

**PERINGKASAN TEKS ARTIKEL BERITA MENGGUNAKAN
ALGORITMA *TEXTRANK* DENGAN WDF DAN LSA**

SKRIPSI

Oleh :
SHAFIRA HALMAHERA
NIM. 210605110008



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

**PERINGKASAN TEKS ARTIKEL BERITA MENGGUNAKAN
ALGORITMA *TEXTRANK* DENGAN WIDF DAN LSA**

SKRIPSI

**Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)**

**Oleh :
SHAFIRA HALMAHERA
NIM. 210605110008**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

HALAMAN PERSETUJUAN

**PERINGKASAN TEKS ARTIKEL BERITA MENGGUNAKAN
ALGORITMA *TEXTRANK* DENGAN WIDF DAN LSA**

SKRIPSI

Oleh :
SHAFIRA HALMAHERA
NIM. 210605110008

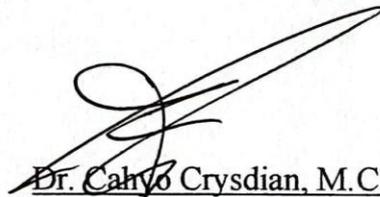
Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 14 April 2025

Pembimbing I,



Prof. Dr. Muhammad Faisal, M.T
NIP. 19740510 200501 1 007

Pembimbing II,



Dr. Cahyo Crysdian, M.CS
NIP. 19740424 200901 1 008

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



Dr. I. Fachrul Kurniawan, M.MT., IPU
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

PERINGKASAN TEKS ARTIKEL BERITA MENGGUNAKAN ALGORITMA *TEXTRANK* DENGAN WIDF DAN LSA

SKRIPSI

Oleh :

SHAFIRA HALMAHERA
NIM. 210605110008

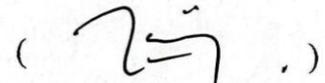
Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 07 Mei 2025

Susunan Dewan Penguji

Ketua Penguji : Dr. Zainal Abidin, M.Kom
NIP. 19760613 200501 1 004



Anggota Penguji I : Ashri Shabrina Afrah, M.T
NIP. 19900430 202012 2 003



Anggota Penguji II : Prof. Dr. Muhammad Faisal, M.T
NIP. 19740510 200501 1 007



Anggota Penguji III : Dr. Cahyo Crysdian, M.CS
NIP. 19740424 200901 1 008



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




Dr. Fachrul Kurniawan, M.MT., IPU
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Shafira Halmahera
NIM : 210605110008
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Peringkasan Teks Artikel Berita Menggunakan Algoritma *TextRank* dengan WIDF dan LSA

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, 20 Mei 2025

Yang membuat pernyataan,



Shafira Halmahera
NIM.210605110008

HALAMAN MOTTO

Allah dulu, Allah lagi, Allah terus, Allah selamanya.

HALAMAN PERSEMBAHAN

Tidak ada bagian dalam laporan skripsi ini yang lebih indah selain lembar persembahan. Dengan penuh rasa syukur kepada Allah SWT atas rahmat, hidayah, dan inayah-Nya, sehingga saya dapat menyelesaikan penelitian ini dengan baik. Melalui kesempatan ini, dengan tulus hati dan rasa terima kasih, penulis persembahkan skripsi ini kepada:

Orang tua saya, Ibu (Yayuk) dan Ayah (Sucipto) yang telah memberikan dukungan moril maupun materil. Ketika orang lain menutup telinga, mereka membuka hati mereka untuk saya. Disaat saya kehilangan kepercayaan diri, mereka selalu ada untuk menguatkan saya. Tidak ada hentinya mendoakan di setiap langkah saya, memberikan cinta, kasih sayang, serta perjuangan dan pengorbanan yang tidak dapat digantikan oleh apapun dan siapapun.

Bapak (Prof. Dr. Muhammad Faisal, M.T) selaku dosen Pembimbing I, Bapak (Dr. Cahyo Crysdiyan, M.CS) selaku dosen Pembimbing II, yang telah membimbing, memberikan banyak saran, dan pelajaran dalam pembuatan skripsi ini.

Terakhir kepada diri saya sendiri, yang telah menghargai setiap proses, berjuang dan bertahan.

KATA PENGANTAR

Assalamualaikum, Wr. Wb.

Alhamdulillah rabbil 'alamin puji syukur ke hadirat Allah *Subhanahu Wa Ta'ala*, yang telah memberikan rahmat, hidayah, dan memberikan kelancaran pada penulis sehingga dapat menyelesaikan skripsi ini dengan baik. Sholawat serta salam tetap tercurahkan kepada Baginda Nabi Muhammad *shallallahu 'alaihi wasallam*.

Penulis ingin mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan, bimbingan, dan dorongan selama proses penyusunan skripsi ini. Ucapan terima kasih penulis sampaikan kepada:

1. Prof. Dr. H. M. Zainudin, M.A., Selaku Rektor Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Harini, M.Si selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang
3. Dr. Ir. Fachrul Kurniawan, M.MT., IPU selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Prof. Dr. Muhammad Faisal, M.T selaku Dosen Pembimbing I yang telah memberi arahan, memotivasi, meluangkan waktu untuk membimbing dengan baik dan sabar selama proses penyusunan skripsi.
5. Dr. Cahyo Crys dian, M.CS selaku Dosen Pembimbing II yang telah membimbing, memberikan arahan serta masukan kepada penulis sehingga penulis dapat menyelesaikan skripsi ini.

6. Dr. Zainal Abidin, M.Kom selaku Penguji I dan Ashri Shabrina Afrah, M.T selaku Penguji II yang telah meluangkan waktunya untuk menguji dan memberikan saran serta kritik selama proses pengujian skripsi ini.
7. Segenap Dosen Program Studi Teknik Informatika yang telah memberikan ilmu yang telah dimiliki kepada penulis, arahan, dan berbagi pengetahuan.
8. Semua sivitas akademika yang telah membantu proses administrasi.
9. Ibu (Yayuk) dan Ayah (Sucipto), cinta pertama dan pintu surgaku. Terima kasih yang tak terhingga atas segala perjuangan, pengorbanan, dan kasih sayang tulus yang kalian berikan. Meskipun mereka tidak pernah merasakan bangku perkuliahan, namun mereka selalu berusaha memberikan yang terbaik sehingga saya dapat merasakan dan menuntaskan perkuliahan S1.
10. Teman-teman Aster 2021 yang telah memberikan dukungan positif dan semangat untuk terus berjuang.
11. Teman teman terdekat selama masa kuliah, Vivin Octavia Cahyani, Eka Mira Novita Subroto, Radina Mutia Haira, yang telah menjadi keluarga penulis selama perantauan.
12. Semua pihak yang telah turut serta baik secara langsung maupun tidak langsung dalam menyelesaikan skripsi ini.
13. Diri sendiri, karena selalu menghargai setiap proses, terus berjuang dan bertahan. Dan tentu saja, bukan karena kemampuan sendiri, tapi karena Allah yang memberikan kemudahan.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, kritik, saran, dan masukan dari berbagai pihak sangat diharapkan untuk

perbaikan dan penyempurnaan skripsi ini. Penulis berharap penelitian ini dapat memberikan manfaat bagi pembaca dan peneliti selanjutnya. Akhir kata, penulis mohon maaf atas segala keterbatasan dan kekurangan dalam skripsi ini. Semoga Allah SWT senantiasa melimpahkan rahmat dan keberkahan dalam setiap langkah perjalanan hidup kita.

Wassalamualaikum Wr. Wb

Malang, 20 Mei 2025

Penulis

DAFTAR ISI

HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
HALAMAN MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
ABSTRAK	xv
ABSTRACT	xvi
مستخلص البحث	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Pernyataan Masalah	5
1.3 Batasan Masalah	6
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	6
BAB II STUDI PUSTAKA	7
2.1 Peringkasan Teks	7
2.2 <i>TextRank</i>	11
2.3 Pembobotan	15
2.3.1 <i>Term Frequency (TF)</i>	15
2.3.2 <i>Inverse Document Frequency (IDF)</i>	15
2.3.3 <i>Term Frequency - Invers Document Frequency (TF-IDF)</i>	16
2.3.4 <i>Weighted Inverse Document Frequency (WIDF)</i>	16
2.4 <i>Latent Semantic Analysis (LSA)</i>	19
BAB III DESAIN DAN IMPLEMENTASI	23
3.1 Data Penelitian	23
3.2 Desain Sistem	24
3.3 <i>Preprocessing</i>	25
3.3.1 <i>Sentence Tokenizing</i>	26
3.3.2 <i>Text Cleaning</i>	27
3.3.3 <i>Stopwords Removal dan Word Tokenizing</i>	28
3.4 Pembobotan Kata	28
3.5 <i>Latent Semantic Analysis</i>	30
3.6 <i>Similarity Matrix</i>	34
3.7 Algoritma <i>TextRank</i>	35
3.7 Pembentukan Ringkasan	38
BAB IV HASIL DAN PEMBAHASAN	40
4.1 Skenario Pengujian	40
4.1.1 <i>Recall</i>	41

4.1.2 <i>Precision</i>	42
4.1.3 <i>F1-Score</i>	42
4.2 Hasil Uji Coba	43
4.3 Pembahasan	48
4.3.1 Aspek Muamalah Ma'a Allah	52
4.3.2 Aspek Muamalah Ma'a An-Nas	54
BAB V KESIMPULAN DAN SARAN	56
5.1 Kesimpulan	56
5.2 Saran	57
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR GAMBAR

Gambar 1.1 Sumber Berita Utama Masyarakat Indonesia (2021-2023).....	1
Gambar 3.1 Bentuk Indosum dataset	23
Gambar 3.2 Struktur File IndoSum Dataset.....	24
Gambar 3.3 Desain Sistem.....	25
Gambar 3.4 Tahap Preprocessing	26
Gambar 3.5 Matriks WIDF	30
Gambar 3.6 Matriks WIDF setelah ditranspos.....	31
Gambar 3.7 Matriks U.....	32
Gambar 3.8 Matriks V.....	32
Gambar 3.9 Matriks Singular	33
Gambar 3.10 Matriks LSA Setelah Reduksi Dimensi	34
Gambar 3.11 Representasi Graf berdasarkan Similarity.....	37
Gambar 4.1 Perbandingan rata-rata hasil evaluasi dengan dan tanpa LSA	49

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait Peringkasan Teks	7
Tabel 2.2 Penelitian Terkait TextRank	12
Tabel 2.3 Penelitian Terkait WIDF	17
Tabel 2.4 Penelitian Terkait Latent Semantic Analysis	20
Tabel 3.1 Hasil sentence tokenizing.....	26
Tabel 3.2 Hasil text cleaning.....	27
Tabel 3.3 Hasil stopwords removal dan word tokenizing.....	28
Tabel 3.4 Pembobotan kata dengan WIDF	29
Tabel 3.5 Matriks cosine similarity.....	35
Tabel 3.6 Skor TextRank Untuk Setiap Kalimat	37
Tabel 3.7 Hasil Pengurutan Skor TextRank.....	38
Tabel 3.8 Hasil Ringkasan Dokumen.....	39
Tabel 4.1 Hasil uji coba	43
Tabel 4.2 Statistik artikel berita	44
Tabel 4.3 Daftar kata hasil ringkasan sistem	44
Tabel 4.4 Daftar kata hasil ringkasan manual	44
Tabel 4.5 Daftar kemunculan kata bersama	45
Tabel 4.6 Statistika jumlah kata ringkasan	45
Tabel 4.7 Hasil evaluasi ROUGE-1 tiap artikel (compression rate 50%).....	47
Tabel 4.8 Hasil evaluasi ROUGE-1 tiap artikel (compression rate 30%).....	47
Tabel 4.9 Rata-rata hasil evaluasi ROUGE-1	49

ABSTRAK

Halmahera, Shafira. 2025. **Peringkasan Teks Artikel Berita Menggunakan Algoritma *TextRank* dengan WIDF dan LSA**. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Prof. Dr. Muhammad Faisal, M.T (II) Dr. Cahyo Crysdiyan, M.CS.

Kata Kunci: Peringkasan Teks Otomatis, *TextRank*, WIDF, LSA.

Pesatnya perkembangan teknologi informasi, khususnya internet, telah mempercepat penyebaran berita secara *online*. Masyarakat kini dapat mengakses berbagai informasi dengan mudah melalui *platform* media *online*. Namun, banyaknya informasi yang tersedia menimbulkan tantangan bagi pembaca untuk menemukan inti berita dengan cepat. Untuk mengatasi hal ini, dibutuhkan peringkasan teks otomatis. Salah satu metode peringkasan adalah peringkasan ekstraktif, yang bertujuan mencari kalimat-kalimat penting dalam dokumen yang mewakili keseluruhan isi dokumen. Penelitian ini menggunakan metode *TextRank* yang dikombinasikan dengan pembobotan WIDF (*Weighted Inverse Document Frequency*) dan reduksi dimensi menggunakan LSA (*Latent Semantic Analysis*). WIDF menghitung bobot kata berdasarkan frekuensi kemunculannya dalam suatu teks dan total kemunculannya di kumpulan teks. LSA mereduksi dimensi matriks WIDF dan merepresentasikan kalimat dalam ruang semantik berdimensi rendah melalui dekomposisi matriks SVD. *TextRank* digunakan untuk memeringkatkan kalimat-kalimat dalam dokumen berdasarkan kemiripan semantik yang dihitung menggunakan *cosine similarity*, kemudian memilih kalimat dengan peringkat tertinggi sebagai ringkasan. Hasil uji coba dengan 14.261 artikel berita menunjukkan bahwa metode *TextRank* dengan pembobotan WIDF tanpa menggunakan LSA menghasilkan performa yang lebih baik dibandingkan dengan kombinasi yang menggunakan LSA. Pada tingkat kompresi 50%, diperoleh nilai rata-rata ROUGE-1 dengan *recall* sebesar 0,6884, *precision* 0,3664, dan *f1-score* 0,4658. Sementara itu, pada tingkat kompresi 30%, diperoleh *recall* sebesar 0,5442, *precision* 0,4345, dan *f1-score* 0,4666.

ABSTRACT

Halmahera, Shafira. 2025. **News Article Text Summarization Using TextRank Algorithm with WIDF and LSA**. Undergraduate Thesis. Informatics Engineering Study Program, Faculty of Science and Technology Maulana Malik Ibrahim State Islamic University Malang. Supervisor: (I) Prof. Dr. Muhammad Faisal, M.T (II) Dr. Cahyo Crys dian, M.CS.

Keywords: Automatic Text Summarization, Textrank, WIDF, LSA.

The rapid development of information technology, especially the internet, has accelerated the spread of news online. People can now access various information easily through online media platforms. However, the sheer amount of information available poses a challenge for readers to find the gist of the news quickly. To overcome this, automatic text summarization is needed. One of the summarization methods is extractive summarization, which aims to find important sentences in the document that represent the entire content of the document. This research uses the TextRank method combined with WIDF (Weighted Inverse Document Frequency) weighting and dimension reduction using LSA (Latent Semantic Analysis). WIDF calculates the weight of a word based on the frequency of its occurrence in a text and its total occurrence in the text set. LSA reduces the dimension of the WIDF matrix and represents sentences in a low-dimensional semantic space through SVD matrix decomposition. TextRank is used to rank the sentences in the document based on semantic similarity calculated using cosine similarity, and then select the highest ranked sentence as the summary. Experimental results with 14,261 news articles show that the TextRank method with WIDF weighting without using LSA produces better performance than the combination using LSA. At 50% compression level, the ROUGE-1 average value is obtained with recall of 0.6884, precision of 0.3664, and f1-score of 0.4658. Meanwhile, at a compression level of 30%, a recall of 0.5442, precision of 0.4345, and f1-score of 0.4666 were obtained.

مستخلص البحث

هالمهيرة، شافيرا 2025. تلخيص نصوص المقالات الإخبارية باستخدام خوارزمية TextRank مع WIDF و LSA. الأطروحة. برنامج دراسة هندسة المعلوماتية، كلية العلوم والتكنولوجيا، الجامعة الإسلامية الحكومية، مولانا مالك إبراهيم مالانج. المشرف: (الأول) أ. د. محمد فيصل، ماجستير (الثاني) د. كاهيو كريسديان، ماجستير في علوم الحاسب.

الكلمات الرئيسية: التلخيص التلقائي للنصوص، تلخيص النص التلقائي، TextRank، WIDF، LSA.

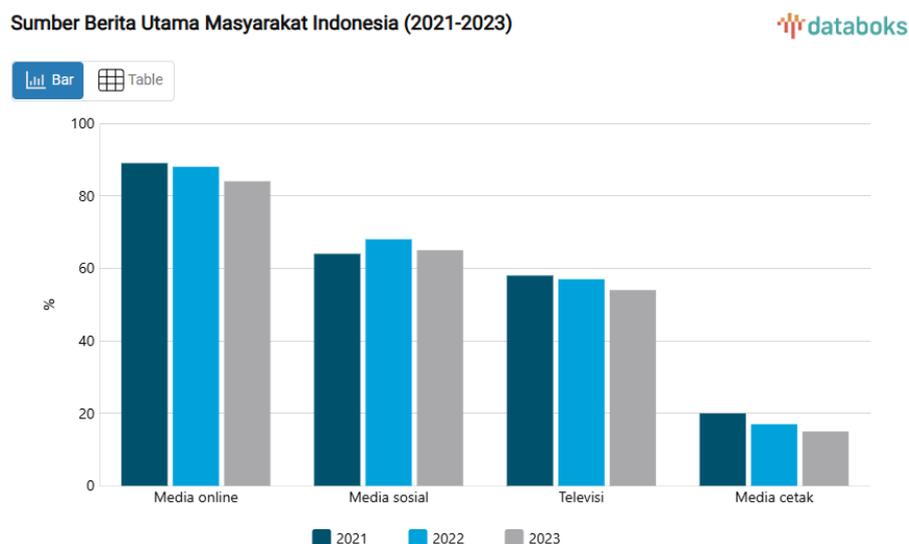
أدى التطور السريع لتكنولوجيا المعلومات، وخاصة الإنترنت، إلى تسريع انتشار الأخبار عبر الإنترنت. ويمكن للأشخاص الآن الوصول إلى المعلومات المختلفة بسهولة من خلال منصات وسائل الإعلام عبر الإنترنت. ومع ذلك، فإن كمية المعلومات المتاحة تشكل تحديًا للقراء للعثور على جوهر الأخبار بسرعة. للتغلب على ذلك، هناك حاجة إلى تلخيص النص تلقائيًا. إحدى طرق التلخيص هي التلخيص الاستخراجي الذي يهدف إلى العثور على الجمل المهمة في المستند التي تمثل محتوى المستند بأكمله. يستخدم هذا البحث طريقة TextRank مع ترجيح WIDF (التردد العكسي المرجح للمستند) وتقليل الأبعاد باستخدام LSA (التحليل الدلالي الكامن). يحسب WIDF وزن الكلمة بناءً على تكرار ورودها في النص وإجمالي ورودها في مجموعة النصوص. يقلل LSA من أبعاد مصفوفة WIDF ويمثل الجمل في فضاء دلالي منخفض الأبعاد من خلال تحليل مصفوفة SVD. يتم استخدام TextRank لترتيب الجمل في المستند بناءً على التشابه الدلالي المحسوب باستخدام تشابه جيب التمام، ثم تحديد الجملة الأعلى مرتبة كملخص. تُظهر النتائج التجريبية التي أجريت على 14,261 مقالة إخبارية أن طريقة TextRank مع ترجيح WIDF دون استخدام LSA تنتج أداءً أفضل من الجمع باستخدام LSA. عند مستوى ضغط بنسبة 50%، يتم الحصول على متوسط قيمة ROUGE-1 مع استدعاء 0.6884، ودقة 0.3664، ودرجة f1 0.4658. وفي الوقت نفسه، عند مستوى ضغط بنسبة 30%، تم الحصول على استدعاء 0.5442، ودقة 0.4345، ودرجة f1 0.4666.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi telah membawa internet menjadi salah satu pendorong utama dalam mempercepat penyebaran informasi. Dengan perkembangan internet yang sangat pesat, masyarakat kini dapat mengakses berbagai informasi dengan mudah, termasuk artikel dan berita yang mencakup topik-topik seperti pendidikan, teknologi, ekonomi, kesehatan, olahraga, dan lain sebagainya (Rozan & Dewi, 2022). Informasi ini tidak hanya tersedia melalui media cetak, tetapi juga melalui media *online*. Saat ini, masyarakat cenderung memilih media *online* sebagai sumber berita utama karena kemudahannya dalam akses serta beragamnya pilihan informasi yang tersedia.



Gambar 1.1 Sumber Berita Utama Masyarakat Indonesia (2021-2023)
Sumber : Katadata.co.id (2023)

Berdasarkan data dari Katadata.co.id (2023), mayoritas masyarakat Indonesia mendapatkan berita melalui berbagai *platform* media. Sekitar 90% responden menyatakan bahwa media *online* adalah sumber utama mereka, sedangkan 70% lainnya menggunakan media sosial untuk memperoleh informasi terkini. Sementara itu, televisi tetap menjadi pilihan bagi 60% responden, dan media cetak digunakan oleh sekitar 20% responden. Meskipun media *online* dan media sosial mendominasi, terdapat tantangan dalam meningkatkan kualitas konten berita yang disajikan.

Kemudahan akses informasi secara *online* menyebabkan jumlah berita yang tersedia semakin banyak dan beragam. Namun, berita biasanya memiliki teks panjang yang membutuhkan waktu untuk memahami inti atau ide utamanya. Oleh karena itu, diperlukan teknologi peringkasan teks otomatis untuk membantu proses ekstraksi informasi penting dari suatu berita. Salah satu pendekatan yang dapat digunakan adalah peringkasan teks (*text summarization*) (Mandar & Gunawan, 2018).

Peringkasan teks otomatis merupakan teknologi yang dirancang untuk menghasilkan ringkasan dari teks asli tanpa mengurangi esensi utamanya. Teknologi ini bertujuan untuk membantu pembaca menghemat waktu dengan memungkinkan mereka memahami informasi penting secara cepat (Halimah *et al.*, 2022). Ringkasan yang mencakup frasa kunci dan semua informasi penting serta relevan dari dokumen asli secara otomatis dihasilkan menggunakan peringkasan teks (Widyassari *et al.*, 2022). Dengan teknologi tersebut, pembaca dapat

memperoleh inti artikel berita dalam format yang lebih singkat tanpa perlu membaca seluruh teks (AL-Hafiidh *et al.*, 2022).

Peringkasan teks menjadi salah satu topik penting dalam bidang *Natural Language Processing* (NLP) dan *Data Mining* (DM) (Naik & Gaonkar, 2017). Teknologi ini memiliki banyak aplikasi yang bermanfaat di berbagai bidang. Sebagai contoh, dalam bisnis, peringkasan teks digunakan untuk meringkas ulasan makanan atau umpan balik pelanggan terhadap produk (Vijaya Saraswathi *et al.*, 2022). Penelitian lain menunjukkan implementasi peringkasan teks pada dokumen esai mahasiswa di sistem *E-Learning* Universitas Udayana oleh Suwija Putra *et al.* (2021) dan aplikasi peringkasan otomatis untuk atlas penyakit menular pada manusia (Caesarardhi *et al.*, 2023).

