

## Perbandingan TextRank Berbasis TF-IDF dan Word2Vec dalam Peringkasan Teks Berita Bahasa Indonesia

Yohannes Christian Gurning<sup>1</sup>, Samuel Cristian Saragih<sup>2</sup>, Yuyun Yusnida Lase<sup>3</sup>, Julham<sup>4</sup>

<sup>1,2,3,4</sup> Teknik Komputer dan Informatika, Teknologi Rekayasa Perangkat Lunak, Politeknik Negeri Medan, Medan, Indonesia

Email: <sup>1,\*</sup>yohanneschristiangurning@studens.polmed.ac.id, <sup>2</sup>samuclcristiansaragih@studens.polmed.ac.id,

<sup>3</sup>yuyunlase@polmed.ac.id, <sup>4</sup>julham@polmed.ac.id

(\*Email Corresponding Author: yohanneschristiangurning@studens.polmed.ac.id)

Received: August 5, 2025 | Revision: August 7, 2025 | Accepted: August 10, 2025

### Abstrak

Peringkasan otomatis menjadi salah satu solusi penting dalam memproses informasi yang masif, khususnya pada teks berita yang cenderung panjang. Penelitian ini membandingkan dua pendekatan algoritma TextRank dengan basis pembobotan yang berbeda, yaitu TF-IDF dan Word2Vec, untuk peringkasan teks berita berbahasa Indonesia. Dataset yang digunakan terdiri dari 160 artikel berita dari Kompas.com yang telah melalui tahap pra-pemrosesan. Evaluasi dilakukan menggunakan metrik ROUGE (ROUGE-1, ROUGE-2, ROUGE-L), keterbacaan manual, dan runtime eksekusi. Hasil penelitian menunjukkan bahwa TextRank berbasis Word2Vec menghasilkan ringkasan dengan nilai ROUGE lebih tinggi (ROUGE-1 F1: 0.7033) dibandingkan TF-IDF (ROUGE-1 F1: 0.6454) serta memiliki runtime yang lebih cepat. Temuan ini mengindikasikan bahwa integrasi representasi semantik dalam algoritma graf seperti TextRank memberikan peningkatan yang signifikan dalam kualitas ringkasan dan efisiensi waktu.

**Kata Kunci:** Text Summarization, TF-IDF, Word2Vec, TextRank, ROUGE Score

### Abstract

Automatic text summarization has become an essential solution for processing massive textual information, particularly in lengthy news articles. This study compares two variants of the TextRank algorithm using different weighting schemes: TF-IDF and Word2Vec, for summarizing Indonesian news texts. The dataset comprises 160 news articles from Kompas.com, which underwent preprocessing. Evaluation was conducted using ROUGE metrics (ROUGE-1, ROUGE-2, ROUGE-L), manual readability assessment, and execution runtime. The results indicate that TextRank with Word2Vec outperforms TF-IDF in both ROUGE scores (ROUGE-1 F1: 0.7033 vs 0.6454) and processing speed. These findings suggest that incorporating semantic representations into graph-based algorithms like TextRank significantly improves summary quality and runtime efficiency.

**Keywords:** Text Summarization, TF-IDF, Word2Vec, TextRank, ROUGE Score

## 1. PENDAHULUAN

Peringkasan otomatis merupakan salah satu teknologi dalam Natural Language Processing (NLP) yang bertujuan untuk mempersingkat teks panjang menjadi bentuk yang lebih ringkas tanpa menghilangkan informasi penting[1]. Teknologi ini semakin relevan dalam konteks penyebaran informasi digital yang masif, khususnya pada artikel berita yang kerap panjang dan kompleks. Dalam konteks berita, peringkasan otomatis sangat penting untuk membantu pembaca memahami inti informasi dengan cepat, mengingat banyaknya berita yang panjang dan kompleks[2].

