0,705426	0,614865	0,65704
13	644	148	155	121	0,780645	0,817568	0,79868
14	534	130	138	97	0,702899	0,746154	0,723881
15	285	75	84	70	0,833333	0,933333	0,880503
16	657	149	142	126	0,887324	0,845638	0,865979
17	799	198	221	170	0,769231	0,858586	0,811456
18	1029	218	260	184	0,707692	0,844037	0,769874
19	340	56	90	45	0,5	0,803571	0,616438
20	695	164	180	139	0,772222	0,847561	0,80814
21	574	138	123	97	0,788618	0,702899	0,743295
22	532	133	138	103	0,746377	0,774436	0,760148
23	321	88	67	55	0,820896	0,625	0,709677
24	373	82	87	46	0,528736	0,560976	0,544379
25	366	70	65	46	0,707692	0,657143	0,681481
26	329	89	89	58	0,651685	0,651685	0,651685
27	383	88	87	53	0,609195	0,602273	0,605714
28	359	74	85	44	0,517647	0,594595	0,553459
29	352	74	76	30	0,394737	0,405405	0,4
30	409	84	88	71	0,806818	0,845238	0,825581
31	592	162	142	97	0,683099	0,598765	0,638158
32	468	124	107	78	0,728972	0,629032	0,675325
33	481	83	84	58	0,690476	0,698795	0,694611
34	331	80	66	34	0,515152	0,425	0,465753
35	364	62	72	44	0,611111	0,709677	0,656716
36	436	86	88	60	0,681818	0,697674	0,689655
37	320	74	65	58	0,892308	0,783784	0,834532
38	350	77	86	54	0,627907	0,701299	0,662577
39	506	130	131	120	0,916031	0,923077	0,91954
40	380	89	90	55	0,611111	0,617978	0,614525
41	417	99	105	80	0,761905	0,808081	0,784314
42	571	124	134	92	0,686567	0,741935	0,713178
43	376	114	80	53	0,6625	0,464912	0,546392
44	666	187	189	153	0,809524	0,818182	0,81383
45	432	84	99	68	0,686869	0,809524	0,743169

No	A	GT	RP	O	Precision	Recall	F-score
46	454	121	103	58	0,563107	0,479339	0,517857
47	472	92	95	51	0,536842	0,554348	0,545455
48	576	143	128	109	0,851563	0,762238	0,804428
49	556	138	119	90	0,756303	0,652174	0,700389
50	405	114	85	64	0,752941	0,561404	0,643216
51	554	143	91	77	0,846154	0,538462	0,65812
52	518	122	114	111	0,973684	0,909836	0,940678
53	393	62	72	50	0,694444	0,806452	0,746269
54	394	105	90	74	0,822222	0,704762	0,758974
55	931	241	226	164	0,725664	0,680498	0,702355
56	390	92	86	53	0,616279	0,576087	0,595506
57	637	142	148	128	0,864865	0,901408	0,882759
58	1103	286	267	176	0,659176	0,615385	0,636528
59	652	179	142	76	0,535211	0,424581	0,47352
60	387	78	72	32	0,444444	0,410256	0,426667
61	395	134	86	48	0,55814	0,358209	0,436364
62	666	137	145	80	0,551724	0,583942	0,567376
63	759	158	166	96	0,578313	0,607595	0,592593
64	465	104	108	55	0,509259	0,528846	0,518868
65	432	113	104	68	0,653846	0,60177	0,626728
66	359	68	87	56	0,643678	0,823529	0,722581
67	278	56	43	13	0,302326	0,232143	0,262626
68	225	52	37	19	0,513514	0,365385	0,426966
69	434	148	115	71	0,617391	0,47973	0,539924
70	494	118	128	90	0,703125	0,762712	0,731707
71	219	65	60	32	0,533333	0,492308	0,512
72	385	74	55	38	0,690909	0,513514	0,589147
73	874	187	175	125	0,714286	0,668449	0,690608
74	523	87	100	70	0,7	0,804598	0,748663
75	517	106	99	54	0,545455	0,509434	0,526829
76	370	88	93	51	0,548387	0,579545	0,563536
77	335	90	101	60	0,594059	0,666667	0,628272
78	478	114	128	62	0,484375	0,54386	0,512397
79	530	151	148	98	0,662162	0,649007	0,655518
80	328	70	82	39	0,47561	0,557143	0,513158
	Rata-rata				0,665382	0,657805	0,656862

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

## Lampiran 24 :

Hasil ROUGE 1 - Textrank dengan *similarity word overlapping*, dengan

Compression Rate 30%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	131	135	122	0,903704	0,931298	0,917293
2	450	152	162	115	0,709877	0,756579	0,732484
3	565	200	179	122	0,681564	0,61	0,643799
4	528	174	184	107	0,581522	0,614943	0,597765
5	657	259	238	195	0,819328	0,752896	0,784708
6	447	144	132	103	0,780303	0,715278	0,746377
7	966	254	347	218	0,628242	0,858268	0,725458
8	509	188	194	146	0,752577	0,776596	0,764398
9	386	157	136	119	0,875	0,757962	0,812287
10	368	144	132	115	0,871212	0,798611	0,833333
11	528	198	183	120	0,655738	0,606061	0,629921
12	513	194	199	129	0,648241	0,664948	0,656489
13	644	221	228	177	0,776316	0,800905	0,788419
14	534	206	197	162	0,822335	0,786408	0,80397
15	285	105	101	77	0,762376	0,733333	0,747573
16	657	246	209	176	0,842105	0,715447	0,773626
17	799	313	298	244	0,818792	0,779553	0,798691
18	1029	313	369	264	0,715447	0,84345	0,774194
19	340	102	115	88	0,765217	0,862745	0,81106
20	695	244	252	175	0,694444	0,717213	0,705645
21	574	203	190	130	0,684211	0,640394	0,661578
22	532	171	181	131	0,723757	0,766082	0,744318
23	321	124	115	102	0,886957	0,822581	0,853556
24	373	153	125	69	0,552	0,45098	0,496403
25	366	121	98	75	0,765306	0,619835	0,684932
26	329	117	114	84	0,736842	0,717949	0,727273
27	383	115	112	80	0,714286	0,695652	0,704846
28	359	118	131	73	0,557252	0,618644	0,586345
29	352	123	125	61	0,488	0,495935	0,491935
30	409	157	158	110	0,696203	0,700637	0,698413
31	592	223	218	190	0,87156	0,852018	0,861678
32	468	155	152	100	0,657895	0,645161	0,651466
33	481	137	133	89	0,669173	0,649635	0,659259
34	331	114	98	64	0,653061	0,561404	0,603774
35	364	98	101	74	0,732673	0,755102	0,743719
36	436	152	136	116	0,852941	0,763158	0,805556
37	320	136	110	88	0,8	0,647059	0,715447
38	350	123	118	99	0,838983	0,804878	0,821577
39	506	212	166	139	0,837349	0,65566	0,73545
40	380	141	140	124	0,885714	0,879433	0,882562
41	417	176	154	139	0,902597	0,789773	0,842424
42	571	151	193	102	0,528497	0,675497	0,593023
43	376	142	118	80	0,677966	0,56338	0,615385
44	666	276	256	239	0,933594	0,865942	0,898496
45	432	160	139	130	0,935252	0,8125	0,869565

No	A	GT	RP	O	Precision	Recall	F-score
46	454	161	151	101	0,668874	0,627329	0,647436
47	472	125	158	75	0,474684	0,6	0,530035
48	576	216	183	154	0,84153	0,712963	0,77193
49	556	216	191	159	0,832461	0,736111	0,781327
50	405	154	116	86	0,741379	0,558442	0,637037
51	554	215	183	157	0,857923	0,730233	0,788945
52	518	162	188	138	0,734043	0,851852	0,788571
53	393	120	120	84	0,7	0,7	0,7
54	394	148	145	105	0,724138	0,709459	0,716724
55	931	364	285	245	0,859649	0,673077	0,755008
56	390	130	129	78	0,604651	0,6	0,602317
57	637	205	204	175	0,857843	0,853659	0,855746
58	1103	448	394	318	0,807107	0,709821	0,755344
59	652	249	223	160	0,717489	0,64257	0,677966
60	387	116	114	86	0,754386	0,741379	0,747826
61	395	189	147	132	0,897959	0,698413	0,785714
62	666	215	207	156	0,753623	0,725581	0,739336
63	759	259	246	188	0,764228	0,725869	0,744554
64	465	151	141	90	0,638298	0,596026	0,616438
65	432	164	149	124	0,832215	0,756098	0,792332
66	359	112	123	100	0,813008	0,892857	0,851064
67	278	105	83	51	0,614458	0,485714	0,542553
68	225	74	61	44	0,721311	0,594595	0,651852
69	434	200	168	156	0,928571	0,78	0,847826
70	494	174	170	131	0,770588	0,752874	0,761628
71	219	93	65	38	0,584615	0,408602	0,481013
72	385	110	117	70	0,598291	0,636364	0,61674
73	874	247	243	198	0,814815	0,801619	0,808163
74	523	170	141	100	0,70922	0,588235	0,643087
75	517	176	166	116	0,698795	0,659091	0,678363
76	370	124	127	68	0,535433	0,548387	0,541833
77	335	146	144	120	0,833333	0,821918	0,827586
78	478	178	192	111	0,578125	0,623596	0,6
79	530	214	207	139	0,671498	0,649533	0,660333
80	328	123	120	98	0,816667	0,796748	0,806584
	Rata-rata				0,74297	0,70651	0,721871

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

## Lampiran 25 :

Hasil ROUGE 1 - Textrank dengan *similarity word overlapping*, dengan

Compression Rate 40%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	164	158	137	0,867089	0,835366	0,850932
2	450	185	194	158	0,814433	0,854054	0,833773
3	565	267	261	216	0,827586	0,808989	0,818182
4	528	239	237	157	0,662447	0,656904	0,659664
5	657	299	295	269	0,911864	0,899666	0,905724
6	447	200	188	159	0,845745	0,795	0,819588
7	966	360	476	313	0,657563	0,869444	0,748804
8	509	246	248	201	0,810484	0,817073	0,813765
9	386	179	182	145	0,796703	0,810056	0,803324
10	368	180	164	141	0,859756	0,783333	0,819767
11	528	271	243	180	0,740741	0,664207	0,700389
12	513	272	252	223	0,884921	0,819853	0,851145
13	644	311	307	262	0,85342	0,842444	0,847896
14	534	263	247	210	0,850202	0,798479	0,823529
15	285	133	136	101	0,742647	0,759398	0,750929
16	657	306	257	212	0,824903	0,69281	0,753108
17	799	414	407	356	0,874693	0,859903	0,867235
18	1029	447	498	382	0,767068	0,854586	0,808466
19	340	132	146	119	0,815068	0,901515	0,856115
20	695	330	336	268	0,797619	0,812121	0,804805
21	574	265	242	178	0,735537	0,671698	0,70217
22	532	247	239	196	0,820084	0,793522	0,806584
23	321	162	137	119	0,868613	0,734568	0,795987
24	373	179	176	103	0,585227	0,575419	0,580282
25	366	159	144	96	0,666667	0,603774	0,633663
26	329	155	156	135	0,865385	0,870968	0,868167
27	383	167	150	108	0,72	0,646707	0,681388
28	359	158	169	112	0,662722	0,708861	0,685015
29	352	171	163	118	0,723926	0,690058	0,706587
30	409	209	194	172	0,886598	0,822967	0,853598
31	592	277	279	240	0,860215	0,866426	0,863309
32	468	224	187	162	0,86631	0,723214	0,788321
33	481	185	187	144	0,770053	0,778378	0,774194
34	331	160	144	123	0,854167	0,76875	0,809211
35	364	158	131	108	0,824427	0,683544	0,747405
36	436	186	183	139	0,759563	0,747312	0,753388
37	320	161	149	114	0,765101	0,708075	0,735484
38	350	158	163	121	0,742331	0,765823	0,753894
39	506	276	227	174	0,76652	0,630435	0,691849
40	380	167	175	131	0,748571	0,784431	0,766082
41	417	231	214	202	0,943925	0,874459	0,907865
42	571	229	260	176	0,676923	0,768559	0,719836
43	376	170	165	103	0,624242	0,605882	0,614925
44	666	386	392	311	0,793367	0,805699	0,799486
45	432	202	195	170	0,871795	0,841584	0,856423

