

PERINGKASAN TEKS OTOMATIS ARTIKEL

BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA *TEXTRANK*

TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

FAIZAH HUSNIAH

NIM. 11850120304



UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU

2023

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, dan pengkajian, dan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSETUJUAN

**PERINGKASAN TEKS OTOMATIS ARTIKEL
BERBAHASA INDONESIA MENGGUNAKAN
ALGORITMA *TEXTRANK***

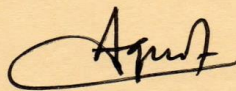
TUGAS AKHIR

Oleh

FAIZAH HUSNIAH
NIM. 11850120304

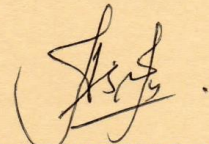
Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 22 Desember 2022

Pembimbing I,



SURYA AGUSTIAN, S.T., M.KOM.
NIP. 19760830 201101 1 003

Pembimbing II,



IIS AFRIANTY S.T., M.SC.
NIP. 19880426 201903 2 009

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN

**PERINGKASAN TEKS OTOMATIS ARTIKEL
BERBAHASA INDONESIA MENGGUNAKAN
ALGORITMA *TEXTRANK***

Oleh

FAIZAH HUSNIAH

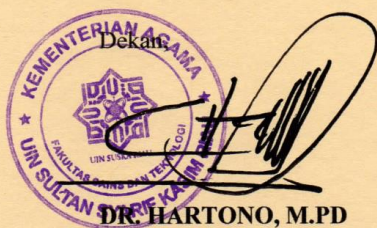
NIM. 11850120304

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 22 Desember 2022

Mengesahkan,

Ketua Jurusan,



DR. HARTONO, M.PD

NIP. 19640301 199203 1 003

IWAN ISKANDAR, M.T.

NIP. 19821216 201503 1 003

DEWAN PENGUJI

Ketua : Reski Mai Candra, S.T., M.Sc.

Pembimbing I : Surya Agustian, S.T, M.Kom

Pembimbing II : Iis Afrianty, ST, M.Sc.

Penguji I : Muhammad Fikry, ST, M.Sc

Penguji II : Febi Yanto, M.Kom

LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

© Hak Cipta dan Hak Milik UIN Suska Riau

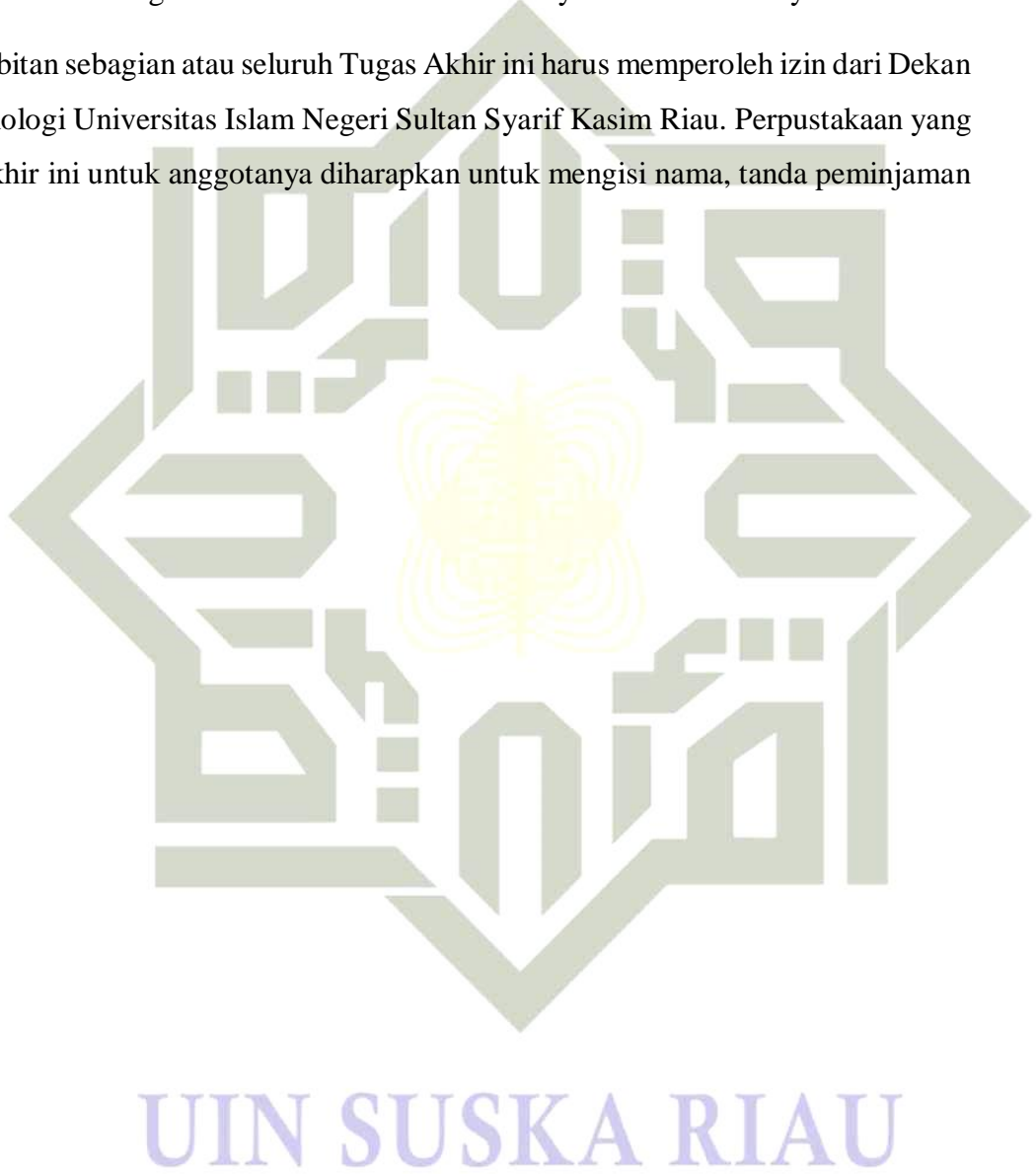
State Islamic University of Sultan Syarif Kasim Riau

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. Perpustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan dengan izin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Penggunaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi nama, tanda peminjaman dan tanggal pinjam.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



LEMBAR PERNYATAAN

Peran ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan didalam daftar pustaka.

Pekanbaru, 10 Januari 2023

Yang membuat pernyataan,

FAIZAH HUSNIAH

11850120304



UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Siapa yang menempuh jalan untuk mencari ilmu, maka Allah akan mudahkan baginya jalan menuju surga."

Alhamdulillah Rabbil'alamin. Puji syukur kita panjatkan kepada Allah Swt. atas segala limpahan rahmat dan karunia-Nya, sehingga setelah berlika-liku perjuangan, Tugas Akhir ini dapat diselesaikan. Persehatian dan selamat beriring salam kita sampaikan untuk nabi besar Muhammad Saw.

Pertama tama saya mengucapkan terimakasih kepada ayah dan ibu saya yang telah bekerja keras memberikan *support* , materi dan dukungan agar dapat menyelesaikan Tugas Akhir ini.

Kemudian saya mengucapkan terimakasih kepada kakak dan adik saya yang senantiasa memberikan *support* dan mendengarkan keluh kesah saya dalam berjuang untuk menyelesaikan Tugas Akhir ini

Kemudian saya mengucapkan terimakasih kepada pembimbing saya Pak Surya Agustian, S.T, M.Kom., dan Ibu Iis Afrianty, ST, M.Sc, yang telah membimbing dengan saya dengan sangat baik dan memberikan dukungan dukungan lainnya sehingga saya dapat bertahan hingga tahap ini

Tidak lupa juga saya mengucapkan terima kasih kepada teman teman yang memberikan dukungan dan *support*, terimakasih kepada Salsa yang bersedia menyediakan kosnya untuk dijadikan basecamp, Akkw, Zhastrri yang bersedia menginap dirumah saya sebelum sidang terimakasih juga kepada Alimah, Ma, Mori, Cai, Faras, Bibah, Kak Tri, Windy, Gunawan dll.

