Seminar Nasional Teknologi dan Inovasi (SITASI)
Vol 1, Nomor 2, Juni 2026, hal. 188-198
P-ISSN: XXXX-XXXX | E-ISSN: XXXX- XXXX
Penerbit Institut Bisnis dan Teknologi Indonesia

Penerapan Explainable Boosting Machine Pada Hasil Pencarian Best Matching 25 Sebagai Rekomendasi Kualitas Buku Dengan Penjelasan Fitur

Bagus Satrio Wicaksono[1*], Intan Yuniar Purbasari[2], Henni Endah Wahanani[3]

[1 2 3]*Universitas Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia*
[1] *bagussatrio232@gmail.com**, [2] intanyuniar.if@upnjatim.ac.id, [3] henniendah@upnjatim.ac.id

INFO ARTIKEL

Article history:
Received Juni 2026
Published Juli 2026

ABSTRAK

Pertumbuhan koleksi buku digital menyebabkan tantangan dalam menemukan buku bagi pembaca. Sistem pencarian dengan BM25 hanya menghasilkan buku yang relevan dengan *keywords*, mengabaikan kualitas buku. Model *re-ranking* umumnya masih bersifat *black-box* sehingga tidak dapat dipahami alasan model menghasilkan prediksi. Pendekatan *post-hoc* menghasilkan penjelasan yang bersifat aproksimasi yang tidak merepresentasikan logika model seutuhnya. Penelitian ini mengusulkan arsitektur *Two-stage ranking* dengan penerapan Explainable Boosting Machine (EBM) sebagai model prediksi dan re-ranker *glass-box* pada hasil pencarian buku BM25. EBM dilatih menggunakan 13,4 juta data interaksi rating Goodreads untuk mengestimasi kualitas berdasarkan metadata buku. Pada evaluasi ranking dengan 50 kueri uji menunjukkan model EBM dengan interaksi fitur mencapai skor NDCG@10=0.9373, melampaui BM25=0.9284 pada kueri genre yang merepresentasikan pencarian eksploratif. EBM terbukti menciptakan pergeseran peringkat dengan memprioritaskan buku berkualitas yang popular. Penjelasan global menunjukkan fitur popularitas memiliki rata-rata kontribusi terbesar pada skor prediksi, serta pada penjelasan lokal menampilkan kontribusi setiap fitur buku pada skor akhir prediksi. Uji ablasi mengkonfirmasi kejujuran pada penjelasan EBM, dimana penghapusan fitur kontribusi terbesar, menurunkan skor prediksi secara konsisten dengan nilai *shape function* fiturnya. Penerapan EBM terbukti meningkatkan kualitas ranking hasil pencarian BM25 berdasarkan tambahan sinyal kualitas buku secara umum, sekaligus menghasilkan penjelasan fitur yang dapat diaudit langsung.

Kata Kunci: Best Matching 25, explainable boosting machine, interpretasi, sistem pencarian buku, Explainable AI

ABSTRAK

The rapid growth of digital book collections creates information overload, complicating book discovery for readers. Traditional BM25 search systems retrieve keyword-relevant documents but completely ignore book quality signals. Furthermore, conventional machine learning re-ranking models act as black-boxes, obscuring their decision-making processes, while post-hoc explanation methods only provide approximations that do not represent the model's exact logic. This study proposes a two-stage ranking architecture integrating an Explainable Boosting Machine (EBM) as a glass-box re-ranker over BM25 initial retrievals. Trained on 13.4 million Goodreads interaction records, the EBM predicts general book quality based on metadata. Evaluation using 50 queries demonstrates that EBM with feature interactions achieves an NDCG@10 of 0.9373, outperforming the BM25 baseline (0.9284) specifically on exploratory genre queries by prioritizing popular, high-quality books. Global interpretability

analysis reveals that popularity contributes most significantly to prediction scores, whereas local explanations precisely map individual feature attributions. Ablation testing confirms the faithfulness of these explanations; removing dominant features consistently reduces prediction scores matching their exact shape function values. In conclusion, integrating EBM successfully enhances BM25 ranking quality through community-based signals while providing fully transparent and auditable feature explanations.

Keywords: Best Matching 25, book search system, Explainable AI, explainable boosting machine, interpretability

1. Pendahuluan

Perkembangan teknologi informasi berdampak pada peningkatan volume informasi yang disebut fenomena information overload, dimana pengguna kesulitan dalam mengakses informasi yang sesuai preferensi serta menilai kualitas kontennya (Bhajantri et al., 2024)(Perera et al., 2025). Pada perpustakaan digital dan platform buku daring seperti Goodreads, fenomena tersebut juga menyebabkan jumlah koleksi buku terus bertambah banyak, yang menghambat pembaca dalam memilih bacaan yang relevan dan berkualitas (Ahmad et al., 2025). Pendekatan sistem pencarian umum digunakan untuk mencari buku dari kata kunci, serta sistem rekomendasi untuk meningkatkan pengalaman pengguna melalui personalisasi konten (Duhan & Arunachalam, 2024). Salah satu tantangan dalam sistem pencarian adalah kemampuan sistem dalam memberikan hasil yang berkualitas dan dapat diandalkan.

Pendekatan BM25 sebagai fungsi dasar pada information retrieval untuk menghasilkan kandidat item yang cepat dalam pencarian di koleksi dokumen besar, masih tidak bisa mempertimbangkan kualitas itemnya (Preminger & Anker, 2025). Teknik pengurutan ulang (reranking) adalah tahap untuk mengubah urutan kandidat pada hasil ranking awal dengan perhitungan dari input sinyal lain yang tidak bisa ditangkap mesin pencari (Preminger & Anker, 2025). Pemanfaatan sinyal lain sebagai reranking, seperti boosting fitur jumlah salinan buku atau tahun terbit pada kandidat hasil pencarian buku BM25, dapat membantu meningkatkan kualitas item pada hasil pencariannya (Nivetha et al., n.d.). Namun, penggunaan pembobotan skor fitur secara manual pada hasil pencarian, masih bergantung pada nilai bobotnya, serta tidak mempelajari pola fitur seperti interaksi jumlah rating, tahun terbit, dan genre pada Goodreads.

Selain kebutuhan akan hasil pencarian yang relevan dan berkualitas, muncul kebutuhan pada kemampuan sistem untuk menjelaskan mengapa atau bagaimana suatu model dalam menghasilkan prediksi. Model black-box seperti gradient boost umumnya memiliki performa dan akurasi yang tinggi, namun proses internalnya cenderung kompleks dan sulit untuk diinterpretasi untuk pengguna atau praktisi yang menyebabkan penurunan kepercayaan dan reliabilitas sistem (Marconi et al., 2022). Pada temu kembali informasi, aspek penjelasan dibutuhkan dalam membuat sistem yang transparan dan dapat dipercaya, yang penting pada pembuatan dan audit sistem (Anand et al., n.d.). Penerapan eXplainable AI (XAI) pada sistem ML dapat meningkatkan transparansi sistem yang dapat membantu pengguna atau praktisi dalam memahami bagaimana model memberikan hasil prediksi yang bermanfaat dalam meningkatakan kepercayaan pada sistem serta mendapatkan wawasan pada perilaku model dalam mengolah data yang berguna untuk audit atau debugging modelnya (Quadir et al., 2024).

