

SISTEM REKOMENDASI PENELUSURAN BUKU BERBASIS CONTENT-BASED FILTERING DENGAN PEMBOBOTAN TF-RF

Sadesty Rahmadhani¹, Lutfi Hakim², Galih Hendra Wibowo³, Sepyan Purnama Kristanto⁴,
Eka Mistiko Rini⁵

^{1,2,4,5}Program Studi Teknologi Rekayasa Perangkat Lunak, ³Program Studi Teknologi Rekayasa Komputer,
Jurusan Bisnis dan Informatika, Politeknik Negeri Banyuwangi
¹sadestyrhmadhani012@gmail.com, ²lutfi@poliwangi.ac.id, ³galih@poliwangi.ac.id,
⁴sepyan@poliwangi.ac.id, ⁵ekamrini@poliwangi.ac.id

Abstrak

Aplikasi manajemen buku di perpustakaan biasanya menyediakan berbagai buku digital yang dapat diakses secara *online*. Ini memungkinkan pengguna menggunakan fitur pencarian yang didukung teknologi sistem rekomendasi untuk menemukan buku sesuai minat mereka. Penelitian ini mengembangkan sistem rekomendasi pencarian buku dengan metode *Content-Based Filtering* yang dikombinasikan dengan algoritma *Term Frequency-Relevance Frequency* (TF-RF) untuk meningkatkan akurasi sistem pencarian pada aplikasi *E-Library* Perpustakaan Politeknik Negeri Banyuwangi. Pada penelitian ini, 6148 judul buku berbahasa Indonesia digunakan sebagai dataset. Proses pengembangan melibatkan beberapa tahap, yaitu pengumpulan data, *preprocessing*, pembobotan TF-RF, perhitungan kemiripan menggunakan *Cosine Similarity*, pengambilan nilai tertinggi dengan *Quick Sort*, dan evaluasi dengan mencari nilai presisi. Setelah melalui tahap evaluasi menunjukkan bahwa sistem rekomendasi yang dikembangkan memiliki nilai presisi rata-rata 86%, dibandingkan dengan 51% pada sistem pencarian buku di aplikasi *E-Library* yang ada. Hasil ini menunjukkan bahwa metode *Content-Based Filtering* dengan algoritma TF-RF dapat secara signifikan memperbaiki akurasi dan relevansi sistem pencarian buku, meskipun masih ada keterbatasan dalam menangani kata kunci yang tidak lengkap atau lebih dari satu suku kata. Penelitian ini menyimpulkan bahwa implementasi metode *Content-Based Filtering* dan TF-RF dapat meningkatkan performa sistem rekomendasi pencarian buku secara signifikan, menawarkan solusi yang lebih efektif dan efisien bagi pengguna dalam menemukan buku yang diinginkan di *E-Library* Perpustakaan Politeknik Negeri Banyuwangi.

Kata kunci : sistem pencarian, buku perpustakaan, analisis sentimen, *cosine similarity*, pemrosesan data teks

1. Pendahuluan

Aplikasi manajemen buku di perpustakaan, sudah sangat umum menyediakan beragam buku dalam bentuk digital yang dapat diakses secara *online*. Sehingga memungkinkan pengguna untuk memanfaatkan fitur pencarian dengan teknologi sistem rekomendasi untuk mencari buku yang sesuai dengan minat mereka. Sistem rekomendasi pencarian buku adalah perangkat lunak yang memberikan masukan atau rekomendasi buku kepada pengguna berdasarkan masukan pertanyaan atau kriteria yang dimasukkan oleh pengguna. Sistem rekomendasi merupakan sebuah teknik yang bertujuan untuk memberi saran kepada pengguna tentang *item* yang sebaiknya dipilih (Murti et al., 2019). Seperti halnya sistem pencarian buku yang terdapat pada aplikasi *E-Library* di Perpustakaan Politeknik Negeri Banyuwangi, yang bertujuan untuk mempermudah pengguna dalam menemukan buku yang diinginkan sesuai dengan preferensi mereka.

Namun untuk menemukan buku yang dicari pada sistem pencarian tersebut, tidak semua kata kunci yang dimasukkan oleh pengguna akan

menghasilkan rekomendasi pencarian buku yang optimal. Melalui analisis yang telah dilakukan pada sistem tersebut, ditemukan bahwa terdapat salah satu kasus ketika pengguna hendak memasukkan kata kunci berupa judul lengkap buku atau kata kunci yang lebih spesifik, maka hasil yang diberikan oleh sistem berupa pesan "Data Kosong" yang berarti buku yang dicari tidak tersedia. Akan tetapi, sebenarnya judul buku tersebut telah tersedia pada database. Hal ini dapat dibuktikan dengan melakukan percobaan pencarian menggunakan beberapa suku kata saja dari judul buku yang diinginkan, maka data buku yang awalnya tidak ada akan masuk ke dalam list data buku yang direkomendasikan sistem. Kasus tersebut terjadi disebabkan oleh metode atau algoritma yang digunakan kurang sesuai sehingga dapat mempengaruhi akurasi yang diberikan oleh sistem menjadi rendah. Untuk mengoptimalkan sistem pencarian buku pada aplikasi *E-Library* penelitian ini menerapkan metode *Content-Based Filtering* dalam pengembangan sistem rekomendasi yang akan dibuat.