Semoga tugas akhir ini bermanfaat bagi pembacanya. Aamiin yaa rabbal 'aalamiin

UIN SUSKA RIAU

Lampiran Surat :
Nomor : Nomor 25/2021
Tanggal : 10 September 2021

SURAT PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : Faizah Husniah
NIM : 11850120304
Tempat/Tgl. Lahir : Pekanbaru/04 Desember 1999
Fakultas/Pascasarjana : Sains dan Teknologi
Prodi : Teknik Informatika
Judul Skripsi :

PERINGKASAN TEKS OTOMATIS ARTIKEL BERBAHASA INDONESIA
MENGUNAKAN ALGORITMA *TEXTRANK*

1. Penulisan Skripsi dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu Skripsi saya ini, saya nyatakan bebas dari plagiat.
4. Apa bila dikemudian hari terbukti terdapat plagiat dalam penulisan Skripsi saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Pekanbaru, 10 Januari 2023
Yang membuat pernyataan



Faizah Husniah
NIM : 11850120304

**pilih salah satu sesuai jenis karya tulis*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Peringkasan Teks Otomatis Artikel Berbahasa Indonesia Menggunakan Algoritma *Textrank*

Faizah Husniah¹⁾, Surya Agustian²⁾, & Iis Afrianty³⁾

^{1,2,3} Teknik Informatika, Fakultas Sains & Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Jl.H.R. Soebrantas No. 155 KM.1,5 Simpang Baru Panam Pekanbaru, 0761-858832, Website: www.uin-suska.ac.id, E-mail:

¹11180120304@students.uin-suska.ac.id, ²surya.agustian@uin-suska.ac.id, ³iis.afrianty@uin-suska.ac.id

Banyaknya informasi yang terdapat pada media online salah satunya adalah artikel. Artikel pada umumnya berisi kumpulan kalimat yang panjang sehingga pembaca membutuhkan waktu yang lebih banyak untuk memperoleh intisari dari artikel. Sistem peringkasan teks otomatis membantu pengguna mengetahui isi artikel tanpa harus membaca keseluruhan dokumen, sehingga dapat menghemat waktu. Penelitian ini menyelidiki algoritma *textrank* untuk peringkasan dokumen. Algoritma *textrank* akan menghitung skor dari setiap kalimat, kemudian memilih sejumlah kalimat sebagai ringkasan. Untuk meningkatkan performa *textrank* dengan menguji pemilihan kelompok skor kalimat yang akan dipilih menjadi ringkasan. *ROUGE* scoring digunakan untuk menghitung performa sistem berdasarkan ringkasan manual yang dibuat oleh manusia (*human annotator*) sebagai *gold standard*. Hasil pengujian dengan *compression rate* 50% pada pilihan ranking kalimat yang terpilih sebagai ringkasan, menghasilkan nilai pada *ROUGE-1*, *ROUGE-2* dan *ROUGE-L* secara berturut-turut adalah 68.76%, 60.60%, dan 68.29%. Sedangkan untuk *compression rate* 30%, rata-rata pada *ROUGE-1*, *ROUGE-2* dan *ROUGE-L* berturut-turut adalah 45.00%, 31.62%, dan 43.59%. Hasil ini sangat kompetitif dibandingkan dengan score pada penelitian tentang sistem peringkasan dokumen lainnya.

Keyword: peringkasan teks otomatis; *compression rate*; *rouge*; skor kalimat; *textrank*

Abstract

The amount of information contained in online media, one of them is articles. Articles generally contain a long collection of sentences so that readers need more time to get the point of the article. The automatic text summarization system helps users to find out the contents of the article without having to read the entire document, thus saving time. This research to inspect the *textrank* algorithm for document summarization. The *textrank* algorithm will calculate the score of each sentence, then choose the sentence number as a summary. To improve the performance of the *textrank* by testing the score of the sentence that will be selected as a group summary. The *ROUGE* scoring is used to calculate system performance based on a human annotator as the gold standard. The results of the test with a *compression level* of 50% on the choice of sentence ratings that were selected as a summary, resulted in values on *ROUGE-1*, *ROUGE-2* and *ROUGE-L* sequentially 68.76%, 60.60%, and 68.29%. As for the *compression rate* 30%, the average s at *ROUGE-1*, *ROUGE-2* and *ROUGE-L* are 45.00%, 31.62%, and 43.59%, sequentially. These results are very competitive compared with the scores on the other about research about document summary systems.

Kata kunci: automatic text summarization; *compression rate*; *rouge*; sentences score; *textrank*

PENDAHULUAN

Pada saat ini, informasi telah menjadi kebutuhan penting bagi manusia. Kebutuhan akan informasi dapat diperoleh lewat berbagai media, salah satunya media digital [1]. Media digital merupakan media elektronik untuk menyimpan dan menerima informasi dalam bentuk digital. Internet sebagai media digital merupakan solusi untuk mendapatkan informasi [2]. Informasi pada internet salah satunya adalah melalui artikel. Dengan banyaknya informasi yang terdapat pada artikel, maka pembaca terkadang kesulitan untuk mengetahui intisari dari artikel tersebut, selain itu

pembaca harus meluangkan banyak waktu untuk membaca artikel tersebut, sehingga dibutuhkan sebuah sistem untuk meringkas dokumen secara otomatis [3].

Peringkasan teks otomatis adalah bidang dari *information extraction*, yang meringkas dokumen teks berbentuk digital, namun tetap mempertahankan isi informasinya. *Information extraction* yaitu mengekstraksi informasi terstruktur secara otomatis dari dokumen tidak terstruktur atau semi terstruktur. Terdapat dua klasifikasi metode *text summarization* yaitu peringkasan dengan cara ekstraktif dan peringkasan dengan cara abstraktif [1, 4].

Peringkasan ekstraktif dilakukan dengan meng-
 ekstraksi sebagian kalimat penting dari dokumen awal
 dengan mengganti struktur kalimat. Peringkasan
 abstraktif dilakukan dengan membuat dan menyusun
 kalimat baru, yang pada dasarnya merupakan kom-
 binasi informasi dari dokumen yang diringkas [3]. Pada
 umumnya, peringkasan teks digolongkan ke dalam
 metode *extractive summarization* bila berdasarkan
 metode ekstraksi kalimat-kalimat yang ada pada
 dokumen sumber [5]. Sedangkan metode peringkasan
 abstraktif bekerja dengan penyusunan kalimat yang
 baru, yang sesuai dengan aturan tata bahasa manusia.
 Metode abstraktif sangat sulit dilakukan secara
 otomatis, karena sistem harus mencocokkan tata bahasa
 yang tepat dengan model tata bahasa manusia [6].
 Peringkasan teks otomatis modul pembelajaran
 bahasa Indonesia menggunakan metode *Cross
 Latent Semantic Analysis (CLSA)* menghasilkan nilai
 rata-rata, *precision*, dan *recall* pada *compression rate*
 20% berturut-turut yaitu 0.3853, 0.432, dan 0.3715 dan
 termasuk rendah [5]. Algoritma *textrank* untuk
automatic text summarization pada dokumen berbahasa
 Indonesia, menyarankan agar melakukan proses
 penghilangan *stopword* agar kata-kata yang tidak
 mempunyai makna penting dalam tugas pengolahan
 teks, seperti kata hubung tidak dimasukkan dalam
 proses penghitungan *similarity* [1]. Berdasarkan
 literatur review oleh [4] menjelaskan bahwa sejauh ini
 peringkasan dokumen teks didominasi oleh teknik yang
 bersifat ekstraktif dan metode-metode *unsupervised*.
 Metode lainnya yang telah diusulkan seperti *machine
 learning*, teruji mampu menghasilkan ringkasan yang
 baik, tetapi masih memerlukan banyak data latihan.
 Metode berbasis *unsupervised* kebanyakan dikembang-
 kan untuk peringkasan teks yang bersifat ekstraktif,
 salah satunya yang terdapat pada algoritma *textrank*.
 Metode kerja algoritma *textrank* ialah dengan cara
 memberikan skor kepada tiap kalimat, kemudian
 memilih kalimat dengan skor terbaik sebagai ringkasan
 akhir [4].
 Keunggulan dari algoritma *textrank* yaitu tidak
 membutuhkan pengetahuan mendalam tentang tata
 bahasa dan tidak memerlukan data latihan untuk dapat
 meringkas teks [7]. Pemilihan kalimat untuk menjadi
 ringkasan didasarkan pada formulasi matematis
 tertentu saja. Penelitian ini menyelidiki teknik
 pemilihan kandidat kalimat sebagai penyusun
 ringkasan, dengan menggunakan algoritma *textrank*.
 Asumsinya *textrank* bekerja akan memilih sejumlah
 kalimat yang dianggap berisi informasi penting dari
 dokumen dan menyusunnya dengan cara yang mirip
 dengan bagaimana manusia memilih kalimat di dalam
 dokumen yang panjang [8].
 Penelitian yang dilakukan pada [9] menerapkan
 algoritma *textrank* untuk *multidocument summary-
 zation*, namun masih ditemukan kalimat serupa dalam
 hasil ringkasan.

