

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan informasi yang pesat telah menyebabkan terjadinya kelebihan informasi (information overload), di mana portal berita menerbitkan sejumlah besar berita setiap harinya. Kondisi ini menyebabkan relevansi waktu suatu berita menjadi sangat singkat karena dengan cepat digantikan oleh berita-berita yang lebih baru. Hal ini tidak sejalan dengan kebutuhan pembaca yang memerlukan waktu lebih lama untuk menemukan dan membaca berita yang sesuai dengan minat mereka. Untuk mengatasi permasalahan tersebut, diperlukan sebuah sistem rekomendasi yang dapat membantu pembaca dalam menemukan dan memilih berita yang relevan. Sistem rekomendasi merupakan sistem yang dirancang untuk memberikan saran informasi yang berguna bagi pengguna dalam mencapai tujuannya, seperti menyarankan artikel untuk dibaca atau memilih produk tertentu sehingga pengguna dapat lebih efektif dalam menentukan pilihan. Sistem ini bekerja dengan menyaring data menggunakan berbagai algoritma untuk merekomendasikan data yang paling relevan sesuai dengan preferensi pengguna.[1]

Sistem rekomendasi merupakan salah satu komponen penting dalam memberikan pengalaman yang bersifat personal kepada pengguna. Sistem ini berfungsi untuk menyaring dan menampilkan informasi yang dianggap paling relevan, sehingga pengguna dapat memperoleh informasi dengan lebih cepat dan efisien. Beberapa metode yang umum digunakan dalam pengembangan sistem rekomendasi antara lain adalah *content-based filtering*, collaborative filtering, dan pendekatan gabungan (*hybrid filtering*). Meskipun telah banyak diterapkan, sistem rekomendasi tetap menghadapi sejumlah tantangan, seperti permasalahan cold-start, keterbatasan dalam skala besar, kecenderungan memberikan rekomendasi yang terlalu sempit (over-specialization), serta perubahan preferensi dan data pengguna yang dinamis[2]

Seiring dengan pesatnya pertumbuhan portal berita digital, ribuan artikel baru diterbitkan setiap harinya mencakup berbagai topik seperti politik, ekonomi, teknologi, dan hiburan. Banyaknya volume informasi yang tersedia seringkali membuat pengguna kesulitan menemukan berita yang benar-benar relevan dengan kebutuhan mereka. Untuk mengatasi permasalahan tersebut, pada penelitian ini dikembangkan sebuah sistem mesin pencari berita yang mampu menelusuri dan menampilkan artikel secara relevan berdasarkan kata kunci (*keyword*) yang dimasukkan pengguna.

Data yang digunakan dalam sistem ini berasal dari portal berita Detik.com, yang dikumpulkan dan diolah untuk membangun basis data berita. Sistem ini menerapkan konsep *Information retrieval* (IR) dengan algoritma TF-IDF dan *cosine similarity*, yang digunakan untuk mencocokkan *query* pengguna terhadap konten berita sehingga dapat menampilkan hasil pencarian yang paling relevan dengan konteks pencarian.

Berdasarkan beberapa penelitian yang telah dikaji, metode *cosine similarity* dinilai efektif dalam meningkatkan relevansi hasil pencarian maupun rekomendasi, terutama dalam domain teks seperti artikel berita. Misalnya, penelitian oleh Arif Akbarul Huda dkk. menunjukkan bahwa penggunaan TF-IDF dan *cosine similarity* dalam sistem rekomendasi artikel berita mampu menghasilkan nilai Recall sebesar 80%, yang menandakan efektivitas metode ini dalam menampilkan artikel yang relevan dengan preferensi pengguna. Selain itu, studi oleh Mohammed Al-Hagery membuktikan bahwa algoritma pencarian berbasis *cosine similarity* dapat menyaring ratusan hasil pencarian menjadi daftar pendek yang lebih akurat dan relevan. Keunggulan ini menjadikan *cosine similarity* sebagai metode yang tepat untuk menangani permasalahan kelebihan informasi dalam sistem rekomendasi berita.

Penelitian oleh Ammar Ismael Kadhim turut memperkuat alasan pemilihan metode TF-IDF. Dalam penelitiannya, dilakukan perbandingan antara teknik feature extraction BM25 dan TF-IDF, dan hasil eksperimen menunjukkan bahwa TF-IDF mampu memberikan nilai F1-measure tertinggi, yaitu 89,77, sedikit lebih tinggi dari BM25 yang memperoleh 89,16. Temuan ini menegaskan bahwa TF-IDF

lebih efektif dalam merepresentasikan pentingnya suatu istilah dalam dokumen, sehingga meningkatkan akurasi proses klasifikasi teks. Dengan

1.2 Perumusan Masalah

1.2.1 Permasalahan

Belum optimalnya sistem pencarian berita dalam menelusuri, menyaring, dan memeringkat artikel berita berdasarkan kata kunci yang dimasukkan pengguna, sehingga hasil pencarian yang ditampilkan sering kali kurang relevan dengan kebutuhan pengguna. Selain itu, sistem pencarian yang ada belum memanfaatkan metode pembobotan kata dan pengukuran kemiripan dokumen secara efektif, khususnya dengan penerapan TF-IDF dan *cosine similarity*, untuk mengukur tingkat kesesuaian antara *query* pengguna dan dokumen berita, sehingga akurasi dan relevansi hasil pencarian belum maksimal.

