Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal. 391-401

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.30540 Link : https://dx.doi.org/10.29408/jit.v8i2.30540

# Sistem Rekomendasi Peminjaman Buku Menggunakan Metode Vector Space Model Berbasis Pembobotan TF-IDF dan FastText (Studi Kasus Perpustakaan SMP Negeri 1 Kartasura)

**Rendi Buana Perdana[1*], Dwi Hartanti[2], Herliyani Hasanah[3]**
[1,2,3] Program Studi Teknik Informatika, Universitas Duta Bangsa Surakarta
*210103115@mhs.udb.ac.id

**Abstrak**

Perpustakaan berperan penting dalam mendukung proses belajar di sekolah. Namun, dalam pemanfaatannya siswa seringkali kesulitan dalam mencari buku yang sesuai dengan kebutuhan dan minat mereka akibat dari pengelolaan koleksi buku yang belum optimal. Permasalahan ini juga terjadi di SMP Negeri 1 Kartasura, di mana belum tersedia sistem yang dapat menghasilkan rekomendasi buku secara otomatis dan relevan. Tujuan dari penelitian ini yaitu untuk membangun sistem rekomendasi buku berbasis Content-Based Filtering dengan metode Vector Space Model yang dikembangkan lebih lanjut dengan mengkombinasikan bobot TF-IDF dan representasi vektor kata dari FastText, sehingga menghasilkan teknik TF-IDF Weighted with FastText. TF-IDF digunakan untuk menentukan bobot pentingnya suatu kata dalam deskripsi buku, sementara FastText digunakan untuk menangkap makna semantik kata secara lebih mendalam. Sistem diimplementasikan menggunakan data koleksi buku yang ada di Perpustakaan SMP Negeri 1 Kartasura, dengan input berupa kata kunci dari pengguna. Representasi vektor dokumen hasil kombinasi dari TF-IDF dan FastText kemudian dihitung kemiripannya menggunakan cosine similarity untuk menentukan buku yang paling relevan. Evaluasi kemudian dilakukan menggunakan metrik Precision@10 terhadap sepuluh kata kunci berbeda, dan sistem menghasilkan rata-rata Precision@10 sebesar 0.93, menunjukkan bahwa 93% rekomendasi yang diberikan sistem ini relevan dengan kata kunci pengguna.

Kata kunci : Content-Based Filtering, Cosine Similarity, FastText, Sistem Rekomendasi, TF-IDF, Vector Space Model

***Abstract***

*Libraries play an important role in supporting the learning process at school. However, in its utilization, students often find it difficult to find books that suit their needs and interests due to the non-optimal management of the book collection. This problem also occurs at SMP Negeri 1 Kartasura, where there is no system that can generate book recommendations automatically and relevantly. The purpose of this research is to build a Content-Based Filtering-based book recommendation system with the Vector Space Model method which is further developed by combining TF-IDF weights and word vector representation from FastText, resulting in the TF-IDF Weighted with FastText technique. TF-IDF is used to determine the importance of a word in a book description, while FastText is used to capture the semantic meaning of the word in more depth. The system is implemented using book collection data in SMP Negeri 1 Kartasura Library, with input in the form of keywords from users. The document vector representation resulting from the combination of TF-IDF and FastText is then calculated using cosine similarity to determine the most relevant book. Evaluation was then conducted using the Precision@10 metric against ten different keywords, and the system produced an average Precision@10 of 0.93, indicating that 93% of the recommendations provided by this system are relevant to the user's keywords.*

Keywords : *Content-Based Filtering, Cosine Similarity, FastText, Recommendation System, TF-IDF, Vector Space Model.*

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal. 391-401

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.30540 Link : https://dx.doi.org/10.29408/jit.v8i2.30540

## 1. Pendahuluan

Perpustakaan merupakan salah satu fasilitas penting yang diberikan sekolah untuk mendukung proses belajar di sekolah. Perpustakaan merupakan salah satu sumber informasi dan pengetahuan yang berperan dalam menunjang kegiatan belajar mengajar antara guru dan siswa di sekolah[1][2]. Dengan hadirnya perpustakaan yang dikelola dengan baik, maka dapat meningkatkan minat baca siswa, membantu dalam menambah wawasan dan mengembangkan kemampuan literasi siswa. Namun, seiring dengan begitu pesatnya perkembangan teknologi, perpustakaan perlu berinovasi supaya tetap relevan dan menarik perhatian para siswa.

Permasalahan utama yang seringkali dihadapi perpustakaan yaitu pengelolaan koleksi buku yang belum optimal, khususnya dalam memberikan pengalaman yang lebih personal bagi pengguna di perpustakaan. Siswa kerap menghadapi kesulitan dalam memilih buku yang sesuai dengan minat dan kebutuhannya, akibat dari keterbatasan informasi serta kurangnya panduan dalam menentukan pilihan buku yang tersedia. Hal ini menyebabkan kurang optimalnya pemanfaatan dari koleksi buku yang tersedia di perpustakaan, sehingga banyak buku yang jarang dipinjam, maupun tidak sesuai dengan preferensi siswa.