No	A	GT	RP	O	Precision	Recall	F-score
46	454	201	203	141	0,694581	0,701493	0,69802
47	472	171	211	134	0,635071	0,783626	0,701571
48	576	272	245	208	0,84898	0,764706	0,804642
49	556	262	248	198	0,798387	0,755725	0,776471
50	405	179	146	116	0,794521	0,648045	0,713846
51	554	274	231	195	0,844156	0,711679	0,772277
52	518	215	240	180	0,75	0,837209	0,791209
53	393	168	179	147	0,821229	0,875	0,847262
54	394	198	184	156	0,847826	0,787879	0,816754
55	931	477	433	394	0,909931	0,825996	0,865934
56	390	174	171	132	0,77193	0,758621	0,765217
57	637	304	290	273	0,941379	0,898026	0,919192
58	1103	561	561	475	0,846702	0,846702	0,846702
59	652	313	287	222	0,773519	0,709265	0,74
60	387	161	161	108	0,670807	0,670807	0,670807
61	395	236	192	172	0,895833	0,728814	0,803738
62	666	281	254	192	0,755906	0,683274	0,717757
63	759	355	325	255	0,784615	0,71831	0,75
64	465	209	197	151	0,766497	0,722488	0,743842
65	432	231	188	163	0,867021	0,705628	0,778043
66	359	152	160	106	0,6625	0,697368	0,679487
67	278	134	116	80	0,689655	0,597015	0,64
68	225	93	85	52	0,611765	0,55914	0,58427
69	434	235	239	225	0,941423	0,957447	0,949367
70	494	240	223	187	0,838565	0,779167	0,807775
71	219	123	99	82	0,828283	0,666667	0,738739
72	385	134	161	102	0,63354	0,761194	0,691525
73	874	349	342	272	0,795322	0,77937	0,787265
74	523	214	183	143	0,781421	0,668224	0,720403
75	517	242	212	174	0,820755	0,719008	0,76652
76	370	172	169	118	0,698225	0,686047	0,692082
77	335	178	178	145	0,814607	0,814607	0,814607
78	478	240	224	180	0,803571	0,75	0,775862
79	530	294	269	212	0,788104	0,721088	0,753108
80	328	173	178	147	0,825843	0,849711	0,837607
	Rata-rata				0,789496	0,759321	0,772427

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

## Lampiran 26 :

Hasil ROUGE 1 - Textrank dengan *similarity word overlapping*, dengan

Compression Rate 50%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	194	199	185	0,929648	0,953608	0,941476
2	450	233	263	196	0,745247	0,841202	0,790323
3	565	308	301	251	0,833887	0,814935	0,824302
4	528	295	302	224	0,741722	0,759322	0,750419
5	657	345	359	295	0,821727	0,855072	0,838068
6	447	221	228	173	0,758772	0,782805	0,770601
7	966	461	561	410	0,730838	0,889371	0,802348
8	509	301	304	269	0,884868	0,893688	0,889256
9	386	218	245	180	0,734694	0,825688	0,777538
10	368	219	203	178	0,876847	0,812785	0,843602
11	528	338	290	244	0,841379	0,721893	0,77707
12	513	320	292	250	0,856164	0,78125	0,816993
13	644	365	375	304	0,810667	0,832877	0,821622
14	534	345	316	298	0,943038	0,863768	0,901664
15	285	160	158	121	0,765823	0,75625	0,761006
16	657	352	337	290	0,860534	0,823864	0,8418
17	799	507	515	462	0,897087	0,911243	0,90411
18	1029	560	585	455	0,777778	0,8125	0,79476
19	340	171	176	140	0,795455	0,818713	0,806916
20	695	404	395	345	0,873418	0,85396	0,863579
21	574	323	301	239	0,79402	0,739938	0,766026
22	532	310	298	248	0,832215	0,8	0,815789
23	321	201	195	158	0,810256	0,78607	0,79798
24	373	204	202	141	0,69802	0,691176	0,694581
25	366	222	228	177	0,776316	0,797297	0,786667
26	329	192	191	172	0,900524	0,895833	0,898172
27	383	214	170	141	0,829412	0,658879	0,734375
28	359	202	207	166	0,801932	0,821782	0,811736
29	352	221	193	142	0,735751	0,642534	0,68599
30	409	247	240	222	0,925	0,898785	0,911704
31	592	356	338	287	0,849112	0,80618	0,827089
32	468	290	253	219	0,865613	0,755172	0,80663
33	481	244	247	194	0,785425	0,795082	0,790224
34	331	197	175	141	0,805714	0,715736	0,758065
35	364	196	164	140	0,853659	0,714286	0,777778
36	436	220	211	164	0,777251	0,745455	0,761021
37	320	186	173	136	0,786127	0,731183	0,75766
38	350	201	186	141	0,758065	0,701493	0,728682
39	506	325	312	257	0,823718	0,790769	0,806907
40	380	219	208	157	0,754808	0,716895	0,735363
41	417	283	249	228	0,915663	0,805654	0,857143
42	571	308	337	248	0,735905	0,805195	0,768992
43	376	184	213	127	0,596244	0,690217	0,639798
44	666	476	425	392	0,922353	0,823529	0,870144
45	432	250	224	202	0,901786	0,808	0,852321

No	A	GT	RP	O	Precision	Recall	F-score
46	454	236	256	174	0,679688	0,737288	0,707317
47	472	222	249	191	0,767068	0,86036	0,81104
48	576	322	287	255	0,888502	0,791925	0,837438
49	556	313	333	266	0,798799	0,84984	0,823529
50	405	210	190	168	0,884211	0,8	0,84
51	554	300	262	228	0,870229	0,76	0,811388
52	518	310	316	286	0,905063	0,922581	0,913738
53	393	213	221	170	0,769231	0,798122	0,78341
54	394	225	208	172	0,826923	0,764444	0,794457
55	931	563	507	429	0,846154	0,761989	0,801869
56	390	202	200	151	0,755	0,747525	0,751244
57	637	380	363	333	0,917355	0,876316	0,896366
58	1103	697	647	571	0,882535	0,819225	0,849702
59	652	374	356	283	0,794944	0,756684	0,775342
60	387	214	215	179	0,832558	0,836449	0,834499
61	395	284	259	246	0,949807	0,866197	0,906077
62	666	364	325	268	0,824615	0,736264	0,777939
63	759	412	391	309	0,790281	0,75	0,769614
64	465	274	258	197	0,763566	0,718978	0,740602
65	432	265	246	197	0,800813	0,743396	0,771037
66	359	191	197	148	0,751269	0,774869	0,762887
67	278	146	140	111	0,792857	0,760274	0,776224
68	225	133	128	91	0,710938	0,684211	0,697318
69	434	268	267	243	0,910112	0,906716	0,908411
70	494	292	253	220	0,869565	0,753425	0,807339
71	219	129	127	105	0,826772	0,813953	0,820313
72	385	191	194	121	0,623711	0,633508	0,628571
73	874	447	426	369	0,866197	0,825503	0,845361
74	523	270	230	205	0,891304	0,759259	0,82
75	517	284	255	225	0,882353	0,792254	0,834879
76	370	211	210	170	0,809524	0,805687	0,807601
77	335	227	220	189	0,859091	0,832599	0,845638
78	478	276	264	220	0,833333	0,797101	0,814815
79	530	339	328	282	0,859756	0,831858	0,845577
80	328	229	218	196	0,899083	0,855895	0,876957
	Rata-rata				0,820596	0,793333	0,805585

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

### Lampiran 27 :

Hasil ROUGE 1 - Textrank dengan Jaccard *similarity*, dengan Compression Rate 10%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	59	35	12	0,342857	0,20339	0,255319
2	450	43	51	23	0,45098	0,534884	0,489362
3	565	101	53	21	0,396226	0,207921	0,272727
4	528	70	45	28	0,622222	0,4	0,486957
5	657	92	57	36	0,631579	0,391304	0,483221
6	447	62	41	28	0,682927	0,451613	0,543689
7	966	38	117	22	0,188034	0,578947	0,283871
8	509	95	77	55	0,714286	0,578947	0,639535
9	386	43	46	34	0,73913	0,790698	0,764045
10	368	77	32	14	0,4375	0,181818	0,256881
11	528	68	43	19	0,44186	0,279412	0,342342
12	513	60	54	20	0,37037	0,333333	0,350877
13	644	72	61	38	0,622951	0,527778	0,571429
14	534	64	64	16	0,25	0,25	0,25
15	285	47	17	8	0,470588	0,170213	0,25
16	657	88	42	28	0,666667	0,318182	0,430769
17	799	94	88	64	0,727273	0,680851	0,703297
18	1029	114	66	44	0,666667	0,385965	0,488889
19	340	18	39	5	0,128205	0,277778	0,175439
20	695	79	82	35	0,426829	0,443038	0,434783
21	574	50	42	22	0,52381	0,44	0,478261
22	532	63	50	29	0,58	0,460317	0,513274
23	321	45	24	12	0,5	0,266667	0,347826
24	373	32	32	7	0,21875	0,21875	0,21875
25	366	33	33	11	0,333333	0,333333	0,333333
26	329	42	37	7	0,189189	0,166667	0,177215
27	383	37	41	18	0,439024	0,486486	0,461538
28	359	42	35	17	0,485714	0,404762	0,441558
29	352	34	46	9	0,195652	0,264706	0,225
30	409	41	34	28	0,823529	0,682927	0,746667
31	592	64	66	37	0,560606	0,578125	0,569231
32	468	62	50	24	0,48	0,387097	0,428571
33	481	33	36	14	0,388889	0,424242	0,405797
34	331	30	32	15	0,46875	0,5	0,483871
35	364	30	26	13	0,5	0,433333	0,464286
36	436	45	43	34	0,790698	0,755556	0,772727
37	320	28	20	8	0,4	0,285714	0,333333
38	350	37	23	9	0,391304	0,243243	0,3
39	506	77	40	24	0,6	0,311688	0,410256
40	380	44	33	10	0,30303	0,227273	0,25974
41	417	52	52	23	0,442308	0,442308	0,442308
42	571	84	77	38	0,493506	0,452381	0,47205
43	376	61	37	15	0,405405	0,245902	0,306122
44	666	72	72	72	1	1	1
45	432	32	48	25	0,520833	0,78125	0,625

No	A	GT	RP	O	Precision	Recall	F-score
46	454	43	50	21	0,42	0,488372	0,451613
47	472	61	48	21	0,4375	0,344262	0,385321
48	576	94	56	36	0,642857	0,382979	0,48
49	556	67	66	31	0,469697	0,462687	0,466165
50	405	43	24	6	0,25	0,139535	0,179104
51	554	94	45	35	0,777778	0,37234	0,503597
52	518	58	41	25	0,609756	0,431034	0,505051
53	393	39	39	39	1	1	1
54	394	57	38	19	0,5	0,333333	0,4
55	931	88	65	39	0,6	0,443182	0,509804
56	390	45	33	20	0,606061	0,444444	0,512821
57	637	62	56	49	0,875	0,790323	0,830508
58	1103	152	123	62	0,504065	0,407895	0,450909
59	652	106	78	24	0,307692	0,226415	0,26087
60	387	30	27	20	0,740741	0,666667	0,701754
61	395	58	28	21	0,75	0,362069	0,488372
62	666	69	60	24	0,4	0,347826	0,372093
63	759	58	66	47	0,712121	0,810345	0,758065
64	465	55	46	26	0,565217	0,472727	0,514851
65	432	70	38	31	0,815789	0,442857	0,574074
66	359	27	27	16	0,592593	0,592593	0,592593
67	278	40	23	9	0,391304	0,225	0,285714
68	225	15	15	15	1	1	1
69	434	70	47	27	0,574468	0,385714	0,461538
70	494	58	58	35	0,603448	0,603448	0,603448
71	219	21	20	3	0,15	0,142857	0,146341
72	385	29	29	21	0,724138	0,724138	0,724138
73	874	84	73	43	0,589041	0,511905	0,547771
74	523	40	40	20	0,5	0,5	0,5
75	517	46	36	29	0,805556	0,630435	0,707317
76	370	32	55	10	0,181818	0,3125	0,229885
77	335	55	48	16	0,333333	0,290909	0,31068
78	478	50	54	14	0,259259	0,28	0,269231
79	530	60	60	19	0,316667	0,316667	0,316667
80	328	44	42	25	0,595238	0,568182	0,581395
	Rata-rata				0,520158	0,440405	0,466398

