TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1351-1360
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9106

Perancangan Skema Evaluasi untuk Sistem Rekomendasi Berita Menggunakan Metrik Precision, Recall, dan F1-Score

Irwan Kurnia Phan[1,*], Yuricha[2]

[1] Fakultas Teknologi Informasi, Program Studi Bisnis Digital, Universitas Widya Dharma Pontianak, Pontianak, Indonesia
[2] Program Studi Sistem dan Teknologi Informasi, Institut Teknologi dan Bisnis Sabda Setia, Pontianak, Indonesia
Email: [1,*] irwanphan@widyadharma.ac.id, [2] yuricha@itbss.ac.id,
Email Penulis Korespondensi: irwanphan@widyadharma.ac.id

Abstrak–Standarisasi evaluasi untuk sistem rekomendasi berita masih minim, meskipun sistem ini menjadi solusi penting menghadapi information overload di era digital. Penelitian ini dirancang untuk mengembangkan skema evaluasi komprehensif bagi sistem rekomendasi berita berbasis konten menggunakan lima metrik evaluasi utama: Precision, Recall, F1-Score, Hit Rate, dan Mean Reciprocal Rank (MRR). Penelitian menggunakan News Category Dataset dari HuffPost yang berisi 209.527 artikel berita dari 41 kategori. Evaluasi dilakukan dengan mensimulasikan feedback pengguna melalui tiga pendekatan: random baseline sebagai pembanding, content-based filtering dengan TF-IDF, dan Approximate Nearest Neighbor (ANN) berbasis Faiss. Untuk evaluasi final, digunakan 10.000 artikel yang dipilih secara random. Hasil menunjukkan bahwa TF-IDF mencapai Precision@10 sebesar 20,20%, Recall@10 sebesar 0,57%, F1-Score@10 sebesar 1,10%, dan Hit Rate@10 sebesar 69%, sementara ANN menghasilkan Precision@10 sebesar 11,50%, Recall@10 sebesar 0,33%, F1-Score@10 sebesar 0,63%, dan Hit Rate@10 sebesar 43%. Metrik Hit Rate@10 menunjukkan bahwa TF-IDF berhasil memberikan minimal satu artikel relevan pada 69% query, dibandingkan ANN yang mencapai 43% dan Random Baseline yang hanya mencapai 27%. TF-IDF mengungguli ANN dengan Precision@10 sebesar 1,76 kali lebih baik (20,20% vs 11,50%) dan Recall@10 sebesar 1,73 kali lebih baik (0,57% vs 0,33%). Dari segi efisiensi komputasi, TF-IDF mencapai runtime 0,0100 detik per rekomendasi, hanya 1,04 kali lebih cepat dibanding ANN yang mencapai 0,0104 detik, menunjukkan perbedaan yang sangat minimal. Kontribusi utama penelitian adalah skema evaluasi terstruktur menggunakan lima metrik komplementer yang dapat diterapkan pada berbagai sistem rekomendasi berita dan memberikan framework untuk perbandingan objektif antar metode.

Kata Kunci: Sistem Rekomendasi; Metrik Evaluasi; Precision; Recall; F1-Score; Content-Based Filtering; Information Retrieval

Abstract–Standardization of evaluation for news recommendation systems remains minimal, despite the importance of these systems in addressing information overload in the digital era. This research was designed to develop a comprehensive evaluation scheme for content-based news recommendation systems using five primary evaluation metrics: Precision, Recall, F1-Score, Hit Rate, and Mean Reciprocal Rank (MRR). The study utilized the News Category Dataset from HuffPost, which contains 209,527 news articles across 41 categories. Evaluation was conducted by simulating user feedback through three approaches: random baseline as a comparison reference, content-based filtering with TF-IDF, and Approximate Nearest Neighbor (ANN) based on Faiss. For the final evaluation, 10,000 randomly selected articles were used. Results demonstrate that TF-IDF achieved Precision@10 of 20.20%, Recall@10 of 0.57%, F1-Score@10 of 1.10%, and Hit Rate@10 of 69%, while ANN yielded Precision@10 of 11.50%, Recall@10 of 0.33%, F1-Score@10 of 0.63%, and Hit Rate@10 of 43%. The Hit Rate@10 metric shows that TF-IDF successfully provides at least one relevant article in 69% of queries, compared to ANN which achieves 43% and Random Baseline which only achieves 27%. TF-IDF surpasses ANN by 1.76 times in terms of Precision@10 (20.20% vs 11.50%) and 1.73 times in terms of Recall@10 (0.57% vs 0.33%). In terms of computational efficiency, TF-IDF achieves a runtime of 0.0100 seconds per recommendation, only 1.04 times faster than ANN which achieves 0.0104 seconds, showing a very minimal difference. The primary contribution of this research is a structured evaluation scheme using five complementary metrics that can be applied to various news recommendation systems and provides a framework for objective comparison among different methods.

Keywords: Recommendation System; Evaluation Metrics; Precision; Recall; F1-Score; Content-Based Filtering; Information Retrieval

1. PENDAHULUAN

Rekomendasi berita telah menjadi solusi kritis dalam menghadapi ledakan volume informasi di era digital. Setiap hari, jutaan artikel berita dipublikasikan dari berbagai sumber media massa, menciptakan tantangan signifikan bagi pengguna untuk menemukan konten yang relevan dan sesuai dengan preferensi personal mereka. Dalam konteks ini, sistem rekomendasi menjadi alat penting yang membantu memberikan pengalaman membaca yang lebih personal, efisien, dan memuaskan (Lops et al., 2011; Liu et al., 2024; Rong et al., 2024). Teknologi sistem rekomendasi tidak hanya meningkatkan *engagement* pengguna terhadap platform media, tetapi juga membantu pengguna menghemat waktu dalam mencari informasi yang relevan.

Terdapat beberapa pendekatan yang telah dikembangkan untuk membangun sistem rekomendasi berita berbasis konten. Menurut Rajaraman dan Ullman (Rajaraman & Ullman, 2012), pendekatan berbasis konten seperti TF-IDF (Term Frequency-Inverse Document Frequency) sangat efektif untuk menangani data teks dalam skala kecil hingga menengah. TF-IDF memanfaatkan representasi numerik dari teks dengan mempertimbangkan relevansi kata-kata dalam korpus untuk mengukur kemiripan antar dokumen. Selain itu, pendekatan Approximate Nearest Neighbor (ANN) telah berkembang sebagai solusi alternatif yang sangat berguna untuk meningkatkan efisiensi pencarian pada dataset besar. ANN memungkinkan sistem untuk menemukan dokumen yang serupa dengan waktu komputasi yang lebih cepat, menjadikannya ideal untuk skenario produksi dengan skala data yang massive (Hou et al., 2022;Johnson et al., 2017;

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1351-1360
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9106

Malkov & Yashunin, 2016). Perkembangan sistem rekomendasi modern telah mencakup berbagai pendekatan dari *content-based* hingga *collaborative filtering* (Bobadilla et al., 2013).

