

Analisis Perbandingan Algoritma BM25 dan TF-IDF untuk Temu Kembali Metadata Jurnal Indonesia pada Temujurnal.com

Muhamad Syarif^{1*}, Rabiatus Sa'adah², M Rafi Ardana Listi³, Ravi Manisha⁴

^{1,2,3,4}Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Indonesia

Email: ¹muhamad.mdx@bsi.ac.id, ²rabiatus.rbh@bsi.ac.id

INFORMASI ARTIKEL

Histori artikel:

Naskah masuk, 28 November 2025

Direvisi, 28 Desember 2025

Diiterima, 31 Desember 2025

ABSTRAK

Abstract- The growth of scientific publications in Indonesia necessitates an efficient journal metadata search mechanism. This study aims to compare the performance of Term Frequency-Inverse Document Frequency (TF-IDF) and Best Matching 25 (BM25) algorithms on the Temujurnal.com aggregator platform. The methodology involves text preprocessing and the implementation of Hybrid Retrieval logic, which integrates textual variables with filters such as Sinta rank, publication month, and publication fees. Testing was performed using 100 journal metadata entries and 20 test queries, evaluated through Precision, Recall, F1-Score, and Mean Average Precision (MAP) metrics. The results indicate that BM25 significantly outperforms TF-IDF, achieving a MAP value of 0.74 compared to 0.59. BM25 is proven to be more stable in managing variations in journal metadata length than TF-IDF. This study recommends the implementation of BM25 on the Temujurnal platform to enhance the accuracy and efficiency of the national scientific literature retrieval system.

Kata Kunci:

BM25,
TF-IDF,
Temu Kembali Informasi,
Metadata Jurnal,
Temujurnal.com

Abstrak- Pertumbuhan publikasi ilmiah di Indonesia menuntut mekanisme pencarian metadata jurnal yang efisien. Penelitian ini bertujuan membandingkan performa algoritma Term Frequency-Inverse Document Frequency (TF-IDF) dan Best Matching 25 (BM25) pada platform agregator Temujurnal.com. Metodologi yang digunakan meliputi pra-pemrosesan teks dan penerapan logika Hybrid Retrieval yang mengintegrasikan variabel teksual dengan filter peringkat Sinta, bulan terbit dan biaya publikasi. Pengujian dilakukan menggunakan 100 metadata jurnal dan 20 kueri uji dengan metrik evaluasi Precision, Recall, F1-Score, dan Mean Average Precision (MAP). Hasil penelitian menunjukkan bahwa BM25 secara signifikan mengungguli TF-IDF dengan nilai MAP 0,74 berbanding 0,59. Algoritma BM25 terbukti lebih stabil dalam menangani variasi panjang metadata jurnal dibandingkan TF-IDF. Penelitian ini merekomendasikan implementasi BM25 pada platform Temujurnal untuk meningkatkan akurasi dan efisiensi sistem temu kembali literatur ilmiah nasional.

Copyright © 2025 LPPM - STMIK IKMI Cirebon
This is an open access article under the CC-BY license

Penulis Korespondensi:

Muhamad Syarif

Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

Universitas Bina Sarana Informatika

Jl. Abdul Rahman Saleh No.18, Kec. Pontianak Tenggara, Kota Pontianak, Kalimantan Barat - Indonesia

Email: muhamad.mdx@bsi.ac.id

1. Pendahuluan

Ekosistem publikasi ilmiah di Indonesia telah mengalami transformasi digital yang masif melalui portal SINTA yang mengintegrasikan ribuan jurnal berbasis *Open Journal Systems* (OJS). Peningkatan

volume dokumen akademik memicu tantangan baru dalam efisiensi penemuan informasi yang relevan bagi peneliti [1]. Kebutuhan untuk menemukan wadah publikasi yang tepat kini tidak lagi terbatas pada kesesuaian *focus* dan *scope*, namun melibatkan

parameter multidimensi yang krusial seperti peringkat Sinta, biaya publikasi (*Article Processing Charge*), serta periode bulan terbit untuk memastikan target luaran riset tercapai tepat waktu [2][3].

Temujurnal.com hadir sebagai *platform* aggregator jurnal Indonesia yang memfasilitasi pencarian komprehensif melalui atribut metadata tekstual (nama jurnal, singkatan, penerbit, deskripsi) serta filter spesifik (peringkat sinta, periode terbit, biaya publikasi). Keberagaman atribut metadata ini menuntut adanya sistem temu kembali informasi (*Information Retrieval*) yang cerdas agar mampu memberikan hasil yang presisi. Namun, banyak sistem pencarian pada repositori menghadapi kendala dalam menentukan relevansi ketika kueri pengguna bersifat ambigu atau sangat spesifik [4][5]. Permasalahan utama terletak pada penggunaan algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) yang secara historis menjadi standar dalam pembobotan dokumen [6].