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

### Lampiran 28 :

Hasil ROUGE 1 - Textrank dengan Jaccard *similarity*, dengan Compression Rate 20%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	101	73	51	0,69863	0,50495	0,586207
2	450	91	87	46	0,528736	0,505495	0,516854
3	565	139	94	64	0,680851	0,460432	0,549356
4	528	126	96	60	0,625	0,47619	0,540541
5	657	152	121	84	0,694215	0,552632	0,615385
6	447	95	58	45	0,775862	0,473684	0,588235
7	966	111	235	89	0,378723	0,801802	0,514451
8	509	157	107	84	0,785047	0,535032	0,636364
9	386	100	82	55	0,670732	0,55	0,604396
10	368	112	67	49	0,731343	0,4375	0,547486
11	528	123	103	79	0,76699	0,642276	0,699115
12	513	148	114	63	0,552632	0,425676	0,480916
13	644	148	142	105	0,739437	0,709459	0,724138
14	534	130	123	74	0,601626	0,569231	0,58498
15	285	75	53	33	0,622642	0,44	0,515625
16	657	149	121	98	0,809917	0,657718	0,725926
17	799	198	199	169	0,849246	0,853535	0,851385
18	1029	218	219	174	0,794521	0,798165	0,796339
19	340	56	62	46	0,741935	0,821429	0,779661
20	695	164	164	107	0,652439	0,652439	0,652439
21	574	138	123	97	0,788618	0,702899	0,743295
22	532	133	124	94	0,758065	0,706767	0,731518
23	321	88	66	41	0,621212	0,465909	0,532468
24	373	82	75	41	0,546667	0,5	0,522293
25	366	70	61	36	0,590164	0,514286	0,549618
26	329	89	75	40	0,533333	0,449438	0,487805
27	383	88	87	53	0,609195	0,602273	0,605714
28	359	74	85	44	0,517647	0,594595	0,553459
29	352	74	78	28	0,358974	0,378378	0,368421
30	409	84	77	64	0,831169	0,761905	0,795031
31	592	162	142	97	0,683099	0,598765	0,638158
32	468	124	107	78	0,728972	0,629032	0,675325
33	481	83	82	58	0,707317	0,698795	0,70303
34	331	80	63	36	0,571429	0,45	0,503497
35	364	62	57	37	0,649123	0,596774	0,621849
36	436	86	80	61	0,7625	0,709302	0,73494
37	320	74	64	49	0,765625	0,662162	0,710145
38	350	77	77	36	0,467532	0,467532	0,467532
39	506	130	92	69	0,75	0,530769	0,621622
40	380	89	76	50	0,657895	0,561798	0,606061
41	417	99	83	57	0,686747	0,575758	0,626374
42	571	124	130	76	0,584615	0,612903	0,598425
43	376	114	80	53	0,6625	0,464912	0,546392
44	666	187	189	153	0,809524	0,818182	0,81383
45	432	84	87	73	0,83908	0,869048	0,853801

No	A	GT	RP	O	Precision	Recall	F-score
46	454	121	103	58	0,563107	0,479339	0,517857
47	472	92	84	47	0,559524	0,51087	0,534091
48	576	143	110	90	0,818182	0,629371	0,711462
49	556	138	119	73	0,613445	0,528986	0,568093
50	405	114	48	29	0,604167	0,254386	0,358025
51	554	143	77	59	0,766234	0,412587	0,536364
52	518	122	103	93	0,902913	0,762295	0,826667
53	393	62	58	43	0,741379	0,693548	0,716667
54	394	105	82	54	0,658537	0,514286	0,57754
55	931	241	148	89	0,601351	0,369295	0,457584
56	390	92	87	47	0,54023	0,51087	0,52514
57	637	142	131	110	0,839695	0,774648	0,805861
58	1103	286	239	176	0,736402	0,615385	0,670476
59	652	179	132	69	0,522727	0,385475	0,44373
60	387	78	68	33	0,485294	0,423077	0,452055
61	395	134	78	35	0,448718	0,261194	0,330189
62	666	137	126	83	0,65873	0,605839	0,631179
63	759	158	142	95	0,669014	0,601266	0,633333
64	465	104	100	52	0,52	0,5	0,509804
65	432	113	84	59	0,702381	0,522124	0,598985
66	359	68	65	55	0,846154	0,808824	0,827068
67	278	56	43	13	0,302326	0,232143	0,262626
68	225	52	37	19	0,513514	0,365385	0,426966
69	434	148	103	74	0,718447	0,5	0,589641
70	494	118	109	88	0,807339	0,745763	0,77533
71	219	65	42	12	0,285714	0,184615	0,224299
72	385	74	55	38	0,690909	0,513514	0,589147
73	874	187	156	127	0,814103	0,679144	0,740525
74	523	87	68	53	0,779412	0,609195	0,683871
75	517	106	92	58	0,630435	0,54717	0,585859
76	370	88	92	57	0,619565	0,647727	0,633333
77	335	90	80	46	0,575	0,511111	0,541176
78	478	114	128	62	0,484375	0,54386	0,512397
79	530	151	135	75	0,555556	0,496689	0,524476
80	328	70	73	41	0,561644	0,585714	0,573427
	Rata-rata				0,653976	0,563944	0,600196

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

### Lampiran 29 :

Hasil ROUGE 1 - Textrank dengan Jaccard *similarity*, dengan Compression Rate 30%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	131	117	97	0,82906	0,740458	0,782258
2	450	152	152	117	0,769737	0,769737	0,769737
3	565	200	169	107	0,633136	0,535	0,579946
4	528	174	142	90	0,633803	0,517241	0,56962
5	657	259	203	155	0,763547	0,598456	0,670996
6	447	144	107	88	0,82243	0,611111	0,701195
7	966	254	332	220	0,662651	0,866142	0,750853
8	509	188	159	118	0,742138	0,62766	0,680115
9	386	157	136	119	0,875	0,757962	0,812287
10	368	144	107	81	0,757009	0,5625	0,645418
11	528	198	174	120	0,689655	0,606061	0,645161
12	513	194	181	135	0,745856	0,695876	0,72
13	644	221	204	150	0,735294	0,678733	0,705882
14	534	206	185	139	0,751351	0,674757	0,710997
15	285	105	96	66	0,6875	0,628571	0,656716
16	657	246	178	145	0,814607	0,589431	0,683962
17	799	313	287	241	0,839721	0,769968	0,803333
18	1029	313	329	268	0,81459	0,85623	0,834891
19	340	102	85	55	0,647059	0,539216	0,588235
20	695	244	234	161	0,688034	0,659836	0,67364
21	574	203	186	130	0,698925	0,640394	0,66838
22	532	171	175	141	0,805714	0,824561	0,815029
23	321	124	95	72	0,757895	0,580645	0,657534
24	373	153	125	69	0,552	0,45098	0,496403
25	366	121	98	75	0,765306	0,619835	0,684932
26	329	117	104	86	0,826923	0,735043	0,778281
27	383	115	118	92	0,779661	0,8	0,7897
28	359	118	118	91	0,771186	0,771186	0,771186
29	352	123	122	44	0,360656	0,357724	0,359184
30	409	157	147	110	0,748299	0,700637	0,723684
31	592	223	191	157	0,82199	0,704036	0,758454
32	468	155	134	96	0,716418	0,619355	0,66436
33	481	137	132	91	0,689394	0,664234	0,67658
34	331	114	100	66	0,66	0,578947	0,616822
35	364	98	91	59	0,648352	0,602041	0,624339
36	436	152	125	105	0,84	0,690789	0,758123
37	320	136	106	75	0,707547	0,551471	0,619835
38	350	123	116	83	0,715517	0,674797	0,694561
39	506	212	166	139	0,837349	0,65566	0,73545
40	380	141	111	75	0,675676	0,531915	0,595238
41	417	176	154	139	0,902597	0,789773	0,842424
42	571	151	193	102	0,528497	0,675497	0,593023
43	376	142	118	80	0,677966	0,56338	0,615385
44	666	276	257	215	0,836576	0,778986	0,806754
45	432	160	146	136	0,931507	0,85	0,888889

No	A	GT	RP	O	Precision	Recall	F-score
46	454	161	148	99	0,668919	0,614907	0,640777
47	472	125	131	74	0,564885	0,592	0,578125
48	576	216	170	142	0,835294	0,657407	0,735751
49	556	216	191	154	0,806283	0,712963	0,756757
50	405	154	88	67	0,761364	0,435065	0,553719
51	554	215	138	110	0,797101	0,511628	0,623229
52	518	162	168	156	0,928571	0,962963	0,945455
53	393	120	120	84	0,7	0,7	0,7
54	394	148	134	104	0,776119	0,702703	0,737589
55	931	364	250	210	0,84	0,576923	0,684039
56	390	130	129	78	0,604651	0,6	0,602317
57	637	205	196	171	0,872449	0,834146	0,852868
58	1103	448	363	296	0,815427	0,660714	0,729963
59	652	249	207	142	0,68599	0,570281	0,622807
60	387	116	114	86	0,754386	0,741379	0,747826
61	395	189	129	108	0,837209	0,571429	0,679245
62	666	215	196	145	0,739796	0,674419	0,705596
63	759	259	212	146	0,688679	0,563707	0,619958
64	465	151	140	85	0,607143	0,562914	0,584192
65	432	164	137	112	0,817518	0,682927	0,744186
66	359	112	116	100	0,862069	0,892857	0,877193
67	278	105	75	29	0,386667	0,27619	0,322222
68	225	74	61	44	0,721311	0,594595	0,651852
69	434	200	144	125	0,868056	0,625	0,726744
70	494	174	159	129	0,811321	0,741379	0,774775
71	219	93	65	38	0,584615	0,408602	0,481013
72	385	110	117	70	0,598291	0,636364	0,61674
73	874	247	227	190	0,837004	0,769231	0,801688
74	523	170	117	102	0,871795	0,6	0,710801
75	517	176	147	100	0,680272	0,568182	0,619195
76	370	124	127	68	0,535433	0,548387	0,541833
77	335	146	151	114	0,754967	0,780822	0,767677
78	478	178	167	114	0,682635	0,640449	0,66087
79	530	214	177	119	0,672316	0,556075	0,608696
80	328	123	120	98	0,816667	0,796748	0,806584
	Rata-rata				0,736792	0,650752	0,687901

