BINER
Jurnal Ilmiah Informatika dan Komputer
Vol. 5 | No. 2 | Juli 2026

p-ISSN : 2828-0210 | e-ISSN : 2828-0229
https://ojs.unsiq.ac.id/index.php/biner

EVALUASI PERFORMA TF-IDF, *WORD2VEC*, DAN INDOBERT PADA RETRIEVAL BERITA TOPIK ASTA CITA

Ayu Lintang Pratiwi [1)], Abdul Rezha Efrat Najaf [2)], Reisa Permatasari [3)]
[1,2,3]Sistem Informasi UPN "Veteran" Jawa Timur
Email: alintang04@gmail.com [1)]

Dikirim: 10 Juni 2026; Disetujui: 9 Juli 2026; Dipublikasikan: 31 Juli 2026

ABSTRAK

Peningkatan pemberitaan digital mengenai Asta Cita menghasilkan beragamnya informasi, sehingga diperlukan proses *retrieval* untuk menghasilkan dokumen yang relevan sesuai dengan kebutuhan pencarian. Perbedaan karakteristik model representasi teks TF-IDF, *Word2Vec,* IndoBERT menjadi dasar dari penelitian ini untuk mengevaluasi dan membandingkan performa dari ketiga model tersebut pada korpus berita Asta Cita. Data yang digunakan berasal dari *web scraping* dengan jumlah 1338 dokumen berita dan rentang tahun 2024-2026. Data dilakukan tahapan *preprocessing* dan diperoleh 1311 dokumen berita. Dokumen direpresentasikan menggunakan TF-IDF, *Word2Vec,* dan IndoBERT. Penghitungan kemiripan menggunakan metrik *cosine similarity.* Evaluasi dilakukan dengan lima skenario yaitu, perbandingan model, variasi jenis kueri, variasi panjang kueri, variasi jumlah hasil pencarian, dan variasi ruang pencarian. Evaluasi diukur dengan metrik *precision, recall,* dan *Mean Average Precision* (MAP). Hasil evaluasi menunjukkan *Word2Vec* memperoleh rata-rata MAP tertinggi sebesar 0,859 dibandingkan dengan TF-IDF sebesar 0,682 dan IndoBERT sebesar 0,646. Berdasarkan hasil tersebut, *Word2Vec* lebih konsisten dibandingkan kedua model lainnya karena unggul pada sebagian besar skenario. Hasil evaluasi tersebut memberikan gambaran mengenai performa masing-masing model representasi teks pada berbagai kondisi *retrieval* dan dapat menjadi pertimbangan pemilihan model representasi teks untuk pengembangan sistem informasi *retrieval.*
Kata Kunci : *Information Retrieval,* Asta Cita, Representasi Teks, Evaluasi, Relevansi Dokumen

ABSTRACT

The increase in digital news coverage of Asta Cita has resulted in a wide variety of information, making a retrieval process necessary to produce documents relevant to the search query. The differences in the characteristics of the TF-IDF, Word2Vec, and IndoBERT text representation models form the basis of this study to evaluate and compare the performance of these three models on the Asta Cita news corpus. The data used was obtained through web scraping, consisting of 1,338 news documents spanning the years 2024–2026. The data underwent preprocessing, resulting in 1,311 news documents. The documents were represented using TF-IDF, Word2Vec, and IndoBERT. Similarity was calculated using the cosine similarity metric. The evaluation was conducted across five scenarios: model comparison, query type variation, query length variation, number of search results variation, and search space variation. Evaluation was measured using precision, recall, and Mean Average Precision (MAP) metrics. The evaluation results showed that Word2Vec achieved the highest average MAP of 0.859, compared to 0.682 for TF-IDF and 0.646 for IndoBERT. Based on these results, Word2Vec was more consistent than the other two models because it outperformed them in most scenarios. These evaluation results provide an overview of the performance of each text representation model under various retrieval conditions and can serve as a basis for selecting a text representation model for the development of information retrieval systems.
Keywords *: Information Retrieval, Asta Cita, Text Representation, Evaluation, Document Relevance*

1. PENDAHULUAN

Perkembangan informasi digital menyebabkan meningkatnya jumlah data digital termasuk data yang berbasis teks. Banyaknya informasi yang tersedia harus dikelola agar pengguna dapat menemukan informasi yang relevan dengan kebutuhan informasi. Kondisi berikut menimbulkan tantangan dalam pengelolaan dan pencarian dokumen karena dalam setiap berita yang memuat topik, istilah, dan konteks yang beragam. Oleh karena itu, pendekatan temu balik informasi (*information retrieval*) yang diperlukan guna membantu proses pencarian pencarian dokumen yang relevan berdasarkan kebutuhan informasi yang diberikan melalui kueri pencarian[1].