SMP Negeri 1 Kartasura sebagai salah satu institusi pendidikan juga menghadapi tantangan yang serupa. Meskipun memiliki koleksi buku yang cukup beragam, perpustakaan pada sekolah ini belum memiliki sistem yang secara efektif dapat membantu siswa dalam menemukan buku yang relevan berdasarkan minat atau kebiasaan baca mereka. Akibatnya, masih banyak siswa yang bingung dalam memilih buku yang ingin dipinjam, yang pada akhirnya menurunkan minat siswa dalam memanfaatkan fasilitas perpustakaan.

Untuk menyelesaikan masalah ini, diperlukan suatu sistem yang dapat memberikan rekomendasi buku secara otomatis kepada siswa. Sistem rekomendasi adalah suatu teknik yang memiliki tujuan untuk memberikan masukan atau saran kepada pengguna mengenai suatu *item* yang sebaiknya dipilih[3]. Metode *Content Based Filtering* dengan pendekatan *Vector Space Model* (VSM) ini dipilih karena mampu memanfaatkan data deskriptif dari koleksi buku yang sudah ada, seperti judul, kategori, sinopsis, maupun topik tertentu, tanpa memerlukan data historis yang kompleks maupun data pengguna lain. Pendekatan ini cocok digunakan ketika pengguna baru terdaftar di sistem, maupun jika tidak tersedianya data dari pengguna lain[4].

Dalam penelitian ini, teknik *Vector Space Model* dikembangkan lebih lanjut dengan mengkombinasikan bobot *Term Frequency-*

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal. 391-401

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.30540 Link : https://dx.doi.org/10.29408/jit.v8i2.30540

*Inverse Document Frequency* (TF-IDF) dan representasi kata dari *FastText*, sehingga menghasilkan teknik TF-IDF *Weighted with FastText*. Metode ini tidak hanya mempertimbangkan pentingnya suatu kata dalam dokumen (TF-IDF), tetapi juga memperkaya representasi semantik kata melalui vektor *FastText* yang mampu menangkap makna dan hubungan antar kata secara lebih mendalam. Dengan demikian, sistem rekomendasi yang dihasilkan diharapkan mampu memberikan saran atau rekomendasi buku yang relevan dan lebih akurat berdasarkan preferensi siswa, sekaligus mendukung pengelolaan perpustakaan yang lebih efektif serta meningkatkan minat baca di lingkungan SMP Negeri 1 Kartasura.

## 2. Tinjauan Pustaka

### 2.1. Penelitian Terkait

Penelitian ini mengacu pada sejumlah riset terdahulu yang memiliki keterkaitan baik dari segi metode maupun objek yang digunakan. Berikut merupakan 5 penelitian yang terkait dengan penelitian yang dilakukan oleh peneliti:

- Penelitian yang dilakukan oleh Hafizh Dharmawan, Tukino, Shofa Shofiah Hilabi, dan Ismi Kaniawulan (2023) yang berjudul "Sistem Rekomendasi Buku Dengan Metode *K-Nearest Neighbor* (K-NN) Pada Gramedia"[5]. Penelitian ini mengembangkan sistem rekomendasi buku pada platform Gramedia memakai algoritma *K-Nearest Neighbor* (K-NN) dan *ball tree* untuk menganalisis kemiripan antar buku berdasarkan preferensi pengguna. Sistem ini diuji menggunakan data koleksi buku Gramedia yang didapat dari *website kaggle* sebanyak 2194 data rating buku, dan berhasil mencapai tingkat presisi rekomendasi sebesar 80%, yang memperlihatkan efektivitas sistem dalam menyajikan buku yang cocok berdasarkan minat pengguna.
- Penelitian yang dilakukan oleh Syaifudin Rosyad, Danang Mahendra, dan Noor Azizah (2023) yang berjudul "Sistem Rekomendasi Buku Di Perpustakaan Daerah Jepara Menggunakan Metode *Item-Based Collaborative Filtering*"[6]. Penelitian ini memanfaatkan data rating buku yang diberikan oleh 100 anggota perpustakaan selama periode April hingga Mei untuk menghasilkan rekomendasi menggunakan metode *Item-Based Collaborative Filtering*. Hasil pengujian menunjukkan bahwa nilai *Mean Absolute Error* (MAE) sebesar 1.2912227520737, menandakan bahwa tingkat akurasi cukup baik untuk sistem rekomendasi berbasis rating pengguna.
- Penelitian yang dilakukan oleh Rasyida Hanun Dhiya Reswara, Evanita, dan Arief Susanto (2025) yang berjudul "Implementasi *Content-Based Filtering* pada Sistem

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal. 391-401

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.30540 Link : https://dx.doi.org/10.29408/jit.v8i2.30540

Rekomendasi Buku Perpustakaan"[7]. Penelitian ini mengembangkan sistem rekomendasi buku untuk Perpustakaan Universitas Muria Kudus menggunakan metode *Content-Based Filtering*. Hasil evaluasi menggunakan metrik *Precision@10* menghasilkan rata-rata nilai sebesar 0,91, yang menunjukkan bahwa sistem mampu memberikan rekomendasi yang sangat relevan terhadap preferensi pengguna.