Meskipun efektif dalam membedakan kata-kata unik, TF-IDF memiliki kelemahan signifikan dalam menangani metadata heterogen, seperti perbedaan panjang karakter antara “Singkatan Jurnal” yang singkat dengan “Deskripsi Jurnal” yang ekspansif. Tanpa mekanisme normalisasi panjang dokumen yang memadai, TF-IDF cenderung memberikan skor yang tidak proporsional pada dokumen yang lebih panjang, sehingga menurunkan kualitas peringkat relevansi bagi pengguna [7][8]. Sebagai solusi, algoritma BM25 (*Best Matching 25*) memiliki mekanisme pendekatan probabilistik yang mengintegrasikan parameter *term saturation* dan *document length normalization*. Mekanisme BM25 memberikan bobot yang lebih stabil, sehingga dokumen pendek seperti “Nama Jurnal” tidak terabaikan oleh dokumen dengan “Deskripsi Jurnal” yang lebih panjang namun memiliki densitas kata kunci yang serupa [4]. Integrasi algoritma berbasis BM25 secara konsisten mampu memberikan nilai akurasi yang lebih tinggi dibandingkan metode ruang vektor tradisional, terutama pada sistem yang mengolah teks berbahasa Indonesia dengan struktur yang tidak seragam [5][9].

Keunggulan BM25 terletak pada dua parameter kalibrasi utama, yaitu k_1 yang berfungsi mengontrol fungsi saturasi frekuensi kata (*term frequency saturation*) dan b yang mengatur tingkat normalisasi panjang dokumen (*document length normalization*) [10].

Penelitian ini bertujuan untuk melakukan analisis perbandingan performa antara algoritma BM25 dan TF-IDF pada *platform* Temujurnal.com. Fokus utamanya adalah melakukan analisis perbandingan performa antara algoritma BM25 dan TF-IDF untuk mengevaluasi efektivitas temu kembali metadata jurnal Indonesia pada *platform*

Temujurnal.com. Hasil penelitian diharapkan dapat memberikan rujukan teknis yang konkret dalam mengoptimalkan mesin pencari metadata jurnal.

2. Studi Literatur

2.1. Information Retrieval

Temu kembali informasi (*Information Retrieval* atau IR) adalah bidang ilmu yang berfokus pada penemuan dokumen atau informasi yang tidak terstruktur (biasanya teks) yang memenuhi kebutuhan informasi dari dalam koleksi yang besar [11]. Proses utama dalam sistem IR melibatkan representasi dokumen melalui *indexing* dan pemrosesan kueri untuk menghasilkan daftar peringkat dokumen berdasarkan tingkat relevansinya. Di era digital saat ini, efisiensi sistem IR pada portal jurnal menjadi kunci bagi peneliti dalam mengelola *information overload* guna mendapatkan referensi yang akurat [6][11].

2.2. Metadata Jurnal

Metadata merupakan data terstruktur yang mendeskripsikan karakteristik suatu dokumen. Dalam konteks jurnal ilmiah di Indonesia, metadata mencakup atribut krusial seperti judul, singkatan, institusi, bidang ilmu, peringkat Sinta, hingga biaya publikasi [8]. Pengolahan metadata yang heterogen memerlukan pendekatan algoritma yang mampu menangani perbedaan panjang teks (seperti nama jurnal yang pendek versus deskripsi jurnal yang panjang) agar informasi filter seperti bidang ilmu dan peringkat tetap terjaga relevansinya [12].

2.3. Algoritma TF-IDF (*Term Frequency-Inverse Document Frequency*)

TF-IDF adalah metode statistik yang digunakan untuk mengevaluasi seberapa penting sebuah kata dalam sebuah dokumen terhadap sekumpulan korpus [7]. Algoritma ini terdiri dari dua komponen utama:

- Term Frequency* (TF): Menghitung jumlah kemunculan kata kunci pada korpus data untuk menentukan bobot relevansi dokumen.
- Inverse Document Frequency* (IDF): Mengurangi bobot kata yang muncul terlalu sering di banyak dokumen (seperti kata hubung) dan meningkatkan bobot kata yang unik.

Secara matematis, bobot TF-IDF dirumuskan sebagai berikut [13]:

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

Komponen-komponen dalam perhitungan skor TF-IDF didefinisikan sebagai berikut:

- $tf_{i,j}$: Jumlah kemunculan *term i* di dalam dokumen *j* (*number of occurrences of i in j*). Nilai ini mewakili aspek *Term Frequency*, di mana semakin sering suatu kata muncul dalam

- dokumen, semakin tinggi bobotnya untuk dokumen tersebut.
- df_i : Jumlah dokumen yang mengandung *term i* (*number of documents containing i*). Nilai ini digunakan untuk menghitung *Inverse Document Frequency* (IDF).
 - N : Total jumlah dokumen dalam koleksi atau korpus (*total number of documents*).
 - $\log(N/df_i)$: Bagian ini merupakan komponen IDF. Fungsinya adalah untuk mengurangi bobot kata-kata yang muncul terlalu umum di banyak dokumen (seperti kata hubung “dan”, “yang”) dan meningkatkan bobot kata-kata yang lebih unik dan jarang muncul di seluruh korpus [13].

Kelemahan utama TF-IDF dalam pencarian metadata yang beragam panjangnya adalah ketiadaan mekanisme normalisasi panjang dokumen yang eksplisit dan seringkali menyebabkan bias pada dokumen dengan teks lebih panjang [14].