Di era digital dan industri 5.0, manusia dituntut untuk menyelesaikan tugas dengan cepat dan efisien, termasuk dalam membaca. Dengan beragam informasi yang melimpah, pembaca perlu memahami teks panjang dalam waktu singkat untuk menghemat waktu. Peringkasan teks membantu pembaca untuk memahami isi teks dengan cepat, sehingga dapat menghemat banyak waktu. Dalam islam, pentingnya memanfaatkan waktu dengan baik juga diajarkan. Allah *Subhanahu wa Ta'ala* bersumpah dengan waktu atau masa dalam QS. Al-Ashr ayat 1-2:

وَالْعَصْرِ ﴿١﴾ إِنَّ الْإِنْسَانَ لَفِي خُسْرٍ

“Demi masa. Sesungguhnya manusia itu benar-benar dalam kerugian.” (QS. Al-Ashr 103: ayat 1-2).

Penafsiran dalam kitab Tafsir Jalalain memberikan penjelasan mengenai ayat tersebut yaitu وَالْعَصْرِ (Demi masa) atau zaman atau waktu yang dimulai dari

tergelincirnya matahari hingga terbenamnya, maksudnya adalah waktu sholat Asar. لَفِي خُسْرٍ إِنَّ الْإِنْسَانَ (sesungguhnya manusia itu) yang dimaksud adalah jenis manusia (benar-benar berada dalam kerugian) di dalam perniagaannya (Al-Mahalli & As-Suyuti, 1469).

Peringkasan teks berdasarkan input terbagi menjadi dua jenis, yaitu *single-document* dan *multi-document*. *Single-document* berfokus pada meringkas satu dokumen, sedangkan *multi-document* menggabungkan beberapa dokumen menjadi satu ringkasan (Akhmetov et al., 2021). Berdasarkan pendekatannya, terdapat dua jenis utama peringkasan teks, yaitu ekstraktif dan abstraktif (Akhmetov et al., 2021). Penelitian ini menggunakan peringkasan ekstraktif karena prosesnya lebih sederhana dibandingkan peringkasan abstraktif. Pendekatan ini memilih kalimat-kalimat dari dokumen asli yang dianggap mewakili inti pembahasan tanpa melakukan *paraphrase* (Madhuri & Ganesh Kumar, 2019).

Salah satu algoritma yang digunakan dalam peringkasan ekstraktif adalah *TextRank*, sebuah metode *unsupervised* yang mengevaluasi relevansi kalimat berdasarkan bobotnya. Kalimat dengan skor tertinggi diprioritaskan untuk membentuk ringkasan akhir (Yuliska & Syaliman, 2020). Keunggulan *TextRank* adalah tidak memerlukan data latih, sehingga lebih ringan secara komputasi (Eris et al., 2017). Namun, dalam beberapa kasus, algoritma ini tidak optimal tanpa ekstraksi fitur yang tepat untuk mengidentifikasi peringkat kalimat (Karo et al., 2024). Meskipun *TextRank* telah banyak digunakan, pengembangan yang melibatkan teknik tambahan seperti WIDF dan LSA masih jarang dilakukan, terutama dalam konteks artikel berita berbahasa Indonesia. Penelitian oleh Karo et

al (2024) menunjukkan bahwa kombinasi *TextRank* dengan teknik ekstraksi fitur seperti TF-IDF mampu meningkatkan akurasi peringkasan (Karo *et al.*, 2024). WIDF (*Weighted Inverse Document Frequency*) merupakan teknik pembobotan alternatif yang memberikan bobot lebih signifikan pada kata-kata penting dalam artikel berita. Metode ini merupakan pengembangan dari *Inverse Document Frequency* (IDF) dalam pendekatan TF-IDF (Susandi & Sholahudin, 2017).

Penelitian peringkasan teks mencakup berbagai bahasa, seperti Kannada, Vietnam, Inggris, dan Arab. Namun sebagian besar masih menggunakan dataset berbahasa Inggris atau non-Indonesia (Hidayattullah & Azizi, 2021). Penggunaan dataset yang berbahasa Indonesia memberikan peluang untuk mengembangkan peringkasan teks yang lebih sesuai dan relevan dengan kebutuhan masyarakat lokal. Penelitian ini akan mengimplementasikan algoritma *TextRank* dengan menggunakan *Weighted Inverse Document Frequency* (WIDF) untuk pembobotan kata dan *Latent Semantic Analysis* (LSA) sebagai metode reduksi dimensi. Diharapkan hasil penelitian ini dapat menghasilkan ringkasan teks yang efektif pada artikel berita.

1.2 Pernyataan Masalah

Seberapa baik performa algoritma *TextRank* yang dikombinasikan dengan WIDF dan LSA dalam menghasilkan ringkasan teks artikel berita berdasarkan evaluasi ROUGE menggunakan *recall*, *precision*, dan *f1-score*?

1.3 Batasan Masalah

1. Penelitian ini berfokus pada peringkasan ekstraktif dengan input berupa *single-dokumen*.
2. Penelitian ini menggunakan dataset *Indonesian Text Summarization* (Indosum).

1.4 Tujuan Penelitian

Mengukur performa algoritma *TextRank* yang dikombinasikan dengan WIDF dan LSA dalam menghasilkan ringkasan teks artikel berita menggunakan evaluasi ROUGE dengan parameter *recall*, *precision*, dan *f1-score*.

1.5 Manfaat Penelitian

1. Perusahaan media dan jurnalistik, peringkasan teks dapat digunakan oleh kantor berita, redaksi surat kabar, dan media *online* untuk menyusun ringkasan artikel berita otomatis, yang dapat mempercepat proses penyajian informasi penting kepada masyarakat.
2. *Platform* berita digital, peringkasan teks dapat diimplementasikan dalam *platform* berita digital untuk menyediakan ringkasan berita otomatis, yang dapat meningkatkan kenyamanan dan efisiensi pengguna dalam mengakses informasi.
3. Akademisi dan peneliti, dapat menjadi referensi bagi akademisi dan peneliti untuk pengembangan lebih lanjut dalam bidang pengolahan bahasa alami (NLP) dan sistem peringkasan teks.

BAB II

STUDI PUSTAKA

2.1 Peringkasan Teks

Peringkasan teks (*Automatic Text Summarization*) merupakan salah satu bidang penting dalam pemrosesan bahasa alami (*Natural Language Processing*). Bidang ini telah mengalami perkembangan yang signifikan sejak pertama kali diperkenalkan oleh Luhn pada tahun 1958 (Kresna & Cholissodin, 2021). Peringkasan teks otomatis memanfaatkan teknologi komputer untuk menyederhanakan informasi dari teks yang panjang tanpa mengurangi inti atau makna utamanya. Proses ini memungkinkan pembaca untuk memahami isi utama dokumen dengan lebih cepat dan efisien (Halimah *et al.*, 2022).

Peringkasan teks otomatis dapat dilakukan melalui dua teknik utama, yaitu ekstraktif dan abstraktif. Teknik ekstraktif menyusun ringkasan dengan memilih poin-poin penting langsung dari teks asli tanpa mengubah kata atau frasa. Sementara itu, teknik abstraktif menyusun ulang informasi dengan kata-kata baru, mirip dengan proses *paraphrase* yang dilakukan manusia (See *et al.*, 2017). Beberapa penelitian terdahulu yang membahas peringkasan teks dengan beberapa metode dan algoritma dirangkum pada Tabel 2.1.

Tabel 2.1 Penelitian Terkait Peringkasan Teks

Sumber	Judul	Metode	Data	Hasil
Victor <i>et al.</i> (2019)	<i>Extractive Text Summarization Applied to Speech-to-text Media</i>	Algoritma peringkasan teks ekstraktif (Luhn, TextRank, LexRank, LSA,	Dataset DUC2001, OWIDSum, dan ICSI Corpus	Algoritma Luhn dan TextRank menunjukkan performa terbaik dalam peringkasan teks hasil transkripsi audio menggunakan metode ekstraktif. Rekomendasi

		<i>SumBasic</i> , dan <i>KLSum</i>)		utama untuk <i>speech-to-text summarization</i> adalah Luhn dan <i>TextRank</i> .
Widyassari <i>et al.</i> (2022)	<i>Review of automatic text summarization techniques & methods</i>	<i>Systematic Literature Review (SLR)</i> terhadap teknik dan metode peringkasan teks	Berbagai dataset publik (DUC 2002 & 2004, TAC, <i>Tweet</i> , <i>News</i> , <i>Multilingual</i>) dan privat (novel, mikroblog, bahasa lokal, dll)	Identifikasi topik populer: <i>multi-document</i> dan ekstraktif; pendekatan favorit: <i>machine learning</i> dan statistik; permasalahan utama: redundansi, kohesi, pemrosesan semantik; rekomendasi: gabungan metode untuk meningkatkan performa dan <i>real-time summarization</i> .
Gulden <i>et al.</i> (2019)	<i>Extractive summarization of clinical trial descriptions</i>	Algoritma peringkasan teks ekstraktif (<i>TextRank</i> , <i>LexRank</i> , <i>LSA</i> , Luhn, <i>SumBasic</i> , <i>KLSum</i>)	Dataset dari <i>clinicaltrials.gov</i>	Algoritma <i>TextRank</i> menunjukkan performa terbaik dalam peringkasan deskripsi uji klinis dengan nilai ROUGE-1 F1 sebesar 0.3531, ROUGE-2 F1 sebesar 0.1723, dan ROUGE-L F1 sebesar 0.3003.
Chouigui <i>et al.</i> (2021)	<i>An Arabic Multi-source News Corpus: Experimenting on Single-document Extractive Summarization</i>	Algoritma peringkasan teks ekstraktif (<i>LexRank</i> , <i>TextRank</i> , Luhn, <i>LSA</i>)	ANT corpus v2.1	Algoritma <i>LexRank</i> menunjukkan performa terbaik berdasarkan evaluasi menggunakan metrik ROUGE dan BLEU. Penghapusan <i>stopwords</i> meningkatkan akurasi peringkasan, sedangkan proses <i>stemming</i> cenderung menurunkan kualitas ringkasan.
Nawaz <i>et al.</i> (2020)	<i>Extractive Text Summarization Models for Urdu Language</i>	Pendekatan <i>Local Weight (Sentence Weight & Term Frequency)</i> dan <i>Global Weight (TF-IDF/Vector Space Model)</i> .	Dataset publik berbahasa Urdu	Pendekatan <i>Local Weight</i> lebih efektif dibanding <i>Global Weight</i> . <i>F-score</i> metode <i>Sentence Weight</i> mencapai 80%, <i>Weighted Term Frequency</i> 76%, sedangkan <i>VSM</i> hanya 62%.

Penelitian yang dilakukan oleh Victor *et al.* (2019) mengevaluasi penerapan algoritma peringkasan teks ekstraktif untuk merangkum hasil transkripsi dari media audio ke teks. Penelitian ini menguji enam algoritma peringkasan populer, yaitu *Luhn*, *TextRank*, *LexRank*, *Latent Semantic Analysis (LSA)*, *SumBasic*, dan *KLSum*. *Luhn* menggunakan frekuensi kata dan relevansi untuk menentukan kalimat penting, sedangkan *TextRank* dan *LexRank* memanfaatkan teknik berbasis *graf* untuk meranking kalimat. *LSA* menggunakan dekomposisi matriks untuk menemukan hubungan semantik antar kalimat. *SumBasic* memilih kalimat berdasarkan frekuensi kata secara iteratif, sementara *KLSum* menggunakan divergensi probabilitas untuk mencocokkan ringkasan dengan teks asli. Penelitian ini menggunakan dataset bahasa Inggris seperti DUC2001, OWIDSum, dan ICSI *Corpus*, dengan hasil menunjukkan bahwa algoritma *Luhn* dan *TextRank* adalah yang paling efektif.

Penelitian selanjutnya oleh Widyassari *et al.* (2022) mengkaji berbagai teknik dan metode peringkasan teks otomatis melalui analisis terhadap 85 publikasi dari tahun 2008 hingga 2019. Penelitian ini membahas tren dalam penelitian, dataset yang sering digunakan, langkah *preprocessing*, fitur, teknik, metode, evaluasi, serta tantangan dalam bidang ini. Pendekatan ekstraktif dan abstraktif menjadi fokus utama, dengan dataset publik seperti DUC2002 dan DUC2004 menjadi yang paling sering digunakan untuk menguji performa metode. Penelitian ini membahas secara mendalam mengenai pentingnya fitur seperti panjang dan posisi kalimat dalam peringkasan ekstraktif serta pendekatan berbasis *graf* seperti *TextRank*.

Penelitian selanjutnya oleh Gulden *et al.* (2019) mengkaji penerapan metode peringkasan teks ekstraktif untuk menghasilkan ringkasan deskripsi uji klinis secara singkat. Dataset penelitian diambil dari *clinicaltrials.gov*, yang mencakup 101.016 catatan uji klinis setelah proses penyaringan. Beberapa algoritma ekstraktif seperti *TextRank*, *LexRank*, *LSA*, *SumBasic*, *KLSum*, dan *Luhn* digunakan untuk mengevaluasi performa peringkasan. Kinerja algoritma diukur dengan metrik ROUGE. Hasilnya menunjukkan bahwa *TextRank* memiliki performa terbaik dengan skor ROUGE-1 sebesar 0,3531, ROUGE-2 sebesar 0,1723, dan ROUGE-L sebesar 0,3003.

Penelitian selanjutnya oleh Chouigui *et al.* (2021) meneliti peringkasan teks otomatis dengan pendekatan *single-document extractive text summarization* pada korpus berita multi-sumber berbahasa Arab. Penelitian ini mengevaluasi empat algoritma peringkasan teks ekstraktif, yaitu *LexRank* dan *TextRank* yang berbasis graf, serta *Luhn* dan *Latent Semantic Analysis (LSA)* yang berbasis statistik. Evaluasi dilakukan menggunakan metrik ROUGE dan BLEU untuk mengukur kesesuaian ringkasan yang dihasilkan dengan abstrak referensi. Hasil penelitian menunjukkan bahwa *LexRank* memberikan performa terbaik, dengan skor ROUGE-1 sebesar 0,972 dan BLEU-1 sebesar 0,851, khususnya pada kategori berita olahraga.

Penelitian selanjutnya oleh Nawaz *et al.* (2020) mengembangkan kerangka kerja peringkasan teks ekstraktif untuk bahasa Urdu menggunakan pendekatan *Local Weighting (LW)* dan *Global Weighting (GW)*. LW meliputi metode bobot kalimat yang menghitung bobot berdasarkan proporsi kata penting dalam kalimat,

serta metode bobot frekuensi *term* yang didasarkan pada jumlah kemunculan kata penting dalam teks. Pendekatan GW menggunakan model ruang vektor (*Vector Space Model*) dengan perhitungan TF-IDF untuk mengukur pentingnya kata secara global dalam dokumen. Dataset yang digunakan terdiri dari 50 artikel berbahasa Urdu yang diringkas secara manual oleh ahli. Evaluasi menggunakan metrik ROUGE-1 dan ROUGE-2 menunjukkan bahwa pendekatan LW lebih unggul dibandingkan GW, dengan metode bobot kalimat mencapai *F-score* 80% dan metode bobot frekuensi *term* mencapai 76%. Sebaliknya, pendekatan TF-IDF menghasilkan akurasi sebesar 62%.

2.2 *TextRank*

TextRank adalah algoritma pemeringkatan berbasis *graf* yang pertama kali diusulkan oleh Mihalcea & Tarau pada tahun 2004. Algoritma ini dirancang untuk memberikan skor pada setiap simpul (*node*) berdasarkan hubungan antar elemen dalam *graf* itu sendiri (Hadyan *et al.*, 2017). Dalam *TextRank*, setiap kalimat dianggap sebagai sebuah *vertex*, dan semakin tinggi skor sebuah *vertex*, semakin penting kalimat tersebut dalam ringkasan (Yuliska & Syaliman, 2020). *TextRank* banyak digunakan dalam pemrosesan bahasa alami, khususnya untuk peringkasan teks dokumen tunggal (*single document*) (Fadhila & Nuryana, 2024). Konsep dasar dari algoritma ini terinspirasi oleh algoritma *PageRank* yang digunakan oleh mesin pencari Google (Zhou *et al.*, 2022).

Dalam penerapannya, setiap kalimat dalam dokumen direpresentasikan sebagai simpul dalam sebuah *graf*. Hubungan antar kalimat dibentuk sebagai sisi (*edge*) yang menunjukkan tingkat kemiripan antar kalimat, baik secara leksikal

maupun semantik. Sisi-sisi tersebut kemudian diberi bobot berdasarkan tingkat kemiripan kalimat, semakin tinggi kemiripan, semakin besar bobotnya. Setelah *graf* terbentuk, algoritma iteratif seperti *PageRank* digunakan untuk menghitung skor relevansi tiap simpul. Kalimat-kalimat dengan skor tertinggi dianggap paling penting dan dipilih sebagai ringkasan dokumen (Zhou *et al.*, 2022).

Beberapa penelitian yang membahas penerapan algoritma *TextRank* dalam peringkasan teks dirangkum pada Tabel 2.2.

Tabel 2.2 Penelitian Terkait *TextRank*

Sumber	Judul	Metode	Data	Hasil
(Zaware <i>et al.</i> , 2021)	<i>Text Summarization using TF-IDF and TextRank algorithm</i>	Kombinasi algoritma TF-IDF dan <i>TextRank</i>	Dataset BBC News Summary	Algoritma yang diusulkan menghasilkan ringkasan ekstraktif yang lebih efisien dibandingkan metode TF-IDF standar.
(Yulianti <i>et al.</i> , 2023)	<i>Enhanced TextRank using weighted word embedding for text summarization</i>	Kombinasi <i>TextRank</i> dengan <i>word embedding</i> (Word2Vec, FastText, BERT) yang diberi bobot menggunakan TF-IDF	Dataset Liputan6	Penggunaan <i>word embedding</i> meningkatkan efektivitas <i>TextRank</i> dalam peringkasan teks, dengan model berbasis BERT menunjukkan performa terbaik. Pembobotan TF-IDF meningkatkan akurasi, terutama pada Word2Vec dan FastText. Pendekatan yang diusulkan meningkatkan skor ROUGE-1 hingga 17,33% dan ROUGE-2 hingga 30,01% dibandingkan <i>TextRank</i> konvensional.
Chen & Song (2021)	<i>News Text Summarization Method based on BART-TextRank Model</i>	Kombinasi <i>TextRank</i> dan model BART	Dataset CNN atau Daily Mail	Model BART- <i>TextRank</i> meningkatkan performa dibanding BART saja. Evaluasi menggunakan metrik ROUGE menunjukkan bahwa model ini meningkatkan skor ROUGE-1 sebesar 1,5%, ROUGE-2 sebesar 0,5%, dan ROUGE-L sebesar 1,3%.
Fakhrezi <i>et al.</i> (2021)	<i>Implementation of Automatic Text Summarization with TextRank Method in the</i>	<i>TextRank</i>	Dataset Tafsir al-Misbah	Evaluasi menunjukkan rata-rata nilai <i>F-Score</i> sebesar 0,6173.

	<i>Development of Al-Qur'an Vocabulary Encyclopedia</i>			
--	---	--	--	--

Penelitian oleh Zaware *et al.* (2021) mengusulkan sistem peringkasan teks otomatis yang menggabungkan algoritma TF-IDF dan *TextRank*. Sistem ini berfokus pada teknik peringkasan ekstraktif dengan tahapan utama meliputi perhitungan matriks TF-IDF untuk menentukan bobot kata dalam dokumen dan pembangunan *graf* menggunakan matriks kesamaan kosinus (*cosine similarity*). *Graf* yang terbentuk kemudian diberi skor menggunakan algoritma *TextRank* untuk menentukan peringkat kalimat yang relevan. Dataset yang digunakan adalah BBC *News Summary Dataset* yang terdiri dari 4450 dokumen. Evaluasi menggunakan metrik ROUGE menunjukkan bahwa kombinasi TF-IDF dan *TextRank* menghasilkan ringkasan yang lebih baik dibandingkan hanya menggunakan TF-IDF.

Penelitian oleh Yulianti *et al.* (2023) memperkenalkan peningkatan pada algoritma *TextRank* dengan menggunakan *embedding* kata berbobot (*weighted word embedding*) yang dipadukan dengan pembobotan TF-IDF untuk peringkasan teks berita Indonesia. Dataset yang digunakan adalah berita dari portal Liputan6 dengan 10.972 dokumen. Hasil penelitian menunjukkan bahwa *TextRank* dengan *Word2Vec*+TF-IDF mencapai ROUGE-1 sebesar 0.4082, ROUGE-2 sebesar 0.3041, dan ROUGE-L sebesar 0.3926, sementara *FastText*+TF-IDF menghasilkan ROUGE-1 sebesar 0.4042 dan ROUGE-2 sebesar 0.2982. Sistem berbasis IndoBERT memberikan performa serupa dengan ROUGE-1 hingga 0.4044 dan

ROUGE-2 sebesar 0.2982. Secara keseluruhan, penggunaan TF-IDF pada *embedding* kata meningkatkan performa *TextRank* hingga 17,33% pada ROUGE-1 dan 30,01% pada ROUGE-2, dengan *Word2Vec* menunjukkan peningkatan paling signifikan.

Penelitian oleh Chen & Song (2021) mengembangkan metode peringkasan teks berita berbasis model BART-*TextRank*. Penelitian ini menggabungkan keunggulan algoritma ekstraktif *TextRank* dan kemampuan generatif BART (*Bidirectional and Auto-Regressive Transformer*). Dataset yang digunakan adalah CNN atau *Daily Mail* yang terdiri dari 13.760 artikel berita. Hasil penelitian menunjukkan bahwa model BART-*TextRank* meningkatkan skor ROUGE secara signifikan dibandingkan dengan model BART tunggal, dengan peningkatan 1,5% pada ROUGE-1, 0,5% pada ROUGE-2, dan 1,3% pada ROUGE-L.

Penelitian oleh Fakhrezi *et al.* (2021) menerapkan *TextRank* untuk pengembangan Ensiklopedia Kosa Kata Al-Qur'an. Penelitian ini bertujuan untuk merangkum interpretasi kata dalam Al-Qur'an dengan dataset yang terdiri dari 500 dokumen dari Tafsir Al-Misbah karya Quraish Shihab. Proses melibatkan *preprocessing* seperti tokenisasi, penghapusan *stopword*, *stemming*, representasi vektor *Word2Vec*, pembangunan matriks kesamaan kosinus (*cosine similarity*), dan algoritma *PageRank* untuk menentukan peringkat kalimat yang relevan. Evaluasi menunjukkan bahwa metode ini menghasilkan rata-rata nilai *F1-Score* sebesar 0,6173, dengan hasil ringkasan mendekati kualitas ringkasan manual.

2.3 Pembobotan

Pembobotan (*weighting*) adalah teknik untuk memberikan nilai atau bobot pada setiap elemen yang terlibat dalam suatu proses untuk menentukan tingkat kepentingannya. Pembobotan dapat dilakukan melalui pendekatan objektif maupun subjektif. Pendekatan objektif menggunakan metode statistik, sementara pendekatan subjektif mempertimbangkan pemahaman dan pengalaman dalam konteks tertentu.

2.3.1 Term Frequency (TF)

Term Frequency (TF) adalah metode pembobotan yang menghitung seberapa sering suatu *term* muncul dalam sebuah dokumen. Semakin sering *term* muncul, semakin tinggi bobotnya. Bobot TF dari *term i* pada dokumen *d* dirumuskan pada Persamaan 2.1.

$$W(i, d) = TF(i, d) \quad (2.1)$$

Dimana $TF(i, d)$ adalah jumlah kemunculan *term i* dalam dokumen *d*.