Salah satu topik yang banyak menjadi perhatian dalam pemberitaan nasional adalah Asta Cita. Asta Cita merupakan delapan agenda prioritas pembangunan menuju visi misi Indonesia Emas 2045[2]. Pemberitaan terkait Asta Cita mencakup berbagai isu, seperti pendidikan, ekonomi, ketahanan pangan, pembangunan sumber daya manusia, hingga transformasi digital. Banyak variasi informasi pada topik tersebut menyebabkan pengguna memekan mekanisme pencarian yang mampu menghasilkan berita relevan sesuai kebutuhan informasi tertentu.

Berbagai pendekatan representasi teks telah dikembangkan untuk meningkatkan kualitas temu balik informasi. TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan salah satu metode pembobotan yang banyak digunakan karena mampu menentukan tingkat kepentingan kata berdasarkan frekuensi kemunculan pada dokumen[3]. Penelitian sebelumnya menerapkan TF-IDF dengan *cosine similarity* untuk menentukan relevansi dokumen berdasarkan kesamaan antara kueri dengan dokumen[4]. Pendekatan tersebut menunjukkan bahwa pembobotan kata menggunakan TF-IDF mampu mendukung proses pencarian dokumen dengan menghasilkan ranking dokumen yang lebih relevan terhadap kata kunci pencarian.

Penelitian lain menggunakan model berbasis *word embedding* seperti *Word2Vec* mampu memetakan hubungan semantik kata dalam bentuk vektor numerik sehingga kata dengan makna serupa memiliki posisi yang berdekatan pada ruang vektor[5]. Penelitian sebelumnya menunjukkan bahwa *Word2Vec* mampu merepresentasikan hubungan semantik antar kata sehingga mendukung dalam peningkatan akurasi pencarian dokumen.

Perkembangan lebih lanjut ditunjukkan oleh model berbasis *transformer* seperti IndoBERT yang berkemampuan dalam memahami konteks bahasa Indonesia secara lebih mendalam dan efektif dalam menangkap relevansi semantik antara kueri dengan dokumen[6]. Penelitian sebelumnya yang menerapkan IndoBERT menunjukkan bahwa IndoBERT mampu menghasilkan performa *retrieval* yang lebih baik dibandingkan pendekatan sekuensial tradisional karena memiliki kemampuan dalam memahami konteks bahasa secara lebih mendalam dan menghasilkan representasi kontekstual yang lebih optimal.

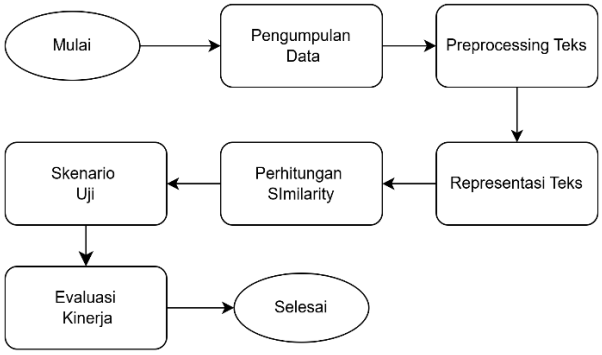
Berdasarkan dari penelitian terdahulu, menunjukkan bahwa TF-IDF, *Word2Vec,* dan IndoBERT mampu meningkatkan kualitas *retrieval* dalam berbagai penelitian mengenai *information retrieval.* Namun, penelitian tersebut masih berfokus pada penerapan masing-masing model secara terpisah sehingga belum memberikan perbandingan performa model pada objek dan kondisi penelitian yang sama. Selain itu, berdasarkan kajian literatur yang digunakan, belum ditemukan penelitian yang secara khusus membandingkan model TF-IDF, *Word2Vec,* dan IndoBERT pada korpus Asta Cita dengan melalui lima skenario pengujian. Oleh karena itu, penelitian ini melakukan evaluasi komparatif terhadap TF-IDF, *Word2Vec,* dan IndoBERT melalui lima skenario pengujian untuk mengevaluasi performa masing-masing model pada berbagai kondisi *retrieval.* Melalui lima skenario pengujian tersebut, diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai performa masing-masing model representasi teks pada korpus berita dengan topik Asta Cita.

Untuk mencapai tujuan tersebut, maka penelitian ini mengevaluasi dan membandingkan tiga pendekatan representasi teks yang memiliki karakteristik berbeda, yaitu TF-IDF sebagai model yang berbasis frekuensi kata, *Word2Vec* sebagai model representasi vektor yang mampu menangkap hubungan semantik antar kata, serta model IndoBERT sebagai model yang berbasis *transformer* yang

mampu memahami konteks bahasa Indonesia secara lebih mendalam. Performa dari ketiga model akan dievaluasi untuk mengetahui model representasi teks yang memberikan hasil *retrieval* yang optimal pada korpus berita dengan topik Asta Cita. Dengan demikian, perbandinga dari penelitian ini dapat menjadi dasar dalam menentukan model representasi teks yang memberikan performa terbaik dalam *information retrieval* pada berita dengan topik Asta Cita.

2. METODE

Penelitian ini melakukan evaluasi dan membandingkan performa dari model representasi teks pada proses temu balik informasi berita dengan topik Asta Cita. Berikut ini alur penelitian yang digunakan dalam penelitian ini.