Penelitian sebelumnya terkait rekomendasi buku telah dilakukan oleh (Alkaff et al., 2020) menggunakan metode *Content-Based Filtering* yang

digunakan untuk mencari kemiripan antara buku yang dicari dan buku yang tersedia. Kemudian merekomendasikan 5 buku paling relevan menggunakan metode *Weighted Tree Similarity*. Pada penelitian tersebut menghasilkan nilai presisi mencapai 88%. Pada penelitian yang dilakukan oleh (Putri et al., 2020) pada pengembangan sistem rekomendasi produk pena menggunakan metode *Content-Based Filtering* dan TF-IDF, yang menggunakan beberapa parameter kriteria pena yang akan dibeli oleh pengguna dan kemudian sistem akan memberikan rekomendasi berdasarkan kecocokan ketersediaan produk dengan kriteria pena yang diinginkan oleh pengguna, dengan nilai akurasi mencapai 96,5%.

Selain itu, penelitian ini memanfaatkan algoritma *Term Frequency-Relevance Frequency* (TF-RF) untuk melakukan evaluasi sebuah dokumen dengan mempertimbangkan seberapa sering kata kunci muncul dalam konteks kategori yang relevan (Joergensen Munthe et al., 2022). Tujuannya adalah untuk mendapatkan tingkat performa yang lebih baik dibandingkan dengan metode pembobotan kata lainnya. Dalam penggunaan algoritma juga telah dilakukan penelitian sebelumnya oleh (Sari et al., 2022), dengan melakukan perbandingan TF-IDF dan TF-RF yang digabungkan dengan metode *Weighted Tree Similarity*. Hasilnya menunjukkan bahwa TF-RF memiliki tingkat akurasi yang lebih tinggi dibanding *Term Frequency – Inverse Document Frequency* (TF-IDF) hingga mencapai 98%.

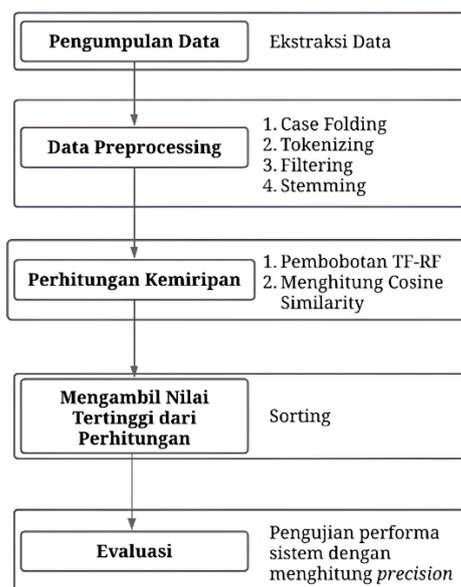
Metode *Content-Based Filtering* dapat digunakan pada beberapa konteks atau aplikasi berbasis teks. Selain itu, metode ini memiliki prinsip memberikan rekomendasi berdasarkan kemiripan item profile (Mondi & Wijayanto, n.d.). Adapun beberapa metrik kesamaan yang umum digunakan dalam hal ini termasuk *cosine similarity*, *adjusted cosine similarity*, dan *pearson coefficient* (Pramesti & Santiyasa, 2022). Berdasarkan tingkat akurasi yang relatif tinggi pada penelitian sebelumnya dan prinsip yang dimiliki oleh metode *Content-Based Filtering* serta algoritma TF-RF, diharapkan dengan mengadopsi kedua pendekatan tersebut dapat memperbaiki akurasi dan relevansi sistem pencarian buku secara signifikan.

Penelitian ini dilakukan dengan tujuan mengembangkan sistem rekomendasi pencarian buku menggunakan metode *Content-Based Filtering* yang dikombinasikan dengan algoritma TF-RF untuk meningkatkan akurasi sistem pencarian pada aplikasi *E-Library* Perpustakaan Politeknik Negeri Banyuwangi. Serta melakukan evaluasi dengan menguji performa (nilai presisi) sistem rekomendasi pencarian buku yang menerapkan kedua pendekatan tersebut untuk mengetahui apakah penerapan kedua pendekatan tersebut dapat meningkatkan akurasi pada sistem pencarian *E-Library*. Adapun batasan yang diberikan pada penelitian ini mencakup fokus pada implementasi dan evaluasi metode *Content-*

Based Filtering dengan algoritma TF-RF untuk sistem rekomendasi menggunakan data buku dari aplikasi *E-Library* di Perpustakaan Politeknik Negeri Banyuwangi, tanpa melibatkan aspek lain dari aplikasi *E-Library* seperti manajemen pengguna dan infrastruktur teknologi informasi. Penggunaan data terbatas pada data buku yang tersedia di Perpustakaan Politeknik Negeri Banyuwangi dengan judul yang berbahasa Indonesia.

2. Metode Penelitian

Proses pembangunan sistem rekomendasi pencarian buku ini menggunakan metode *Content-Based Filtering* yang digabungkan dengan algoritma TF-RF. Metode *Content-Based Filtering* adalah salah satu teknik sistem rekomendasi yang menggunakan fitur, atribut, atau karakteristik sebuah *item* untuk memberikan rekomendasi kepada pengguna dan berprinsip memberikan rekomendasi berdasarkan kemiripan *item profile* (Mondi & Wijayanto, n.d.), (Pramesti & Santiyasa, 2022). Dalam penerapan metode ini, terdapat beberapa tahapan yang harus dilewati mulai dari pengumpulan data, data *preprocessing*, perhitungan kemiripan, mengambil nilai tertinggi dari perhitungan, dan evaluasi, seperti yang tertera pada Gambar 1. Beberapa tahapan tersebut menggunakan bahasa pemrograman PHP yang menyesuaikan dengan sistem *E-Library* guna mempermudah dalam proses implementasi. Alur metode *Content-Based Filtering* dapat dilihat pada Gambar 1 di bawah ini.