- Penelitian yang dilakukan oleh Nova Noor Kamala Sari, Ressa Priskila, dan Putu Bagus Adidyana Anugrah Putra (2024) yang berjudul "Implementasi *Content-Based Filtering* Menggunakan TF-IDF *and Cosine Similarity* Untuk Sistem Rekomendasi Resep Masakan"[8]. Penelitian ini mengembangkan sistem rekomendasi resep masakan berdasarkan kesesuaian bahan menggunakan TF-IDF dan *cosine similarity*. Hasil evaluasi menunjukkan sistem mampu memberikan rekomendasi akurat dengan nilai RMSE sebesar 0,356.
- Penelitian yang dilakukan oleh Zahir Abbas Khan dan Dr. Rekha V (2023) yang berjudul "*Fake News Detection Using TF-IDF Weighted with Word2Vec: An Ensemble Approch*"[9]. Penelitian ini menerapkan pendekatan yang serupa dengan mengkombinasikan bobot TF-IDF dengan *Word2Vec* untuk mendeteksi berita palsu pada *dataset FakeNewsNet*. Representasi vektor dokumen dihasilkan dari penggabungan bobot TF-IDF dan *Word2Vec*, lalu diklasifikasikan menggunakan Random Forest dan Decision Tree dengan majority voting. Model ini mencapai presisi 94,24%, menunjukkan efektivitas pendekatan dalam tugas klasifikasi dan potensinya untuk analisis teks lainnya, termasuk sistem rekomendasi.

## 2.2. Landasan Teori

### 1. Sistem Rekomendasi

Sistem rekomendasi merupakan sebuah sistem yang berfungsi dalam menyaring dan mengenali *item*, seperti produk, barang, informasi, maupun layanan, yang kemungkinan besar akan diminati, dibeli, maupun dipakai oleh pengguna[10]. Sistem rekomendasi dirancang untuk merekomendasikan suatu *item* yang paling cocok berdasarkan preferensi, minat dan keinginan pengguna, sekaligus mempermudah pengguna dalam mengambil suatu keputusan[11].

### 2. Content Based Filtering

*Content Based Filtering* merupakan salah satu metode pada sistem rekomendasi yang menghasilkan rekomendasi dengan mengukur kemiripan berdasarkan suatu atribut atau karakteristik dari *item* tersebut kemudian dibandingkan dengan *item* lain[12][13].

Metode sistem rekomendasi ini dipilih karena memiliki sifat *user independence*, di mana hasil

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal. 391-401

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.30540 Link : https://dx.doi.org/10.29408/jit.v8i2.30540

rekomendasi tidak bergantung pada data historis yang kompleks dan data pengguna lain[4][7].

3. *Vector Space Model*

*Vector Space Model* (VSM) merupakan salah satu pendekatan pendekatan matematis yang digunakan untuk merepresentasikan teks sebagai vektor dalam suatu ruang multidimensi, di mana setiap dimensi berkorelasi dengan istilah atau kata unik dalam korpus. VSM diterapkan untuk mengukur kesamaan antar buku menggunakan fitur *metadata* seperti judul, kategori, dan deskripsi. Bobot istilah dalam vektor dihitung menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk menilai seberapa pentingnya suatu kata dalam dokumen atau *query*. Model ini mengukur kesamaan buku dengan menggunakan *cosine similarity* untuk menemukan tingkat similaritas atau kesamaan antara vektor dokumen dengan *query*. Keunggulan dari VSM terletak pada kemudahan untuk diimplementasi, fleksibel, dan kemampuan menangani data tekstual secara intuitif[14][15].

4. *FastText*

*FastText* merupakan suatu *library* untuk mempelajari tentang *word embedding* dan klasifikasi teks yang dikembangkan oleh *Facebook's Artificial Intelligence Research* (FAIR). *FastText* adalah salah satu metode *word embedding* yang merupakan pengembangan dari metode *Word2Vec* dimana representasi vektor didapatkan melalui karakter n-gram yang muncul dalam kata[16]. *FastText* merupakan model yang mampu memahami makna kata dalam teks dengan memperhitungkan bagian-bagian kecil dari kata, seperti sub-kata atau n-gram. Pendekatan ini menjadikannya lebih efektif dalam menangani kata-kata yang belum dikenal maupun kata yang salah tulis. Dengan menerapkan *FastText*, setiap kata dalam teks dikonversi menjadi sebuah vektor yang mencerminkan makna kata tersebut dalam ruang vektor[17].

5. *Precision@N*

*Precision@N* merupakan salah satu metrik evaluasi yang digunakan pada sistem temu kembali informasi dan sistem rekomendasi untuk menilai seberapa besar proporsi *item* yang relevan di antara N *item* teratas yang diambil atau direkomendasikan oleh sistem[7].