Pendekatan untuk menjelaskan model black-box umumnya menggunakan metode post-hoc, seperti metode LIME, menunjukkan penjelasan lokalnya pada hasil model cukup baik, namun terdapat keterbatasan pada penjelasan yang berbentuk aproksimasi (tebakan) terhadap hasil dari modelnya yang tidak merefleksikan logika model secara akurat (Ahmed & Alpkoçak, 2022). Keterbatasan tersebut mendorong penerapan model transparan atau model glass-box, yang secara internal prosesnya dapat diinterpretasi tanpa model tambahan seperti post-hoc, yang dapat memudahkan pengguna atau praktisi untuk memahami bagaimana model belajar datanya dan membuat keputusan (Rudin, 2019). Aspek faithfulness (kejujuran) pada penjelasan berguna untuk mengetahui apakah penjelasan yang dihasilkan merefleksikan logika model secara akurat (Pawlicki et al., 2024).

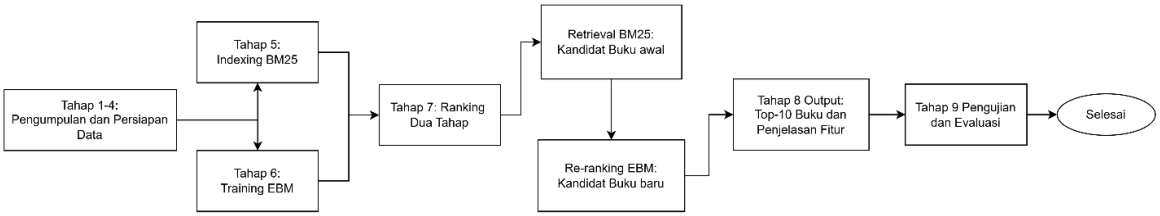
Model Explainable Boosting Machine (EBM) memiliki kelebihan pada proses internalnya yang berdasar pada Generalized Additive Model (GAM) dengan struktur aditif, yang dapat dijelaskan hasil perhitungannya dengan mengurai skor kontribusi dari setiap fitur pada suatu prediksi item. EBM dapat menunjukkan pengaruh dan kontribusi setiap fitur pada skor prediksi secara lokal untuk menjelaskan skor prediksi item individual dan global untuk menjelaskan pengaruh setiap fitur pada prediksi di keseluruhan data (Jenkins & Caruana, n.d.).

EBM sebagai model prediksi dimana skor akhirnya dapat digunakan sebagai pengurut ulang kandidat BM25 dengan mempelajari pola interaksi rating user umum terhadap fitur bukunya. Praktik dua tahap ini sejalan dengan pola Learning to Rank (LtR) yang digunakan pada praktik sistem mesin pencari, dimana model LtR

mempelajari sinyal fitur dari kandidat awal untuk meningkatkan urutan item yang lebih relevan (Preminger & Anker, 2025)(Nivetha et al., n.d.). Penerapan reranking pada retrieval BM25 dengan penambahan sinyal fitur metadata lain, serta intergrasi model black-box seperti XGBoost, menunjukkan bahwa BM25 masih digunakan sebagai dasar fungsi pencarian yang performanya baik, namun belum diterapkan model reranker yang bersifat glass-box (Preminger & Anker, 2025)(Kwon & Bak, 2025)(Nivetha et al., n.d.). Sementara EBM memiliki akurasi prediksi yang sering mendekati hingga setara dengan XGBoost, namun memiliki penjelasan bawaan yang berguna untuk transparansi sistem yang akurat (Schwartz et al., n.d.)(Hoang & Tran, 2023).

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan pendekatan two-stage ranking dengan menggabungkan BM25 sebagai metode retrieval dan EBM sebagai model reranking untuk meningkatkan kualitas hasil pencarian buku. Selain meningkatkan kualitas ranking, pendekatan ini juga menyediakan interpretasi terhadap kontribusi fitur, sehingga memungkinkan analisis yang lebih transparan terhadap hasil prediksi model.

2. Metode Penelitian



Gambar 1. Flowchart sistem pencarian
Sumber: Penulis, 2026

Pada alur sistem **Gambar 1** tersebut, pada tahapan awal dilakukan pengumpulan dataset yang dibutuhkan. Kemudian dilakukan persiapan data (prapemrosesan, gabung teks, dan feature engineering). Setelah itu, dilakukan pengindeksan teks pada koleksi buku untuk mencatat daftar token yang ada di seluruh data buku, dan mendaftarkan tiap buku yang memiliki token individual tersebut. Kemudian melatih EBM untuk memprediksi rating (kualitas) pada setiap buku berdasarkan riwayat pola rating dari kebanyakan pengguna pada data interaksi. Setelah data dan model siap, sistem akan melalui 2 tahap utama dalam proses utamanya, serta 1 tahap untuk penjelasan fitur pada prediksi.

- Tahap 1 (Retrieval): Tahapan awal adalah mencari buku yang relevan dengan kueri bahas inggris dan menghasilkan 100 kandidat buku. BM25 berperan sebagai *first-stage ranking* yang cepat untuk mengambil kandidat buku yang relevan secara teks terhadap kueri dari seluruh koleksi buku yang besar. BM25 hanya dirancang untuk relevansi leksikal, tanpa pertimbangan fitur lain.
- Tahap 2 (Re-ranking): Setelah mendapatkan kandidat ranking awal dari BM25, EBM berperan sebagai *second-stage re-ranker* yang mengurutkan ulang kandidat BM25 berdasarkan kualitas bukunya dengan prediksi rating setiap buku pada kandidat awal. EBM memprediksi rating buku menggunakan fitur pada metadata buku (popularitas, tahun terbit, genre). Proses ini bertujuan untuk menempatkan buku-buku dengan karakteristik tertentu yang berkualitas menurut pengguna umum Goodreads pada data historis, sehingga buku dengan fitur populer, tahun rilis dan genrenya biasanya disukai pengguna akan cenderung berada di posisi lebih atas.
- Penjelasan Fitur: Hasil prediksi EBM dijelaskan dengan menguraikan kontribusi fitur dari setiap buku. Aspek tersebut juga menjelaskan mengapa fitur buku mempengaruhi urutan kandidat BM25 berubah berdasarkan skor prediksi EBM-nya. Bentuk penjelasan yang disediakan adalah lokal dan global secara visual.

Selain itu, dilakukan pengujian pada 3 skenario dari hasil kandidat dari pencarian baseline BM25 saja, hasil reranking dengan EBM tanpa deteksi interaksi antar fitur dan tambahan interaksi antar fitur, dan penggabungan skor BM25 dan EBM untuk melihat hasil kandidat dengan pertimbangan dua sinyal relevansi dan kualitas buku.

