

BAB II

TINJAUAN PUSTAKA

Terdapat beberapa penelitian yang bisa membangun sebuah *Job Recommendation System*. Penggunaan metode CBF dalam sistem rekomendasi artikel berita. Dengan menggunakan *dataset* dari portal berita Kabar Informatika Universitas Amikom Yogyakarta, sistem yang dikembangkan mampu mencapai skor *Recall@5* sebesar 73% dan *Recall@10* sebesar 80% [9]. Studi ini menekankan bahwa pendekatan TF-IDF dan *cosine similarity* dapat memberikan rekomendasi yang relevan berdasarkan teks artikel, namun masih memerlukan optimasi lebih lanjut untuk menangani isu *sparsity* dalam data rekomendasi.

Selanjutnya, algoritma sistem rekomendasi berbasis CBF yang telah ditingkatkan dengan mengintegrasikan informasi semantik menggunakan kombinasi TF-IDF dan *word embedding* [10]. Dalam penelitian tersebut, TF-IDF digunakan untuk mengekstraksi kata kunci dari korpus teks, sementara *word embedding* dilatih dengan Word2Vec atau FastText dan digunakan untuk merepresentasikan makna semantik dari kata-kata tersebut. Item *similarity* dihitung dengan pendekatan *weighted similarity* antara TF-IDF dan vektor semantik, dan eksperimen menunjukkan bahwa kombinasi bobot optimal adalah 0.7 untuk TF-IDF dan 0.3 untuk *word embedding*. Sistem diuji menggunakan *dataset* MOOC dan menunjukkan bahwa pendekatan ini menghasilkan akurasi rekomendasi yang lebih tinggi dibandingkan TF-IDF klasik, dengan peningkatan akurasi hingga 7.6% dibandingkan pendekatan *word embedding* saja. Kekurangannya terdapat pada potensi ketergantungan terhadap kualitas data *keyword* dan parameter pelatihan *word vector* yang digunakan dalam representasi semantik.

Di sisi lain, sistem rekomendasi film berbasis *Content-Based Filtering* dengan pendekatan kombinasi antara TF-IDF, Word2Vec sebagai metode *query*, *expansion*, dan *cosine similarity* untuk menghitung tingkat kemiripan antar sinopsis film [11]. *Dataset* yang digunakan terdiri dari 150 data film berbahasa Inggris, dan pengujian dilakukan menggunakan 30 data uji. Penelitian ini menunjukkan bahwa metode gabungan tersebut mampu menghasilkan nilai *precision@10* tertinggi

sebesar 0,47 dan *Mean Average Precision* (MAP) sebesar 0,7096. Meskipun demikian, hasil eksperimen menunjukkan bahwa penambahan kata dari *Word2Vec* tidak selalu memberikan hasil yang lebih baik karena banyak kata tambahan yang tidak relevan secara semantik terhadap *query* awal. Hal ini menandakan bahwa efektivitas metode ini bergantung pada kualitas dan ukuran *dataset*.

Sistem rekomendasi lowongan pekerjaan telah dikembangkan menggunakan TF-IDF dan *cosine similarity* untuk mencocokkan keterampilan pengguna dengan deskripsi pekerjaan [12]. Sistem ini memanfaatkan fitur *key skills*, pengalaman, kategori peran, dan industri, dengan *dataset* yang dikumpulkan dari berbagai sumber rekrutmen. Hasil eksperimen menunjukkan akurasi tinggi dengan skor *cosine similarity* mencapai 0,89 pada pekerjaan yang paling relevan. Keunggulan utama sistem ini adalah kemampuannya dalam memproses rekomendasi dengan menggunakan representasi vektor teks, serta integrasi dengan antarmuka berbasis Streamlit untuk meningkatkan pengalaman pengguna. Namun, penelitian ini juga mencatat kekurangan atribut *dataset* terhadap variasi penulisan deskripsi pekerjaan. Sistem ini dikembangkan menggunakan Python dan Scikit-learn sebagai *framework* utama untuk pemrosesan teks.

Selanjutnya, sistem rekomendasi pekerjaan berbasis *Content-Based Filtering* dengan algoritma utama *cosine similarity* yang didukung oleh teknik representasi teks *word embedding* menggunakan GloVe [13]. Sistem ini dirancang untuk mencocokkan deskripsi profil pengguna dari LinkedIn dengan lowongan pekerjaan berdasarkan kesamaan konten. *Dataset* yang digunakan terdiri dari 1.638 data gabungan antara profil pengguna LinkedIn dan deskripsi pekerjaan. Kelebihan dari pendekatan ini terletak pada kemampuannya memberikan hasil yang relevan berbasis konten, serta kesederhanaan dalam implementasi, cocok untuk platform yang memiliki keterbatasan data pengguna lain.

Kemudian, sistem rekomendasi berbasis *Content-Based Filtering* (CBF) dengan menggabungkan representasi BERT dan TF-IDF [14]. Dalam penelitian tersebut, BERT digunakan untuk menangkap informasi laten dan semantik dari teks berita menggunakan model varian KOBERT. Sementara itu, TF-IDF digunakan

untuk menyoroti kata-kata penting secara statistik dalam dokumen. *Similarity* antar item dihitung menggunakan pendekatan *cosine similarity*, dari kedua vektor tersebut. Melalui eksperimen pada data berita ekspor produk K-wave, pendekatan terbaik ditemukan pada kombinasi bobot 0.3 untuk BERT dan 0.7 untuk TF-IDF, yang menghasilkan peningkatan akurasi rekomendasi secara signifikan dibandingkan penggunaan salah satu metode secara tunggal. Meskipun model ini menunjukkan performa yang unggul, keterbatasan terletak pada jumlah kategori produk yang masih terbatas serta fokus yang hanya pada data berita sebagai sumber informasi.