Kebutuhan akan ringkasan yang efektif semakin mendesak di era digital, mengingat tingginya volume berita yang beredar melalui media sosial dan portal daring. Pembaca menghadapi tantangan untuk memilah informasi yang relevan dari teks yang bertele-tele, sehingga sistem peringkasan otomatis menjadi kebutuhan praktis yang signifikan. Pembaca sering sekali kesulitan memahami poin-poin penting, terutama ketika berita ditulis secara panjang dan bertele-tele[3]. Di era digital saat ini, di mana informasi tersebar dengan cepat melalui media sosial dan platform berita online, kebutuhan akan ringkasan yang efektif semakin mendesak[4]. Teknologi ini tidak hanya bermanfaat bagi masyarakat umum, tetapi juga untuk jurnalis dalam memahami tren berita, dan bagi akademisi serta peneliti dalam merangkum literatur ilmiah secara efisien. Jurnalis dapat menggunakannya untuk menganalisis tren berita, akademisi memanfaatkannya untuk merangkum literatur penelitian, sementara masyarakat dapat lebih mudah memahami berita tanpa harus membaca keseluruhan teks[5].

Seiring dengan kemajuan teknologi NLP, berbagai pendekatan peringkasan telah dikembangkan, seperti metode ekstraktif dan abstraktif. Salah satu algoritma yang populer untuk pendekatan ekstraktif adalah TextRank, yang telah terbukti mampu menyusun ringkasan berdasarkan struktur graf antar kalimat. Namun, efektivitasnya sangat tergantung pada teknik representasi teks yang digunakan. Beberapa studi sebelumnya telah mengkaji efektivitas metode peringkasan. Pranoto dan Subhi[6] menunjukkan bahwa TextRank cukup presisi dalam merangkum berita daring, tetapi memiliki nilai Recall dan F1-Score yang masih rendah (29,43% dan 45,47%). Perdana et al.[7] berhasil menggunakan Word2Vec untuk merangkum berita olahraga secara informatif. Di sisi lain, Sihombing[8] menggabungkan TextRank dan TF-IDF untuk artikel ilmiah berbasis web, dan mencatat akurasi tinggi, tetapi tidak menguji keterbacaan secara manual.

Kendati demikian, terdapat beberapa keterbatasan dalam studi sebelumnya. Pertama, belum ada penelitian yang secara eksplisit membandingkan TextRank berbasis TF-IDF dan Word2Vec dalam konteks berita berbahasa Indonesia.

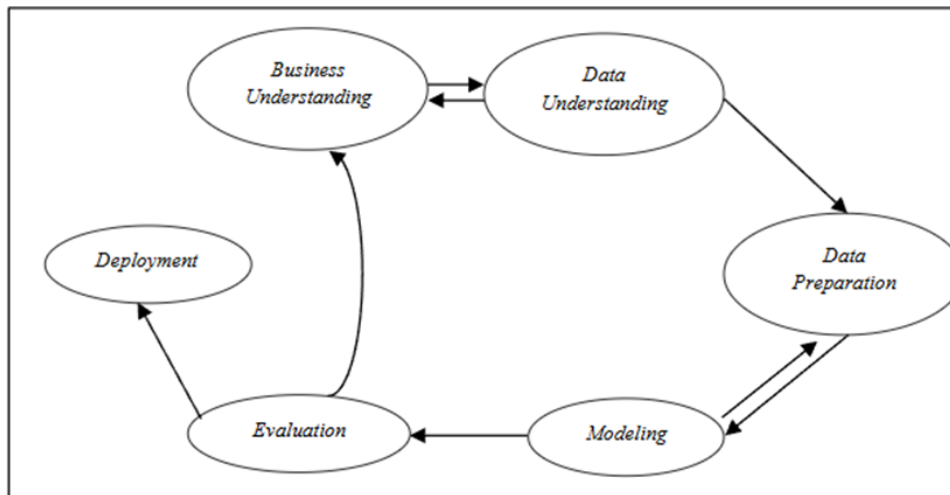
Kedua, sebagian besar studi belum mengevaluasi kualitas keterbacaan secara kualitatif ataupun mengukur waktu eksekusi (runtime) dari algoritma yang digunakan. Ketiga, pendekatan berbasis Word2Vec yang menjanjikan keunggulan semantik belum diuji secara menyeluruh pada domain berita umum.