Dalam memprediksi bagaimana sebuah buku akan dirating oleh pengguna umum pada platform buku, diperlukan data interaksi antara buku dengan pengguna dari data riwayat interaksi pembaca. Serta untuk proses pengindeksan, diperlukan fitur yang berisi teks yang mewakili setiap buku, seperti judul, deskripsi, dan genre. Dataset yang digunakan pada penelitian ini adalah dataset sekunder yang telah dikumpulkan oleh tim dari University of California San Diego yang bernama *Goodreads Book Graph Datasets* yang tersedia secara publik di internet. Dataset ini terdiri dari 3 tabel utama, yaitu metadata buku (2,3 juta), interaksi antara pengguna dan buku (228 juta), dan list genre setiap buku. Dataset tersebut dipilih karena memiliki metadata buku lengkap, dengan judul, deskripsi lengkap, daftar genre tiap buku, serta data interaksi yang cocok digunakan sebagai proxy kualitas buku dan memberikan sinyal fitur untuk pembelajaran model EBM dan data teks untuk diproses BM25.

Tabel 1. Dataset UCSD

Komponen	Atribut	Jumlah	Deskripsi
Metadata buku	book_id, title, description, average_rating, ratings_count,	2,360,655 buku	Digunakan untuk indexing BM25 dan fitur pelatihan EBM.

	publication_year, num_pages, language_code		
Interaksi pengguna dan buku	user_id, book_id, rating,	228,648,342 interaksi dengan rating 1-5	Riwayat rating pengguna umum terhadap buku. Digunakan sebagai target pelatihan EBM.
Genre Buku	book_id, genres: [], shelf_count	2,3 Juta baris genre per id buku	Daftar genre per buku. Sebagai fitur genre untuk pelatihan EBM.

Untuk mengurangi beban komputasi EBM dari dataset besar, serta mengurangi data interaksi yang kosong (sparse) yang berpengaruh pada pembelajaran pola oleh EBM. Dataset direduksi menjadi subset agar ukuran data dapat dikelola lebih ringan dan lebih padat. Filter interaksi menggunakan iterasi K-Core dengan nilai k = 5 untuk memastikan setiap interaksi antara user dan buku memiliki riwayat minimal 5 interaksi, sehingga dataset menjadi lebih padat dan sinyal interaksi user dan bukunya cukup untuk pembelajaran EBM yang lebih stabil dan mengurangi noise (interaksi minim). Kemudian dilakukan proses *cleaning* yang terdiri dari menghapus baris duplikat pada tabel buku (*book_id* sama), menghapus baris interaksi duplikat dengan *user_id* dan *book_id* yang sama.

Pada persiapan korpus BM25 dengan indexing, digunakan input token dan dokumen teks buku. Setiap buku merepresentasikan dokumen teks yang dibentuk dari penggabungan judul, deskripsi, dan genre buku sebagai satu kolom.

Fitur-fitur dari buku diekstrak untuk mempersiapkan matriks input (*X*) serta target y untuk pelatihan model EBM. Atribut yang digunakan untuk EBM adalah *ratings_count, publication_year,* dan genres, serta target interaksi rating pengguna umum terhadap buku. Fitur *ratings_count* dan *num_pages* ditransformasi untuk menjaga kestabilan pembelajaran model dan mencegah dominasi buku-buku yang sangat populer (jumlah rating tinggi) dalam proses training. Kemudian data genre diubah ke pembobotan berdasarkan diksi setiap buku, yang berisi pasangan {nama_genre: jumlah} yang dimana 'jumlah' mempresentasikan seberapa sering buku dikaitkan dengan genre tertentu berdasarkan shelves pengguna user Goodreads. Lalu, dilakukan penggabungan tabel *interactions* dan *books* dengan *left join* untuk memastikan setiap interaksi rating memiliki informasi lengkap fitur buku yang di-rating. Data gabungan tersebut akan di-*split* menjadi dua proporsi 80% untuk data latih dan 20% untuk data uji, menggunakan *Group Shuffle Split* dari fungsi *'GroupShuffleSplit'* oleh scikit-learn. Terakhir dilakukan normalisasi ke rentang 0-1 pada nilai fitur numerik metadata buku pada data latih.

Tahapan offline pada sistem ini terdiri dari proses indexing untuk BM25 dan pelatihan model Explainable Boosting Machine (EBM) menggunakan data yang telah melalui tahap prapemrosesan.

Pada tahap indexing, setiap buku direpresentasikan sebagai dokumen teks hasil penggabungan judul, deskripsi, dan genre. Teks kemudian ditokenisasi untuk membentuk kamus token serta menghitung statistik penting seperti panjang dokumen, frekuensi term (TF), dan document frequency (DF). Nilai inverse document frequency (IDF) dihitung untuk merepresentasikan tingkat kelangkaan token dalam korpus. Seluruh informasi ini disimpan dalam struktur inverted index untuk mempercepat proses pencarian tanpa perlu menghitung ulang statistik saat query dijalankan. Pada tahap pelatihan, model EBM digunakan untuk mempelajari hubungan antara fitur metadata buku dan rating pengguna sebagai estimasi kualitas buku. Fitur yang digunakan meliputi popularitas (jumlah rating), tahun terbit, serta representasi genre, dengan target berupa rating pengguna.

Model EBM dilatih menggunakan pendekatan *Generalized Additive Model* (GAM) berbasis boosting iteratif, di mana setiap fitur dipelajari secara bertahap untuk meminimalkan error prediksi.

$$\hat{y} = f_0 + \sum_j f_j(x_j) + \sum_{(j,k)\in\mathcal{I}} f_{jk}(x_j, x_k) \quad (1)$$

Proses ini dimulai dengan nilai dasar f_0 (*intercept*) yang berupa rata-rata rating user di data interaksi, kemudian diperbaiki melalui iterasi kontribusi fitur hingga konvergen, dimana dibuat fungsi $f_j(x_j)$ untuk mempelajari setiap fitur bukunya terhadap skor prediksi rating. Selain itu, dilakukan eksplorasi interaksi antar fitur $f_{jk}(x_j, x_k)$ untuk meningkatkan kemampuan model dalam menangkap hubungan kompleks antar atribut buku. Hasil pelatihan berupa fungsi kontribusi (shape function) untuk setiap fitur yang digunakan dalam proses prediksi. Model kemudian dievaluasi menggunakan metrik RMSE dan MAE untuk memastikan akurasi prediksi sebelum digunakan pada tahap reranking.

Sistem yang diusulkan menerapkan pendekatan two-stage ranking yang terdiri dari tahap retrieval menggunakan BM25 dan tahap reranking menggunakan Explainable Boosting Machine (EBM).

Pada tahap pertama, BM25 digunakan untuk mengambil kandidat buku yang relevan secara teks terhadap kueri pengguna. Setiap kueri diproses menjadi token dan dicocokkan dengan inverted index untuk menghitung skor relevansi berbasis frekuensi term dan kelangkaan kata. Hasil dari tahap ini berupa Top-K kandidat buku yang memiliki kecocokan leksikal tertinggi terhadap kueri. Pada tahap kedua, kandidat hasil BM25 diurutkan ulang menggunakan model EBM berdasarkan estimasi kualitas buku. Untuk setiap kandidat, diekstrak fitur metadata seperti popularitas, tahun terbit, dan representasi genre. Fitur-fitur tersebut digunakan sebagai input untuk

memprediksi rating buku menggunakan model EBM yang telah dilatih sebelumnya. EBM menghasilkan skor prediksi berdasarkan pendekatan aditif, di mana nilai akhir merupakan kombinasi antara nilai dasar (intercept) dan kontribusi masing-masing fitur. Skor ini kemudian digunakan untuk mengurutkan ulang kandidat buku secara menurun, sehingga buku dengan estimasi kualitas lebih tinggi ditempatkan pada posisi teratas. Pendekatan ini memungkinkan sistem untuk menyeimbangkan relevansi teks dari BM25 dengan kualitas buku berbasis metadata. Selain itu, EBM juga menyediakan interpretasi kontribusi fitur terhadap skor prediksi, sehingga perubahan peringkat dapat dijelaskan secara transparan.