2.4. Algoritma BM25 (Best Matching 25)

Algoritma BM25 merupakan pengembangan dari model sistem temu kembali informasi berbasis probabilistik, dirancang untuk menyempurnakan kelemahan pada metode TF-IDF, khususnya dalam hal normalisasi panjang dokumen dan efek saturasi frekuensi *term* secara non-linier [10]. Skor relevansi dokumen d terhadap kueri Q dihitung dengan rumus sebagai berikut:

$$Score_{BM25}(Q, d) = \sum_{i=1}^n IDF(q_i) \cdot \frac{f(q_i, d) \cdot (k_1 + 1)}{f(q_i, d) + k_1 \cdot (1 - b + b \cdot \frac{|d|}{avgdl})}$$

Komponen-komponen dalam perhitungan skor BM25 adalah sebagai berikut [15]:

- $Score_{BM25}(Q, d)$: Skor akhir relevansi dokumen d terhadap kueri Q .
- $IDF(q_i)$: Nilai *Inverse Document Frequency* dari kata ke- i dalam kueri. Nilai ini memberikan bobot lebih tinggi pada kata-kata yang jarang muncul di seluruh korpus.
- $f(q_i, d)$: Frekuensi kemunculan kata q_i di dalam dokumen d .
- $|d|$: Panjang dokumen d yang sedang dihitung skornya (jumlah kata).
- $avgdl$: Rata-rata panjang dokumen (*average document length*) dalam keseluruhan koleksi atau *database*.
- k_1 : Parameter bebas (*free parameter*) yang berfungsi untuk mengontrol fungsi saturasi frekuensi kata. Nilai yang umum digunakan berada pada rentang 1.2 hingga 2.0.
- b : Parameter bebas yang berfungsi untuk mengatur tingkat normalisasi panjang dokumen. Nilai b berada di antara 0 hingga 1; jika $b=1$, sistem melakukan normalisasi panjang dokumen secara penuh, sedangkan jika $b=0$, tidak ada normalisasi panjang dokumen yang dilakukan.
- n : Jumlah total kata kunci yang terdapat dalam kueri Q .

Keunggulan utama BM25 dalam pencarian metadata pada terletak pada penggunaan parameter b . Parameter ini memastikan bahwa dokumen dengan judul yang singkat namun sangat relevan tidak kalah bersaing dengan dokumen yang memiliki deskripsi panjang, sehingga hasil pencarian menjadi lebih relevan dan akurat [3][15].

2.5. Text Preprocessing

Dalam memproses teks berbahasa Indonesia, tahap *preprocessing* sangat menentukan akurasi *indexing*. Tahapan ini umumnya meliputi *case folding*, *punctuation and number removal*, *tokenizing*, *Stopwords Removal*, dan *Stemming* [16]. Mengingat bahasa Indonesia memiliki morfologi yang kompleks dengan imbuhan awalan, sisipan, dan akhiran, penggunaan algoritma *stemming* yang tepat seperti algoritma Nazief-Adriani atau Sastrawi sangat diperlukan untuk mengembalikan kata ke bentuk dasarnya agar dapat dicocokkan oleh algoritma BM25 maupun TF-IDF [17].

Pipeline pra-pemrosesan yang diterapkan pada setiap dokumen metadata adalah sebagai berikut:

- Case Folding*: Melakukan penyeragaman seluruh karakter teks menjadi huruf kecil (*lowercase*) guna menghindari redundansi pada term yang sama.
- Punctuation Removal and Number Removal*: Mengeliminasi elemen *noise* berupa tanda baca dan angka.
- Tokenization*: Melakukan segmentasi teks dengan memecah rangkaian kalimat menjadi unit-unit *term* yang mandiri.
- Stopword Removal*: Menghapus kata-kata fungsional yang memiliki frekuensi kemunculan tinggi namun minim kandungan informasi (misal: “dan”, “di”, “yang”).
- Stemming*: Mereduksi kata ke bentuk dasar.

2.6. Metrik Evaluasi Sistem

Untuk mengukur kinerja perbandingan antara kedua algoritma, penelitian ini menggunakan metrik standar dalam sistem temu kembali informasi, yaitu *Precision*, *Recall*, dan *F1-Score* [18].

- Precision*: Rasio dokumen relevan yang ditemukan terhadap total dokumen yang diambil.
- Recall*: Rasio dokumen relevan yang ditemukan terhadap total dokumen relevan yang ada dalam *database*.
- F1-Score*: Rata-rata antara *score Precision* dan *Recall* untuk melihat stabilitas performa algoritma [19][20].

Metode seleksi fitur secara umum dapat dibagi menjadi tiga kategori: *filter*, *wrapper*, dan *embedded*. Metode *filter* adalah yang paling sederhana dan paling cepat secara komputasi. Metode ini mengevaluasi relevansi fitur berdasarkan

karakteristik statistik data itu sendiri, sebelum model dilatih.

