

Teks Ringkas Otomatis pada Portal Berita CNN Indonesia Menggunakan Algoritma Textrank

Lizza Nur Fadhila¹, I Kadek Dwi Nuryana²

^{1,2} Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya

lizza.17051214019@mhs.unesa.ac.id

dwinuryana@unesa.ac.id

Abstrak— Peringkasan teks otomatis dapat membantu pembaca artikel *website* dalam memahami dengan singkat inti dari sebuah artikel. Konsep dari peringkasan teks otomatis yaitu dapat mencari bagian penting informasi yang berasal dari teks artikel asli, dengan program tersebut hasil dari rangkuman dapat di jadikan lebih ringkas, padat, dan jelas sesuai dengan topik utama dari artikel yang akan di ringkas. Program ini akan menggunakan bahasa pemrograman Python dengan menggunakan Platform Jupyter untuk menjalankan Algoritma *Textrank*. Untuk bahasa yang digunakan dalam program ini yaitu bahasa Indonesia dengan melakukan proses *cleaning* terlebih dahulu seperti menghilangkan tanda baca pada teks agar program dapat berjalan dengan baik. Untuk penggunaan program ini penulis telah membuat format file .txt dari sumber artikel berita CNN Indonesia dan menjalankan program dengan menggunakan jupyter pada browser untuk melihat hasil perangkuman teks artikel tersebut. Penelitian ini bertujuan untuk (1) Mengetahui pengimplementasian Algoritma *textrank* melalui platform Jupyter Notebook, (2) Mengetahui presentase perbandingan yang dapat diperoleh dari teks artikel asli dan hasil peringkasan dengan Algoritma *textrank*, dan (3) Mengetahui Algoritma *textrank* dapat efektif dalam meringkas text berita dalam bahasa Indonesia. Berdasarkan hasil pengujian dihasilkan bahwa rata-rata presentase ringkasan sebesar 6,9% lebih ringkas daripada artikel asli, dengan catatan penulis membatasi jumlah kalimat sebesar 3 kalimat dan hasil analisis Algoritma *textrank* efektif dalam meringkas *text* berita dalam bahasa Indonesia.

Kata Kunci— *Text*, *TextRank*, *Alogaritme*, *Jupyter Notebook*, *Phyton*,

I. PENDAHULUAN

Berita teks merupakan salah satu contoh media yang beralih ke internet berupa artikel pada website. Berita teks pada internet sangat lengkap dan lebih mudah di akses kapan saja dan dimana saja melalui internet. Namun terkadang berita teks berupa artikel pada website memiliki kalimat yang Panjang dengan paragraph yang banyak. Sehingga pembaca kesulitan menemukan inti dari suatu topik berita pada artikel

tersebut.[1] Pada sebuah website artikel sering muncul iklan yang sengaja dibuat oleh pembuat website, iklan tersebut terkadang muncul di tiap batas paragraph artikel, sehingga mengganggu fokus pembaca, hal ini mengganggu kenyamanan para pembaca artikel untuk mencerna inti dari artikel text tersebut. Text Summarization ataupun ringkasan bacaan awal kali diperkenalkan pada tahun 1950 (Gaikwad & amp; Mahender, 2016), ialah sesuatu tata cara dalam membuat ringkasan yang pendek, akurat, serta bisa di cerna dari sesuatu dokumen bacaan yang panjang. Sebaliknya automatic text summarization merupakan sesuatu langkah buat membuat ringkasan secara otomatis dengan tidak mengganti inti dari sesuatu dokumen yang diringkaskan [2].

Algoritma *Textrank* bergantung pada bentuk graf untuk menentukan usulan penting atau tidaknya menggunakan simpul (node/vertex). Dalam mengimplementasikan algoritma ini, langkah pertama adalah membuat grafik yang terdiri dari titik-titik yang mewakili setiap himpunan. Setiap himpunan digabungkan berdasarkan kemiripan (*similarity*) antar himpunan, sering disebut sebagai edge [2].