Gambar 1. Alur Metode *Content-Based Filtering*

Tahapan pertama adalah pengumpulan data yang diambil dari database dengan menggunakan metode ekstraksi data. kemudian memproses data sebelum dilakukan perhitungan kemiripan dengan melewati beberapa tahapan di dalamnya seperti *Case Folding*, *Tokenizing*, *Filtering*, dan *Stemming*,

sehingga data siap untuk dilanjutkan ke proses selanjutnya. Tahapan berikutnya adalah pembobotan setiap judul dokumen dengan menggunakan algoritma TF-RF. Selain itu, pada tahapan ini dilakukan pengukuran kemiripan tiap judul dokumen dengan memanfaatkan perhitungan *cosine similarity*. Pada tahap ke empat ini merupakan proses pengambilan nilai tertinggi dari hasil perhitungan dengan memanfaatkan algoritma pengurutan data (*sorting*). Di tahap terakhir yaitu tahap evaluasi yang berisi dari serangkaian pengujian yang dilakukan untuk mengevaluasi hasil dari algoritma yang telah diterapkan. Pada tahap ini dilakukan pengujian dengan menghitung nilai presisi.

2.1 Pengumpulan Data

Ekstraksi data merupakan teknik pengambilan data dengan mengekstrak data dari *database* menggunakan *Structured Query Language (SQL)* (Amna et al., 2023). Pada tahap ini dilakukan pengambilan data buku sebagai pendukung sistem rekomendasi dalam penelitian ini. Data yang digunakan merupakan kumpulan informasi mengenai buku-buku yang terdapat dalam *database* aplikasi *E-Library* di Perpustakaan Politeknik Negeri Banyuwangi. Teknik ekstraksi data ini dapat melibatkan penggunaan berbagai bahasa pemrograman dan perangkat lunak seperti SQL untuk mengambil data dari *database* relasional.

2.2 Data Pre-Processing

Pre-processing merupakan langkah awal dalam klasifikasi teks yang bertujuan untuk menyiapkan data teks sebelum diolah ke dalam proses berikutnya (Khairunnisa et al., 2021), (Kristanto et al., 2022). Tujuan utama dari pra pemrosesan data adalah untuk menghapus variabilitas atau efek yang tidak diinginkan sehingga informasi yang berkaitan dengan properti yang diminati dapat digunakan untuk pemodelan yang efisien (Mishra et al., 2020). Pada tahap ini, data teks disusun kembali untuk menghasilkan informasi teks dengan kualitas yang optimal dan siap untuk digunakan dalam proses selanjutnya. Untuk mendukung *Pre-processing* data tersebut, memanfaatkan teknologi *Natural Language Processing (NLP)*. NLP adalah kecerdasan buatan, yang ditujukan agar komputer dapat memahami pernyataan atau kata-kata yang ditulis dalam bahasa manusia (Khurana et al., 2023). Dalam penerapan NLP mencakup beberapa tahapan, seperti *case folding*, *tokenizing*, *filtering*, dan *stemming*. Tahapan-tahapan yang dilalui pada *pre-processing* data ini dilakukan dengan bahasa pemrograman PHP sehingga memanfaatkan salah satu library yang dapat mendukungnya yaitu *NLPTools*.

1. Case folding

Pada tahap *case folding* dilakukan perubahan terhadap huruf kapital menjadi huruf kecil

(Laksono et al., 2020). Hal ini bertujuan untuk menciptakan keseragaman antar kata dalam judul dokumen data buku.

2. Tokenizing

Tahap *tokenizing* adalah tahap pemotongan kalimat berdasarkan tiap kata yang menyusunnya. Selain itu, spasi digunakan untuk memisahkan antar kata (Mukhtar et al., 2022). Pada langkah ini, data dari setiap buku akan mengalami tahap pembersihan untuk menghilangkan simbol, tanda baca, *backslash*, URL, serta mengganti *multiple whitespace* menjadi *single whitespace* dan *single char*. Dengan membersihkan data, judul dokumen yang akan di-*tokenize* dapat memiliki ukuran yang lebih kecil, memungkinkan proses berlangsung lebih efisien.

3. Filtering

Filtering adalah langkah yang diambil untuk mengekstrak kata-kata kunci yang terdapat dalam judul dokumen data buku. Pada tahap ini, dilakukan eliminasi kata-kata yang tidak memiliki nilai deskriptif, termasuk kata penghubung, kata depan, dan kata ganti seperti "di", "yang", "dan", "ke", dan "dari". Tujuan dari proses *filtering* adalah untuk mengurangi jumlah kata dalam judul dokumen, memungkinkan proses pengolahan lebih akurat dengan memfokuskan pada kata-kata yang memiliki signifikansi di dalam judul dokumen.