Meskipun beberapa penelitian telah mengeksplorasi sistem rekomendasi berita, masih terdapat gap signifikan yang perlu diisi. Pertama, Liu et al., (2024) mengimplementasikan *news recommendation* dengan *attention mechanism* yang fokus pada pendekatan *deep learning*, namun belum menyediakan framework evaluasi komprehensif yang membandingkan metode tradisional (TF-IDF) dengan *modern methods* secara sistematis menggunakan *multiple evaluation metrics*. Kedua, Raza & Ding, (2021) melakukan review terhadap *news recommender systems* dan mengidentifikasi challenges dalam domain berita, tetapi review tersebut lebih bersifat deskriptif tanpa memberikan skema evaluasi praktis yang dapat langsung diimplementasikan untuk perbandingan antar metode dengan metrik terstandarisasi. Ketiga, (S. Wu et al., 2023) menyajikan survei komprehensif tentang *graph neural networks* dalam *recommender systems* dengan fokus pada arsitektur advanced, namun penelitian tersebut kurang memberikan perhatian pada *baseline evaluation metrics* dan tidak mengeksplorasi trade-off antara *interpretability dan performance* untuk *content-based methods* yang lebih sederhana. Keempat, Hou et al., (2022) mengembangkan CORE framework untuk *session-based recommendation* dengan *consistent representation space*, akan tetapi penelitian tersebut lebih fokus pada pola sequential dan belum mengeksplorasi perbandingan sistematis antara *similarity-based methods* (Cosine Similarity) dan *approximate methods* (ANN/Faiss) dalam konteks news domain dengan mempertimbangkan aspek skalabilitas dan efisiensi. Keempat penelitian tersebut menunjukkan bahwa meskipun telah ada kemajuan dalam pengembangan algoritma rekomendasi berita, masih kurang adanya framework evaluasi terstruktur yang mengintegrasikan *multiple complementary metrics* (Precision, Recall, F1-Score, Hit Rate, MRR) untuk memberikan perbandingan objektif antara *content-based methods* dengan kompleksitas berbeda, yang menjadi fokus utama penelitian ini.

Penelitian ini dirancang untuk mengembangkan skema evaluasi komprehensif bagi sistem rekomendasi berita berbasis konten yang mengintegrasikan tiga metrik evaluasi utama: Precision, Recall, dan F1-Score. Precision mengukur proporsi rekomendasi yang relevan dari total rekomendasi yang diberikan sistem, memberikan indikasi tentang seberapa akurat sistem dalam memprediksi item yang relevan. Recall mengukur kemampuan sistem untuk mengidentifikasi dan mencakup semua item relevan yang tersedia dalam dataset, menunjukkan coverage sistem terhadap items yang seharusnya direkomendasi. F1-Score, sebagai harmonic mean dari Precision dan Recall, memberikan gambaran performa keseluruhan yang seimbang ketika kedua metrik sama pentingnya dalam evaluasi. Kombinasi ketiga metrik ini memungkinkan evaluasi yang lebih holistik dan objektif terhadap performa sistem rekomendasi.

Penelitian ini menggunakan News Category Dataset dari HuffPost yang terdiri dari 209.527 entri berita dengan 41 kategori unik (Misra, 2022). Dataset ini mencakup kolom-kolom penting seperti headline, deskripsi singkat, kategori, penulis, link artikel, dan tanggal publikasi. Evaluasi dilakukan dengan mensimulasikan feedback pengguna terhadap rekomendasi yang diberikan oleh tiga metode: random baseline sebagai pembanding, content-based filtering dengan TF-IDF menggunakan Cosine Similarity, dan Approximate Nearest Neighbor (ANN) berbasis Faiss untuk pencarian yang lebih efisien.

Kontribusi utama penelitian ini adalah: (1) merancang skema evaluasi terstruktur yang dapat diterapkan secara konsisten pada berbagai sistem rekomendasi berita, (2) melakukan analisis perbandingan performa antara metode TF-IDF dan ANN menggunakan tiga metrik evaluasi yang berbeda, dan (3) mengidentifikasi trade-off antara akurasi rekomendasi dan efisiensi komputasi dalam konteks sistem rekomendasi real-time. Hasil penelitian diharapkan menjadi acuan bagi pengembang sistem rekomendasi berita dalam memilih pendekatan yang paling sesuai dengan kebutuhan spesifik mereka, baik dari segi akurasi maupun efisiensi komputasi.

2. METODOLOGI PENELITIAN

2.1 Kerangka Dasar Penelitian

Penelitian ini menggunakan pendekatan Comparative Experimental Research untuk membandingkan efektivitas dua metode rekomendasi berita berbasis konten dalam hal akurasi dan efisiensi komputasi. Penelitian bersifat deskriptif-komparatif yang berfokus pada perbandingan performa antar metode tanpa menguji hipotesis formal. Penelitian membandingkan tiga pendekatan rekomendasi: (1) Random Baseline sebagai pembanding dasar, (2) TF-IDF dengan Cosine Similarity sebagai pendekatan berbasis konten tradisional, dan (3) Approximate Nearest Neighbor (ANN) dengan Faiss sebagai pendekatan yang dioptimalkan untuk efisiensi. Desain penelitian ini memungkinkan identifikasi trade-off antara akurasi rekomendasi dan kecepatan komputasi, sehingga memberikan insight tentang kapan masing-masing metode paling sesuai untuk diterapkan.

2.2 Sumber Data dan Dataset Penelitian

Penelitian menggunakan News Category Dataset dari Kaggle yang merupakan dataset publik untuk tugas klasifikasi dan analisis kategori berita. Dataset ini terdiri dari 209.527 entri artikel berita dengan 41 kategori unik yang mencakup berbagai topik seperti Politics, Entertainment, Business, dan Health. Setiap artikel dalam dataset mengandung enam kolom utama: (1) link berisi URL artikel asli, (2) headline merupakan judul artikel berita, (3) category adalah kategori klasifikasi artikel, (4) shortdescription adalah deskripsi singkat konten artikel, (5) authors memuat nama penulis artikel, dan (6) date menunjukkan tanggal publikasi artikel.

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1351-1360
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9106

Distribusi kategori dalam dataset menunjukkan ketidakseimbangan yang natural, dengan kategori Politics mencapai 18,7% dari total artikel, Entertainment 14,9%, Business 12,3%, dan Health 9,4%. Analisis karakteristik teks menunjukkan bahwa panjang rata-rata headline adalah 10 kata, sementara shortdescription memiliki rata-rata 23 kata. Statistik ini menunjukkan bahwa dataset memiliki informasi teks yang cukup padat untuk diproses dengan pendekatan TF-IDF maupun ANN.

Penelitian menggunakan dataset yang disampling menjadi 10.000 artikel secara random dengan seed tertentu (random_state=42) untuk memastikan reproduktibilitas hasil penelitian. Dataset ini digunakan untuk seluruh tahapan penelitian, dari preprocessing hingga evaluasi final, memastikan konsistensi pada semua tahapan penelitian.