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

### Lampiran 30 :

Hasil ROUGE 1 - Textrank dengan Jaccard *similarity*, dengan Compression Rate 40%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	164	132	111	0,840909	0,676829	0,75
2	450	185	194	158	0,814433	0,854054	0,833773
3	565	267	221	179	0,809955	0,670412	0,733607
4	528	239	220	163	0,740909	0,682008	0,71024
5	657	299	248	193	0,778226	0,645485	0,705667
6	447	200	154	119	0,772727	0,595	0,672316
7	966	360	435	324	0,744828	0,9	0,815094
8	509	246	209	164	0,784689	0,666667	0,720879
9	386	179	182	145	0,796703	0,810056	0,803324
10	368	180	164	141	0,859756	0,783333	0,819767
11	528	271	226	157	0,69469	0,579336	0,631791
12	513	272	233	199	0,854077	0,731618	0,788119
13	644	311	295	250	0,847458	0,803859	0,825083
14	534	263	229	191	0,834061	0,726236	0,776423
15	285	133	127	86	0,677165	0,646617	0,661538
16	657	306	235	190	0,808511	0,620915	0,702403
17	799	414	376	320	0,851064	0,772947	0,810127
18	1029	447	417	344	0,82494	0,769575	0,796296
19	340	132	117	99	0,846154	0,75	0,795181
20	695	330	336	268	0,797619	0,812121	0,804805
21	574	265	242	178	0,735537	0,671698	0,70217
22	532	247	239	196	0,820084	0,793522	0,806584
23	321	162	129	103	0,79845	0,635802	0,707904
24	373	179	146	84	0,575342	0,469274	0,516923
25	366	159	144	108	0,75	0,679245	0,712871
26	329	155	152	123	0,809211	0,793548	0,801303
27	383	167	147	130	0,884354	0,778443	0,828025
28	359	158	163	112	0,687117	0,708861	0,697819
29	352	171	148	101	0,682432	0,590643	0,633229
30	409	209	194	172	0,886598	0,822967	0,853598
31	592	277	249	207	0,831325	0,747292	0,787072
32	468	224	187	162	0,86631	0,723214	0,788321
33	481	185	181	130	0,718232	0,702703	0,710383
34	331	160	144	118	0,819444	0,7375	0,776316
35	364	158	129	107	0,829457	0,677215	0,745645
36	436	186	176	132	0,75	0,709677	0,729282
37	320	161	139	96	0,690647	0,596273	0,64
38	350	158	158	117	0,740506	0,740506	0,740506
39	506	276	213	161	0,755869	0,583333	0,658487
40	380	167	157	109	0,694268	0,652695	0,67284
41	417	231	214	202	0,943925	0,874459	0,907865
42	571	229	262	179	0,683206	0,781659	0,729124
43	376	170	165	103	0,624242	0,605882	0,614925
44	666	386	298	265	0,889262	0,686528	0,774854
45	432	202	195	170	0,871795	0,841584	0,856423

No	A	GT	RP	O	Precision	Recall	F-score
46	454	201	192	146	0,760417	0,726368	0,743003
47	472	171	198	137	0,691919	0,80117	0,742547
48	576	272	245	208	0,84898	0,764706	0,804642
49	556	262	242	186	0,768595	0,709924	0,738095
50	405	179	139	115	0,827338	0,642458	0,72327
51	554	274	182	147	0,807692	0,536496	0,644737
52	518	215	212	174	0,820755	0,809302	0,814988
53	393	168	179	147	0,821229	0,875	0,847262
54	394	198	167	140	0,838323	0,707071	0,767123
55	931	477	337	284	0,84273	0,595388	0,697789
56	390	174	171	132	0,77193	0,758621	0,765217
57	637	304	272	245	0,900735	0,805921	0,850694
58	1103	561	510	417	0,817647	0,743316	0,778711
59	652	313	267	201	0,752809	0,642173	0,693103
60	387	161	157	105	0,66879	0,652174	0,660377
61	395	236	160	139	0,86875	0,588983	0,70202
62	666	281	254	192	0,755906	0,683274	0,717757
63	759	355	307	233	0,758958	0,656338	0,703927
64	465	209	197	151	0,766497	0,722488	0,743842
65	432	231	174	146	0,83908	0,632035	0,720988
66	359	152	152	118	0,776316	0,776316	0,776316
67	278	134	116	80	0,689655	0,597015	0,64
68	225	93	85	52	0,611765	0,55914	0,58427
69	434	235	204	182	0,892157	0,774468	0,829157
70	494	240	217	185	0,852535	0,770833	0,809628
71	219	123	86	67	0,77907	0,544715	0,641148
72	385	134	136	100	0,735294	0,746269	0,740741
73	874	349	302	250	0,827815	0,716332	0,768049
74	523	214	176	144	0,818182	0,672897	0,738462
75	517	242	189	145	0,767196	0,599174	0,672854
76	370	172	161	123	0,763975	0,715116	0,738739
77	335	178	178	145	0,814607	0,814607	0,814607
78	478	240	224	180	0,803571	0,75	0,775862
79	530	294	269	212	0,788104	0,721088	0,753108
80	328	173	145	123	0,848276	0,710983	0,773585
	Rata-rata				0,787676	0,707797	0,743319

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

### Lampiran 31 :

Hasil ROUGE 1 - Textrank dengan Jaccard *similarity*, dengan Compression Rate 50%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	194	185	169	0,913514	0,871134	0,891821
2	450	233	234	198	0,846154	0,849785	0,847966
3	565	308	264	215	0,814394	0,698052	0,751748
4	528	295	287	225	0,783972	0,762712	0,773196
5	657	345	293	242	0,825939	0,701449	0,758621
6	447	221	187	145	0,775401	0,656109	0,710784
7	966	461	498	415	0,833333	0,900217	0,865485
8	509	301	272	239	0,878676	0,79402	0,834206
9	386	218	239	182	0,761506	0,834862	0,796499
10	368	219	195	165	0,846154	0,753425	0,797101
11	528	338	290	244	0,841379	0,721893	0,77707
12	513	320	288	252	0,875	0,7875	0,828947
13	644	365	370	294	0,794595	0,805479	0,8
14	534	345	316	298	0,943038	0,863768	0,901664
15	285	160	147	109	0,741497	0,68125	0,710098
16	657	352	309	236	0,763754	0,670455	0,71407
17	799	507	480	444	0,925	0,87574	0,899696
18	1029	560	528	408	0,772727	0,728571	0,75
19	340	171	169	137	0,810651	0,80117	0,805882
20	695	404	382	327	0,856021	0,809406	0,832061
21	574	323	299	238	0,795987	0,736842	0,765273
22	532	310	298	248	0,832215	0,8	0,815789
23	321	201	184	156	0,847826	0,776119	0,81039
24	373	204	175	119	0,68	0,583333	0,627968
25	366	222	194	156	0,804124	0,702703	0,75
26	329	192	191	172	0,900524	0,895833	0,898172
27	383	214	167	149	0,892216	0,696262	0,782152
28	359	202	207	166	0,801932	0,821782	0,811736
29	352	221	184	127	0,690217	0,574661	0,62716
30	409	247	240	222	0,925	0,898785	0,911704
31	592	356	305	259	0,84918	0,727528	0,783661
32	468	290	239	206	0,861925	0,710345	0,778828
33	481	244	236	183	0,775424	0,75	0,7625
34	331	197	175	148	0,845714	0,751269	0,795699
35	364	196	164	140	0,853659	0,714286	0,777778
36	436	220	202	156	0,772277	0,709091	0,739336
37	320	186	163	121	0,742331	0,650538	0,69341
38	350	201	184	138	0,75	0,686567	0,716883
39	506	325	274	224	0,817518	0,689231	0,747913
40	380	219	200	167	0,835	0,762557	0,797136
41	417	283	248	230	0,927419	0,812721	0,86629
42	571	308	326	235	0,720859	0,762987	0,741325
43	376	184	213	127	0,596244	0,690217	0,639798
44	666	476	347	313	0,902017	0,657563	0,760632
45	432	250	228	203	0,890351	0,812	0,849372

No	A	GT	RP	O	Precision	Recall	F-score
46	454	236	257	173	0,673152	0,733051	0,701826
47	472	222	245	198	0,808163	0,891892	0,847966
48	576	322	292	260	0,890411	0,807453	0,846906
49	556	313	315	250	0,793651	0,798722	0,796178
50	405	210	172	132	0,767442	0,628571	0,691099
51	554	300	203	162	0,79803	0,54	0,644135
52	518	310	316	286	0,905063	0,922581	0,913738
53	393	213	186	152	0,817204	0,713615	0,761905
54	394	225	209	178	0,851675	0,791111	0,820276
55	931	563	451	381	0,844789	0,676732	0,751479
56	390	202	198	154	0,777778	0,762376	0,77
57	637	380	363	333	0,917355	0,876316	0,896366
58	1103	697	635	559	0,880315	0,802009	0,839339
59	652	374	356	283	0,794944	0,756684	0,775342
60	387	214	215	179	0,832558	0,836449	0,834499
61	395	284	225	209	0,928889	0,735915	0,821218
62	666	364	325	268	0,824615	0,736264	0,777939
63	759	412	383	293	0,765013	0,711165	0,737107
64	465	274	250	205	0,82	0,748175	0,782443
65	432	265	225	175	0,777778	0,660377	0,714286
66	359	191	197	148	0,751269	0,774869	0,762887
67	278	146	141	112	0,794326	0,767123	0,780488
68	225	133	110	85	0,772727	0,639098	0,699588
69	434	268	236	213	0,902542	0,794776	0,845238
70	494	292	253	220	0,869565	0,753425	0,807339
71	219	129	107	84	0,785047	0,651163	0,711864
72	385	191	162	124	0,765432	0,649215	0,70255
73	874	447	417	363	0,870504	0,812081	0,840278
74	523	270	222	195	0,878378	0,722222	0,792683
75	517	284	236	208	0,881356	0,732394	0,8
76	370	211	204	165	0,808824	0,781991	0,795181
77	335	227	220	189	0,859091	0,832599	0,845638
78	478	276	264	220	0,833333	0,797101	0,814815
79	530	339	328	282	0,859756	0,831858	0,845577
80	328	229	194	169	0,871134	0,737991	0,799054
	Rata-rata				0,82356	0,754345	0,785813

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

### Lampiran 32 :

Hasil ROUGE 1 - Textrank dengan BoW Cosine *similarity*, dengan Compression

Rate 10%

No	A	GT	RP	O	Precision	Recall	F-Score
1	331	59	48	34	0,708333	0,576271	0,635514
2	450	43	59	15	0,254237	0,348837	0,294118
3	565	101	60	38	0,633333	0,376238	0,47205
4	528	70	37	27	0,72973	0,385714	0,504673
5	657	92	109	52	0,477064	0,565217	0,517413
6	447	62	49	41	0,836735	0,66129	0,738739
7	966	38	138	23	0,166667	0,605263	0,261364
8	509	95	77	55	0,714286	0,578947	0,639535
9	386	43	46	34	0,73913	0,790698	0,764045
10	368	77	53	46	0,867925	0,597403	0,707692
11	528	68	57	34	0,596491	0,5	0,544
12	513	60	64	21	0,328125	0,35	0,33871
13	644	72	58	35	0,603448	0,486111	0,538462
14	534	64	66	22	0,333333	0,34375	0,338462
15	285	47	28	22	0,785714	0,468085	0,586667
16	657	88	51	31	0,607843	0,352273	0,446043
17	799	94	125	45	0,36	0,478723	0,410959
18	1029	114	140	104	0,742857	0,912281	0,818898
19	340	18	30	5	0,166667	0,277778	0,208333
20	695	79	82	35	0,426829	0,443038	0,434783
21	574	50	52	25	0,480769	0,5	0,490196
22	532	63	62	51	0,822581	0,809524	0,816
23	321	45	23	12	0,521739	0,266667	0,352941
24	373	32	28	19	0,678571	0,59375	0,633333
25	366	33	33	11	0,333333	0,333333	0,333333
26	329	42	36	5	0,138889	0,119048	0,128205
27	383	37	41	18	0,439024	0,486486	0,461538
28	359	42	43	32	0,744186	0,761905	0,752941
29	352	34	36	15	0,416667	0,441176	0,428571
30	409	41	41	20	0,487805	0,487805	0,487805
31	592	64	62	28	0,451613	0,4375	0,444444
32	468	62	56	38	0,678571	0,612903	0,644068
33	481	33	36	14	0,388889	0,424242	0,405797
34	331	30	32	15	0,46875	0,5	0,483871
35	364	30	26	13	0,5	0,433333	0,464286
36	436	45	48	38	0,791667	0,844444	0,817204
37	320	28	29	9	0,310345	0,321429	0,315789
38	350	37	37	6	0,162162	0,162162	0,162162
39	506	77	69	60	0,869565	0,779221	0,821918
40	380	44	33	10	0,30303	0,227273	0,25974
41	417	52	54	31	0,574074	0,596154	0,584906
42	571	84	72	25	0,347222	0,297619	0,320513
43	376	61	38	24	0,631579	0,393443	0,484848
44	666	72	59	44	0,745763	0,611111	0,671756
45	432	32	48	25	0,520833	0,78125	0,625