2.3.2 Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) mengukur pentingnya suatu *term* berdasarkan jumlah dokumen tempat *term* tersebut muncul. Semakin jarang *term* muncul dalam kumpulan dokumen, semakin tinggi bobotnya. Nilai IDF dituliskan dalam persamaan 2.2.

$$IDF(t) = \log\left(\frac{D}{df_i}\right) \quad (2.2)$$

Dimana nilai *D* adalah total jumlah teks atau dokumen dan $df(i)$ adalah jumlah dokumen yang mengandung *term i* (Wedel *et al.*, 2022).

2.3.3 *Term Frequency - Invers Document Frequency (TF-IDF)*

TF-IDF merupakan gabungan dari TF dan IDF. Metode ini menghitung bobot suatu *term* dengan memperhatikan frekuensi kemunculan kata di satu dokumen (TF) dan distribusinya di seluruh dokumen (IDF). Rumus TF-IDF sebagai berikut.

$$TF-IDF(i,d) = TF(i,d) \times IDF(i) \quad (2.3)$$

Metode ini terbukti memberikan performa yang lebih baik dibandingkan penggunaan TF atau IDF secara terpisah (Septiani & Isabela, 2022).

2.3.4 *Weighted Inverse Document Frequency (WIDF)*

WIDF (*Weighted Inverse Document Frequency*) merupakan pengembangan dari metode IDF yang dikembangkan oleh Tokunaga dan Iwayama. Keunggulan dari WIDF adalah kemampuannya untuk membedakan masing-masing dokumen. Pada IDF, semua *term* diperlakukan sama di seluruh dokumen. Kemunculan *term* pada sebuah dokumen tidak dibedakan dengan dokumen lainnya. Dengan kata lain, WIDF merupakan bentuk normalisasi dari *Term Frequency* (Armianti *et al.*, 2019). Nilai WIDF dirumuskan pada Persamaan 2.4.

$$WIDF(d,t) = \frac{TF(d,t)}{\sum_i TF(i,t)} \quad (2.4)$$

Dimana $TF(d,t)$ adalah *term frequency* dari *term t* di dalam *text d* dan i menyatakan jumlah *text*. WIDF dari *term t* menjumlahkan semua *term frequency* dari semua kumpulan *text* (Pratama *et al.*, 2020). Beberapa penelitian yang mengimplementasikan metode WIDF dapat dilihat pada Tabel 2.3.

Tabel 2.3 Penelitian Terkait WIDF

Sumber	Judul	Metode	Data	Hasil
Deolika <i>et al.</i> (2019)	Analisis Pembobotan Kata Pada Klasifikasi Text Mining	Pembobotan kata menggunakan TF-IDF, TF.RF, dan WIDF dengan klasifikasi <i>Naïve Bayes</i>	Dokumen dari layanan pengaduan masyarakat LAPOR!	Hasil penelitian menunjukkan bahwa pembobotan TF.RF memiliki kinerja terbaik dalam klasifikasi teks dibandingkan TF-IDF dan WIDF. Model <i>Naïve Bayes</i> yang digunakan mencapai akurasi sebesar 98,67%, <i>precision</i> 93,81%, dan <i>recall</i> 96,67%, sehingga efektif dalam klasifikasi teks pada <i>text mining</i> .
Armianti <i>et al.</i> (2019)	Klasifikasi Emosi Lagu Berdasarkan Lirik pada Teks Berbahasa Indonesia Menggunakan <i>K-Nearest Neighbor</i> dengan Pembobotan WIDF	<i>K-Nearest Neighbor</i> (KNN) dengan pembobotan <i>Weighted Inverse Document Frequency</i> (WIDF)	Data lirik lagu berbahasa Indonesia	Metode KNN dengan pembobotan WIDF menghasilkan akurasi sebesar 66%, dengan nilai rata-rata <i>precision</i> 0,49 dan <i>recall</i> 0,53.
Pratama <i>et al.</i> (2020)	<i>Weighted inverse document frequency and vector space model for hadith search engine</i>	Metode <i>Weighted Inverse Document Frequency</i> (WIDF) dan <i>Vector Space Model</i> (VSM)	Data dari Kitab <i>Bulughul Maram</i>	Model pencarian berbasis WIDF dan VSM memiliki tingkat <i>recall</i> sebesar 96%, menunjukkan efektivitas dalam menemukan hadis sesuai kata kunci. Namun, nilai <i>precision</i> yang rendah (35,46%) menunjukkan bahwa banyak hasil pencarian yang kurang relevan. Hal ini disebabkan oleh penggunaan representasi teks <i>bag-of-words</i> (1-gram) yang kurang mampu mempertahankan konteks makna dalam teks.
(Susandi & Sholahudin, 2017)	Pemanfaatan Vector Space Model pada Penerapan Algoritma Nazief Adriani, KNN dan Fungsi <i>Similarity Cosine</i> untuk Pembobotan	<i>Vector Space Model</i> (VSM) dengan pembobotan TF-IDF dan WIDF, serta algoritma Nazief Adriani.	9 dokumen teks berbahasa Indonesia dalam kategori Hukum, Kesehatan, dan Pendidikan	Sistem yang dikembangkan mampu mengklasifikasikan dokumen teks berbahasa Indonesia berdasarkan kategori yang ditentukan. Evaluasi menunjukkan bahwa pembobotan TF-IDF memiliki akurasi lebih tinggi dibandingkan WIDF, dengan nilai <i>precision</i> dan <i>recall</i> masing-masing sebesar 70,7%

	IDF dan WIDF pada Prototipe Sistem Klasifikasi Teks Bahasa Indonesia			untuk TF-IDF dan 70,6% untuk WIDF
--	--	--	--	-----------------------------------

Penelitian oleh Pratama *et al.* (2020) mengembangkan sebuah mesin pencari hadith menggunakan metode WIDF dan *Vector Space Model* (VSM). Pendekatan ini mempertimbangkan frekuensi kata dalam koleksi dokumen untuk memberikan bobot yang lebih spesifik. Dalam penelitian tersebut, VSM digunakan untuk mengukur kemiripan antara dokumen dan kueri berdasarkan representasi vektor. Sistem diuji menggunakan 380 teks hadith dari kitab *Bulughul Maram*, dan hasilnya menunjukkan nilai *recall* yang tinggi sebesar 96%, meskipun *precision*-nya hanya mencapai 35,46%. Keterbatasan ini dikaitkan dengan representasi teks berbasis *bag-of-words*.

Penelitian oleh Armianti *et al.* (2019) membahas klasifikasi emosi pada lirik lagu berbahasa Indonesia menggunakan metode *K-Nearest Neighbor* (KNN) dengan pembobotan *Weighted Inverse Document Frequency* (WIDF). Sebelum proses klasifikasi dilakukan, data melalui tahap *preprocessing* yang mencakup *tokenizing*, *stopword removal*, dan *stemming* untuk meningkatkan kualitas teks yang akan diklasifikasikan. Penelitian ini menggunakan 108 data lirik lagu, yang terbagi menjadi 90 data latih dan 18 data uji, dengan jumlah data yang seimbang di setiap kategori emosi. Evaluasi dilakukan dengan menguji beberapa nilai *K* secara acak sebanyak enam kali untuk menentukan performa terbaik. Hasil penelitian

menunjukkan bahwa metode KNN dengan pembobotan WIDF menghasilkan akurasi 66%, dengan nilai rata-rata *precision* sebesar 0,49 dan *recall* sebesar 0,53.

Penelitian lain oleh Deolika *et al.* (2019) membandingkan tiga metode pembobotan kata dalam *text mining*, yaitu TF-IDF, TF-RF, dan WIDF. Hasil penelitian menunjukkan bahwa TF-RF memberikan hasil terbaik dengan akurasi 98,67%, *precision* 93,81%, dan *recall* 96,67%. Sementara itu, WIDF memberikan keunggulan dalam memperhitungkan distribusi frekuensi kata di seluruh dokumen untuk menghasilkan pembobotan yang lebih akurat dibandingkan TF-IDF, meskipun performanya tidak sebaik TF-RF.

Penelitian selanjutnya oleh (Susandi & Sholahudin, 2017) membahas penggunaan WIDF dalam klasifikasi dokumen berbahasa Indonesia. Penelitian ini menggunakan *Vector Space Model* (VSM) dengan pembobotan TF-IDF dan WIDF serta algoritma *K-Nearest Neighbor* (KNN) untuk menghitung kesamaan antar dokumen. Pengujian dilakukan pada 9 dokumen dalam kategori hukum, kesehatan, dan pendidikan. Hasil evaluasi menunjukkan bahwa TF-IDF dan WIDF memiliki performa yang hampir setara, dengan *precision* dan *recall* masing-masing sebesar 70,7% dan 70,6%. WIDF menunjukkan potensi untuk menghasilkan pembobotan yang lebih spesifik meskipun tidak selalu unggul dibandingkan metode lain.

2.4 Latent Semantic Analysis (LSA)

Latent Semantic Analysis merupakan algoritma yang digunakan untuk menganalisis hubungan antara kalimat atau *frasa* dengan sekumpulan dokumen (Gotami *et al.*, 2018). LSA bekerja dengan mengidentifikasi pola hubungan

semantik antar kata dalam teks, sehingga dapat menggambarkan keterkaitan makna antar kata maupun kalimat dalam suatu dokumen.

Konsep LSA direalisasikan melalui dua komponen utama, yaitu pembentukan matriks dan penerapan *Singular Value Decomposition* (SVD). Proses dimulai dengan membentuk matriks *term-document* yang diperoleh melalui teknik pembobotan *Weighted Inverse Document Frequency* (WIDF). Matriks ini merepresentasikan hubungan antar kata dalam dokumen dalam bentuk numerik.

Selanjutnya, SVD digunakan untuk mengolah komponen-komponen dalam matriks tersebut guna menemukan pola keterkaitan atau kesamaan makna antara kata dan kalimat. SVD memiliki kemampuan dalam mereduksi *noise* atau gangguan pada data, sehingga dapat meningkatkan akurasi hasil pada proses peringkasan (Mandar & Gunawan, 2018). Metode LSA pertama kali diperkenalkan oleh Scott Deerwester, Thomas Landauer, dan Susan Dumais pada tahun 1998, dan telah banyak digunakan dalam berbagai aplikasi seperti pencarian informasi, analisis semantik, serta peringkasan teks (Sianturi *et al.*, 2023). Beberapa penelitian yang menerapkan LSA dalam pemrosesan teks dirangkum pada Tabel 2.4.

Tabel 2.4 Penelitian Terkait *Latent Semantic Analysis*

Sumber	Judul	Metode	Data	Hasil
Mujilawati (2023)	Kombinasi Algoritma Data Reduksi untuk Optimalisasi Dokumen Cluster	<i>K-Means</i> untuk <i>clustering</i> , TF-IDF dan <i>Hashing</i> untuk ekstraksi fitur, serta LSA untuk reduksi dimensi	Dokumen abstrak skripsi mahasiswa	Kombinasi <i>LSA-Hashing</i> menghasilkan <i>F-Measure</i> tertinggi (0,393), sedangkan LSA-TF-IDF memiliki waktu eksekusi paling efisien. Penambahan LSA meningkatkan homogenitas dan efektivitas hasil <i>clustering</i> .
Gotami <i>et al.</i> (2018)	Peringkasan Teks Otomatis Secara Ekstraktif pada	Metode LSA berbasis TF-IDF	Artikel berita kesehatan dari kompas.com	Peringkasan dengan LSA menunjukkan performa terbaik pada <i>compression rate</i> 40% dengan nilai <i>precision</i> 0.75,

	Artikel Berita Kesehatan Berbahasa Indonesia dengan Menggunakan Metode <i>Latent Semantic Analysis</i>			<i>recall</i> 0.6, <i>f-measure</i> 0.667, dan akurasi 0.727.
Rozi <i>et al.</i> (2021)	Otomatisasi Peringkasan Teks pada Dokumen Hukum Menggunakan Metode <i>Latent Semantic Analysis</i>	<i>Latent Semantic Analysis</i> (LSA) dan TF-IDF untuk pembobotan kata	Dokumen nota pembelaan hukum	Hasil terbaik diperoleh pada <i>compression rate</i> 25%, dengan <i>precision</i> 51%, <i>recall</i> 79%, <i>f-measure</i> 61%, dan <i>accuracy</i> 75%.

Penelitian yang dilakukan oleh Mujilahwati (2023) mengkombinasikan algoritma *K-Means* dengan LSA, *Term Frequency-Inverse Document Frequency* (TF-IDF), dan *Hashing* untuk mengoptimalkan hasil *clustering* data abstrak skripsi. Studi ini menggunakan 229 data abstrak skripsi Teknik Informatika yang dikelompokkan ke dalam 4 *cluster*. TF-IDF digunakan untuk memberikan bobot berbasis frekuensi *term*, sementara *Hashing* digunakan untuk mengubah teks menjadi vektor. LSA dengan TF-IDF diterapkan untuk mengurangi dimensi data, sedangkan kombinasi LSA dengan *Hashing* dimanfaatkan untuk meningkatkan efisiensi evaluasi *clustering*. Hasil penelitian menunjukkan bahwa LSA dengan TF-IDF memberikan waktu eksekusi yang lebih efisien, sedangkan LSA dengan *Hashing* menghasilkan nilai *F-measure* yang lebih tinggi.

Penelitian yang dilakukan oleh Gotami *et al.* (2018) menggunakan LSA untuk pengembangan sistem peringkasan teks otomatis pada artikel berita kesehatan berbahasa Indonesia. Proses penelitian meliputi *preprocessing* data

seperti *parsing*, tokenisasi, *filtering*, *stemming*, dan pembobotan *term* dengan TF-IDF, serta penerapan LSA untuk mendekomposisi matriks menjadi tiga komponen utama. Metode *Cross-LSA* digunakan untuk menyusun urutan ringkasan berdasarkan nilai rata-rata panjang vektor matriks. Hasil evaluasi menunjukkan bahwa sistem ini mampu menghasilkan ringkasan yang akurat. Pada *Compression rate* 50%, nilai *precision*, *recall*, *f-measure*, dan akurasi masing-masing mencapai 0.668, 0.743, 0.700, dan 0.690. Sedangkan pada *compression rate* 40%, nilai-nilai tersebut adalah 0.696, 0.605, 0.642, dan 0.663.

Penelitian yang dilakukan oleh Rozi *et al.* (2021) mengembangkan sistem peringkasan otomatis pada dokumen hukum menggunakan metode *Latent Semantic Analysis* (LSA). Proses dimulai dengan *preprocessing* teks, pembobotan kata menggunakan TF-IDF, dan pemilihan kalimat utama melalui metode *Singular Value Decomposition* (SVD). Pengujian dilakukan pada sepuluh dokumen nota pembelaan, dan hasil terbaik diperoleh pada *compression rate* 25% dengan nilai *precision* sebesar 51%, *recall* 79%, *f-measure* 61%, dan akurasi 75%. Hasil ini menunjukkan bahwa metode LSA dapat digunakan secara efektif untuk meringkas dokumen hukum secara otomatis.

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Data Penelitian

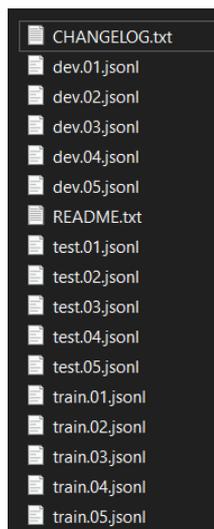
Penelitian ini menggunakan dataset “*Indonesian Text Summarization* (IndoSum)”, yang merupakan dataset *benchmark* untuk peringkasan teks berbahasa Indonesia, khususnya pada domain berita. Dataset ini terdapat minimal 14.261 berita, yang mencakup judul, kategori, *gold label*, sumber berita, alamat sumber berita, dan ringkasan. Ringkasan pada dataset dibuat secara manual oleh dua penutur asli Bahasa Indonesia. Terdapat 6 buah kategori berita, yaitu hiburan/*entertainment*, inspirasi/*inspiration*, olahraga/*sport*, gosip/*showbiz*, berita utama/*headline* dan teknologi/*tech*. Dataset ini juga telah digunakan sebelumnya oleh Kurniawan & Louvan (2019) dalam penelitiannya yang berjudul “INDOSUM: *A New Benchmark Dataset for Indonesian Text Summarization*”. Gambaran bentuk dari dataset IndoSum dapat dilihat pada Gambar 3.1.

category	gold_labels	id	paragraphs	source	source_url	summary	
14257	tajuk utama	[[True], [True], [False], [False, False], [Fal...	1497645345-as-kirimkan-peluncur-rudal-ke-suria...	[[[Jakarta, ,, CNN, Indonesia, -, -, Amerika, ...	crn indonesia	http://www.cnnindonesia.com/internasional/2017...	[[Amerika, Serikat, dilaporkan, telah, mengiri...
14258	olahraga	[[True, True], [True, False], [False], [False...	1495406700-borneo-bersyukur-tahan-persib-di-gbla	[[[Bandung, ,, CNN, Indonesia, -, -, Borneo, F...	crn indonesia	http://www.cnnindonesia.com/olahraga/201705202...	[[Borneo, FC, menahan,imbang, Persib, Bandung...
14259	tajuk utama	[[True], [False, False], [True, True, False], ...	1513941815-mantan-dirjen-pertubungan-laut-sege...	[[[JAKARTA, (, Pos, Kota,), -, Komisi, Pember...	poskotanews	http://poskotanews.com/2017/12/21/mantan-dirje...	[[Komisi, Pemberantasan, Korupsi, (, KPK,), s...
14260	olahraga	[[True, True], [True, True], [False, False], [...	1496440800-ranking-ffifa-indonesia-naik-dua-pe...	[[[Merdeka.com, -, Sebuah, kabar, gembira, dat...	merdeka	https://www.merdeka.com/sepakbola/ranking-ffif...	[[Kabar, gembira, datang, bagi, sepakbola, ...
14261	teknologi	[[True, False], [False, False, False, F...	1505421900-apple-tv-4k-usung-resolusi-super-be...	[[[Ada, satu, perangkat, menarik, lainnya, yan...	dailysocial.id	https://dailysocial.id/post/apple-tv-4k-hdr	[[Ada, satu, perangkat, menarik, lainnya, yang...

Gambar 3.1 Bentuk Indosum dataset

Dataset IndoSum memiliki lima bentuk data dan terbagi menjadi tiga subset utama, yaitu *training subset*, *test subset*, dan *development subset*. Setiap bentuk data berisi 14.261 berita dengan tujuh fitur. Dalam penelitian ini, hanya dua fitur utama yang digunakan, yaitu “*paragraphs*” dan “*summary*”. Fitur “*paragraphs*” berisi

daftar paragraf dari artikel asli, sementara fitur “*summary*” berisi ringkasan manual dalam bentuk daftar kalimat. Adapun rata-rata panjang ringkasan manual pada dataset ini adalah 3,48 kalimat per artikel. Struktur file dataset IndoSum dapat dilihat pada Gambar 3.2.

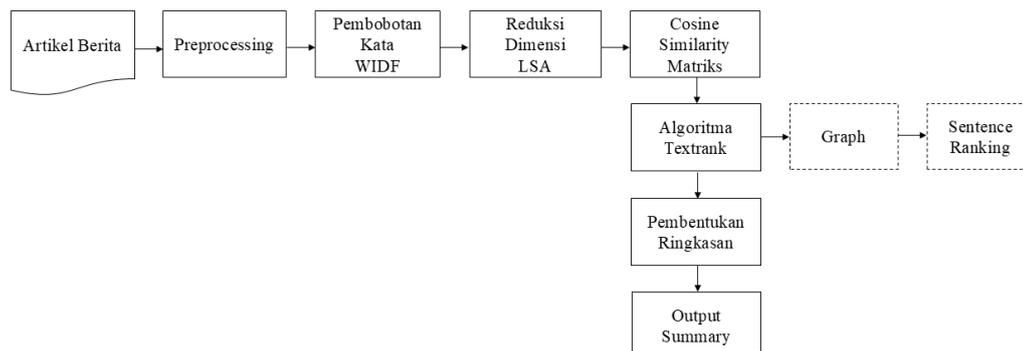


Gambar 3.2 Struktur File IndoSum Dataset

3.2 Desain Sistem

Desain sistem dalam penelitian ini terdiri dari beberapa tahapan. Pertama, dilakukan *preprocessing*. Selanjutnya, dilakukan pembobotan kata menggunakan metode WIDF. Setelah pembobotan, tahap berikutnya adalah reduksi dimensi menggunakan LSA, yang bertujuan untuk menangkap hubungan semantik antar kalimat serta mengurangi kompleksitas data. Hasil dari LSA digunakan untuk membangun matriks kesamaan antar kalimat menggunakan *cosine similarity*. Matriks kesamaan ini kemudian digunakan untuk membangun *graf* hubungan antar kalimat. Selanjutnya, algoritma *TextRank* diterapkan pada *graf* tersebut untuk menghitung skor relevansi setiap kalimat melalui iterasi hingga mencapai konvergensi. Skor relevansi ini mencerminkan pentingnya setiap kalimat dalam

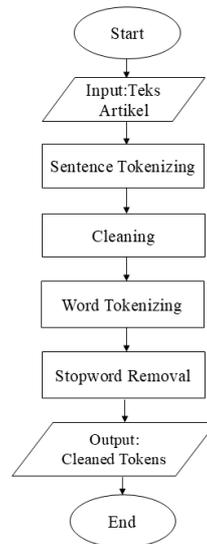
konteks dokumen. Berdasarkan skor tersebut, kalimat-kalimat diurutkan kemudian kalimat dengan skor tertinggi dipilih untuk membentuk ringkasan yang mewakili inti artikel. Ringkasan yang dipilih berdasarkan skor relevansi ini menjadi *output* dari sistem. Gambar 3.3 menunjukkan desain sistem dalam penelitian ini.



Gambar 3.3 Desain Sistem

3.3 *Preprocessing*

Proses *preprocessing* dilakukan untuk mempersiapkan data teks agar dapat dipahami oleh sistem. Tahap *preprocessing* meliputi tokenisasi kalimat, pembersihan teks, dan penghapusan *stopwords* serta tokenisasi kata. Tahap *preprocessing* dapat dilihat pada Gambar 3.4.

Gambar 3.4 Tahap *Preprocessing*

3.3.1 *Sentence Tokenizing*

Pada tahap ini, teks artikel dipecah menjadi unit-unit kalimat untuk mempermudah pengelolaan dan analisis setiap kalimat secara terpisah. Hasil penerapan *sentence tokenizing* dapat dilihat pada Tabel 3.1.

