

BAB II
TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian sebelumnya tentang topik ini akan dibahas secara menyeluruh dalam subbagian ini. Pada Tabel 2.1 dibawah ini merupakan hasil dari berbagai penelitian sejenis yang menjadi acuan penelitian

Tabel 2.1 Penelitian Terdahulu

| No | Penulis | Tahun | Ringkasan Singkat | Hasil |
|----|---------|-------|--|--|
| 1 | [11] | 2021 | Online Learning Video Recommendation Menggunakan Content-Based Filtering | Hasil pengujian menunjukkan bahwa persentase kinerja rata-rata adalah 81,13% dalam mencapai tujuan sistem rekomendasi, yang meliputi relevansi, novelty, serendipity, dan peningkatan variasi rekomendasi. |
| 2 | [12] | 2022 | Sistem Rekomendasi Artikel Berita dengan TF-IDF | Pada penelitian ini didapat skor Recall@5 sekitar 73% dan Recall@10 sekitar 80% didapatkan dengan menggunakan sistem rekomendasi Content-based Filtering. |
| 3 | [10] | 2021 | Rekomendasi Produk Emina Cosmetics dengan Content-Based Filtering | Dengan metode content-based filtering menghasilkan sepuluh produk yang disarankan dengan nilai similaritas 0,7195. |

2.2 E-Commerce

E-commerce adalah alat yang digunakan untuk melakukan promosi dan jual beli secara daring (online) dengan menggunakan website sebagai platform untuk bertransaksi. Dengan memanfaatkan kemudahan akses internet, bisnis dapat mencapai pasar global tanpa batasan geografis. Konsep ini telah menjadi pendorong utama transformasi usaha di masa digital.

Dengan berjalannya waktu, inovasi pada teknologi internet memainkan peran kunci dalam evolusi e-commerce. Pada awalnya e-commerce lebih terfokus pada transaksi elektronik sederhana, namun seiring dengan kemajuan teknologi, model bisnis e-commerce pun berkembang, termasuk B2C (Business to Consumer), B2B (Business to Business), dan C2C (Consumer to Consumer).

Namun, evolusi e-commerce tidak hanya sebatas pada perkembangan teknologi dan model bisnis. Penelitian yang dilakukan oleh [13] menyoroti peran perubahan perilaku konsumen, di mana preferensi belanja online dan ekspektasi konsumen terhadap pengalaman digital semakin mendominasi.

2.3 Sistem Rekomendasi

Sistem rekomendasi bertujuan untuk memecahkan masalah kelebihan informasi sambil mempersonalisasi pengalaman pengguna dengan memberikan rekomendasi item / produk yang akurat dan dipersonalisasi kepada pengguna sesuai dengan preferensi mereka[3]. Sistem rekomendasi bertujuan untuk memprediksi apakah suatu item akan berguna bagi pengguna berdasarkan informasi yang diberikan. Penggunaan sistem rekomendasi tidak dapat ditebak, mengingat pengaruhnya yang potensial untuk meningkatkan banyak tantangan atas pilihan. Berbagai aplikasi telah mengadopsi sistem rekomendasi, termasuk e-commerce, kesehatan, transportasi, pertanian, dan media.

Sistem rekomendasi dapat diklasifikasikan menjadi *hybrid*, *content-based*, *utility-based*, *demographic-based*, *knowledge-based*, dan *collaborative filtering*. [14]. Pendekatan filtrasi yang paling sering digunakan menggunakan *content-based and collaborative filtering*. Metode *Collaborative filtering* bekerja dengan cara menggunakan pendapat orang lain untuk menilai atau menyaring item [5]. *Content-Based Filtering* (CF) adalah metode rekomendasi produk berdasarkan deskripsi dan preferensi pengguna.

2.4 *Preprocessing Data*

Preprocessing data adalah proses transformasi di mana data mentah dari *big data* diperiksa, disaring, dan diproses menjadi format yang dapat dipahami dan dapat diterapkan pada aplikasi analitik seperti pembelajaran mesin. Proses *preprocessing* data melibatkan penilaian kualitas data dengan menilai keakuratan, kelengkapan, konsistensi, kepercayaan, dan interpretabilitas. Kualitas data sendiri sangat penting karena jika input data dalam aplikasi analitik berkualitas tinggi, outputnya juga akan berkualitas tinggi, yang menunjukkan bahwa kualitas data sangat penting untuk mendapatkan output yang diinginkan [15].

Bergantung pada kebutuhan dan tujuan analisis tertentu, ada banyak teknik yang dapat digunakan untuk *preprocessing* data. Teknik yang paling umum adalah data cleaning, yang berguna untuk menghilangkan data yang hilang, tidak valid, atau tidak relevan dari dataset, dan normalisasi, yang digunakan untuk menyesuaikan skala data agar sama dan memastikan bahwa data dapat dibandingkan dengan benar dan tidak dimanipulasi. Secara keseluruhan, tujuan *preprocessing* data adalah menyiapkan data agar siap untuk dianalisis dan dapat digunakan dengan efektif untuk menjawab pertanyaan penelitian atau menyelesaikan masalah yang ada.