4. Stemming

Stemming merupakan langkah untuk menemukan akar kata dari setiap kata yang telah melewati proses *filtering*. Pada tahap ini, dilakukan upaya untuk mengembalikan berbagai bentuk kata ke dalam representasi yang seragam. Proses pencarian kata dasar dilakukan dengan menghapus semua kata imbuhan, baik itu awalan, sisipan, atau akhiran, dari kata-kata tersebut.

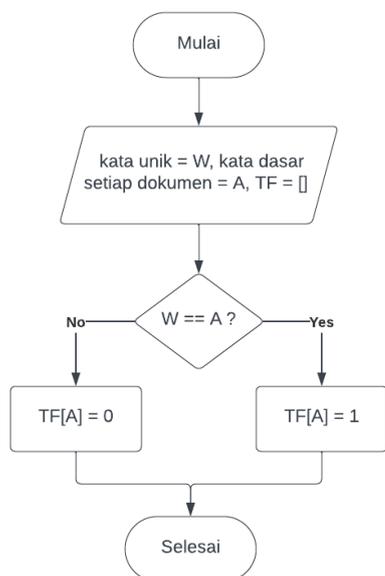
2.3 Perhitungan Kemiripan

Penghitungan skor kemiripan antar buku dilakukan melalui dua tahap. Tahap pertama melibatkan penggunaan algoritma TF-RF untuk memberikan bobot pada setiap term dalam tiap judul buku. Dan tahap kedua melibatkan perhitungan kemiripan antara buku dengan menggunakan algoritma *Cosine Similarity*.

1. Pembobotan TF-RF

TF-RF adalah metode yang menggabungkan *Term Frequency (TF)* dan *Relevance Frequency (RF)* untuk mengevaluasi relevansi dokumen dengan mempertimbangkan seberapa sering kata muncul dalam kategori yang bersangkutan (Dananjaya & Indradewi, 2023). Pembobotan ini dilakukan secara spesifik pada kriteria judul yang telah melalui tahap *pre-processing*, melalui beberapa tahap perhitungan hingga

diperoleh nilai similarity dari setiap dokumen (judul buku). Adapun diagram alir pembobotan nilai *Term Frequency* (TF) seperti yang terlihat pada Gambar 2 berikut.



Gambar 2. Diagram Alir Pembobotan TF

Berdasarkan gambar diagram alir di atas, nilai TF dihitung dengan mencocokkan setiap judul buku dengan kata unik dari keseluruhan judul buku. Setelah itu, dilakukan proses perhitungan RF dan TF-RF dengan mengimplementasikan persamaan (1) dan (2) menggunakan data yang relevan.

$$RF = \log \left(2 + \frac{b}{\max(1,c)} \right) \quad (1)$$

$$TF - RF(t, d) = TF(t, d) * RF \quad (2)$$

TF = frekuensi kemunculan istilah t dalam pada judul dokumen (d)

RF = relevansi istilah t dalam seluruh koleksi pada judul dokumen

t = istilah atau kata kunci yang sedang dipertimbangkan

d = judul dokumen yang sedang dievaluasi

b = jumlah judul dokumen yang mengandung istilah atau kata kunci (*term*)

c = jumlah judul dokumen yang tidak mengandung istilah atau kata kunci (*term*)

2. Cosine Similarity

Cosine Similarity adalah teknik yang digunakan untuk mengukur tingkat kesamaan antara dua objek, yang direpresentasikan dalam bentuk vektor, dengan menggunakan kata kunci (*keyword*) dari sebuah dokumen sebagai acuan (Apriani et al., 2021), (Sanjaya et al., 2023). Sebelum melanjutkan ke perhitungan *cosine similarity*, perlu dilakukan perhitungan untuk mendapatkan nilai DxQ (TF-RF setiap kata pada judul buku dikalikan dengan TF-RF setiap kata

pada kata kunci) sebagai langkah awal dalam menentukan tingkat kemiripan antara dokumen dan *query* (Q) dan perhitungan untuk menemukan nilai akar kuadrat dari total kuadrat TF-RF setiap judul buku sebagai langkah penting untuk membantu meningkatkan akurasi dan relevansi. Setelah melalui kedua perhitungan awal tersebut, selanjutnya dilakukan perhitungan *cosine similarity* yang dihitung berdasarkan persamaan (3).

$$\cos(\theta_{ij}) = \frac{\sum_k (d_{ik}d_{jk})}{\sqrt{\sum_k d_{ik}^2} \sqrt{\sum_k d_{jk}^2}} \quad (3)$$

$\cos(\theta_{ij})$ = Nilai *cosine similarity* antara dua judul dokumen i dan j

$\sum_k (d_{ik}d_{jk})$ = Hasil penjumlahan perkalian titik antara komponen vektor dari judul dokumen i dan j (DxQ).

$\sqrt{\sum_k d_{ik}^2} \sqrt{\sum_k d_{jk}^2}$ = Norma *Euclidean* dari masing – masing vektor judul dokumen i dan j (Kuadrat TF-RF).

2.4 Mengambil Nilai Tertinggi

Pada tahap ini, dilakukan pengambilan nilai tertinggi dari hasil perhitungan *similarity* pada setiap judul buku dengan memanfaatkan teknik struktur data *Quick sort* yang merupakan algoritma pengurutan dan memungkinkan untuk menyusun hasil perhitungan *similarity* secara terurut berdasarkan nilai – nilai tertinggi hingga terendah. Metode ini menggunakan konsep *divide and conquer*, di mana data dibagi menjadi bagian-bagian kecil dan menggunakan *pivot* sebagai referensi untuk pengurutan (Al Rivian, 2017). Dari pengurutan tersebut didapatkan data buku yang telahurut berdasarkan nilai *similarity* dengan kata kunci yang dimasukkan.