No	A	GT	RP	O	Precision	Recall	F-Score
46	454	43	50	21	0,42	0,488372	0,451613
47	472	61	60	44	0,733333	0,721311	0,727273
48	576	94	85	74	0,870588	0,787234	0,826816
49	556	67	67	43	0,641791	0,641791	0,641791
50	405	43	48	11	0,229167	0,255814	0,241758
51	554	94	48	36	0,75	0,382979	0,507042
52	518	58	50	33	0,66	0,568966	0,611111
53	393	39	39	39	1	1	1
54	394	57	55	43	0,781818	0,754386	0,767857
55	931	88	87	40	0,45977	0,454545	0,457143
56	390	45	39	24	0,615385	0,533333	0,571429
57	637	62	48	38	0,791667	0,612903	0,690909
58	1103	152	122	68	0,557377	0,447368	0,49635
59	652	106	76	31	0,407895	0,292453	0,340659
60	387	30	33	13	0,393939	0,433333	0,412698
61	395	58	47	23	0,489362	0,396552	0,438095
62	666	69	62	25	0,403226	0,362319	0,381679
63	759	58	67	33	0,492537	0,568966	0,528
64	465	55	46	26	0,565217	0,472727	0,514851
65	432	70	43	32	0,744186	0,457143	0,566372
66	359	27	46	15	0,326087	0,555556	0,410959
67	278	40	23	9	0,391304	0,225	0,285714
68	225	15	15	15	1	1	1
69	434	70	40	12	0,3	0,171429	0,218182
70	494	58	55	45	0,818182	0,775862	0,79646
71	219	21	20	3	0,15	0,142857	0,146341
72	385	29	25	18	0,72	0,62069	0,666667
73	874	84	78	58	0,74359	0,690476	0,716049
74	523	40	44	30	0,681818	0,75	0,714286
75	517	46	45	24	0,533333	0,521739	0,527473
76	370	32	55	10	0,181818	0,3125	0,229885
77	335	55	48	16	0,333333	0,290909	0,31068
78	478	50	45	20	0,444444	0,4	0,421053
79	530	60	65	22	0,338462	0,366667	0,352
80	328	44	47	24	0,510638	0,545455	0,527473
Rata-rata				0,542083	0,504017	0,513604	

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

### Lampiran 33 :

Hasil ROUGE 1 - Textrank dengan BoW Cosine *similarity*, dengan Compression

Rate 20%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	101	90	76	0,844444	0,752475	0,795812
2	450	91	106	70	0,660377	0,769231	0,71066
3	565	139	94	64	0,680851	0,460432	0,549356
4	528	126	96	60	0,625	0,47619	0,540541
5	657	152	157	92	0,585987	0,605263	0,595469
6	447	95	76	52	0,684211	0,547368	0,608187
7	966	111	207	86	0,415459	0,774775	0,540881
8	509	157	110	83	0,754545	0,528662	0,621723
9	386	100	83	57	0,686747	0,57	0,622951
10	368	112	99	78	0,787879	0,696429	0,739336
11	528	123	103	79	0,76699	0,642276	0,699115
12	513	148	137	70	0,510949	0,472973	0,491228
13	644	148	142	105	0,739437	0,709459	0,724138
14	534	130	124	92	0,741935	0,707692	0,724409
15	285	75	64	53	0,828125	0,706667	0,76259
16	657	149	135	123	0,911111	0,825503	0,866197
17	799	198	211	138	0,654028	0,69697	0,674817
18	1029	218	236	191	0,809322	0,876147	0,84141
19	340	56	87	45	0,517241	0,803571	0,629371
20	695	164	165	116	0,70303	0,707317	0,705167
21	574	138	114	94	0,824561	0,681159	0,746032
22	532	133	125	82	0,656	0,616541	0,635659
23	321	88	73	69	0,945205	0,784091	0,857143
24	373	82	75	41	0,546667	0,5	0,522293
25	366	70	71	37	0,521127	0,528571	0,524823
26	329	89	77	40	0,519481	0,449438	0,481928
27	383	88	87	53	0,609195	0,602273	0,605714
28	359	74	82	43	0,52439	0,581081	0,551282
29	352	74	78	37	0,474359	0,5	0,486842
30	409	84	87	60	0,689655	0,714286	0,701754
31	592	162	142	97	0,683099	0,598765	0,638158
32	468	124	100	78	0,78	0,629032	0,696429
33	481	83	87	55	0,632184	0,662651	0,647059
34	331	80	63	36	0,571429	0,45	0,503497
35	364	62	61	31	0,508197	0,5	0,504065
36	436	86	83	75	0,903614	0,872093	0,887574
37	320	74	71	54	0,760563	0,72973	0,744828
38	350	77	86	54	0,627907	0,701299	0,662577
39	506	130	107	89	0,831776	0,684615	0,751055
40	380	89	90	55	0,611111	0,617978	0,614525
41	417	99	105	80	0,761905	0,808081	0,784314
42	571	124	133	77	0,578947	0,620968	0,599222
43	376	114	80	53	0,6625	0,464912	0,546392
44	666	187	189	153	0,809524	0,818182	0,81383
45	432	84	92	71	0,771739	0,845238	0,806818

No	A	GT	RP	O	Precision	Recall	F-score
46	454	121	108	60	0,555556	0,495868	0,524017
47	472	92	99	58	0,585859	0,630435	0,60733
48	576	143	114	94	0,824561	0,657343	0,731518
49	556	138	128	90	0,703125	0,652174	0,676692
50	405	114	84	67	0,797619	0,587719	0,676768
51	554	143	77	59	0,766234	0,412587	0,536364
52	518	122	88	66	0,75	0,540984	0,628571
53	393	62	58	43	0,741379	0,693548	0,716667
54	394	105	82	54	0,658537	0,514286	0,57754
55	931	241	174	107	0,614943	0,443983	0,515663
56	390	92	97	58	0,597938	0,630435	0,613757
57	637	142	128	114	0,890625	0,802817	0,844444
58	1103	286	247	174	0,704453	0,608392	0,652908
59	652	179	141	89	0,631206	0,497207	0,55625
60	387	78	71	43	0,605634	0,551282	0,577181
61	395	134	78	35	0,448718	0,261194	0,330189
62	666	137	120	81	0,675	0,591241	0,63035
63	759	158	141	97	0,687943	0,613924	0,648829
64	465	104	108	66	0,611111	0,634615	0,622642
65	432	113	102	72	0,705882	0,637168	0,669767
66	359	68	87	56	0,643678	0,823529	0,722581
67	278	56	37	19	0,513514	0,339286	0,408602
68	225	52	39	22	0,564103	0,423077	0,483516
69	434	148	87	58	0,666667	0,391892	0,493617
70	494	118	116	89	0,767241	0,754237	0,760684
71	219	65	42	12	0,285714	0,184615	0,224299
72	385	74	55	38	0,690909	0,513514	0,589147
73	874	187	169	129	0,763314	0,68984	0,724719
74	523	87	81	55	0,679012	0,632184	0,654762
75	517	106	103	57	0,553398	0,537736	0,545455
76	370	88	92	57	0,619565	0,647727	0,633333
77	335	90	93	58	0,623656	0,644444	0,63388
78	478	114	122	63	0,516393	0,552632	0,533898
79	530	151	126	71	0,563492	0,470199	0,512635
80	328	70	82	39	0,47561	0,557143	0,513158
	Rata-rata				0,664942	0,611371	0,631611

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

### Lampiran 34 :

Hasil ROUGE 1 - Textrank dengan BoW Cosine *similarity*, dengan Compression

Rate 30%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	131	117	97	0,82906	0,740458	0,782258
2	450	152	146	124	0,849315	0,815789	0,832215
3	565	200	210	153	0,728571	0,765	0,746341
4	528	174	172	103	0,598837	0,591954	0,595376
5	657	259	203	155	0,763547	0,598456	0,670996
6	447	144	122	92	0,754098	0,638889	0,691729
7	966	254	360	217	0,602778	0,854331	0,70684
8	509	188	165	119	0,721212	0,632979	0,674221
9	386	157	135	111	0,822222	0,707006	0,760274
10	368	144	135	115	0,851852	0,798611	0,824373
11	528	198	174	120	0,689655	0,606061	0,645161
12	513	194	190	122	0,642105	0,628866	0,635417
13	644	221	209	169	0,808612	0,764706	0,786047
14	534	206	183	139	0,759563	0,674757	0,714653
15	285	105	101	77	0,762376	0,733333	0,747573
16	657	246	191	163	0,853403	0,662602	0,745995
17	799	313	283	226	0,798587	0,722045	0,758389
18	1029	313	329	262	0,796353	0,837061	0,816199
19	340	102	110	78	0,709091	0,764706	0,735849
20	695	244	252	175	0,694444	0,717213	0,705645
21	574	203	186	130	0,698925	0,640394	0,66838
22	532	171	182	131	0,71978	0,766082	0,74221
23	321	124	103	89	0,864078	0,717742	0,784141
24	373	153	125	69	0,552	0,45098	0,496403
25	366	121	113	71	0,628319	0,586777	0,606838
26	329	117	104	66	0,634615	0,564103	0,597285
27	383	115	105	86	0,819048	0,747826	0,781818
28	359	118	124	73	0,58871	0,618644	0,603306
29	352	123	126	65	0,515873	0,528455	0,522088
30	409	157	148	119	0,804054	0,757962	0,780328
31	592	223	198	173	0,873737	0,775785	0,821853
32	468	155	136	99	0,727941	0,63871	0,680412
33	481	137	133	86	0,646617	0,627737	0,637037
34	331	114	98	55	0,561224	0,482456	0,518868
35	364	98	91	59	0,648352	0,602041	0,624339
36	436	152	125	105	0,84	0,690789	0,758123
37	320	136	121	105	0,867769	0,772059	0,817121
38	350	123	118	99	0,838983	0,804878	0,821577
39	506	212	166	139	0,837349	0,65566	0,73545
40	380	141	125	101	0,808	0,716312	0,759398
41	417	176	160	137	0,85625	0,778409	0,815476
42	571	151	193	89	0,46114	0,589404	0,517442
43	376	142	109	72	0,66055	0,507042	0,573705
44	666	276	253	211	0,833992	0,764493	0,797732
45	432	160	146	136	0,931507	0,85	0,888889

No	A	GT	RP	O	Precision	Recall	F-score
46	454	161	163	102	0,625767	0,63354	0,62963
47	472	125	147	79	0,537415	0,632	0,580882
48	576	216	183	154	0,84153	0,712963	0,77193
49	556	216	191	154	0,806283	0,712963	0,756757
50	405	154	108	75	0,694444	0,487013	0,572519
51	554	215	138	110	0,797101	0,511628	0,623229
52	518	162	144	118	0,819444	0,728395	0,771242
53	393	120	116	101	0,87069	0,841667	0,855932
54	394	148	147	115	0,782313	0,777027	0,779661
55	931	364	250	210	0,84	0,576923	0,684039
56	390	130	129	78	0,604651	0,6	0,602317
57	637	205	193	168	0,870466	0,819512	0,844221
58	1103	448	382	318	0,832461	0,709821	0,766265
59	652	249	206	140	0,679612	0,562249	0,615385
60	387	116	114	86	0,754386	0,741379	0,747826
61	395	189	129	108	0,837209	0,571429	0,679245
62	666	215	196	145	0,739796	0,674419	0,705596
63	759	259	233	168	0,72103	0,648649	0,682927
64	465	151	141	79	0,560284	0,523179	0,541096
65	432	164	151	122	0,807947	0,743902	0,774603
66	359	112	115	100	0,869565	0,892857	0,881057
67	278	105	83	51	0,614458	0,485714	0,542553
68	225	74	54	34	0,62963	0,459459	0,53125
69	434	200	139	125	0,899281	0,625	0,737463
70	494	174	170	133	0,782353	0,764368	0,773256
71	219	93	65	38	0,584615	0,408602	0,481013
72	385	110	111	76	0,684685	0,690909	0,687783
73	874	247	244	199	0,815574	0,805668	0,810591
74	523	170	125	93	0,744	0,547059	0,630508
75	517	176	147	100	0,680272	0,568182	0,619195
76	370	124	127	68	0,535433	0,548387	0,541833
77	335	146	151	114	0,754967	0,780822	0,767677
78	478	178	152	105	0,690789	0,589888	0,636364
79	530	214	194	135	0,695876	0,630841	0,661765
80	328	123	120	98	0,816667	0,796748	0,806584
	Rata-rata				0,738444	0,671159	0,70025