Gambar 1. Alur Penelitian

Tahapan dalam penelitian ini ditunjukkan dalam gambar 1. Penelitian ini diawali dengan pengumpulan data berita Asta Cita yang dikumpulkan dengan teknik *web scraping* melalui *Google News* dengan rentang tahun publikasi 2024 hingga 2026. *Web scraping* merupakan teknik yang digunakan untuk mengambil informasi dari halaman website yang dilakukan secara otomastis dengan menngambil konten dari dokumen semi-terstruktur untuk diekstraksi bagian data yang dibutuhkan[7]. Data dikumpulkan dari berbagai media berita daring nasional. Lima sumber media daring dengan jumlah dokumen terbanyak diantaranya RRI sebanyak 186 dokumen, Detik News 53 dokumen, Antara News 41 dokumen, Tempo 30 dokumen, dan Liputan6.com 25 dokumen. Setelah data terkumpul kemudian dilakukan tahapan preprocessing yang meliputi *cleaning, case folding,* tokenisasi, *stopword removal,* dan *stemming.* Tahapan ini dilakukan agar data yang digunakan dalam penelitian dapat lebih optimal.

Selanjutnya dilakukan representasi teks menggunakan tiga pendekatan, yaitu TF-IDF, *Word2Vec* dan IndoBERT. TF-IDF dalam penelitian temu balik informasi digunakan untuk merepresentasikan dokumen dengan pembobotan kata berdasarkan Tingkat kepentingan kemunculan term pada dokumen dan keseluruhan korpus[8]. Untuk representasi teks dengan pendekatan *Word2Vec* dalam penelitian ini digunakan untuk mendukung proses *retrieval* dengan menghasilkan representasi teks semantik antara dokumen dengan kueri yang digunakan dalam menghitung tingkat kemiripan dokumen dengan kebutuhan informasi pengguna[9]. Pada penelitian ini, model *Word2Vec* dilatih menggunakan korpus berita Asta Cita dari hasil *preprocessing*. Model *Word2Vec* dilatih menggunakan arsitektur Skip-Gram dengan parameter vector_size 100, window 5, min_count 2, dan epoch 20. Sementara itu, pendekatan representasi teks dengan IndoBERT dalam penelitian ini untuk menghasilkan representasi kontekstual bahasa Indonesia yang mampu memahami hubungan semantic antara dokumen dengan kueri yang mendukung dalam pencocokan semantik pada proses *retrieval*[6]. Pada penelitian ini, model IndoBERT yang digynakan menggunakan model *pretrained* bahasa Indonesia untuk menghasilkan representasi dokumen dan kueri. Model yang digunakan adalah *pretrained indobenchmark/indobert-base-p1* tanpa proses fine-tuning.

Selanjutnya, setelah tahapan representasi teks, untuk menghitung tingkat kemiripan antara kueri yang diinputkan dengan korpus berita menggunakan *cosine similarity*. Metode perhitungan ini mengukur kemiripan berdasarkan sudut kosinus antara dua vektor dalam ruang multidimensi[10]. *Cosine similarity* digunakan karena mampu mengukur tingkat kemiripan antar representasi vektor tanpa dipengaruhi oleh panjang dokumen sehingga dapat diterapkan secara konsisten pada TF-IDF, *Word2Vec*, maupun IndoBERT. Perhitungan *cosine similarity* dinyatakan dalam persamaan berikut.

$$cos(\theta) = \frac{Q.D}{|Q||D|} = \frac{\sum_{i=1}^{n}(wq_i \, x \, wd_i)}{\sqrt{\sum_{i=1}^{n}(wq1)^2} x \sqrt{\sum_{i=1}^{n}(wd_i)^2}} \quad (1)$$

Nilai wq_i merupakan bobot term ke-i pada kueri, sedangkan wd_i menyatakan bobot term ke-i pada korpus dokumen. Nilai n menyatakan jumlah term yang dibandingkan antara kueri dengan dokumen.

Untuk melakukan pengujian performa dari masing-masing model representasi teks, maka diterapkan beberapa skenario pengujian. Skenario pengujian dalam penelitian disusun untuk mengevaluasi performa model dari berbagai aspek, skenario dalam penelitian ini ada lima yaitu, membandingkan ketika model dengan kueri yang identik, variasi jenis kueri yang sesuai topik Asta Cita, variasi panjang dari kueri yang digunakan, variasi jumlah dokumen hasil pencarian (Top-K), dan variasi ruang pencarian. Tujuan dari pengujian kelima skenario tersebut untuk mengetahui pengaruh kondisi pencarian terhadap performa *retrieval* dari masing-masing model representasi teks.