Penelitian ini diharapkan dapat membantu pembaca artikel website dalam memahami dengan singkat inti dari sebuah artikel. Dengan menggunakan algoritma *textrank*, pembaca artikel *website* dapat mencari bagian penting informasi yang berasal dari teks artikel asli menjadi lebih ringkas, padat, dan jelas sesuai dengan topik utama dari artikel yang dibaca

II. KAJIAN TEORI

A. Penelitian Terdahulu

Penelitian sebelumnya dilakukan dengan tujuan untuk memperoleh landasan dan pedoman yang menjadi dasar dalam penulisan skripsi ini. Selain itu, hal ini juga dilakukan untuk menghindari kesan bahwa penelitian ini sama seperti yang telah dilakukan sebelumnya. Oleh karena itu, dalam tinjauan pustaka ini, peneliti memasukkan hasil-hasil dari penelitian sebelumnya sebagai referensi berikut:

Penelitian pertama yaitu oleh Deepali Jain, Malaya Dutta Borah, Anupam Biswas pada tahun 2020 dengan judul *Fine-Tuning Textrank for Legal Document Summarization : A Bayesian Optimization Based Approach* Hasil eksperimen

terhadap dua kumpulan data peringkasan dokumen hukum, BillSum dan Forum of Information Retrieval Evaluation (FIRE), mengungkapkan bahwa pendekatan yang diusulkan dapat mencapai perbaikan yang signifikan dibandingkan pendekatan mutakhir saat ini. Lebih khusus lagi, ini mencapai peningkatan skor F1 metrik ROUGE sebesar (1–6)% dan (6–12)% untuk set pengujian BillSum dan FIRE masing-masing. Berdasarkan Hasil ringkasan yang mengesankan ini menunjukkan kegunaan pendekatan yang diusulkan dalam menemukan inti dari sebuah dokumen hukum yang panjang, sehingga memberikan bantuan penting bagi para praktisi hukum [3].

Penelitian kedua yang dilakukan oleh Mingxi Zhang, Xuemin Li, Shuibo Yue, Liuqian Yang pada tahun 2020 dengan judul *An Empirical Study of TextRank For Keyword Extraction*. Berdasarkan studi empiris tentang *Textrank* untuk menguji efektivitas pada pengaturan parameter yang berbeda, termasuk jendela kejadian bersama, nomor iterasi, peluruhan factor, peringkat dalam daftar yang dikembalikan dan panjang teks menunjukkan bahwa *Textrank* bekerja baik saat jendela kejadian bersama ukuran $w = 3$, nomor iterasi $t = 20$, faktor peluruhan $c = 0,9$ dan peringkat $k = 10$ masing-masing, dan hasilnya independen dari panjang teks [4].

Penelitian ketiga dilakukan oleh K. G. Subhawa, I. Atastina, A. Nursikuwagus pada tahun 2019 dengan judul *Penerapan Algoritma Text Mining dan Lexrank dalam Meringkas Teks Secara Otomatis*. Hasil penelitian yang dilakukan yaitu Penerapan algoritma text mining yang diterapkan untuk meringkas teks secara otomatis masih terbilang lemah dikarenakan masih memiliki nilai yang sama. Adanya nilai bobot dari suatu term dapat memudahkan paragraf mana yang akan diawali untuk diringkas Hasil ringkasan dari sistem peringkasan teks otomatis pada penelitian ini adalah kalimat inti yang mirip dengan query dan berdasarkan urutan bobot [5].

B. Landasan Teori

Landasan teori untuk penelitian mengenai peringkasan teks otomatis pada portal berita menggunakan algoritma *textrank* mencakup peringkasan teks otomatis, berita, CNN Indonesia, algoritma *textrank*. Berikut adalah penjelasan singkat tentang setiap landasan teori tersebut:

1) *Peringkasan Teks Otomatis*: Text Summarization ataupun ringkasan bacaan awal kali diperkenalkan pada tahun 1950 (Gaikwad& Mahender, 2016), ialah sesuatu tata cara dalam membuat ringkasan yang pendek, akurat, serta bisa di cerna dari sesuatu dokumen bacaan yang panjang. Sebaliknya automatic text summarization merupakan sesuatu langkah buat membuat ringkasan secara otomatis dengan tidak mengganti inti dari sesuatu dokumen yang diringkas.

Automatic text summarization ataupun biasa diucap automatic summarization sangat diperlukan pada masa big informasi saat ini ini, di mana jumlah informasi bacaan tiap hari senantiasa meningkat dengan sangat banyak serta tidak terstruktur sehingga buat menolong

menciptakan data yang relevan dengan lebih kilat diperlukan automatic summarization [6].