2.3 Variabel Penelitian

Penelitian ini mengidentifikasi variabel-variabel berikut:
Variabel Independen (IV):

1. Metode rekomendasi: terdiri dari tiga kategori yaitu (1) Random Baseline yang melakukan pemilihan artikel secara random sebagai kontrol, (2) TF-IDF dengan Cosine Similarity yang menggunakan representasi bobot kata untuk pengukuran kemiripan, dan (3) ANN dengan Faiss yang mengoptimalkan pencarian nearest neighbor menggunakan indexing.
2. Parameter metode: Pada TF-IDF, digunakan max_features=1.000 untuk membatasi jumlah feature dan stopwords='English' untuk filtering kata-kata umum. Pada Faiss, digunakan IndexFlatL2 untuk metrik L2 distance dan float32 precision untuk efisiensi memori.

Variabel Dependen (DV):

1. Akurasi rekomendasi: diukur melalui tiga metrik yaitu Precision (proporsi rekomendasi relevan dari total rekomendasi), Recall (proporsi item relevan yang teridentifikasi dari total item yang seharusnya relevan), dan F1-Score (harmonic mean dari Precision dan Recall).
2. Efisiensi komputasi: diukur melalui Runtime, yaitu waktu eksekusi dalam detik untuk menghasilkan satu set rekomendasi.

2.4 Tahapan Penelitian

Penelitian terdiri dari lima tahapan utama yang disusun secara sekuensial:

2.4.1 Pengumpulan dan Persiapan Data

Tahap pertama melibatkan loading dataset News Category Dataset dari file JSON dan pemeriksaan kualitas data. Dataset diverifikasi untuk mengecek keberadaan missing values menggunakan fungsi isnull() dan sum(). Hasil pemeriksaan menunjukkan bahwa dataset tidak memiliki nilai yang hilang pada semua kolom, sehingga tidak diperlukan imputasi atau penghapusan data. Setelah verifikasi, dataset digunakan sebesar 10.000 artikel untuk penelitian.

2.4.2 Preprocessing dan Pembersihan Teks

Preprocessing dilakukan untuk memastikan kualitas teks sebelum vectorization. Langkah-langkah preprocessing mencakup: (1) Konversi ke lowercase menggunakan method .lower() untuk normalisasi teks, (2) Penghapusan karakter non-alfabet seperti simbol, angka, dan tanda baca menggunakan Regular Expressions (regex), (3) Penghapusan spasi berlebih untuk konsistensi format teks. Preprocessing dilakukan pada kolom headline dan shortdescription secara terpisah untuk memastikan bahwa setiap elemen teks dibersihkan dengan baik.

2.4.3 Feature Engineering dan Vectorization

Setelah preprocessing, dilakukan dua operasi feature engineering:

1. Penggabungan Kolom:
   Kolom headline dan shortdescription digabungkan menjadi satu kolom baru bernama 'content' dengan pemisah spasi. Contohnya, artikel dengan headline "WHO Chief Urges Halt To COVID-19 Booster Shots" dan *short description* "WHO chief says booster priority undermines global equity" akan menjadi "who chief urges halt to covid booster shots who chief says booster priority undermines global equity". Tujuan penggabungan adalah untuk menciptakan representasi teks yang lebih lengkap dan menangkap konteks artikel secara lebih holistik.
2. TF-IDF Vectorization:
   Kolom 'content' ditransformasi menjadi vektor numerik menggunakan TfidfVectorizer dari scikit-learn dengan parameter max_features=1.000 dan stopwords='English'. TF-IDF menghitung bobot setiap kata berdasarkan frekuensi kemunculannya dalam dokumen (TF) dan keunikannya dalam corpus (IDF). Output dari tahap ini adalah sparse matrix dengan dimensi (n_articles, 1.000) yang siap untuk kedua pendekatan rekomendasi.

2.4.4 Implementasi Model Rekomendasi

Penelitian mengimplementasikan dua pendekatan rekomendasi:

1. Pendekatan 1 - TF-IDF dengan Cosine Similarity:

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1351-1360
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9106

Tahapan implementasi mencakup: (1) Perhitungan cosine similarity antara TF-IDF vector dari artikel query dan semua artikel lainnya dalam corpus menggunakan formula $\cos(\theta) = (A \cdot B) / (||A|| \times ||B||)$, (2) Sorting hasil similarity berdasarkan score tertinggi, (3) Seleksi top-5 artikel dengan similarity score tertinggi sebagai rekomendasi. Kelebihan pendekatan ini adalah interpretability tinggi karena dapat dijelaskan melalui similarity score dan bobot kata, serta implementasi yang sederhana dan mudah dipahami. Kekurangan utama adalah kompleksitas waktu O(n) per query, menjadikannya kurang efisien untuk dataset besar.

2. Pendekatan 2 - Approximate Nearest Neighbor dengan Faiss:
   Tahapan implementasi mencakup: (1) Konversi TF-IDF sparse matrix menjadi *dense array* dengan format float32 yang kompatibel dengan Faiss, (2) Pembuatan Faiss IndexFlatL2 untuk indexing vektor artikel, (3) Pencarian top-5 nearest neighbors menggunakan L2 distance metric. L2 distance menghitung jarak Euclidean antar vektor: $d(A,B) = \sqrt{(\Sigma(ai - bi)^2)}$. Kelebihan pendekatan ini adalah kecepatan pencarian yang jauh lebih cepat dengan kompleksitas waktu lebih rendah, dan skalabilitas baik untuk dataset besar. Kekurangan adalah interpretability lebih rendah dibanding *cosine similarity*.

2.4.5 Evaluasi dan Pengukuran Metrik

Evaluasi dilakukan melalui dua komponen utama:

1. Simulasi Interaksi Pengguna:
   Untuk mengevaluasi akurasi rekomendasi, penelitian mensimulasikan interaksi pengguna dengan cara: (1) Memilih secara random 10 artikel dari 10.000 dataset sebagai *'clicked articles'* yang merepresentasikan preferensi pengguna, (2) Menyimpan kategori dari *clicked articles* sebagai ground_truth label, (3) Membangkitkan 10 kategori prediksi secara random dari sistem sebagai predicted_categories, (4) Menghitung jumlah *true positives* (TP), *false positives* (FP), dan *false negatives* (FN) dengan membandingkan ground_truth dan *predicted categories*. Simulasi ini dilakukan berulang kali (minimal 100 iterasi) untuk mendapatkan rata-rata metrik yang stabil.
2. Perhitungan Metrik Evaluasi:
   Precision dihitung sebagai $P = TP / (TP + FP)$, mengukur proporsi rekomendasi yang benar-benar relevan. Recall dihitung sebagai $R = TP / (TP + FN)$, mengukur proporsi item relevan yang berhasil diidentifikasi. F1-Score dihitung sebagai $F1 = 2 \times (P \times R) / (P + R)$, memberikan harmonic mean dari Precision dan Recall untuk gambaran performa keseluruhan yang seimbang.
3. Pengukuran Runtime:
   Efisiensi komputasi diukur menggunakan time.time() untuk mencatat waktu awal dan akhir eksekusi fungsi rekomendasi. Runtime dihitung sebagai selisih antara waktu akhir dan waktu awal dalam satuan detik. Pengukuran dilakukan untuk 100+ iterasi pencarian rekomendasi untuk setiap metode, sehingga diperoleh runtime rata-rata yang representatif.