Penjelasan fitur pada EBM dibagi menjadi dua bagian, yaitu penjelasan lokal dan global. Penjelasan lokal digunakan untuk memahami alasan di balik skor prediksi pada setiap kandidat buku. Untuk setiap buku, skor prediksi diuraikan menjadi kontribusi dari masing-masing fitur, termasuk kontribusi individu dan interaksi antar fitur jika tersedia. Nilai kontribusi dapat bersifat positif atau negatif, yang masing-masing menunjukkan peningkatan atau penurunan terhadap skor prediksi akhir. Penjelasan ini memungkinkan analisis perubahan peringkat pada hasil reranking, dengan mengidentifikasi fitur mana yang paling berpengaruh terhadap posisi buku dalam ranking.

Penjelasan global digunakan untuk menganalisis perilaku model secara keseluruhan pada seluruh dataset. Analisis ini mencakup identifikasi fitur yang paling berpengaruh terhadap prediksi (feature importance) serta hubungan antara nilai fitur dan kontribusinya terhadap skor prediksi melalui fungsi bentuk (*shape function*). Shape function menggambarkan pola hubungan antara nilai fitur dan dampaknya terhadap prediksi, sehingga dapat digunakan untuk memahami kecenderungan model dalam menilai kualitas buku. Dengan demikian, penjelasan global memberikan gambaran umum mengenai faktor dominan yang digunakan model dalam proses prediksi.

Pengujian dilakukan untuk mengevaluasi performa model EBM dalam dua aspek, yaitu akurasi prediksi dan kualitas ranking pada sistem two-stage ranking. Eksperimen dilakukan dengan membandingkan dua konfigurasi model EBM, yaitu tanpa interaksi fitur dan dengan interaksi terbatas. Parameter interaksi mengontrol kompleksitas model dalam menangkap hubungan antar fitur. Model dilatih menggunakan data latih dan divalidasi untuk memperoleh konfigurasi terbaik berdasarkan metrik akurasi RMSE dan MAE. Evaluasi dilakukan dengan beberapa skenario, meliputi:

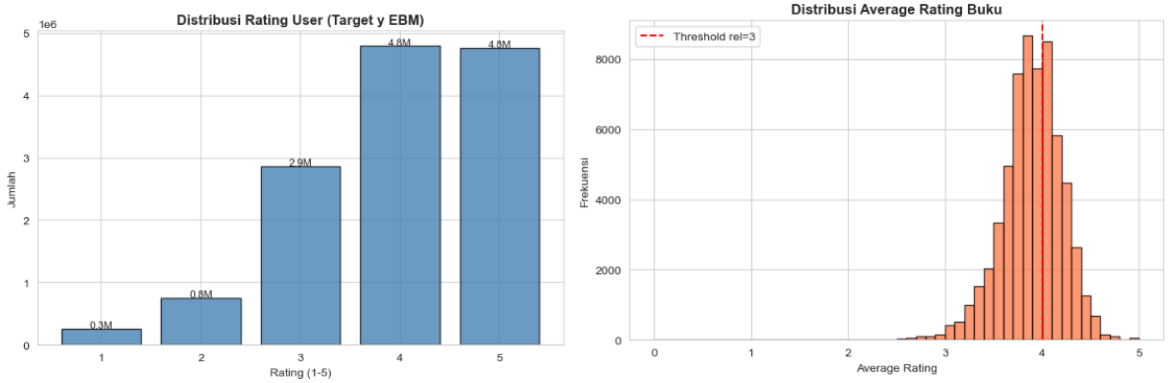
a. BM25 (baseline) sebagai tahap retrieval awal dan dievaluasi dengan metrik Precision@100
b. BM25 + EBM dengan model interaksi 0 dan 5 untuk reranking kandidat

Pengujian menggunakan 50 kueri uji yang merepresentasikan berbagai tipe pencarian, seperti berbasis genre, judul, dan deskripsi.

Akurasi model EBM diukur menggunakan Root Mean Squared Error (RMSE) dan Mean Absolute Error (MAE) pada data uji. Kedua metrik digunakan untuk mengevaluasi kemampuan model dalam memprediksi rating berdasarkan fitur metadata buku. Kualitas ranking dievaluasi menggunakan Precision@10 dan NDCG@10. Ground truth relevansi ditentukan berdasarkan nilai *average rating* buku yang dikonversi ke dalam skala relevansi bertingkat. Evaluasi ini bertujuan untuk mengukur kemampuan sistem dalam menempatkan buku berkualitas tinggi pada posisi teratas hasil pencarian. Selain performa, dilakukan analisis terhadap penjelasan model EBM pada tingkat lokal dan global. Analisis ini bertujuan untuk memahami kontribusi fitur terhadap prediksi serta mengevaluasi konsistensi model melalui perbandingan konfigurasi dan uji ablasi.

3. Hasil dan Pembahasan

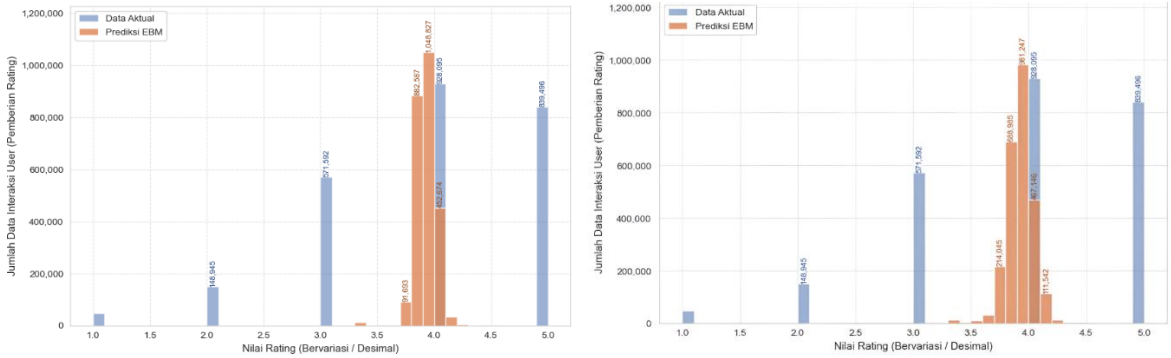
Proses reduksi data dengan *5-core filtering* menunjukkan bahwa adanya bias positif pada dataset, yang ditunjukkan di distribusi rating user pada data interaksi dan data rata-rata rating buku. Hal ini menunjukkan banyak user merating buku yang mereka suka, sehingga mayoritas rating user dan nilai *avg rating* bukunya berada pada rentang 4 sampai 5 bintang.