Pada umumnya penelitian mengenai peringkasan teks
 otomatis mengambil ringkasan pada rangking teratas
 skor *textrank* saja, namun pada penelitian ini mencoba
 menyelidiki pengambilan hasil ringkasan, tidak hanya
 bagian awal saja (*ranking* teratas), namun ringkasan
 dibagi menjadi 3 bagian, yakni ringkasan pada *ranking*
 awal, ringkasan pada *ranking* tengah, dan ringkasan
 pada *ranking* akhir yangn diurutan berdasarkan skor
 yang didapat dari algoritma *textrank* untuk setiap
 kalimat.

Hasil ringkasan yang dibentuk dari ketiga cara
 pemilihan kalimat ringkasan tersebut diharapkan lebih
 maksimal dalam menyampaikan fakta yang lebih
 banyak dan kaya informasi. Pertimbangannya,
 ringkasan yang berisi kalimat-kalimat yang mirip atau
 ide pokok yang diulang-ulang tentu memboroskan
 waktu membacanya. Hasil ringkasan tersebut dieva-
 luasi menggunakan *ROUGE (Recall-Oriented Under-
 study for Gisting Evaluation)*. *ROUGE* dihitung dengan
 membandingkan hasil ringkasan oleh sistem dengan
 ringkasan manual oleh manusia (*gold standard*).

Penerapan algoritma *textrank* pada peringkasan teks
 otomatis diharapkan dapat menghasilkan ringkasan
 yang lebih akurat sesuai dengan yang diharapkan.
 Penelitian ini bertujuan untuk menerapkan algoritma
textrank dalam peringkasan artikel berbahasa Indo-
 nesia, menyelidiki cara pemilihan kandidat kalimat
 ringkasan berdasarkan posisi kalimat dari ranking
textrank, dan kemudian mengukur relevansi hasil
 peringkasan menggunakan *ROUGE score* pada 300
 dokumen berbahasa Indonesia.

2 LANDASAN TEORI

2.1 Peringkasan Teks Otomatis

Peringkasan teks adalah proses menghasilkan
 ringkasan dari suatu artikel atau dokumen yang sangat
 panjang oleh komputer. Peringkasan ekstraktif memilih
 beberapa kalimat dari dokumen asli untuk
 merepresentasikan dokumen secara keseluruhan tanpa
 perubahan dari struktur kalimat-kalimat tersebut.
 Sedangkan peringkasan abstraktif mengatur ulang
 kalimat-kalimat menjadi ringkasan berdasarkan kata-
 kata inti yang terdapat pada dokumen asli.

2.2 Text preprocessing

Text preprocessing merupakan langkah pertama
 dalam proses penambangan teks (*text mining*). Pada
 umumnya dokumen/artikel mempunyai bentuk yang
 tidak terstruktur. Oleh karena itu, diperlukan suatu
 proses yang dapat mengubah bentuk data yang
 sebelumnya tidak terstruktur ke dalam bentuk data
 yang terstruktur. Proses pengubahan ini dikenal dengan
 istilah *text preprocessing* [10]. Tahapan-tahapan dalam
text preprocessing yang dapat dilakukan antara lain
 adalah [11] sebagai berikut:

1. *Case Folding*: proses perubahan huruf seluruh teks menjadi huruf kecil (*lower case*).
 2. *Tokenization*: proses pemecahan kalimat menjadi token-token (kata-kata tunggal) dengan spasi sebagai delimiter.
 3. *Remove Punctuation*: proses penghapusan tanda baca, angka, dan simbol.
 4. *Stemming*: proses mengubah kata-kata yang terdapat dalam suatu dokumen menjadi kata dasar (*root*).
 5. *Stopword Removal*: penghapusan kata-kata umum yang sering muncul dan tidak memiliki arti seperti yang, di, ke, dari, dan dll.
- Graf adalah sebuah struktur diskrit yang terdiri atas titik-titik (*vertex*) dan suatu himpunan pasangan tak berurutan yang menghubungkan dari titik-titik tersebut yang disebut sisi (*edge*) [6]. Graf dapat merepresentasikan hubungan antar kalimat di dalam suatu dokumen, berdasarkan kemiripan antar dokumen-dokumen tersebut,

2.1. ROUGE

ROUGE sendiri merupakan akronim dari *Recall-Oriented Understudy for Gisting Evaluation* atau dalam bahasa Indonesia berarti evaluasi intisari yang berorientasi *recall*. *ROUGE* merupakan standar pengujian peringkasan teks berbahasa indonesa untuk mengukur kualitas hasil ringkasan dengan menghitung unit-unit token (*gram*) yang *overlap* seperti urutan kata dan pasangan-pasangan kata antara ringkasan manual yang dibuat oleh manusia sebagai referensi (*human annotator*) dan ringkasan yang *degenerate* oleh mesin.

Ada beberapa macam jenis *ROUGE* yaitu *ROUGE-N*, *ROUGE-L*, dan *ROUGE-SU* [4].

2.2. TextRank

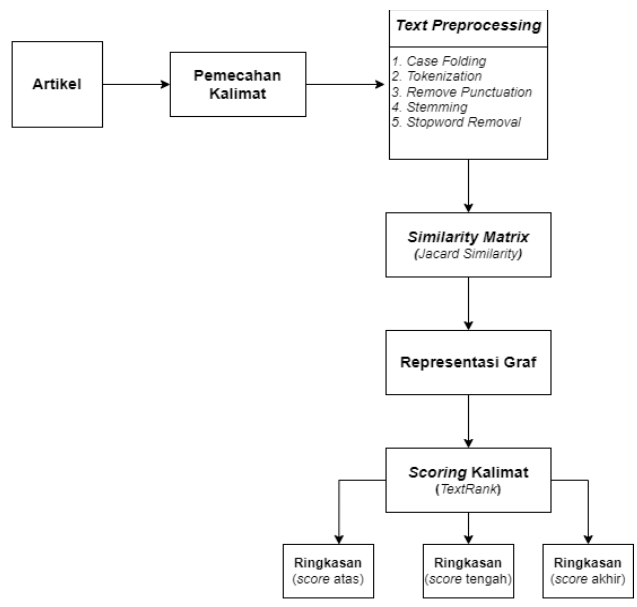
TextRank adalah salah satu algoritma yang berasal dari metode *pagerank*. *pagerank* adalah metode yang digunakan untuk memperkirakan seberapa penting sebuah node pada jaringan dan berbasis graf [12]. Algoritma perankingam berbasis graf adalah cara untuk menentukan seberapa penting sebuah *vertex* dalam graf, berdasarkan informasi yang diambil dari struktur graf keseluruhan secara rekursif. *textrank* mengekstraksi kalimat-kalimat beranking tinggi untuk dimasukkan ke dalam ringkasan.

3. METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Tahapan yang dilakukan dalam penelitian ini dimulai dengan pengumpulan dokumen, sebanyak 300

buah artikel dari berbagai topik melalui media online. Setiap artikel diproses dalam tahapan seperti pada Gambar 1.



Gambar 1. Tahapan Penelitian

3.1.2 Pemecahan Kalimat

Pemecahan kalimat membagi teks pada sebuah dokumen menjadi beberapa kalimat. Tanda pemisah (*delimiter*) yang dipakai pada proses *sentences splitting* adalah tanda titik (.), tanda tanya (?) dan tanda seru (!).

3.1.3 Text Preprocessing

Text preprocessing bertujuan untuk mengubah bentuk teks (kata atau kalimat) agar sesuai dengan kebutuhan *task*/tugas yang dikerjakan (problem yang akan dipecahkan). Tahapan-tahapan umum dalam *text preprocessing* meliputi *case folding*, *tokenization*, *remove punctuation*, *stemming*, dan *stopword removal*.

Tabel 1 di bawah ini menunjukkan contoh kalimat dari satu artikel di dalam korpus, yang telah melalui tahap *text preprocessing*, dengan judul artikel “*Anak Di Bawah Umur Banyak Terlibat Tindak Kriminal Ini Kata Kapolres Pekanbaru*” (sumber riaulink.com).

3.1.4 Pembentukan Similarity Matrix

Tahap pembentukan matriks kemiripan (*similarity matrix*) adalah metode untuk menghitung nilai kesamaan antar kalimat. Salah satu metode yang dapat digunakan untuk menghitung kemiripan kalimat adalah metode *jaccard similarity* [13].

Tabel 1. Hasil Preprocessing

Kalimat Awal	Hasil Preprocessing
<p>PEKANBARU, RIAULINK.COM - Sangat disayangkan mulai banyak kasus pencurian dengan kekerasan (Curas) maupun pencurian sepeda motor (Curanmor) yang didalamnya ikut terlibat anak-anak di bawah umur harus menjadi perhatian serius.</p> <p>Dimana diketahui dari beberapa kali gelar press release kasus curas, maupun curanmor yang telah dilakukan pihak Polresta Pekanbaru beserta jajaran Polsek Polresta Pekanbaru masih ada mendapati anak di bawah umur terlibat melakukan tindak pidana tersebut.</p> <p>Kapolresta Pekanbaru Kombes Pol Nandang Mu'min Wijaya saat wawancara mengungkapkan melihat masih adanya anak di bawah umur yang terlibat melakukan tindak pidana, mengatakan bahwa anak-anak tersebut melakukan tindak kejahatan karena salah pergaulan.</p> <p>"Kita harus imbau kepada para orangtua agar benar-benar bisa mengawasi anaknya di lingkungan pergaulan, karena salah pergaulan bisa membawa anak-anak menjadi pelaku kriminal, karna memang tantangan di kota besar ini banyak sekali, oleh karena itu perlu pengawasan orangtua, di lingkungan rumah dulu yang paling utama," ucap Kapolresta, saat wawancara riaulink.com, Kamis (16/1/2020).</p> <p>Termasuk juga di lingkungan pendidikan, benar-benar di ajarkan pelajaran-pelajaran terkait dengan keagamaan maupun moralitas itu yang paling penting, agar anak-anak kita ini nanti bisa menjadi generasi harapan bangsa," tambah Kapolresta Pekanbaru.</p> <p>Lebih lanjut Kapolresta menyebutkan untuk meminimalisir angka kejahatan yang dilakukan anak di bawah umur, pihak Polresta sudah melakukan upaya sosialisasi dengan cara program pembinaan.</p> <p>"Kita (Polresta Pekanbaru, red) sudah banyak program-program pembinaan, pembinaan ketertiban masyarakat yang dilaksanakan oleh jajaran satuan binmas maupun polsek dan para bhabin-kamtibmas untuk melakukan kegiatan pembinaan, penyuluhan dan sosialisai terkait untuk mencegah kejahatan baik kejahatan jalan, ataupun konvensional, narkoba maupun kekerasan didalam rumah tangga dan lainnya," jelasnya.</p> <p>Kegiatan yang paling penting itu adalah kegiatan yang bersifat bekerjasama dengan dunia pendidikan, disekolah-sekolah ya kita bekerjasama untuk melakukan kegiatan penyuluhan kesekolah-sekolah untuk menghimbau dan mengingatkan kepada pelajar untuk menghindari narkoba kemudian menghindari tindakan-tindakan kriminalitas.</p> <p>Saat disinggung soal motif yang tidak asing didengar lagi dari para pelaku kriminal melakukan kejahatan hanya karna untuk bermain game online, Kapolresta Pekanbaru tegas mengatakan bahwa pihaknya saat ini sering melaksanakan razia rutin yang ditingkatkan dilokasi warnet.</p> <p>"Saat ini kita akan sering melaksanakan kegiatan rutin yang ditingkatkan baik jajaran jajaran satuan fungsi pre-emptif preventif baik satuan binmas, satuan sabhara termasuk juga jajaran polsek bersama-sama melakukan kegiatan rutin yang ditingkatkan, dilokasi-lokasi warnet yang jam operasionalnya melebihi batas, itu kita himbau dan ingatkan pada pemilik warnet untuk mematuhi aturan yang ada," katanya.</p> <p>11. "Jangan sampai warnet-warnet ini disalahgunakan yang seharusnya izinya untuk browsing, tapi disalahgunakan menjadi kegiatan permainan game, dan biasanya memang terindikasi lewat tengah malam, padahal sebenarnya izin warnet ditentukan tidak sampai lewat larut malam ataupun sampai pagi, dan kegiatan rutin yang ditingkatkan ini sering kita lakukan baik dari jajaran Polresta maupun jajaran Polsek," tutup Kapolresta</p>	<ol style="list-style-type: none"> pekanbaru riaulinkcom sayang banyak curi keras curas curi sepeda motor curanmor dalam libat anak-anak umur perhati serius mana kali gelar press release curas curanmor polresta pekanbaru serta jajaran polsek polresta pekanbaru dapat anak umur libat tindak pidana kapolresta pekanbaru kombes pol nandang mumin wijaya wawancara tanggap anak bawah umur libat tindak pidana anak-anak tindak jahat salah gaul imbau orangtua awas anak lingkungan gaul salah gaul bawa anak-anak laku kriminal karna tantang kota awas orangtua lingkungan rumah utama kapolresta wawancara riaulinkcom Kamis 16/1/2020 lingkung didik ajar pelajaran-pelajaran kait agama moralitas anak-anak generasi harap bangsa kapolresta pekanbaru kapolresta meminimalisir angka jahat anak bawah umur polresta upaya sosialisasi program bina polresta pekanbaru red program-program bina bina tertib masyarakat laksana jajar satu binmas polsek bhabin-kamtibmas giat bina suluh sosialisai kait cegah jahat-jahat jalan konvensional narkoba keras dalam rumah tangga giat-giat sifat bekerjasama dunia didik disekolah-sekolah ya kerjasama giat suluh kesekolah-sekolah menghimbau ajar hindar narkoba hindar tindak-tindak kriminalitas singgung motif asing dengar laku kriminal tindak jahat karna main judi online kapolresta pekanbaru laksana razia rutin tingkat lokasi warnet laksana giat rutin tingkat jajar-jajar satu fungsi pre-emptif preventif satu binmas satu sabhara jajaran polsek bersamasama giat rutin tingkat dilokasilokasi warnet jam operasional lebih batas himbau ingat milik warnet patuh atur warnet-warnet disalahgunakan izin browsing disalahgunakan giat main game indikasi malam izin warnet tentu larut malam pagi giat rutin tingkat laku jajar polresta jajaran polsek tutup kapolresta