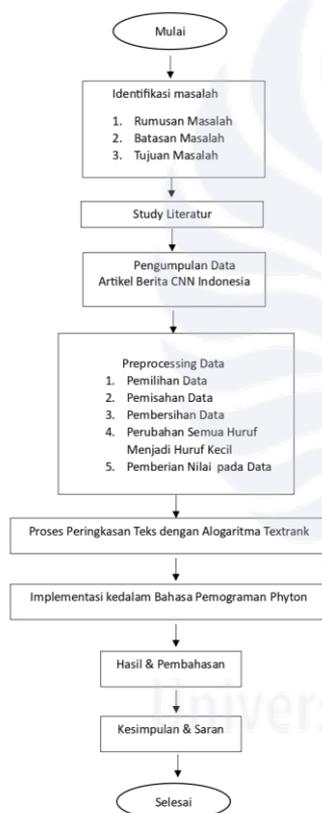
- 2) *Berita*: Berita menurut Adi Bajuri adalah laporan suatu peristiwa yang sudah terjadi, gagasan atau pendapat seseorang atau kelompok atau temuan baru di segala bidang yang dipandang penting untuk diliput wartawan yang bertujuan untuk dimuat dalam media.
- 3) *CNN Indonesia*: CNN Indonesia adalah sebuah platform berita digital gratis dan berbayar milik Trans Media dengan menambil lisensi nama CNN dari Warner Bris. Saluran ini disiarkan pertama kalinya pada Senin, 17 Agustus 2015. Peluncuran utamanya berlangsung sejak 15 Desember 2015 dalam rangka memperingati ulang tahun Trans Media ke-14. Menurut sebuah laporan yang dikeluarkan Reuters Institute for the Study of Journalism dan Universitas Oxford pada tahun 2021, CNN Indonesia merupakan media yang paling dipercaya masyarakat dengan skor kepercayaan mencapai 69%.
- 4) *Project Jupyter* : proyek dan komunitas yang tujuannya adalah untuk "mengembangkan perangkat lunak sumber terbuka, standar terbuka, dan layanan untuk komputasi interaktif di puluhan bahasa pemrograman". Itu dipisahkan dari IPython pada tahun 2014 oleh Fernando Pérez. Nama Project Jupyter adalah referensi ke tiga bahasa pemrograman inti yang didukung oleh Jupyter, yaitu Julia, Python dan R, dan juga penghormatan kepada notebook Galileo yang merekam penemuan bulan Jupiter. Project Jupyter telah mengembangkan dan mendukung produk komputasi interaktif Jupyter Notebook, JupyterHub, dan JupyterLab [7].
- 5) *Python* adalah bahasa pemrograman berorientasi objek interaktif yang menyediakan struktur data tingkat tinggi. Python adalah bahasa pemrograman interpretatif multiguna dengan perancangan yang fokus pada keterbacaan kode agar sintaks yang digunakan lebih mudah dipahami. Python dikatakan sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaks kode yang sangat jelas. Bahasa pemrograman Python khusus dirancang untuk memudahkan programmer dalam membuat program baik itu dari efisiensi waktu, mudah dalam pengembangan dan kompalibilitasnya dengan sistem. Python dapat digunakan untuk membuat aplikasi standalone (berdiri sendiri) atau pemrograman script [2].
- 6) *Textrank*: Menurut Pinandhita, 2013, *Textrank* merupakan *graph-based ranking algorithm* (graf dengan model pemeringkatan) untuk pemrosesan teks dari dokumen bahasa alami atau manusia. Dokumen

yang diolah berupa dokumen tunggal (single-document). Menurut Mihalcea & Tarau (2004), ada 4 tahapan dalam mengimplementasikan algoritma berbasis graf dalam pemrosesan teks bahasa alami, yaitu:

- Identifikasi unit teks yang paling sesuai untuk dijadikan simpul pada graf.
- Pemberian sisi antar simpul unit teks baik dengan bobot atau tidak berarah atau tidak berarah.
- Proses menggunakan algoritma hingga objek graf bertemu satu dengan lainnya (convergence).
- Urutkan simpul berdasarkan skornya. Nilai skor selanjutnya dimasukkan kedalam simpul pada saat proses Algoritma [8].