Gambar 2. Distribusi rating dataset pasca filtering
Sumber: Penulis, 2026

Model EBM merefleksikan tren populasi rating tersebut melalui nilai dasar *intercept* yang konvergen di nilai rata-rata 3.97. Oleh sebab itu, fungsi kontribusi aditif dari EBM (terutama fitur seperti tahun, jumlah halaman,

genre) cenderung berfungsi sebagai penggeser halus nilai akhir prediksi, sehingga penyebaran nilai prediksi berkumpul di rentang nilai 3.5 hingga 4.5. Sekitar 23.665 dari 61.893 buku pada korpus, sebesar 38% sudah memenuhi ambang relevansi (*avg_rating* ≥ 4.0). Pada hasil pencarian buku, hal ini menyebabkan metrik NDCG@10 BM25 cukup tinggi, karena kemungkinan BM25 mengambil buku yang berkualitas dari korpus ini sudah tinggi. Serta hasil NDCG@10 pada semua skenario terlihat tinggi juga. Selain itu, penggunaan fitur *avg_rating* mentah sebagai proxy relevansi memiliki keterbatasan mendasar akibat bias popularitas. Mayoritas buku dalam korpus sudah memiliki gain relevansi ≥ level 2 (nilai *avg_rating* ≥ 3.5), sehingga retrieval acak pun akan menghasilkan NDCG yang sudah cukup tinggi. Sesuai dengan distribusi prediksi rating EBM (**Gambar 3**) yang mayoritas berada di rentang rating 4 dan 5.



Gambar 3. Distribusi Prediksi EBM
Sumber: Penulis, 2026

Distribusi prediksi rating model EBM int0 berada pada rentang 3.5 hingga 4.5 seperti pada **Gambar 3 Kiri** (batang oranye), dimana komunitas Goodreads cenderung memberikan rating tinggi terhadap sebuah buku yang mereka sukai, sehingga nilai Intercept EBM menarik sebagian besar skor prediksi buku ke arah nilai rating 4. Hal ini terjadi karena nilai Intercept pada pelatihan EBM didapat dari rata-rata global rating komunitas. Serta fitur metadata buku hanya merepresentasikan karakteristik umum buku, sehingga tidak cukup dalam menjelaskan keberagaman selera personal user individual. Pada distribusi model int5 (**Gambar 3 Kanan**) juga memiliki rentang yang sama dengan prediksi int0, hanya saja terdapat sedikit perbedaan jumlah frekuensi dimana model int5 memiliki skor prediksi yang lebih menyebar dibanding int0 pada rentang 3.5-4.3.

Kueri uji yang diekstrak dari data UCSD berdasarkan pasangan genre yang paling sering muncul pada buku, menghasilkan 20 kueri genre yang didominasi oleh kombinasi genre populer seperti *historical fiction*, *biography*, *thriller*, dan *mystery*, yang merupakan genre paling umum pada korpus UCSD Goodreads. Pendekatan ini menghasilkan beberapa genre populer tersebut muncul lebih sering dalam kueri uji. Keterbatasan ini diakui sebagai area untuk penelitian lanjutan dengan korpus yang lebih beragam secara genre. Adapun kueri deskripsi menunjukkan variasi yang lebih luas karena proses ekstraksi kata menggunakan seleksi kata-kata naratif seperti *murder, quest, magic*, dst, yang mendorong keberagaman konteks cerita pada buku.

Pada skenario a, fungsi pencarian BM25 menunjukkan performa yang kuat sebagai baseline dalam menghasilkan kandidat yang relevan secara teks. Pengujian dengan 50 kueri uji dengan metrik Precision@100 menunjukkan kandidat pencarian BM25 memiliki kualitas rata-rata hanya sebesar 34% dari 100 kandidat, sehingga dibutuhkan fungsi reranking kualitas. Pada kueri uji judul dengan judul buku yang populer, Precision@100 BM25 memiliki skor tinggi karena buku-buku tersebut sudah memiliki nilai rating yang sudah cukup tinggi dari korpusnya.

Validasi performa prediksi EBM menunjukkan kemampuan yang memadai dalam menggeneralisasi ekspektasi *rating* komunitas. Evaluasi membandingkan pengaturan EBM tanpa deteksi interaksi (*int0*) dan pendeteksian 5 interaksi antar fitur (*int5*). Hasil *int0* menghasilkan nilai akurasi lebih presisi (RMSE: 0.9782) dibandingkan model *int5* (RMSE: 0.9820). Hal ini mengisyaratkan bahwa dalam menaksir ekspektasi *rating* murni secara regresi linear, pendekatan fitur tunggal masih lebih reliabel. Pengurutan ulang oleh EBM pada 100 buku kandidat BM25 menghasilkan skor evaluasi ranking yang bervariasi. Pada tahap 7, skenario b = BM25 + EBM, dikarenakan skala penilaian relevansinya serta bentuk kueri uji yang diekstrak dari beberapa buku populer. Hal tersebut menyebabkan hasil skor NDCG@10 pada hasil kueri uji jenis judul, ranking kandidat BM25 memiliki skor lebih besar, sedangkan EBM berada sedikit di bawahnya.

Tabel 2. NDCG@10 skenario b

Jenis	BM25	Interactions 0 (int0)	Interactions 5 (int5)
Deskripsi	0.9166	0.9084	**0.9360**
Genre	0.9088	0.9282	**0.9336**
Judul	**0.9662**	0.9332	0.9435

Skor NDCG model int5 pada kueri genre menunjukkan lebih unggul 0.02 poin dibandingkan skor ranking BM25. Model ini menujukkan bahwa pola interaksi antar fitur membantu untuk mengestimasi kualitas buku dengan lebih baik dibanding int0 dan relevansi teks saja. Meskipun pada kueri judul, skor BM25 masih yang tertinggi, hal ini terjadi seperti yang dibahas di hasil skenario a.

Sebagai contoh pada model EBM int5 dengan NDCG tertinggi, kandidat buku hasil pencarian dengan kueri jenis genre '*fantasy paranormal magic*' diprediksi rating bukunya dengan mencocokkan nilai fitur kandidat buku dengan fungsi bentuk dari hasil pelatihan EBM.

Peringkat	Geser Peringkat	Judul	Genre	Jumlah Rating	Skor BM25	Skor EBM	Avg rating
1	—	The Magic Books:...	fantasy, paranorm...	83	9.93	—	4.11
2	—	The Time of the ...	fantasy, paranorm...	43	9.07	—	4.01
3	—	Magic by the Lak...	fantasy, paranorm...	113	8.84	—	4.07
4	—	Magic Knight Ray...	fantasy, paranorm...	421	8.8	—	4.2
5	—	The Magic Goes ...	fantasy, paranorm...	295	8.75	—	3.88
6	—	Rifts World Book ...	fiction fantasy, ...	120	8.74	—	3.46
7	—	Magic Kingdom F...	fantasy, paranorm...	119	8.69	—	3.87
8	—	Mountain Magic (...	romance fantasy, ...	10	8.62	—	4.2
9	—	The Magic Stone	children young-a...	6	8.56	—	3.5
10	—	The magic of Sha...	non-fiction fanta...	45	8.43	—	3.36

Gambar 4. Peringkat pencarian BM25
Sumber: Penulis, 2026

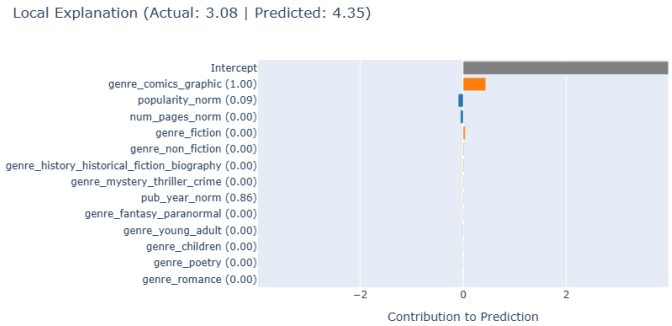
Dapat dilihat buku pada peringkat 1 (**Gambar 3**) memiliki jumlah rating rendah sebanyak 83 dengan nilai rating 4.11. Kemudian pada **Gambar 4** dapat dilihat perubahan ranking beberapa buku ke peringkat top 10.