Tabel 2.1. Perbandingan Penelitian

Penulis	Algoritma	Dataset	Akurasi	Kelebihan	Kekurangan	Bahasa Pemrograman	Machine Learning Library	Implementasi Model
Huda et al. (2022)	<i>Content-Based Filtering</i> (CBF), TF-IDF, <i>Cosine Similarity</i>	Portal berita Kabar Informatika Universitas Amikom Yogyakarta	<i>Recall@5</i> : 73%, <i>Recall@10</i> : 80%	Meningkatkan relevansi rekomendasi berita	Memerlukan optimasi <i>sparsity</i>	Python	Scikit-learn, NLTK	-
Huang (2023)	<i>Content-Based Filtering</i> (CBF), TF-IDF, <i>Weighted Similarity</i> (TF-IDF + <i>Word Embedding</i>)	MOOC dataset	Akurasi prediksi 0.893 (<i>top-1</i>), stabil hingga 0.780 (<i>top-10</i>)	Menggabungkan keunggulan statistik (TF-IDF) dan semantik (<i>embedding</i>)	Hasil sangat bergantung pada kualitas <i>keyword</i> dan parameter <i>word vector</i>	Python	Gensim, Scikit-learn, Pandas, NLTK	-
Laili et al. (2019)	<i>Content-Based Filtering</i> , TF-IDF, Word2Vec, <i>Cosine Similarity</i>	150 data film berbahasa Inggris dari situs The Movie	<i>Precision@10</i> : 0,47; <i>Mean Average Precision</i> (MAP): 0,7096	Mampu meningkatkan perluasan kata kunci dan menangkap kemiripan sinopsis film secara lebih semantik	Perubahan kata dari Word2Vec terkadang tidak relevan secara semantik	Python	Gensim, Scikit-learn, Pandas, NLTK	-

		Database (IMDb)						
Inamdar et al. (2024)	TF-IDF, <i>Cosine Similarity</i>	Data lowongan pekerjaan dari berbagai sumber rekrutmen	<i>Cosine similarity</i> 0,89	integrasi dengan Streamlit untuk UI interaktif	Kekurangan atribut <i>dataset</i> terhadap variasi penulisan deskripsi pekerjaan	Python	Scikit-learn	Streamlit
Habibi & Albanna (2022)	<i>Content-Based Filtering</i> , <i>Cosine Similarity</i> , GloVe	1.638 data dari profil LinkedIn dan <i>dataset</i> job	<i>Cosine similarity</i> 0,31	Memberikan hasil rekomendasi berbasis konten tanpa perlu data pengguna lain, cocok untuk sistem sederhana	Belum ada metrik evaluasi formal.	Python	Pandas, CSV, GloVe, numpy	FlaskAPI
Lee et al. (2023)	CBF, TF-IDF, dan BERT	Data berita ekspor dari KOTRA dan KITA sebanyak 20.756 data	P@1=0,5; P@5=0,28; P@10=0,54; P@20=0,51; MRR=0,5722	Memanfaatkan TF-IDF dan BERT secara bersamaan	Kategori item masih terbatas hanya <i>clothing</i> , <i>cosmetics</i> , dan <i>electronics</i>	Python	-	-

Berdasarkan penelitian sebelumnya, pendekatan *Content-Based Filtering* (CBF) dapat membangun sistem rekomendasi pekerjaan karena mampu mencocokkan profil pengguna dengan deskripsi pekerjaan berdasarkan kesamaan konten teks. Penerapan TF-IDF dan *cosine similarity* pada sistem rekomendasi artikel berita mampu menghasilkan skor *Recall@5* sebesar 73% dan *Recall@10* sebesar 80% [9]. Penelitian lain menyatakan pendekatan CBF dengan menggabungkan TF-IDF dan *Word Embedding* melalui *weighted similarity*, yang terbukti mampu meningkatkan akurasi rekomendasi hingga 7.6% dibandingkan TF-IDF klasik [10]. Sementara itu, Word2Vec dan *cosine similarity* untuk sistem rekomendasi film dan memperoleh *Precision@10* sebesar 0,47, menunjukkan bahwa integrasi semantik dalam representasi teks memberikan hasil yang lebih relevan [11].

Dalam konteks domain pekerjaan, TF-IDF dan *cosine similarity* untuk mencocokkan keterampilan dan kategori peran pengguna dengan deskripsi pekerjaan, mencapai skor kemiripan hingga 0,89 [12]. Pendekatan serupa diterapkan dengan menggunakan GloVe dalam sistem rekomendasi profil LinkedIn dan lowongan kerja [13]. Sementara itu, model rekomendasi berbasis CBF dengan menggabungkan TF-IDF dan BERT, dan menghasilkan MRR tertinggi 0,5722 pada bobot 0.7 (TF-IDF) dan 0.3 (BERT) [14]. Dalam konteks implementasi di PT XYZ, pendekatan CBF dengan TF-IDF, *word embedding*, BERT, dan *cosine similarity* dipilih karena tidak memerlukan data interaksi pengguna lain, sehingga dapat dilakukan sistem rekomendasi meskipun histori penggunaan sistem masih terbatas. Dengan demikian, metode CBF diharapkan dapat memberikan personalisasi yang lebih tepat dalam sistem rekomendasi pekerjaan yang dikembangkan di PT XYZ.