

**PENERAPAN EXPLAINABLE BOOSTING MACHINE PADA HASIL
PENCARIAN BEST MATCHING 25 SEBAGAI REKOMENDASI KUALITAS
BUKU DENGAN PENJELASAN FITUR**

Bagus Satrio Wicaksono¹, Intan Yuniar Purbasari², Henni Endah Wahanani³

^{1,2,3}Informatika, Ilmu Komputer, Universitas Pembangunan Nasional Veteran
Jawa Timur

¹22081010252@student.upnjatim.ac.id, ²intanyuniar.if@upnjatim.ac.id*,

[³henniendah@upnjatim.ac.id](mailto:henniendah@upnjatim.ac.id)

ABSTRACT

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

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 populer. 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, goodreads, interpretasi, sistem pencarian buku

A. 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 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). Pemanfaatan sinyal lain sebagai *reranking* pada hasil pencarian, 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 buku 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 meningkatkan

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 lokal yang cukup baik pada hasil keputusan model, 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, yang dapat memudahkan pengguna atau praktisi untuk memahami bagaimana model belajar datanya dan membuat keputusan (Rudin, 2019).

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.).

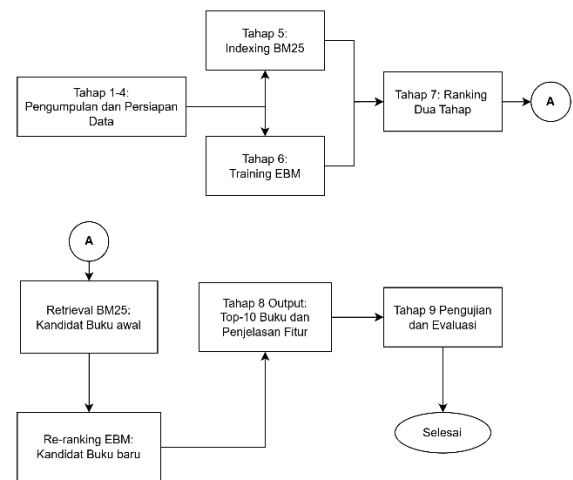
EBM sebagai model prediksi dimana skor akhirnya dapat digunakan sebagai pengurut ulang kandidat BM25 dengan mempelajari pola interaksi rating user umum terhadap fitur bukunya. Praktik dua tahap ini sejalan dengan pola *Learning to Rank* (LtR) yang digunakan pada praktik sistem mesin pencari, dimana model LtR mempelajari sinyal fitur dari kandidat awal untuk meningkatkan urutan item yang lebih relevan (Preminger & Anker, 2025)(Nivetha et al., n.d.).

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.

B. Metode Penelitian

Pada alur sistem **Gambar 1** tersebut, pada tahapan awal dilakukan pengumpulan dataset yang dibutuhkan.



Gambar 1. Alur sistem penelitian

Kemudian dilakukan persiapan data (prapemrosesan, gabung teks, dan feature engineering). Setelah itu, melakukan 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 umum) 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. 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. 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, dan 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 genre yang biasanya disukai pengguna akan cenderung berada di posisi lebih atas.

Hasil prediksi EBM kemudian 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.

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.

Tabel 1. Dataset UCSD

Product	Atribut	Jumlah
Metadata Buku	title, description, avg_rating, ratings_count, pub_year, num_pages	2,360,655 buku
Interaksi User-Books	user_id, book_id, rating	228,648,342
Genre	Book_id, genres: []	2.3 Juta

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).

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. Setelah teks digabung, akan dilakukan prapemrosesan tokenisasi terdiri dari; casefolding, stopwords, dan pemisahan setiap kata. 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. 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 seperti panjang dokumen, *term frequency* (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 rating setiap pengguna. Model EBM dilatih menggunakan pendekatan *Generalized Additive Model* (GAM) berbasis *boosting* iteratif, di mana setiap fitur dipelajari secara bertahap untuk meminimalkan error skor prediksi.

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

Di mana:

f_0 = intercept, yaitu nilai dasar target yang akan diprediksi

$f_j(x_j)$ = fungsi untuk mempelajari fitur j , Dimana fitur dipelajari melalui boosting iteratif per fitur

$f_{jk}(x_j, x_k)$ = fungsi interaksi untuk pasangan fitur (j,k), dimana pengaruh antar fitur dihitung terhadap kontribusinya pada prediksi (deteksi otomatis oleh EBM)

Proses ini dimulai dengan nilai dasar (intercept) berupa rata-rata rating dari data interaksi user Goodreads yang digunakan sebagai Target prediksi. Kemudian nilai Intercept tersebut diperbaiki melalui iterasi kontribusi fitur hingga konvergen. Selain itu, dilakukan eksplorasi interaksi antar fitur 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.

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.