Tabel 3.1 Hasil *sentence tokenizing*

Teks berita	
	<p>Tanjungbalai , Sumut (ANTARA News) - Polres Tanjungbalai menggagalkan upaya perdagangan orang berkedok pengiriman tenaga kerja Indonesia / TKI ilegal yang akan diperkerjakan sebagai pelayan restoran dan buruh kebun di Malaysia . Kapolres Tanjungbalai AKBP Tri Setyadi Artono di Tanjungbalai , Selasa , mengatakan , personel polisi di Kesatuan Pelaksana Pengamanan Pelabuhan (KP3) Teluk Nibung curiga terhadap seorang laki-laki yang diduga agen TKI ilegal karena selalu mendampingi empat orang calon penumpang kapal cepat tujuan Malaysia . Kemudian petugas menghampiri target beserta empat orang calon penumpang tersebut dan memboyongnya ke Polsek Teluk Nibung untuk dilakukan penyelidikan lebih lanjut . Hasil pemeriksaan terhadap mereka , laki-laki yang dicurigai sebagai agen TKI ilegal tersebut bernama Adian Supriadi warga Jalan Lugu Lingkungan III Sendang Rejo , Kabupaten Langkat . " Saat petugas Polsek Teluk Nibung menginterogasi keempat penumpang , terungkap bahwa mereka dijanjikan akan diperkerjakan sebagai TKI di Malaysia oleh Adian Supriadi , " ujar Tri Setyadi . Menurut Kapolres , sesuai pengakuan para calon TKI ilegal itu mereka dijanjikan diperkerjakan di sektor restoran dan perkebunan di Malaysia oleh tersangka (Adian Supriadi) dengan syarat mereka harus membayar biaya administrasi kepada agen tersebut , seperti pengurusan paspor dan tiket pemberangkatan sebesar Rp 3,5 hingga 4 juta sesuai dengan jenis pekerjaan . Karena desakan ekonomi dan tergiur iming-iming dari tersangka , para calon TKI ini pun menyanggupi dan membayar biaya yang diminta tersangka . Sementara itu , kepada petugas tersangka mengakui bahwa dalam</p>

Lanjutan Tabel 3.1

Teks berita	
	merekrut calon TKI ilegal ini dirinya tidak sendirian , tetapi berkerja sama dengan seseorang wanita bernama Evi (agen) yang berada di Malaysia . Agen di Malaysia itu menjanjikan tersangka upah sebesar RM1500 atau sekitar Rp 4,5 juta per orang apabila berhasil mengirim calon TKI tersebut . " Adian Supriadi dinyatakan sebagai tersangka human trafficking atau penjualan orang dan dilakukan penahanan untuk menyidikan lebih lanjut , " kata Kapolres . Berdasarkan catatan , empat orang calon TKI ilegal dan hampir menjadi korban human trafficking , yaitu Azmain (24) , Ahmadsyah (28) dan Bambang Syaputra (24) ketiganya warga Jalan Abadi Desa Pertumbukan Kabupaten Langkat , serta Idris (44) warga Percut Sei Tuan Kabupaten Deli Serdang . Polisi juga menyita empat buah pasport dan barang lainnya milik calon TKI tersebut dan selanjutnya akan berkoordinasi dengan pihak Imigrasi .
Setelah proses <i>sentence tokenizing</i>	
D1	'Tanjungbalai , Sumut (ANTARA News) - Polres Tanjungbalai menggagalkan upaya perdagangan orang berkedok pengiriman tenaga kerja Indonesia / TKI ilegal yang akan diperkerjakan sebagai pelayan restoran dan buruh kebun di Malaysia .'
D2	'Kapolres Tanjungbalai AKBP Tri Setyadi Artono di Tanjungbalai , Selasa , mengatakan , personel polisi di Kesatuan Pelaksana Pengamanan Pelabuhan (KP3) Teluk Nibung curiga terhadap seorang laki-laki yang diduga agen TKI ilegal karena selalu mendampingi empat orang calon penumpang kapal cepat tujuan Malaysia .'
D3	'Kemudian petugas menghampiri target beserta empat orang calon penumpang tersebut dan memboyongnya ke Polsek Teluk Nibung untuk dilakukan penyelidikan lebih lanjut .'
...	...
D12	'Polisi juga menyita empat buah pasport dan barang lainnya milik calon TKI tersebut dan selanjutnya akan berkoordinasi dengan pihak Imigrasi .'

3.3.2 Text Cleaning

Tahap *text cleaning* bertujuan untuk membersihkan teks dari karakter-karakter khusus, simbol, atau elemen lain yang tidak relevan untuk analisis. Selain itu, keseluruhan teks juga diubah menjadi huruf kecil agar tidak ada redundansi dalam pembedaan huruf besar dan kecil. Hasil dari proses ini ditampilkan pada Tabel 3.2.

Tabel 3.2 Hasil *text cleaning*

Setelah proses <i>text cleaning</i>	
D1	'tanjungbalai sumut antara news polres tanjungbalai menggagalkan upaya perdagangan orang berkedok pengiriman tenaga kerja indonesia tki ilegal yang akan diperkerjakan sebagai pelayan restoran dan buruh kebun di malaysia '
D2	'kapolres tanjungbalai akbp tri setyadi artono di tanjungbalai selasa mengatakan personel polisi di kesatuan pelaksana pengamanan pelabuhan kp3 teluk nibung

Lanjutan Tabel 3.2

Setelah proses <i>text cleaning</i>	
	curiga terhadap seorang lakilaki yang diduga agen tki ilegal karena selalu mendampingi empat orang calon penumpang kapal cepat tujuan malaysia '
D3	'kemudian petugas menghampiri target beserta empat orang calon penumpang tersebut dan memboyongnya ke polsek teluk nibung untuk dilakukan penyelidikan lebih lanjut '
...	...
D12	'polisi juga menyita empat buah pasport dan barang lainnya milik calon tki tersebut dan selanjutnya akan berkoordinasi dengan pihak imigrasi '

3.3.3 *Stopwords Removal dan Word Tokenizing*

Tahap ini melibatkan penghapusan kata-kata umum (*stopwords*) yang tidak memiliki nilai analisis dan pemecahan setiap kalimat menjadi kata-kata (*tokens*). Hasil dari tahap penghapusan *stopwords* dan tokenisasi kata dapat dilihat pada Tabel 3.3.

Tabel 3.3 Hasil *stopwords removal dan word tokenizing*

Setelah proses <i>stopwords removal dan word tokenizing</i>	
D1	'tanjungbalai', 'sumut', 'news', 'polres', 'tanjungbalai', 'menggagalkan', 'upaya', 'perdagangan', 'orang', 'berkedok', 'pengiriman', 'tenaga', 'kerja', 'indonesia', 'tki', 'ilegal', 'diperkerjakan', 'pelayan', 'restoran', 'buruh', 'kebun', 'malaysia'
D2	'kapolres', 'tanjungbalai', 'akbp', 'tri', 'setyadi', 'artono', 'tanjungbalai', 'selasa', 'personel', 'polisi', 'kesatuan', 'pelaksana', 'pengamanan', 'pelabuhan', 'kp3', 'teluk', 'nibung', 'curiga', 'lakilaki', 'diduga', 'agen', 'tki', 'ilegal', 'mendampingi', 'orang', 'calon', 'penumpang', 'kapal', 'cepat', 'tujuan', 'malaysia'
D3	'petugas', 'menghampiri', 'target', 'beserta', 'orang', 'calon', 'penumpang', 'memboyongnya', 'polsek', 'teluk', 'nibung', 'penyelidikan'
...	...
D12	'polisi', 'menyita', 'buah', 'pasport', 'barang', 'milik', 'calon', 'tki', 'berkoordinasi', 'imigrasi'

3.4 **Pembobotan Kata**

Pada tahap ini, *term* yang diperoleh dari hasil *preprocessing* akan dihitung bobotnya. Pembobotan dilakukan dengan menggunakan *Weighted Inverse Document Frequency* (WIDF), yaitu membagi frekuensi kemunculan *term t* dalam dokumen *d* dengan total frekuensi kemunculan *term t* dalam semua dokumen. Tabel

3.4 adalah tabel frekuensi kemunculan *term* (TF) pada tiap dokumen, total kemunculan *term* di semua dokumen beserta hasil pembobotan WIDF.

Tabel 3.4 Pembobotan kata dengan WIDF

No	Term	TF(d,t)					TF(i,t)					TF(d,t) / TF(i,t)				
		d1	d2	...	d11	d12	d1	d2	...	d11	d12	d1	d2	...	d11	d12
1	24	0	0	...	2	0	2	0	0	...	1	0				
2	28	0	0	...	1	0	1	0	0	...	1	0				
3	35	0	0	...	0	0	1	0	0	...	0	0				
4	4	0	0	...	0	0	1	0	0	...	0	0				
5	44	0	0	...	1	0	1	0	0	...	1	0				
6	45	0	0	...	0	0	1	0	0	...	0	0				
7	abadi	0	0	...	1	0	1	0	0	...	1	0				
8	adian	0	0	...	0	0	4	0	0	...	0	0				
9	administrasi	0	0	...	0	0	1	0	0	...	0	0				
10	agen	0	1	...	0	0	5	0	0.2	...	0	0				
...				
132	tiki	1	1	...	1	1	10	0.1	0.1	...	0.1	0.1				
133	trafficking	0	0	...	1	0	2	0	0	...	0.5	0				
134	tri	0	1	...	0	0	2	0	0.5	...	0	0				
135	tuan	0	0	...	1	0	1	0	0	...	1	0				
136	tujuan	0	1	...	0	0	1	0	1	...	0	0				
137	upah	0	0	...	0	0	1	0	0	...	0	0				
138	upaya	1	0	...	0	0	1	1	0	...	0	0				
139	wanita	0	0	...	0	0	1	0	0	...	0	0				
140	warga	0	0	...	2	0	3	0	0	...	0.667	0				

Keterangan kolom tabel :

Term : kata unik hasil *preprocessing*.

$TF(d, t)$: frekuensi kemunculan kata *t* dalam dokumen *d*.

$TF(i, t)$: total frekuensi kemunculan kata *t* di semua dokumen.

Berikut adalah perhitungan pembobotan WIDF berdasarkan *term* pada tiap dokumen dan jumlah *term* pada semua dokumen, sebagai contoh diambil data dari *term* 1 dokumen 1, *term* 7 dokumen 11 dan *term* 10 dokumen 2.

$$WIDF(d, t) = \frac{TF(d, t)}{\sum_i TF(i, t)}$$

$$WIDF(11,7) = \frac{1}{1} = 1$$

$$WIDF(2,10) = \frac{1}{5} = 0.2$$

$$WIDF(1,1) = \frac{0}{2} = 0$$

Setelah nilai WIDF dihitung, hasil tersebut direpresentasikan dalam bentuk matriks WIDF. Setiap baris mewakili sebuah kata yang unik, sedangkan setiap kolom mewakili dokumen tempat kata tersebut muncul (Suhartono, 2015; Pratama *et al.*, 2020). Visualisasi hasil pembobotan ini ditampilkan pada Gambar 3.5 dalam bentuk matriks *NumPy array*.

```
[[0.  0.  0.  ... 0.  1.  0. ]
 [0.  0.  0.  ... 0.  1.  0. ]
 [0.  0.  0.  ... 0.  0.  0. ]
 ...
 [1.  0.  0.  ... 0.  0.  0. ]
 [0.  0.  0.  ... 0.  0.  0. ]
 [0.  0.  0.  ... 0.  0.667 0. ]]
```

Gambar 3.5 Matriks WIDF

3.5 *Latent Semantic Analysis*

Setelah diperoleh matriks bobot kata melalui pembobotan WIDF, tahap berikutnya adalah menerapkan *Latent Semantic Analysis* (LSA) untuk mereduksi dimensi dari matriks tersebut. Matriks WIDF yang dihasilkan pada tahap sebelumnya berbentuk *term-document matrix*, di mana setiap baris mewakili *term* (kata), dan setiap kolom mewakili dokumen atau kalimat. Namun, agar dapat diterapkan pada LSA, matriks ini perlu diubah menjadi *document-term matrix* di mana setiap baris merepresentasikan dokumen atau kalimat, dan setiap kolom mewakili *term*. Oleh karena itu, matriks WIDF harus ditranspos terlebih dahulu untuk memudahkan proses dekomposisi menggunakan LSA, seperti yang ditunjukkan pada Gambar 3.6.

$$\begin{bmatrix}
 [\theta. & \theta. & \theta. & \dots & 1. & \theta. & \theta. &] \\
 [\theta. & \theta. & \theta. & \dots & \theta. & \theta. & \theta. &] \\
 [\theta. & \theta. & \theta. & \dots & \theta. & \theta. & \theta. &] \\
 \dots & & & & & & & \\
 [\theta. & \theta. & \theta. & \dots & \theta. & \theta. & \theta. &] \\
 [1. & 1. & \theta. & \dots & \theta. & \theta. & 0.667 &] \\
 [\theta. & \theta. & \theta. & \dots & \theta. & \theta. & \theta. &]
 \end{bmatrix}$$

Gambar 3.6 Matriks WIDF setelah ditranspos

LSA bekerja dengan teknik dekomposisi matriks yang disebut *Singular Value Decomposition* (SVD). Proses ini memecah matriks A menjadi tiga komponen utama, ditunjukkan pada Persamaan 3.1.

$$A = U\Sigma V^T \quad (3.1)$$

Dimana A adalah matriks input berukuran $m \times n$, mewakili kalimat atau kata dalam dokumen. Matriks U menggambarkan matriks ortogonal $m \times k$. Matriks Σ (*sigma*) merupakan matriks diagonal berukuran $k \times k$ berisi nilai-nilai *singular*. Matriks V adalah matriks ortogonal berukuran $k \times n$ (Gotami *et al.*, 2018).

Berdasarkan penelitian oleh Geetha J K & Deepamala N (2015), langkah pertama dalam penerapan SVD adalah menghitung $A \times A^T$ untuk memperoleh *eigenvector*, yang akan membentuk matriks U . Hasil dari matriks U dapat dilihat pada Gambar 3.7.

```

=== Eigenvectors (U) dari AAT ===
[[ 1.16197050e-02  1.96257378e-01 -1.89211862e-01 -9.61960480e-01
  2.29094206e-03 -7.96959465e-03 -2.32779524e-03 -3.13454787e-05
 -5.24295491e-03 -7.40600889e-03  4.66231376e-03  8.19678628e-04]
 [ 2.39270852e-02  9.11071002e-01  3.83644713e-01  1.11789637e-01
  4.55569395e-02 -4.62043400e-02 -1.59502902e-02 -2.88558335e-02
 -1.52231910e-02 -6.41440290e-02 -5.29975791e-03  5.02991079e-03]
 [ 3.49485365e-03  3.64300349e-02  1.06685707e-02  3.01083992e-03
 -1.04507108e-03 -4.09521247e-01 -1.15978801e-01  3.06003537e-03
  7.94711375e-02  8.62642879e-01  2.33719525e-01 -1.10794182e-01]
 [ 7.87597543e-02  4.71252750e-02 -9.95188245e-03  1.10357491e-02
 -9.91610536e-01 -1.28989073e-02 -2.05540926e-02 -1.68478251e-03
 -2.43050286e-02 -3.27054547e-02  7.27449970e-02 -2.28017601e-02]
 [ 3.38452183e-03  8.37233872e-02 -1.46018222e-02  9.05480096e-03
 -1.96423006e-02  8.91331713e-01  1.37768192e-01 -4.98045359e-03
  2.9577977e-02  4.15707296e-01  4.79206411e-02 -5.32637282e-02]
 [ 1.56597145e-02  3.41750109e-01 -9.00191080e-01  2.47849531e-01
  3.29390101e-02 -3.51337519e-02 -2.51797065e-02 -5.51973998e-03
 -7.12067554e-02 -2.07646305e-02 -1.25094145e-02 -5.04615055e-02]
 [ 2.36441815e-03  2.00793291e-02 -4.99451397e-02  1.56088588e-02
  1.01327151e-03  2.39962203e-02 -4.11281708e-02  1.56111296e-02
  4.89554159e-02  3.79424587e-02  3.11715035e-01  9.45165843e-01]
 [ 5.30072787e-03  1.82464451e-02 -1.08880781e-02 -4.40812807e-04
 -8.18997874e-02 -5.28329339e-02 -1.89754689e-02  1.46021834e-02
  1.05000605e-01  2.46320828e-01 -9.14416985e-01  2.85640576e-01]
 [ 4.77043242e-03  3.30750933e-02 -5.63880941e-02  1.35634572e-02
 -1.13605558e-02  1.12333921e-02 -1.55113776e-02  3.45023870e-02
  9.85681819e-01 -1.14635178e-01  6.22737584e-02 -7.24260614e-02]
 [ 3.15075819e-02  1.82029276e-02 -1.71745687e-02  6.15576516e-03
 -1.60147832e-02 -1.75114518e-01  9.81397873e-01 -1.55645704e-03
  2.23470298e-02  4.57968326e-02  1.82746462e-02  3.66958948e-02]
 [ 9.95874531e-01 -3.46334773e-02  8.91422047e-03  3.42320598e-03
  7.78451037e-02  6.88538436e-03 -2.84136396e-02 -1.60390886e-03
 -2.92871784e-03 -2.19412501e-03 -3.22923332e-03 -1.50438479e-03]
 [ 2.28644024e-03  2.68531431e-02  8.85658971e-03  3.93305925e-03
  1.40498040e-03  3.89596004e-03  3.34762670e-03  9.98722947e-01
 -3.72927103e-02 -2.75990751e-03  5.79253356e-03 -1.64915943e-02]]

```

Gambar 3.7 Matriks U

Selanjutnya, dilakukan perhitungan $A^T \times A$ untuk memperoleh *eigenvector* yang membentuk matriks V . Hasilnya dapat dilihat pada Gambar 3.8.

```

=== Eigenvectors (V) dari ATA ===
[[-2.11984347e-01+0.j  8.36761976e-03+0.j  2.22338229e-03+0.j ...
  2.76991976e-21+0.j  2.20665779e-25+0.j -3.45126646e-30+0.j]]
 [-2.11984347e-01+0.j  8.36761976e-03+0.j  2.22338229e-03+0.j ...
  9.68038491e-09+0.j  7.76767203e-13+0.j -2.76698689e-16+0.j]]
 [-3.33336605e-03+0.j -8.25685201e-02+0.j -2.24525399e-01+0.j ...
  3.32882985e-10+0.j  2.66472723e-14+0.j -3.20266691e-18+0.j]]
 ...
 [-2.47339950e-03+0.j -4.74167552e-02+0.j -4.71931679e-02+0.j ...
  1.11881911e-09+0.j  8.97820506e-14+0.j  1.02995004e-16+0.j]]
 [-1.12832621e-03+0.j -4.40843157e-03+0.j -2.71570129e-03+0.j ...
  8.83715875e-10+0.j  7.09074566e-14+0.j -1.72126836e-16+0.j]]
 [-1.46976304e-01+0.j  1.78976429e-03+0.j  6.56424837e-04+0.j ...
 -6.16158953e-08+0.j -4.94402192e-12+0.j  1.64034191e-15+0.j]]

```

Gambar 3.8 Matriks V

Setelah matriks U dan V diperoleh, nilai *eigen* yang dihasilkan dari $A \times A^T$ dan $A^T \times A$ kemudian diakarkan untuk mendapatkan nilai *singular* (Σ) yang

diurutkan dari nilai terbesar ke terkecil. Hasil nilai *singular* dapat dilihat pada Gambar 3.9.

```
=== Nilai Singular (Sigma) ===
[4.69786824 4.13898794 4.00930624 3.81461034 3.0590847 2.69613055
 2.57345915 2.46476301 2.37128399 2.32591549 2.18521768 2.07294825]
```

Gambar 3.9 Matriks *Singular*

Dari hasil dekomposisi matriks, yang digunakan adalah matriks hasil reduksi dimensi, yaitu U_k, Σ_k, V_k^T . Ketiga matriks ini hanya memuat k komponen utama dengan nilai *singular* terbesar. Matriks U_k terdiri dari k kolom pertama dari matriks U yang menggambarkan keterkaitan antara dokumen dan dimensi laten. Matriks Σ_k adalah matriks diagonal yang berisi k nilai *singular* terbesar, menunjukkan pentingnya masing-masing dimensi. V_k^T adalah k baris pertama dari *transpos* matriks V yang merepresentasikan hubungan antara *term* dan dimensi laten.

Dalam penelitian ini digunakan nilai $k=2$, sehingga setiap kalimat direpresentasikan dalam ruang semantik berdimensi dua. Representasi akhir ini diperoleh dari hasil perkalian U_k dan Σ_k . Hasil perkalian ini membentuk matriks LSA, yang memuat representasi vektor dua dimensi untuk setiap kalimat yang ditampilkan pada Gambar 3.10.

Matriks LSA ini selanjutnya digunakan untuk menghitung kemiripan antar kalimat menggunakan metode *cosine similarity*. Implementasi LSA dilakukan menggunakan fungsi *TruncatedSVD* dari modul *Scikit-learn*, yang secara otomatis melakukan dekomposisi serta memilih komponen utama untuk mereduksi dimensi data.

```

[[ 0.05458784  0.81230692]
 [ 0.11240629  3.77091189]
 [ 0.01641836  0.15078348]
 [ 0.37000295  0.19505094]
 [ 0.01590004  0.34653009]
 [ 0.07356728  1.41449958]
 [ 0.01110772  0.0831081 ]
 [ 0.02490212  0.07552182]
 [ 0.02241086  0.13689741]
 [ 0.14801847  0.0753417 ]
 [ 4.67848733 -0.14334755]
 [ 0.01074139  0.11114484]]

```

Gambar 3.10 Matriks LSA Setelah Reduksi Dimensi

3.6 Similarity Matrix

Setelah dilakukan reduksi dimensi menggunakan *Latent Semantic Analysis* (LSA), tahap selanjutnya adalah menghitung tingkat kemiripan antar kalimat menggunakan metode *cosine similarity*. Nilai *cosine similarity* berkisar antara 0 hingga 1. Semakin mendekati nilai 1, tingkat kemiripan antar kalimat semakin tinggi, sedangkan semakin mendekati nilai 0, kemiripan antar kalimat semakin rendah. Rumus perhitungan *similarity* sebagai berikut (Jayakodi *et al.*, 2016).

$$\text{Similarity}(A, B) = \cos \theta = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (3.2)$$

Keterangan :
A dan B : vektor representasi kalimat (dalam penelitian ini hasil dari LSA)
A · B : hasil perkalian *dot product* dari dua vektor.
 $\|A\|$ dan $\|B\|$: panjang (*magnitudo*) dari vektor A dan B.

Perhitungan *cosine similarity* antara Kalimat 1 (A) dan kalimat 2 (B):

Kalimat 1 (A) = [0.05458784, 0.81230692]

Kalimat 2 (B) = [0.11240629, 3.77091189]

Langkah-langkah perhitungan adalah sebagai berikut:

1. Hitung Dot Product $A \cdot B$:

$$A \cdot B = (0.05458784 \times 0.11240629) + (0.81230692 \times 3.77091189)$$

$$A \cdot B = 0.006134389 + 3.062712405 = 3.068846794$$

2. Hitung Panjang Vektor $\|A\|$ dan $\|B\|$:

$$\begin{aligned} \|A\| &= \sqrt{(0.05458784)^2 + (0.81230692)^2} = \sqrt{0.00297984 + 0.65984636} \\ &= \sqrt{0.6628262} = 0.814164 \end{aligned}$$

$$\begin{aligned} \|B\| &= \sqrt{(0.11240629)^2 + (3.77091189)^2} = \sqrt{0.01263511 + 14.21078022} \\ &= \sqrt{14.22341533} = 3.771319 \end{aligned}$$

3. Hitung *Cosine similarity*:

$$\text{Similarity}(A, B) = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{3.068846794}{0.814164 \times 3.771319} = \frac{3.068846794}{3.070804} = 0.99936$$

Diperoleh nilai *similarity* antara kalimat 1 dan kalimat 2 sebesar **0,99936** yang menunjukkan tingkat kesamaan yang sangat tinggi. Nilai *cosine similarity* untuk semua pasangan kalimat ditampilkan pada Tabel 3.5.