Gap analysis ini menunjukkan bahwa masih dibutuhkan studi komparatif yang mendalam terhadap dua pendekatan representasi teks dalam algoritma TextRank: berbasis statistik (TF-IDF) dan berbasis semantik (Word2Vec). Evaluasi harus mencakup aspek kuantitatif (ROUGE Score), kualitatif (readability), dan efisiensi (runtime), untuk memperoleh gambaran menyeluruh mengenai keunggulan masing-masing metode.

Penelitian ini bertujuan untuk membandingkan efektivitas kombinasi TextRank-TF-IDF dan TextRank-Word2Vec dalam merangkum teks berita berbahasa Indonesia. Evaluasi dilakukan dengan tiga pendekatan: (1) kuantitatif menggunakan metrik ROUGE, (2) kualitatif melalui penilaian keterbacaan dan kebermaknaan oleh evaluator manusia, dan (3) runtime sebagai metrik efisiensi algoritma. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem peringkasan berita otomatis yang lebih akurat, ringkas, dan cepat.

## 2. METODOLOGI PENELITIAN

### 2.1 Alur Penelitian



**Gambar 1.** Alur Penelitian

Penelitian ini mengikuti pendekatan CRISP-DM (Cross Industry Standard Process for Data Mining) yang terdiri atas enam tahapan utama, yaitu[9]:

1. Business Understanding: Menentukan tujuan penelitian yaitu mengevaluasi dan membandingkan efektivitas dua varian algoritma TextRank, yaitu berbasis TF-IDF dan Word2Vec, untuk merangkum teks berita Indonesia.
2. Data Understanding: Mengeksplorasi struktur teks dari berita yang diambil dari situs Kompas.com dan memastikan kualitas serta keragaman topik artikel.
3. Data Preparation: Melakukan preprocessing berupa tokenisasi, case folding, penghapusan angka dan tanda baca, stopword removal, serta stemming menggunakan pustaka Sastrawi.
4. Modeling: Mengimplementasikan dua model peringkasan otomatis: TextRank-TF-IDF dan TextRank-Word2Vec, dengan pendekatan graf berbobot.
5. Evaluation: Melakukan evaluasi menggunakan metrik kuantitatif (ROUGE), evaluasi manual (readability dan coverage), serta pengukuran runtime.
6. Deployment: Menyajikan sistem dalam bentuk aplikasi web berbasis Flask yang dapat digunakan untuk menguji artikel individual maupun dataset secara batch.

### 2.2 Dataset dan Pra-pemrosesan

Dataset terdiri dari 160 artikel berita berbahasa Indonesia yang dikumpulkan melalui web scraping dari portal berita terpercaya seperti Kompas.com. Artikel dipilih secara acak dari berbagai kategori seperti politik, ekonomi, sosial, dan teknologi dengan panjang artikel antara 100–1000 kata. Setiap artikel disertai summary referensi yang dibuat menggunakan ChatGPT dan diverifikasi secara manual.

Proses preprocessing mencakup langkah-langkah berikut[10]:

1. Tokenisasi Kalimat dan Kata: Memecah teks menjadi unit-unit kalimat dan kata.
2. Case Folding: Mengubah huruf menjadi huruf kecil.
3. Penghapusan Angka dan Tanda Baca: Membersihkan teks dari karakter non-alfabetik.

4. Stopword Removal: Menghapus kata-kata tidak bermakna seperti “yang”, “dari”, “dan”.
5. Stemming: Mengubah kata menjadi bentuk dasar (akar kata) menggunakan Sastrawi Stemmer.