Metode *Jaccard similarity* menghitung kesamaan antar 2 kalimat berdasarkan jumlah kata unik yang sama berbanding dengan total kata di kedua kalimat. Metode ini diambil dengan pertimbangan bahwa dalam satu artikel, hanya sedikit jumlah kata unik yang bisa

dihasilkan, sehingga perhitungan menggunakan *Jaccard similarity* dapat dilakukan dengan mudah dan efisien, dengan formula pada persamaan (1) berikut.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

Dimana $|A \cap B|$ adalah jumlah kata unik yang sama antara kalimat A dan B; $|A \cup B|$ adalah jumlah seluruh kata pada kalimat A dan B, dan $J(A,B)$ adalah nilai *jaccard similarity* yang akan dihitung pada kalimat A dan B.

Sebagai contoh, penerapan *jaccard similarity* untuk kalimat 2 dan kalimat 3 hasil *preprocessing* Tabel 1, dapat dilihat dengan cara sebagai berikut.

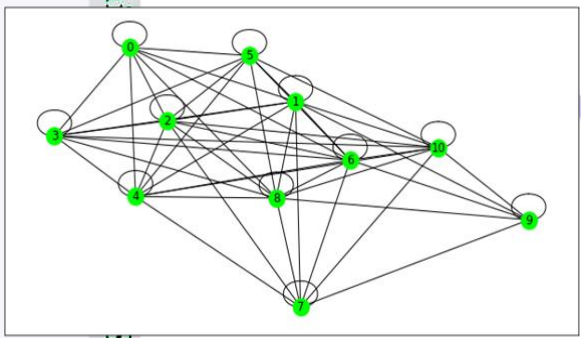
Kalimat 2:

Tabel 2. Perhitungan Jaccard Similarity

S _i	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	S ₈	S ₉	S ₁₀	S ₁₁
S ₁	1	0.214	0.133	0.058	0.074	0.076	0.076	0	0.028	0	0
S ₂	0.214	1	0.2	0.027	0.068	0.111	0.1	0.03	0.055	0.048	0.083
S ₃	0.133	0.2	1	0.151	0.107	0.2	0.047	0.03	0.117	0	0.026
S ₄	0.058	0.027	0.151	1	0.1	0.066	0.022	0	0.111	0	0.03
S ₅	0.074	0.068	0.107	0.1	1	0.086	0.054	0.074	0.064	0	0.03
S ₆	0.076	0.111	0.2	0.066	0.086	1	0.117	0	0.066	0	0.064
S ₇	0.076	0.1	0.047	0.022	0.054	0.117	1	0.05	0.069	0.133	0.093
S ₈	0	0.03	0.03	0	0.074	0	0.05	1	0.028	0.025	0.027
S ₉	0.028	0.055	0.117	0.111	0.064	0.066	0.069	0.028	1	0.09	0.171
S ₁₀	0	0.048	0	0	0	0	0.133	0.025	0.097	1	0.15
S ₁₁	0	0.083	0.026	0.051	0.03	0.064	0.093	0.027	0.171	0.15	1

3.1.5 Representasi Graf

Pada tahap pembentukan graf, *vertex* merepresentasikan kalimat-kalimat yang berada dalam dokumen sedangkan *edge* adalah representasi dari *similarity* antara dua kalimat yang saling berelasi. *Similarity* menyatakan hubungan antar *vertex* yang dihitung berdasarkan persamaan (1). Pada gambar 2. adalah contoh representasi graf dari dokumen yang terdiri dari 11 kalimat (indeks kalimat dimulai dari nol).



Gambar 2. Representasi Graf berdasarkan similarity pada Tabel 2

“mana kali gelar press release curas curanmor polresta **pekanbaru** serta jajar polsek polresta **pekanbaru** dapat **anak umur libat tindak pidana**”

Kalimat 3:

“kapolresta **pekanbaru** kombes pol nandang mumin wijaya wawancara tanggap **anak** bawah **umur libat tindak pidana anak anak tindak jahat salah gaul**”

Bila kata yang tercetak tebal adalah kata yang sama (beririsan), ada di kedua kalimat, maka hasil perhitungan kemiripan antara kalimat 2 dan 3 menggunakan *jaccard similarity* adalah:

$$Sim(S_2, S_3) = 6/30 = 0.2$$

3.1.6 Scoring Kalimat

Tahap *Scoring Kalimat* bertujuan mendapatkan skor akhir dari suatu *vertex*. Semakin besar skor akhir dari suatu *vertex*, maka semakin banyak informasi di dalamnya. Pada algoritma *textrank*, proses perankingan dilakukan dengan menggunakan rumus untuk graf tidak berarah dan berbobot [1]. Kalimat yang sudah memperoleh skor, kemudian diberi peringkat berdasarkan skor tertinggi ke skor yang terendah. Berikut merupakan rumus tahap perankingan kalimat menggunakan algoritma *textrank*

$$WS(S_i) = (1 - d) + d * \sum_{V_k \in Adj(V_j)} \frac{W_{ji}}{\sum_{V_k \in Adj(V_j)} W_{jk}} WS(S_j) \quad (2)$$

Dimana *WS* adalah *weight sentence* (skor kalimat); *d* adalah *dumping factor* yang diisi dengan 0.85; *V_i* adalah *vertex* yang akan dihitung skornya; *V_j* adalah *vertex* yang bertetangga dengan *V_i*; *V_k* adalah *vertex* yang bertetangga dengan *V_j*; *W_{ji}* adalah nilai *jaccard similarity* kalimat yang akan dihitung; *W_{jk}* adalah kalimat yang terhubung dengan *W_j*, *S_j* (bernilai = 1).

3.1.7 Penentuan Ringkasan

Pembentukan ringkasan adalah proses penggabungan dari kalimat-kalimat yang terpilih sebagai ringkasan berdasarkan skor pada persamaan (2). Jumlah kalimat yang dipilih sesuai dengan *compression rate* yang diinginkan, yaitu seberapa padat ringkasan yang akan dibandingkan dokumen aslinya. Dalam penelitian ini digunakan *compression rate* 30% dan 50%. Sebagai kontribusi penelitian, peneliti mencoba untuk melihat apakah hasil ringkasan akan lebih baik bila menggunakan kelompok kalimat dengan skor rendah, atau skor menengah, selain metode asli yang hanya mengambil *top-n* kalimat dengan score tinggi.

Dalam kalimat-kalimat kandidat ringkasan, dikelompokkan menjadi tiga bagian, yaitu bagian awal, bagian tengah dan bagian akhir. Kalimat-kalimat yang memperoleh ranking tertinggi pada proses pemeringkatan dikategorikan sebagai ringkasan awal (*top-n*/skor tertinggi), kalimat-kalimat yang mendapat ranking pertengahan sebagai ringkasan tengah (*middle-n*/skor menengah), dan kalimat-kalimat yang mendapat ranking terendah sebagai ringkasan akhir (*bottom-n*/skor terendah). Pertimbangannya adalah, kalimat yang skornya tinggi akan memiliki kesamaan yang tinggi, sehingga kurang bermanfaat kalau dijadikan ringkasan karena informasinya akan relative sama. Sebaliknya, kalimat-kalimat dengan skor rendah, berarti kurang mirip bahkan berbeda, mungkin akan berguna bila terpilih sebagai ringkasan karena membawa informasi baru.