Tabel 3.5 Matriks *cosine similarity*

	Dokumen										
	1	2	3	4	5	...	8	9	10	11	12
1	1.000	0.999	0.999	0.525	1.000	...	0.969	0.995	0.512	0.036	1.000
2	0.999	1.000	0.997	0.492	1.000	...	0.959	0.991	0.480	-0.001	0.998
3	0.999	0.997	1.000	0.559	0.998	...	0.978	0.999	0.547	0.078	1.000
4	0.525	0.492	0.559	1.000	0.506	...	0.720	0.603	1.000	0.870	0.549
5	1.000	1.000	0.998	0.506	1.000	...	0.963	0.993	0.494	0.015	0.999
...
8	0.969	0.959	0.978	0.720	0.963	...	1.000	0.988	0.710	0.284	0.975
9	0.995	0.991	0.999	0.603	0.993	...	0.988	1.000	0.592	0.131	0.998
10	0.512	0.480	0.547	1.000	0.494	...	0.710	0.592	1.000	0.877	0.537
11	0.036	-0.001	0.078	0.870	0.015	...	0.284	0.131	0.877	1.000	0.066
12	1.000	0.998	1.000	0.549	0.999	...	0.975	0.998	0.537	0.066	1.000

3.7 Algoritma *TextRank*

Algoritma *TextRank* adalah metode berbasis *graf* yang digunakan untuk menentukan tingkat kepentingan setiap kalimat dalam sebuah dokumen berdasarkan hubungan semantik antar kalimat. Algoritma ini dimulai dengan

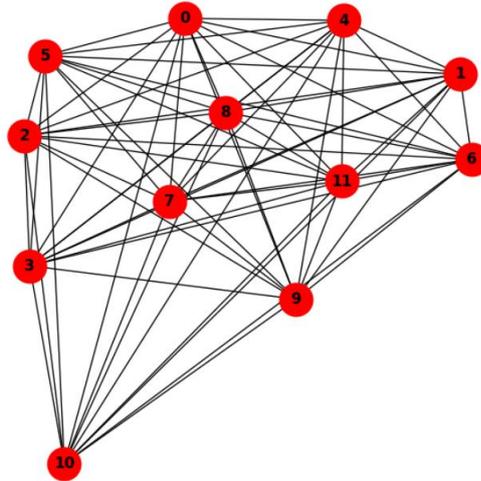
menghitung nilai kemiripan antar kalimat menggunakan *cosine similarity*. Setelah itu, matriks kemiripan tersebut dikonversi menjadi sebuah *graf*. Pada *graf* ini, setiap kalimat direpresentasikan sebagai *node* (*vertex*), sementara hubungan antar kalimat direpresentasikan sebagai *edge* (sisi) dengan bobot sesuai nilai *similarity* antar kalimat. Algoritma *TextRank* memanfaatkan metode *PageRank* untuk menghitung skor kepentingan setiap kalimat dalam dokumen. Skor ini dihitung berdasarkan bobot hubungan antar *node* dalam *graf*, di mana bobot *edge* mewakili nilai *similarity* antar kalimat. Perhitungan skor pada *TextRank* menggunakan persamaan (3.3) (Ramadhan *et al.*, 2020):

$$WS(V_i) = (1 - d) + d \times \sum_{V_j \in in(V_i)} \frac{sim_{ij}}{\sum_{V_k \in out(V_j)} sim_{jk}} WS(V_j) \quad (3.3)$$

$WS(V_i)$ adalah bobot untuk *vertex* i (V_i) atau bobot untuk kalimat i . sim_{ij} adalah nilai kemiripan antara kalimat i dan kalimat j . $In(V_i)$ adalah semua *edge* (sisi) yang mengarah ke *vertex* i (V_i). $Out(V_j)$ adalah semua *edge* yang mengarah keluar dari *vertex* j (V_j). d adalah *damping factor*.

Proses *TextRank* diawali dengan membangun *graf* berdasarkan matriks kemiripan, di mana setiap *node* dalam *graf* merepresentasikan satu kalimat, dan setiap *edge* memiliki bobot sesuai dengan nilai *similarity* antar kalimat. Semua *node* diinisialisasi dengan nilai skor awal yang sama, yaitu 1. *TextRank* kemudian bekerja secara iteratif dengan memperbarui skor setiap *node* berdasarkan kontribusi dari *node-node* yang terhubung hingga skor mencapai konvergensi, yaitu tidak ada lagi perubahan signifikan dalam skor antar iterasi.

Gambar 3.11 menunjukkan representasi *graf* dari dokumen yang terdiri dari 12 kalimat (index kalimat dimulai dari nol). Dalam *graf* ini, *vertex* merepresentasikan kalimat-kalimat yang berada dalam dokumen, sedangkan *edge* menggambarkan *similarity* antar kalimat yang saling terhubung.



Gambar 3.11 Representasi *Graf* berdasarkan *Similarity*

Setelah *graf* terbentuk dan algoritma *TextRank* menghitung skor secara iteratif, diperoleh skor akhir untuk setiap kalimat ditampilkan dalam Tabel 3.6.

Tabel 3.6 Skor *TextRank* Untuk Setiap Kalimat

Index Kalimat	Skor <i>TextRank</i>
0	1.064515578389433
1	1.051707574847881
2	1.077344568836578
3	0.910002479099716
4	1.057314902842321
5	1.059427943213710
6	1.084172254093206
7	1.117656677937309
8	1.091657479119497
9	0.901251075975554
10	0.512385866842285
11	1.073749833756793

3.7 Pembentukan Ringkasan

Proses pembentukan ringkasan dimulai dengan mengurutkan skor akhir dari setiap *vertex* (kalimat) berdasarkan nilai *TextRank* dalam urutan menurun (*descending*). Pengurutan ini menghasilkan peringkat pentingnya kalimat dalam sebuah dokumen. Jumlah kalimat yang akan dipilih sebagai ringkasan ditentukan berdasarkan nilai *compression rate*, yaitu rasio kompresi yang menunjukkan proporsi kalimat yang diambil dibandingkan dengan jumlah total kalimat dalam dokumen asli. *Compression rate* umumnya dinyatakan dalam bentuk persentase, dan menjadi acuan untuk mengatur panjang ringkasan yang dihasilkan secara otomatis. Panjang ringkasan dihitung menggunakan Persamaan 3.4.

$$n_{Ringkasan} = CR \times N \quad (3.4)$$

Keterangan :
 $n_{Ringkasan}$: banyaknya kalimat yang diambil menjadi ringkasan.
 CR : *compression rate* (persentase panjang ringkasan).
 N : banyaknya kalimat pada dokumen asli sebelum diringkas.

Pada Tabel 3.7 menunjukkan hasil pengurutan kalimat berdasarkan skor *TextRank*, di mana 50% dari total kalimat dipilih sebagai ringkasan. Kalimat-kalimat yang terpilih kemudian diurutkan kembali sesuai urutan aslinya dalam dokumen awal, seperti pada Tabel 3.8 (Sihombing *et al.*, 2024).

Tabel 3.7 Hasil Pengurutan Skor *TextRank*

Index Kalimat	Kalimat ke	Skor <i>TextRank</i>
7	8	1.117656677937309
8	9	1.091657479119497
6	7	1.084172254093206
2	3	1.077344568836578
11	12	1.073749833756793
0	1	1.064515578389433
5	6	1.059427943213710
4	5	1.057314902842321
1	2	1.051707574847881

3	4	0.910002479099716
9	10	0.901251075975554
10	11	0.512385866842285

Tabel 3.8 Hasil Ringkasan Dokumen

Index Kalimat	Kalimat ke	Isi kalimat
0	1	Tanjungbalai , Sumut (ANTARA News) - Polres Tanjungbalai menggagalkan upaya perdagangan orang berkedok pengiriman tenaga kerja Indonesia / TKI ilegal yang akan diperkerjakan sebagai pelayan restoran dan buruh kebun di Malaysia .
2	3	Kemudian petugas menghampiri target beserta empat orang calon penumpang tersebut dan memboyongnya ke Polsek Teluk Nibung untuk dilakukan penyelidikan lebih lanjut .
6	7	Karena desakan ekonomi dan tergiur iming-iming dari tersangka , para calon TKI ini pun menyanggupi dan membayar biaya yang diminta tersangka .
7	8	Sementara itu , kepada petugas tersangka mengakui bahwa dalam merekrut calon TKI ilegal ini dirinya tidak sendirian , tetapi berkerja sama dengan seseorang wanita bernama Evi (agen) yang berada di Malaysia .
8	9	Agen di Malaysia itu menjanjikan tersangka upah sebesar RM1500 atau sekitar Rp 4,5 juta per orang apabila berhasil mengirim calon TKI tersebut . "
11	12	Polisi juga menyita empat buah pasport dan barang lainnya milik calon TKI tersebut dan selanjutnya akan berkoordinasi dengan pihak Imigrasi .

BAB IV

HASIL DAN PEMBAHASAN

4.1 Skenario Pengujian

Pengujian dilakukan pada 14.261 artikel berita berbahasa Indonesia yang diambil dari subset *train.01.jsonl* dalam dataset IndoSum. Tahapan pertama dalam pengujian ini adalah *preprocessing* teks, yang bertujuan untuk menghilangkan elemen-elemen yang tidak relevan dan menghasilkan data yang lebih terstruktur serta siap dianalisis. Selanjutnya, setiap kata dalam dokumen diberi bobot menggunakan metode *Weighted Inverse Document Frequency* (WIDF), yang digunakan untuk merepresentasikan pentingnya kata dalam dokumen tersebut. Hasil dari proses ini adalah matriks WIDF, yang kemudian akan diuji dengan reduksi dimensi menggunakan metode *Latent Semantic Analysis* (LSA) dan tanpa reduksi dimensi menggunakan LSA. Pada penelitian ini, metode LSA digunakan untuk membandingkan hasil evaluasi ringkasan yang dihasilkan dengan dan tanpa menggunakan LSA.

Setelah reduksi dimensi, matriks *similarity* dibangun menggunakan metode *cosine similarity*. Matriks ini mengukur tingkat kesamaan antar kalimat dalam dokumen dan berfungsi sebagai dasar untuk penerapan algoritma *TextRank*. Dalam algoritma *TextRank*, setiap kalimat direpresentasikan sebagai simpul (*node*) dalam sebuah *graf*, di mana setiap *edge* antar simpul memiliki bobot yang mencerminkan nilai kesamaan antar kalimat. Proses iteratif dalam *TextRank* menghasilkan skor untuk setiap kalimat, yang kemudian digunakan untuk menentukan kalimat-kalimat

yang akan dimasukkan ke dalam ringkasan. Kalimat dengan skor tertinggi dipilih karena dianggap mewakili inti informasi dari artikel dan diprioritaskan untuk dijadikan ringkasan.

Pengujian sistem ringkasan dilakukan dengan membandingkan hasil ringkasan otomatis dengan ringkasan manual (*ground truth*). Pada pengujian ini, tingkat kompresi ringkasan yang digunakan adalah 30% dan 50% dari panjang teks asli. Evaluasi dilakukan menggunakan metode *Recall-Oriented Understudy of Gisting Evaluation* (ROUGE), yang merupakan metrik utama untuk mengukur kualitas ringkasan teks. Metode ROUGE mengukur tingkat kemiripan antara ringkasan sistem dan ringkasan manual yang dibuat oleh manusia (Zamzam, 2020). Ringkasan manual dianggap sebagai *gold standard* karena umumnya disusun oleh ahli yang memiliki pemahaman mendalam terhadap teks (Gunawan *et al.*, 2019).

Dalam penelitian ini, digunakan ROUGE-1, yang mengukur irisan kata individu (*unigram*) antara ringkasan sistem dan ringkasan manual (Lin, 2004). Pengukuran dilakukan secara kata per kata untuk menilai sejauh mana ringkasan sistem mencakup kata-kata penting dari ringkasan manual (Gulati *et al.*, 2023). Evaluasi menggunakan ROUGE-1 dilakukan dengan tiga parameter yaitu *recall*, *precision*, dan *f1-score*.

4.1.1 Recall

Recall digunakan untuk mengukur sejauh mana kata-kata dalam ringkasan manual berhasil dicakup oleh ringkasan sistem. Nilai *recall* dihitung dengan membagi jumlah kata yang sama (*overlap*) antara ringkasan sistem dan ringkasan

manual dengan jumlah total kata pada ringkasan manual. Rumus untuk menghitung *recall* adalah sebagai berikut:

$$Recall = \frac{\text{Jumlah kata overlap antara ringkasan sistem \& ringkasan manual}}{\text{Banyaknya kata pada ringkasan manual}} \quad (4.1)$$

Recall merepresentasikan kelengkapan (*completeness*) dari ringkasan sistem. Nilai *recall* yang tinggi menunjukkan bahwa sebagian besar informasi penting dalam ringkasan manual berhasil dimasukkan ke dalam ringkasan sistem.

4.1.2 Precision

Precision digunakan untuk mengukur sejauh mana kata-kata dalam ringkasan sistem sesuai atau relevan dengan kata-kata yang ada dalam ringkasan manual. Nilai *precision* dihitung dengan membagi jumlah kata yang sama (*overlap*) antara ringkasan sistem dan ringkasan manual dengan jumlah total kata dalam ringkasan sistem. Rumus untuk menghitung *precision* adalah sebagai berikut:

$$Precision = \frac{\text{Jumlah kata overlap antara ringkasan sistem \& ringkasan manual}}{\text{Banyaknya kata pada ringkasan sistem}} \quad (4.2)$$

Precision merepresentasikan akurasi dari ringkasan sistem. Nilai *precision* yang tinggi menunjukkan bahwa sebagian besar kata dalam ringkasan sistem relevan dengan ringkasan manual.

4.1.3 F1-Score

F1-Score adalah metrik gabungan yang mengukur keseimbangan antara *recall* dan *precision*. Rumus untuk menghitung *f1-score* adalah sebagai berikut:

$$F1-Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (4.3)$$

Nilai *f1-score* yang tinggi menunjukkan bahwa ringkasan sistem memiliki performa yang baik dengan kombinasi *recall* dan *precision*.

4.2 Hasil Uji Coba

Berdasarkan skenario uji coba yang diuraikan pada subbab 4.1, pengujian dilakukan dengan membandingkan hasil ringkasan manual dari referensi dataset dengan hasil ringkasan sistem yang dihasilkan oleh algoritma. Salah satu artikel berita ke lima digunakan dalam pengujian ini, dan tabel 4.1 menampilkan hasil ringkasan manual serta hasil ringkasan sistem berdasarkan tingkat kompresi (*compression rate*) sebesar 50%.

Tabel 4.1 Hasil uji coba

Artikel berita	Merdeka.com - Presiden Joko Widodo (Jokowi) memimpin upacara penurunan bendera di Halaman Istana Merdeka , Jakarta . Usai prosesi penurunan bendera dilakukan , Jokowi kembali bagi-bagi sepeda kepada tamu undangan yang mengenakan pakaian adat terbaik . Berbeda dengan saat upacara pengibaran bendera , bagi-bagi sepeda kali ini diumumkan oleh Wakil Presiden Jusuf Kalla (JK) . Hal berbeda juga terjadi bagi mereka yang menerima . Apabila sepeda diberikan ke pejabat negara mau pun keluarga , kali ini sepeda diberikan kepada masyarakat biasa . " Seperti tadi pagi , sore ini juga panitia membentuk tim penilai untuk menilai siapa yang berbusana tradisional adat - adat daerah yang paling baik sore ini , " kata JK . Kelima penerima sepeda di antaranya , Frans Maksim yang merupakan Kepala Suku Arfak , Papua , Ratna Dewi Budiono yang mengenakan pakaian adat Dayak , Yusak Rumambi yang mengenakan pakaian adat Sulawesi Utara , Teuku Johan Marzuki yang mengenakan pakaian adat Aceh . Terakhir , ada Sumahartarti yang mengenakan pakaian adat asal Bengkulu . " Silakan semua datang ke panggung . Mendapat sepeda dari Bapak Presiden bisa langsung dipakai keliling - keliling nanti , " ujarnya . Pada Upacara Pengibaran Bendera , lima orang menggunakan pakaian adat terbaik , yakni Menteri Hukum dan Hak Asasi Manusia (Menkum HAM) Yasonna Hamonangan Laoly dengan pakaian adat Nias , Ketua DPD RI Oesman Sapta Odang yang menggunakan pakaian adat Minang , dan Asisten Ajudan Presiden Syarif Muhammad Fidriansyah dengan pakaian adat Dayak , Kalimantan Barat . Dua orang lainnya adalah istri Kapolri Jenderal Pol Tito Karnavian , Tri Suswati yang mengenakan pakaian adat Papua , dan istri Wakil Ketua MPR Mahyudin , Agati Suli menggunakan baju adat Dayak . [bal]
Ringkasan manual atau referensi	Jokowi memimpin upacara penurunan bendera . Usai prosesi tersebut , Jokowi bagi-bagi sepeda kepada tamu undangan yang mengenakan pakaian adat terbaik . Berbeda dengan saat upacara pengibaran bendera , bagi-bagi sepeda kali ini diumumkan oleh Wakil Presiden Jusuf Kalla (JK) . Hal berbeda juga terjadi

	bagi mereka yang menerima . Apabila sepeda diberikan ke pejabat negara mau pun keluarga , kali ini sepeda diberikan kepada masyarakat biasa .
Ringkasan sistem	Usai prosesi penurunan bendera dilakukan , Jokowi kembali bagi-bagi sepeda kepada tamu undangan yang mengenakan pakaian adat terbaik . Hal berbeda juga terjadi bagi mereka yang menerima . Seperti tadi pagi , sore ini juga panitia membentuk tim penilai untuk menilai siapa yang berbusana tradisional adat - adat daerah yang paling baik sore ini , " kata JK . Terakhir , ada Sumahartarti yang mengenakan pakaian adat asal Bengkulu . " Mendapat sepeda dari Bapak Presiden bisa langsung dipakai keliling - keliling nanti , " ujarnya . Dua orang lainnya adalah istri Kapolri Jenderal Pol Tito Karnavian , Tri Suswati yang mengenakan pakaian adat Papua , dan istri Wakil Ketua MPR Mahyudin , Agati Suli menggunakan baju adat Dayak .

Statistik terkait dokumen artikel berita, ringkasan manual, dan ringkasan sistem, meliputi jumlah kata dan jumlah kalimat ditampilkan dalam tabel 4.2.

Tabel 4.2 Statistik artikel berita

Statistik	Jumlah Kata	Jumlah Kalimat
Artikel berita	289	13
Ringkasan manual	52	5
Ringkasan sistem	85	6

Berdasarkan Tabel 4.2, perbandingan dilakukan untuk menghitung nilai *recall* menggunakan metrik ROUGE-1. Metrik ini menghitung jumlah kata yang dianggap cocok (*overlap*) antara ringkasan sistem dan ringkasan manual. Tabel 4.3 menampilkan daftar kata pada hasil ringkasan sistem, dan tabel 4.4 menampilkan daftar kata-kata dalam ringkasan manual.

Tabel 4.3 Daftar kata hasil ringkasan sistem

Tito	Papua	terjadi	membentuk	Terakhir
Suli	dari	orang	lainnya	sepeda
MPR	Dayak	mereka	kembali	Suswati
kepada	sore	dan	prosesi	bagi
panitia	siapa	Jokowi	menggunakan	JK
...
istri	Kapolri	penurunan	bendera	berbeda
undangan	pakaian	Wakil	adat	terbaik

Tabel 4.4 Daftar kata hasil ringkasan manual

keluarga	menerima	memimpin	prosesi	Berbeda
bagi	terjadi	kali	Wakil	pun
biasa	oleh	diumumkan	bendera	Jokowi

upacara	pengibaran	pakaian	masyarakat	sepeda
penurunan	terbaik	undangan	negara	tamu
...
Usai	berbeda	JK	mengenakan	kepada
Apabila	Presiden	mau	pejabat	adat

Tabel 4.5 menampilkan daftar kata-kata yang muncul bersama (*overlapping words*) pada ringkasan manual dan ringkasan sistem.

Tabel 4.5 Daftar kemunculan kata bersama

menerima	prosesi	bagi	terjadi	Wakil
bendera	Jokowi	pakaian	juga	ini
sepeda	penurunan	Usai	terbaik	mereka
...
bagi-bagi	undangan	tamu	yang	berbeda
JK	mengenakan	kepada	Presiden	adat

Tabel 4.6 menampilkan jumlah kata yang terdapat di tabel 4.3, tabel 4.4, dan tabel 4.5, yaitu kata-kata dalam ringkasan manual, ringkasan sistem, serta jumlah kata yang sama (*overlap*) antara keduanya.

Tabel 4.6 Statistika jumlah kata ringkasan

Kata unik	Jumlah
Ringkasan manual	52
Ringkasan sistem	85
Kemunculan yang sama pada ringkasan manual dan ringkasan sistem	28

Berdasarkan data dalam tabel 4.6, nilai ROUGE-1 dihitung menggunakan rumus sesuai dengan persamaan (4.1), (4.2), dan (4.3). Adapun perhitungannya, sebagai berikut:

Recall: *recall* mengukur seberapa banyak kata pada ringkasan manual berhasil dicakup oleh ringkasan sistem. Dengan jumlah kata *overlap* sebanyak 28 dan jumlah total kata pada ringkasan manual sebanyak 52, nilai *recall* dihitung sebagai berikut:

$$Recall = \frac{\text{Jumlah kata overlap antara ringkasan sistem \& ringkasan manual}}{\text{Banyaknya kata pada ringkasan manual}} = \frac{28}{52} = 0.538$$

Precision: *precision* mengukur seberapa relevan kata-kata dalam ringkasan sistem dengan kata-kata pada ringkasan manual. Dengan jumlah kata *overlap* sebanyak 28 dan jumlah total kata dalam ringkasan sistem sebanyak 85, nilai *precision* dihitung sebagai berikut:

$$Precision = \frac{\text{Jumlah kata overlap antara ringkasan sistem \& ringkasan manual}}{\text{Banyaknya kata pada ringkasan sistem}} = \frac{28}{85} = 0.329$$

F1-Score: *f1-score* adalah rata-rata yang mempertimbangkan keseimbangan antara *recall* dan *precision*. Dengan nilai *recall* sebesar 0.538 dan *precision* sebesar 0.329, *f1-score* dihitung sebagai berikut:

$$F1-Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} = \frac{2 \times (0.329 \times 0.538)}{0.329 + 0.538} = 0.408$$

Pengujian dilakukan pada 14.261 artikel berita dengan dua tingkat kompresi, yaitu 50% dan 30%. Hasil evaluasi menggunakan metrik ROUGE-1 dirangkum dalam tabel 4.7 untuk tingkat kompresi 50% dan tabel 4.8 untuk tingkat kompresi 30%. Kedua tabel tersebut menunjukkan hasil evaluasi menggunakan tiga metrik utama, yaitu *recall*, *precision*, dan *f1-score* pada setiap artikel dalam dataset.