III. METODOLOGI PENELITIAN

A. Tahapan Penelitian



Gbr 1. Rancangan penelitian

Berikut penjelasan dari tahapan penelitian :

- Identifikasi Masalah**
Tahap ini merupakan identifikasi masalah untuk menemukan suatu masalah yang dihadapi oleh para pembaca artikel yang kesulitan menemukan inti teks artikel suatu berita.
- Study Literatur**

Pada tahapan ini penulis mencari informasi terkait permasalahan yang akan diteliti dengan cara mengumpulkan beberapa referensi dari beberapa sumber yang diperoleh yaitu dari skripsi lampau, jurnal penelitian, paper, buku, dan beberapa sumber lainnya.

- Data dan Informasi**
Setelah identifikasi masalah, tahap selanjutnya adalah mencari data dan informasi yang berguna sebagai dasar dalam melakukan pengembangan sistem peringkasan teks otomatis menggunakan algoritma textrank untuk menghasilkan inti teks berita yang singkat, padat, dan jelas.
- Proses Pengembangan Sistem**
Pada tahapan ini peneliti mengimplementasikan hasil dari analisis sistem dan perancangan sistem pada tahap sebelumnya kedalam kode program sistem aplikasi menggunakan bahasa pemrograman Python sebagai back-end untuk memproses Algoritma textrank dan menampilkannya kedalam platform Jupyter Notebook. Hasil dan Pembahasan
- Analisis dan Pembahasan**
Setelah melakukan pengembangan sistem, maka sistem akan dilakuka uji coba untuk mengetahui kelemahan ataupun kesalahan dari sistem yang dimana hasil dari pengujian ini dapat berguna sebagai acuan untuk memperbaiki sistem dan pengujian sistem telah dibuat dengan hasil manual ekstraksi kata kunci dari ahli yang telah dipilih.
- Kesimpulan**
Pada tahap ini menunjukkan hasil dari peringkasan teks otomatis pada portal berita CNN Indonesia menggunakan algoritma textrank secara keseluruhan.

B. Metode Pengumpulan Data

Pada penelitian ini, penulis mengumpulkan data yakni berupa :

- Sumber Data:** Data utama yang di gunakan dalam penelitian ini yaitu teks artikel berita dari *website* CNN Indonesia.
- Jenis Data:** Data Kualitatif berupa teks artikel berita dari *website* CNN Indonesia yang tiap artikelnya berjumlah antara 100 – 1000 kata.

Efek Buruk Sering Nyetir Kendaraan Saat Jarum Bensin di Posisi E
Jakarta, CNN Indonesia -- Malas isi bensin dan sering mengemudikan kendaraan ketika posisi jarum indikator bahan bakar di posisi 'E' (empty) adalah kebiasaan buruk. Hal ini bisa merusak kendaraan untuk jangka panjang. Saat di posisi 'E' (empty), sebenarnya kendaraan anda tak akan langsung mati dan tak bisa digunakan. Menjelang mati bakal muncul gejala-gejala yang menunjukkan kendaraan anda lagi kelaparan. Gejala itu misalnya mesin batuk-batuk, tenaga berkurang dan mungkin juga muncul juga letupan-letupan di knalpot. Kendaraan yang dipaksa tetap bekerja saat bensin tiris bakal berdampak negatif. Ketika ini dilakukan ada kemungkinan pompa bahan bakar bakal bekerja keras sampai menyedot kotoran yang ada di dasar tangki bahan bakar. Kotoran itu dapat tersangkut dan menutup aliran bahan bakar atau bisa juga merusak pompa bahan bakar dan injektor.

menghitung hubungan similarity antar kalimat. Berikut tabel hasil pengelompokan kata

Tabel 1. Hasil Pengelompokan Kata Pada Artikel CNN Indonesia

Selain itu pompa bahan bakar membutuhkan aliran bahan bakar sebagai pelumas dan pendinginan dan bila tidak tersedia maka udara juga bisa masuk ke sistem jalur bahan bakar ke mesin.
 Jika Anda tak mau merusak kendaraan untuk jangka panjang dan boncos biaya perawatan besar, sebaiknya segera berhenti mengemudi bila bensin sudah habis.
 Pastikan Anda menggunakan tetes-tetes bahan bakar terakhir buat mengarahkan kendaraan ke tempat paling aman sambil menunggu bantuan. Anda juga bisa memasang lampu hazard untuk menunjukkan kondisi darurat. Satu hal lagi yang perlu diperhatikan, bila pompa bahan bakar sudah rusak karena kehabisan bahan bakar maka anda kemungkinan tak bisa menyalakan mobil ketika tangki diisi bahan bakar.
 Ini terjadi karena aliran bahan bakar bisa jadi tersumbat atau sudah disusupi udara. Cara mengatasinya Anda perlu memanggil mekanik.