2.5 Teknik Analisis Data

Penelitian menggunakan tiga teknik analisis data utama:

1. Analisis Deskriptif:
   Statistik dataset dianalisis untuk memahami karakteristik dasar, mencakup jumlah total artikel (209.527), jumlah kategori unik (41), distribusi kategori (Politics 18,7%, Entertainment 14,9%, Business 12,3%, Health 9,4%), panjang rata-rata headline (10 kata), dan panjang rata-rata short description (23 kata). Analisis preprocessing juga dilakukan untuk menunjukkan jumlah feature yang dihasilkan setelah vectorization (1.000 features), serta statistik dasar tentang sparsity dari TF-IDF matrix.
2. Analisis Perbandingan (Comparative Analysis):
   Performa ketiga metode rekomendasi dibandingkan secara side-by-side menggunakan metrik yang sama. Untuk runtime, dihitung rata-rata waktu eksekusi per rekomendasi untuk TF-IDF dan ANN, kemudian dibandingkan untuk melihat perbedaan efisiensi. Untuk akurasi, nilai Precision, Recall, dan F1-Score dari kedua metode dibandingkan untuk menentukan metode mana yang memberikan hasil lebih baik. Perbandingan juga melibatkan analisis trade-off antara akurasi dan kecepatan, serta identifikasi kondisi penggunaan yang optimal untuk setiap metode.
3. Evaluasi dan Interpretasi Hasil:
   Hasil metrik evaluasi diinterpretasikan dalam konteks *problem statement* dan *research objectives*. Nilai-nilai metrik dianalisis untuk memahami apa yang mereka representasikan dalam konteks sistem rekomendasi nyata. Perbandingan dengan baseline random membantu menunjukkan seberapa baik masing-masing metode dalam memberikan rekomendasi yang relevan. Diskusi juga mencakup implikasi praktis dari hasil penelitian, termasuk rekomendasi tentang metode mana yang paling sesuai untuk berbagai skenario (dataset kecil vs besar, prioritas akurasi vs kecepatan).

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1351-1360
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9106

3. HASIL DAN PEMBAHASAN

3.1 Hasil Evaluasi Runtime

Pengukuran runtime dilakukan untuk mengevaluasi efisiensi komputasi kedua metode rekomendasi. Pengukuran dilakukan pada dataset yang sama (10.000 artikel) dengan jumlah iterasi yang konsisten untuk memastikan perbandingan yang adil dan akurat. Hasil pengukuran runtime menunjukkan perbedaan antara kedua metode:

Tabel 1. Perbandingan Runtime Kedua Metode Rekomendasi

Metode Rekomendasi	Runtime (detik)
TF-IDF + Cosine Sim.	0.01
ANN + Faiss	0.0104
Selisih	0.0004 detik
Rasio (ANN/TF-IDF)	1.04x

Dari hasil pada Tabel 1, TF-IDF dengan Cosine Similarity mencapai runtime rata-rata 0.01 detik per rekomendasi, sementara ANN dengan Faiss mencapai 0.0104 detik. Ini berarti TF-IDF lebih cepat 1.04 kali dibanding ANN pada dataset kecil hingga menengah (10.000 artikel). Interpretasi hasil runtime menunjukkan bahwa pada dataset dengan ukuran 10.000 artikel, pendekatan TF-IDF + Cosine Similarity memiliki sedikit keunggulan dalam hal kecepatan eksekusi. Hal ini disebabkan oleh beberapa faktor: (1) TF-IDF menggunakan matriks yang sudah pre-computed, sehingga perhitungan cosine similarity dapat dilakukan secara langsung tanpa overhead tambahan, (2) Kompleksitas waktu TF-IDF adalah O(n) untuk setiap query, di mana n adalah jumlah artikel, (3) ANN memerlukan proses indexing awal dan overhead komputasi untuk approximate search, meskipun keuntungan ANN akan menjadi jelas pada dataset yang jauh lebih besar. Penting untuk dicatat bahwa runtime ANN tidak jauh lebih lambat (hanya 0.0004 detik lebih lama), sehingga perbedaan ini dapat diakui sebagai sangat kecil. Baseline ini memberikan perspektif berbeda tentang apa yang diharapkan "mirip" dalam hasil evaluasi metrik berbasis konten. Evaluasi akurasi dilakukan dengan mensimulasikan interaksi pengguna pada dataset 10.000 artikel. Simulasi ini melibatkan pemilihan acak 100 artikel sebagai query, dan sistem merekomendasikan 10 artikel teratas untuk setiap query. Relevansi didefinisikan berdasarkan kesamaan kategori antara artikel query dan artikel yang direkomendasikan. Hasil evaluasi akurasi untuk baseline (*random prediction*) adalah sebagai berikut:

Tabel 2. Hasil Evaluasi Akurasi Metrik Baseline

Metrik Evaluasi	Nilai Baseline
Precission@10	0,0402 (4,20%)
Recall@10	0,0007 (0,07%)
F1-Score@10	0,0014 (0,14%)
Hit Rate@10	0.2700 (27,00%)
MRR@10	0.0934

Interpretasi hasil baseline menunjukkan bahwa dengan strategi random selection, sistem hanya mampu mencapai precision sebesar 4.20%, yang berarti dari 10 rekomendasi yang diberikan, hanya sekitar 0-1 artikel yang relevan dengan kategori query. Hal ini menunjukkan bahwa random baseline tidak efektif sebagai strategi rekomendasi.

Tabel 3. Hasil Evaluasi Akurasi Kedua Metode

Metrik Evaluasi	TF-IDF Cosine	ANN (Faiss)	Nilai Baseline
Precission@10	0,2020(20,20%)	0,1150(11,50%)	0,0402 (4,20%)
Recall@10	0,0057(0,57%)	0,0033(0,33%)	0,0007 (0,07%)
F1-Score@10	0,0110(1,10%)	0,0063(0,63%)	0,0014 (0,14%)
Hit Rate@10	0,6900(69,00%)	0,4300(43,00%)	0.2700 (27,00%)
MRR@10	0,3719	0,2420	0.0934

Dari hasil evaluasi pada Tabel 3 tersebut, TF-IDF dengan Cosine Similarity menunjukkan performa yang lebih baik dibandingkan dengan ANN berbasis Faiss. Precision@10 TF-IDF mencapai 20.20%, yang berarti rata-rata 2 artikel dari 10 rekomendasi relevan dengan kategori query. Recall@10 untuk TF-IDF adalah 0.57%, menunjukkan bahwa sistem mampu mengidentifikasi sebagian kecil dari total artikel yang sebenarnya relevan di dataset. Hit Rate@10 sebesar 69% berarti dalam 69% dari test queries, sistem berhasil menemukan minimal satu artikel yang relevan dalam top-10 rekomendasi. MRR@10 (Mean Reciprocal Rank) sebesar 0.3719 menunjukkan posisi rata-rata dari artikel relevan pertama di dalam ranking, dengan nilai ini sistem cukup baik dalam menempatkan artikel relevan di posisi awal.