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

### Lampiran 35 :

Hasil ROUGE 1 - Textrank dengan BoW Cosine *similarity*, dengan Compression

Rate 40%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	164	151	128	0,847682	0,780488	0,812698
2	450	185	194	158	0,814433	0,854054	0,833773
3	565	267	250	205	0,82	0,76779	0,793037
4	528	239	239	159	0,665272	0,665272	0,665272
5	657	299	262	216	0,824427	0,722408	0,770053
6	447	200	200	154	0,77	0,77	0,77
7	966	360	458	304	0,663755	0,844444	0,743276
8	509	246	225	186	0,826667	0,756098	0,789809
9	386	179	182	145	0,796703	0,810056	0,803324
10	368	180	171	135	0,789474	0,75	0,769231
11	528	271	226	158	0,699115	0,583026	0,635815
12	513	272	252	223	0,884921	0,819853	0,851145
13	644	311	295	250	0,847458	0,803859	0,825083
14	534	263	229	191	0,834061	0,726236	0,776423
15	285	133	127	86	0,677165	0,646617	0,661538
16	657	306	244	197	0,807377	0,643791	0,716364
17	799	414	395	352	0,891139	0,850242	0,87021
18	1029	447	443	347	0,783296	0,776286	0,779775
19	340	132	142	125	0,880282	0,94697	0,912409
20	695	330	336	268	0,797619	0,812121	0,804805
21	574	265	239	178	0,74477	0,671698	0,706349
22	532	247	234	191	0,816239	0,773279	0,794179
23	321	162	140	114	0,814286	0,703704	0,754967
24	373	179	176	103	0,585227	0,575419	0,580282
25	366	159	154	109	0,707792	0,685535	0,696486
26	329	155	152	122	0,802632	0,787097	0,794788
27	383	167	155	124	0,8	0,742515	0,770186
28	359	158	163	112	0,687117	0,708861	0,697819
29	352	171	158	119	0,753165	0,695906	0,723404
30	409	209	194	172	0,886598	0,822967	0,853598
31	592	277	244	208	0,852459	0,750903	0,798464
32	468	224	182	151	0,82967	0,674107	0,743842
33	481	185	180	137	0,761111	0,740541	0,750685
34	331	160	147	116	0,789116	0,725	0,7557
35	364	158	131	108	0,824427	0,683544	0,747405
36	436	186	176	132	0,75	0,709677	0,729282
37	320	161	160	132	0,825	0,819876	0,82243
38	350	158	163	131	0,803681	0,829114	0,816199
39	506	276	205	155	0,756098	0,561594	0,644491
40	380	167	157	109	0,694268	0,652695	0,67284
41	417	231	214	202	0,943925	0,874459	0,907865
42	571	229	262	179	0,683206	0,781659	0,729124
43	376	170	165	103	0,624242	0,605882	0,614925
44	666	386	364	307	0,843407	0,795337	0,818667
45	432	202	185	154	0,832432	0,762376	0,795866

46	454	201	201	147	0,731343	0,731343	0,731343
47	472	171	194	141	0,726804	0,824561	0,772603
48	576	272	238	202	0,848739	0,742647	0,792157
49	556	262	242	186	0,768595	0,709924	0,738095
50	405	179	139	115	0,827338	0,642458	0,72327
51	554	274	170	134	0,788235	0,489051	0,603604
52	518	215	211	175	0,829384	0,813953	0,821596
53	393	168	157	127	0,808917	0,755952	0,781538
54	394	198	184	156	0,847826	0,787879	0,816754
55	931	477	365	320	0,876712	0,67086	0,760095
56	390	174	171	132	0,77193	0,758621	0,765217
57	637	304	282	251	0,890071	0,825658	0,856655
58	1103	561	510	431	0,845098	0,768271	0,804855
59	652	313	273	203	0,74359	0,648562	0,692833
60	387	161	161	108	0,670807	0,670807	0,670807
61	395	236	182	163	0,895604	0,690678	0,779904
62	666	281	254	192	0,755906	0,683274	0,717757
63	759	355	316	239	0,756329	0,673239	0,71237
64	465	209	203	146	0,719212	0,698565	0,708738
65	432	231	188	163	0,867021	0,705628	0,778043
66	359	152	153	121	0,79085	0,796053	0,793443
67	278	134	116	80	0,689655	0,597015	0,64
68	225	93	76	55	0,723684	0,591398	0,650888
69	434	235	239	225	0,941423	0,957447	0,949367
70	494	240	214	184	0,859813	0,766667	0,810573
71	219	123	86	67	0,77907	0,544715	0,641148
72	385	134	137	111	0,810219	0,828358	0,819188
73	874	349	337	265	0,78635	0,759312	0,772595
74	523	214	180	144	0,8	0,672897	0,730964
75	517	242	200	162	0,81	0,669421	0,733032
76	370	172	161	123	0,763975	0,715116	0,738739
77	335	178	181	141	0,779006	0,792135	0,785515
78	478	240	207	166	0,801932	0,691667	0,742729
79	530	294	269	212	0,788104	0,721088	0,753108
80	328	173	160	125	0,78125	0,722543	0,750751
Rata-rata					0,790081	0,732264	0,758052

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

### Lampiran 36 :

Hasil ROUGE 1 - Textrank dengan BoW Cosine *similarity*, dengan Compression

Rate 50%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	194	180	167	0,927778	0,860825	0,893048
2	450	233	234	198	0,846154	0,849785	0,847966
3	565	308	291	249	0,85567	0,808442	0,831386
4	528	295	289	212	0,733564	0,718644	0,726027
5	657	345	305	256	0,839344	0,742029	0,787692
6	447	221	236	178	0,754237	0,80543	0,778993
7	966	461	547	414	0,756856	0,898048	0,821429
8	509	301	278	245	0,881295	0,813953	0,846287
9	386	218	238	165	0,693277	0,756881	0,723684
10	368	219	210	190	0,904762	0,86758	0,885781
11	528	338	292	250	0,856164	0,739645	0,793651
12	513	320	287	241	0,839721	0,753125	0,794069
13	644	365	372	299	0,803763	0,819178	0,811398
14	534	345	316	298	0,943038	0,863768	0,901664
15	285	160	147	109	0,741497	0,68125	0,710098
16	657	352	309	236	0,763754	0,670455	0,71407
17	799	507	480	448	0,933333	0,883629	0,907801
18	1029	560	582	443	0,761168	0,791071	0,775832
19	340	171	176	140	0,795455	0,818713	0,806916
20	695	404	395	326	0,825316	0,806931	0,81602
21	574	323	299	238	0,795987	0,736842	0,765273
22	532	310	297	249	0,838384	0,803226	0,820428
23	321	201	195	158	0,810256	0,78607	0,79798
24	373	204	196	131	0,668367	0,642157	0,655
25	366	222	217	179	0,824885	0,806306	0,81549
26	329	192	192	172	0,895833	0,895833	0,895833
27	383	214	183	163	0,89071	0,761682	0,821159
28	359	202	207	166	0,801932	0,821782	0,811736
29	352	221	189	144	0,761905	0,651584	0,702439
30	409	247	240	222	0,925	0,898785	0,911704
31	592	356	303	253	0,834983	0,710674	0,76783
32	468	290	224	183	0,816964	0,631034	0,712062
33	481	244	236	183	0,775424	0,75	0,7625
34	331	197	175	141	0,805714	0,715736	0,758065
35	364	196	168	136	0,809524	0,693878	0,747253
36	436	220	206	162	0,786408	0,736364	0,760563
37	320	186	184	145	0,788043	0,77957	0,783784
38	350	201	184	143	0,777174	0,711443	0,742857
39	506	325	272	220	0,808824	0,676923	0,737018
40	380	219	199	161	0,809045	0,73516	0,770335
41	417	283	249	228	0,915663	0,805654	0,857143
42	571	308	326	235	0,720859	0,762987	0,741325
43	376	184	213	127	0,596244	0,690217	0,639798
44	666	476	436	410	0,940367	0,861345	0,899123
45	432	250	224	202	0,901786	0,808	0,852321

No	A	GT	RP	O	Precision	Recall	F-score
46	454	236	255	179	0,701961	0,758475	0,729124
47	472	222	250	191	0,764	0,86036	0,809322
48	576	322	290	257	0,886207	0,798137	0,839869
49	556	313	315	250	0,793651	0,798722	0,796178
50	405	210	184	145	0,788043	0,690476	0,736041
51	554	300	248	212	0,854839	0,706667	0,773723
52	518	310	294	247	0,840136	0,796774	0,817881
53	393	213	171	135	0,789474	0,633803	0,703125
54	394	225	207	179	0,864734	0,795556	0,828704
55	931	563	484	406	0,838843	0,721137	0,775549
56	390	202	199	151	0,758794	0,747525	0,753117
57	637	380	359	336	0,935933	0,884211	0,909337
58	1103	697	633	563	0,889415	0,807747	0,846617
59	652	374	366	290	0,79235	0,775401	0,783784
60	387	214	215	179	0,832558	0,836449	0,834499
61	395	284	225	209	0,928889	0,735915	0,821218
62	666	364	325	268	0,824615	0,736264	0,777939
63	759	412	388	305	0,786082	0,740291	0,7625
64	465	274	244	196	0,803279	0,715328	0,756757
65	432	265	238	196	0,823529	0,739623	0,779324
66	359	191	197	148	0,751269	0,774869	0,762887
67	278	146	141	112	0,794326	0,767123	0,780488
68	225	133	119	100	0,840336	0,75188	0,793651
69	434	268	267	243	0,910112	0,906716	0,908411
70	494	292	253	220	0,869565	0,753425	0,807339
71	219	129	107	84	0,785047	0,651163	0,711864
72	385	191	168	143	0,85119	0,748691	0,796657
73	874	447	429	386	0,899767	0,863535	0,881279
74	523	270	222	195	0,878378	0,722222	0,792683
75	517	284	269	238	0,884758	0,838028	0,860759
76	370	211	202	176	0,871287	0,834123	0,8523
77	335	227	220	189	0,859091	0,832599	0,845638
78	478	276	254	202	0,795276	0,731884	0,762264
79	530	339	328	282	0,859756	0,831858	0,845577
80	328	229	202	180	0,891089	0,786026	0,835267
	Rata-rata				0,824938	0,773695	0,797181

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

### Lampiran 37 :

Hasil ROUGE 1 - Textrank dengan TF-IDF Cosine *similarity*, dengan

Compression Rate 10%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	59	48	34	0,708333	0,576271	0,635514
2	450	43	63	24	0,380952	0,55814	0,45283
3	565	101	60	39	0,65	0,386139	0,484472
4	528	70	41	15	0,365854	0,214286	0,27027
5	657	92	81	66	0,814815	0,717391	0,763006
6	447	62	49	41	0,836735	0,66129	0,738739
7	966	38	118	26	0,220339	0,684211	0,333333
8	509	95	64	42	0,65625	0,442105	0,528302
9	386	43	46	34	0,73913	0,790698	0,764045
10	368	77	32	14	0,4375	0,181818	0,256881
11	528	68	57	34	0,596491	0,5	0,544
12	513	60	64	21	0,328125	0,35	0,33871
13	644	72	58	35	0,603448	0,486111	0,538462
14	534	64	85	15	0,176471	0,234375	0,201342
15	285	47	25	21	0,84	0,446809	0,583333
16	657	88	66	52	0,787879	0,590909	0,675325
17	799	94	109	61	0,559633	0,648936	0,600985
18	1029	114	103	78	0,757282	0,684211	0,718894
19	340	18	30	5	0,166667	0,277778	0,208333
20	695	79	78	52	0,666667	0,658228	0,66242
21	574	50	43	22	0,511628	0,44	0,473118
22	532	63	56	31	0,553571	0,492063	0,521008
23	321	45	23	12	0,521739	0,266667	0,352941
24	373	32	28	9	0,321429	0,28125	0,3
25	366	33	33	11	0,333333	0,333333	0,333333
26	329	42	33	17	0,515152	0,404762	0,453333
27	383	37	37	35	0,945946	0,945946	0,945946
28	359	42	41	31	0,756098	0,738095	0,746988
29	352	34	33	10	0,30303	0,294118	0,298507
30	409	41	41	20	0,487805	0,487805	0,487805
31	592	64	62	40	0,645161	0,625	0,634921
32	468	62	56	38	0,678571	0,612903	0,644068
33	481	33	46	21	0,456522	0,636364	0,531646
34	331	30	32	15	0,46875	0,5	0,483871
35	364	30	26	13	0,5	0,433333	0,464286
36	436	45	48	38	0,791667	0,844444	0,817204
37	320	28	20	8	0,4	0,285714	0,333333
38	350	37	23	9	0,391304	0,243243	0,3
39	506	77	46	39	0,847826	0,506494	0,634146
40	380	44	33	10	0,30303	0,227273	0,25974
41	417	52	40	18	0,45	0,346154	0,391304
42	571	84	63	38	0,603175	0,452381	0,517007
43	376	61	38	24	0,631579	0,393443	0,484848
44	666	72	69	26	0,376812	0,361111	0,368794
45	432	32	48	25	0,520833	0,78125	0,625