Salah satu metode filter yang paling umum adalah analisis korelasi, sering kali menggunakan koefisien korelasi *Pearson*. Pendekatan ini mengukur kekuatan hubungan linier antara setiap fitur independen dan variabel target misalnya, *diagnosed_diabetes*. Dengan menetapkan ambang batas misalnya peneliti dapat dengan cepat menyaring sejumlah besar fitur dan hanya mempertahankan fitur-fitur yang memiliki hubungan statistik terkuat dengan hasil yang ingin diprediksi.

3. Celah Penelitian

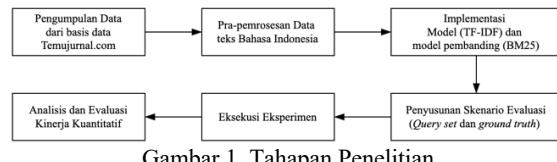
Berdasarkan tinjauan literatur yang telah dipaparkan pada bab sebelumnya, terdapat beberapa kesenjangan penelitian (*research gaps*) yang menjadi landasan utama dilakukannya penelitian ini:

- Karakteristik Metadata Pendek (Short-Text Metadata): Sebagian besar penelitian mengenai perbandingan TF-IDF dan BM25 di Indonesia berfokus pada dokumen teks panjang seperti berita daring atau dokumen skripsi secara utuh [4][7]. Masih sangat terbatas penelitian yang secara spesifik menguji performa kedua algoritma tersebut pada metadata jurnal yang bersifat heterogen, di mana terdapat perbedaan panjang yang kontras antara singkatan jurnal (sangat pendek) dengan deskripsi jurnal (sedang).
- Integrasi Atribut Non-Tekstual: Penelitian terdahulu umumnya hanya berfokus pada relevansi berbasis teks semata [15][20]. Belum ada kajian mendalam yang meneliti bagaimana algoritma *ranking* seperti BM25 berinteraksi dengan filter yang sering digunakan peneliti Indonesia, seperti peringkat Sinta dan biaya publikasi (*APC*). Penelitian ini mengisi celah tersebut dengan menguji bagaimana algoritma menangani kueri pada platform Temujurnal.com yang memiliki beberapa atribut filter.

Dengan mengidentifikasi celah-celah tersebut, penelitian ini bermaksud untuk memberikan kontribusi baru berupa evaluasi empiris yang lebih relevan bagi kebutuhan akademisi di Indonesia dalam menemukan wadah publikasi ilmiah yang sesuai berdasarkan hasil perbandingan metode terbaik.

4. Metode Penelitian

Penelitian ini dilaksanakan melalui enam tahapan sistematis untuk memastikan perbandingan yang objektif. Tahapan penelitian diilustrasikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

4.1. Pengumpulan Data

Tahap pertama adalah pengumpulan data (*data acquisition*) dari basis data Temujurnal.com. *Dataset* terdiri dari 100 metadata jurnal Indonesia yang mencakup variabel teks dan non-teks.

- Variabel Teks: Nama Jurnal, Singkatan, Bidang Ilmu, Penerbit, dan Deskripsi.
- Variabel Non-Teks (Filter): Peringkat Sinta, Biaya Publikasi, dan Periode Terbit.

4.2. Pra-pemrosesan Data

Data teks yang bersifat tidak terstruktur diproses melalui tahap NLP (*Natural Language Processing*) untuk menghasilkan indeks yang bersih. Tahapan ini meliputi *Case Folding*, *Punctuation Removal and Number Removal*, *Tokenization*, *Stopword Removal*, dan *Stemming*.

4.3. Implementasi Model

Kedua algoritma diterapkan untuk menghitung bobot relevansi dokumen terhadap kueri pengguna.

- Implementasi TF-IDF
Mengacu pada rumus pembobotan frekuensi *term* dikalikan logaritma frekuensi dokumen terbalik.
- Implementasi BM25
Digunakan sebagai model pembanding yang menyertakan parameter normalisasi panjang dokumen guna menangani variasi teks dan non teks (filter). Parameter diatur pada nilai $k_1 = 1.2$ dan $b = 0.75$.

4.4. Penyusunan Skenario Evaluasi

Disusun skenario evaluasi yang meliputi:

- Query Set*: Penentuan 20 kueri uji yang mencakup pencarian nama jurnal dan instansi Penerbit (misal: "Universitas Bina Sarana Informatika").
- Ground Truth*: Penetapan daftar "dokumen relevan" dilakukan melalui *Expert Judgment*. Dokumen dianggap relevan jika metadata memenuhi kriteria teks kueri dan batasan filter Sinta atau Biaya.

4.5. Eksesi Eksperimen

Tahap eksesi eksperimen merupakan inti dari pengujian sistem, di mana kueri diolah melalui mekanisme *Hybrid Retrieval*. Mekanisme ini menggabungkan efisiensi sistem berbasis aturan (*rule-based filtering*) dengan akurasi pemeringkatan berbasis teks (*probabilistic ranking*). Proses eksesi dibagi menjadi tiga fase operasional utama.

4.5.1. Fase Penyaringan

Pada fase pertama, sistem menerima masukan filter non-teks yang mencakup Peringkat Sinta, batasan Biaya Publikasi, dan Periode Terbit. Atribut-atribut ini diperlakukan sebagai *Hard Constraints*, di mana sistem melakukan operasi logika Boolean untuk membatasi ruang pencarian (*search space*).