Secara spesifik, *Precision@N* mengukur persentase *item* relevan yang terdapat pada N hasil teratas yang diberikan oleh sistem. Metrik ini sering dimanfaatkan untuk mengevaluasi performa suatu sistem[7]. Adapun rumus perhitungan *Precision@N* adalah sebagai berikut:

$$Precission@N = \frac{r}{n} \quad (6)$$

Keterangan:

$r$ : Jumlah *item* yang relevan dalam N rekomendasi teratas

$n$ : Jumlah total *item* yang direkomendasikan

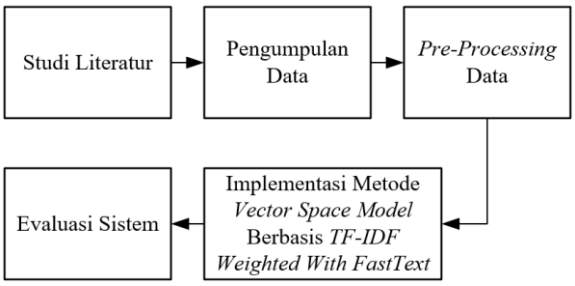
Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal. 391-401

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.30540 Link : https://dx.doi.org/10.29408/jit.v8i2.30540

## 3. Metode Penelitian

Pada penelitian ini terdapat beberapa proses yang harus dilalui untuk mengimplementasikan sistem rekomendasi buku dengan metode *Vector Space Model* yang dikembangkan lebih lanjut dengan mengkombinasikan bobot *Term Frequency-Inverse Document Frequency* (TF-IDF) dan representasi kata dari *FastText*. Penelitian ini menggunakan data koleksi buku yang ada di Perpustakaan SMP Negeri 1 Kartasura. Tahapan utama yang harus dilalui dalam penelitian ini dapat dilihat pada Gambar 1 berikut ini :




Gambar 1 Tahapan Penelitian

a) Studi Literatur

Studi literatur pada penelitian ini bertujuan untuk mengkaji penerapan metode *Content-Based Filtering* dengan pendekatan *Vector Space Model* berbasis TF-IDF yang dikombinasikan dengan representasi vektor kata dari *FastText* untuk mendapatkan rekomendasi buku. Studi literatur dilakukan dengan menelaah sumber yang relevan seperti jurnal ilmiah, buku, artikel, serta informasi dari internet guna memahami efektivitas penerapan metode ini dalam sistem rekomendasi. Salah satu penelitian terdahulu menunjukkan bahwa kombinasi bobot TF-IDF dengan teknik *Word Embedding*, seperti *Word2Vec* dapat secara efektif untuk menghasilkan representasi vektor dokumen yang tidak hanya memperhatikan pentingnya kata dalam korpus, tetapi juga makna semantik dari setiap kata. Teknik ini terbukti mampu meningkatkan kualitas representasi teks yang digunakan dalam tugas klasifikasi, dan berpotensi untuk diterapkan pada sistem rekomendasi berbasis konten. Berdasarkan temuan tersebut, penelitian ini juga menerapkan konsep *Word Embedding* yang serupa namun menggunakan *library* berbeda yaitu *FastText*. Sehingga harapannya, penelitian ini mampu mengembangkan sistem rekomendasi buku yang dapat memberikan saran yang lebih terpersonalisasi kepada siswa berdasarkan kata kunci yang mereka masukkan sebagai *query*.

b) Pengumpulan Data

Data yang dipakai dalam penelitian ini diperoleh dari Perpustakaan SMP Negeri 1 Kartasura, dimana data tersebut didapat dari *file excel* yang telah dimiliki oleh pihak perpustakaan pada saat pendataan buku periode tahun 2024. Data ini berisi informasi mengenai koleksi buku yang tersedia seperti judul, pengarang, penerbit, tahun terbit, ISBN, kategori dan deskripsi. Namun, pada penelitian ini data yang digunakan hanya judul, pengarang, dan deskripsi. Hal ini dikarenakan ketiga atribut ini dianggap paling

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal. 391-401

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.30540 Link : https://dx.doi.org/10.29408/jit.v8i2.30540

merepresentasikan konten atau topik dari setiap buku dibandingkan dengan atribut lain.

Dengan menggunakan data yang sudah terdokumentasi dalam format *excel* ini, proses pengumpulan data dapat dilakukan secara efisien tanpa perlu melakukan *input* data manual, sekaligus memastikan bahwa informasi yang digunakan mencerminkan kondisi aktual dari koleksi buku yang dimiliki perpustakaan. Contoh sampel data yang dipakai dalam penelitian ini ditunjukkan dalam Tabel 1 berikut ini.