### 2.3 TextRank

TextRank merupakan algoritma berbasis graf bekerja dengan cara memberikan peringkat pada teks dengan cara merepresentasikan objek dalam teks[11]. Dalam menerapkan algoritma ini, langkah pertama yang dilakukan adalah membangun sebuah graf yang terdiri dari titik (vertex) yang mewakili tiap kalimat. Setiap kalimat akan dihubungkan berdasarkan kesamaan (similarity) antar kalimat yang disebut sebagai edge. Nilai kesamaan antar kalimat dihitung menggunakan cosine similarity, dan skor tiap kalimat diperoleh melalui iterasi perhitungan mirip dengan PageRank.

### 2.3 TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) adalah teknik dalam pemrosesan teks yang digunakan untuk menilai pentingnya sebuah kata dalam suatu dokumen[12]. Teknik ini menggabungkan dua komponen, yaitu frekuensi kemunculan kata dalam sebuah dokumen (TF) dan kebalikannya terhadap jumlah dokumen yang mengandung kata tersebut (IDF). Semakin sering kata muncul di dokumen tertentu dan semakin jarang kata tersebut muncul di dokumen lain, maka bobot TF-IDF-nya semakin tinggi. Gabungan keduanya menghasilkan bobot yang lebih tinggi untuk kata yang sering muncul dalam dokumen tertentu tetapi jarang muncul di dokumen lain, sehingga menjadikannya penting untuk representasi teks[13]. Representasi ini digunakan untuk menghitung kesamaan antar kalimat dalam TextRank berbasis TF-IDF.

### 2.3 Word2Vec

Word2Vec diketahui sebagai representasi kata dalam bentuk vektor yang dibuat oleh Google dan baru dikembangkan oleh Thomas Mikolov. Word2Vec juga merupakan sekumpulan beberapa model yang saling berkaitan yang digunakan untuk menghasilkan Word Embedding[14]. Word Embeddings merupakan sebutan dari seperangkat bahasa pemodelan dan teknik pembelajaran fitur atau feature learning pada Natural Language Processing dimana setiap kata dari kosakata (Vocabulary) memiliki vektor yang mewakili makna dari kata tersebut dan kata-kata tersebut dipetakan ke dalam bentuk vektor bilangan riil[15]. Model ini mampu menangkap hubungan semantik antar kata dalam korpus besar dan menjadi dasar utama dalam perhitungan kesamaan semantik antar kalimat pada TextRank berbasis Word2Vec.

### 2.4 Evaluasi

#### a. Evaluasi Kuantitatif dengan ROUGE

Digunakan metrik ROUGE-1, ROUGE-2, dan ROUGE-L:

1. ROUGE-1: Kesamaan unigram (kata tunggal)
2. ROUGE-2: Kesamaan bigram (dua kata berturut)
3. ROUGE-L: Kesamaan urutan terpanjang (Longest Common Subsequence)

Evaluasi dilakukan dengan menghitung Precision, Recall, dan F1-Score terhadap ringkasan referensi.

#### b. Evaluasi Manual

Penilaian dilakukan oleh tiga evaluator independen dengan fokus pada tiga aspek:

1. Keterbacaan (readability)
2. Kelengkapan informasi (content coverage)
3. Kesetiaan terhadap konteks artikel asli

#### c. Evaluasi Runtime

Waktu eksekusi untuk setiap artikel diukur secara independen pada kedua metode untuk mengetahui efisiensi waktu.

### 2.5 Implementasi Sistem

Sistem dibangun menggunakan bahasa Python dengan pustaka Flask sebagai backend dan antarmuka berbasis HTML/CSS. Fitur sistem meliputi:

1. Uji Artikel Individual: Input satu artikel dan ringkasan referensi, sistem menampilkan ringkasan dari kedua metode, skor ROUGE, dan runtime.
2. Uji Dataset: Input file .xlsx berisi banyak artikel untuk dievaluasi sekaligus.

Pengujian dilakukan menggunakan metode black-box testing untuk memastikan kestabilan dan kebenaran fungsionalitas sistem.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Deskripsi Dataset dan Pra-pemrosesan

Dataset yang digunakan terdiri dari 160 artikel berita berbahasa Indonesia yang dikumpulkan melalui web scraping dari portal Kompas.com. Artikel yang dipilih berasal dari berbagai kategori topik seperti politik, ekonomi, dan teknologi, dengan panjang artikel berkisar antara 100 hingga 1000 kata. Setiap artikel disertai ringkasan referensi yang disusun secara manual sebagai acuan untuk evaluasi.