Hasil dari pengujian masing-masing skenario selanjutnya akan dievaluasi menggunakan metrik evaluasi *precision, recall,* dan *Mean Average Precision* (MAP). Sebelum proses evaluasi dilakukan, dokumen relevan ditentukan berdasarkan kesesuaian poin Asta Cita antara kueri dan dokumen. Sebelum proses evaluasi dilakukan, dokumen relevan ditentukan berdasarkan kesesuaian poin Asta Cita antara kueri dan dokumen. *Precision* digunakan untuk mengukur ketepatan dengan membandingkan jumlah dokumen relevan dengan seluruh dokumen yang terambil[11]. *Recall* merupakan metrik yang mengukur kinerja model dengan menunjukkan dokumen yang relevan dengan kueri pengguna yang berhasil diambil oleh sistem[12]. Sedangkan untuk metrik MAP digunakan untuk mengukur kualitas pemeringkatkan dokumen hasil pencarian dari model[13]. Persamaan metrik evaluasi ditunjukkan sebagai berikut pada persamaan (2), persamaan (3), dan persamaan (4).

Nilai *Precision* dihitung dengan persamaan :

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Nilai *Recall* dihitung dengan persamaan :

$$recall = \frac{TP}{TP+FN} \quad (3)$$

Nilai MAP dihitung dengan persamaan :

$$MAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \quad (4)$$

Nilai TP menunjukkan jumlah dokumen relevan yang dapat ditemukan, nilai FP menunjukkan jumlah dokumen yang tidak relevan, nilai FN menunjukkan jumlah dokumen relevan yang tidak dapat ditemukan. Sementara itu, untuk nilai AP_i menunjukkan nilai *average precision* untuk kueri ke-i, nilai n menunjukkan jumlah kueri. Pada evaluasi ini penentuan dokumen relevan dilakukan berdasarkan kesesuaian kategori poin Asta Cita antara kueri dan dokumen. Dokumen dengan kategori yang sama sebagai target kueri diberi label relevan, sedangkan dokumen dengan kategori berbeda diberi label tidak relevan. Pada evaluasi ini dokumen dinyatakan relevan.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini, data yang berhasil dikumpulkan dengan teknik *web scraping* dari berbagai sumber berita pada rentang tahun 2024 hingga 2026 sebanyak 1338 berita yang berkaitan dengan topik Asta Cita. Selanjutnya, data tersebut akan melalui tahapan *preprocessing* yang meliputi tahapan *cleaning, case folding,* tokenisasi, *stopword removal,* dan *stemming* untuk menghasilkan data yang lebih bersih dan siap digunakan untuk proses representasi teks. Jumlah data setelah dilakukan tahapan *preprocessing* menjadi sebanyak 1311 dokumen berita. Data tersebut yang akan digunakan sebagai ruang pencarian pada penelitian.

Setelah melalui tahapan *preprocessing* dilakukan, data direpresentasikan ke dalam bentuk numerik agar dapat diproses pada tahapan perhitungan kemiripan dokumen. Dokumen direpresentasikan menggunakan masing-masing model yaitu, TF-IDF, *Word2Vec,* dan IndoBERT.

Dokumen direpresentasikan menggunakan TF-IDF untuk menghasilkan bobot setiap term berdasarkan Tingkat kepentingan dokumen. Berikut adalah contoh hasil pembobotan dokumen dengan model TF-IDF pada dokumen berita indeks ke-140.

Tabel 1. Contoh Hasil Pembobotan TF-IDF

Kata	Bobot TF-IDF
Arah	0,037
Besar	0,030
Ekonomi	0,026
Gizi	0,050
Nasional	0,094

Tabel 1 menampilkan beberapa hasil dari pembobotan kata yang dihasilkan oleh model TF-IDF. Pada tabel terlihat nilai pembobotan

antar kata memiliki nilai yang berbeda-beda sesuai Tingkat kepentingannya dalam dokumen. *Term* dengan nilai pembobotan TF-IDF yang lebih tinggi menunjukkan kontribusi yang llebih besar dalam mempresentasikan dokumen, sedangkan *term* dengan bobot yang lebih rendah memiliki pengaruh yang lebih kecil.

Pada model *Word2Vec,* setiap kata direpresentasikan ke dalam bentuk vektor numerik yang menggambarkan hubungan semantik antar kata. Berikut adalah contoh hasil vektor dokumen model *Word2Vec.*

Tabel 2. Contoh Hasil Vektor *Word2Vec*

Dokumen	V1	V2	V3	V4	V5
7	-0.12	0.28	-0.05	0.47	-0.19
590	-0.11	0.28	0.15	0.33	-0.08
1298	-0.24	0.20	-0.16	-0.03	0.34

Tabel 2 menampilkan contoh hasil vektor numerik dalam bentuk vektor berdimensi 100. dari model *Word2Vec.* Representasi dokumen dalam vektor tersbeut digunakan sebagai masukan untuk proses menghitung Tingkat kemiripan dokumen dan kueri.

Pada model IndoBERT menghasilkan representasi kontekstual dalam bentuk vektor yang digunakan untuk mempertimbangkan hubungan antar kata dalam suatu kalimat. Berikut adalah contoh hasil vektor dokumen model IndoBERT.