No	A	GT	RP	O	Precision	Recall	F-score
46	454	43	50	21	0,42	0,488372	0,451613
47	472	61	53	31	0,584906	0,508197	0,54386
48	576	94	74	54	0,72973	0,574468	0,642857
49	556	67	69	42	0,608696	0,626866	0,617647
50	405	43	52	10	0,192308	0,232558	0,210526
51	554	94	46	25	0,543478	0,265957	0,357143
52	518	58	50	33	0,66	0,568966	0,611111
53	393	39	47	27	0,574468	0,692308	0,627907
54	394	57	55	43	0,781818	0,754386	0,767857
55	931	88	60	33	0,55	0,375	0,445946
56	390	45	33	20	0,606061	0,444444	0,512821
57	637	62	56	49	0,875	0,790323	0,830508
58	1103	152	130	66	0,507692	0,434211	0,468085
59	652	106	76	31	0,407895	0,292453	0,340659
60	387	30	34	20	0,588235	0,666667	0,625
61	395	58	28	21	0,75	0,362069	0,488372
62	666	69	73	29	0,39726	0,42029	0,408451
63	759	58	66	47	0,712121	0,810345	0,758065
64	465	55	48	25	0,520833	0,454545	0,485437
65	432	70	67	51	0,761194	0,728571	0,744526
66	359	27	46	15	0,326087	0,555556	0,410959
67	278	40	23	9	0,391304	0,225	0,285714
68	225	15	15	15	1	1	1
69	434	70	52	11	0,211538	0,157143	0,180328
70	494	58	64	50	0,78125	0,862069	0,819672
71	219	21	20	3	0,15	0,142857	0,146341
72	385	29	25	18	0,72	0,62069	0,666667
73	874	84	78	61	0,782051	0,72619	0,753086
74	523	40	36	29	0,805556	0,725	0,763158
75	517	46	36	29	0,805556	0,630435	0,707317
76	370	32	49	10	0,204082	0,3125	0,246914
77	335	55	40	20	0,5	0,363636	0,421053
78	478	50	47	31	0,659574	0,62	0,639175
79	530	60	69	33	0,478261	0,55	0,511628
80	328	44	51	30	0,588235	0,681818	0,631579
	Rata-rata				0,559784	0,508584	0,522804

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

### Lampiran 38 :

Hasil ROUGE 1 - Textrank dengan TF-IDF Cosine *similarity*, dengan

Compression Rate 20%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	101	68	54	0,794118	0,534653	0,639053
2	450	91	96	68	0,708333	0,747253	0,727273
3	565	139	99	68	0,686869	0,489209	0,571429
4	528	126	100	66	0,66	0,52381	0,584071
5	657	152	157	89	0,566879	0,585526	0,576052
6	447	95	105	63	0,6	0,663158	0,63
7	966	111	235	89	0,378723	0,801802	0,514451
8	509	157	104	83	0,798077	0,528662	0,636015
9	386	100	90	69	0,766667	0,69	0,726316
10	368	112	99	78	0,787879	0,696429	0,739336
11	528	123	104	71	0,682692	0,577236	0,625551
12	513	148	123	77	0,626016	0,52027	0,568266
13	644	148	144	91	0,631944	0,614865	0,623288
14	534	130	131	82	0,625954	0,630769	0,628352
15	285	75	64	53	0,828125	0,706667	0,76259
16	657	149	135	123	0,911111	0,825503	0,866197
17	799	198	213	174	0,816901	0,878788	0,846715
18	1029	218	241	186	0,771784	0,853211	0,810458
19	340	56	59	31	0,525424	0,553571	0,53913
20	695	164	164	102	0,621951	0,621951	0,621951
21	574	138	121	87	0,719008	0,630435	0,671815
22	532	133	115	90	0,782609	0,676692	0,725806
23	321	88	63	49	0,777778	0,556818	0,649007
24	373	82	75	41	0,546667	0,5	0,522293
25	366	70	88	46	0,522727	0,657143	0,582278
26	329	89	73	58	0,794521	0,651685	0,716049
27	383	88	78	61	0,782051	0,693182	0,73494
28	359	74	75	57	0,76	0,77027	0,765101
29	352	74	82	24	0,292683	0,324324	0,307692
30	409	84	88	71	0,806818	0,845238	0,825581
31	592	162	142	97	0,683099	0,598765	0,638158
32	468	124	98	68	0,693878	0,548387	0,612613
33	481	83	97	56	0,57732	0,674699	0,622222
34	331	80	75	36	0,48	0,45	0,464516
35	364	62	65	44	0,676923	0,709677	0,692913
36	436	86	88	68	0,772727	0,790698	0,781609
37	320	74	64	49	0,765625	0,662162	0,710145
38	350	77	86	54	0,627907	0,701299	0,662577
39	506	130	95	80	0,842105	0,615385	0,711111
40	380	89	90	55	0,611111	0,617978	0,614525
41	417	99	97	69	0,71134	0,69697	0,704082
42	571	124	120	73	0,608333	0,58871	0,598361
43	376	114	84	53	0,630952	0,464912	0,535354
44	666	187	189	153	0,809524	0,818182	0,81383
45	432	84	99	68	0,686869	0,809524	0,743169

No	A	GT	RP	O	Precision	Recall	F-score
46	454	121	111	68	0,612613	0,561983	0,586207
47	472	92	92	55	0,597826	0,597826	0,597826
48	576	143	125	106	0,848	0,741259	0,791045
49	556	138	121	76	0,628099	0,550725	0,586873
50	405	114	87	66	0,758621	0,578947	0,656716
51	554	143	77	59	0,766234	0,412587	0,536364
52	518	122	96	74	0,770833	0,606557	0,678899
53	393	62	64	43	0,671875	0,693548	0,68254
54	394	105	82	54	0,658537	0,514286	0,57754
55	931	241	166	105	0,63253	0,435685	0,515971
56	390	92	80	54	0,675	0,586957	0,627907
57	637	142	148	116	0,783784	0,816901	0,8
58	1103	286	268	183	0,682836	0,63986	0,66065
59	652	179	136	82	0,602941	0,458101	0,520635
60	387	78	66	44	0,666667	0,564103	0,611111
61	395	134	73	34	0,465753	0,253731	0,328502
62	666	137	122	82	0,672131	0,59854	0,633205
63	759	158	157	102	0,649682	0,64557	0,647619
64	465	104	98	73	0,744898	0,701923	0,722772
65	432	113	97	65	0,670103	0,575221	0,619048
66	359	68	87	56	0,643678	0,823529	0,722581
67	278	56	43	13	0,302326	0,232143	0,262626
68	225	52	39	22	0,564103	0,423077	0,483516
69	434	148	114	76	0,666667	0,513514	0,580153
70	494	118	129	78	0,604651	0,661017	0,631579
71	219	65	45	13	0,288889	0,2	0,236364
72	385	74	55	38	0,690909	0,513514	0,589147
73	874	187	161	111	0,689441	0,593583	0,637931
74	523	87	93	69	0,741935	0,793103	0,766667
75	517	106	86	60	0,697674	0,566038	0,625
76	370	88	92	62	0,673913	0,704545	0,688889
77	335	90	80	50	0,625	0,555556	0,588235
78	478	114	122	63	0,516393	0,552632	0,533898
79	530	151	135	75	0,555556	0,496689	0,524476
80	328	70	80	43	0,5375	0,614286	0,573333
	Rata-rata				0,66387	0,610925	0,630851

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

### Lampiran 39 :

Hasil ROUGE 1 - Textrank dengan TF-IDF Cosine *similarity*, dengan

Compression Rate 30%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	131	112	95	0,848214	0,725191	0,781893
2	450	152	144	117	0,8125	0,769737	0,790541
3	565	200	140	106	0,757143	0,53	0,623529
4	528	174	139	95	0,683453	0,545977	0,607029
5	657	259	211	171	0,810427	0,660232	0,72766
6	447	144	156	110	0,705128	0,763889	0,733333
7	966	254	333	230	0,690691	0,905512	0,783646
8	509	188	141	110	0,780142	0,585106	0,668693
9	386	157	132	115	0,871212	0,732484	0,795848
10	368	144	132	117	0,886364	0,8125	0,847826
11	528	198	165	103	0,624242	0,520202	0,567493
12	513	194	190	152	0,8	0,783505	0,791667
13	644	221	209	169	0,808612	0,764706	0,786047
14	534	206	181	146	0,80663	0,708738	0,754522
15	285	105	90	68	0,755556	0,647619	0,697436
16	657	246	180	162	0,9	0,658537	0,760563
17	799	313	300	254	0,846667	0,811502	0,828711
18	1029	313	328	277	0,844512	0,884984	0,864275
19	340	102	96	70	0,729167	0,686275	0,707071
20	695	244	245	153	0,62449	0,627049	0,625767
21	574	203	183	129	0,704918	0,635468	0,668394
22	532	171	166	155	0,933735	0,906433	0,919881
23	321	124	105	90	0,857143	0,725806	0,786026
24	373	153	125	69	0,552	0,45098	0,496403
25	366	121	130	107	0,823077	0,884298	0,85259
26	329	117	99	89	0,89899	0,760684	0,824074
27	383	115	117	108	0,923077	0,93913	0,931034
28	359	118	121	90	0,743802	0,762712	0,753138
29	352	123	116	72	0,62069	0,585366	0,60251
30	409	157	156	119	0,762821	0,757962	0,760383
31	592	223	176	140	0,795455	0,627803	0,701754
32	468	155	125	95	0,76	0,612903	0,678571
33	481	137	140	99	0,707143	0,722628	0,714801
34	331	114	106	74	0,698113	0,649123	0,672727
35	364	98	92	70	0,76087	0,714286	0,736842
36	436	152	129	109	0,844961	0,717105	0,775801
37	320	136	110	88	0,8	0,647059	0,715447
38	350	123	112	98	0,875	0,796748	0,834043
39	506	212	181	159	0,878453	0,75	0,80916
40	380	141	115	98	0,852174	0,695035	0,765625
41	417	176	160	137	0,85625	0,778409	0,815476
42	571	151	181	101	0,558011	0,668874	0,608434
43	376	142	125	91	0,728	0,640845	0,681648
44	666	276	253	211	0,833992	0,764493	0,797732
45	432	160	134	122	0,910448	0,7625	0,829932