Tabel 4.7 Hasil evaluasi ROUGE-1 tiap artikel (*compression rate 50%*)

Artikel	Asli		Ringkasan sistem		Ringkasan manual		Jumlah kata		Evaluasi		
	Kata	Kalimat	Kata	Kalimat	Kata	Kalimat	Overlapping	Referensi unik	Recall	Precision	F1-Score
1	329	18	107	9	46	3	46	0	1	0.429907	0.601307
2	837	40	246	20	47	3	38	9	0.808511	0.154472	0.259386
3	595	24	176	12	42	2	38	4	0.904762	0.215909	0.348624
4	214	11	76	5	57	2	31	26	0.54386	0.407895	0.466165
5	289	13	85	6	52	5	28	24	0.538462	0.329412	0.408759
...
14258	444	26	137	13	61	3	36	25	0.590164	0.262774	0.363636
14259	281	14	90	7	57	3	31	26	0.54386	0.344444	0.421769
14260	198	13	47	6	46	4	12	34	0.26087	0.255319	0.258065
14261	400	22	138	11	53	3	51	2	0.962264	0.369565	0.534031
Rata-rata									0.6276	0.3446	0.4330

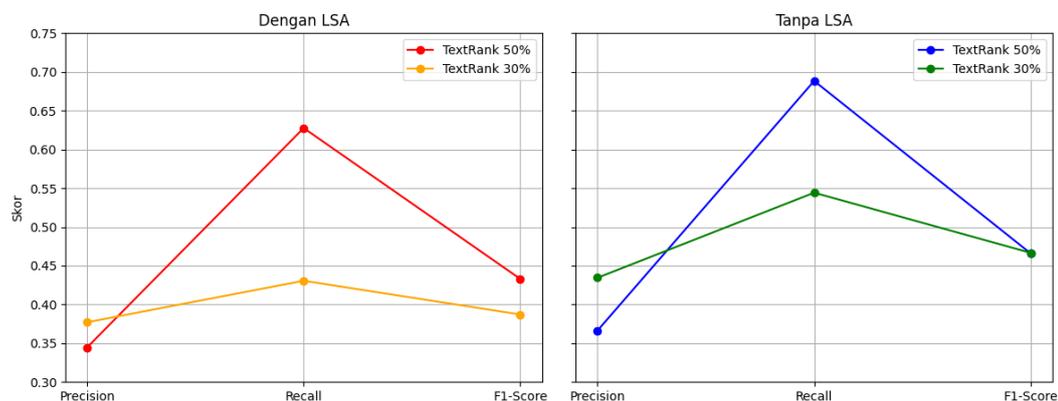
Tabel 4.8 Hasil evaluasi ROUGE-1 tiap artikel (*compression rate 30%*)

Artikel	Asli		Ringkasan sistem		Ringkasan manual		Jumlah kata		Evaluasi		
	Kata	Kalimat	Kata	Kalimat	Kata	Kalimat	Overlapping	Referensi unik	Recall	Precision	F1-Score
1	329	18	76	5	46	3	27	19	0.586957	0.355263	0.442623
2	837	40	158	12	47	3	37	10	0.787234	0.234177	0.360976
3	595	24	117	7	42	2	38	4	0.904762	0.324786	0.477987
4	214	11	42	3	57	2	15	42	0.263158	0.357143	0.30303
5	289	13	62	3	52	5	12	40	0.230769	0.193548	0.210526
...
14258	444	26	85	7	61	3	34	27	0.557377	0.4	0.465753
14259	281	14	65	4	57	3	21	36	0.368421	0.323077	0.344262
14260	198	13	29	3	46	4	10	36	0.217391	0.344828	0.266667
14261	400	22	69	6	53	3	31	22	0.584906	0.449275	0.508197
Rata-rata									0.4306	0.3771	0.3871

4.3 Pembahasan

Dalam penelitian ini menggunakan metrik ROUGE-1 untuk mengevaluasi performa sistem peringkasan teks dengan mengukur nilai *recall*, *precision*, dan *f1-score*. Metrik *recall* digunakan untuk menilai sejauh mana hasil ringkasan sistem mencakup informasi dari ringkasan manual. Semakin relevan ringkasan sistem terhadap ringkasan manual, maka semakin tinggi nilai *recall* yang dihasilkan. Jika *recall* bernilai 1, berarti seluruh informasi pada ringkasan manual berhasil dimasukkan ke dalam ringkasan sistem. Sebaliknya, *precision* mengukur sejauh mana informasi dalam ringkasan sistem relevan terhadap ringkasan manual. Nilai *precision* sebesar 1 menunjukkan bahwa semua informasi pada ringkasan sistem terdapat dalam ringkasan manual. Kombinasi dari kedua metrik ini adalah *f1-score*, memberikan gambaran menyeluruh mengenai kemampuan sistem dalam menghasilkan ringkasan yang relevan dan mencakup informasi penting.

Evaluasi dilakukan pada dua pendekatan sistem peringkasan, yaitu *TextRank* dengan pembobotan WIDF, serta *TextRank* yang dikombinasikan dengan WIDF dan reduksi dimensi menggunakan *Latent Semantic Analysis* (LSA). Uji coba dilakukan pada 14.261 artikel berita dengan dua tingkat kompresi ringkasan, yaitu 50% dan 30%. Gambar 4.1 menyajikan perbandingan performa dari kedua metode tersebut berdasarkan rata-rata nilai *recall*, *precision*, dan *f1-score*.



Gambar 4.1 Perbandingan rata-rata hasil evaluasi dengan dan tanpa LSA

Berdasarkan Gambar 4.1, dapat dilihat bahwa metode *TextRank* dengan WIDF menghasilkan nilai evaluasi yang lebih tinggi dibandingkan dengan metode yang menggunakan kombinasi WIDF dan LSA, baik pada tingkat kompresi 50% maupun 30%. Hal ini menunjukkan bahwa penambahan proses reduksi dimensi menggunakan LSA tidak memberikan peningkatan performa sistem pada data artikel berita yang digunakan.

Rata-rata hasil evaluasi lebih detail ditampilkan pada Tabel 4.9 yang menyajikan rata-rata nilai *recall*, *precision*, dan *f1-score* dari kedua metode pada masing-masing tingkat kompresi.

Tabel 4.9 Rata-rata hasil evaluasi ROUGE-1

Metode	Compression rate					
	50%			30%		
	Rata-rata Recall	Rata-rata Precision	Rata-rata F1-score	Rata-rata Recall	Rata-rata Precision	Rata-rata F1-score
<i>TextRank</i> + WIDF	0.6884	0.3664	0.4658	0.5442	0.4345	0.4666
<i>TextRank</i> + WIDF + LSA	0.6276	0.3446	0.4330	0.4306	0.3771	0.3871

Pada tingkat kompresi 30%, metode *TextRank* dengan WIDF menghasilkan nilai rata-rata *recall* sebesar 0.5442, *precision* 0.4345, dan *f1-score* 0.4666.

Sementara itu, pada *compression rate* 50%, nilai rata-rata *recall* adalah 0.6884, *precision* 0.3664, dan *f1-score* 0.4658. Pada tingkat kompresi 50%, ringkasan sistem cenderung lebih panjang dibandingkan dengan ringkasan manual, sehingga nilai *recall* menjadi tinggi. Namun, *precision* menurun karena banyak informasi dalam ringkasan sistem yang tidak terdapat pada ringkasan manual. Hal ini menyebabkan ketidakseimbangan antara nilai *recall* dan *precision*. Sebaliknya, pada tingkat kompresi 30%, panjang ringkasan sistem lebih mendekati panjang ringkasan manual, sehingga nilai *recall* dan *precision* tidak berbeda jauh.

Pada metode *TextRank* yang dikombinasikan dengan WIDF dan LSA pada tingkat kompresi 30%, diperoleh nilai rata-rata *recall* sebesar 0.4306, *precision* 0.3771, dan *f1-score* 0.3871. Sedangkan pada tingkat kompresi 50%, didapati nilai rata-rata *recall* 0.6276, *precision* 0.3446, dan *f1-score* 0.4330. Hasil yang diperoleh pada *compression rate* 30% serupa dengan metode tanpa LSA, di mana nilai *recall* dan *precision* tidak berbeda jauh. Namun, pada *compression rate* 50%, nilai *f1-score* mengalami peningkatan.

Pada metode *TextRank* dengan WIDF pada tingkat kompresi 50%, nilai *recall* dan *precision* tertinggi terdapat pada artikel berita ke-36 dan artikel berita ke-831 dengan nilai 1. Nilai *f1-score* tertinggi ditemukan pada artikel berita ke-598 dengan nilai 0,980. Nilai terendah pada ketiga metrik (*recall*, *precision*, dan *f1-score*) terdapat pada artikel ke-11536 dengan nilai 0. Hal tersebut dikarenakan jumlah kalimat pada teks asli artikel yang sangat sedikit. Pada tingkat kompresi 30%, nilai *recall* dan nilai *precision* tertinggi terdapat pada artikel berita ke-36 dan artikel berita ke-100 dengan nilai 1. Nilai *f1-score* tertinggi ditemukan pada artikel

berita ke-3683 dengan nilai 1. Pada artikel ke-2218 memperoleh nilai 0 pada semua metrik. Hal ini disebabkan oleh jumlah kalimat dalam artikel tersebut yang hanya terdiri dari 3 kalimat.

Pada metode *TextRank* yang dikombinasikan dengan WIDF dan LSA pada tingkat kompresi 50%, nilai *recall* dan *precision* tertinggi terdapat pada artikel berita ke-1 dan artikel berita ke-1649 dengan nilai 1. Nilai *f1-score* tertinggi ditemukan pada artikel berita ke-2312 dengan nilai 0,991. Nilai terendah pada ketiga metrik (*recall*, *precision*, dan *f1-score*) terdapat pada artikel ke-11536 dengan nilai 0. Hal tersebut dikarenakan jumlah kalimat pada teks asli artikel yang sangat sedikit. Pada tingkat kompresi 30%, nilai *recall* dan nilai *precision* tertinggi terdapat pada artikel berita ke-36 dan artikel berita ke-125 dengan nilai 1. Nilai *f1-score* tertinggi ditemukan pada artikel berita ke-3605 dengan nilai 1. Pada artikel ke-2218 memperoleh nilai 0 pada semua metrik. Hal ini disebabkan oleh jumlah kalimat dalam artikel tersebut yang hanya terdiri dari 3 kalimat.

Peringkasan teks artikel berita memberikan manfaat dengan menyajikan informasi utama secara lebih ringkas, sehingga memudahkan pengguna dalam memahami isi artikel. Berdasarkan hasil penelitian ini, metode *TextRank* dan WIDF tanpa penambahan LSA menunjukkan kinerja yang lebih optimal dalam menghasilkan ringkasan dibandingkan dengan metode yang menggunakan LSA.

Peringkasan teks otomatis (*automatic text summarization*) ini juga selaras dengan nilai-nilai Islam, terutama melalui dua konsep muamalah, yakni muamalah kepada Allah (*mu'amalah ma'a Allah*) dan muamalah kepada sesama manusia (*mu'amalah ma'a an-nas*). Dalam Islam, memanfaatkan waktu dengan baik serta

menghindari untuk mengikuti prasangka yang tidak berdasar merupakan hal yang dianjurkan. Dengan adanya sistem peringkasan ini, pengguna dapat memperoleh pemahaman yang lebih cepat tanpa harus membaca keseluruhan isi teks, sehingga waktu yang tersedia dapat digunakan untuk hal-hal lain yang lebih bermanfaat, baik dalam kehidupan sehari-hari maupun dalam menjalankan ibadah.

4.3.1 Aspek Muamalah Ma'a Allah

Muamalah ma'a Allah mencakup segala hal yang berhubungan antara manusia dengan Sang Pencipta, diwujudkan melalui ketaatan dan ketundukan terhadap perintah-Nya. Dalam Islam, Allah *Subhanahu wa Ta'ala* mengajarkan pentingnya menjalankan perintah-Nya dengan sebaik-baiknya, termasuk dalam memanfaatkan waktu secara efektif dan menghindari untuk mengikuti prasangka atau dugaan yang tidak berdasar. Sebagaimana firman Allah *Subhanahu wa Ta'ala* dalam Surah An-Nisa' ayat 59:

يَا أَيُّهَا الَّذِينَ آمَنُوا أَطِيعُوا اللَّهَ وَأَطِيعُوا الرَّسُولَ وَأُولَى الْأَمْرِ مِنْكُمْ فَإِنْ تَنَازَعْتُمْ فِي شَيْءٍ فَرُدُّوهُ إِلَى اللَّهِ وَالرَّسُولِ إِنْ كُنْتُمْ تُؤْمِنُونَ بِاللَّهِ وَالْيَوْمِ الْآخِرِ ذَلِكَ خَيْرٌ وَأَحْسَنُ تَأْوِيلًا

“Wahai orang-orang yang beriman, taatilah Allah dan taatilah Rasul (Nabi Muhammad) serta ulul amri (pemegang kekuasaan) di antara kamu. Jika kamu berbeda pendapat tentang sesuatu, kembalikanlah kepada Allah (Al-Qur'an) dan Rasul (sunahnya) jika kamu beriman kepada Allah dan hari Akhir. Yang demikian itu lebih baik (bagimu) dan lebih bagus akibatnya (di dunia dan di akhirat).” (QS. An-Nisa' 4:59)

Ayat ini menekankan pentingnya menaati Allah *Subhanahu wa Ta'ala* dan Rasul-Nya, termasuk dalam mengelola waktu dengan sebaik-baiknya. Dalam penelitian ini, sistem peringkasan teks berperan dalam membantu pengguna

menghemat waktu dengan menyediakan ringkasan informatif, sehingga pembaca tidak perlu membaca keseluruhan teks untuk memahami inti informasi.

Ajaran dalam Al-Qur'an mengenai pentingnya menghindari untuk mengikuti prasangka yang tidak berdasar tercermin dalam Surah Yunus ayat 36:

وَمَا يَتَّبِعْ أَكْثَرُهُمْ إِلَّا ظَنًّا إِنَّ الظَّنَّ لَا يُغْنِي مِنَ الْحَقِّ شَيْئًا إِنَّ اللَّهَ عَلِيمٌ بِمَا يَفْعَلُونَ

"Kebanyakan mereka hanya mengikuti dugaan. Sesungguhnya dugaan itu tidak sedikit pun berguna menyangkut (perolehan) kebenaran. Sesungguhnya Allah Maha Mengetahui apa yang mereka lakukan." (QS. Yunus 10:36)

Ayat ini menunjukkan bahwa mengingatkan kita untuk tidak mengikuti prasangka yang tidak berdasar. Dalam penelitian ini, penggunaan LSA sebagai metode reduksi dimensi tidak selalu meningkatkan performa. Berdasarkan hasil pengujian, peringkasan tanpa menggunakan LSA ternyata memberikan kinerja lebih baik, yang mana selaras dengan ajaran Al-Qur'an dalam ayat ini.

Selain itu, peringkasan teks juga mencerminkan salah satu kekhususan Rasulullah *Shallallahu 'Alaihi wa Sallam*, yaitu *Jawami' al-Kalim*, yakni kemampuan menyampaikan pesan yang ringkas tetapi bermakna luas. Sebagaimana dalam hadist berikut:

فُضِّلْتُ عَلَى الْأَنْبِيَاءِ بِسِتِّ أَعْطَيْتُ جَوَامِعَ الْكَلِمِ وَنُصِرْتُ بِالرُّعْبِ وَأُحِلَّتْ لِي الْغَنَائِمُ وَجُعِلَتْ لِي الْأَرْضُ طَهُورًا
وَمَسْجِدًا وَأُرْسِلْتُ إِلَى الْخَلْقِ كَافَّةً وَخُتِمَ بِي النَّبِيُّونَ

"Aku diberikan keutamaan dibandingkan nabi-nabi sebelumku dengan enam hal: aku diberi Jawami' al-Kalim, aku ditolong dengan rasa takut yang disematkan di hati para musuh, dihalalkan aneka hewan ternak, dijadikan bumi yang suci sebagai tempat sujud, Aku diutus kepada seluruh makhluk dan menutup kenabian denganku" (HR. Muslim 1195 & Turmudzi 1640).

Dalam hadist ini disebutkan bahwa *Jawami' al-Kalim* merupakan bagian dari keistimewaan Nabi *Shallallahu 'Alaihi wa Sallam*. Disebut sebagai keistimewaan Nabi Muhammad *Shallallahu 'Alaihi wa Sallam*, karena *Jawami' al-Kalim* tidak diberikan kepada nabi-nabi sebelum beliau. Prinsip ini menjadi dasar dalam sistem peringkasan teks, di mana ringkasan yang dihasilkan tetap padat, bermakna, dan relevan sehingga pengguna tetap dapat memahami inti informasi dengan lebih efisien tanpa kehilangan esensi utama dokumen.

4.3.2 Aspek Muamalah Ma'a An-Nas

Muamalah ma'a an-nas mencakup hubungan antara manusia dengan sesama, yang dalam Islam didasarkan pada prinsip ta'awun atau tolong-menolong dalam kebaikan. Prinsip ini dianjurkan dalam Al-Qur'an, sebagaimana firman Allah *Subhanahu wa Ta'ala* dalam QS. Al-Ma'idah ayat 2:

يَا أَيُّهَا الَّذِينَ آمَنُوا لَا تَحِلُّوا شَعَائِرَ اللَّهِ وَلَا الشَّهْرَ الْحَرَامَ وَلَا الْهَدْيَ وَلَا الْقَلَائِدَ وَلَا آيَاتِ الْبَيْتِ الْحَرَامِ يَبْتَغُونَ فَضْلًا
مِنْ رَبِّهِمْ وَرِضْوَانًا وَإِذَا حَلَلْتُمْ فَاصْطَادُوا وَلَا يَجْرِمَنَّكُمْ شَنَا نُ قَوْمٍ أَنْ صَدُّوكُمْ عَنِ الْمَسْجِدِ الْحَرَامِ أَنْ تَعْتَدُوا وَتَعَاوَنُوا
عَلَى الْبِرِّ وَالتَّقْوَىٰ وَلَا تَعَاوَنُوا عَلَى الْإِثْمِ وَالْعُدْوَانِ وَاتَّقُوا اللَّهَ إِنَّ اللَّهَ شَدِيدُ الْعِقَابِ

“Wahai orang-orang yang beriman, janganlah kamu melanggar syiar-syiar (kesucian) Allah, jangan (melanggar kehormatan) bulan-bulan haram, jangan (mengganggu) hadyu (hewan-hewan kurban) dan qalā'id (hewan-hewan kurban yang diberi tanda), dan jangan (pula mengganggu) para pengunjung Baitulharam sedangkan mereka mencari karunia dan rida Tuhannya! Apabila kamu telah bertahalul (menyelesaikan ihram), berburulah (jika mau). Janganlah sekali-kali kebencian(-mu) kepada suatu kaum, karena mereka menghalang-halangi dari Masjidilharam, mendorongmu berbuat melampaui batas (kepada mereka). Tolong-menolonglah kamu dalam (mengerjakan) kebajikan dan takwa, dan jangan tolong-menolong dalam berbuat dosa dan permusuhan. Bertakwalah kepada Allah, sesungguhnya Allah sangat berat siksaan-Nya.” (QS. Al-Ma'idah 5:2)

Ayat 2 dalam QS. Al-Ma'idah mengajarkan agar umat Islam saling membantu dalam kebaikan dan ketakwaan serta menghindari perbuatan dosa dan permusuhan. Larangan terhadap tindakan yang dapat memicu konflik menggarisbawahi pentingnya menjaga keharmonisan hubungan antarindividu dan membentuk landasan etika sosial dalam interaksi manusia.

Al-Qur'an tidak secara spesifik menjelaskan mengenai sistem peringkasan teks. Akan tetapi, prinsip *ta'awun* atau saling tolong-menolong dalam ayat ini dapat dijadikan landasan konseptual dalam pengembangan sistem peringkasan teks. Peringkasan teks bertujuan untuk menghemat waktu dan meningkatkan efisiensi dalam memahami informasi dengan cara menyajikannya secara ringkas namun tetap mencakup isi yang penting. Selaras dengan konsep *muamalah ma'a an-nas* yang menekankan kemudahan dalam berbagi informasi, diharapkan pengguna dapat lebih mudah mengakses informasi yang dibutuhkan dan memahami isi teks secara lebih efektif.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil implementasi dan evaluasi menggunakan metrik ROUGE-1, algoritma *TextRank* yang dikombinasikan dengan pembobotan WIDF tanpa penambahan LSA menunjukkan performa yang lebih baik dibandingkan dengan kombinasi yang menggunakan LSA pada kedua tingkat kompresi, yaitu 30% dan 50%. Evaluasi dilakukan dengan menggunakan parameter *recall*, *precision*, dan *f1-score*.

Pada tingkat kompresi 30%, metode *TextRank* dengan WIDF menghasilkan nilai rata-rata *recall* sebesar 0,5442 *precision* sebesar 0,4345, dan *f1-score* sebesar 0,4666. Sedangkan pada tingkat kompresi 50%, diperoleh nilai *recall* sebesar 0,6884, *precision* 0,3664, dan *f1-score* 0,4658. Sementara itu, metode *TextRank* dengan kombinasi WIDF dan LSA pada tingkat kompresi 30% menghasilkan *recall* sebesar 0,4306, *precision* 0,3771, dan *f1-score* 0,3871. Pada tingkat kompresi 50%, diperoleh *recall* sebesar 0,6276, *precision* 0,3446, dan *f1-score* 0,4330.

Hasil tersebut menunjukkan bahwa penambahan proses reduksi dimensi menggunakan LSA tidak selalu meningkatkan kualitas ringkasan. Dengan demikian, dapat disimpulkan bahwa algoritma *TextRank* dengan pembobotan WIDF tanpa LSA merupakan metode yang lebih efektif dalam menghasilkan ringkasan teks artikel berita berdasarkan metrik ROUGE-1.

5.2 Saran

Peneliti menyadari adanya beberapa keterbatasan dalam penelitian ini. Untuk pengembangan lebih lanjut dalam membangun sistem peringkasan otomatis berbahasa Indonesia, berikut beberapa saran:

1. Menggunakan data ringkasan manual yang lebih panjang agar tidak terjadi kehilangan informasi krusial yang terdapat dalam artikel berita.
2. Menggunakan ekstraksi fitur *word embedding* sebagai alternatif dari metode pembobotan, untuk menangkap konteks dan makna kata secara lebih baik.

DAFTAR PUSTAKA

- Akhmetov, I., Gelbukh, A., & Mussabayev, R. (2021). Greedy Optimization Method for Extractive Summarization of Scientific Articles. *IEEE Access*, 9, 168141–168153. <https://doi.org/10.1109/ACCESS.2021.3136302>
- AL-Hafiidh, D. F., Rozi, I. F., & Putri, I. K. (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>
- Al-Mahalli, I. J., & As-Suyuti, I. J. (1469). *Tafsir Jalalin Jilid 2*. Sinar Baru Algesindo. <https://tedisobandi.blogspot.com/>
- Armianti, D. N., Indriati, & Adinugroho, S. (2019). Klasifikasi Emosi Lagu Berdasarkan Lirik pada Teks Berbahasa Indonesia Menggunakan K-Nearest Neighbor dengan Pembobotan WIDF. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3.
- Caesarardhi, M. R., Vinarti, R. A., & Kusumawardani, R. P. (2023). Implementasi Aplikasi Peringkasan Teks Otomatis Untuk Atlas Penyakit Menular Pada Manusia Menggunakan Metode Ordered Abstractive Summarization. *Jurnal Linguistik Komputasional (JLK)*, 6(2), 70–86. <https://doi.org/10.26418/jlk.v6i2.182>
- Chen, Y., & Song, Q. (2021). News Text Summarization Method based on BART-TextRank Model. *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2005–2010. <https://doi.org/10.1109/IAEAC50856.2021.9390683>
- 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>
- Deolika, A., Kusrini, K., & Luthfi, E. T. (2019). ANALISIS PEMBOBOTAN KATA PADA KLASIFIKASI TEXT MINING. *JURNAL TEKNOLOGI INFORMASI*, 3(2), 179. <https://doi.org/10.36294/jurti.v3i2.1077>
- Eris, Christanti M, V., & Pragantha, J. (2017). PENERAPAN ALGORITMA TEXTRANK UNTUK AUTOMATIC SUMMARIZATION PADA DOKUMEN BERBAHASA INDONESIA. *Jurnal Ilmu Teknik Dan Komputer*, 1.
- Fadhila, L. N., & Nuryana, I. K. D. (2024). *Teks Ringkas Otomatis pada Portal Berita CNN Indonesia Menggunakan Algoritma Textrank*. 05(01).