Tabel 3. Contoh Hasil Vektor IndoBERT

Dokumen	V1	V2	V3	V4	V5
7	-4.44	1.82	-4.97	4.64	1.67
590	-1.32	1.51	-5.52	1.61	1.15
1298	-8.80	1.43	-4.44	3.15	7.95

Berdasarkan tabel 3, IndoBERT menghasilkan representasi dokumen dalam bentuk vektor berdimensi 768. Pada tabel tersebut hanya disajikan sebagian dimensi untuk memudahkan penyajian. Representasi vektor tersebut yang akan digunakan pada proses perhitungan tingkat kemiripan kueri dengan dokumen.

Pengujian kemudian dilakukan dengan menggunakan lima skenario untuk mengevaluasi performa model TF-IDF, *Word2Vec,* dan IndoBERT dalam menghasilkan dokumen yang relevan terhadap kueri pencarian penggguna. Evaluasi pada skenario dilakukan dengan beberapa kueri uji pada setiap skenario dan hasil performa dari masing-masing model diukur menggunakan metrik *precision, recall,* dan MAP.

3.1. Skenario Perbandingan Antar Model

Skenario ini dilakukan untuk mengetahui performa dari masing-masing model dalam menghasilkan dokumen relevan dengan menggunakan kueri yang identik. Hasil evaluasi ditunjukkan pada tabel berikut.

Tabel 4. Hasil Evaluasi Skenario Pertama

Skenario Perbandingan Antar Model			
Model	**Precision**	***Recall***	**MAP**
TF-IDF	0,267	0,008	0,508
Word2Vec	0,8	0,022	0,833
IndoBERT	0,467	0,015	0,631

Tabel 4 merupakan hasil evaluasi pada skenario perbandingan antar model, terlihat masing-masing model menunjukkan performa yang berbeda. Berdasarkan tabel tersebut, terlihat bahwa model *Word2Vec* memperoleh hasil tertinggi pada seluruh metrik dengan nilai precision sebesar 0,800, *recall* sebesar 0,022, dan MAP sebesar 0,833. Hasil ini menunjukkan bahwa *Word2Vec* lebih mampu menempatkan dokumen relevan pada urutan awal dibandingkan TF-IDF dan IndoBERT. IndoBERT berada pada urutan kedua dengan MAP sebesar 0,631, sedangkan TF-IDF memperoleh MAP sebesar 0,508. Nilai *recall* pada ketiga model masih rendah karena jumlah dokumen yang ditampilkan dalam hasil pencarian lebih sedikit dibandingkan keseluruhan dokumen relevan yang terdapat dalam korpus. Meskipun demikian, nilai MAP *Word2Vec* yang tinggi menunjukkan bahwa sebagian besar dokumen relevan berhasil ditempatkan pada peringkat awal hasil pencarian.

3.2. Skenario Variasi Jenis Kueri

Skenario ini untuk menguji kemampuan dari model untuk menangani formulasi kueri yang beragam. Hasil evaluasi pada skenario ini ditunjukkan dalam tabel dibawah ini.

Tabel 5. Hasil Evaluasi Skenario Kedua

Skenario Variasi Jenis Kueri			
Model	**Precision**	***Recall***	**MAP**
TF-IDF	0,6	0,017	0,591
Word2Vec	0,733	0,03	0,854

Skenario Variasi Jenis Kueri			
IndoBERT	0,3	0,018	0,48

Tabel 5 menunjukkan hasil evaluasi pada skenario variasi jenis kueri, menunjukkan bahwa model *Word2Vec* kembali memperoleh nilai tertinggi dengan *precision* sebesar 0,733, *recall* sebesar 0,030, dan MAP sebesar 0,854. Hasil tersebut menunjukkan bahwa *Word2Vec* mampu mempertahankan hasil pencarian ketika kueri disusun menggunakan pilihan kata yang berbeda, tetapi masih memiliki hubungan makna yang sama. TF-IDF memperoleh MAP sebesar 0,591, sedangkan IndoBERT memperoleh MAP sebesar 0,480. TF-IDF masih dapat menemukan dokumen yang memiliki kesamaan kata secara langsung dengan kueri, sedangkan hasil IndoBERT menunjukkan bahwa representasi kontekstual yang digunakan belum selalu mampu menempatkan dokumen relevan pada urutan awal untuk setiap jenis kueri.

3.3. Skenario Variasi Panjang Kueri

Skenario ini dilakukan untuk menguji pengaruh dari panjang kueri yang diberikan terhadap performa *retrieval* pada masing-masing model. Hasil evaluasi pada skenario ini ditunjukkan pada tabel berikut.

