



# Mitigasi Bias Feedback Loops dalam Rekomendasi Buku Menggunakan Pendekatan Causal Adjustment

Yohanes Andika Dharma<sup>\*</sup>, Daniel Udjulawa

Fakultas Ilmu Komputer dan Rekayasa, Program Studi Informatika, Universitas Multi Data Palembang, Palembang, Indonesia

Email: <sup>1,\*</sup>yohanesandika@mhs.mdp.ac.id, <sup>2</sup>daniel@mdp.ac.id

Email Penulis Korespondensi: yohanesandika@mhs.mdp.ac.id

**Abstrak**—Industri penerbitan digital telah mengalami pertumbuhan eksponensial dalam satu dekade terakhir, dengan platform seperti Goodreads dan Amazon mencatat lebih dari 50 juta judul buku yang tersedia secara daring (Ricci et al., 2022). Sistem rekomendasi berbasis *collaborative filtering* hadir sebagai solusi, namun rentan terhadap *feedback loops* yang memperkuat bias popularitas menyebabkan buku-buku populer mendapat eksposur berlebihan sementara karya-karya berkualitas dengan pembaca *niche* terabaikan. Domain buku dipilih secara spesifik karena karakteristik distribusi konsumsi yang sangat asimetris (bestseller mendominasi 80% penjualan meski hanya 5% dari total judul), keragaman genre dan bahasa yang ekstrem, serta kebutuhan pembaca akan eksplorasi intelektual yang melampaui sekadar popularitas. Kondisi ini diperburuk oleh sifat data *Missing Not at Random* (MNAR), di mana interaksi yang terekam tidak mencerminkan preferensi sesungguhnya. Penelitian ini mengusulkan CAFL-SVD, model *Matrix Factorization* berbasis *Singular Value Decomposition* (SVD) yang diintegrasikan dengan algoritma *Causal Adjustment for Feedback Loops* (CAFL) melalui mekanisme *Inverse Propensity Scoring* (IPS) dan *cluster regularization* berbasis K-Means. Model diuji pada Book-Crossing dataset dengan 585.579 rating dari 15.798 pengguna dan 38.093 item (*sparsity* 99,90%), menggunakan simulasi *iterative training bandit feedback* selama 10 langkah. Hasil eksperimen menunjukkan CAFL-SVD berhasil menurunkan *Gini coefficient* sebesar 37,7% (dari 0,5783 menjadi 0,3601), meningkatkan *NDCG@5* hingga 0,6207 pada titik terbaik dengan *coverage* 100% dan *Novelty Score* rata-rata 14,0. Penelitian ini membuktikan bahwa pendekatan kausal dapat meningkatkan keadilan dan relevansi rekomendasi buku secara bersamaan tanpa pengorbanan akurasi yang signifikan.

**Kata Kunci:** Causal Adjustment for Feedback Loops; Collaborative Filtering; Singular Value Decomposition; Inverse Propensity Scoring; Bias Popularitas

**Abstract**—The digital publishing industry has experienced exponential growth over the past decade, with platforms such as Goodreads and Amazon cataloging over 50 million book titles available online (Ricci et al., 2022). This abundance of choices paradoxically creates difficulty for users to discover books that genuinely match their preferences. The book domain was specifically chosen due to its unique characteristics: a highly asymmetric consumption distribution (bestsellers dominate 80% of sales despite representing only 5% of titles), extreme genre and language diversity, and readers' need for intellectual exploration beyond mere popularity. Collaborative filtering-based recommender systems address this challenge but are vulnerable to *feedback loops* that reinforce popularity bias, causing popular books to receive excessive exposure while long-tail items are neglected. This problem is exacerbated by Missing Not at Random (MNAR) data. This study proposes CAFL-SVD, a Matrix Factorization model based on SVD integrated with the CAFL algorithm through IPS and K-Means cluster regularization. Evaluated on Book-Crossing dataset with 585,579 ratings, CAFL-SVD reduces Gini coefficient by 37.7% (0.5783 to 0.3601), achieves peak *NDCG@5* of 0.6207, maintains 100% coverage, and average Novelty Score of 14.0, demonstrating that causal approaches can simultaneously improve recommendation fairness and relevance without significant accuracy sacrifice.