2.5 Evaluasi

Dalam tahap evaluasi penting untuk mengukur kinerja sistem dengan menghitung nilai *precision*. *Precision* adalah salah satu metrik penting dalam evaluasi sistem pengambilan informasi yang mengukur tingkat ketepatan atau akurasi sistem dalam menemukan dokumen yang relevan dengan kebutuhan pengguna. Dalam konteks ini, *precision* memberikan gambaran tentang seberapa efektif sistem dalam memberikan hasil buku yang relevan bagi pengguna, dengan menghitung proporsi dari data judul buku *True Positive* (TP) dibagi dengan data judul buku *True Positive* (TP) ditambah data judul buku *False Positive* (FP). Judul buku yang relevan dengan kata kunci adalah judul buku yang mengandung term (kata yang muncul) pada kata kunci. Dengan kata lain, semakin tinggi nilai *precision*, semakin sedikit dokumen yang tidak relevan yang akan disertakan dalam hasil pencarian,

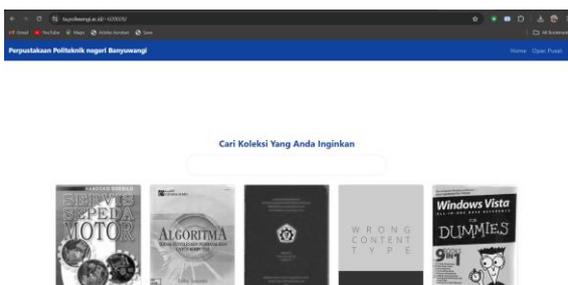
yang secara langsung meningkatkan pengalaman pengguna dan kemudahan dalam menemukan informasi yang diinginkan. Nilai presisi yang diharapkan dapat dihitung menggunakan persamaan (4).

$$Precision = \frac{TP (True Positive)}{TP (True Positive) + FP (False Positive)} \quad (4)$$

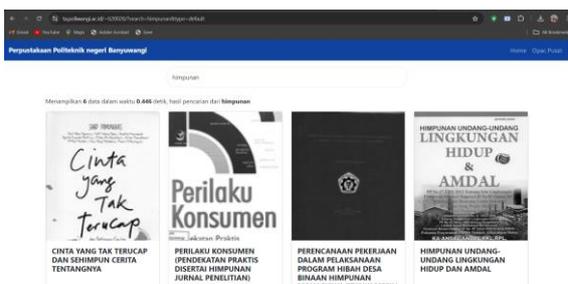
TP = data judul buku relevan yang muncul
 FP = data judul buku tidak relevan yang muncul

3. Hasil dan Pembahasan

Data yang digunakan pada penelitian ini, sejumlah 6148 data judul buku murni berbahasa Indonesia yang tersedia pada aplikasi *E-Library* di Perpustakaan Politeknik Negeri Banyuwangi yang telah melalui tahap penyaringan secara manual. Berikut merupakan tampilan sistem dari hasil penerapan metode *Content-Based Filtering* seperti yang terlihat pada Gambar 3 dan 4 di bawah ini.



Gambar 3. Halaman Utama Sistem Rekomendasi Pencarian Buku



Gambar 4. Halaman Hasil Pencarian

Gambar di atas merupakan hasil penerapan metode *Content-Based Filtering* pada *website* sederhana sistem rekomendasi pencarian buku. Pada halaman pertama pengguna disajikan tampilan sederhana yang berisi kotak inputan untuk memasukkan kata kunci atau judul buku yang ingin dicari oleh pengguna. Di bawahnya diikuti dengan 5 rekomendasi buku yang diberikan sistem kepada pengguna diambil dari pencarian terakhir. Pada saat pengguna memasukkan kata kunci dan menekan tombol “enter” di halaman kedua sistem memberikan hasil rekomendasi berdasarkan kata kunci yang dimasukkan pada halaman pertama. Selain itu, sistem juga menampilkan jumlah buku yang relevan atau ditampilkan dan jumlah waktu eksekusi yang

dibutuhkan pada saat melakukan pencarian dengan kata kunci tersebut.

Untuk mendapatkan hasil seperti yang terlihat pada Gambar 4, data judul buku yang digunakan pada penelitian ini melalui beberapa tahapan sehingga mendapatkan hasil similaritas dengan menerapkan metode *Content-Based Filtering* yang dikombinasikan dengan algoritma TF-RF. Pengolahan data pada penelitian ini dilakukan menggunakan bahasa pemrograman PHP dan memanfaatkan beberapa *library* seperti *NLPTools* yang digunakan untuk *pre-processing* dengan *corpus sastra* dan *voku*. Sebelum data diproses perhitungan kemiripan, data tersebut perlu dilakukan *pre-processing* dengan hasil berupa kata dasar dari judul buku. Proses tersebut untuk mempermudah pada tahap perhitungan kemiripan agar teks yang diolah dapat diidentifikasi secara konsisten. Kemudian, dilakukan perhitungan kemiripan yang diawali dari pembobotan TF-RF pada setiap kata yang terdiri dari judul buku dan kata kunci yang dimasukkan “himpunan”. Adapun hasil Pre-Processing dapat lihat pada Tabel 1 di bawah ini.