Tabel 6. Hasil Evaluasi Skenario Ketiga

Skenario Variasi Panjang Kueri			
Model	**Precision**	***Recall***	**MAP**
TF-IDF	0,7	0,023	0,770
Word2Vec	0,733	0,024	0,761
IndoBERT	0,333	0,011	0,548

Tabel 6 menunjukkan hasil evaluasi yang beragam pada masing-masing model. *Word2Vec* memperoleh *precision* dan *recall* tertinggi, yaitu sebesar 0,733 dan 0,024. Namun, nilai MAP tertinggi diperoleh TF-IDF sebesar 0,770, sedikit lebih tinggi dibandingkan *Word2Vec* yang sebesar 0,761. Hasil tersebut menunjukkan bahwa bertambahnya kata pada kueri dapat meningkatkan peluang kesamaan term secara langsung antara kueri dan dokumen. Kondisi ini menguntungkan TF-IDF karena pembobotannya didasarkan pada kemunculan dan tingkat kepentingan *term*. Sementara itu, *Word2Vec* lebih banyak menemukan dokumen relevan dalam hasil pencarian, tetapi urutan dokumen relevannya tidak selalu lebih baik dibandingkan dengan TF-IDF. Perbedaan MAP antara kedua model juga relatif kecil, sehingga hasil ini hanya menunjukkan kecenderungan performa pada skenario variasi panjang kueri.

3.4. Skenario Variasi Hasil Pencarian Top-K

Skenario ini dilakukan untuk menguji kualitas dan konsistensi model dalam menampilkan hasil perankingan pada kedalaman hasil pencarian dengan jumlah dokumen hasil yang beragam. Berikut ini hasil evaluasi yang ditunjukkan pada tabel.

Tabel 7. Hasil Evaluasi Skenario Keempat

Skenario Variasi Jumlah Hasil Pencarian			
Model	**Precision**	***Recall***	**MAP**
TF-IDF	0,683	0,08	0,875
Word2Vec	0,867	0,106	0,952
IndoBERT	0,49	0,061	0,571

tabel 7 terlihat bahwa model *Word2Vec* menunjukkan nilai MAP tertinggi pada skenario variasi jumlah hasil pencarian dengan *precision* sebesar 0,867, *recall* sebesar 0,106, dan MAP sebesar 0,952. TF-IDF berada pada urutan kedua dengan MAP sebesar 0,875, sedangkan IndoBERT memperoleh MAP sebesar 0,571. Nilai *recall* pada skenario ini lebih tinggi dibandingkan beberapa skenario sebelumnya. Hal tersebut terjadi karena semakin banyak dokumen yang ditampilkan dalam hasil pencarian, semakin besar pula peluang dokumen relevan untuk ditemukan. Hasil ini menunjukkan bahwa rendahnya nilai recall pada skenario sebelumnya turut dipengaruhi oleh pembatasan jumlah hasil pencarian, bukan hanya oleh kemampuan model dalam menemukan dokumen relevan.

3.5. Skenario Variasi Ruang Pencarian

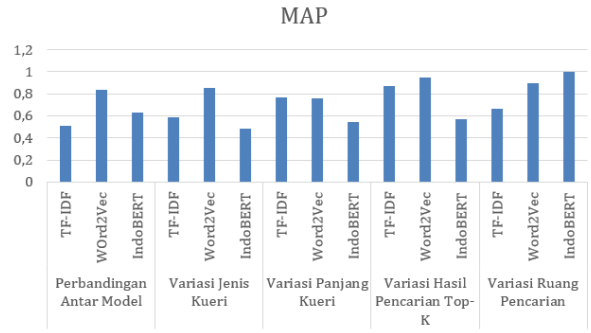
Skenario ini dilakukan untuk menguji pengaruh jumlah korpus dokumen terhadap performa *retrieval* yang dihasilkan model. Berikut ini tabel yang akan menyajikan hasil evaluasi model pada skenario ini.

Tabel 8. Hasil Evaluasi Skenario Kelima

Skenario Variasi Ruang Pencarian			
Model	**Precision**	***Recall***	**MAP**
TF-IDF	0,733	0,034	0,666
Word2Vec	0,8	0,037	0,897
IndoBERT	1,0	0,046	1,0

Tabel 8 diatas menunjukkan bahwa model IndoBERT memperoleh hasil tertinggi dengan *precision* dan MAP sebesar 1,0 serta *recall* sebesar 0,046. Nilai tersebut menunjukkan bahwa dokumen yang ditempatkan pada urutan awal hasil pencarian IndoBERT sesuai dengan target kueri pada skenario ini. Meskipun

precision dan MAP telah mencapai nilai maksimal, *recall* masih rendah karena dokumen yang ditemukan hanya mencakup sebagian kecil dari seluruh dokumen relevan yang tersedia. *Word2Vec* juga menunjukkan hasil yang baik dengan MAP sebesar 0,897, sedangkan TF-IDF memperoleh MAP sebesar 0,666. Keunggulan IndoBERT pada skenario ini menunjukkan bahwa representasi kontekstual mampu memberikan hasil yang baik ketika jumlah dokumen dalam ruang pencarian mengalami perubahan.



Gambar 2 Grafik Nilai MAP

Berdasarkan Gambar 2, nilai MAP ketiga model menunjukkan hasil yang berbeda pada setiap skenario pengujian. *Word2Vec* memperoleh nilai MAP tertinggi pada tiga dari lima skenario, yaitu skenario perbandingan antar model, variasi jenis kueri, dan variasi jumlah hasil pencarian. Sementara itu, TF-IDF memperoleh nilai tertinggi pada skenario variasi panjang kueri, sedangkan IndoBERT unggul pada skenario variasi ruang pencarian. Hasil tersebut menunjukkan bahwa setiap model memiliki kemampuan yang berbeda sesuai dengan kondisi pengujian yang digunakan.