Gbr 2. Data Teks Artikel Berita CNN Indonesia

C. Metode Analisis Data

Pada sub bab ini akan di jelaskan penerapan metode algoritma textrank untuk menghasilkan nilai pada setiap kalimat teks artikel berita untuk peringkasan teks otomatis. Berikut adalah tahapan algoritma textrank:

- 1) Persiapan data teks artikel berita yang diambil dari website CNN Indonesia dari berbagai kategori berita yang pada website tersebut.
- 2) Tahap pre-processing, pada tahap ini terdiri dari beberapa langkah, yaitu : *sentence splitting, cleaning, case folding, tokenizing.*
- 3) Tahap Processing: pada tahap ini kalimat yang telah diberi nilai token awal akan dihitung nilai similarity-nya. Kemudian dari hasil score tersebut akan diurutkan dari score tertinggi yang akan menjadi hasil teks peringkasan. Adapun langkah-langkah yang diperlukan dalam tahap processing ini yaitu :
 - a) Mempresentasikan graph
 - b) Menghitung similarity antar kalimat
 - c) Menjumlahkan score pada tiap kalimat
 - d) Memilih kalimat yang paling penting

Sebelum mepresentasikan graph, diperlukan proses pengelompokan kata untuk mempermudah peneliti dalam

No.	Token
1.	jakarta cnn indonesia malas isi bensin dan sering mengemudi kendaraan ketika posisi jarum indikator bahan bakar di posisi empty adalah kebiasaan buruk
2.	hal ini bisa merusak kendaraan untuk jangka panjang
3.	saat di posisi empty sebenarnya kendaraan anda tak akan langsung mati dan tak bisa digunakan
4.	menjelang mati bakal muncul gejala gejala yang menunjukkan kendaraan anda lagi kelaparan
5.	gejala itu misalnya mesin batuk batuk tenaga berkurang dan mungkin juga muncul juga letupan letupan di knalpot
6.	kendaraan yang dipaksa tetap bekerja saat bensin tiris bakal berdampak negatif
7.	ketika ini dilakukan ada kemungkinan pompa bahan bakar bakal bekerja keras sampai menyedot kotoran yang ada di dasar tangki bahan bakar
8.	kotoran itu dapat tersangkut dan menutup aliran bahan bakar atau bisa juga merusak pompa bahan bakar dan injektor
9.	selain itu pompa bahan bakar membutuhkan aliran bahan bakar sebagai pelumas dan pendinginan dan bila tidak tersedia maka udara juga bisa masuk ke sistem jalur bahan bakar ke mesin
10.	jika anda tak mau merusak kendaraan untuk jangka panjang dan boncos biaya perawatan besar sebaiknya segera berhenti mengemudi bila bensin sudah habis
11.	pastikan anda menggunakan tetes tetes bahan bakar terakhir buat mengarahkan kendaraan ke tempat paling aman sambil menunggu bantuan
12.	anda juga bisa memasang lampu hazard untuk menunjukkan kondisi darurat
13.	satu hal lagi yang perlu diperhatikan bila pompa bahan bakar sudah rusak karena kehabisan bahan bakar maka anda kemungkinan tak bisa menyalakan mobil ketika tangki diisi bahan bakar
14.	ini terjadi karena aliran bahan bakar bisa jadi tersumbat atau sudah disusupi udara
15.	cara mengatasinya anda perlu memanggil mekanik

Untuk nilai overlap (W_k) pada vertex i dan vertex j :

$$\text{Similarity}(S_i, S_j) = \frac{|\{W_k | W_k \in S_i \& W_k \in S_j\}|}{\log(|S_i|) + \log(|S_j|)}$$

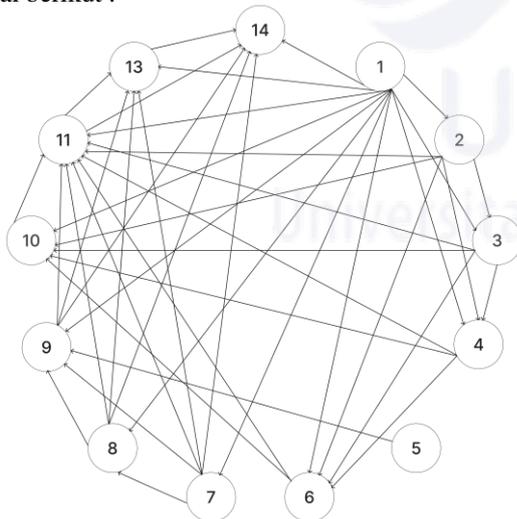
- 5) pemeringkatan antar kalimat sesuai dengan besarnya nilai pada vertex tersebut.