Peringkat	Geser Peringkat	Judul	Genre	Jumlah Rating	Skor BM25	Skor EBM	Avg rating
1	↑32	The Magic School...	children fiction ...	1952	8.13	4.11	4.38
2	↑51	Split Infinity (App...	fantasy, paranorm...	13491	7.94	4.11	3.94
3	↑54	The Books of Ma...	comics, graphic f...	43	7.91	4.1	3.97
4	↑84	The Magic Finger	children fantasy,...	19574	7.63	4.1	3.66
5	↑90	Waldo and Magic,...	fiction fantasy, ...	6968	7.59	4.1	3.89
6	↑19	Lost in the Solar S...	children fiction ...	8800	8.18	4.08	4.28
7	↑19	Through The Mag...	fantasy, paranorm...	87	8.18	4.05	3.98
8	↑39	Bloodstorm	fantasy, paranorm...	553	7.98	4.05	4.15
9	↓3	Rifts World Book ...	fiction fantasy, ...	120	8.74	4.04	3.46
10	↑45	Evie The Mist Fair...	children fantasy,...	1113	7.93	4.03	3.79

Gambar 5. Peringkat pencarian reranking EBM
Sumber: Penulis, 2026

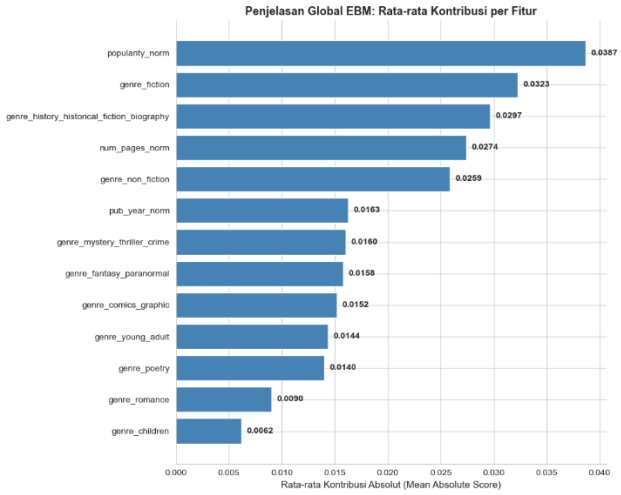
Perubahan ranking pada **Gambar 4** ditunjukkan pada buku peringkat 1 dengan jumlah rating 1952 dengan rating 4.38, mengalami kenaikan 32 tingkat dari yang sebelumnya peringkat 33 pada kandidat BM25 (ditunjukkan tanda panah buku naik). Serta buku pada kandidat BM25 yang berada di peringkat ke 6 (*Rifts World Book*), diurutkan ulang menjadi turun ke peringkat 9. Mayoritas kandidat buku pada peringkat top-10 jumlah ratingnya meningkat. Hal tersebut menunjukkan prediksi EBM menaikkan buku yang cenderung populer (jumlah rating lebih banyak). Sehingga kandidat bukunya lebih menyesuaikan kualitas umum. EBM dapat menurunkan peringkat buku yang memiliki rating tinggi namun popularitas rendah dan menaikkan buku dengan prediksi berkualitas menurut komunitas.

Penjelasan fitur yang dihasilkan EBM adalah penjelasan lokal dan global. Penjelasan lokal pada setiap kandidat buku menunjukkan fitur mana yang mempengaruhi skor prediksi pada setiap buku kandidat EBM. Hal ini menjelaskan penyusunan skor prediksi yang mempengaruhi perubahan peringkat buku dari kandidat BM25 menjadi kandidat EBM.



Gambar 6. Kontribusi fitur sebuah buku secara lokal
Sumber: Penulis, 2026

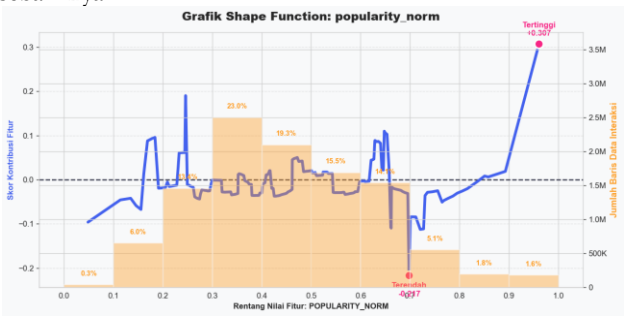
Pada penjelasan fitur, ditunjukkan bahwa fitur yang dipelajari EBM memiliki kontribusi yang kecil. Kontribusi kecil tersebut disebabkan karena metadata yang digunakan pada pelatihan masih terbatas pada variabel agregat (numerik dan genre) dari fitur utama buku, sementara target *y* per baris interaksi usernya memiliki banyak variasi (13 juta baris interaksi). Sehingga input metadata fitur buku tersebut kurang bisa menjelaskan variasi rating setiap baris interaksi user, yang memiliki preferensi rating beragam pada tiap usernya. Nilai aktual fitur buku tidak berubah antara user satu dan user lain untuk buku yang sama, maka dari itu kontribusi fiturnya hanya bisa mengubah nilai intercept secara kecil pada skor prediksi total.



Gambar 7. Rata-rata kontribusi fitur global buku
Sumber: Penulis, 2026

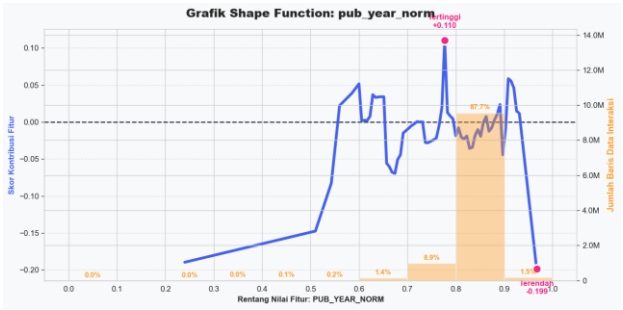
Grafik penjelasan global (**Gambar 7**) menunjukkan fitur utama yang paling berkontribusi secara global dari hasil training EBM adalah fitur popularitas dengan nilai rata-rata kontribusi (Mean Absolute Score) sekitar 0.038 terhadap prediksi model di seluruh data, diikuti oleh genre *fiction* sebesar 0.032. Setiap fitur EBM memiliki fungsi kontribusi aditif terhadap skor prediksi, sehingga nilai ini merepresentasikan seberapa besar pengaruh rata-rata fitur dalam mengubah nilai prediksi dari Intercept. Prediksi akhir yang dihasilkan adalah ekspektasi kualitas dari komunitas secara umum, bukan preferensi personal. Oleh karena itu, nilai utama EBM dalam sistem ini adalah kemampuannya menilai buku yang dianggap berkualitas berdasarkan rating komunitas.