Tabel 1. Hasil *Pre-Processing* dari Setiap Judul Buku yang Dicari

	Judul Buku	Hasil <i>Pre-processing</i>
A1	Himpunan Undang-Undang Perdata Islam & Peraturan Pelaksanaan Lainnya Di Negara Hukum Indonesia	himpun, undang, undang, perdata, islam, atur, laksana, negara, hukum, indonesia
A2	Perilaku Konsumen (Pendekatan Praktis Disertai Himpunan Jurnal Penelitian)	perilaku, konsumen, dekat, praktis, serta, himpun, jurnal, teliti
A3	Himpunan Peraturan Perundang-Undangan Republik Indonesia Tentang Ketenagakerjaan	himpun, atur, undang, undang, republik, indonesia, ketenagakerjaan
A4	Perencanaan Pekerjaan Dalam Pelaksanaan Program Hibah Desa Binaan Himpunan Mahasiswa Teknik Mesin Tahun 2020	rencana, kerja, laksana, program, hibah, desa, bina, himpun, mahasiswa, teknik, mesin, tahun, 2020
A5	Cinta Yang Terucap Dan Sehimpun Cerita Tentangnya	cinta, ucap, himpun, cerita, tentang
A6	Himpunan Undang-Undang Lingkungan Hidup Amdal	himpun, undang, undang, lingkungan, hidup, amdal

Sebelum melakukan perhitungan *Cosine Similarity*, tahap yang krusial dalam pengembangan sistem pencarian ini adalah tahap pembobotan TF-RF yang dapat dilihat pada Tabel 2 dan 3 yang tertera pada akhir naskah. Setelah melewati tahapan pembobotan TF-RF seluruh kata pada setiap dokumen memiliki nilai sebagai bobot yang dihasilkan dari perhitungan TF-RF, yang akan digunakan untuk perhitungan *Cosine Similarity* berdasarkan persamaan (3). Kemudian hasil nilai similaritas diurutkan menggunakan struktur data *sorting (Quick Sort)* untuk merekomendasikan judul buku yang memiliki nilai similaritas paling tinggi

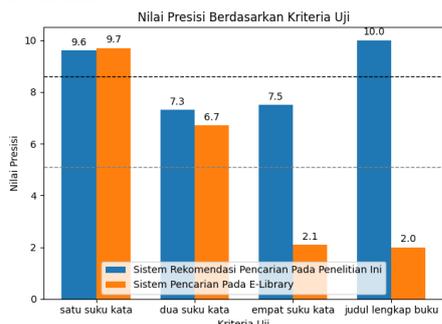
diurutan teratas. Hasil Cosine Similarity yang telah diurutkan dapat dilihat pada Tabel 4 di bawah ini.

Tabel 4. Hasil *Cosine Similarity* dengan Kata Kunci Himpunan

	Cosine Similarity
A5	0.81777406423499
A2	0.7318979095121
A4	0.62983093088692
A6	0.61178342766905
A3	0.59074585568585
A1	0.55284036445038

Tabel di atas menunjukkan nilai *Cosine Similarity* antara kata kunci yang dimasukkan dengan enam judul buku yang tersedia (A1 hingga A6). Berdasarkan hasil perhitungan, terdapat 6 judul buku yang direkomendasikan dengan kata kunci "himpunan". A5 (dokumen ke lima) merupakan buku yang dianggap paling relevan dengan nilai similaritas tertinggi 0,817774. Kemudian diikuti oleh A2 (0.7319), A4 (0.6298), A6 (0.6118), A3 (0.5907), dan A1 dengan nilai terendah (0.5528). Semakin tinggi nilai Cosine Similarity, semakin mirip dokumen/item tersebut. Nilai berkisar antara 0 dan 1, di mana 1 menunjukkan kesamaan sempurna.

Untuk meninjau lebih jauh terkait relevansi hasil rekomendasi yang diberikan oleh sistem yang telah dibuat, evaluasi adalah tahap yang krusial dalam kesesuaian sistem merekomendasikan buku kepada pengguna. Sistem rekomendasi pada penelitian ini telah diuji dengan beberapa tahapan evaluasi termasuk mencari nilai presisi berdasarkan persamaan (4). Dalam mengukur nilai presisi melibatkan beberapa kriteria uji dari memasukkan kata kunci "satu suku kata", "dua suku kata", "empat suku kata", dan "judul lengkap buku" yang dipilih secara random. Selain itu, tahap ini dievaluasi dengan melakukan analisis secara manual pada 10 rekomendasi buku teratas yang disarankan oleh sistem dengan catatan jika sistem menampilkan rekomendasi buku tidak sampai 10, maka jumlah buku yang tidak di tampilkan akan masuk ke dalam *False Positive* (FP). Namun, berbeda dengan kata kunci "judul lengkap buku", pada kriteria uji ini hanya menganalisis 1 buku yang relevan dengan kata kunci. Perbedaan hasil perbandingan yang signifikan dari sistem pencarian yang telah dikembangkan dan sistem pencarian *E-library* dapat dilihat pada Gambar 5 di bawah ini.