Tabel 1. Contoh Sampel Data

| Judul | Kategori | Deskripsi |
|---|---|---|
| Bermain dengan Matematika | Matematika | Buku ini menyajikan materi mengenai bilangan, pengukuran, geometri dan pengolahan data. |
| Tercepat! Supercar | Teknologi | Buku ini berisi informasi tentang mobil mulai yang paling lambat hingga yang paling cepat. |
| Pengetahuan Tentang Satwa | Binatang | Buku ini menjelaskan kenekaragaman satwa dan habitatnya. |

Pada Tabel 1 ditunjukkan contoh sampel data yang terdiri dari judul, kategori, dan deskripsi buku.

c) Pre-Processing Data

Sebelum data dapat digunakan untuk implementasi sistem rekomendasi, perlu dilakukan *pre-processing* data untuk membersihkan dan mempersiapkan data agar siap untuk dianalisis dan digunakan untuk implementasi sistem rekomendasi. Tahapan yang dilakukan pada *pre-processing* data terdiri dari *Cleaning*, *Case Folding*, *Tokenizing*, *Stopword Removal*, dan *Stemming*.

d) Implementasi Metode *Vector Space Model* Berbasis TF-IDF *Weighted With FastText*

Dalam penelitian ini metode *Content-Based Filtering* diimplementasikan dengan pendekatan *Vector Space Model* yang dikembangkan lebih lanjut dengan mengkombinasikan bobot *Term Frequency-Inverse Document Frequency* (TF-IDF) dengan representasi vektor kata dari *FastText*, sehingga menghasilkan teknik TF-IDF *Weighted with FastText*. Metode ini tidak hanya mempertimbangkan pentingnya suatu kata dalam dokumen (TF-IDF), tetapi juga memperkaya representasi semantik kata melalui vektor *FastText* yang mampu menangkap makna dan hubungan antar kata secara lebih mendalam.

Penggunaan algoritma ini digunakan untuk mengubah data teks pada *dataset* koleksi buku perpustakaan SMP Negeri 1 Kartasura menjadi representasi vektor numerik yang dapat dihitung tingkat kesamaannya menggunakan *cosine similarity*.

e) Evaluasi Sistem

Tahapan evaluasi ini dilakukan untuk mengetahui kinerja dan efektivitas sistem rekomendasi yang sudah dikembangkan. Evaluasi memiliki tujuan untuk mengukur seberapa baiknya sistem dalam

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal. 391-401

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.30540 Link : https://dx.doi.org/10.29408/jit.v8i2.30540

memberikan rekomendasi yang relevan sesuai dengan kebutuhan pengguna. Pada penelitian ini, proses evaluasi dilakukan dengan menggunakan metrik *Precision@N*, yang menghitung persentase *item* yang relevan di antara N *item* yang direkomendasikan oleh sistem[7]..

## 4. Hasil dan Pembahasan

### 4.1. Hasil Penelitian

Data buku yang diperoleh dari *file excel* yang dimiliki oleh pihak Perpustakaan SMP Negeri 1 Kartasura adalah sebanyak 2.079 buku dengan atribut berupa judul, pengarang, penerbit, tahun_terbit, ISBN, kategori dan deskripsi buku. Dari keseluruhan atribut, dipilih tiga atribut utama yang dinilai paling mewakili isi atau topik buku, atribut tersebut terdiri dari kolom judul, kategori, dan deskripsi. Ketiga atribut ini kemudian diproses dan digabungkan menjadi satu kolom baru yang disebut *book_keywords*, yang berfungsi sebagai representasi awal isi buku dalam bentuk teks.

Penggabungan kolom ini bertujuan untuk menyatukan informasi penting yang dapat merepresentasikan isi dan karakteristik dari masing-masing buku secara lebih utuh dalam satu kesatuan teks. Hasil penggabungan ini menjadi dasar dalam proses ekstraksi fitur pada tahap selanjutnya. Kolom *book_keywords* digunakan untuk membentuk representasi vektor dari setiap buku, yang nantinya dipakai dalam proses perhitungan kemiripan menggunakan metode TF-IDF dan *FastText*.

### 4.2. Pre-Processing Data

Sebelum dilakukan ekstraksi fitur, data teks pada kolom judul, kategori, dan deskripsi diproses melalui tahapan *pre-processing* untuk menghasilkan teks yang bersih dan seragam. Proses ini melibatkan penghilangan karakter tidak relevan (*cleaning)*, normalisasi huruf (*case folding)*, penghapusan kata-kata umum (*stopword removal*), dan penyederhanaan kata ke bentuk dasarnya (*stemming*) menggunakan *library Sastrawi* [18].

Hasil dari tahap ini adalah kolom baru bernama *book_keywords*, yang berisi representasi teks bersih dari setiap buku. Contoh hasil data buku sebelum dan sesudah dilakukan *pre-processing* disajikan pada Tabel 2.

Tabel 2. Contoh Hasil *Pre-Processing*

| Sebelum *Pre-Processing* | Setelah *Pre-Processing* |
|---|---|
| Bermain dengan Matematika, Matematika, Buku ini menyajikan materi mengenai bilangan, pengukuran, geometri dan pengolahan data. | main matematika matematika buku saji materi kena bilang ukur geometri olah data |
| Tercepat! Supercar, Teknologi, Buku ini berisi informasi tentang mobil mulai yang paling lambat hingga yang paling cepat. | cepat supercar teknologi buku isi informasi mobil mulai paling lambat hingga paling cepat |

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal. 391-401

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.30540 Link : https://dx.doi.org/10.29408/jit.v8i2.30540

| | |
|---|---|
| Pengetahuan Tentang Satwa, Binatang, Buku ini menjelaskan kenekaragaman satwa dan habitatnya. | tahu satwa binatang buku jelas kenekaragaman satwa habitat |

Hasil *pre-processing* ini digunakan sebagai *input* utama dalam pembentukan vektor teks pada tahap selanjutnya. Dengan teks yang telah dibersihkan dan disederhanakan, proses representasi semantik menjadi lebih optimal, serta meningkatkan akurasi dalam sistem rekomendasi.