Berdasarkan hasil pengelompokan kata, maka selanjutnya ditemukan hubungan antar kalimat, sebagai berikut :

Table 2. Hubungan Antar Kalimat

Vertex	Hubungan
1 =	2 3 4 6 7 8 9 10 11 13 14
2 =	3 4 6 10 11
3 =	4 6 10 11
4 =	6 10 11
5 =	9
6 =	10 11
7 =	8 9 11 13 14
8 =	9 11 13 14
9 =	11 13 14
10 =	11
11 =	13 14
13 =	14

Hubungan kata similarity antar kalimat dapat digambarkan sebagai berikut :



Gbr 3. Graph Similarity Antar Kalimat

- 4) Menghitung nilai similarity antar kalimat:

D. Algoritma Textrank

Algoritma ini bergantung pada bentuk graf untuk menentukan usulan penting atau tidaknya menggunakan simpul (node/vertex). Dalam mengimplementasikan algoritma ini, langkah pertama adalah membuat grafik yang terdiri dari titik-titik yang mewakili setiap himpunan. Setiap himpunan digabungkan berdasarkan kemiripan (*similarity*) antar himpunan, sering disebut sebagai edge.

Menurut Mihalcea & Tarau (2004), ada 4 tahapan dalam mengimplementasikan algoritma berbasis graf dalam pemrosesan teks bahasa alami, yaitu:

1. Identifikasi unit teks yang paling sesuai untuk dijadikan simpul padagraf
2. Pemberian sisi antar simpul unit teks baik dengan bobot atau tidak berarah atau tidak berarah.
3. Proses menggunakan algoritma hingga objek graf bertemu satu dengan lainnya (*convergence*).
4. Urutkan simpul berdasarkan skornya. Nilai skor selanjutnya dimasukkan kedalam simpul pada saat proses Algoritma.

Dan untuk menghitung nilai bobot dari *similarity* kalimat menurut (Mihalcea & Tarau, 2004) menggunakan rumus *Weighted graph* dapat dilihat pada persamaan berikut

$$WS(V_i) = (1 - d) + d * \sum_{V_j \in In(V_i)} \frac{W_{ji}}{\sum_{V_k \in Out(V_i)} W_{jk}} WS(V_i)$$

Keterangan :

WS = Weight Sentence (Bobot Kalimat).

w = nilai *content overlap similarity*.

V_i = vertex yang dihitung skor-nya

V_j = vertex yang bertetangga dengan V_i .

V_k = vertex yang bertetangga dengan V_j .

d = *damping factor* yang nilainya antara 0 dan 1, biasanya 0,85.

Dimana nilai d digunakan sebagai faktor yang dapat diberi nilai 0 dan 1, biasanya diberi nilai konstanta 0.85. Pada vertex (V_i), rumus $In(V_i)$ dapat digunakan sebagai simpul yang akan mengarah ke simpul yang lainnya (*predecessors*) dan rumus $Out(V_j)$ digunakan sebagai simpul yang mengarah ke vertex V_j (*successors*) dan $WS(V_i)$ adalah hasil dari jumlah score simpul V_i .

Algoritma textrank bisa berguna sebagai menemukan keyword dalam peringkasan teks berita secara otomatis dengan mempertimbangkan kalimat-kalimat yang penting (*sentences extraction*). Berikut untuk menghitung nilai

similarity kalimat menurut (Mihalca, et al.2004) dapat dilihat pada persamaan :

$$\text{Similarity}(S_i, S_j) = \frac{|W_k | W_k \in S_i \& W_k \in S_j|}{\log(|S_i|) + \log(|S_j|)}$$

Keterangan :

- Wk = Jumlah kata (term) yang sama antara kalimat Si dan Sj.
- Si = Panjang kalimat Si.
- Sj = Panjang kalimat Sj.