Gambar 5. Diagram Hasil Perbandingan Sistem Penelitian Ini dengan Sistem Pencarian *E-Library*

Diagram tersebut menggambarkan perbandingan nilai presisi antara sistem rekomendasi pencarian yang dikembangkan dalam penelitian ini dengan sistem pencarian pada aplikasi *E-Library* berdasarkan kriteria uji yang berbeda. Pada kriteria "satu suku kata," sistem pencarian pada *E-Library* menunjukkan nilai presisi sebesar 9.7, sedikit lebih tinggi dibandingkan sistem yang dikembangkan dalam penelitian ini yang memiliki nilai presisi 9.6. Namun, untuk kriteria "dua suku kata," sistem yang dikembangkan memiliki nilai presisi lebih tinggi sebesar 7.3 dibandingkan dengan 6.7 pada sistem *E-Library*. Perbedaan yang signifikan terlihat pada kriteria "empat suku kata," di mana sistem yang dikembangkan menunjukkan nilai presisi sebesar 7.5, jauh lebih tinggi dari sistem *E-Library* yang hanya mencapai 2.1. Pada kriteria "judul lengkap buku," sistem yang dikembangkan mencapai nilai presisi maksimal sebesar 10.0, sedangkan sistem *E-Library* hanya mencapai nilai 2.0. Adapun rata-rata nilai presisi yang dihasilkan pada evaluasi seluruh kriteria uji dari sistem rekomendasi pencarian buku yang telah dikembangkan mencapai 8.6 atau jika dalam presentase 86%, sedangkan sistem pencarian *E-Library* memperoleh hasil rata-rata nilai presisi sebesar 5.1 atau jika dalam presentase 51%. Nilai presisi yang dihasilkan oleh sistem yang telah dikembangkan tergantung pada kata kunci yang dimasukkan dan ketersediaan buku pada dataset yang digunakan. Selain itu, panjang suku kata dalam judul buku dan banyaknya *term* (kata yang muncul) juga dapat mempengaruhi nilai similaritas yang dihasilkan oleh setiap judul buku.

Setelah melalui tahap evaluasi, dapat diketahui bahwa sistem rekomendasi pencarian buku yang telah dikembangkan pada penelitian ini dapat memberikan akurasi dan relevansi yang lebih baik sebesar 86% dari pada sistem pencarian buku pada aplikasi *E-Library* yaitu 51%. Terdapat perbedaan yang signifikan pada sistem rekomendasi pencarian buku yang telah dibuat dengan sistem pencarian buku pada *E-Library*. Dapat ditunjukkan bahwa *E-Library* mungkin memiliki keterbatasan dalam menangani kata kunci yang lebih kompleks atau spesifik. Kemudian hasil pencarian buku yang ditampilkan pada sistem rekomendasi pencarian buku yang telah dibuat, terdapat kecenderungan untuk memberikan rekomendasi meskipun relevansinya rendah dan jika kata kunci yang dimasukkan tidak sesuai dengan data buku yang ada. Di sisi lain, *E-Library* terlihat lebih baik dalam menangani kata kunci yang lebih umum tetapi tidak optimal pada kata kunci yang lebih spesifik dan menunjukkan bahwa banyak buku yang relevan tidak ditampilkan.

4. Kesimpulan dan Saran

Kesimpulan yang dapat diambil dari penelitian pengembangan sistem rekomendasi pencarian buku menggunakan metode *Content-Based Filtering* dengan algoritma TF-RF menunjukkan keberhasilan

implementasi metode ini pada data dari *E-Library* Perpustakaan Politeknik Negeri Banyuwangi dengan dataset sebanyak 6148 judul buku. Tahap Evaluasi menunjukkan bahwa sistem ini memiliki nilai presisi sebesar 86%, yang lebih baik dibandingkan nilai presisi sistem pencarian *E-Library* yang hanya sebesar 51%. Sistem ini menunjukkan bahwa nilai presisi bervariasi tergantung pada panjang kata kunci yang digunakan, dengan kata kunci yang lebih panjang atau judul lengkap buku cenderung memberikan hasil yang lebih akurat. Namun, sistem rekomendasi ini juga memiliki beberapa keterbatasan, terutama dalam menangani kata kunci dengan lebih dari satu suku kata tetapi bukan judul lengkap buku, yang seringkali menghasilkan rekomendasi dengan relevansi rendah. Selain itu, tahap *preprocessing*, khususnya proses *stemming*, mempengaruhi nilai presisi karena dapat merekomendasikan judul buku yang memiliki kata dasar yang sama dengan kata kunci meskipun kata tersebut tidak relevan. Secara keseluruhan, metode *Content-Based Filtering* menggunakan algoritma TF-RF memberikan hasil yang lebih baik dalam hal presisi dibandingkan sistem pencarian *E-Library*, meskipun masih ada ruang untuk perbaikan dalam menangani variasi kata kunci. Pada penelitian selanjutnya, dapat mempertimbangkan kembali dalam penggunaan teknik *stemming* pada saat *preprocessing* untuk menghindari perubahan ke dalam kata dasar dari setiap judul buku. Kemudian untuk meningkatkan cakupan rekomendasi, sistem dapat diperluas agar mampu memberikan rekomendasi buku dalam berbagai bahasa seperti bahasa Inggris, atau tidak hanya buku yang berbahasa Indonesia.