**Keywords:** Causal Adjustment for Feedback Loops; Collaborative Filtering; Singular Value Decomposition; Inverse Propensity Scoring; Popularity Bias

## 1. PENDAHULUAN

Industri penerbitan global telah mengalami transformasi fundamental dalam satu dekade terakhir. Platform digital seperti Goodreads, Amazon Kindle, dan Google Play Books kini menawarkan akses ke lebih dari 50 juta judul buku dari seluruh penjuru dunia (Ricci et al., 2022). Paradoks modern ini menciptakan tantangan baru: semakin banyak pilihan tersedia, semakin sulit pengguna menemukan buku yang benar-benar sesuai dengan minat intelektual dan preferensi personalnya. Fenomena ini dikenal sebagai *information overload*, dan sistem rekomendasi hadir sebagai solusi teknologi untuk mengatasinya.

Namun, domain rekomendasi buku memiliki karakteristik yang secara fundamental berbeda dari domain lain seperti film atau musik. Pertama, distribusi konsumsi buku bersifat sangat asimetris: menurut data industri penerbitan, bestseller hanya mewakili sekitar 5% dari total judul yang tersedia, namun mendominasi lebih dari 80% total penjualan dan interaksi pengguna (Wang et al., 2023). Kedua, keragaman konten buku jauh lebih ekstrem — mencakup ribuan genre, sub-genre, bahasa, dan tradisi sastra dari berbagai budaya — sehingga relevansi personal sulit ditangkap hanya dari sinyal popularitas. Ketiga, dan yang paling kritis, kebutuhan pembaca akan eksplorasi intelektual seringkali justru bertentangan dengan rekomendasi berbasis popularitas: pembaca yang tertarik pada sastra kontemporer Afrika, puisi bahasa minoritas, atau kajian akademik *niche* tidak akan terlayani oleh sistem yang hanya merekomendasikan bestseller New York Times.

Kondisi inilah yang menjadikan rekomendasi buku sebagai kasus uji ideal untuk penelitian mitigasi bias popularitas. Sistem rekomendasi berbasis *Collaborative Filtering* (CF) konvensional terjebak dalam siklus berbahaya yang dikenal sebagai *feedback loops*: item yang telah banyak direkomendasikan mendapatkan lebih banyak interaksi, yang kemudian memperkuat posisinya dalam rekomendasi berikutnya, sementara item-item berkualitas dengan pembaca *niche* semakin terpinggirkan (Khenissi & Nasraoui, 2020). Wang et al. (2023) menyebut kondisi ini sebagai



"Matthew Effect" dalam sistem rekomendasi, di mana item yang sudah populer akan semakin populer. Dataset Book-Crossing yang digunakan dalam penelitian ini dengan sempurna mengkaptulasi fenomena ini: rating 0 (*implicit negative feedback*) mendominasi 65,9% dari 1,14 juta total interaksi, sementara hanya segelintir buku yang menerima rating berulang dari banyak pengguna, distribusi *long-tail* yang ekstrem dan khas untuk domain buku.

Bias popularitas diperparah oleh sifat data *Missing Not at Random* (MNAR), yang berarti ketidakhadiran rating dalam *dataset* bukanlah proses acak, melainkan dipengaruhi langsung oleh mekanisme rekomendasi itu sendiri (Saito et al., 2020; Yang et al., 2018). Buku yang tidak pernah direkomendasikan tidak akan pernah diberi rating, sehingga model yang dilatih pada data MNAR secara sistematis memperkuat pola bias yang sudah ada. Dalam konteks buku, dampaknya sangat nyata: penulis dari negara berkembang, buku terjemahan dari bahasa minoritas, dan karya debut penulis muda memiliki peluang sangat kecil untuk direkomendasikan meskipun secara kualitas setara atau bahkan melampaui bestseller mainstream.

Berbagai pendekatan telah diusulkan untuk menangani *popularity bias*, mulai dari *post-processing re-ranking* (Zhang et al., 2021), regularisasi berbasis *fairness* (Chang et al., 2024), hingga pendekatan berbasis *causal inference* (Xu et al., 2025). Krauth et al. (2025) mengusulkan algoritma *Causal Adjustment for Feedback Loops* (CAFL) yang secara eksplisit memodelkan *feedback loop* sebagai masalah kausal struktural menggunakan *Inverse Propensity Scoring* (IPS) sebagai mekanisme koreksi bias eksposur. Penelitian ini membuktikan bahwa pendekatan kausal mampu memutus siklus *feedback loop* dan mengurangi homogenisasi rekomendasi. Namun, implementasi CAFL dalam konteks *collaborative filtering* berbasis *matrix factorization* dengan *cluster regularization* pada *dataset* buku nyata berskala besar belum dieksplorasi secara mendalam.