ANN dengan Faiss menunjukkan nilai Precision@10 sebesar 11.50%, lebih rendah dari TF-IDF, dengan Recall@10 0.33% dan Hit Rate@10 43%. Meskipun ANN memiliki performa lebih rendah pada dataset kecil hingga menengah ini, metode ini masih mampu memberikan rekomendasi yang jauh lebih baik dari random baseline. Perbandingan dengan random baseline menunjukkan peningkatan signifikan: TF-IDF meningkatkan precision hingga 380.95%, recall hingga 696.70%, dan hit rate hingga 155.56%.

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1351-1360
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9106

Kesimpulannya, TF-IDF dengan Cosine Similarity adalah pilihan yang lebih optimal untuk dataset berukuran 10.000 artikel, dengan memberikan hasil rekomendasi yang lebih relevan dan akurat dibandingkan dengan ANN berbasis Faiss pada skala dataset ini.

3.2 Hasil Rekomendasi Kualitatif

Penelitian juga menganalisis hasil rekomendasi secara kualitatif dengan melihat top-5 artikel yang direkomendasikan oleh kedua metode untuk artikel yang sama (index 0).

Tabel 4. Top-5 Rekomendasi dengan TF-IDF + Cosine Similarity

Ranking	Headline Rekomendasi
1	This Might Be The Democratic Party's Costliest...
2	WHO Chief Urges Halt To COVID-19 Booster Shots...
3	Taylor Swift Teams Up With Ryan Reynolds To De...
4	Puerto Ricans Desperate For Water After Hurric...
5	Why Kamala Harris Is A Historic Vice President...

Tabel 5. Top-5 Rekomendasi dengan ANN + Faiss

Ranking	Headline Rekomendasi
1	Kimmel Host Nikki Glaser Absolutely Goes The...
2	Mariah Carey Brings Big, Big Energy To Lattos...
3	Dog On The Mend After Rescuing Owner From Moun...
4	Oscars 2022 See The Complete Winners List
5	Haiti's Health Professionals Go On Strike Over...

Perbandingan hasil rekomendasi pada Tabel 4 dan Tabel 5 menunjukkan perbedaan signifikan dalam jenis artikel yang direkomendasikan oleh kedua metode:

1. TF-IDF + Cosine Similarity menghasilkan rekomendasi yang lebih berfokus pada topik-topik spesifik seperti Democratic Party politics, COVID-19 *vaccines*, dan *political news*. Hasil ini menunjukkan bahwa TF-IDF berhasil menangkap kesamaan semantik berbasis kata-kata kunci yang muncul dalam artikel query.
2. ANN + Faiss menghasilkan rekomendasi yang lebih beragam dalam kategori, mencakup entertainment (talk shows, music, awards), human interest (dog rescue), dan world news (Haiti). Perbedaan ini mungkin disebabkan oleh cara ANN mengukur jarak Euclidean dalam ruang vektor, yang mungkin memberikan weight berbeda pada fitur-fitur tertentu dibanding cosine similarity.

Kedua metode menghasilkan artikel-artikel yang memiliki teks dan konteks berbeda, menunjukkan bahwa setiap metode memiliki karakteristik unik dalam menemukan kesamaan antar dokumen. Tidak ada metode yang "lebih benar", tetapi keduanya memberikan perspektif berbeda tentang apa yang dianggap "mirip".

3.3 Analisis Perbandingan Antar Metode

Perbandingan komprehensif antara TF-IDF + Cosine Similarity dan ANN + Faiss dapat dilihat dari berbagai dimensi:

Tabel 6. Perbandingan Karakteristik Kedua Metode

Nama	TF-IDF	ANN
Runtime (detik)	0.01	0.0104
Kompleksitas Waktu	O(n)	O(log n) approx.
Interpretabilitas	Tinggi	Rendah
Scalability	Sedang	Tinggi
Feature Importance	Jelas (TF-IDF)	Implicit
Preprocessing Overhead	Rendah	Tinggi (indexing)
Memory Usage	Sedang	Tinggi

TF-IDF lebih cepat hanya 1.04 kali dibanding ANN pada dataset 10.000 artikel, yang berarti perbedaannya sangat minimal (hanya 0.0004 detik). Perbedaan kecil ini adalah konsekuensi dari overhead yang diperlukan ANN untuk menjalankan indexed search. Namun, keunggulan runtime TF-IDF pada dataset kecil ini akan berkurang signifikan pada dataset yang lebih besar karena kompleksitas O(n) akan mendominasi waktu eksekusi. TF-IDF menawarkan interpretabilitas yang lebih tinggi karena setiap rekomendasi dapat dijelaskan melalui TF-IDF scores dan kata-kata yang memberikan kontribusi terbesar pada similarity score. ANN, di sisi lain, bersifat black-box karena sulit untuk menjelaskan mengapa vektor tertentu dianggap sebagai nearest neighbor. Untuk dataset dengan jutaan artikel, ANN memiliki keunggulan signifikan karena kompleksitas waktu aproksimasi yang lebih baik. Meskipun ANN lebih lambat pada dataset kecil (dengan margin yang sangat kecil), gap ini akan menyusut seiring pertumbuhan dataset, dan ANN akan menjadi lebih cepat pada dataset yang sangat besar, terutama jika menggunakan approximate indexing methods yang lebih sophisticated seperti HNSW atau IVF.

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1351-1360
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9106

3.4 Implikasi dan Rekomendasi

Hasil evaluasi menunjukkan bahwa kedua metode content-based recommendation yang diimplementasikan memiliki trade-off berbeda yang harus dipertimbangkan dalam konteks aplikasi praktis.

1. TF-IDF + Cosine Similarity menawarkan kecepatan yang lebih tinggi (0.0100 detik) dan interpretabilitas yang lebih baik, menjadikannya pilihan ideal untuk sistem rekomendasi yang memerlukan penjelasan kepada end-user atau untuk deployment dengan resource terbatas. Dalam konteks sistem rekomendasi berita di portal media, TF-IDF dapat digunakan untuk memberikan rekomendasi real-time dengan kemampuan explainability yang tinggi. Pengguna dapat dipahami bahwa artikel A direkomendasikan karena memiliki kesamaan kata-kata kunci tertentu dengan artikel yang mereka baca sebelumnya.
2. ANN + Faiss menawarkan scalability yang lebih baik dan efisiensi untuk dataset besar, meskipun dengan keunggulan runtime yang minimal pada dataset kecil (0.0104 detik, hanya 1.04 kali lebih lambat dari TF-IDF). Pada sistem berita dengan miliaran artikel dan jutaan pengguna concurrent, ANN menjadi pilihan yang lebih sesuai karena dapat menangani search load yang masif dengan efisiensi yang lebih baik. Keunggulan ANN akan semakin terlihat jika dataset berkembang ke ratusan juta atau miliaran artikel, di mana kompleksitas O(n) dari TF-IDF akan menjadi bottleneck performa.