1.2.2 Metode Usulan

Metode usulan penelitian ini terdiri dari beberapa langkah utama. Pertama, data dikumpulkan melalui web *crawling* dari situs berita Detik.com. Selanjutnya, dilakukan *Preprocessing* teks yang mencakup *case folding*, tokenisasi, *stopword removal*, dan lemmatisasi untuk menyiapkan data. Setelah itu, fitur diekstraksi menggunakan metode TF-IDF untuk menghitung bobot kata. Kemudian, tingkat kemiripan antara *query* pengguna dan dokumen berita diukur dengan *Cosine similarity*. Terakhir, evaluasi sistem dilakukan menggunakan metrik presisi untuk menilai relevansi hasil pencarian.

1.3 Pertanyaan Penelitian

1. Bagaimana metode TF-IDF dapat diimplementasikan untuk meningkatkan relevansi hasil pencarian dalam sistem pencarian berita berbasis konten?
2. Bagaimana metode *Cosine similarity* dapat diimplementasikan untuk meningkatkan efisiensi dalam proses pencocokan dan perankingan hasil pencarian berita berbasis konten?

1.4 Tujuan

1. Merancang dan mengimplementasikan mesin pencari berita yang efektif menggunakan model TF-IDF dan *cosine similarity*.
2. Pembaca akan lebih mudah menemukan berita yang sesuai dengan minat mereka. Sistem yang dirancang akan membantu mereka mendapatkan informasi yang relevan dengan lebih cepat, sehingga pengalaman membaca menjadi lebih menyenangkan.
3. Metode dan temuan yang ada bisa digunakan sebagai dasar untuk penelitian lebih lanjut, membuka peluang untuk inovasi baru.

1.5 Manfaat Penelitian

1. Penelitian ini membantu peneliti memahami lebih dalam tentang teknik teknik seperti TF-IDF dan *Cosine similarity*, dan meningkatkan keterampilan dan pengetahuan mereka di bidang teknologi.
2. Pembaca jadi lebih mudah menemukan berita yang sesuai dengan minat mereka, sehingga pengalaman membaca menjadi lebih menyenangkan.
3. Penelitian ini bisa menjadi acuan bagi peneliti lain yang ingin mengembangkan sistem pencarian atau rekomendasi, membuka jalan untuk inovasi baru.

1.6 Batasan Masalah

Penelitian ini akan dibatasi pada:

1. Penggunaan *dataset* berita dari detik.com.
2. Penggunaan metode TF-IDF dan *cosine similarity* sebagai algoritma utama dalam pengembangan sistem.
3. Fokus pada sistem rekomendasi berita tanpa memperhitungkan faktor lain seperti media sosial atau platform berita lainnya.

1.7 Sistematika Penulisan Proposal Skripsi

Sistematika penulisan skripsi ini disusun sebagai berikut:

BAB I: PENDAHULUAN

Berisi latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, metodologi penelitian, dan sistematika penulisan.

BAB II: TINJAUAN PUSTAKA

Membahas teori-teori dan penelitian terdahulu yang berkaitan, termasuk konsep dasar sistem rekomendasi, TF-IDF, *cosine similarity*, dan web scraping

BAB III: METODE USULAN

Menjelaskan tahapan-tahapan teknis yang dilakukan dalam pengembangan sistem, mulai dari pengumpulan data hingga evaluasi.

BAB IV HASIL DAN PEMBAHASAN

Bab 4 hasil dan pembahasan berisikan tentang hasil penelitian sistem pencarian berita dengan menerapkan TF-IDF dan *Cosine similarity* serta menganalisis hasil kesesuaian dengan query pengguna dan dokumen berita menggunakan berita dari detik.com.

serta melakukan evaluasi sistem dengan menggunakan matrik presisi untuk melihat seberapa relevan hasil yang dicari dengan hasil yang diinginkan.

BAB V KESIMPULAN

Bab 5 berisikan tentang kesimpulan dari hasil penerapan TF-IDF dan cosine, seberapa relevan hasil pencarian menggunakan TF-IDF dan *Cosine similarity*, serta seberapa efektif penerapan metode ini. Selain itu, bab ini juga berisikan tentang saran yang ditujukan untuk penelitian selanjutnya.