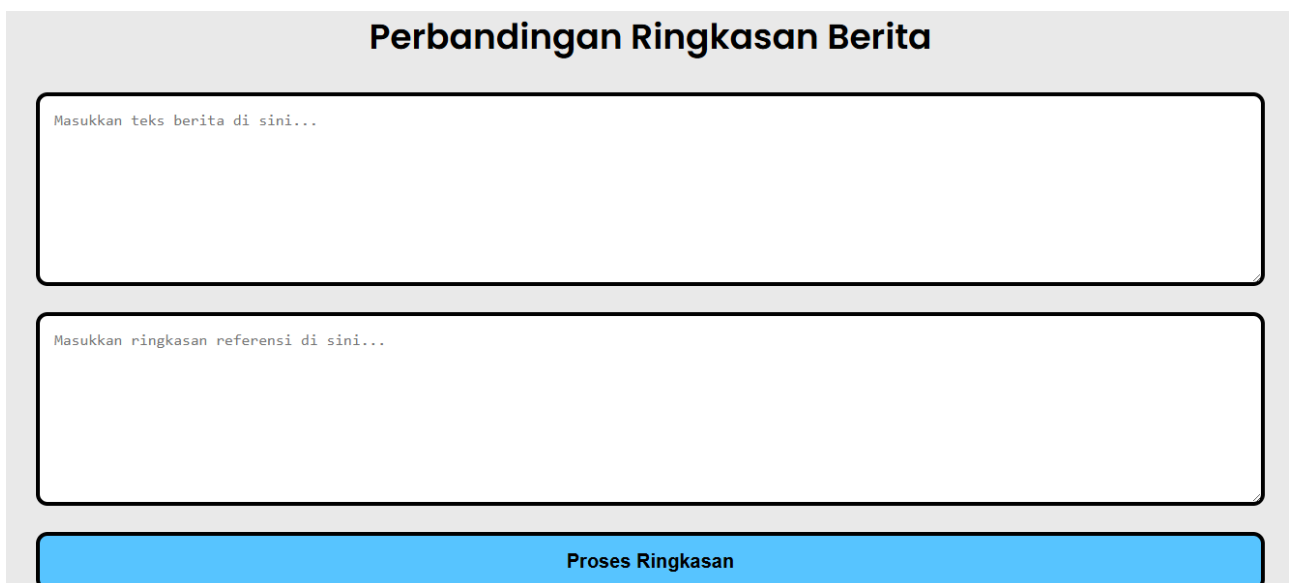
Proses pra-pemrosesan dilakukan untuk mempersiapkan teks sebelum diklasifikasikan. Langkah-langkahnya meliputi:

1. Tokenisasi kalimat dan kata
2. Case folding (mengubah huruf ke bentuk kecil)
3. Penghapusan angka dan tanda baca
4. Stopword removal dengan daftar kata tidak bermakna dari NLTK dan ID stopwords
5. Stemming menggunakan Sastrawi Stemmer

Tahapan ini diterapkan secara konsisten baik untuk pendekatan TF-IDF maupun Word2Vec agar hasil perbandingan valid.