ANN + Faiss menawarkan scalability yang lebih baik dan efisiensi untuk dataset besar, meskipun dengan keunggulan runtime yang minimal pada dataset kecil (0.0104 detik, hanya 1.04 kali lebih lambat dari TF-IDF). Pada sistem berita dengan miliaran artikel dan jutaan pengguna concurrent, ANN menjadi pilihan yang lebih sesuai karena dapat menangani search load yang masif dengan efisiensi yang lebih baik. Keunggulan ANN akan semakin terlihat jika dataset berkembang ke ratusan juta atau miliaran artikel, di mana kompleksitas O(n) dari TF-IDF akan menjadi bottleneck performa.

Perbedaan hasil rekomendasi antara TF-IDF dan ANN (seperti terlihat di Tabel 4 dan 5) menunjukkan bahwa kedua metode menggunakan perspektif berbeda dalam mendefinisikan "kesamaan" antar artikel. TF-IDF fokus pada kesamaan *term-level* (kata-kata yang muncul dalam dokumen), sementara ANN fokus pada kesamaan *vector-level* (jarak Euclidean dalam ruang vektor berdimensi tinggi). Kedua perspektif ini valid dan dapat memberikan insight berbeda:

1. Rekomendasi berbasis TF-IDF lebih konsisten dengan topik spesifik (*politics*, *vaccines*, Democratic Party) dari artikel query
2. Rekomendasi berbasis ANN lebih beragam dalam kategori tetapi tetap mempertahankan relevansi semantik melalui similarity vector

Dalam praktiknya, sistem rekomendasi (Gomez-Uribe & Hunt, 2015, He et al., 2020) *commercial* sering menggunakan *ensemble* dari kedua metode ini untuk mendapatkan keuntungan dari keduanya.

3.4.1 Implikasi Dataset dan Preprocessing

Dataset News Category dengan 209.527 artikel dan 41 kategori menunjukkan bahwa distribusi kategori tidak uniform, dengan kategori seperti Politics dan Entertainment mendominasi (18.7% dan 14.9%). Ini mencerminkan distribusi konten real-world di platform berita, di mana beberapa topik lebih populer daripada yang lain. Implikasi praktis adalah bahwa sistem rekomendasi harus mempertimbangkan class imbalance ini dalam evaluasi metrik, terutama ketika mengukur per-category performance.

Preprocessing yang dilakukan (*lowercase conversion*, *special character removal*, *column merging*) adalah standard dalam NLP tasks dan memastikan bahwa teks diolah dengan konsisten. Penggunaan TF-IDF dengan max_features=1.000 dan *English stopwords removal* adalah praktik baik yang mengurangi noise sambil mempertahankan fitur yang informatif. Penggabungan *headline* dan *short description* menciptakan representasi teks yang lebih kaya, meningkatkan kemampuan sistem untuk menangkap konteks artikel secara lebih holistik.

3.4.2 Implikasi Metodologi Evaluasi

Dalam simulasi interaksi pengguna, metodologi evaluasi yang digunakan merupakan versi simplifikasi dari evaluasi real-world yang lebih kompleks. Evaluasi pada tahap produksi sebenarnya memerlukan integrasi beberapa komponen penting yang tidak sepenuhnya tercakup dalam simulasi ini, termasuk *real user interactions* dengan *explicit* dan *implicit* feedback untuk pengukuran yang lebih autentik, pertimbangan *temporal dynamics* yang membedakan artikel lama dan baru sesuai preferensi kesegaran berita, *personalization* yang mengakomodasi preferensi individual setiap pengguna, serta context awareness yang mencakup faktor-faktor kontekstual seperti waktu akses, jenis device, dan lokasi geografis pengguna.

Meskipun demikian, simulasi yang digunakan dalam penelitian ini telah memadai untuk memvalidasi superioritas sistem rekomendasi berbasis konten dibandingkan *random selection*, sekaligus memberikan perbandingan karakteristik yang jelas antara metode TF-IDF dan ANN.

3.4.3 Generalisasi

Temuan dalam penelitian ini menunjukkan bahwa *content-based filtering* dengan TF-IDF dan ANN dapat memberikan hasil yang signifikan lebih baik daripada *random selection*. Namun, efektivitas kedua metode dapat bervariasi

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1351-1360
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9106

tergantung pada beberapa faktor kunci. Pertama, karakteristik domain aplikasi memiliki pengaruh signifikan terhadap performa metode rekomendasi, di mana sistem rekomendasi berita memiliki karakteristik berbeda dibandingkan sistem untuk produk e-commerce atau musik dalam hal pola konsumsi, *temporal dynamics*, dan preferensi pengguna. Kedua, distribusi dan kualitas data dalam dataset sangat mempengaruhi hasil rekomendasi, termasuk keseimbangan kategori, kelengkapan informasi teks, dan konsistensi metadata yang tersedia. Ketiga, skalabilitas sistem dan ketersediaan resource komputasi menjadi pertimbangan penting dalam pemilihan metode, di mana TF-IDF lebih cocok untuk infrastruktur terbatas sementara ANN memerlukan resource lebih besar namun menawarkan skalabilitas lebih baik. Keempat, trade-off antara interpretability dan akurasi harus dipertimbangkan sesuai kebutuhan aplikasi, di mana beberapa sistem memerlukan explainability tinggi untuk user trust sementara yang lain memprioritaskan akurasi maksimal.

Untuk implementasi di production environment, pertimbangan praktis seperti *deployment complexity*, *monitoring requirements*, dan *maintenance overhead* juga menjadi faktor penting selain performa metrik evaluasi, mengingat sistem yang lebih sederhana cenderung lebih mudah dioperasikan dan dipelihara dalam jangka panjang.

3.4.4 Rekomendasi

Berdasarkan hasil penelitian, untuk sistem rekomendasi skala kecil hingga menengah dengan dataset kurang dari 1 juta artikel, pendekatan TF-IDF dengan Cosine Similarity menjadi pilihan yang paling optimal. Metode ini menawarkan keunggulan dalam hal kecepatan eksekusi dan interpretabilitas hasil yang tinggi, membuatnya ideal untuk deployment dengan resource infrastruktur terbatas. Implementasi TF-IDF relatif *straightforward* menggunakan library scikit-learn yang sudah mature dan teruji, sehingga dapat di-deploy dengan infrastruktur sederhana seperti single server atau small cluster tanpa memerlukan konfigurasi kompleks. Keunggulan utama pendekatan ini adalah kemampuan *explainability* yang memungkinkan sistem memberikan penjelasan transparan kepada pengguna, misalnya dengan menginformasikan bahwa artikel Y direkomendasikan karena memiliki kesamaan kata kunci tertentu dengan artikel X yang baru saja dibaca pengguna.

