

BAB I PENDAHULUAN

1.1 Latar Belakang

Industri buku terus mengalami pertumbuhan jumlah terbitan yang signifikan baik di tingkat global maupun nasional. Menurut laporan tahun 2022 oleh *World Intellectual Property Organization* (WIPO), jumlah ISBN yang didaftarkan menunjukkan dominasi Amerika Serikat dengan lebih dari 3,3 juta judul, diikuti Jepang (902 ribu), Republik Korea (338 ribu), India (281 ribu), dan Jerman (277 ribu). Sementara itu, data legal deposit mencatat sekitar 2,7 juta buku disimpan di repository nasional pada tahun yang sama. Namun jika dibandingkan dengan tahun 2021, terlihat adanya penurunan di beberapa negara besar seperti Inggris, Indonesia, dan Ukraina, sedangkan negara Jepang dan Amerika Serikat justru peningkatan yang signifikan [1].

Namun, tren tidak selalu menunjukkan peningkatan terutama pada tahun 2019 hingga 2023, di Indonesia sendiri mengalami nilai fluktuatif yang cenderung turun dari 135 ribu pada 2020 hingga 99 ribu judul pada 2023 [2]. Kondisi ini menunjukkan bahwa tantangan utama bukan lagi sekedar kuantitas, melainkan bagaimana pembaca menemukan buku yang relevan di tengah keterbatasan fluktuasi produksi.

Sebuah studi empiris besar menunjukkan bahwa metadata konvensional seringkali tidak memadai untuk kebutuhan pencarian buku yang kompleks, sehingga pencarian berbasis kata kunci rentan menghasilkan hasil yang tidak akurat [3]. Meskipun sistem rekomendasi buku bisa menjadi cara untuk menyaring dan mengarahkan bacaan yang lebih relevan, berdasarkan penelitian oleh [4], kelemahan atau *sparsity* metadata item memperburuk performa hasil rekomendasi, terutama dalam sistem yang bergantung pada representasi konten. Metadata seperti judul, genre, atau kategori sering kali tidak cukup untuk merepresentasikan makna semantik dari isi buku, terutama ketika informasi deskripsi seperti sinopsis tidak tersedia secara lengkap.

Dalam konteks permasalahan tersebut, sistem rekomendasi dikembangkan untuk menghasilkan preferensi yang dipersonalisasi dengan menggunakan teknik dan algoritma penyaringan informasi dengan tujuan mendukung proses pengambilan keputusan pengguna [5]. Salah satu pendekatannya adalah metode CBF (*Content-Based Filtering*) yang menghasilkan *output* relevan dengan menggunakan data statis spesifik walaupun tidak menggunakan data pengguna. Cara kerja CBF adalah dengan menyaring informasi dan mencari *item* yang mirip dengan item yang disukai atau diketahui pengguna. Selain itu, CBF memiliki banyak opsi implementasi seperti algoritma representasi teks seperti TF-IDF maupun algoritma *word embedding* seperti Word2Vec sehingga fleksibel diterapkan sesuai ketersediaan dataset dan lingkup rekomendasi seperti buku.

Perbedaan cara kerja dan karakteristik representasi teks antara algoritma representasi teks TF-IDF dengan metode yang lebih modern seperti Word2Vec menjadi alasan perlunya dilakukan penelitian komparatif untuk menilai efektivitas menghasilkan rekomendasi buku. TF-IDF bekerja sebagai metode berbasis frekuensi yang menghitung bobot tiap kata secara eksplisit, sehingga cocok dengan data yang punya struktur teks jelas dan tidak terlalu kompleks. Sebaliknya, algoritma Word2Vec membangun representasi semantik yang lebih dalam dengan mencari konteks antar kata, sehingga menghasilkan hubungan makna antar data. Namun, efektivitas kedua metode tersebut bergantung pada ketersediaan data, karakteristik dokumen, serta alur yang ditentukan. Oleh karena itu pengujian performa antara TF-IDF dan Word2Vec dapat berbeda secara signifikan tergantung banyak faktor terutama dalam pembuatan rekomendasi buku [6].