Peringkat merupakan contoh pengelompokan skor kalimat algoritma *textrank* dengan *compression rate* 50% dari 11 kalimat dokumen awal, dipilih 5 kalimat sebagai ringkasan.

Ranking ke-	Skor Kalimat	Nomor Kalimat
1	0.10258990411959588	3
2	0.0997119800434317	2
3	0.0954155363618634	9
4	0.09388190433304869	7
5	0.0931112015022097	6
6	0.0912895197435418	11
7	0.08946987602886734	5
8	0.08801099005183073	1
9	0.08576150784919453	4
10	0.08259628217578834	10
11	0.07816129779062853	8

Gambar 3. Pengelompokan Skor Algoritma *Textrank*

Kemudian kalimat-kalimat dalam masing-masing kelompok diurutkan berdasarkan ID kalimat, agar tampil berkesinambungan antara kalimat yang satu dengan kalimat yang lainnya di dalam ringkasan. Dengan

demikian, ringkasan awal terdiri dari kalimat dengan ID {2,3,6,7,9}. Sedangkan ringkasan tengah terdiri dari kalimat {1,5,6,7,11}, dan ringkasan akhir terdiri dari kalimat {1,4,5,8,11}.

3.2 Evaluasi

Tahap evaluasi membandingkan hasil ringkasan oleh manusia (*gold-standard*) dengan hasil ringkasan oleh sistem, menggunakan metrik evaluasi *ROUGE*. *ROUGE* mengukur kualitas hasil ringkasan dengan menghitung jumlah kata yang beririsan antara ringkasan sistem dan ringkasan hasil yang menjadi referensi [3]. Pada penelitian ini mengambil nilai *precision*, *recall*, dan pada *ROUGE-1*, *ROUGE-2* dan *ROUGE-L*, sebagaimana persamaan (3-9).

$$ROUGE - 1 recall = \frac{\text{jumlah unigram kata yang sama}}{\text{Total kata di ringkasan manual}} \quad (3)$$

$$ROUGE - 1 precision = \frac{\text{jumlah unigram kata yang sama}}{\text{Total kata di ringkasan sistem}} \quad (4)$$

$$ROUGE - 2 recall = \frac{\text{jumlah bigram kata yang sama}}{\text{Total kata di ringkasan manual}} \quad (5)$$

$$ROUGE - 2 precision = \frac{\text{jumlah bigram kata yang sama}}{\text{Total kata di ringkasan sistem}} \quad (6)$$

$$ROUGE - L recall = \frac{LCS(\text{sistem}, \text{manual})}{\text{Total kata di ringkasan manual}} \quad (7)$$

$$ROUGE - L precision = \frac{LCS(\text{sistem}, \text{manual})}{\text{Total kata di ringkasan sistem}} \quad (8)$$

$$f1 scores = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (9)$$

4 HASIL DAN PEMBAHASAN

4.1 Pemilihan Model *Textrank*

Dari dataset yang ada, dibagi menjadi 150 artikel sebagai *data training* untuk memilih model *textrank* yang terbaik di antara ketiga kelompok ringkasan (awal, tengah atau akhir). Model dipilih berdasarkan nilai *F1-score* tertinggi untuk *ROUGE-L* dari ringkasan 150 dokumen dibandingkan dengan *gold-standard*.

Model dipilih berdasarkan jumlah terbanyak dari kelompok ringkasan yang memperoleh nilai (persamaan 9) tertinggi. Kemudian model yang terpilih, diterapkan untuk *data testing*, yaitu sebanyak 150 artikel lainnya dalam dataset yang belum digunakan.

4.2 Model *Textrank* untuk CR 50%

Tabel 3 dibawah ini adalah cuplikan hasil ringkasan dengan *compression rate* 50% menggunakan 150 artikel, dengan data nilai *precision*, *recall*, dan dari *ROUGE-1*, *ROUGE-2*, dan *ROUGE-L*. Nilai maksimum dari tiap dokumen ditandai. Kemudian dilakukan pemilihan kelompok skor *textrank* berdasarkan nilai maksimal dari *ROUGE-L*.

Tabel 3. Data Evaluasi Ringkasan Model (CR=50%) pada data training

Judul Artikel	Kelompok Skor Textrank	ROUGE-1 (%)			ROUGE-2 (%)			ROUGE-L (%)			S max pada-
		R	P	F	R	P	F	R	P	F	
Situs Berita Online	Awal	53,66	64,23	58,47	37,66	52,94	44,01	53,66	64,23	58,47	Akhir
	Tengah	53,05	56,86	54,89	40,17	44,44	42,20	52,44	56,21	54,26	
	Akhir	70,12	66,09	68,05	60,7	57,8	59,2	70,1	66,1	68,0	
Jurnal Kesehatan Masyarakat	Awal	58,33	72,06	64,47	47,17	64,10	54,35	58,33	72,06	64,47	Tengah
	Tengah	63,10	76,81	69,28	53,77	67,86	60,00	61,90	75,36	67,97	
	Akhir	61,90	66,67	64,20	52,83	59,67	56,00	60,71	65,38	62,96	
		:	:	:	:	:	:	:	:	:	
Jurnal Reser Kriminologi	Awal	30,13	44,34	35,88	20,20	34,19	25,40	28,21	42,51	33,59	Akhir
	Tengah	62,82	68,06	65,33	55,56	63,22	59,14	60,90	65,97	63,33	
	Akhir	84,62	85,16	84,89	83,33	83,76	83,54	84,62	85,16	84,89	
Seorang Wn Ins Tersandung	Awal	52,90	71,30	60,74	40,18	61,54	48,62	52,26	70,43	60,00	Akhir
	Tengah	65,81	76,12	70,59	59,36	70,65	64,52	65,16	75,37	69,90	
	Akhir	69,68	74,48	72,00	61,64	67,84	64,59	69,68	74,48	72,00	

Setelah itu dihitung total perolehan kelompok skor terhadap 150 dokumen tersebut. Kelompok yang terbanyak dipilih sebagai model metode *textrank* yang akan digunakan untuk pengujian terhadap *data testing*.

dipilih sebagai model untuk sistem *deployment* pada evaluasi ringkasan data testing.

4.3 Evaluasi Ringkasan pada Data Testing (50%)

Tabel 5 berikut merupakan data hasil evaluasi ringkasan uji dengan *compression rate* 50% menggunakan 150 artikel. Terlihat nilai untuk masing-masing *ROUGE-1*, *ROUGE-2* dan *ROUGE-L* mencapai di atas 60%, dengan rata-rata berturut-turut adalah 68.76%, 60.60% dan 68.29%.

4.4 Model *Textrank* untuk CR 30%

Tabel 6 berikut merupakan data pemilihan model dengan *CR=30%* menggunakan 150 artikel *data training*. Setelah itu dilakukan pemilihan skor *textrank* yang maksimal berdasarkan nilai dari *ROUGE-L*, berdasarkan posisi kelompok skor awal (*top-n*), menengah (*middle-n*) atau akhir (*bottom-n*).