Tabel 9. Rata-Rata Hasil Evaluasi

Rata-Rata Hasil Evaluasi			
Model	**Precision**	***Recall***	**MAP**
TF-IDF	0,597	0,032	0,682
Word2Vec	0,787	0,044	0,859
IndoBERT	0,518	0,030	0,646

Tabel 9 menunjukkan rata-rata hasil evaluasi dari seluruh skenario pengujian. *Word2Vec* memperoleh rata-rata nilai tertinggi pada seluruh metrik evaluasi, yaitu *precision* sebesar 0,787, *recall* sebesar 0,044, dan MAP sebesar 0,859. TF-IDF memperoleh rata-rata MAP sebesar 0,682, sedangkan IndoBERT sebesar 0,646. Hasil tersebut menunjukkan bahwa secara keseluruhan *Word2Vec* lebih baik dalam menemukan dan menempatkan dokumen relevan pada urutan awal hasil pencarian.

Hasil *Word2Vec* yang lebih tinggi kemungkinan dipengaruhi oleh proses pelatihannya yang menggunakan korpus berita Asta Cita. Dengan menggunakan korpus penelitian, hubungan antarkata yang dipelajari menjadi lebih sesuai dengan istilah dan pembahasan yang terdapat pada data. Hal ini sesuai dengan konsep representasi distribusional, yaitu kata yang sering muncul dalam konteks yang sama akan memiliki posisi vektor yang berdekatan. Kondisi tersebut membantu Word2Vec menemukan dokumen yang tetap berkaitan dengan kueri meskipun tidak menggunakan susunan kata yang sama.

Sementara itu, hasil IndoBERT masih lebih rendah dibandingkan *Word2Vec* pada sebagian besar skenario. Pada penelitian ini, IndoBERT digunakan dalam bentuk *pretrained* tanpa *fine-tuning* pada korpus berita Asta Cita. Oleh karena itu, representasi yang dihasilkan masih bersifat umum dan belum secara khusus menyesuaikan dengan tugas pencarian berita pada penelitian ini. Selain itu, representasi dokumen diambil dari token CLS dengan panjang masukan maksimal 512 token, sehingga isi berita yang lebih panjang belum tentu dapat terwakili secara keseluruhan. Meskipun demikian, IndoBERT tetap menunjukkan hasil terbaik pada skenario variasi ruang pencarian.

Rata-rata nilai *recall* ketiga model masih rendah, yaitu antara 0,030 hingga 0,044. Hal ini terjadi karena jumlah dokumen yang ditampilkan pada hasil pencarian lebih sedikit dibandingkan dengan jumlah seluruh dokumen relevan yang terdapat dalam korpus. Akibatnya, meskipun beberapa dokumen relevan sudah berada pada urutan awal, jumlah dokumen yang ditemukan masih mencakup sebagian kecil dari seluruh dokumen relevan.

Secara keseluruhan, hasil lima skenario menunjukkan bahwa *Word2Vec* memberikan hasil paling baik pada sebagian besar kondisi pengujian. *Word2Vec* memperoleh nilai MAP tertinggi pada tiga dari lima skenario serta menghasilkan rata-rata MAP tertinggi sebesar 0,859. Hasil tersebut menunjukkan bahwa penggunaan korpus berita Asta Cita pada proses pelatihan *Word2Vec* membantu model membentuk representasi kata yang lebih sesuai dengan karakteristik data penelitian. Meskipun demikian, TF-IDF dan IndoBERT tetap menunjukkan keunggulan pada kondisi tertentu,

yaitu TF-IDF pada variasi panjang kueri dan IndoBERT pada variasi ruang pencarian. Namun, hasil perbandingan ini masih terbatas pada nilai evaluasi yang diperoleh dari skenario pengujian dan belum didukung oleh uji signifikansi statistik.

4. PENUTUP

4.1. Kesimpulan

Penelitian ini telah berhasil melakukan evaluasi dan membandingkan performa dari model representasi teks TF-IDF, *Word2Vec,* dan IndoBERT pada proses *retrieval* berita dengan topik Asta Cita yang diperoleh dari hasil *web scraping* dengan rentang tahun 2024 hingga 2026. Penlitian ini melakukan pengujian dengan menggunakan lima skenario pengujian. Kemudian dilakukan evaluasi untuk melihat performa dari masing-masing model. Hasil evaluasi menunjukkan bahwa masing-masing model menghasilkan performa yang berbeda pada setiap kondisi pengujian. Untuk model *Word2Vec* menunjukkan performa yang konsisten dengan memperoleh nilai *precision, recall,* dan MAP tertinggi pada sebagian besar skenario. Sedangkan pada model TF-IDF dan IndoBERT menunjukkan performa yang lebih baik pada kondisi tertentu sesuai dengan karakteriktik masing-masing model.