- a. Logika Filter: Jurnal hanya akan diloloskan ke tahap berikutnya jika memenuhi kriteria filter.
$$\text{Status} = (\text{Sinta} == \text{Target}) \text{ and } (\text{Cost} \leq \text{Limit}) \text{ and } (\text{Month} == \text{Target})$$
- b. Dokumen yang tidak memenuhi kriteria tersebut akan langsung dieliminasi, sehingga algoritma pembobotan hanya bekerja pada sub-set data yang sudah valid.

4.5.2. Fase Pemeringkatan Probabilistik

Setelah sub-set dokumen terbentuk, sistem menerapkan algoritma TF-IDF dan BM25 secara terpisah untuk menghitung skor relevansi tekstual terhadap kueri pengguna. Pada tahap ini, sistem melakukan pencocokan kueri terhadap empat variabel teks utama: Nama Jurnal, Singkatan, Penerbit, dan Deskripsi.

- a. *Multi-field Weighting*: Mengingat metadata memiliki kepentingan yang berbeda, sistem memberikan bobot prioritas (*boost*), pada atribut Nama Jurnal dan Singkatan diberikan bobot lebih tinggi dibandingkan Deskripsi, karena mencerminkan relevansi langsung terhadap jurnal.
- b. Perhitungan Skor: Untuk BM25, skor akhir dihitung dengan menjumlahkan kontribusi dari setiap kata kueri (q_i) yang ditemukan pada seluruh variabel teks metadata tersebut sesuai dengan parameter saturasi k_1 dan normalisasi panjang dokumen b .

4.5.3. Fase Integrasi dan Urutan Luaran

Hasil akhir dari eksekusi ini adalah daftar jurnal yang diurutkan secara menurun (*descending order*) berdasarkan total skor relevansi yang dihasilkan pada fase kedua.

- a. Penanganan Relevansi Nol: Jika terdapat dokumen yang lolos filter Boolean namun memiliki skor teks nol (tidak ada kata kueri yang cocok), dokumen tersebut akan tetap ditampilkan di posisi paling bawah atau diabaikan tergantung pada pengaturan ambang batas (*threshold*) relevansi.
- b. Validasi Komparatif: Setiap luaran dari model TF-IDF dan BM25 dicatat secara sistematis untuk kemudian dibandingkan nilai akurasinya pada tahap analisis kuantitatif. Proses ini memastikan bahwa pengaruh normalisasi panjang dokumen pada BM25 benar-benar teruji.

4.6. Analisis dan Evaluasi

Tahap akhir dari metodologi ini adalah melakukan pengukuran efektivitas sistem untuk

menentukan algoritma mana yang paling optimal dalam melakukan temu kembali metadata. Evaluasi dilakukan dengan membandingkan luaran sistem terhadap *ground truth* yang telah ditetapkan sebelumnya. Performa diukur menggunakan empat metrik standar, yaitu *Precision*, *Recall*, *F1-Score*, dan *Mean Average Precision* (MAP).

4.6.1. Precision

Precision digunakan untuk mengukur tingkat keakuratan sistem dalam menyajikan dokumen yang relevan dari keseluruhan dokumen yang berhasil diambil (*retrieved*). Nilai presisi yang tinggi menunjukkan bahwa sistem mampu meminimalisir kemunculan jurnal yang tidak relevan pada daftar hasil pencarian. Perhitungan *Precision* dilakukan dengan rumus:

$$\text{Precision} = \frac{\text{Jumlah Metadata Relevan yang Terambil}}{\text{Total Jumlah Metadata yang Terambil}}$$

4.6.2. Recall

Recall mengukur kemampuan sistem dalam menemukan kembali seluruh dokumen relevan yang terdapat di dalam basis data. Metrik ini menunjukkan seberapa lengkap sistem dalam menyajikan jurnal yang sesuai dengan kriteria filter (Sinta, Biaya, Bidang Ilmu) dan kueri pengguna. Rumus yang digunakan adalah:

$$\text{Recall} = \frac{\text{Jumlah Metadata Relevan yang Terambil}}{\text{Total Jumlah Metadata Relevan dalam Basis Data}}$$

4.6.3. F1-Score

Mengingat sering terjadinya *trade-off* antara *Precision* dan *Recall*, maka digunakan *F1-Score* sebagai rata-rata untuk melihat keseimbangan performa algoritma secara keseluruhan. Metrik ini sangat berguna untuk memberikan penilaian tunggal yang objektif terhadap efektivitas algoritma BM25 dan TF-IDF. Perhitungan *F1-Score* adalah sebagai berikut:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.6.4. Mean Average Precision (MAP)

Untuk mengevaluasi kualitas urutan peringkat (*ranking quality*), penelitian ini menerapkan *Mean Average Precision* (MAP). MAP memberikan gambaran mengenai kemampuan algoritma dalam menempatkan dokumen yang paling relevan pada posisi teratas di setiap kueri. MAP dihitung dengan merata-ratakan nilai *Average Precision* (AP) dari seluruh kueri uji (Q):

$$\text{MAP} = \frac{1}{|Q|} \sum_{q=1}^{|Q|} \text{AP}(q)$$

4.6.5. Prosedur Validasi Statistik

Data hasil perhitungan dari kedua algoritma kemudian ditabulasi untuk dilakukan analisis komparatif. Analisis ini bertujuan untuk membuktikan secara empiris apakah penggunaan parameter normalisasi panjang dokumen pada BM25 memberikan peningkatan performa yang signifikan

dibandingkan TF-IDF, terutama pada metadata yang melibatkan variabel teks yang bervariasi.