Daftar Pustaka:

- Al Rivan, M. E. (2017). Perbandingan Kecepatan Gabungan Algoritma Quick Sort dan Merge Sort dengan Insertion Sort, Bubble Sort dan Selection Sort. *Jurnal Teknik Informatika Dan Sistem Informasi*, 3(2). <https://doi.org/10.28932/jutisi.v3i2.629>
- Alkaff, M., Khatimi, K., & Eriady, A. (2020). Sistem Rekomendasi Buku Menggunakan Weighted Tree Similarity dan Content Based Filtering. *Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20.
- Amna, S., W., Sudipa, I. G. I., Putra, T. andi E., Wahidin, A. J., Syukrilla, W. A., Wardhani, A. K., Heryana, N., Indriyani, T., & Santoso, L. W. (2023). *Data Mining* (D. S. Kom., M. Kom. Ediana & A. M. Pd. Yanto, Eds.). PT Global Eksklusif Teknologi.
- Apriani, A., Zakiyudin, H., & Marzuki, K. (2021). Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta. *Jurnal Bumigora Information Technology (BITE)*, 3(1), 19–27. <https://doi.org/10.30812/bite.v3i1.1110>
- Dananjaya, I. K. W., & Indradewi, I. G. A. A. D. (2023). Perbandingan Metode Pembobotan TF-RF Dan TF-ABS Pada Kategorisasi Berita Di BDI Denpasar. *SINTECH (Science and Information Technology) Journal*, 6(1), 16–25. <https://doi.org/10.31598/sintechjournal.v6i1.1252>
- Joergensen Munthe, C. E., Astuti Hasibuan, N., & Hutabarat, H. (2022). RESOLUSI: Rekayasa Teknik Informatika dan Informasi Penerapan Algoritma Text Mining Dan TF-RF Dalam Menentukan Promo Produk Pada Marketplace. *Media Online*, 2(3), 110–115. <https://djournals.com/resolusi>
- Khairunnisa, S., Adiwijaya, A., & Faraby, S. Al. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *Jurnal Media Informatika Budidarma*, 5(2), 406. <https://doi.org/10.30865/mib.v5i2.2835>
- Khurana, D., Koli, A., Khatker, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, 82(3), 3713–3744. <https://doi.org/10.1007/s11042-022-13428-4>
- Kristanto, S. P., Hakim, L., Yusuf, D., Haq, E. S., & Asyari, A. R. (2022). Classification of Public Opinion on Vaccine Administration Using Convolutional Neural Network. *2022 Fifth International Conference on Vocational Education and Electrical Engineering (ICVEE)*, 65–71. <https://doi.org/10.1109/ICVEE57061.2022.9930412>
- Laksono, E., Basuki, A., & Bachtiar, F. (2020). Optimization of K Value in KNN Algorithm for Spam and Ham Email Classification. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(2), 377–383. <https://doi.org/10.29207/resti.v4i2.1845>
- Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., & Rutledge, D. N. (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC Trends in Analytical Chemistry*, 132, 116045. <https://doi.org/10.1016/j.trac.2020.116045>
- Mondi, R. H., & Wijayanto, A. (n.d.). *Recommendation System With Content-Based Filtering Method For Culinary Tourism In Mangan Application*.
- Mukhtar, H., Al Amien, J., & Rucyat, M. A. (2022). Filtering Spam Email menggunakan Algoritma Naïve Bayes. *Jurnal CoSciTech (Computer Science and Information Technology)*, 3(1), 9–19. <https://doi.org/10.37859/coscitech.v3i1.3652>

- Murti, H., Lestariningsih, E., & Sugiyamta. (2019). Perancangan Sistem Rekomendasi Buku Pada Katalog Perpustakaan Menggunakan Pendekatan Content-Based Filtering Dan Algoritma Fp-Growth. *Proceeding SINTAK 2019*.
- Pramesti, D. A. P. D., & Santiyasa, I. W. (2022). Penerapan Metode Content-Based Filtering dalam Sistem Rekomendasi Video Game. *Jurnal Nasioanal Teknologi Informasi Dan Aplikasinya, 1*.
- Putri, M. W., Muchayan, A., & Kamisutara, M. (2020). Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode Content-Based Filtering dan TF-IDF. *JOINTECS (Journal of Information Technology and Computer Science), 5(3), 229*.
<https://doi.org/10.31328/jointecs.v5i3.1563>
- Sanjaya, A., Bagus Setiawan, A., Mahdiyah, U., Nur Farida, I., Risky Prasetyo, A., & Nusantara PGRI Kediri, U. (2023). Pengukuran Kemiripan Makna Menggunakan Cosine Similarity Dan Basis Data Sinonim Kata Measurement Of Meaning Similarity Using Cosine Similarity And Word Synonyms Database. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK), 10(4)*.
<https://doi.org/10.25126/jtiik.2023106864>
- Sari, Y., Baskara, A. R., Prakoso, P. B., & Royani, N. (2022). Perbandingan Metode Pembobotan TF-RF Dan TF-IDF Dengan Dikombinasikan Dengan Weighted Tree Similarity Untuk Sistem Rekomendasi Buku. *Jurnal Teknologi Informasi Dan Ilmu Komputer, 9*.
<https://doi.org/10.25126/jtiik.202295709>

0	0	0	0,335792	0	0	0
0	0	0	0,335792	0	0	0
0	0	0	0,335792	0	0	0
0	0	0	0,335792	0	0	0
0	0	0	0	0,335792	0	0
0	0	0	0	0,335792	0	0
0	0	0	0	0,335792	0	0
0	0	0	0	0,335792	0	0
0	0	0	0	0	0,335792	0
0	0	0	0	0	0,335792	0
0	0	0	0	0	0,335792	0