Untuk sistem rekomendasi skala besar dengan dataset lebih dari 1 juta artikel, pendekatan ANN dengan Faiss atau *approximate indexing methods* lainnya seperti Hierarchical Navigable Small World (HNSW) dan Inverted File (IVF) menjadi pilihan yang lebih sesuai. Metode ANN menawarkan skalabilitas yang linear atau bahkan sublinear dengan pertumbuhan jumlah artikel, memungkinkan sistem tetap responsif meskipun dataset berkembang hingga jutaan atau miliaran artikel. Namun, implementasi ANN memerlukan infrastruktur yang lebih *sophisticated* termasuk *distributed indexing*, GPU *acceleration*, dan sistem monitoring yang lebih *advanced*, serta tim *engineering* yang terampil untuk *maintenance*. *Trade-off* interpretabilitas perlu diterima dalam pendekatan ini, di mana sistem mengorbankan *explainability* untuk mendapatkan *performance* optimal pada skala masif, yang seringkali merupakan pertukaran yang dapat diterima untuk aplikasi produksi berskala besar.

Pendekatan hybrid yang mengkombinasikan kekuatan kedua metode merupakan strategi yang direkomendasikan untuk mencapai keseimbangan optimal antara kecepatan, akurasi, dan interpretabilitas (Boka et al., 2024; Burke, 2002). Dalam arsitektur hybrid ini, TF-IDF dapat digunakan untuk tahap *initial candidate generation* dengan melakukan fast filtering terhadap corpus besar untuk menghasilkan subset kandidat artikel yang relevan, kemudian ANN digunakan untuk tahap re-ranking dan refinement untuk mengurutkan ulang kandidat dengan presisi yang lebih tinggi. *Scores* dari kedua metode dapat dikombinasikan menggunakan *weighted ensemble* yang disesuaikan dengan karakteristik aplikasi spesifik, sehingga menghasilkan rekomendasi yang optimal dalam hal kecepatan response time, akurasi relevansi, dan interpretabilitas hasil untuk user trust.

3.5 Limitasi dan Future Work

Penelitian ini memiliki beberapa limitasi yang perlu dipertimbangkan dalam interpretasi hasil. Pertama, evaluasi dilakukan menggunakan *simulated user interactions* berbasis kategori matching, bukan *real user feedback* dengan pola preferensi yang kompleks dan dinamis. Kedua, baseline metrics yang masih relatif rendah menunjukkan bahwa kedua metode *content-based* yang diimplementasikan masih memerlukan peningkatan signifikan, misalnya melalui *incorporating collaborative filtering* atau *matrix factorization* untuk memanfaatkan pola interaksi pengguna (Koren et al., 2009; Walker et al., 2022). Ketiga, penelitian belum mempertimbangkan *temporal dynamics* atau *news freshness* yang menjadi faktor krusial dalam konteks rekomendasi berita real-time, di mana artikel yang lebih baru cenderung lebih relevan bagi pengguna. Keempat, tidak adanya penelitian pengguna untuk memvalidasi kualitas rekomendasi secara subjektif membuat penelitian ini belum dapat mengkonfirmasi apakah hasil rekomendasi benar-benar meningkatkan kepuasan pengguna dan *engagement* dalam penggunaan aktual.

Untuk penelitian masa depan, beberapa arah pengembangan untuk mengatasi limitasi dan meningkatkan performa sistem rekomendasi. Pertama, implementasi *content-based filtering* dengan deep learning seperti word embeddings dan transformers dapat meningkatkan pemahaman semantic terhadap konten artikel (Le & Mikolov, 2014; Mikolov et al., 2013; Zhang et al., 2020). Kedua, *incorporation of collaborative filtering* dapat memanfaatkan pola interaksi pengguna untuk memberikan rekomendasi yang lebih personal dan akurat (Raza & Ding, 2021; Walker et al., 2022; C. Wu et al., 2019; S. Wu et al., 2023). Ketiga, A/B testing dengan *real users* diperlukan untuk memvalidasi peningkatan dalam kepuasan pengguna dan mengukur dampak aktual terhadap *engagement metrics*. Keempat, *experiment* dengan *approximate nearest neighbor methods* yang lebih *advanced* seperti HNSW dan IVF-PQ dapat memberikan peningkatan skalabilitas untuk deployment berskala masif. Kelima, *consideration of temporal dynamics*

TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1351-1360
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9106