IV. HASIL DAN PEMBAHASAN

Program yang akan di uji meliputi peringkasan yang dilakukan program terhadap hasil output yang lebih ringkas dan padat. Terdapat jumlah artikel yang diujikan dengan varian jumlah kata dari 100 sampai dengan 1000 kata. Pada pengambilan artikel di CNN Indonesia penulis memiliki keterbatasan dalam mengambil sumber artikel dikarenakan penulis menggunakan versi gratis pada website CNN Indonesia, sehingga memiliki akses terbatas untuk artikel *Efek Buruk Sering Nyetir Kendaraan Saat Jarum Bensin di Posisi E* Jakarta, CNN Indonesia -- Malas isi bensin dan sering mengemudi kendaraan ketika posisi jarum indikator bahan bakar di posisi 'E' (empty) adalah kebiasaan buruk. Hal ini bisa merusak kendaraan untuk jangka panjang. Saat di posisi 'E' (empty), sebenarnya kendaraan anda tak akan langsung mati dan tak bisa digunakan. Menjelang mati bakal muncul gejala-gejala yang menunjukkan kendaraan anda lagi kelaparan. Gejala itu misalnya mesin batuk-batuk, tenaga berkurang dan mungkin juga muncul juga letupan-letupan di knalpot. Kendaraan yang dipaksa tetap bekerja saat bensin tiris bakal berdampak negatif. Ketika ini dilakukan ada kemungkinan pompa bahan bakar bakal bekerja keras sampai menyedot kotoran yang ada di dasar tangki bahan bakar. Kotoran itu dapat tersangkut dan menutup aliran bahan bakar atau bisa juga merusak pompa bahan bakar dan injektor. Selain itu pompa bahan bakar membutuhkan aliran bahan bakar sebagai pelumas dan pendinginan dan bila tidak tersedia maka udara juga bisa masuk ke sistem jalur bahan bakar ke mesin. Jika Anda tak mau merusak kendaraan untuk jangka panjang dan boncos biaya perawatan besar, sebaiknya segera berhenti mengemudi bila bensin sudah habis. Pastikan Anda menggunakan tetes-tetes bahan bakar terakhir buat mengarahkan kendaraan ke tempat paling aman sambil menunggu bantuan. Anda juga bisa memasang lampu hazard untuk menunjukkan kondisi darurat. Satu hal lagi yang perlu diperhatikan, bila pompa bahan bakar sudah rusak karena kehabisan bahan bakar maka anda kemungkinan tak bisa menyalakan mobil ketika tangki diisi bahan bakar. Ini terjadi karena aliran bahan bakar bisa jadi tersumbat atau sudah disusupi udara. Cara mengatasinya Anda perlu memanggil mekanik.

Gbr 4. Artikel teks berita CNN Indonesia

artikel berita tersebut terdiri dari 15 kalimat. Setelah dilakukan pemeringkatan menggunakan algoritma textrank, didapatkan nilai top-rank dari setiap kalimat yang dapat dilihat pada table 4.1 sebagai berikut :

Tabel 3. Hasil Pemeringkatan Vertex Teks Artikel CNN Indonesia

No.	Vertex	Nilai
1	V ₁₁	0.8813
2	V ₁	0.5902
3	V ₈	0.4002
4	V ₇	0.3995
5	V ₁₄	0.3634
6	V ₉	0.3504
7	V ₁₃	0.3307
8	V ₆	0.3162
9	V ₃	0.2725
10	V ₂	0.2422
11	V ₄	0.2203
13	V ₁₀	0.219
14	V ₅	0.0217

Serta menghitung hasil presentase ringkasan sebagai berikut :

Tabel 4. Hasil Presentase Peringkasan Teks Otomatis

No.	Jumlah Teks Asli	Frekuensi Berita	Rata-rata Ringkasan (Kata/ 3 Kalimat)	Presentase Hasil Ringkasan
1	100 – 200	1	65	34%
2	200 – 300	5	78,8	5,8%
3	300 – 400	8	84	3,2%
4	400 – 500	5	98,6	4,5%
5	500 – 600	3	108	6,3%
6	600 – 700	0	0	0%
7	700 – 800	0	0	0%
8	800 – 900	1	72	8%
9	900 – 1000	0	0	0%
10	> 1000	2	104	4,9%