- Fakhrezi, M. F., Bijaksana, Moch. 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, 391–398. <https://doi.org/10.1016/j.procs.2021.01.021>
- Geetha J K, & Deepamala N. (2015). KANNADA TEXT SUMMARIZATION USING LATENT SEMANTIC ANALYSIS. *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*.
- Gotami, N. S. W., Indriati, & Dewi, R. K. (2018). Peringkasan Teks Otomatis Secara Ekstraktif Pada Artikel Berita Kesehatan Berbahasa Indonesia Dengan Menggunakan Metode Latent Semantic Analysis. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*.
- Gulati, V., Kumar, D., Popescu, D. E., & Hemanth, J. D. (2023). Extractive Article Summarization Using Integrated TextRank and BM25+ Algorithm. *Electronics*, 12(2), 372. <https://doi.org/10.3390/electronics12020372>
- 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, 114–121. <https://doi.org/10.1016/j.ijmedinf.2019.05.019>
- Gunawan, D., Harahap, S. H., & Fadillah Rahmat, R. (2019). Multi-document Summarization by using TextRank and Maximal Marginal Relevance for Text in Bahasa Indonesia. *2019 International Conference on ICT for Smart Society (ICISS)*, 1–5. <https://doi.org/10.1109/ICISS48059.2019.8969785>
- Hadyan, F., Shaufiah, & Bijaksana, Moch. A. (2017). Comparison of Document Index Graph Using TextRank and HITS Weighting Method in Automatic Text Summarization. *Journal of Physics: Conference Series*, 801, 012076. <https://doi.org/10.1088/1742-6596/801/1/012076>
- Halimah, Surya Agustian, & Siti Ramadhani. (2022). Peringkasan teks otomatis (automated text summarization) pada artikel berbahasa indonesia menggunakan algoritma lexrank. *Jurnal CoSciTech (Computer Science and Information Technology)*, 3(3), 371–381. <https://doi.org/10.37859/coscitech.v3i3.4300>
- Hidayattullah, M. F., & Azizi, A. (2021). Peringkasan Otomatis Teks Berbahasa Arab Menggunakan Algoritma TextRank. *Jurnal Ilmiah Informatika*, 6(1), 33–42. <https://doi.org/10.35316/jimi.v6i1.1231>
- Jayakodi, K., Bandara, M., & Meedeniya, D. (2016). An automatic classifier for exam questions with WordNet and Cosine similarity. *2016 Moratuwa Engineering Research Conference (MERCon)*, 12–17. <https://doi.org/10.1109/MERCon.2016.7480108>

- Karo, I. M. K., Perdana, A., & Dewi, S. (2024). Automatic Text Review Summarization of Digital Library System Application using TextRank Algorithm and TF-IDF. *2024 4th International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, 570–575. <https://doi.org/10.1109/ICSINTESA62455.2024.10747952>
- Katadata.co.id. (2023). *Meski Trennya Turun, Media Online Tetap Jadi Sumber Berita Utama Masyarakat Indonesia | Databoks*. <https://databoks.katadata.co.id/media/statistik/22dd8cfc6a8f5e3/meski-trennya-turun-media-online-tetap-jadi-sumber-berita-utama-masyarakat-indonesia>
- Kresna, Y. A., & Cholissodin, I. (2021). Peringkasan Teks menggunakan metode Maximum Marginal Relevance terhadap Artikel Berita terkait COVID-19. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*.
- Kurniawan, K., & Louvan, S. (2019). *IndoSum: A New Benchmark Dataset for Indonesian Text Summarization* (arXiv:1810.05334). arXiv. <http://arxiv.org/abs/1810.05334>
- Lin, C.-Y. (2004). *Looking for a Few Good Metrics: ROUGE and its Evaluation*.
- Madhuri, J. N., & Ganesh Kumar, R. (2019). Extractive Text Summarization Using Sentence Ranking. *2019 International Conference on Data Science and Communication (IconDSC)*, 1–3. <https://doi.org/10.1109/IconDSC.2019.8817040>
- Mandar, G., & Gunawan, G. (2018). Peringkasan dokumen berita Bahasa Indonesia menggunakan metode Cross Latent Semantic Analysis. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, 3(2), 94–104. <https://doi.org/10.26594/register.v3i2.1161>
- Mujilahwati, S. (2023). Kombinasi Algoritma Data Reduksi untuk Optimalisasi Dokumen Cluster. *Jurnal Eksplora Informatika*, 12(2), 113–119. <https://doi.org/10.30864/eksplora.v12i2.819>
- Naik, S. S., & Gaonkar, M. N. (2017). Extractive text summarization by feature-based sentence extraction using rule-based concept. *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 1364–1368. <https://doi.org/10.1109/RTEICT.2017.8256821>
- Nawaz, A., Bakhtyar, M., Baber, J., Ullah, I., Noor, W., & Basit, A. (2020). Extractive Text Summarization Models for Urdu Language. *Information Processing & Management*, 57(6), 102383. <https://doi.org/10.1016/j.ipm.2020.102383>

- Pratama, S. E., Darmalaksana, W., Maylawati, D. S., Sugilar, H., Mantoro, T., & Ali Ramdhani, M. (2020). Weighted inverse document frequency and vector space model for hadith search engine. *Indonesian Journal of Electrical Engineering and Computer Science*, 18(2), 1004. <https://doi.org/10.11591/ijeecs.v18.i2.pp1004-1014>
- Ramadhan, M. R., Endah, S. N., & Mantau, A. B. J. (2020). Implementation of Textrank Algorithm in Product Review Summarization. *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, 1–5. <https://doi.org/10.1109/ICICoS51170.2020.9299005>
- Rozan, Z. R., & Dewi, A. O. P. (2022). Penggunaan Internet sebagai Sumber Informasi pada Generasi Baby boomer berdasarkan Kemampuan Literasi Informasi. *Anuva: Jurnal Kajian Budaya, Perpustakaan, dan Informasi*, 6(1), 23–42. <https://doi.org/10.14710/anuva.6.1.23-42>
- Rozi, I., Batubulan, K., & Rusbandi, M. (2021). Otomatisasi Peringkasan Teks Pada Dokumen Hukum Menggunakan Metode Latent Semantic Analysis. *JIP (Jurnal Informatika Polinema)*.
- See, A., Liu, P. J., & Manning, C. D. (2017). *Get To The Point: Summarization with Pointer-Generator Networks* (arXiv:1704.04368). arXiv. <http://arxiv.org/abs/1704.04368>
- Septiani, D., & Isabela, I. (2022). ANALISIS TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) DALAM TEMU KEMBALI INFORMASI PADA DOKUMEN TEKS. 01(2).
- Sianturi, M. H. F., Ridok, A., & Santoso, E. (2023). Peringkasan Teks Otomatis menggunakan Metode Latent Semantic Analysis pada Artikel Berita Ekonomi berbahasa Indonesia. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*.
- Sihombing, J. J., Arnita, A., Al Idrus, S. I., & Niska, D. Y. (2024). Implementation of text summarization on indonesian scientific articles using textrank algorithm with TF-IDF web-based. *Journal of Soft Computing Exploration*, 5(3), 310–319. <https://doi.org/10.52465/josce.v5i3.475>
- Suhartono, D. (2015). *Penggunaan Latent Semantic Analysis (LSA) dalam Pemrosesan Teks*. <https://socs.binus.ac.id/2015/08/03/penggunaan-latent-semantic-analysis-lsa-dalam-pemrosesan-teks/>
- Susandi, D., & Sholahudin, U. (2017). Pemanfaatan Vector Space Model pada Penerapan Algoritma Nazief Adriani, KNN dan Fungsi Similarity Cosine untuk Pembobotan IDF dan WIDF pada Prototipe Sistem Klasifikasi Teks Bahasa Indonesia. *ProTekInfo(Pengembangan Riset dan Observasi Teknik Informatika)*, 3, 22–29. <https://doi.org/10.30656/protkinfo.v3i0.54>

- Suwija Putra, I. M., Adiwinata, Y., Singgih Putri, D. P., & Sutramiani, N. P. (2021). Extractive Text Summarization of Student Essay Assignment Using Sentence Weight Features and Fuzzy C-Means. *International Journal of Artificial Intelligence Research*, 5(1). <https://doi.org/10.29099/ijair.v5i1.187>
- 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. In H. Pérez García, L. Sánchez González, M. Castejón Limas, H. Quintián Pardo, & E. Corchado Rodríguez (Eds.), *Hybrid Artificial Intelligent Systems* (Vol. 11734, pp. 540–550). Springer International Publishing. https://doi.org/10.1007/978-3-030-29859-3_46
- Vijaya Saraswathi, R., Chunchu, R. V., Kunchala, S., Varun, M., Begari, T., & Bodduru, S. (2022). A LSTM based Deep Learning Model for Text Summarization. *2022 6th International Conference on Electronics, Communication and Aerospace Technology*, 1063–1068. <https://doi.org/10.1109/ICECA55336.2022.10009541>
- Wedel, I., Palk, M., & Voß, S. (2022). A Bilingual Comparison of Sentiment and Topics for a Product Event on Twitter. *Information Systems Frontiers*, 24(5), 1635–1646. <https://doi.org/10.1007/s10796-021-10169-x>
- Widyassari, A. P., Rustad, S., Shidik, G. F., Noersasongko, E., Syukur, A., Affandy, A., & Setiadi, D. R. I. M. (2022). Review of automatic text summarization techniques & methods. *Journal of King Saud University - Computer and Information Sciences*, 34(4), 1029–1046. <https://doi.org/10.1016/j.jksuci.2020.05.006>
- Yulianti, E., Pangestu, N., & Jiwanggi, M. A. (2023). Enhanced TextRank using weighted word embedding for text summarization. *International Journal of Electrical and Computer Engineering (IJECE)*, 13(5), 5472. <https://doi.org/10.11591/ijece.v13i5.pp5472-5482>
- Yuliska, Y., & Syaliman, K. U. (2020). Literatur Review Terhadap Metode, Aplikasi dan Dataset Peringkasan Dokumen Teks Otomatis untuk Teks Berbahasa Indonesia. *IT Journal Research and Development*, 5(1), 19–31. [https://doi.org/10.25299/itjrd.2020.vol5\(1\).4688](https://doi.org/10.25299/itjrd.2020.vol5(1).4688)
- Zamzam, M. A. (2020). SISTEM AUTOMATIC TEXT SUMMARIZATION MENGGUNAKAN ALGORITMA TEXTRANK. *MATICS*, 12(2), 111–116. <https://doi.org/10.18860/mat.v12i2.8372>
- Zaware, S., Patadiya, D., Gaikwad, A., Gulhane, S., & Thakare, A. (2021). Text Summarization using TF-IDF and Textrank algorithm. *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, 1399–1407. <https://doi.org/10.1109/ICOEI51242.2021.9453071>

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, 1–19. <https://doi.org/10.1155/2022/5649994>

LAMPIRAN

LAMPIRAN

Lampiran 1: Kode program WIDF (*Weighted Inverse Document Frequency*)

```
from collections import Counter
from tqdm import tqdm

widf_matrix = []

# Iterasi setiap paragraf
for paragraf in tqdm(df["final_preprocessing"]):

    tf_kalimat = [Counter(kalimat) for kalimat in paragraf]

    unique_words = sorted(set(kata for kalimat in paragraf for kata in kalimat))

    total_kemunculan_kata = {
        kata: sum(tf.get(kata, 0) for tf in tf_kalimat)
        for kata in unique_words
    }

    widf = np.array([
        [
            (tf.get(kata, 0) / total_kemunculan_kata[kata]) if total_kemunculan_kata[kata] != 0 else 0
            for kata in unique_words
        ]
        for tf in tf_kalimat
    ])

    # Save matriks WIDF
    widf_matrix.append(widf)
```

Lampiran 2: Kode program LSA (*Latent Semantic Analysis*)

```
from sklearn.decomposition import TruncatedSVD

lsa_model = TruncatedSVD(n_components=2, random_state=42)

lsa_result = []

# Proses LSA pada setiap matriks WIDF
for matrix in widf_matrix:
    if matrix.shape[0] > 1 and matrix.shape[1] > 1:
        reduced_matrix = lsa_model.fit_transform(matrix)
    else:
        reduced_matrix = np.zeros((matrix.shape[0], 2))

    lsa_result.append(reduced_matrix)
```

Lampiran 3: Kode program *Cosine Similarity*

```
from sklearn.metrics.pairwise import cosine_similarity

similarity_matrices = []

# Menghitung cosine similarity untuk setiap hasil LSA
for reduced_matrix in lsa_result:
    similarity_matrix = cosine_similarity(reduced_matrix)
    similarity_matrices.append(similarity_matrix)
```

Lampiran 4: Kode program *TextRank*

```
# Fungsi TextRank
def textrank(kalimat_list, similarity_matrix, damping=0.85, max_iter=100, tol=1e-6):
    jumlah_kalimat = len(kalimat_list)
    skor = np.ones(jumlah_kalimat) / jumlah_kalimat

    for iterasi in range(max_iter):
        skor_baru = np.zeros(jumlah_kalimat)
        for i in range(jumlah_kalimat):
            total_kontribusi = 0
            for j in range(jumlah_kalimat):
                if similarity_matrix[j, i] > 0:
                    jumlah_keluar = np.sum(similarity_matrix[j])
                    if jumlah_keluar > 0:
                        kontribusi = (similarity_matrix[j, i] / jumlah_keluar) * skor[j]
                        total_kontribusi += kontribusi
            skor_baru[i] = (1 - damping) + damping * total_kontribusi

        if np.linalg.norm(skor_baru - skor, ord=1) < tol:
            break

    skor = skor_baru

    return skor
```

Lampiran 5: Kode program pembentukan ringkasan

```
CR = 0.5 # Compression Rate
all_selected_indices = []

for i in range(len(df)):
    kalimat = df["sentences"].iloc[i]
    similarity = similarity_matrices[i]

    # Skor tiap kalimat
    skor_kalimat = textrank(kalimat, similarity)

    # Urutkan berdasarkan skor
    peringkat_kalimat = sorted(enumerate(skor_kalimat), key=lambda x: x[1], reverse=True)

    # Mengambil kalimat terpilih sesuai CR
    jumlah_ringkasan = int(CR * len(skor_kalimat))
    kalimat_terpilih_idx = [idx for idx, _ in peringkat_kalimat[:jumlah_ringkasan]]
    kalimat_terpilih_idx.sort()

    all_selected_indices.append(kalimat_terpilih_idx)
```

Lampiran 6: Data teks artikel berita asli dan ringkasan sistem *compression rate* 50% dan 30%

No Berita	Teks Berita	Ringkasan Sistem	
		50%	30%
1	<p>Jakarta , CNN Indonesia - - Dokter Ryan Thamrin , yang terkenal lewat acara Dokter Oz Indonesia , meninggal dunia pada Jumat (4 / 8) dini hari . Dokter Lula Kamal yang merupakan selebriti sekaligus rekan kerja Ryan menyebut kawannya itu sudah sakit sejak setahun yang lalu . Lula menuturkan , sakit itu membuat Ryan mesti vakum dari semua kegiatannya , termasuk menjadi pembawa acara Dokter Oz Indonesia . Kondisi itu membuat Ryan harus kembali ke kampung halamannya di Pekanbaru , Riau untuk menjalani istirahat . " Setahu saya dia orangnya sehat , tapi tahun lalu saya dengar dia sakit . (Karena) sakitnya , ia langsung pulang ke Pekanbaru , jadi kami yang mau jenguk juga susah . Barangkali mau istirahat , ya betul juga , kalau di Jakarta susah isirahatnya , " kata Lula kepada CNNIndonesia.com , Jumat (4 / 8) . Lula yang mengenal Ryan sejak sebelum aktif berkarier di televisi mengaku belum sempat membesuk Ryan lantaran lokasi yang jauh . Dia juga tak tahu penyakit apa yang diderita Ryan . " Itu saya enggak tahu , belum sempat jenguk dan enggak selamanya bisa dijenguk juga . Enggak tahu berat sekali apa bagaimana , " tutur Ryan . Walau sudah setahun menderita sakit , Lula tak mengetahui apa penyebab pasti kematian Dr Oz Indonesia itu . Meski demikian , ia mendengar beberapa kabar yang menyebut bahwa penyebab Ryan meninggal adalah karena jatuh di kamar mandi . Saya tidak tahu , barangkali penyakit yang dulu sama yang sekarang berbeda , atau penyebab kematiannya beda dari penyakit sebelumnya . Kita kan enggak bisa mengambil kesimpulan , " kata Lula . Ryan Thamrin terkenal sebagai dokter yang rutin membagikan tips</p>	<p>Jakarta , CNN Indonesia - - Dokter Ryan Thamrin , yang terkenal lewat acara Dokter Oz Indonesia , meninggal dunia pada Jumat (4 / 8) dini hari . Dokter Lula Kamal yang merupakan selebriti sekaligus rekan kerja Ryan menyebut kawannya itu sudah sakit sejak setahun yang lalu . Lula menuturkan , sakit itu membuat Ryan mesti vakum dari semua kegiatannya , termasuk menjadi pembawa acara Dokter Oz Indonesia . Kondisi itu membuat Ryan harus kembali ke kampung halamannya di Pekanbaru , Riau untuk menjalani istirahat . " Barangkali mau istirahat , ya betul juga , kalau di Jakarta susah isirahatnya , " kata Lula kepada CNNIndonesia.com , Jumat (4 / 8) . Enggak tahu berat sekali apa bagaimana , " tutur Ryan . Walau sudah setahun menderita sakit , Lula tak mengetahui apa penyebab pasti kematian Dr Oz Indonesia itu . Meski demikian , ia mendengar beberapa kabar yang menyebut bahwa penyebab Ryan meninggal adalah karena jatuh di kamar mandi .</p>	<p>Jakarta , CNN Indonesia - - Dokter Ryan Thamrin , yang terkenal lewat acara Dokter Oz Indonesia , meninggal dunia pada Jumat (4 / 8) dini hari . Lula menuturkan , sakit itu membuat Ryan mesti vakum dari semua kegiatannya , termasuk menjadi pembawa acara Dokter Oz Indonesia . Barangkali mau istirahat , ya betul juga , kalau di Jakarta susah isirahatnya , " kata Lula kepada CNNIndonesia.com , Jumat (4 / 8) . Walau sudah setahun menderita sakit , Lula tak mengetahui apa penyebab pasti kematian Dr Oz Indonesia itu . Meski demikian , ia mendengar beberapa kabar yang menyebut bahwa penyebab Ryan meninggal adalah karena jatuh di kamar mandi .</p>

	<p>dan informasi kesehatan lewat tayangan Dokter Oz Indonesia . Ryan menempuh Pendidikan Dokter pada tahun 2002 di Fakultas Kedokteran Universitas Gadjah Mada . Dia kemudian melanjutkan pendidikan Klinis Kesehatan Reproduksi dan Penyakit Menular Seksual di Mahachulalongkornrajavidyalaya University , Bangkok , Thailand pada 2004 .</p>	<p>Kita kan enggak bisa mengambil kesimpulan , " kata Lula .</p>	
2	<p>Selfie ialah salah satu tema terpanas di kalangan produsen smartphone , bahkan menjadi senjata andalan beberapa brand terkenal . Anda mungkin berpikir bahwa saat ini , pasar handset spesialis selfie sudah sangat sesak . Tapi Asus masih melihat adanya peluang besar menanti di sana . Dari data mereka , sebanyak 71 persen orang Indonesia setidaknya mengambil selfie atau wefie setiap minggu . Setelah mulai menyelami ranah swafoto dua tahun silam lewat ZanFone Selfie , sang produsen hardware asal Taiwan itu akhirnya membawa sepasang pewarisnya ke tanah air . Handset - handset ini merupakan anggota keluarga ZenFone generasi keempat dan keduanya sama-sama dibekali setup kamera ganda di depan . Mereka adalah Asus ZenFone 4 Selfie Pro ZD552KL dan ZenFone 4 Selfie ZD553KL . CEO Asus Jerry Shen menjelaskan bahwa sudah waktunya bagi Asus untuk memberikan penawaran baru buat penggemar self - portrait demi menunjukkan keseriusan mereka di segmen itu . Walaupun telah tersedia banyak pilihan , Asus berpendapat bahwa konsumen di Indonesia membutuhkan solusi yang lebih profesional . Dua ZenFone Selfie anyar ini kabarnya diracik sedemikian rupa sebagai jawaban atas kekurangan yang ada di perangkat - perangkat kompetitor , khususnya pada aspek jangkauan lensa dan performa di kondisi low-light . ZenFone 4 Selfie Pro merupakan produk swafoto pamungkas Asus . Tubuh berbahan aluminiumnya dibuat melalui teknik nano molding dan tiap lekukannya dibentuk</p>	<p>Anda mungkin berpikir bahwa saat ini , pasar handset spesialis selfie sudah sangat sesak . Dari data mereka , sebanyak 71 persen orang Indonesia setidaknya mengambil selfie atau wefie setiap minggu . Mereka adalah Asus ZenFone 4 Selfie Pro ZD552KL dan ZenFone 4 Selfie ZD553KL . Walaupun telah tersedia banyak pilihan , Asus berpendapat bahwa konsumen di Indonesia membutuhkan solusi yang lebih profesional . Dua ZenFone Selfie anyar ini kabarnya diracik sedemikian rupa sebagai jawaban atas kekurangan yang ada di perangkat - perangkat kompetitor , khususnya pada aspek jangkauan lensa dan performa di kondisi low-light . ZenFone 4 Selfie Pro merupakan produk swafoto pamungkas Asus . Tubuh berbahan aluminiumnya dibuat melalui teknik nano molding dan tiap lekukannya dibentuk secara presisi . Atraksi utama dari ZenFone 4 Selfie Pro</p>	<p>Anda mungkin berpikir bahwa saat ini , pasar handset spesialis selfie sudah sangat sesak . Mereka adalah Asus ZenFone 4 Selfie Pro ZD552KL dan ZenFone 4 Selfie ZD553KL . Walaupun telah tersedia banyak pilihan , Asus berpendapat bahwa konsumen di Indonesia membutuhkan solusi yang lebih profesional . Dua ZenFone Selfie anyar ini kabarnya diracik sedemikian rupa sebagai jawaban atas kekurangan yang ada di perangkat - perangkat kompetitor , khususnya pada aspek jangkauan lensa dan performa di kondisi low-light . ZenFone 4 Selfie Pro merupakan produk swafoto pamungkas Asus . Tubuh berbahan aluminiumnya dibuat melalui teknik nano molding dan tiap lekukannya dibentuk secara presisi . Ia dibantu sistem electronic image stabilization , phase detection autofocus , serta LED dual - tone flash . Di dalam , Asus mempersenjatai ZD552KL</p>