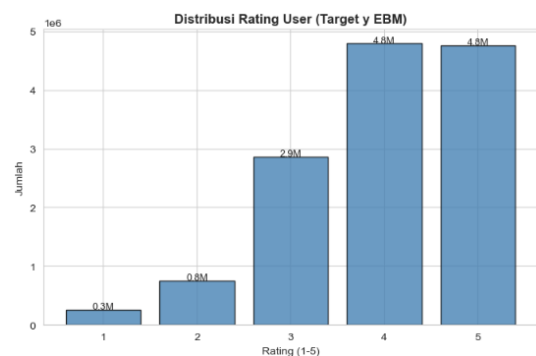
Pengujian dilakukan untuk mengevaluasi performa model EBM dalam dua aspek, yaitu akurasi prediksi dan kualitas ranking pada sistem two-stage ranking. Eksperimen dilakukan dengan membandingkan dua konfigurasi model EBM, yaitu tanpa interaksi fitur dan dengan interaksi terbatas. Evaluasi dilakukan dengan beberapa skenario, meliputi; BM25 (baseline) sebagai retrieval awal; BM25 + EBM dengan model interaksi 0 dan 5 untuk reranking kandidat; dan hybrid scoring yang menggabungkan skor kandidat BM25 dan EBM. Pengujian menggunakan 50 kueri uji yang merepresentasikan berbagai tipe pencarian, seperti berbasis genre, judul, dan deskripsi.

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.

C. Hasil Penelitian dan Pembahasan

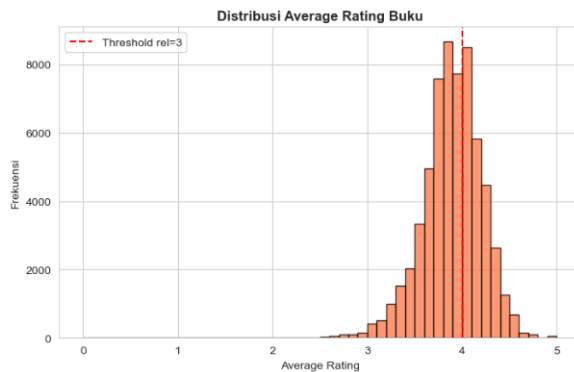
Hasil 5-core filtering pada dataset menunjukkan adanya bias positif yang ditunjukkan di distribusi rating user pada data interaksi dan data rata-rata rating buku. Hal ini menunjukkan banyak user merating buku yang mereka suka, sehingga mayoritas rating user dan nilai avg rating bukunya berada pada rentang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



Gambar 2. Distribusi rating user goodreads

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.



Gambar 3. Distribusi nilai rating buku

Sekitar 23.665 dari 61.893 buku pada korpus, sebesar 38% sudah memenuhi ambang relevansi ($\text{avg_rating} \geq 4.0$). 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 ≥ 2 ($\text{avg_rating} \geq 3.5$), sehingga retrieval acak pun akan menghasilkan NDCG yang tinggi. Sesuai dengan distribusi prediksi rating EBM (Gambar 2) yang mayoritas berada di rentang rating 4 dan 5.

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.

Seluruh 50 kueri digunakan secara konsisten untuk semua skenario evaluasi (Skenario a–c) untuk perbandingan evaluasi yang adil.

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.

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	int0	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 menunjukkan 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.

Rank	Perubahan Rank	Judul	Genre	Tahun	Avg Rating	Int Rating	Skor BM25
1	—	The Magic Books: Fur Magic: Steel M...	Fantasy, Paranorm...	1988	4.11	83	9.93
2	—	The Time of the Warlock	Fantasy, Paranorm...	2001	4.01	43	9.07
3	—	Magic by the Lake (Tales of Magic)	Fantasy, Paranorm...	1999	4.07	113	8.84
4	—	Magic Knight Rayearth I: Bowed Set	Fantasy, Paranorm...	2002	4.20	421	8.80
5	—	The Magic Goes Away Collection: The	Fantasy, Paranorm...	2005	3.88	295	8.75
6	—	Rifts World Book 3: England	Fiction Fantasy, ...	1997	3.46	120	8.74
7	—	Magic Kingdom For Sale/Sold (Magic	Fantasy, Paranorm...	2001	3.87	119	8.69
8	—	Mountain Magic (Magic Mountain, #1-	Romance Fantasy...	2007	4.20	10	8.62
9	—	The Magic Stone	Children Young-Ad...	1964	3.50	6	8.56
10	—	The magic of Shapeshifting	Non-Fiction Fanta...	2001	3.36	45	8.43

Gambar 4. Hasil pencarian BM25 Sebagai contoh pada model EBM int5 dengan $NDCG$ terbesar, 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. Dapat dilihat buku pada peringkat 1 (Gambar 4) memiliki jumlah rating rendah sebanyak 83 dengan nilai rating 4.11. Kemudian pada Gambar 5 dapat dilihat perubahan ranking beberapa buku ke peringkat top 10.