5. Hasil dan Pembahasan

Pengujian dilakukan terhadap 100 metadata jurnal yang mencakup variabel Nama Jurnal, Singkatan, Penerbit, Bidang Ilmu, hingga Deskripsi. Berdasarkan skenario evaluasi yang melibatkan 20 kueri uji, kinerja algoritma diukur menggunakan metrik *Precision*, *Recall*, *F1-Score*, dan *Mean Average Precision* (MAP).

5.1. Hasil Pengujian Kinerja Algoritma

Hasil perbandingan algoritma TF-IDF dan BM25 disajikan pada Tabel 1.

Tabel 1. Perbandingan Performa Rata-rata Algoritma TF-IDF dan BM25

Algoritma	Precision	Recall	F1-Score	MAP
TF-IDF	0.62	0.58	0.60	0.59
BM25	0.78	0.72	0.75	0.74

Data pada Tabel 1 menunjukkan bahwa BM25 mengungguli TF-IDF di seluruh metrik. Nilai MAP sebesar 0,74 mengindikasikan kemampuan BM25 lebih tinggi dalam menempatkan jurnal yang paling relevan pada urutan teratas, yang sangat krusial bagi pengguna di *platform* Temujurnal.com.

5.2. Analisis Pengaruh Karakteristik Metadata dan Variabel Penerbit

Perbedaan signifikan performa kedua algoritma berakar pada cara mereka menangani atribut metadata tekstual yang heterogen.

1. Berdasarkan Rumus TF-IDF, algoritma ini menghitung bobot $w_{i,j}$ secara linear berdasarkan frekuensi kemunculan term ($tf_{i,j}$). Hal ini menyebabkan bias pada atribut Deskripsi dan Penerbit yang cenderung panjang. Jurnal dengan deskripsi yang berulang-ulang mendapatkan skor lebih tinggi meskipun secara substansi kurang relevan.
2. Keunggulan Probabilistik BM25 adalah penggunaan parameter b untuk normalisasi panjang dokumen $|d|$ sangat efektif dalam menangani metadata Singkatan yang pendek. Parameter ini memastikan jurnal dengan singkatan yang cocok secara eksis (seperti: "ECODEMICA") tetap mendapatkan peringkat tinggi tanpa kalah bersaing dengan jurnal yang memiliki deskripsi teks yang ekspansif.

5.3. Dampak Integrasi Hybrid Retrieval

Sesuai dengan alur penelitian pada Gambar 3, eksekusi eksperimen melibatkan logika *Hybrid Retrieval*. Temuan menunjukkan bahwa integrasi filter Sinta, Periode, dan Biaya sebagai *hard constraints* secara drastis meningkatkan efisiensi pencarian.

1. Filter berhasil mengeliminasi metadata yang tidak memenuhi syarat akreditasi, periode terbit atau anggaran pengguna sebelum algoritma pemeringkatan bekerja.
2. Pada tahap *scoring*, Di dalam sub-set data yang sudah difilter, BM25 terbukti lebih akurat dalam mengolah variabel Penerbit. Penambahan variabel ini memberikan konteks institusi yang membantu pengguna membedakan jurnal dengan nama serupa namun dikelola oleh institusi yang berbeda.

5.4. Analisis Sensitivitas Parameter BM25

Untuk memastikan akuntabilitas hasil, dilakukan pengujian sensitivitas terhadap parameter normalisasi b . Hasil pengujian menunjukkan bahwa nilai $b = 0,75$ adalah titik optimal bagi korpus metadata Temujurnal.com.

Tabel 2. Pengaruh Variasi Parameter b Terhadap Nilai MAP

Nilai b	MAP	Keterangan
0,00	0,65	Tanpa normalisasi panjang dokumen
0,50	0,71	Normalisasi tingkat menengah
0,75	0,74	Nilai Optimal
1,00	0,70	Normalisasi penuh

5.5. Pembahasan

Hasil penelitian ini secara empiris membuktikan bahwa algoritma BM25 memiliki stabilitas dan akurasi yang lebih tinggi dibandingkan TF-IDF dalam mengolah metadata jurnal pada platform Temujurnal.com. Keunggulan ini tercermin pada nilai MAP sebesar 0,74 yang dicapai oleh BM25, mengungguli TF-IDF yang hanya mencapai 0,59. Perbedaan signifikan sebesar 0,15 poin dipicu oleh kemampuan BM25 dalam menangani heterogenitas panjang dokumen melalui parameter normalisasi $b=0,75$. Mekanisme ini mencegah terjadinya bias pada atribut Deskripsi dan Penerbit yang memiliki jumlah kata lebih banyak, sehingga nama jurnal dengan teks yang singkat namun relevan tetap mampu menempati peringkat teratas dalam hasil pencarian.