1.2 Rumusan Masalah

Berdasar latar belakang masalah yang telah dijelaskan sebelumnya, rumusan masalah pada penelitian ini adalah sebagai berikut :

1. Bagaimana proses penerapan metode TF-IDF dan Word2Vec dalam menghasilkan sistem rekomendasi berbasis *Content-Based Filtering* dengan menggunakan dataset yang tersedia di publik ?

2. Algoritma manakah yang mempunyai akurasi dan performa lebih baik berdasarkan hasil pengujian dan evaluasi pada sistem rekomendasi buku ?

1.3 Batasan Masalah

Dengan banyaknya keterbatasan dalam proses pengujian dan implementasi, penelitian ini membatasi beberapa hal sebagai berikut :

1. Penelitian hanya menggunakan pendekatan *Content-Based Filtering* dalam proses pengujian metode sistem rekomendasi.
2. Algoritma yang diuji terbatas pada model *Term Frequency-Inverse Document Frequency* (TF-IDF) dan Word2Vec sebagai representasi teks.
3. Dataset buku yang digunakan terbatas pada data statis berupa judul, sinopsis/deskripsi, genre, dan series sehingga tidak menggunakan data seperti rating atau ulasan.
4. Sumber dataset buku yang digunakan dibatasi dari dua sumber yaitu <https://kaggle.com> dengan data puluhan ribu buku dan dataset buku pelajaran lokal dari kemdikbudristek, sehingga hasil penelitian ini diharapkan dapat diterapkan di beberapa kasus serupa walaupun dataset yang digunakan berisikan buku dari berbagai bahasa.
5. Fokus penelitian hanya pada metode CBF, selain itu hanya akan menguji efektifitas dari 2 metode yaitu TF-IDF dan Word2Vec.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitiannya adalah menghasilkan :

1. Mengimplementasikan sistem rekomendasi buku dengan pendekatan *Content Based Filtering* menggunakan dua model representasi teks, yaitu TF-IDF dan Word2Vec.
2. Membandingkan nilai efektivitas dari dua pendekatan TF-IDF dan Word2Vec berbasis *Cosine Similarity* dalam mengukur kemiripan antara data buku untuk mengetahui metode paling efektif.

3. Menguji pengaruh pemilihan metode representasi teks terhadap kualitas rekomendasi, sehingga dapat mengetahui model mana yang memberikan rekomendasi yang lebih baik pada dataset buku yang digunakan.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut :

1. Penelitian ini diharapkan dapat digunakan untuk penelitian dengan studi kasus pada perpustakaan yang baru saja dibuka dan ingin membuat sebuah sistem rekomendasi yang efektif dengan data yang terbatas.
2. Menjadi referensi bagi penelitian selanjutnya terkait efektivitas representasi teks dalam sistem rekomendasi yang mungkin punya konteks masalah berbeda.

1.6 Sistematika Penulisan

Sistematika penulisan naskah tugas akhir pada skripsi adalah sebagai berikut :

BAB I PENDAHULUAN, berisi Latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, dan manfaat penelitian.

BAB II TINJAUAN PUSTAKA, berisi tinjauan pustaka, dan juga dasar teori yang diterapkan seperti TF-IDF, Word2Vec, Sistem Rekomendasi, Python.

BAB III METODE PENELITIAN, didalamnya menjelaskan tentang alur kerja yang akan dilalui selama proses penelitian meliputi apa metode yang digunakan, perangkat keras dan perangkat lunak yang dipakai, tahapan perancangan, hingga proses pengujiannya nanti.

BAB IV HASIL DAN PEMBAHASAN, bab ini menyajikan proses implementasi serta hasil yang mencakup proses pengolahan data, implementasi tahap awal hingga akhir, pengujian metode, dan evaluasi sebagai pembandingan dari metode yang diujikan.

BAB V PENUTUP, berisi kesimpulan dan saran yang ditemukan selama penelitian berlangsung serta saran untuk penelitian selanjutnya.