Secara keseluruhan, penelitian ini menunjukkan bahwa model *Word2Vec* menjadi model representasi teks yang paling konsisten dalam proses *retrieval* berita Asta Cita.

4.2. Saran

Penelitian selanjutnya dapat mengembangkan evaluasi *retrieval* dengan menggunakan korpus berita yang lebih besar dan beragam, mengeksplorasi metode representasi teks lainnya, serta menerapkan variasi scenario pengujian maupun metrik evaluasi tambahan untuk memperoleh analisis performa yang lebih komprehensif. Penelitian selanjutnya juga diharapkan untuk memperluas variasi skenario uji yang dilakukan dengan berbagai kondisi.

5. DAFTAR PUSTAKA

[1] I. Cholissodin *et al.*, "Pemeringkatan Pencarian Pada Buku Pedoman Akademik Filkom Ub Menuju Merdeka Belajar Dan Free E-Book Pembelajaran Sebagai Prototype Local Smart Micro Search Engine Menggunakan Algoritma Pagerank Dan Tf-Idf", doi: 10.25126/jtiik.202184384.

[2] R. E. Sakti, "Astacita, Visi Pemerintahan Prabowo-Gibran Menuju Indonesia Emas 2045," Kompas.id. Accessed: Dec. 08, 2025. [Online]. Available: https://www.kompas.id/artikel/astacita-visi-pemerintahan-prabowo-gibran-menuju-indonesia-emas-2045

[3] E. Prayitno, T. Suprawoto, and B. F. Riyanto, "Optimasi Hasil Pencarian Pada Web Scrapping Menggunakan Pembobotan Kata TF-IDF," *Journal of Innovation Research and Knowledge*, vol. 1, no. 7, pp. 241–246, Dec. 2021.

[4] I. Putu Gede Hendra Suputra, I. Made Widhi Wirawan, D. Ketut Puri Trisnantya Aji, A. Agung Sinta Trisnajayanti, and F. Matematika dan Ilmu Pengetahuan Alam, "Sistem Temu Kembali Informasi Untuk Mendukung Layanan Frequently Asked Questions (Faq) Pada Platform Sistem Informasi Terintegrasi," *Jurnal Pendidikan Teknologi dan Kejuruan*, vol. 22, no. 2, pp. 205–215, Jul. 2025.

[5] D. A. Suryaningrum, R. Syaifudin, and H. R. P. Putra, "Integrasi Word Embeddings Dan Inverse Book Frequency Dalam Pembobotan Term Untuk Peningkatan Pencarian Dokumen," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 4, pp. 2529–2537, Dec. 2024, doi: 10.29100/jipi.v9i4.7557.

[6] N. S. Sunendar and I. Saputra, "Comparative Performance Of Transformer And LSTM Models For Indonesian Information Retrieval With IndoBERT," *Jurnal Pilar Nusa Mandiri*, vol. 21, no. 2, pp. 228–233, Sep. 2025, doi: 10.33480/pilar.v21i2.6920.

[7] L. Tiara, H. Syaputra, W. Cholil, and A. H. Mirza, "Web Scraper Dan Graphql API Untuk Data Perguruan Tinggi Di Indonesia Berdasarkan Website Kementrian Ristekdikti," *Jurnal Nasional Ilmu Komputer*, vol. 2, no. 3, Aug. 2021.

[8] D. Septiani and I. Isabela, "Analisis Term Frequency Inverse Document Frequency

(Tf-Idf) Dalam Temu Kembali Informasi Pada Dokumen Teks," *SINTESIA: Jurnal Sistem dan Teknologi Informasi Indonesia*, vol. 1, no. 2, Mar. 2022.

[9] Y. C. Gurning, S. C. Saragih, Y. Y. Lase, and J. Julham, "Perbandingan TextRank Berbasis TF-IDF dan *Word2Vec* dalam Peringkasan Teks Berita Bahasa Indonesia," *Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI)*, vol. 4, no. 2, pp. 922–928, Aug. 2025, doi: 10.62712/juktisi.v4i2.552.

[10] M. Dzikry Afandi, A. Homaidi, A. Ghofur, and A. Zubairi, "Penerapan Information Retrieval dalam Sistem Analisis Kemiripan Proposal Skripsi menggunakan Cosine Similarity," *JURNAL SWABUMI*, vol. 12, no. 1, p. 2023, 2024.

[11] S. Bahri, "Aplikasi Pencarian Bahan Pustaka Di Perpustakaan Menggunakan Metode Vector Space Model," *JIMP-Jurnal Informatika Merdeka Pasuruan*, vol. 5, 2020.

[12] D. Jurafsky and J. H. Martin, "Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models Third Edition draft."

[13] A. A. Kalinin *et al.*, "A versatile information retrieval framework for evaluating profile strength and similarity," *Nature Communications* , vol. 16, no. 1, Dec. 2025, doi: 10.1038/s41467-025-60306-2.