Tabel 5. Hasil Evaluasi Ringkasan Data Testing (CR=50%)

No	Judul Artikel	ROUGE-1 (%)			ROUGE-2 (%)			ROUGE-L (%)		
		R	P	F	R	P	F	R	P	F
1	(HUT) Harian Pekanbaru MX	81,67	72,27	76,88	74,72	63,50	68,65	81,33	71,98	76,37
2	5 Kondisi Kesehatan Gigi dan Mulut yang Menantang Pikiran Sedang Stres	75,00	65,69	70,04	67,95	58,89	63,10	74,17	64,96	69,26
		:	:	:	:	:	:	:	:	:
149	Yuk Kenali Produk Vegan	64,41	67,86	66,09	62,50	60,81	61,64	62,71	66,07	64,35

150	Yuk Mengenal Masjid Quba	80,47	60,59	69,13	72,35	51,46	60,15	80,47	60,59	69,13
	RATA-RATA	72,32	66,15	68,76	65,08	57,29	60,60	71,81	65,72	68,29

Tabel 6. Data Evaluasi Ringkasan Model (CR=30%) pada data training

Judul Artikel	Kelompok Skor Textrank	ROUGE-1 (%)			ROUGE-2 (%)			ROUGE-L (%)			S max pada-
		R	P	F	R	P	F	R	P	F	
7 Situs Berita Online...	Awal	29,41	40,70	34,15	14,74	24,21	18,33	28,57	39,53	33,17	Akhir
	Tengah	34,45	38,68	36,44	19,23	21,28	20,20	32,77	36,79	34,67	
	Akhir	31,93	41,30	36,02	18,59	23,97	20,94	31,09	40,22	35,07	
Agen Bahasa Rusia...	Awal	14,00	17,50	15,56	1,00	1,00	1,00	12,00	15,00	13,33	Tengah
	Tengah	58,00	54,72	56,31	42,28	44,44	46,28	58,00	54,72	56,31	
	Akhir	36,00	34,62	35,29	18,97	20,00	19,47	34,00	32,69	33,33	
∴ ∴		∴ ∴			∴ ∴						
Satuan Reserse Kriminal ...	Awal	8,51	16,67	11,27	5,26	11,32	7,19	8,51	16,67	11,27	Tengah
	Tengah	73,40	69,00	71,13	67,54	68,14	67,84	73,40	69,00	71,13	
	Akhir	44,68	42,42	43,52	31,58	32,14	31,86	43,62	41,41	42,49	
Seorang Wn Inggris Tersandung ...	Awal	18,10	27,94	21,97	4,48	8,22	5,80	17,14	26,47	20,81	Tengah
	Tengah	64,76	72,34	68,34	53,73	62,07	57,60	61,90	69,15	65,33	
	Akhir	43,14	43,14	43,14	26,79	28,30	27,52	41,18	41,18	41,18	

Tabel 7. Statistik perolehan kelompok skor CR=30%

Posisi Kelompok Skor dgn tertinggi	Jumlah dokumen
Awal	38
Tengah	65
Akhir	47

Tabel 7. diatas merupakan total hasil dari penjumlahan posisi kelompok skor *textrank* dengan tertinggi. Model dengan CR=30%, yaitu kelompok skor tengah sebanyak 65 artikel, digunakan sebagai evaluasi ringkasan pada data *testing* untuk pengujian akhir sebagai model *deployment*.

Tabel 8. Hasil Evaluasi Ringkasan Data Testing (CR=30%)

No	Judul Artikel	ROUGE-1 (%)			ROUGE-2 (%)			ROUGE-L (%)		
		R	P	F	R	P	F	R	P	F
1	(HUMAS) Harian Pekanbaru MX	34,44	56,36	42,76	24,04	39,68	29,94	34,44	56,36	42,76

4.5 Evaluasi Ringkasan pada Data Testing (30%)

Tabel 8 berikut adalah data hasil eksperimen menggunakan model terpilih CR=30%, menggunakan model skor tengah. Secara umum nilai rata-rata untuk *ROUGE-1*, *ROUGE-2* dan *ROUGE-L* menurun dibandingkan dengan CR=50%. Hal ini karena jumlah kalimat yang dibangkitkan sebagai hasil ringkasan lebih sedikit. Dari hasil yang diperoleh, metode ini sudah cukup baik karena dapat menghasilkan angka h seperti rata-rata penelitian serupa, yaitu di kisaran angka 30-50% untuk CR yang rendah (misalnya 30%). Dan nilai pada *ROUGE-1* dan *ROUGE-L* dapat melebihi 40%.

8	5 Kondisi Kesehatan Gigi dan Mulut yang Menandakan Pikiran Sedang Stres	53,42	51,32	52,35	42,68	44,32	43,58	52,05	50,00	51,01
9	Yuk Kenali Produk Vegan	53,42	51,32	52,35	42,68	44,32	43,58	52,05	50,00	51,01
10	Yuk Mengenal Masjid Quba	53,42	51,32	52,35	42,68	44,32	43,58	52,05	50,00	51,01
	RATA-RATA	43,28	48,02	45,00	30,24	34,05	31,62	41,91	46,52	43,59

SIMPULAN

Berdasarkan hasil yang diperoleh dari algoritma *textrank* pada dokumen Bahasa Indonesia di atas, pemilihan skor kalimat tidak harus mengambil *top-n* saja. Dari pengujian yang sudah dilakukan, untuk CR=50%, terlihat bahwa hasil ringkasan lebih baik bila mengambil kelompok kalimat dengan skor *textrank* di posisi akhir (*bottom-n*), dengan rata-rata s 68,76% pada *ROUGE-1*, 60,60% pada *ROUGE-2*, dan 68,29% pada *ROUGE-L*. Sedangkan untuk CR=30%, hasil ringkasan lebih baik dengan mengambil kelompok skor *textrank* bagian

PUSTAKAAN

- J. Pragantha, T. Informatika, F. T. Informasi, and U. Tarumanagara, "Penerapan Algoritma *Textrank* Untuk Automatic Summarization Pada Dokumen Berbahasa Indonesia," vol. 1, no. 1, pp. 71–78, 2017.
- J. Kurniawati and S. Baroroh, "Literasi Media Digital Mahasiswa Universitas Muhammadiyah Bengkulu," *J. Komun.*, vol. 8, no. 2, pp. 51–66, 2016, [Online]. Available: <https://journal.umy.ac.id/index.php/jkm/article/download/2069/2586>
- D. State Ismi and F. Ardianto, "Peringkasan Efektif Teks Bahasa Indonesia dengan Pendekatan Unsupervised Menggunakan Metode Clustering," *Cybernetics*, vol. 3, no. 02, p. 90, 2020, doi: 10.29406/cbn.v3i02.2290.
- Y. Yuliska and K. U. Syaliman, "Literatur Review Terhadap Metode, Aplikasi dan Dataset Peringkasan Dokumen Teks Otomatis untuk Teks Berbahasa Indonesia," *IT J. Res. Dev.*, vol. 5, no. 1, pp. 19–31, 2020, doi: 10.2299/itjrd.2020.vol5(1).4688.
- Y. M. Sari and N. S. Fatonah, "Peringkasan Teks Otomatis pada Modul Pembelajaran Berbahasa Indonesia Menggunakan Cross Late Semantic Analysis (CLSA)," vol. 7, no. 2, pp. 153–159, 2021.
- F. K. Youmarsy, "Evaluation of Automatic Text Summarizations based on Human Summaries," *Procedia - Soc. Behav. Sci.*, vol. 192, pp. 83–91, 2015, doi: 10.1016/j.sbspro.2015.06.013.
- M. Akademi and M. Pematangsiantar, "Penerapan algoritma dalam model peringkasan teks untuk mempercepat pekerjaan

tengah (*middle-n*), dengan rata-rata rata s 45.00%, 31.62% dan 43.59% pada *ROUGE-1*, *ROUGE-2*, dan *ROUGE-L* secara berturut-turut.

Didapatkan kesimpulan bahwa peringkasan teks otomatis dengan *jaccard similarity* pada metode *textrank*, kurang baik bila mengambil kelompok skor awal, disebabkan kalimat-kalimat yang mirip lebih banyak terpilih sebagai ringkasan. Hal ini kurang dapat memberikan lebih banyak fakta dari artikel yang diringkaskan.