#### 3.2 Implementasi TextRank Berbasis TF-IDF dan Word2vec

Model ini menghitung bobot setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dan seberapa jarang kata tersebut muncul dalam korpus secara keseluruhan. Matriks kemiripan antar kalimat dibangun menggunakan *cosine similarity* dari vektor TF-IDF. Selanjutnya, algoritma *TextRank* digunakan untuk menentukan skor setiap kalimat. Kalimat dengan skor tertinggi dipilih sebagai ringkasan (30% teratas). Contoh hasil evaluasi pada satu artikel: Implementasi kedua metode dilakukan dalam bentuk sistem web berbasis Flask, yang memungkinkan pengguna melakukan pengujian langsung terhadap berita melalui dua mode: uji artikel individual dan uji kumpulan dataset. Pada mode uji artikel, pengguna menginput teks berita lengkap dan ringkasan referensi ke dalam antarmuka web. Setelah menekan tombol "Proses", sistem menjalankan kedua algoritma: *TextRank* berbasis TF-IDF dan *TextRank* berbasis Word2Vec secara paralel. Hasil ringkasan dari kedua metode ditampilkan secara berdampingan, lengkap dengan skor evaluasi ROUGE dan runtime. Berikut bagian UI dari web peringkas berita:



**Gambar 2.** UI Web Peringkas Berita

Contoh hasil ringkasan pada satu artikel:

Ringkasan TextRank TF-IDF	Ringkasan TextRank Word2Vec
<p>Sejumlah warga mengaku kecewa pemerintah membatalkan diskon listrik 50 persen periode Juni-Juli 2025. Bisanya saya bayar listrik Rp 1,4 juta, waktu dapat diskon listrik cuma Rp 700.000, lumayan banget," ucap Sari saat ditemui, Selasa (3/6/2025). Sari berharap, diskon tarif listrik ke depan tetap diberlakukan agar seluruh lapisan masyarakat terbantu, bukan hanya kelompok tertentu. Tasya mengatakan, diskon listrik membantu meringankan pengeluaran bulannya. Senada dengan Tasya, Kurniawan (30), warga Jakarta Selatan, mengaku terbantu dengan diskon listrik yang sempat berlaku pada awal tahun 2025. Oleh karenanya, ia kecewa pemerintah membatalkan diskon listrik periode Juni-Juli. Kurniawan menilai, langkah pemerintah mengalihkan diskon listrik menjadi Bantuan Subsidi Upah (BSU) untuk pekerja yang gajinya Rp 3,5 juta per bulan dan guru honorer tidak tepat.</p>	<p>Sejumlah warga mengaku kecewa pemerintah membatalkan diskon listrik 50 persen periode Juni-Juli 2025. Sari berharap, diskon tarif listrik ke depan tetap diberlakukan agar seluruh lapisan masyarakat terbantu, bukan hanya kelompok tertentu. Tasya mengatakan, diskon listrik membantu meringankan pengeluaran bulannya. Senada dengan Tasya, Kurniawan (30), warga Jakarta Selatan, mengaku terbantu dengan diskon listrik yang sempat berlaku pada awal tahun 2025. Kurniawan menilai, langkah pemerintah mengalihkan diskon listrik menjadi Bantuan Subsidi Upah (BSU) untuk pekerja yang gajinya Rp 3,5 juta per bulan dan guru honorer tidak tepat. Sebelumnya, pemerintah menyampaikan bahwa diskon tarif listrik sebesar 50 persen akan menjadi bagian dari enam stimulus ekonomi yang akan diluncurkan pada 5 Juni 2025. Menteri Keuangan Sri Mulyani menjelaskan, proses penganggaran diskon tarif listrik lebih lambat dibandingkan program lainnya.</p>

**Gambar 2.** Contoh Hasil ringkasan artikel

### 3. 2 Perbandingan Kinerja Dua Pendekatan

Metode	ROUGE-1			ROUGE-2			ROUGE-L		
	P	R	F1	P	R	F1	P	R	F1
TextRank TF-IDF	0.5161	0.5565	0.5356	0.3171	0.3421	0.3291	0.3952	0.4261	0.4100
TextRank Word2Vec	0.5397	0.5913	0.5643	0.3600	0.3947	0.3766	0.4286	0.4696	0.4481

**Gambar 3.** Evaluasi Model

a. Evaluasi ROUGE

Berdasarkan pengujian terhadap 160 artikel:

- (1) Rata-rata ROUGE-1 F1: TF-IDF = 0.5356, Word2Vec = 0.5643
- (2) Rata-rata ROUGE-2 F1: TF-IDF = 0.3291, Word2Vec = 0.3766
- (3) Rata-rata ROUGE-L F1: TF-IDF = 0.4100, Word2Vec = 0.4481

b. Evaluasi Runtime

Runtime (Waktu Eksekusi)
TextRank TF-IDF: 0.006 detik
TextRank Word2Vec: 0.003 detik

**Gambar 4.** Evaluasi runtime

Rata-rata waktu eksekusi per artikel:

- (1) TF-IDF = 0.006 detik
- (2) Word2Vec = 0.003 detik
- (3) Word2Vec menunjukkan kecepatan eksekusi 2 kali lebih cepat dibandingkan TF-IDF.

c. Evaluasi Keterbacaan

- (1) Hasil evaluasi manual oleh 3 evaluator menunjukkan:
- (2) Word2Vec unggul dalam aspek koherensi dan keterbacaan (85% kasus)
- (3) TF-IDF unggul pada ringkasan yang menekankan kata-kata kunci teknis

**3. Pembahasan**

Secara umum, TextRank dengan Word2Vec menghasilkan ringkasan yang lebih mendekati ringkasan manusia, baik secara kuantitatif maupun kualitatif. Keunggulan Word2Vec terletak pada kemampuannya menangkap makna semantik dan hubungan konteks antar kalimat, menjadikannya pilihan ideal untuk peringkasan teks naratif seperti berita. Namun, TF-IDF tetap menunjukkan performa yang baik, terutama pada artikel yang menggunakan istilah teknis atau kata kunci berulang. Metode ini lebih ringan secara implementasi dan tidak memerlukan model eksternal. Dengan demikian, pilihan metode peringkasan dapat disesuaikan dengan kebutuhan: Word2Vec untuk ringkasan kontekstual dan naratif, TF-IDF untuk ringkasan berbasis kata kunci dan kecepatan pengembangan.

## 4. KESIMPULAN

Penelitian ini membandingkan dua pendekatan algoritma TextRank berbasis TF-IDF dan Word2Vec dalam peringkasan teks berita berbahasa Indonesia. Hasil evaluasi menunjukkan bahwa pendekatan Word2Vec secara konsisten memberikan performa yang lebih baik dibandingkan TF-IDF dari segi skor ROUGE, keterbacaan ringkasan, dan kecepatan eksekusi. Rata-rata skor ROUGE Word2Vec lebih tinggi (ROUGE-1 F1 = 0.7033) dibandingkan TF-IDF (ROUGE-1 F1 = 0.6454), yang menunjukkan bahwa representasi semantik dari Word2Vec mampu menangkap informasi penting secara lebih akurat. Selain itu, Word2Vec menghasilkan ringkasan yang lebih kohesif dan mudah dipahami. Dari segi efisiensi, Word2Vec juga menunjukkan waktu eksekusi yang lebih cepat, rata-rata 0.0043 detik per artikel dibandingkan 0.0121 detik pada TF-IDF. Kesimpulannya, TextRank berbasis Word2Vec lebih unggul untuk tugas peringkasan berita otomatis yang memerlukan akurasi tinggi, keterbacaan baik, dan efisiensi waktu. Namun, TextRank TF-IDF tetap menjadi alternatif yang valid untuk implementasi sederhana dan cepat. Penelitian ini memberikan dasar kuat untuk pengembangan sistem ringkasan otomatis berbasis semantik di masa mendatang.