No	A	GT	RP	O	Precision	Recall	F-score
46	454	161	146	100	0,684932	0,621118	0,651466
47	472	125	135	75	0,555556	0,6	0,576923
48	576	216	184	156	0,847826	0,722222	0,78
49	556	216	197	154	0,781726	0,712963	0,745763
50	405	154	111	72	0,648649	0,467532	0,543396
51	554	215	138	109	0,789855	0,506977	0,617564
52	518	162	144	130	0,902778	0,802469	0,849673
53	393	120	116	101	0,87069	0,841667	0,855932
54	394	148	126	89	0,706349	0,601351	0,649635
55	931	364	246	208	0,845528	0,571429	0,681967
56	390	130	128	80	0,625	0,615385	0,620155
57	637	205	211	157	0,744076	0,765854	0,754808
58	1103	448	382	326	0,853403	0,727679	0,785542
59	652	249	213	164	0,769953	0,658635	0,709957
60	387	116	105	91	0,866667	0,784483	0,823529
61	395	189	125	108	0,864	0,571429	0,687898
62	666	215	204	157	0,769608	0,730233	0,749403
63	759	259	236	181	0,766949	0,698842	0,731313
64	465	151	137	98	0,715328	0,649007	0,680556
65	432	164	146	109	0,746575	0,664634	0,703226
66	359	112	135	96	0,711111	0,857143	0,777328
67	278	105	75	29	0,386667	0,27619	0,322222
68	225	74	54	34	0,62963	0,459459	0,53125
69	434	200	144	125	0,868056	0,625	0,726744
70	494	174	174	125	0,718391	0,718391	0,718391
71	219	93	50	21	0,42	0,225806	0,293706
72	385	110	123	85	0,691057	0,772727	0,729614
73	874	247	235	186	0,791489	0,753036	0,771784
74	523	170	165	117	0,709091	0,688235	0,698507
75	517	176	148	110	0,743243	0,625	0,679012
76	370	124	126	78	0,619048	0,629032	0,624
77	335	146	134	101	0,753731	0,691781	0,721429
78	478	178	172	129	0,75	0,724719	0,737143
79	530	214	189	141	0,746032	0,658879	0,699752
80	328	123	106	84	0,792453	0,682927	0,733624
	Rata-rata				0,761871	0,684783	0,71799

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

## Lampiran 40 :

Hasil ROUGE 1 - Textrank dengan TF-IDF Cosine *similarity*, dengan

Compression Rate 40%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	164	131	107	0,816794	0,652439	0,725424
2	450	185	174	139	0,798851	0,751351	0,774373
3	565	267	214	164	0,766355	0,614232	0,681913
4	528	239	217	162	0,746544	0,677824	0,710526
5	657	299	276	242	0,876812	0,809365	0,841739
6	447	200	197	164	0,832487	0,82	0,826196
7	966	360	423	322	0,761229	0,894444	0,822478
8	509	246	190	158	0,831579	0,642276	0,724771
9	386	179	168	141	0,839286	0,787709	0,81268
10	368	180	167	141	0,844311	0,783333	0,81268
11	528	271	219	151	0,689498	0,557196	0,616327
12	513	272	252	214	0,849206	0,786765	0,816794
13	644	311	273	238	0,871795	0,765273	0,815068
14	534	263	237	210	0,886076	0,798479	0,84
15	285	133	130	88	0,676923	0,661654	0,669202
16	657	306	251	211	0,840637	0,689542	0,75763
17	799	414	402	372	0,925373	0,898551	0,911765
18	1029	447	474	349	0,736287	0,780761	0,757872
19	340	132	135	120	0,888889	0,909091	0,898876
20	695	330	318	257	0,808176	0,778788	0,79321
21	574	265	240	181	0,754167	0,683019	0,716832
22	532	247	235	206	0,876596	0,834008	0,854772
23	321	162	134	118	0,880597	0,728395	0,797297
24	373	179	181	115	0,635359	0,642458	0,638889
25	366	159	167	134	0,802395	0,842767	0,822086
26	329	155	150	122	0,813333	0,787097	0,8
27	383	167	154	139	0,902597	0,832335	0,866044
28	359	158	163	112	0,687117	0,708861	0,697819
29	352	171	160	133	0,83125	0,777778	0,803625
30	409	209	198	173	0,873737	0,827751	0,850123
31	592	277	215	172	0,8	0,620939	0,699187
32	468	224	197	173	0,878173	0,772321	0,821853
33	481	185	196	151	0,770408	0,816216	0,792651
34	331	160	143	113	0,79021	0,70625	0,745875
35	364	158	141	122	0,865248	0,772152	0,816054
36	436	186	171	133	0,777778	0,715054	0,745098
37	320	161	141	118	0,836879	0,732919	0,781457
38	350	158	150	117	0,78	0,740506	0,75974
39	506	276	215	186	0,865116	0,673913	0,757637
40	380	167	150	128	0,853333	0,766467	0,807571
41	417	231	204	187	0,916667	0,809524	0,85977
42	571	229	245	172	0,702041	0,751092	0,725738
43	376	170	162	114	0,703704	0,670588	0,686747
44	666	386	366	310	0,846995	0,803109	0,824468
45	432	202	171	142	0,830409	0,70297	0,761394

No	A	GT	RP	O	Precision	Recall	F-score
46	454	201	190	153	0,805263	0,761194	0,782609
47	472	171	196	124	0,632653	0,725146	0,675749
48	576	272	238	188	0,789916	0,691176	0,737255
49	556	262	246	178	0,723577	0,679389	0,700787
50	405	179	141	102	0,723404	0,569832	0,6375
51	554	274	170	134	0,788235	0,489051	0,603604
52	518	215	212	174	0,820755	0,809302	0,814988
53	393	168	159	148	0,930818	0,880952	0,905199
54	394	198	176	149	0,846591	0,752525	0,796791
55	931	477	337	284	0,84273	0,595388	0,697789
56	390	174	173	135	0,780347	0,775862	0,778098
57	637	304	288	243	0,84375	0,799342	0,820946
58	1103	561	513	425	0,82846	0,757576	0,791434
59	652	313	272	204	0,75	0,651757	0,697436
60	387	161	151	120	0,794702	0,745342	0,769231
61	395	236	170	150	0,882353	0,635593	0,738916
62	666	281	249	188	0,75502	0,669039	0,709434
63	759	355	315	239	0,75873	0,673239	0,713433
64	465	209	196	146	0,744898	0,698565	0,720988
65	432	231	191	159	0,832461	0,688312	0,753555
66	359	152	165	126	0,763636	0,828947	0,794953
67	278	134	125	99	0,792	0,738806	0,764479
68	225	93	91	66	0,725275	0,709677	0,717391
69	434	235	211	185	0,876777	0,787234	0,829596
70	494	240	219	184	0,840183	0,766667	0,801743
71	219	123	86	67	0,77907	0,544715	0,641148
72	385	134	159	104	0,654088	0,776119	0,709898
73	874	349	324	278	0,858025	0,796562	0,826152
74	523	214	188	153	0,81383	0,714953	0,761194
75	517	242	187	150	0,802139	0,619835	0,699301
76	370	172	155	120	0,774194	0,697674	0,733945
77	335	178	167	129	0,772455	0,724719	0,747826
78	478	240	197	172	0,873096	0,716667	0,787185
79	530	294	271	224	0,826568	0,761905	0,79292
80	328	173	160	137	0,85625	0,791908	0,822823
	Rata-rata				0,805543	0,735032	0,766482

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.

## Lampiran 41 :

Hasil ROUGE 1 - Textrank dengan TF-IDF Cosine *similarity*, dengan

Compression Rate 50%

No	A	GT	RP	O	Precision	Recall	F-score
1	331	194	154	141	0,915584	0,726804	0,810345
2	450	233	216	178	0,824074	0,763948	0,792873
3	565	308	263	223	0,847909	0,724026	0,781086
4	528	295	276	218	0,789855	0,738983	0,763573
5	657	345	319	284	0,890282	0,823188	0,855422
6	447	221	226	191	0,845133	0,864253	0,854586
7	966	461	553	415	0,750452	0,900217	0,81854
8	509	301	253	227	0,897233	0,754153	0,819495
9	386	218	213	164	0,769953	0,752294	0,761021
10	368	219	203	178	0,876847	0,812785	0,843602
11	528	338	280	228	0,814286	0,674556	0,737864
12	513	320	283	239	0,844523	0,746875	0,792703
13	644	365	343	288	0,83965	0,789041	0,813559
14	534	345	284	258	0,908451	0,747826	0,82035
15	285	160	155	125	0,806452	0,78125	0,793651
16	657	352	347	299	0,861671	0,849432	0,855508
17	799	507	483	443	0,917184	0,873767	0,894949
18	1029	560	578	463	0,801038	0,826786	0,813708
19	340	171	171	141	0,824561	0,824561	0,824561
20	695	404	350	290	0,828571	0,717822	0,769231
21	574	323	290	246	0,848276	0,76161	0,80261
22	532	310	298	250	0,838926	0,806452	0,822368
23	321	201	195	158	0,810256	0,78607	0,79798
24	373	204	196	131	0,668367	0,642157	0,655
25	366	222	216	194	0,898148	0,873874	0,885845
26	329	192	191	172	0,900524	0,895833	0,898172
27	383	214	181	159	0,878453	0,742991	0,805063
28	359	202	192	153	0,796875	0,757426	0,77665
29	352	221	182	145	0,796703	0,656109	0,719603
30	409	247	234	219	0,935897	0,88664	0,910603
31	592	356	270	227	0,840741	0,63764	0,72524
32	468	290	261	238	0,911877	0,82069	0,863884
33	481	244	242	208	0,859504	0,852459	0,855967
34	331	197	169	139	0,822485	0,705584	0,759563
35	364	196	164	140	0,853659	0,714286	0,777778
36	436	220	197	163	0,827411	0,740909	0,781775
37	320	186	167	135	0,808383	0,725806	0,764873
38	350	201	180	139	0,772222	0,691542	0,729659
39	506	325	273	238	0,871795	0,732308	0,795987
40	380	219	175	154	0,88	0,703196	0,781726
41	417	283	243	228	0,938272	0,805654	0,86692
42	571	308	321	244	0,760125	0,792208	0,775835
43	376	184	201	126	0,626866	0,684783	0,654545
44	666	476	389	356	0,915167	0,747899	0,823121
45	432	250	219	191	0,872146	0,764	0,814499

No	A	GT	RP	O	Precision	Recall	F-score
46	454	236	254	190	0,748031	0,805085	0,77551
47	472	222	235	176	0,748936	0,792793	0,770241
48	576	322	284	240	0,84507	0,745342	0,792079
49	556	313	304	242	0,796053	0,773163	0,784441
50	405	210	174	142	0,816092	0,67619	0,739583
51	554	300	208	168	0,807692	0,56	0,661417
52	518	310	267	248	0,928839	0,8	0,859619
53	393	213	204	193	0,946078	0,906103	0,925659
54	394	225	201	163	0,810945	0,724444	0,765258
55	931	563	424	352	0,830189	0,625222	0,713273
56	390	202	208	160	0,769231	0,792079	0,780488
57	637	380	356	326	0,91573	0,857895	0,88587
58	1103	697	634	550	0,867508	0,789096	0,826446
59	652	374	342	276	0,807018	0,737968	0,77095
60	387	214	187	169	0,903743	0,78972	0,842893
61	395	284	222	201	0,905405	0,707746	0,794466
62	666	364	332	287	0,864458	0,788462	0,824713
63	759	412	387	304	0,78553	0,737864	0,760951
64	465	274	249	211	0,84739	0,770073	0,806883
65	432	265	225	182	0,808889	0,686792	0,742857
66	359	191	202	169	0,836634	0,884817	0,860051
67	278	146	140	110	0,785714	0,753425	0,769231
68	225	133	132	132	1	0,992481	0,996226
69	434	268	236	222	0,940678	0,828358	0,880952
70	494	292	255	225	0,882353	0,770548	0,822669
71	219	129	90	68	0,755556	0,527132	0,621005
72	385	191	193	132	0,683938	0,691099	0,6875
73	874	447	424	410	0,966981	0,917226	0,941447
74	523	270	228	205	0,899123	0,759259	0,823293
75	517	284	236	199	0,84322	0,700704	0,765385
76	370	211	201	164	0,81592	0,777251	0,796117
77	335	227	215	181	0,84186	0,797357	0,819005
78	478	276	244	217	0,889344	0,786232	0,834615
79	530	339	301	265	0,880399	0,781711	0,828125
80	328	229	218	201	0,922018	0,877729	0,899329
	Rata-rata				0,843542	0,769151	0,80288

Keterangan : A merupakan jumlah kata pada teks asli, GT merupakan jumlah kata pada *ground truth*, RP merupakan jumlah kata pada ringkasan prediksi, O merupakan jumlah kata yang saling tumpang tindih (*overlap*) pada ringkasan prediksi dan *ground truth*.