### 4.3. Implementasi Metode Vector Space Model Berbasis TF-IDF *Weighted With FastText*

Setelah proses *pre-processing* menghasilkan data teks bersih dalam kolom book_keywords, tahap selanjutnya adalah merepresentasikan setiap buku ke dalam bentuk vektor numerik. Representasi ini dibangun menggunakan kombinasi metode TF-IDF dan *FastText*.

TF-IDF berperan dalam memberikan bobot terhadap setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dan dalam seluruh korpus. Sementara itu, *FastText* digunakan untuk menghasilkan vektor kata yang dapat menangkap konteks semantik secara lebih mendalam, karena mampu memperhitungkan karakter sub-kata (*subword information*) yang umum dalam bahasa alami.

Model *FastText* yang digunakan dalam penelitian ini merupakan *pre-trained* model berbahasa Indonesia yang telah tersedia secara publik. Penggunaan model ini mempercepat proses implementasi dan tetap memungkinkan sistem untuk menangkap makna semantik dari kata-kata dalam koleksi buku, tanpa perlu melakukan pelatihan ulang[17].

Dengan menggabungkan kedua pendekatan tersebut, setiap buku direpresentasikan sebagai sebuah vektor berdimensi 300, yang menggambarkan makna semantik dari konten buku secara keseluruhan. Seluruh vektor buku kemudian disusun dalam sebuah matriks representasi (*book_vectors_matrix*) dan dinormalisasi untuk memudahkan dalam proses perhitungan kemiripan.

Proses pembentukan vektor dari setiap buku ini dilakukan dengan cara menghitung bobot TF-IDF dari setiap kata dan mengalikannya dengan vektor *FastText* yang sesuai. Kode untuk proses ini ditunjukkan pada Gambar 2. Hasil akhirnya adalah representasi vektor dokumen yang memperhitungkan baik pentingnya kata dalam konteks dokumen maupun makna semantiknya.

```
def document_vector_fasttext(text):
    words = text.split()
    vec = np.zeros(300)
    weight_sum = 0
    for word in words:
        if word in fasttext.key_to_index and word in tfidf_vocab:
            tfidf_weight = tfidf_vocab[word] * (words.count(word) / len(words))
            vec += fasttext[word] * tfidf_weight
            weight_sum += tfidf_weight
    if weight_sum == 0:
        return vec
    return vec / weight_sum
# Hitung Vektor Seluruh Buku
book_vectors = buku_df['book_keywords'].apply(document_vector_fasttext)
book_vectors_matrix = np.vstack(book_vectors.values)
book_vectors_matrix = normalize(book_vectors_matrix)
```

Gambar 2. Kode Untuk Menghitung Representasi Vektor Dari Setiap Buku

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal. 391-401

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.30540 Link : https://dx.doi.org/10.29408/jit.v8i2.30540

Tahap akhir dari proses sistem rekomendasi ini adalah dengan melakukan pencocokan antara kata kunci dari *input query* pengguna dengan seluruh koleksi buku menggunakan pendekatan *cosine similarity*. Proses perhitungan *similarity* dan pengambilan daftar rekomendasi buku dapat dilihat pada Gambar 3.

```
def recommend_with_fasttext(user_input, top=10):
    # Tahap pre-processing kata kunci pengguna hingga perhitungan vektornya
    cleaned_input = clean_text(user_input)
    input_vec = document_vector_fasttext(cleaned_input)
    input_vec = normalize([input_vec])

    # Menghitung Cosine Similarity
    similarity_scores = cosine_similarity(input_vec, book_vectors_matrix)[0]
    top_indexes = similarity_scores.argsort()[::-1][:top]

    # Format hasil
    recommended_buku = []
    for i in top_indexes:
        judul_buku = buku_df.index[i]
        skor = similarity_scores[i]
        recommended_buku.append(f"{judul_buku} - {skor:.4f}")

    return recommended_buku
```

Gambar 3 Kode Perhitungan Rekomendasi Buku

Pada Gambar 3 ditunjukkan fungsi *recommend_with_fasttext* yang menerima masukan berupa kata kunci dari pengguna, lalu mengembalikan daftar buku yang paling relevan berdasarkan tingkat kemiripan. Kata kunci tersebut juga harus melewati proses *pre-processing* hingga perhitungan vektor seperti data buku. Tahap terakhir pada kode tersebut adalah menghitung nilai *cosine similarity* antara vektor dari kata kunci pengguna dengan seluruh vektor buku di *book_vectors_matrix*. Hasil *similarity* ini kemudian disimpan dalam bentuk skor, yang kemudian diurutkan dari yang tertinggi hingga terendah. Sepuluh skor tertinggi akan dipilih untuk membentuk daftar rekomendasi buku. Adapun hasil rekomendasi dapat dilihat pada Tabel 3.