Rank	Perubahan Rank	Judul	Genre	Tahun	Avg Rating	Int Rating	Skor BM25	Skor EBM
1	↑ 32	The Magic School Bus Going Batty: A	Children Fiction...	1996	4.38	1952	8.13	4.11
2	↑ 55	The Books of Magic: Volume 5: Girl	Comics, Graphic F...	2001	3.97	43	7.91	4.10
3	↑ 85	The Magic Finger	Children Fantasy...	2001	3.66	19,574	7.63	4.10
4	↑ 2	The Time of the Warlock	Fantasy, Paranorm...	2001	4.01	43	9.07	4.00
5	↑ 20	Lost in the Solar System (The Magic	Children Fiction...	1992	4.28	8,800	8.18	4.08
6	↑ 41	Bloodstorm	Fantasy, Paranorm...	2005	4.15	553	7.96	4.05
7	↑ 19	Through The Magic Mirror	Fantasy, Paranorm...	2001	3.98	87	8.18	4.05
8	↑ 10	Magic in the Park	Fantasy, Paranorm...	1986	4.08	206	8.25	3.94
9	↑ 46	Evie The Mist Fairy (Rainbow Magic	Children Fantasy...	2004	3.79	1,113	7.93	4.03
10	↑ 68	Wild Magic (Drinker of Souls Wild	Fantasy, Paranorm...	1991	3.69	100	7.70	4.01

Gambar 5. Kandidat reranking EBM int0

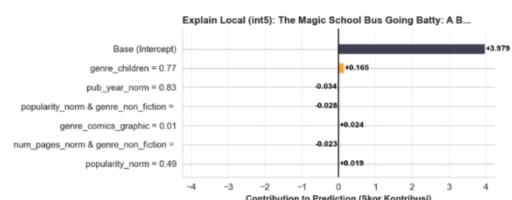
Pada kandidat reranking oleh model int0, menghasilkan buku peringkat 1 dengan jumlah rating 1952 dengan rating 4.38, yang mengalami kenaikan 32 tingkat dari yang sebelumnya peringkat 33 pada kandidat BM25 (ditunjukkan tanda panah buku naik). Kandidat BM25

yang berada di peringkat ke 2 (The Time of The Warlock), diurutkan ulang menjadi turun ke peringkat 4. Mayoritas kandidat buku pada peringkat top-10 EBM 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.

Pada skenario c hybrid scoring, parameter α 0 menjaga relevansi teks murni oleh BM25 yang menghasilkan kandidat yang cocok dengan kueri saja. Terdapat perubahan ranking kandidat buku setelah direrank dengan skor EBM dan α 0.5 yang mengkombinasi kedua skor dengan seimbang. Pada kueri 'harry potter' kandidat BM25 berisi buku yang secara relevansi teks memiliki skor BM25 yang besar, jadi buku pada peringkat teratas bukan buku harry potter seri individual melainkan koleksi set. Sedangkan pada hybrid EBM buku harry potter individual (seri utama) yang lebih

populer dan valid secara komunitas, naik ke peringkat lebih atas.

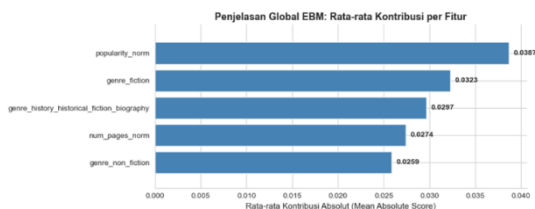
Penjelasan lokal pada setiap kandidat buku menjelaskan penyusunan skor prediksi yang mempengaruhi perubahan peringkat buku dari kandidat BM25 menjadi kandidat EBM. Pada penjelasan fitur lokal (Gambar 6), ditunjukkan bahwa fitur yang dipelajari EBM memiliki kontribusi yang kecil. Kontribusi kecil tersebut 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 user-nya 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 user-nya.



Gambar 6. Penjelasan lokal satu buku

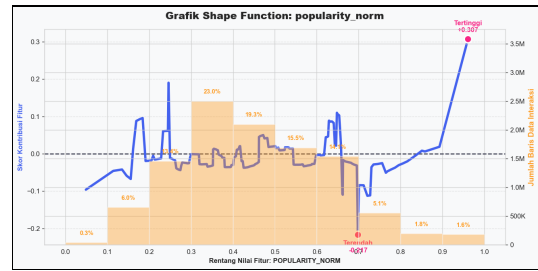
Nilai aktual fitur buku tidak berubah antara user satu dan user lain untuk buku yang sama, maka dari itu

kontribusi fiturnya hanya bisa menyesuaikan nilai intercept secara kecil pada skor prediksi total. 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.



Gambar 7. Rata-rata kontribusi fitur global buku

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.



Gambar 8. Shape function fitur popularitas

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.

Mayoritas genre dan fitur lain memiliki 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.

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.

D. 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 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-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. Penjelasan EBM bersifat jujur (*faithful*) terhadap model itu sendiri, sehingga model dapat diaudit secara langsung.

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).
<https://doi.org/10.1109/CSECS64665.2025.11009793>
- 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>
- Jenkins, S., & Caruana, R. (n.d.). *InterpretML: A Unified Framework for Machine Learning*. 1–8.
- 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>
- 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. *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>