dan *news decay* perlu diintegrasikan ke dalam sistem untuk meningkatkan *freshness* rekomendasi dan memastikan bahwa artikel terbaru mendapat prioritas yang sesuai dalam ranking hasil.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan skema evaluasi komprehensif untuk sistem rekomendasi berita berbasis konten menggunakan lima metrik evaluasi: Precision, Recall, F1-Score, Hit Rate, dan Mean Reciprocal Rank (MRR). Evaluasi terhadap tiga metode (Random Baseline, TF-IDF Cosine Similarity, dan ANN Faiss) pada dataset 10.000 artikel dari News Category Dataset menghasilkan temuan kuantitatif yang signifikan. Dari segi akurasi rekomendasi, TF-IDF Cosine Similarity menunjukkan performa terbaik dengan Precision@10 sebesar 18,7%, Recall@10 sebesar 0,77%, F1-Score@10 sebesar 1,44%, Hit Rate@10 sebesar 64%, dan MRR@10 sebesar 0,288. TF-IDF mengungguli ANN Faiss secara signifikan: 1,76 kali lebih baik dalam hal Precision@10 (18,7% vs 8,7%), dan 1,73 kali lebih baik dalam hal Recall@10 (0,77% vs 0,33%), serta 2,25 kali lebih baik dalam hal F1-Score@10 (1,44% vs 0,64%). Metrik Hit Rate@10 menunjukkan bahwa 64% query dengan TF-IDF berhasil mendapat minimal satu artikel relevan, dibandingkan ANN yang hanya mencapai 36% (sama dengan Random Baseline). Dari segi efisiensi komputasi, TF-IDF mencapai *runtime* rata-rata 0.0100 detik per rekomendasi, hanya 1.04 kali lebih cepat dibanding ANN yang mencapai 0.0104 detik. Keunggulan kecepatan TF-IDF pada dataset kecil hingga menengah ini adalah konsekuensi dari *overhead* yang diperlukan ANN untuk menjalankan *indexed search*. Namun, keunggulan runtime TF-IDF akan berkurang signifikan pada dataset yang lebih besar karena kompleksitas O(n) akan mendominasi waktu eksekusi. Perbandingan dengan Random Baseline menunjukkan efektivitas metode berbasis konten. TF-IDF meningkatkan Precision sebesar 345% (dari 4,2% menjadi 18,7%), Recall sebesar 714% (dari 0,07% menjadi 0,77%), F1-Score sebesar 696% (dari 0,14% menjadi 1,44%), dan Hit Rate sebesar 155% (dari 27% menjadi 64%) dibandingkan baseline. *Improvement* signifikan ini memvalidasi pentingnya menggunakan *content-based filtering* dibandingkan *random selection*. Kontribusi utama penelitian mencakup: (1) skema evaluasi terstruktur menggunakan lima metrik komplementer yang dapat diterapkan pada berbagai sistem rekomendasi berita, (2) temuan empiris bahwa TF-IDF unggul dari ANN pada dataset kecil hingga menengah (2,15x), recall (2,26x), dan kecepatan (2,35x), dengan performa lebih baik dalam *accuracy trade-off* antara akurasi rekomendasi dan efisiensi komputasi, dan (4) *baseline metrics* yang dapat menjadi acuan untuk penelitian selanjutnya dalam evaluasi sistem rekomendasi berita. Penelitian ini memiliki beberapa limitasi yang perlu dipertimbangkan. Evaluasi menggunakan simulasi interaksi pengguna berbasis kategori matching, bukan real user feedback dengan preferensi kompleks. Nilai Recall@10 yang rendah (0,77% untuk TF-IDF) menunjukkan bahwa kedua metode masih perlu peningkatan. Dataset terbatas pada 10.000 artikel dari satu domain (berita), sehingga generalisasi ke domain lain perlu validasi lebih lanjut. Penelitian belum mempertimbangkan *temporal dynamics* atau *news freshness* yang penting dalam konteks berita *real-time*. Untuk penelitian masa depan, beberapa arah pengembangan disarankan: (1) implementasi *content-based filtering* dengan *deep learning* (*word embeddings*, *transformers*), (2) *incorporation of collaborative filtering* untuk memanfaatkan *user interaction patterns*, (3) A/B testing dengan *real users* untuk memvalidasi peningkatan dalam kepuasan pengguna, (4) *experiment* dengan metode *approximate nearest neighbor* yang lebih advanced (HNSW, IVF-PQ) untuk peningkatan skalabilitas, dan (5) kombinasi dengan pendekatan *hybrid* yang menggabungkan kekuatan TF-IDF dan ANN untuk optimal performa pada berbagai scale dataset.

REFERENCES

Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender Systems Survey. *Knowledge-Based Systems*, *46*, 109–132. https://doi.org/10.1016/j.knosys.2013.03.012

Boka, T. F., Niu, Z., & Neupane, R. B. (2024). A Survey of Sequential Recommendation Systems: Techniques, Evaluation, And Future Directions. *Information Systems*, *125*. https://doi.org/10.1016/J.IS.2024.102427

Burke, R. (2002). Hybrid Recommender Systems: Survey And Experiments. *User Modelling and User-Adapted Interaction*, *12*(4), 331–370. https://doi.org/10.1023/A:1021240730564

Gomez-Uribe, C. A., & Hunt, N. (2015). The Netflix Recommender System: Algorithms, Business Value, And Innovation. *ACM Transactions on Management Information Systems*, *6*(4), 1. https://doi.org/10.1145/2843948

He, X., Deng, K., Wang, X., Li, Y., Zhang, Y. D., & Wang, M. (2020). LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. *SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 639–648. https://doi.org/10.1145/3397271.3401063

Hou, Y., Hu, B., Zhang, Z., & Zhao, W. X. (2022). CORE: Simple And Effective Session-Based Recommendation Within Consistent Representation Space. *SIGIR 2022 - Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1796–1801. https://doi.org/10.1145/3477495.3531955

Johnson, J., Douze, M., & Jegou, H. (2017). Billion-Scale Similarity Search with GPUs. *IEEE Transactions on Big Data*, *7*(3), 535–547. https://doi.org/10.1109/TBDATA.2019.2921572

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. *Computer*, *42*(8), 30–37. https://doi.org/10.1109/MC.2009.263


TIN: Terapan Informatika Nusantara
Vol 6, No 8, January 2026, page 1351-1360
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i8.9106

Le, Q., & Mikolov, T. (2014). Distributed Representations of Sentences and Documents. *Proceedings of the 31 St International Conference on Machine Learning*, 1188–1196. https://proceedings.mlr.press/v32/le14.html

Liu, T., Xu, *, Qiao, Yuxin, Jiang, & Chen. (2024). News Recommendation with Attention Mechanism. *Journal of Industrial Engineering and Applied Science*, *2*(1). https://doi.org/10.5281/zenodo.10635481

Lops, P., de Gemmis, M., & Semeraro, G. (2011). Content-based Recommender Systems: State of the Art and Trends. *Recommender Systems Handbook*, 73–105. https://doi.org/10.1007/978-0-387-85820-3_3

Malkov, Y. A., & Yashunin, D. A. (2016). Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(4), 824–836. https://doi.org/10.1109/TPAMI.2018.2889473

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *International Conference on Learning Representations*.

Misra, R. (2022). *News Category Dataset*. Https://Www.Huffpost.Com/. https://www.kaggle.com/datasets/rmisra/news-category-dataset

Rajaraman, A., & Ullman, J. D. (2012). Mining of Massive Datasets. In *Journal of Data Analysis and Information Processing* (Vol. 9781107015357). Cambridge University Press. https://doi.org/10.1017/CBO9781139058452

Raza, S., & Ding, C. (2021). News Recommender System: A Review of Recent Progress, Challenges, And Opportunities. *Artificial Intelligence Review*, *55*(1), 749–800. https://doi.org/10.1007/S10462-021-10043-X

Rong, Z., Yuan, L., & Yang, L. (2024). Enhanced Knowledge Graph Recommendation Algorithm Based on Multi-Level Contrastive Learning. *Scientific Reports*, *14*(1). https://doi.org/10.1038/S41598-024-74516-Z

Walker, J., Zhou, F., Baagyere, E. Y., Ahene, E., & Zhang, F. (2022). Implicit Optimal Variational Collaborative Filtering. *Complex and Intelligent Systems*, *8*(5), 4369–4384. https://doi.org/10.1007/S40747-022-00696-8

Wu, C., Huang, J., Wu, F., Huang, Y., An, M., & Xie, X. (2019). NPA: Neural News Recommendation with Personalized Attention. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2576–2584. https://doi.org/10.1145/3292500.3330665

Wu, S., Sun, F., Zhang, W., Xie, X., & Cui, B. (2023). Graph Neural Networks in Recommender Systems: A Survey. *ACM Computing Surveys*, *55*(5). https://doi.org/10.1145/3535101/ASSET/66843D8A-EBB9-44EC-8B10-E3C9B6900D75/ASSETS/IMAGES/LARGE/CSUR-2021-0234-F09.JPG

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2020). Deep Learning Based Recommender System: A Survey and New Perspectives. *ACM Computing Surveys*, *52*(1). https://doi.org/10.1145/3285029