

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Terdapat beberapa penelitian terkait dengan penelitian yang dilakukan penulis:

1. *Recommendation System for Property Search Using Content Based Filtering Method*

Penelitian dengan judul “*Recommendation System for Property Search Using Content Based Filtering Method*” dibuat oleh Tessy Badriyah dan rekan. Penelitian ini menggunakan metode *Content-based filtering* dan TF-IDF untuk membuat sistem rekomendasi pencarian properti. Hasil dari penelitian ini menunjukkan bahwa sistem rekomendasi yang dibuat telah berhasil memberikan rekomendasi iklan properti berdasarkan data perilaku penggunaan sistem dan sistem dapat memberikan rekomendasi dengan lebih dari satu kata kunci preferensi pengguna [11]. Penelitian yang dibuat akan menggunakan *Content-based filtering* dan TF-IDF untuk merekomendasikan artikel berita berdasarkan perilaku dan aktivitas pengguna situs Kompas.Id.

2. *Recommendation system: Principles, methods and evaluation*

Penelitian dengan judul “*Recommendation system: Principles, methods and evaluation*” dibuat oleh F.O. Isinkaye, Y.O. Folajimi, dan B.A. Ojokoh. Penelitian ini membahas fase proses rekomendasi, kelebihan dan kekurangan teknik *content-based filtering* dan teknik *filtering* sistem rekomendasi lainnya, dan metrik evaluasi untuk algoritma rekomendasi [2].

Penelitian ini merekomendasi *content-based filtering* jika anda ingin menyaring dokumen seperti *website*, publikasi, atau berita. Penelitian yang dibuat akan menggunakan *content-based filtering* sebagai teknik penyaringan rekomendasi artikel berita dan precision, recall, dan F-measure untuk mengevaluasi tingkat keakuratan dari sistem rekomendasi.

3. *Personalized News Recommendation Based on Consumers' Click Behavior*

Penelitian dengan judul “*Personalized News Recommendation Based on Consumers' Click Behavior*” dibuat oleh Yuqi Wang dan Wenqian Shang. Penelitian ini mengusulkan algoritma rekomendasi baru berdasarkan *collaborative filtering* dan *browsing time* pengguna. Hasil dari penelitian ini adalah algoritma rekomendasi yang diusulkan oleh peneliti lebih baik dibandingkan algoritma rekomendasi tradisional [13]. Penelitian yang dibuat akan menggunakan *scroll depth* pembaca artikel Kompas.id untuk menghasilkan rekomendasi yang lebih baik.

2.2 ***Content-Based Filtering (CBF)***

Content-based filtering adalah metode penyaringan dengan menyimpan informasi mengenai apa yang telah dibaca oleh pengguna sehingga dapat membuat preferensi pengguna sesuai dengan kesukaanya. Dimana kesukaan pengguna terlihat dari interaksi antara pengguna dan artikel yang dibacanya, interaksi terjadi ketika pengguna membuka, membaca, dan mengevaluasi artikel tersebut. Pada CBF informasi di filter berdasarkan preferensi pengguna yang dapat dinyatakan sendiri oleh pengguna atau dilihat dari perilakunya pada zaman dahulu.

Tujuan dari CBF adalah merekomendasikan item yang mirip dengan item yang disukai oleh pengguna di masa lalu. Untuk menentukan suatu item mirip dan

relevan dengan item yang ingin direkomendasikan atau tidak, penulis akan menggunakan metode *cosine similarity* dan *vector space model* TF-IDF. Item yang memiliki nilai *cosine similarity* dan TF-IDF yang tinggi akan direkomendasikan ke pengguna. Dibandingkan *collaborative filtering*, teknik CBF tidak memerlukan banyak pengguna yang serupa untuk memberikan rekomendasi yang akurat.

2.3 Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency atau TF-IDF adalah metode untuk menentukan bobot atau seberapa penting suatu kata dalam suatu dokumen teks dan merupakan *vector space model* yang digunakan untuk mencari similarity antara dokumen teks. Frekuensi munculnya sebuah kata di dalam dokumen menunjukkan seberapa penting kata tersebut di dalam dokumen. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata itu. Jika bobot hubungan antara sebuah kata dan sebuah dokumen tinggi, maka kata tersebut memiliki frekuensi yang tinggi di dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut rendah pada kumpulan dokumen [14].

Pembobotan model TF-IDF adalah probabilistik atau teoretis pada sifatnya dan *term* yang independen adalah asumsi implisit dari model tersebut. Perkiraan nilai relevansi dokumen teks kemudian diperoleh dengan membagi informasi berdasarkan ukuran dari kedua vektor [15]. Rumus TF-IDF adalah [16]:

$$TF - IDF = TF(t) \times IDF(t) \quad (2.1)$$

$$TF(t) = \frac{\text{term } t \text{ muncul pada dokumen}}{\text{Total term pada dokumen}} \quad (2.2)$$

$$IDF(t) = \log \left(\frac{\text{Total dokumen}}{\text{Total dokumen yang mengandung term } t} \right) \quad (2.3)$$

2.4 *Cosine Similarity*

Cosine Similarity adalah metode untuk mengukur kesamaan antara dua *vector* yang merupakan dokumen teks. Suatu dokumen teks dapat direpresentasikan oleh ribuan atribut, masing-masing atribut tersebut merekam frekuensi kata tertentu atau frasa dalam dokumen. Dengan demikian, setiap dokumen teks adalah objek yang direpresentasikan oleh *term-frequency vector* [8]. *Cosine similarity* dapat digunakan untuk membandingkan dokumen berdasarkan *vector* kata pada dokumen tersebut untuk mencari kesamaan. Jika nilai cosine adalah 0, maka kedua dokumen tersebut tidak mirip sama sekali. Jika nilainya mendekati 1, maka kedua dokumen tersebut memiliki kesamaan. Rumus cosine similarity adalah [8]:

$$\cos \text{sim}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} \quad (2.4)$$

A = Vektor A

B = Vektor B

$A \cdot B$ = Dot product antara vektor A dan vektor B

$\|A\|$ = Panjang vektor A

$\|B\|$ = Panjang vektor B

$|A||B|$ = Cross product antara panjang vektor A dan panjang vektor B