Selain faktor normalisasi, penggunaan parameter saturasi $k_1=1,2$ pada BM25 terbukti efektif dalam meredam dominasi kata yang muncul berulang kali pada bagian deskripsi jurnal. Hal ini berdampak langsung pada nilai *Precision* sebesar 0,78 menunjukkan sistem mampu meminimalisir luaran jurnal yang tidak relevan dibandingkan model TF-IDF yang *Precision* hanya mencapai nilai 0,62. Integrasi variabel Penerbit memberikan dimensi pencarian yang lebih kuat, mekanisme pembobotan IDF pada BM25 mampu mengenali keunikan nama instansi atau universitas sebagai entitas penting, didukung oleh proses *stemming* bahasa Indonesia

yang akurat untuk memastikan kata dasar institusi diberikan bobot yang tepat secara statistik.

Dalam *Hybrid Retrieval*, algoritma ini bersinergi secara optimal dengan fitur filter seperti peringkat Sinta, periode terbit dan biaya publikasi. Meskipun filter tersebut berfungsi sebagai pembatas kaku (*hard constraint*), algoritma BM25 memegang peranan vital dalam mengurutkan hasil akhir dengan nilai F1-Score mencapai 0,75. Tingginya nilai efektivitas menunjukkan bahwa sistem tidak hanya sekadar menyaring data berdasarkan kategori, tetapi juga mampu memposisikan jurnal yang paling sesuai dengan kebutuhan riset peneliti di posisi peringkat teratas (*top-rank*). Hasil ini memberikan rekomendasi teknis bahwa implementasi BM25 sangat layak digunakan untuk meningkatkan efisiensi proses penemuan wadah publikasi ilmiah pada agregator jurnal nasional.

6. Kesimpulan dan Saran

6.1. Kesimpulan

Berdasarkan hasil analisis dan eksperimen yang telah dilakukan dalam membandingkan algoritma BM25 dan TF-IDF untuk temu kembali metadata jurnal pada *platform* Temujurnal.com, maka dapat ditarik beberapa kesimpulan.

1. Algoritma BM25 terbukti lebih unggul dibandingkan TF-IDF dalam seluruh metrik evaluasi yang digunakan. Hal ini dibuktikan dengan perolehan nilai MAP sebesar 0,74 yang melampaui TF-IDF dengan nilai 0,59. BM25 menunjukkan tingkat akurasi yang lebih tinggi dengan nilai *Precision* sebesar 0,78 dan F1-Score sebesar 0,75.
2. Keunggulan BM25 terletak pada penggunaan parameter normalisasi panjang dokumen ($b=0,75$) yang mampu menangani heterogenitas metadata jurnal secara stabil. Mekanisme ini memastikan bahwa atribut yang singkat seperti Nama Jurnal dan Singkatan tidak terdominasi oleh atribut yang lebih panjang seperti Penerbit dan Deskripsi, sehingga peringkat relevansi yang dihasilkan menjadi lebih adil dan akurat.
3. Implementasi logika *Hybrid Retrieval* yang menggabungkan filter Peringkat Sinta, Periode Terbit dan Biaya Publikasi sebagai *hard constraints* dengan algoritma pembobotan probabilistik terbukti sangat efektif. Sistem mampu menyaring data sesuai kriteria pengguna, sementara BM25 berperan dalam mengurutkan hasil akhir berdasarkan kedekatan kueri terhadap metadata teksual jurnal.

6.2. Saran

Meskipun penelitian memberikan hasil yang positif, terdapat beberapa saran pengembangan yang dapat dilakukan untuk penelitian selanjutnya guna meningkatkan kualitas sistem temu kembali informasi pada domain metadata jurnal ilmiah.

1. Mengingat penelitian ini menggunakan sampel 100 metadata jurnal, disarankan agar penelitian selanjutnya memperluas cakupan dataset hingga mencakup seluruh populasi jurnal nasional yang terindeks di SINTA atau Garuda. Perluasan ini penting untuk menguji konsistensi, efisiensi komputasi, dan skalabilitas algoritma BM25 pada volume data yang jauh lebih besar dan heterogen.
2. Disarankan untuk melakukan optimasi parameter k_1 dan b pada BM25 menggunakan metode pencarian parameter otomatis (seperti *Grid Search*) agar mendapatkan nilai yang paling presisi sesuai dengan karakteristik bahasa Indonesia yang terus berkembang.
3. Keterbatasan algoritma berbasis frekuensi kata dalam menangani sinonim, pengembangan sistem di masa depan dapat mengintegrasikan pendekatan *Semantic Search* berbasis *Deep Learning* (seperti BERT atau Word2Vec) untuk menangkap makna kontekstual di balik kueri pengguna.
4. Untuk memperkuat akuntabilitas perbandingan, penelitian masa depan dapat menyertakan uji signifikansi statistik seperti *Paired Sample T-test* atau *Wilcoxon Signed-Rank Test*.