<p>secara presisi . Dipadu layar berlapis Corning Gorilla Glass 5 2.5D , saya akui penampilan handset ini sangat menawan , terutama untuk varian berwarna merahnya . Sebagai jendela akses konten , smartphone menghadirkan layar AMOLED 1080p berkepadatan 401 ppi seluas 5,5 - inci . Atraksi utama dari ZenFone 4 Selfie Pro tentu saja adalah kamera depannya . Di sana , Asus mencantumkan sistem DuoPixel 24Mp , berisi kombinasi sepasang sensor Sony Exmor RS IMX362 1,4 m 12 - megapixel dengan aperture f/1.8 , ditambah sensor Omnivision 5670 1,12 m f/2.2 yang memiliki lensa wide-angle 120 derajat , sehingga kamera bisa merangkul objek dua kali lebih banyak memungkinkan Anda ber-selfie bersama kawan ataupun keluarga tanpa bantuan monopod . Selain itu , Asus melengkapi ZenFone 4 Selfie Pro dengan bundel perkakas khusus swafoto bernama SelfieMaster . Tool ini berisi fitur-fitur krusial semisal beautify (buat foto maupun video) , serta kolase dan BeautyLive . Produsen juga tak lupa menyiapkan flash LED softlight untuk membantu pengambilan foto di kondisi temaram . Kamera belakangnya sendiri mengandalkan sensor Sony Exmor IMX351 1 m 16 - megapixel berlensa 26 mm f/2.2 . Ia dibantu sistem electronic image stabilization , phase detection autofocus , serta LED dual - tone flash . Di dalam , Asus mempersenjatai ZD552KL dengan chip Qualcomm Snapdragon 625 (ada prosesor octa-core Cortex - A53 2 GHz dan GPU Adreno 506) , RAM sebesar 4 GB , ROM 64 GB , dan baterai 3.000 mAh . Smartphone berjalan di sistem operasi Android 7.1.1 Nougat plus interface ZenUI 4.0 . ZD553KL ialah alternatif lebih terjangkau dari saudaranya di atas . Handset mengusung arahan desain serupa ZenFone 4 Selfie Pro , namun konstruksi tubuhnya terbuat dari plastik , dan layar IPS 5,5 - incinya menyajikan resolusi 720p . Tapi jangan cemas soal penampilannya , smartphone tetap memanfaatkan kaca 2.5D sehingga memberi kesan menyambung pada lekukan di sisi</p>	<p>tentu saja adalah kamera depannya . Selain itu , Asus melengkapi ZenFone 4 Selfie Pro dengan bundel perkakas khusus swafoto bernama SelfieMaster . Ia dibantu sistem electronic image stabilization , phase detection autofocus , serta LED dual - tone flash . Di dalam , Asus mempersenjatai ZD552KL dengan chip Qualcomm Snapdragon 625 (ada prosesor octa-core Cortex - A53 2 GHz dan GPU Adreno 506) , RAM sebesar 4 GB , ROM 64 GB , dan baterai 3.000 mAh . ZD553KL ialah alternatif lebih terjangkau dari saudaranya di atas . Sudut jangkauan jepretan dan sejumlah kelengkapannya tak berbeda dari Selfie 4 Pro . Anda kembali dihindangkan SelfieMaster , flash LED softlight , mode panorama , serta HDR . Handset menyimpan RAM 4 GB , memori internal 64 GB , lalu tenaganya dipasok oleh baterai 3.000 mAh non - removable di dalam . Satu keunikan yang membuat ZD553KL lebih unggul dari ZenFone 4 Selfie Pro adalah dukungan tiga slot kartu : dua untuk SIM , dan satu lagi buat microSD (storage dapat diekspansi sampai 2 TB) . ZenFone 4 Selfie Pro ZD552KL dibanderol seharga Rp 5 juta , sedangkan ZenFone 4 Selfie</p>	<p>dengan chip Qualcomm Snapdragon 625 (ada prosesor octa-core Cortex - A53 2 GHz dan GPU Adreno 506) , RAM sebesar 4 GB , ROM 64 GB , dan baterai 3.000 mAh . ZD553KL ialah alternatif lebih terjangkau dari saudaranya di atas . Sudut jangkauan jepretan dan sejumlah kelengkapannya tak berbeda dari Selfie 4 Pro . Anda kembali dihindangkan SelfieMaster , flash LED softlight , mode panorama , serta HDR . Satu keunikan yang membuat ZD553KL lebih unggul dari ZenFone 4 Selfie Pro adalah dukungan tiga slot kartu : dua untuk SIM , dan satu lagi buat microSD (storage dapat diekspansi sampai 2 TB) .</p>
---	--	--

<p>samping . Kapabilitas swafoto ZenFone 4 Selfie bersandar pada sensor Omnivision 20880 1 m 20 - megapixel f/2.0 dan sensor Omnivision 8856 1,12 m f/2.4 dengan lensa wide-angle 120 derajat . Sudut jangkauan jepretan dan sejumlah kelengkapannya tak berbeda dari Selfie 4 Pro . Anda kembali dihindangkan SelfieMaster , flash LED softlight , mode panorama , serta HDR . Ukuran megapixel kamera belakangnya setara ZenFone 4 Selfie Pro , namun jenis sensornya berbeda . Smartphone tersebut menggunakan Omnivision 16880 1 m 16Mp dengan aperture lensa f/2.2 . Meski demikian , fitur-fitur penunjang fotografi seperti PDAF , EIS , dan flash LED juga tetap ada di sana . ZenFone 4 Selfie diotaki system-on-chip Qualcomm Snapdragon 430 , berisi CPU octa-core Cortex - A53 1,4 GHz dan GPU Adreno 505 . Handset menyimpan RAM 4 GB , memori internal 64 GB , lalu tenaganya dipasok oleh baterai 3.000 mAh non - removable di dalam . Satu keunikan yang membuat ZD553KL lebih unggul dari ZenFone 4 Selfie Pro adalah dukungan tiga slot kartu : dua untuk SIM , dan satu lagi buat microSD (storage dapat diekspansi sampai 2 TB) . Smartphone juga beroperasi di platform Android 7.1.1 Nougat dengan ZenUI 4.0 . ZenFone 4 Selfie Pro ZD552KL dibanderol seharga Rp 5 juta , sedangkan ZenFone 4 Selfie ZD553KL dipatok di harga Rp 3,5 juta . Kedua perangkat sudah mulai dipasarkan mulai tanggal 25 November , dan sampai tanggal 10 November nanti , paket penjualan turut dibundel bersama Gong Yoo Special Gift Box serta speaker Bluetooth khusus ZenFone 4 Selfie Pro selama persediaan masih ada . DailySocial.id adalah portal berita startup dan inovasi teknologi . Kamu bisa menjadi member komunitas startup dan inovasi DailySocial.id , mengunduh laporan riset dan statistik seputar teknologi secara cuma-cuma , dan mengikuti berita startup Indonesia dan gadget terbaru .</p>	<p>ZD553KL dipatok di harga Rp 3,5 juta . Kedua perangkat sudah mulai dipasarkan mulai tanggal 25 November , dan sampai tanggal 10 November nanti , paket penjualan turut dibundel bersama Gong Yoo Special Gift Box serta speaker Bluetooth khusus ZenFone 4 Selfie Pro selama persediaan masih ada . DailySocial.id adalah portal berita startup dan inovasi teknologi . Kamu bisa menjadi member komunitas startup dan inovasi DailySocial.id , mengunduh laporan riset dan statistik seputar teknologi secara cuma-cuma , dan mengikuti berita startup Indonesia dan gadget terbaru .</p>	
--	---	--

3	<p>Jakarta , CNN Indonesia - - Dinas Pariwisata Provinsi Bengkulu kembali menggelar kegiatan Bimbingan Teknis (Bimtek) SDM Kepariwisataaan dalam menyongson " Visit 2020 Wonderful Bengkulu " . Kegiatan yang berlangsung pada 8 hingga 10 November kemarin tersebut sebagai bagian dari upaya Pemerintah Provinsi Bengkulu dalam Hadir sebagai pemateri kegiatan pada 8 - 10 November itu adalah Plt. Asdep Strategi Pemasaran Pariwisata Nusantara , Deputi Bidang Pengembangan Pemasaran Pariwisata Nusantara Hariyanto serta perwakilan dari Deputi Bidang Pengembangan Kelembagaan Kementerian Pariwisata , Faizal . Kepala Dinas Pariwisata Provinsi Bengkulu Yudi Satria mengatakan , kegiatan Bimtek diikuti 250 peserta yang terdiri dari aparatur Pemerintah Provinsi , ASN Kabupaten / Kota , Kelompok Sadar Wisata serta pihak terkait sektor pariwisata di Bengkulu . " Kegiatan ini dimaksudkan untuk memberikan pembekalan kepada peserta di bidang kepariwisataan , " ujar Yudi Satria . Ia mengatakan , Pemprov telah menetapkan pariwisata sebagai salah satu sektor yang akan dikembangkan dan akan menjadi sektor unggulan dalam meningkatkan pertumbuhan ekonomi daerah serta masyarakat . Hal itu , jelas Yudi , tidak lepas dari potensi pariwisata di Bengkulu yang besar memiliki kekayaan alam yang indah serta budaya yang tinggi . " Karena itu pula Pemprov telah menetapkan program ' Visit 2020 Wonderful Bengkulu " yang akan menjadi tujuan besar pariwisata Bengkulu . Salah satu poin utamanya adalah upaya menghasilkan SDM pariwisata yang andal yang akan diwujudkan melalui Bimtek ini , " ujar Yudi . Selain itu , dalam menunjang proses ' Visit 2020 Wonderful Bengkulu ' , Pemprov juga telah menyiapkan 52 acara yang akan digelar dalam satu tahun ke depan yang bertujuan mengangkat potensi lokal ke kelas dunia . " Salah satu yang dirancang secara besar adalah Sail Bengkulu yang akan menjual</p>	<p>Jakarta , CNN Indonesia - - Dinas Pariwisata Provinsi Bengkulu kembali menggelar kegiatan Bimbingan Teknis (Bimtek) SDM Kepariwisataaan dalam menyongson " Visit 2020 Wonderful Bengkulu " . Kegiatan yang berlangsung pada 8 hingga 10 November kemarin tersebut sebagai bagian dari upaya Pemerintah Provinsi Bengkulu dalam Hadir sebagai pemateri kegiatan pada 8 - 10 November itu adalah Plt. Kepala Dinas Pariwisata Provinsi Bengkulu Yudi Satria mengatakan , kegiatan Bimtek diikuti 250 peserta yang terdiri dari aparatur Pemerintah Provinsi , ASN Kabupaten / Kota , Kelompok Sadar Wisata serta pihak terkait sektor pariwisata di Bengkulu . " Ia mengatakan , Pemprov telah menetapkan pariwisata sebagai salah satu sektor yang akan dikembangkan dan akan menjadi sektor unggulan dalam meningkatkan pertumbuhan ekonomi daerah serta masyarakat . Hal itu , jelas Yudi , tidak lepas dari potensi pariwisata di Bengkulu yang besar memiliki kekayaan alam yang indah serta budaya yang tinggi . " Salah satu yang dirancang secara besar adalah Sail Bengkulu yang akan menjual keindahan alam laut Bengkulu yang berhadapan langsung</p>	<p>Jakarta , CNN Indonesia - - Dinas Pariwisata Provinsi Bengkulu kembali menggelar kegiatan Bimbingan Teknis (Bimtek) SDM Kepariwisataaan dalam menyongson " Visit 2020 Wonderful Bengkulu " . Kegiatan yang berlangsung pada 8 hingga 10 November kemarin tersebut sebagai bagian dari upaya Pemerintah Provinsi Bengkulu dalam Hadir sebagai pemateri kegiatan pada 8 - 10 November itu adalah Plt. Ia mengatakan , Pemprov telah menetapkan pariwisata sebagai salah satu sektor yang akan dikembangkan dan akan menjadi sektor unggulan dalam meningkatkan pertumbuhan ekonomi daerah serta masyarakat . Salah satu yang dirancang secara besar adalah Sail Bengkulu yang akan menjual keindahan alam laut Bengkulu yang berhadapan langsung dengan Samudera Hindia , " kata dia . Sementara Hariyanto dalam kesempatan itu menyampaikan materi tentang pengembangan SDM sektor pariwisata dan pengemasan serta pemasaran satu acara . Deputi Pengembangan Pemasaran Pariwisata Nusantara , Esthy Reko Astuti mengatakan , Bimtek kali ini juga bertujuan memberi perspektif dan arah yang sama tentang program</p>
---	---	---	--

<p>keindahan alam laut Bengkulu yang berhadapan langsung dengan Samudera Hindia , " kata dia . Ajang lainnya tentunya Festival Tabot Muharam yang selalu menyedot minat ribuan wisatawan setiap tahun serta Festival Bumi Rafflesia . " Semua kegiatan ini harus disinkronkan , penguatan SDM dan bagaimana promosi pemasarannya agar semua acara ini layak jual untuk wisatawan , " ujarnya . Sementara Hariyanto dalam kesempatan itu menyampaikan materi tentang pengembangan SDM sektor pariwisata dan pengemasan serta pemasaran satu acara . Dalam paparannya ia menjelaskan , bahwa Bimtek penting untuk meningkatkan kompetensi SDM kepariwisataan sehingga mampu berperan dalam peningkatan pembangunan kepariwisataan . Selain itu ia menegaskan bahwa dalam pengembangan pariwisata , promosi juga hal yang harus diperhatikan . Misalnya penekanan pada mekanisme promosi pariwisata secara digital atau daring , penyelenggaraan acara juga harus terkurasi dengan baik sehingga memiliki daya tarik yang kuat . " Bagaimana juga meningkatkan efektivitas partisipasi Dinas Pariwisata pada penyelenggaraan acara dan bagaimana cara memasarkan dan mempromosikan agar bisa dikenal dan menjadi daya tarik wisatawan , terang Hariyanto . Deputi Pengembangan Pemasaran Pariwisata Nusantara , Esthy Reko Astuti mengatakan , Bimtek kali ini juga bertujuan memberi perspektif dan arah yang sama tentang program promosi Pariwisata di Bengkulu . " Selain itu juga untuk memahami potensi destinasi - destinasi wisata di Bengkulu , " ujar Esthy . Menteri Pariwisata Arief Yahya mengapresiasi kegiatan Bimtek sebagai salah satu upaya dan komitmen dari Pemerintah Provinsi Bengkulu dalam mewujudkan pariwisata sebagai salah satu sektor utama . Ia pun berkomitmen akan mendukung Penprov dalam mewujudkan " Visit 2020 Wonderful Bengkulu " . " Obyek wisata andalan Bengkulu , Benteng Marlborough , Rumah Bung Karno ,</p>	<p>dengan Samudera Hindia , " kata dia . Sementara Hariyanto dalam kesempatan itu menyampaikan materi tentang pengembangan SDM sektor pariwisata dan pengemasan serta pemasaran satu acara . Selain itu ia menegaskan bahwa dalam pengembangan pariwisata , promosi juga hal yang harus diperhatikan . Bagaimana juga meningkatkan efektivitas partisipasi Dinas Pariwisata pada penyelenggaraan acara dan bagaimana cara memasarkan dan mempromosikan agar bisa dikenal dan menjadi daya tarik wisatawan , terang Hariyanto . Deputi Pengembangan Pemasaran Pariwisata Nusantara , Esthy Reko Astuti mengatakan , Bimtek kali ini juga bertujuan memberi perspektif dan arah yang sama tentang program promosi Pariwisata di Bengkulu . " Selain itu juga untuk memahami potensi destinasi - destinasi wisata di Bengkulu , " ujar Esthy . Obyek wisata andalan Bengkulu , Benteng Marlborough , Rumah Bung Karno , Pantai Panjang yang memiliki pasir putih yang indah dan bersih .</p>	<p>promosi Pariwisata di Bengkulu . " Selain itu juga untuk memahami potensi destinasi - destinasi wisata di Bengkulu , " ujar Esthy .</p>
---	---	--

	Pantai Panjang yang memiliki pasir putih yang indah dan bersih . Semuanya world class , ditambah dengan puluhan atraksi didalamnya , tinggal kita akan dukung dan promosikan ' Bengkulu Visit ' sehingga lebih mendunia , " kata Arief (syahb)		
...
14260	Merdeka.com - Sebuah kabar gembira datang bagi sepakbola Indonesia . Posisi Timnas Indonesia di ranking FIFA terbaru mengalami peningkatan , di mana Skuat Garuda naik dua peringkat dibanding capaian bulan lalu . FIFA selaku otoritas sepakbola tertinggi dunia baru-baru ini merilis ranking FIFA untuk periode bulan Juni . Dalam ranking tersebut , Indonesia menunjukkan prestasi yang bagus dengan naik dua peringkat . Indonesia saat ini berada di peringkat 175 dengan mengoleksi 105 poin . Peringkat ini naik dua peringkat dari bulan lalu , di mana Indonesia menduduki peringkat ke 177 . Meski naik dua peringkat , pencapaian Indonesia ini masih terbilang kalah dari negara-negara tetangga di ASEAN . Sebagai contoh , Indonesia masih kalah satu peringkat dari Kamboja yang berada di peringkat 174 dengan raihan 107 poin . Indonesia juga tertinggal dari Malaysia yang berada di peringkat 155 dengan raihan 145 poin . Sementara itu ranking tertinggi negara ASEAN masih dipegang oleh Filipina yang berada di ranking 126 dengan raihan 251 poin . Sementara itu Thailand berada di peringkat 127 dengan raihan 249 poin . Semoga dengan naiknya peringkat Indonesia ini memacu para pemain Timnas kita untuk bermain lebih baik di masa depan . (FIFA / dub)	Indonesia saat ini berada di peringkat 175 dengan mengoleksi 105 poin . Meski naik dua peringkat , pencapaian Indonesia ini masih terbilang kalah dari negara-negara tetangga di ASEAN . Sebagai contoh , Indonesia masih kalah satu peringkat dari Kamboja yang berada di peringkat 174 dengan raihan 107 poin . Indonesia juga tertinggal dari Malaysia yang berada di peringkat 155 dengan raihan 145 poin . Sementara itu Thailand berada di peringkat 127 dengan raihan 249 poin . (FIFA / dub)	Indonesia saat ini berada di peringkat 175 dengan mengoleksi 105 poin . Sebagai contoh , Indonesia masih kalah satu peringkat dari Kamboja yang berada di peringkat 174 dengan raihan 107 poin . (FIFA / dub)
14261	Ada satu perangkat menarik lainnya yang tersaji dalam suguhan keynote Apple yang digelar semalam , yaitu Apple TV , perangkat rumah yang menawarkan lebih banyak hiburan nan memanjakan . Mata menjadi salah satu bagian	Ada satu perangkat menarik lainnya yang tersaji dalam suguhan keynote Apple yang digelar semalam , yaitu Apple TV , perangkat rumah yang	Bermodalkan chipset yang lebih baik , maka performa baik core maupun grafisnya tentu akan melampaui apa yang ditawarkan oleh generasi Apple

<p>vital pengguna yang akan paling dimanjakan oleh Apple TV , berkat adanya dukungan resolusi 4 K HDR super bening , lebih bening dari resolusi 4K. Generasi kelima Apple TV ini menyuguhkan sisi visual yang sangat optimal , menampilkan setiap lekukan objek dalam resolusi terbaik . Gambar yang dihasilkan akan lebih tajam , lebih detail , dan nyata . Dampaknya sudah barang tentu memberi rasa nyaman sekaligus memuaskan tuntutan sisi visual pengguna modern . Untuk mengimbangi kemampuan tersebut , Apple membenamkan chip Apple A10X yang juga diadopsi oleh iPad Pro . Bermodalkan chipset yang lebih baik , maka performa baik core maupun grafisnya tentu akan melampaui apa yang ditawarkan oleh generasi Apple TV sebelumnya . Secara default , Apple TV akan menampilkan gambar dalam resolusi terbaik . Namun pengguna tetap mempunyai keleluasaan untuk menentukan pilihan yang menurutnya paling pas untuk perangkat televisi yang mereka punyai . Sementara di bagian piranti lunak , Apple masih menggunakan tvOS namun dengan versi yang sudah ditingkatkan . Peningkatan dapat dijumpai di beberapa bagian , antara lain bagian tatap muka yang mengakomodasi konten - konten 4 K dan HDR . Kemudian dukungan sejumlah aplikasi baru , salah satunya aplikasi siaran langsung program-program olahraga . Integrasi asisten pintar Siri membuat Apple TV semakin fungsional dan cerdas . Selain remote yang dirombak , Siri dapat membantu pengguna menemukan cara yang paling mudah untuk memenuhi keinginan mereka , mulai dari film , musik , hiburan , dan lain-lain hanya dengan input suara . Aplikasi TV mendukung lebih dari 60 layanan di Apple TV dan perangkat iOS . Memungkinkan pengguna untuk membagikan foto dan video dari iPad atau iPhone ke layar yang lebih lega . AirPlay juga tersedia bagi yang ingin menonton film yang tersimpan di iPhone atau iPad . Pre - order Apple TV 4 K akan mulai</p>	<p>menawarkan lebih banyak hiburan nan memanjakan . Mata menjadi salah satu bagian vital pengguna yang akan paling dimanjakan oleh Apple TV , berkat adanya dukungan resolusi 4 K HDR super bening , lebih bening dari resolusi 4K. Bermodalkan chipset yang lebih baik , maka performa baik core maupun grafisnya tentu akan melampaui apa yang ditawarkan oleh generasi Apple TV sebelumnya . Secara default , Apple TV akan menampilkan gambar dalam resolusi terbaik . Namun pengguna tetap mempunyai keleluasaan untuk menentukan pilihan yang menurutnya paling pas untuk perangkat televisi yang mereka punyai . Sementara di bagian piranti lunak , Apple masih menggunakan tvOS namun dengan versi yang sudah ditingkatkan . Kemudian dukungan sejumlah aplikasi baru , salah satunya aplikasi siaran langsung program-program olahraga . Integrasi asisten pintar Siri membuat Apple TV semakin fungsional dan cerdas . Selain remote yang dirombak , Siri dapat membantu pengguna menemukan cara yang paling mudah untuk memenuhi keinginan mereka , mulai dari film , musik , hiburan , dan lain-lain hanya dengan input suara .</p>	<p>TV sebelumnya . Secara default , Apple TV akan menampilkan gambar dalam resolusi terbaik . Sementara di bagian piranti lunak , Apple masih menggunakan tvOS namun dengan versi yang sudah ditingkatkan . Kemudian dukungan sejumlah aplikasi siaran langsung program-program olahraga . Integrasi asisten pintar Siri membuat Apple TV semakin fungsional dan cerdas . Aplikasi TV mendukung lebih dari 60 layanan di Apple TV dan perangkat iOS .</p>
---	--	---

	<p>dibuka pada tanggal 15 September mendatang . Harga jualnya di kisaran US\$ 179 untuk varian 32 GB dan \$ 199 untuk varian 64 GB . Sumber berita Apple . DailySocial.id adalah portal berita startup dan inovasi teknologi . Kamu bisa menjadi member komunitas startup dan inovasi DailySocial.id , mengunduh laporan riset dan statistik seputar teknologi secara cuma-cuma , dan mengikuti berita startup Indonesia dan gadget terbaru .</p>	<p>Aplikasi TV mendukung lebih dari 60 layanan di Apple TV dan perangkat iOS . AirPlay juga tersedia bagi yang ingin menonton film yang tersimpan di iPhone atau iPad .</p>	
--	---	---	--