Analisa skor kontribusi setiap fitur buku yang dipelajari dalam bentuk fungsi bentuk (*shape function*) menghasilkan grafik yang beragam. Grafik fitur popularitas (**Gambar 8**) menunjukkan bahwa buku dengan nilai popularitas tinggi (0.9 – 1.0) memiliki skor kontribusi terbesar dengan skor mencapai 0.3, namun rentang nilai fitur ini hanya berdasarkan 1.6% data interaksi, sehingga kurang merepresentasikan mayoritas pengguna pada hasil prediksi. Pada batang oranye, populasi baris interaksi pada nilai popularitas 0.2-0.7 menunjukkan skor kontribusi yang dapat dipercaya karena pembelajaran model didasarkan data jutaan baris interaksi dari komunitas Goodreads. Skor kontribusi (garis biru) pada nilai rentang tersebut cenderung naik-turun kecil karena EBM menangkap pola interaksi secara detail yang didukung jutaan baris interaksi. Sedangkan pada rentang <0.2 dan >0.8 terdapat beberapa lonjakan besar karena dipelajari dari jumlah data yang sedikit, sehingga EBM cenderung overfitting karena jika terdapat buku populer yang kebetulan mendapat rating sangat tinggi, EBM akan menarik skor kontribusi yang melonjak tinggi, demikian sebaliknya.



Gambar 8. Grafik Fungsi fitur popularitas
Sumber: Penulis, 2026

Konsistensi penjelasan dibuktikan pada penjelasan lokal (**Gambar 6**), di mana buku dengan nilai fitur popularitas yang menurut tren grafik fungsi (**Gambar 8**) bahwa nilai tersebut berkontribusi negatif, mendapat kontribusi negatif juga pada skor prediksi akhir.



Gambar 9. Grafik Fungsi fitur tahun publikasi
Sumber: Penulis, 2026

Pada **Gambar 9**, rentang nilai fitur 0.5 s.d 0.9 memiliki pola kontribusi lebih stabil karena buku di situ lebih banyak muncul di data (batang oranye). Pada nilai fitur 0.0-0.49 s.d 0.9-1 (menunjukkan tahun buku sangat lama dan sangat baru), kontribusinya cenderung negatif karena contoh buku pada dataset lebih sedikit sehingga model kurang bisa menangkap pola interaksi terhadap kontribusi skornya pada prediksi. Grafik garis biru menunjukkan fitur tersebut menunjukkan skor kontribusi negatif sebesar -0.2 pada nilai 0.0. Hal ini disebabkan populasi data interaksi pada rentang nilai fitur buku tersebut 0%, namun pada nilai *publication_year* sebesar 0.48 baru kontribusi mulai meningkat ke positif, diikuti jumlah populasi yang meningkat. Pada kepadatan data di rentang fitur 0.8-0.9, terdapat 9.5 juta baris interaksi, menunjukkan rating komunitas cenderung banyak pada buku dengan tahun muda.

Mayoritas fitur genre memilik kontribusi minimum terhadap skor prediksi. Dalam menentukan kualitas buku secara global, kontribusinya lebih kecil dibandingkan dengan fitur popularitas. Fitur fitur yang dipelajari EBM menunjukkan fungsi grafik yang naik turun karena model memetakan hubungan kompleks antar fitur buku dengan target interaksi rating user Goodreads. Pada uji ablasi, diambil kandidat pencarian '*fantasy adventure magic*' dengan buku peringkat 1 yaitu '*Legend of Lemnear 3*' dengan skor EBM 4.34. Pada buku tersebut, setiap fitur buku dinolkan satu per satu untuk mengukur seberapa besar penurunan skor yang terjadi.

Tabel 3. Hasil uji ablasi fitur buku

Buku	Skor EBM	Fitur yang dihapus	Skor Ablasi	Selisih
Legend of Lemnear 3	4.3491	genre_comics_graphic	3.9079	-0.4411
		pub_year_norm	4.1519	-0.1970

Fitur yang paling berpengaruh adalah *genre_comics_graphic,* buku ini memiliki nilai genre comics = 1 (buku manga/komik), sehingga menghapus informasi genre tersebut menurunkan skor prediksi sebesar -0.4411 poin menjadi 3.90 rating dari yang awalnya 4.34. Hasil ini membuktikan bahwa penjelasan EBM bersifat faithful (jujur), dimana fitur yang menurut shape function berperan besar (*genre_comics_graphic*) memang mengubah prediksi secara signifikan saat dihapus. Uji ablasi membuktikan penurunan skor prediksi jika fitur dengan kontribusi terbesar dihapus, yang membuktikan bahwa penjelasan fitur oleh EBM itu jujur (faithful) dan sesuai pembelajaran model.

Skor prediksi yang dibangun dengan nilai Intercept merepresentasikan estimasi awal model terhadap rating buku secara umum. Nilai ini mendekati rata-rata rating user di data interaksi, sehingga nilai tersebut adalah rata rata ekspektasi komunitas terhadap suatu buku. Kontribusi setiap fitur buku ditambahkan secara aditif terhadap Intercept.

Namun dengan pembelajaran 3 fitur numerik dan 10 kolom genre pada distribusi data interaksi yang sempit (**Gambar 2**), kontribusi fitur terhadap pergeseran nilai Intercept cenderung kecil. Hal ini disebabkan variasi pola rating user terhadap buku (target y) tinggi, sehingga sinyal fitur bukunya terbatas dalam menjelaskan perbedaan selera setiap user. Hasil skor prediksi tersebut sesuai dengan yang disebutkan pada hasil distribusi prediksi EBM, dimana EBM memprediksi buku berdasarkan kualitas secara umum (nilai dasar Intercept merupakan representasi rata-rata rating user komunitas). Pola pembelajaran fitur ini mengkonfirmasi bahwa EBM berhasil mempelajari sinyal kualitas komunitas dari data historis secara otomatis, sinyal yang tidak dapat ditangkap oleh fungsi pencarian BM25 yang hanya mempertimbangkan kecocokan teks.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa BM25 sebagai baseline dalam menghasilkan kandidat yang relevan secara teks sudah cukup baik, sementara EBM berhasil memberikan kontribusi sinyal tambahan pada peringkat hasil pencarian dalam bentuk estimasi kualitas berbasis komunitas. Namun, peningkatan performa bersifat terbatas dan bergantung pada jenis kueri, serta dipengaruhi oleh ketersediaan fitur pada dataset dan definisi ground truth yang digunakan pada metrik evaluasi. Penjelasan fitur pada model *glass-box* tersebut dapat membantu dalam audit model dan dataset, seperti temuan kontribusi fitur yang kecil pada skor prediksi, fitur-fitur yang paling berpengaruh dan korelasi interaksi antar fitur yang dapat meningkatkan atau menurunkan skor prediksi. Penelitian penerapan EBM ini berhasil sebagai reranker kandidat buku berdasarkan kualitas rating secara umum, serta penjelasan prediksi yang langsung dari modelnya.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa sistem pencarian buku berbasis two-stage ranking yang menggabungkan BM25 dan Explainable Boosting Machine (EBM) mampu mengintegrasikan dua aspek utama, yaitu relevansi teks dan estimasi kualitas berbasis rating komunitas. Berdasarkan eksperimen menggunakan 61.893 buku dari dataset UCSD Goodreads dan 50 kueri uji, diperoleh tiga kesimpulan utama yang menjawab rumusan masalah penelitian.