Dapat dilihat ketika menggunakan algoritma textrank pada peringkasan teks otomatis terhadap artikel berita menghasilkan ringkasan dengan rata-rata ringkasan 6,9%, hal ini menjadi lebih ringkas daripada teks artikel berita yang asli. Pada pengujian 25 teks artikel berita yang sudah melalui proses cleaning. Program textrank yang di buat penulis menempatkan batasan maksimal sebesar 3 kalimat untuk hasil output. Perhitungan yang di ambil pada output merupakan hasil ringkasan dari nilai bobot perhitungan nilai *similarity* yang sering muncul pada artikel asli dan nilai kata yang penting pada *textrank*, sehingga mendapatkan kalimat output dari ringkasan utama dengan memiliki jumlah rata-rata kata yang beragam namun maksimal kalimat di rangkum hanya 3 kalimat.

Pada tabel 4.1 dapat dilihat ketika menggunakan algoritma textrank pada peringkasan teks otomatis terhadap artikel berita menghasilkan ringkasan dengan rata-rata ringkasan 6,9%, hal ini menjadi lebih ringkas daripada teks artikel berita yang asli. Pada pengujian 25 teks artikel berita yang sudah melalui proses cleaning. Program textrank yang di buat penulis menempatkan batasan maksimal sebesar 3 kalimat untuk hasil output. Perhitungan yang di ambil pada output merupakan hasil ringkasan dari nilai bobot perhitungan nilai *similarity* yang sering muncul pada artikel asli dan nilai kata yang penting pada *textrank*, sehingga mendapatkan kalimat output dari ringkasan utama dengan memiliki jumlah rata-rata kata yang beragam namun maksimal kalimat di rangkum hanya 3 kalimat.

V. KESIMPULAN

Langkah langkah peringkasan teks otomatis menggunakan algoritma textrank adalah memilih teks artikel dari portal berita CNN Indonesia, kemudian melakukan preprosesing (sentence spilling, cleaning, case folding, tokenizing) berdasarkan algoritma textrank, hasil akhir yang diperoleh dari proses peringkasan teks otomatis menggunakan algoritma textrank yaitu penjumlahan score tiap kalimat sehingga diperoleh nilai terbesar yang dipilih sebagai kalimat penting sebagai hasil peringkasan teks, Hasil pengujian dari 25 artikel berita dengan bahasa Indonesia dari portal berita CNN Indonesia menggunakan algoritma textrank mendapatkan hasil rata-rata presentase ringkasan sebesar 6,9% lebih ringkas daripada artikel asli, dengan catatan penulis membatasi jumlah kalimat sebesar 3 kalimat saja. Hasil kalimat ringkasan dapat diubah jumlahnya dengan menyesuaikan kebutuhan penelitian maupun ringkasan lainnya dan dari hasil analisis Algoritma textrank efektif dalam meringkas text berita dalam bahasa Indonesia.

REFERENSI

- [1] Media Sucahya. Drs. (2008). Jurnal Komunikasi, Volume 2, Teknologi Komunikasi dan Media.
- [2] Muhammad Adib Zamzam, Cahyo Crysdian, Khadijah Fahmi Hayati Holle,2020. *SISTEM AUTOMATIC TEXT SUMMARIZATION MENGGUNAKAN ALGORITMA TEXTRANK*

- [3] Malaya Dutta Borah, Deepali Jain dan Anupam Biswas, 2020. Fine-Tuning Taxtrank for Legal Document Summarization : A Bayesian
- [4] Mingxi Zhang, Xuemin Li, Shuibo Yue, Liuqian Yang,2020. An Empirical Study of TextRank For Keyword Extraction
- [5] K. G. Subhawa, I. Atastina, A. Nursikuwagus, 2019. Penerapan Algoritma Text Mining dan Lexrank dalam Meringkas Teks Secara Otomatis
- [6] Nurdiansyah, Asep. 2019. Peringkat Teks Dokumen Menggunakan Metode Simple Additive Weighting (SAW). Universitas Komputer Indonesia. Bandung
- [7] Aminullah, Muhammad. 2021. Perbandingan Performa Kalsifikasi Manchine Learning Dengan Teknik Resampling pada Dataset tidak Seimbang. Universitas Islam Negeri Syarif Hidayatullah. Jakarta
- [8] Pratama, I. S., Alam, G., & Tinaliah, T. Penerapan Algoritma Centroid-Based Summarization untuk Sistem Peringkasan Dokumen Berbahasa Indonesia.