- akademik," vol. 1, no. 2, pp. 19–26, 2021.
- M. A. Zamzam, "Sistem Automatic Text Summarization Menggunakan Algoritma *Textrank*," *Matics*, vol. 12, no. 2, pp. 111–116, 2020, doi: 10.18860/mat.v12i2.8372.
- D. Gunawan, S. H. Harahap, and R. F. Rahmat, "Peringkasan multi-dokumen dengan menggunakan *Textrank* dan Relevansi Marjinal Maksimal untuk Teks di Bahasa Indonesia," 2019.
- F. S. Jumeilah, "Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 1, pp. 19–25, 2017, doi: 10.29207/resti.v1i1.11.
- M. Kannan, S., Gurusamy, V., Vijayarani, S., Ilamathi, J. & Nithya, "Preprocessing Techniques for Text Mining Preprocessing Techniques for Text Mining," *Int. J. Comput. Sci. Commun. Networks*, vol. 5, no. October 2014, pp. 7–16, 2015.
- M. F. Fakhrezi, M. A. Bijaksana, and A. F. Huda, "Implementation of Automatic Text Summarization with *Textrank* Method in the Development of Al-Qur'an Vocabulary Encyclopedia," *Procedia Comput. Sci.*, vol. 179, no. 2020, pp. 391–398, 2021, doi: 10.1016/j.procs.2021.01.021.
- A. F. Zulfikar, "Pengembangan Algoritma Stemming Bahasa Indonesia dengan Pendekatan Dictionary Base Stemming untuk Menentukan Kata Dasar dari Kata yang Berimbuhan," *J. Inform. Univ. Pamulang*, vol. 2, no. 3, p. 143, 2017, doi: 10.32493/informatika.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LETTER OF ACCEPTANCE



Kolaborasi Sains dan Teknologi Dalam Mengatasi
Problematika Pangan dan Energi

SURAT PENERIMAAN MAKALAH

Kepada,

Faizah Husniah

Assalamu'alaikum wr wb.

Salam dari TEKNOKA!

Dengan sukacita kami sampaikan bahwa setelah melakukan tinjauan terhadap makalah anda yang berjudul,

Peringkasan Teks Otomatis Artikel Berbahasa Indonesia Menggunakan Algoritma Textrank

Dinyatakan **DITERIMA** dengan tidak mengubah isi makalah untuk diterbitkan pada Journal Elektronik Serial Prosiding Seminar Nasional TEKNOKA, ISSN (Cetak) 2502-8782, ISSN (Online) 2580-6408. Kami mengingatkan kembali, acara Seminar Nasional Teknoka Ke-7 2022 akan diselenggarakan pada tanggal 3 Desember 2022

Hormat Kami,



Yos Nofendri, S.Pd., MSME

Ketua Panitia Seminar Nasional TEKNOKA 7

FAKULTAS TEKNIK UHAMKA

Jl. Tanah Merdeka No. 6 Kp. Rambutan, Ps. Rebo, Jakarta Timur

(021) 87782739

(021) 8400941

teknoka@uhamka.ac.id

teknoka.uhamka.ac.id



SERTIFIKAT

Diberikan Kepada :

Faizah Husniah

Sebagai :

Pemakalah

Makalah Berjudul :

Peringkasan Teks Otomatis Artikel Berbahasa Indonesia
Menggunakan Algoritma Textrank

Dalam Seminar Nasional Teknoka-7 2022, dengan Tema
KOLABORASI SAINS DAN TEKNOLOGI DALAM MENGATASI
PROBLEMATIKA PANGAN DAN ENERGI DI INDONESIA

Sabtu, 3 Desember 2022



Dekan
Fakultas Teknologi Industri dan Informatika

Dr. Dan Mugisidi, S.T., M.Si.



Ketua Panitia
Seminar Nasional Teknoka-7
Yos Nofendi, S.Pd., MSME.

UIN SUSKA RIAU

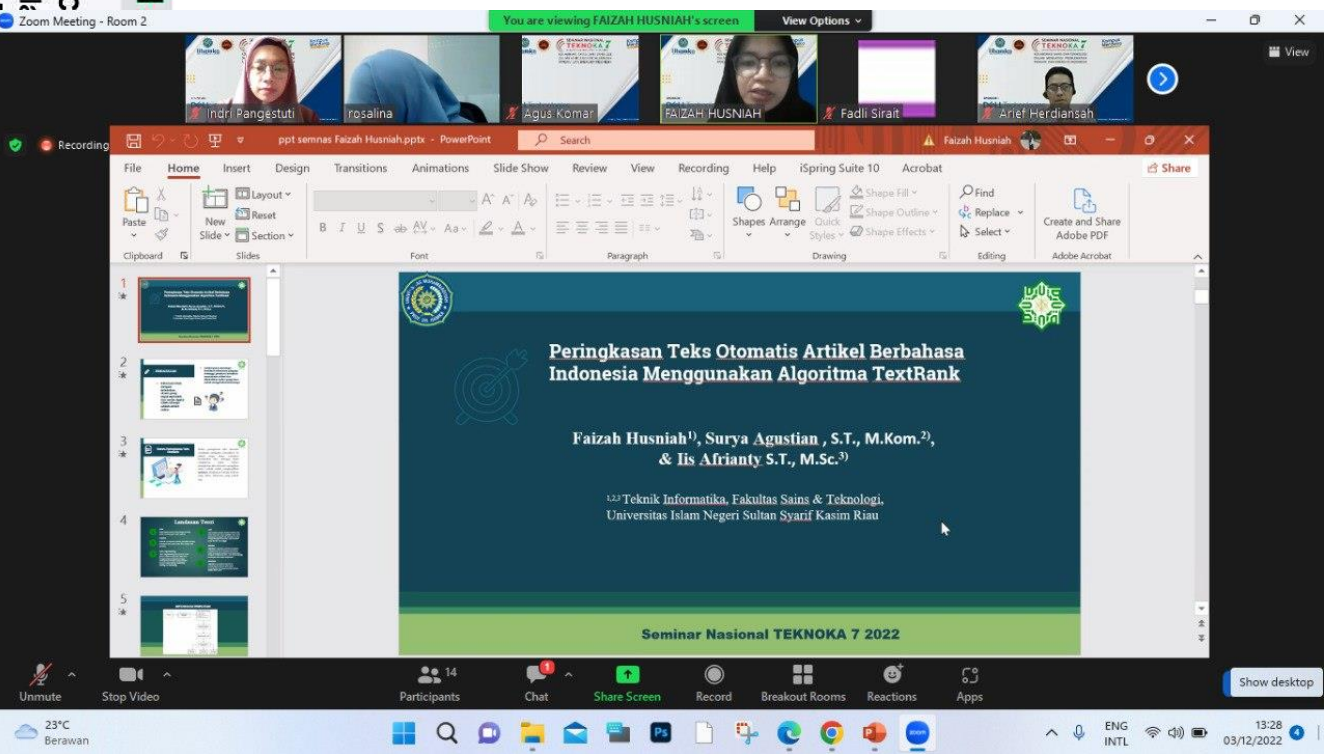
BUKTI PRESENTASI



- Hak C**
1. Dilk
 2. Diarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
1. Dilk
- a. Fungtuhan nanya unuk kepentinngan penairnkan, penenian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Penguipaan tidak merugikan kepentinngan yang wajar UIN Suska Riau.
2. Diarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

mencantumkan dan menyebutkan sumber:

State Islamic University of Sultan Syarif Kasim Riau



Link Presentasi :

[https://drive.google.com/drive/folders/1ZmAIDHCL1QSbdYKkFbhh1CtVr5VrSLXXB?usp=share link](https://drive.google.com/drive/folders/1ZmAIDHCL1QSbdYKkFbhh1CtVr5VrSLXXB?usp=share_link)



UIN SUSKA RIAU