2.5 *Content Based Filtering*

Sistem rekomendasi berbasis konten, juga dikenal sebagai *Content-based filtering*, menggunakan ketersediaan konten suatu item sebagai dasar untuk memberikan rekomendasi. Fitur, atribut, atau karakteristik sering digunakan dalam filter ini. *Content-based filtering* pada *e-commerce* menyediakan rekomendasi produk berdasarkan analisis kesamaan atribut antara produk yang telah dinilai atau dilihat oleh pengguna. Metode ini membentuk profil pengguna berdasarkan atribut produk seperti kategori, merek, atau fitur tertentu yang dianggap relevan. Sebagai contoh, untuk sebuah platform *e-commerce*, atribut pemformatan bisa mencakup jenis produk, merek tertentu, atau kriteria lain yang dianggap penting oleh pengguna. Proses algoritma *content-based filtering* dalam konteks *e-commerce* melibatkan langkah-langkah seperti [16]:

1. suatu produk dibagi menurut vektor dari komponennya.

2. Profil pengguna dibuat oleh sistem berdasarkan berat vektor dari komponen yang membentuk item. Pengguna dapat membuat profil dengan menggunakan algoritma TF-IDF (*terms frequency-inverse document Frequency*). TF-IDF dapat dihitung dengan persamaan berikut:

$$w_{ij} = tf_{ij} \times idf_i$$

$$w_{ij} = tf_{ij} \times \left(\log \frac{D}{df_i} + 1 \right) \quad [11]$$

Pada persamaan diatas digunakan untuk menghitung nilai **TF-IDF**, yang merupakan kombinasi dari dua komponen utama. Komponen pertama adalah tf_{ij} , yaitu frekuensi kemunculan istilah i pada dokumen j , yang memperlihatkan bahwa seberapa sering kata itu muncul dalam dokumen tertentu. Komponen kedua adalah idf_i , yaitu nilai kebalikan dari frekuensi dokumen yang mengandung istilah i , yang berfungsi untuk mengurangi bobot istilah yang terlalu umum dan muncul di banyak dokumen. Dengan mengalikan kedua komponen ini, TF-IDF memberikan bobot lebih tinggi pada kata-kata yang sering muncul tetapi jarang muncul di dokumen lain, sehingga menyoroti kata-kata yang relevan untuk dokumen tersebut.

Berdasarkan profil pengguna yang dibentuk oleh content-based filtering, sistem akan memperkirakan preferensi pengguna terhadap produk berdasarkan analisis kesamaan atribut dari produk yang telah dinilai atau dilihat oleh pengguna. Jika sistem memprediksi bahwa suatu produk memiliki kesamaan atribut yang tinggi dengan produk yang disukai oleh pengguna dalam profilnya, maka pelanggan akan melihat produk ini sebagai pilihan yang mungkin.

2.6 Cosine Similarity

Cosine similarity adalah metode yang digunakan untuk mengukur tingkat kemiripan antara dua vektor n-dimensi, yang umumnya diaplikasikan dalam bidang pencarian informasi untuk membandingkan teks atau dokumen [17]. Nilai cosinus similarity diperoleh dengan menggunakan rumus pada persamaan berikut.

$$Similarity(q, d_j) = \frac{q \cdot d_j}{|q||d_j|} = \frac{\sum_{i=1}^t w_{iq} \times w_{ij}}{\sqrt{\sum_{i=1}^t (w_{iq})^2} \times \sqrt{\sum_{i=1}^t (w_{ij})^2}} \quad [11]$$

Di mana:

- a. q adalah vektor representasi dari kueri.
- b. d_j adalah vektor representasi dari dokumen j .
- c. t adalah jumlah istilah atau kata kunci yang digunakan.
- d. w_{iq} adalah bobot dari istilah i dalam kueri q .
- e. w_{ij} adalah bobot dari istilah i dalam dokumen d_j .

Nilai cosine similarity berada di antara 0 dan 1, jadi nilainya meningkat jika dua teks atau dokumen semakin mirip, dan nilainya menurun jika keduanya semakin tidak mirip.

2.7 Black-box Testing

Black-Box Testing merupakan sebuah pendekatan yang dirancang untuk menguji fungsionalitas dan kegunaan sistem tanpa memperhatikan struktur internal atau implementasi kode. Pendekatan ini bertujuan untuk menguji kelancaran dan kinerja aplikasi secara keseluruhan. Proses pengujian dilakukan dengan cara memasukkan input ke dalam sistem kemudian mengamati apakah output yang dihasilkan sesuai dengan yang diinginkan.

Pengujian *black-box* adalah komponen penting dari pengujian perangkat lunak yang membantu dalam validasi fungsionalitas keseluruhan sistem. Pengujian *black-box* dilakukan berdasarkan kebutuhan pelanggan, sehingga setiap persyaratan yang tidak lengkap atau tidak dapat diprediksi dapat dengan mudah diidentifikasi dan ditangani kemudian. Pentingnya pengujian *black-box* ini terletak pada fakta bahwa pengujian dalam kotak hitam dilakukan berdasarkan perspektif pengguna akhir [18].