Tabel 3. Hasil Rekomendasi Buku

| No | Judul Buku | *Cosine Similarity* |
|---|---|---|
| 1 | Waktu | 0.9107 |
| 2 | Berat | 0.9102 |
| 3 | Mengenal Pengukuran 3 | 0.8735 |
| 4 | Ayo, Mengukur Tinggi | 0.8650 |
| 5 | Isi | 0.8530 |
| 6 | Kecepatan | 0.8453 |
| 7 | Ayo, Mengukur Kelajuan | 0.8274 |
| 8 | Ayo Mengukur Jarak | 0.8088 |
| 9 | Ayo, Mengukur Luas | 0.8083 |
| 10 | Ayo Mengukur Berat | 0.8081 |

Pada Tabel 3 menunjukkan hasil rekomendasi buku berdasarkan kata kunci “pengukuran dalam pelajaran matematika”. Disini sistem menghasilkan daftar sepuluh buku yang memiliki nilai similarity tertinggi berdasarkan kata kunci tersebut.

### 4.4. Evaluasi Dengan Precision@N

Evaluasi dilakukan dengan menggunakan 10 kata kunci berbeda yang merepresentasikan berbagai topik dalam koleksi buku di Perpustakaan SMP Negeri 1 Kartasura. Hasil rekomendasi tersebut kemudian dianalisis relevansinya berdasarkan kecocokan isi buku terhadap topik yang dimaksud. Hasil evaluasi masing-masing kata kunci dihitung menggunakan *Precision@N*, dan disajikan pada Tabel 4.

Tabel 4. Hasil Evaluasi *Precision@N*

| Kata Kunci | Jumlah Rekomendasi | Jumlah Buku Relevan | *Precision* |
|---|---|---|---|
| pengukuran dalam pelajaran matematika | 10 | 10 | 1 |

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal. 391-401

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.30540 Link : https://dx.doi.org/10.29408/jit.v8i2.30540

| | | | |
|---|---|---|---|
| kamus bahasa Indonesia | 10 | 7 | 0.7 |
| buku bahasa inggris | 10 | 8 | 0.8 |
| bagian tubuh manusia | 10 | 9 | 0.9 |
| buku menjadi anak yang sukses | 10 | 9 | 0.9 |
| kumpulan puisi | 10 | 10 | 1 |
| mengenal bumi | 10 | 10 | 1 |
| kisah orang yang percaya cinta dan mimpi | 10 | 10 | 1 |
| buku aneka kreasi | 10 | 10 | 1 |
| hujan | 10 | 10 | 1 |
| Rata – rata *Precision@N* | | | 0.93 |

Berdasarkan hasil evaluasi yang disajikan pada Tabel 4, sistem rekomendasi yang dikembangkan menunjukkan kinerja yang sangat baik dengan rata-rata *Precision@10* mencapai 93%. Hal ini menunjukkan bahwa sebagian besar buku yang direkomendasikan oleh sistem bersifat relevan terhadap kata kunci yang diberikan.

Namun demikian, masih ada kata kunci yang menunjukkan nilai *Precision@10* paling rendah, yaitu "kamus bahasa Indonesia" dengan skor 0.7. Hal ini dapat disebabkan oleh beberapa faktor, seperti keterbatasan data deskripsi pada koleksi buku yang tersedia, atau kata kunci yang terlalu umum sehingga sistem kesulitan dalam menyesuaikan konteks secara spesifik.

## 5. Kesimpulan

Penelitian ini berhasil menerapkan metode *Content-Based Filtering* dengan pendekatan *Vector Space Model* yang dikembangkan lebih lanjut dengan mengkombinasikan bobot TF-IDF dan representasi vektor *FastText*. Hasil evaluasi menggunakan *Precision@10* menunjukkan rata-rata akurasi mencapai 93%, yang menunjukkan bahwa sebagian besar hasil rekomendasi cukup relevan terhadap kata kunci pengguna. Sistem ini diharapkan mampu membantu pengguna menemukan buku yang sesuai dengan kebutuhan atau minat mereka secara lebih efisien. Untuk pengembangan selanjutnya, sistem ini dapat disempurnakan dengan mempertimbangkan informasi tambahan seperti data riwayat peminjaman, atau mengkombinasikan dengan teknik rekomendasi lain untuk dapat memberikan hasil rekomendasi yang lebih personal dan akurat.

## 6. Daftar Pustaka

[1] A. Rosita, N. Puspitasari, and V. Z. Kamila, “Rekomendasi Buku Perpustakaan Kampus Dengan Metode Item-Based Collaborative Filtering,” *Sebatik*, vol. 26, no. 1, pp. 340–346, 2022, doi: 10.46984/sebatik.v26i1.1551.