Hasil prediksi rating EBM pada hasil pencarian buku BM25 berhasil digunakan untuk mengurutkan ulang kandidat berdasarkan kualitas buku secara umum dengan 50 kueri uji (genre, judul, deskripsi). Pada skenario **b** dengan kueri eksploratif (genre), Model EBM int5 (skenario **b2**) menghasilkan NDCG@10 terbesar secara keseluruhan sebesar 0.9373, melampaui BM25 sebesar 0.9284. Perubahan ranking cenderung menunjukkan buku dengan jumlah rating lebih banyak akan mendapat prediksi rating lebih tinggi, yang menunjukkan skor prediksi dipengaruhi fitur *ratings count* sebagai estimasi buku berkualitas berdasarkan komunitas. Penjelasan fitur EBM menunjukkan kontribusi yang beragam. Nilai intercept yang dipelajari EBM sebesar 3.9789, yang mencerminkan rata-rata rating user komunitas Goodreads. Penjelasan global menunjukkan kontribusi fitur buku dengan kontribusi terbesar adalah popularitas dengan rata-rata kontribusi 0.03, diikuti oleh fitur *genre fiction*. Kontribusi fitur cenderung kecil karena variasi metadata yang kurang kuat dibandingkan jumlah data interaksi user, sehingga dengan fitur-fitur yang nilainya identik namun target *y*-nya sangat beragam pada setiap buku, skor prediksi EBM sebagian besar berada di nilai Intercept saja sebagai ekspektasi rata-rata. Hal ini juga sesuai dengan tujuan EBM pada penelitian ini sebagai reranker berdasarkan kualitas umum.

Pada penjelasan global, fitur popularitas berkontribusi paling banyak yang menunjukkan bahwa pada dataset ini popularitas buku sangat mempengaruhi kualitas buku. Uji ablasi membuktikan bahwa menghilangkan fitur buku dengan kontribusi terbesar (fitur popularitas) pada skor prediksi menyebabkan penurunan skor prediksi. Hal tersebut membuktikan penjelasan EBM yang bersifat jujur (*faithful*) terhadap model itu sendiri, sehingga model dapat diaudit secara langsung. Penelitian selanjutnya dapat memperbaiki beberapa aspek, meliputi perbaikan label relevansi yang lebih dinamis (tidak mentah pakai 1 fitur saja), peningkatan dan perbaikan ekstraksi kueri uji. Serta tambahan variasi fitur metadata buku yang lebih luas, seperti aspek personalisasi histori bacaan user, data preferensi user individual juga bisa ditambahkan pada target pelatihan model EBM sehingga EBM bisa menjadi rekomendasi kualitas yang lebih mempertimbangkan personal.

Daftar Pustaka

Ahmad, M., Imran, M., Iqbal, M., & Khan, A. M. (2025). Digital Information Overload and Its Impact on Library Users: Assessing User Experience and Service Adaptations in Female Higher Education Students. *Research Journal for Social Affairs*, *3*(2), 89–96. https://doi.org/10.71317/rjsa.003.02.0098

Ahmed, N., & Alpkoçak, A. (2022). A quantitative evaluation of explainable AI methods using the depth of decision tree. *Turkish Journal of Electrical Engineering and Computer Sciences*, *30*(6), 2054–2072. https://doi.org/10.55730/1300-0632.3924

Anand, A., Lyu, L., Idahl, M., Wang, Y., Wallat, J., & Zhang, Z. (n.d.). *Explainable Information Retrieval : A Survey*. *1*(1).

Bhajantri, A., Nagesh, K., Goudar, R. H., Dhananjaya, G. M., Kaliwal, R. B., Rathod, V., Kulkarni, A., & Govindaraja, K. (2024). Personalized Book Recommendations: A Hybrid Approach Leveraging Collaborative Filtering, Association Rule Mining, and Content-Based Filtering. *EAI Endorsed Transactions on Internet of Things*, *10*, 1–6. https://doi.org/10.4108/eetiot.6996

Duhan, A., & Arunachalam, N. (2024). Book Recommendation System using Machine Learning. *AIP Conference Proceedings*, *3075*(1). https://doi.org/10.1063/5.0217051

Hoang, T., & Tran, A. (2023). *Final Project Explainable Artificial Intelligence in Job Recommendation Systems*.

Jenkins, S., & Caruana, R. (n.d.). *InterpretML : A Unified Framework for Machine Learning*. 1–8.

Kwon, Y., & Bak, G. (2025). *A Hybrid Recommendation System Based on Similar-Price Content in a Large-Scale E-Commerce Environment*. 1–30.

Marconi, L., Matamoros, R. A. A., & Epifania, F. (2022). Discovering the Unknown Suggestion: a Short Review on Explainability for Recommender Systems. *CEUR Workshop Proceedings*, *3463*.

Nivetha, K., Karthik, S., Yogieswaran, V., & Indumathy, M. (n.d.). *IntelliRec : Frequently Asked Questions Hybrid Recommendation System For Personalized Placement Preparation* (Issue Icisd 2025). Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-866-0

Pawlicki, M., Pawlicka, A., Uccello, F., & Szelest, S. (2024). Neurocomputing Evaluating the necessity of the multiple metrics for assessing explainable AI : A critical examination. *Neurocomputing*, *602*(June), 128282. https://doi.org/10.1016/j.neucom.2024.128282

Perera, L. U. S. L., Chathuranga, H. M. S., & Thilakarathna, T. (2025). Personalized Book Recommendation Engine with Emotional Understanding and Contextual Relevance Based on Hybrid Filtering - A Novel Method Providing Recommendation Explanation. *CSECS 2025 - Proceedings of 2025 7th International Conference on Software Engineering and Computer Science*, 1–6. https://doi.org/10.1109/CSECS64665.2025.11009793

Preminger, M., & Anker, A. (2025). Reranking Hits in Public Library Catalog Search with Popularity and Freshness Boosting Reranking Hits in Public Library Catalog Search with Popularity and Freshness Boosting. *Cataloging & Classification Quarterly*, *0*(0), 1–17. https://doi.org/10.1080/01639374.2025.2562079

Quadir, H. M. S., Rahman, A., Mahmood, M., Tasnim, N., Islam, T., Rashed, G., & Das, D. (2024). The Role of XAI in User-Centric Recommender Systems Using Collaborative and Content-Based Approaches. *2024 27th International Conference on Computer and Information Technology (ICCIT)*, *December*, 3016–3021. https://doi.org/10.1109/ICCIT64611.2024.11022562

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(May). https://doi.org/10.1038/s42256-019-0048-x

Schwartz, S., Wang, Q., & Fang, F. (n.d.). *Enhancing ML Interpretability for Credit Scoring*.