Daftar Pustaka

- [1] M. R. K. Muluk and A. Widjaja, "Challenges and Opportunities for Research Discovery," *Journal of Academic Librarianship*, vol. 48, no. 2, pp. 102-110, 2022.
- [2] A. Nasution et al., "Implementasi Algoritma Lexical Chain pada Aplikasi Pencarian Jurnal SINTA Berbasis Web," *Jurnal Teknologi Terapan & Sains 4.0*, vol. 6, no. 2, pp. 112-125, 2025.
- [3] H. Santoso, "Analisis Faktor Pemilihan Jurnal Publikasi pada Dosen di Indonesia," *Jurnal Riset Teknologi Informasi*, vol. 7, no. 1, pp. 45-58, 2023.
- [4] T. Kheng, J. S. Asri, S. Wahyu, and Y. Yulhendri, "Penerapan Algoritma BM25 dalam Pencarian Lowongan Pekerjaan pada Website Job Portal," *Bulletin of Computer Science Research*, vol. 5, no. 5, pp. 1029-1038, 2025.
- [5] A. M. Diningrat, "Optimasi Algoritma Pencarian Dokumen Akademik Menggunakan BM25 dan TF-IDF," *Infomans: Jurnal Ilmu-ilmu Informatika dan Manajemen*, vol. 19, no. 1, pp. 1-10, 2025.
- [6] R. Pratama, "Perkembangan Sistem Temu Kembali Informasi di Era Big Data Indonesia," *Jurnal Informatika dan Sistem Informasi*, vol. 3, no. 2, pp. 88-97, 2022.
- [7] Indriati, I. Rahayudi, and N. H. L. Dewi, "Perbandingan Algoritma TF-IDF dan BM25

- dalam Menentukan Relevansi Dokumen pada Sistem Temu Kembali Informasi,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK)*, vol. 5, no. 11, pp. 4820-4828, 2021.
- [8] F. Aris, “Pentingnya Metadata dalam Manajemen Reposisi Institusi,” *Jurnal Dokumentasi dan Informasi*, vol. 45, no. 1, pp. 12-25, 2024.
- [9] N. Hidayat, “Pengaruh Morfologi Bahasa Indonesia terhadap Akurasi Temu Kembali Informasi pada Dokumen Teks,” *Jurnal Sistem Cerdas*, vol. 7, no. 1, pp. 30-42, 2024.
- [10] S. E. Robertson and H. Zaragoza, “The Probabilistic Relevance Framework: BM25 and Beyond,” *Foundations and Trends in Information Retrieval*, vol. 14, no. 4, pp. 333-469, 2021.
- [11] R. Pratama and S. Fitriani, “Analisis Kinerja Sistem Temu Kembali Informasi pada Dokumen Teks Berbahasa Indonesia,” *Jurnal Informatika dan Sistem Informasi*, vol. 4, no. 1, pp. 22-34, 2023.
- [12] F. Aris, “Pentingnya Struktur Metadata dalam Manajemen Reposisi Institusi Digital,” *Jurnal Dokumentasi dan Informasi*, vol. 45, no. 1, pp. 12-25, 2024.
- [13] H. Santoso, “Implementasi Pembobotan TF-IDF untuk Pencarian Katalog Buku Perpustakaan,” *Jurnal Riset Teknologi Informasi*, vol. 7, no. 1, pp. 45-58, 2023.
- [14] A. P. Utomo and R. Hartanto, “Evaluasi Algoritma Penentuan Relevansi Dokumen pada Search Engine,” *Jurnal Teknologi Informasi*, vol. 18, no. 2, pp. 120-132, 2022.
- [15] S. Saadah and H. L. H. S. Warnars, “The Performance of BM25 and TF-IDF in Indonesian Information Retrieval System,” *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 2, pp. 431-442, 2023.
- [16] N. Hidayat, “Pengaruh Morfologi Bahasa Indonesia terhadap Akurasi Temu Kembali Informasi,” *Jurnal Sistem Cerdas*, vol. 7, no. 1, pp. 30-42, 2024.
- [17] M. Adi and L. Hakim, “Efektivitas Algoritma Stemming Sastrawi pada Dokumen Berita Online,” *Jurnal Komputer dan Teknologi*, vol. 5, no. 3, pp. 210-218, 2021.
- [18] R. Kurniawan, “Metrik Evaluasi pada Sistem Temu Kembali Informasi: Sebuah Tinjauan Literatur,” *Jurnal Ilmiah Teknik Informatika*, vol. 15, no. 2, pp. 88-99, 2022.
- [19] J. Siswanto, “Perbandingan Performa BM25 dan VSM pada Sistem Pencarian Skripsi,” *Jurnal Pengembangan Teknologi Informasi*, vol. 8, no. 1, pp. 12-21, 2024.
- [20] D. Rahmawati and E. Winarko, “Performance Comparison of Information Retrieval Models for Indonesian Documents,” *Indonesian Journal of Computing and Cybernetics Systems (IJCCS)*, vol. 16, no. 4, pp. 345-356, 2022.