[2] H. Sulaiman, Y. Yuliani, K. Panggalih, M. I. Alifudin, and K. Widianto, “Pengelompokan Keaktifan Anggota Perpustakaan Menggunakan Algoritma K-Means,” *Infotek J. Inform. dan Teknol.*, vol. 8, no. 1, pp. 56–65, 2025, doi: https://dx.doi.org/10.29408/jit.v8i1.27978.

[3] S. Rahmadhani *et al.*, “JIP (Jurnal Informatika Polinema) sistem rekomendasi penelusuran buku berbasis content-based filtering dengan pembobotan TF-RF,” pp. 491–500, 2020.

[4] C. Koloman, R. Maulana, R. Dwi, Z. Putri,

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal. 391-401

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.30540 Link : https://dx.doi.org/10.29408/jit.v8i2.30540

and W. A. Harahap, "Sistem Rekomendasi Pekerjaan di bidang IT Menggunakan Algoritma Content-Based Filtering," *J. Creat. Student Res.*, vol. 1, no. 6, pp. 78–88, 2023, [Online]. Available: https://doi.org/10.55606/jcsrpolitama.v1i6.2992

[5] H. Dharmawan, Tukino, S. Shofiah Hilabi, and I. Karniawulan, "Sistem Rekomendasi Buku Dengan Metode K-Nearest Neighbor (K-Nn) Pada Gramedia," *Zo. J. Sist. Inf.*, vol. 5, no. 1, pp. 16–25, 2023, doi: 10.31849/zn.v5i1.12203.

[6] S. Rosyad, D. Mahendra, and N. Azizah, "Sistem Rekomendasi Buku Di Perpustakaan Daerah Jepara Menggunakan Metode Item-Based Collaborative Filtering," *Biner J. Ilm. Inform. dan Komput.*, vol. 2, no. 2, pp. 76–81, 2023, doi: 10.32699/biner.v2i2.3934.

[7] R. Rasyida Hanun Dhiya, Evanita, and A. Susanto, "Implementasi Content-Based Filtering Pada Sistem Rekomendasi Buku Perpustakaan," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 2, pp. 3243–3250, 2025,

[8] N. N. K. Sari, R. Priskila, and P. B. A. A. Putra, "Implementasi Content-Based Filtering Menggunakan Tf-Idf and Cosine Similarity," *J. Teknol. Inf.*, vol. 18, no. 1, pp. 43–51, 2024.

[9] Z. A. Khan and V. Rekha, "Fake News Detection Using TF-IDF Weighted with Word2Vec: An Ensemble Approach," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 3, pp. 1065–1076, 2023.

[10] H. H. Arfisko and A. T. Wibowo, "Sistem Rekomendasi Film Menggunakan Metode Hybrid Collaborative Filtering dan Content-Based Filtering," *e-Proceeding Eng.*, vol. 9, no. 3, pp. 2149–2159, 2022.

[11] V. Atina and D. Hartanti, "Knowledge Based Recommendation Modeling for Clothing Product Selection Recommendation System," *J. Tek. Inform.*, vol. 3, no. 5, pp. 1407–1413, 2022, doi: 10.20884/1.jutif.2022.3.5.584.

[12] D. A. N. Safitri, R. Halilintar, and L. S. Wahyuniar, "Sistem Rekomendasi Skincare Menggunakan Metode Content-Based Filtering dan Algoritma Apriori," *Semin. Nas. Inov. Teknol. (SEMNAS INOTEK)*, pp. 242–248, 2021, [Online].

[13] P. Dewa Ayu Putri Diah and I. W. Santiyasa, "Penerapan Metode Content-Based Filtering dalam Sistem Rekomendasi Video Game," *Jnatia*, vol. 1, no. 1, pp. 229–234, 2022.

[14] D. T. Santoso, V. Atina, and D. Hartanti, "Prototipe Sistem Rekomendasi Film Indonesia Menggunakan Pendekatan Content Based Filtering dan Metode Vector Space Model Infotek : Jurnal Informatika dan Teknologi Film " *Infotek J. Inform. dan Teknol.*, vol. 7, no. 2, pp. 444–445, 2024, doi: https://doi.org/10.29408/jit.v7i2.26083.

[15] R. Huang, "Improved content recommendation algorithm integrating semantic information," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00776-7.

[16] I. F. Ramadhy and Y. Sibaroni, "Analisis Trending Topik Twitter dengan Fitur Ekspansi FastText Menggunakan Metode Logistic Regression," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 1, p. 1, 2022, doi: 10.30865/jurikom.v9i1.3791.

[17] B. R. Ermawan and N. Cahyono, "Optimasi Metode Klasifikasi Menggunakan FastText dan Grid Search pada Aanalisi Sentimen Ulasan Aplikasi SeaBank," *JIKO (Jurnal Inform. dan Komputer)*, vol. 9, no. 1, pp. 226–238, 2025, doi: 10.26798/jiko.v9i1.1523.

[18] R. Ardiansyah, M. Ari Bianto, and B. D. Saputra, "Sistem Rekomendasi Buku Perpustakaan Sekolah menggunakan Metode Content-Based Filtering," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 4, no. 2, pp. 510–518, 2023, doi: 10.37859/coscitech.v4i2.5131.