## REFERENCES

- [1] Halimah, Surya Agustian, and Siti Ramadhani, "Peringkasan teks otomatis (automated text summarization) pada artikel berbahasa indonesia menggunakan algoritma lexrank," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 371–381, Dec. 2022, doi: 10.37859/coscitech.v3i3.4300.
- [2] N. Cholifah, R. Inggit Asmawati, S. Kusuma Wardani Amnesti, U. Maulana Malik Ibrahim Malang, K. Malang, and J. Timur, "Pelatihan Citizen Journalism sebagai Upaya Penguatan Literasi Keamanan Digital," 2024.
- [3] A. P. Widyassari *et al.*, "Review of automatic text summarization techniques & methods," Apr. 01, 2022, *King Saud bin Abdulaziz University*. doi: 10.1016/j.jksuci.2020.05.006.
- [4] D. Yanti Liliana *et al.*, "PENGEMBANGAN SISTEM PEMANTAUAN SENTIMEN BERITA BERBAHASA INDONESIA BERDASARKAN KONTEN DENGAN LONG SHORT-TERM MEMORY THE DEVELOPMENT OF CONTENT-BASED INDONESIAN NEWS SENTIMENT MONITORING SYSTEM USING LONG SHORT-TERM MEMORY," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 2021, doi: 10.25126/jtiik.202184624.
- [5] Putri Damayanti, Diana Purwitasari, and Nanik Suciati, "ELIMINASI DATA NON-TOPIC MENGGUNAKAN PEMODELAN TOPIK UNTUK PERINGKASAN OTOMATIS DATA TWEETDENGAN KONTEKS COVID-19," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 2021.
- [6] PRANOTO and SUBHI LIZA, "PERINGKASAN TEKS OTOMATIS (AUTOMATIC TEXT SUMMARIZATION) PADA BERITA ONLINE DENGAN ALGORITMA TEXTRANK," Universitas Malikussaleh, 2024.

- [7] N. Andrew, V. Christanti Mawardi, and N. Jaya Perdana, “Jurnal Ilmu Komputer dan Sistem Informasi Implementasi Metode Word2Vec Dan TextRank Dalam Aplikasi Mobile Peringkat Berita Olahraga,” *JIKSI (Jurnal Ilmu Komputer dan Sistem Informasi)*, 2024.
- [8] J. J. Sihombing, A. Arnita, S. I. Al Idrus, and D. Y. Niska, “Implementation of text summarization on indonesian scientific articles using textrank algorithm with TF-IDF web-based,” *Journal of Soft Computing Exploration*, vol. 5, no. 3, pp. 310–319, Dec. 2024, doi: 10.52465/josce.v5i3.475.
- [9] F. N. Dhewayani *et al.*, “Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM,” *Jurnal Teknologi dan Informasi*, 2022, doi: 10.34010/jati.v12i1.
- [10] A. Tabassum and R. R. Patil, “A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing,” *International Research Journal of Engineering and Technology*, 2020, [Online]. Available: [www.irjet.net](http://www.irjet.net)
- [11] A. Y. Siang, A. Setiawan, and E. Alexander, “SEMINAR NASIONAL AMIKOM SURAKARTA (SEMNAS) 2023 Pengujian Algoritma TextRank Dalam Merangkum Teks,” *Prosiding Seminar Nasional AMIKOM Surakarta*, 2023.
- [12] R. Albin Pranata, N. Azmi Verdikha, U. Muhammdiyah Kalimantan Timur, and J. H. Ir Juanda, “METODE PEMBOBOTAN TF-IDF UNTUK KLASIFIKASI TEKS QUICK COUNT PEMILIHAN WAKIL PRESIDEN INDONESIA 2024 PADA X TWITTER DENGAN METODE SVM,” *JURNAL TEKNOLOGI INFORMASI*, vol. 18, no. 2, p. 126, 2024, doi: 10.47111/JTI.
- [13] D. Septiani and I. Isabela, “SINTESIA: Jurnal Sistem dan Teknologi Informasi Indonesia ANALISIS TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) DALAM TEMU KEMBALI INFORMASI PADA DOKUMEN TEKS,” *Jurnal Sistem dan Teknologi Informasi Indonesia*, 2022.
- [14] Ilham Amirul Fathoni, “IMPLEMENTASI PERINGKASAN TEKS OTOMATIS DENGAN ALGORITMA TEXTRANK UNTUK BERITA ONLINE,” 2023.
- [15] Satvika, V. Thada, and J. Singh, “A Primer on Word Embedding,” 2021, pp. 525–541. doi: 10.1007/978-981-15-8530-2\_42.