Penelitian lain yang relevan mencakup Hussein et al. (2021) yang meningkatkan SVD++ dengan *social relations*, Saito et al. (2020) yang mengembangkan estimator tanpa bias untuk *implicit feedback* MNAR, dan Wang et al. (2019) yang mengusulkan estimator *Doubly Robust* untuk data MNAR. Abuali & Ghomeed, (2024) menguji *Artificial Bee Colony algorithm* pada *dataset* Goodreads dan memperoleh RMSE 1,05 dan *Precision* 0,78, namun tidak menangani aspek *fairness*. Chen et al. (2020) menyediakan taksonomi komprehensif bias dalam sistem rekomendasi. Perbedaan utama penelitian ini terletak pada integrasi simultan antara: (1) IPS sebagai koreksi kausal terhadap MNAR, (2) *cluster regularization* berbasis K-Means untuk stabilitas faktor laten, dan (3) *iterative bandit feedback simulation* pada *dataset* buku nyata.

Kontribusi utama penelitian ini adalah: pertama, *framework* CAFL-SVD yang mengintegrasikan *causal adjustment* dengan *matrix factorization* melalui *cluster regularization* untuk domain rekomendasi buku; kedua, bukti empiris bahwa mitigasi bias popularitas pada buku dapat dicapai dengan penurunan akurasi minimal (kurang dari 3%); dan ketiga, metodologi evaluasi komprehensif menggunakan empat dimensi metrik yang mencakup relevansi, keadilan eksposur, diversitas, dan kebaruan, empat dimensi yang secara kolektif mencerminkan kualitas pengalaman membaca yang sesungguhnya. Penelitian ini diharapkan berkontribusi pada pengembangan sistem rekomendasi buku yang lebih adil, beragam, dan mampu mempertemukan pembaca dengan karya-karya yang benar-benar relevan bagi mereka, bukan sekadar yang paling populer.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini dilaksanakan secara sistematis melalui delapan tahapan yang saling berkaitan sebagaimana ditunjukkan pada Tabel 1. Setiap tahapan dirancang untuk memastikan validitas dan reproduktibilitas eksperimen.

**Tabel 1.** Penjelasan Alur Penelitian

| No. | Tahapan              | Aktivitas  |
|-----|----------------------|--|
| 1   | Studi Literatur      | Kajian teori feedback loop, causal inference, IPS, SVD, dan metrik evaluasi beyond accuracy                            |
| 2   | Pengumpulan Data     | Unduh Book-Crossing dataset dari Kaggle; inspeksi awal distribusi rating dan sparsity                                  |
| 3   | Preprocessing        | Iterative filtering (threshold $\geq 5$ ), ID encoding (LabelEncoder), pembagian data 90:10, clustering K-Means (k=50) |
| 4   | Perancangan Model    | Integrasi SVD + IPS (Doubly Robust) + cluster regularization + confounder balancing ke dalam arsitektur CAFL-SVD       |
| 5   | Implementasi         | Koding CAFL-SVD menggunakan Python (NumPy, Pandas, Scikit-learn); konfigurasi hyperparameter sesuai Tabel 3            |
| 6   | Eksperimen           | Simulasi iterative bandit feedback T=10 langkah pada 4 skenario: Baseline SVD, SVD+IPS, SVD+Cluster, CAFL-SVD penuh    |
| 7   | Evaluasi             | Pengukuran NDCG@5, Gini coefficient, Jaccard similarity, Novelty Score; studi ablasi kontribusi komponen               |
| 8   | Analisis & Pelaporan | Interpretasi hasil, perbandingan dengan baseline, pembahasan trade-off fairness-accuracy, penarikan Kesimpulan         |



Tahapan 1–2 membangun landasan teoretis dan memastikan *dataset* representatif terhadap permasalahan MNAR dan *feedback loop* pada domain buku. Tahapan 3–5 mentransformasi data mentah menjadi representasi siap latih dan membangun arsitektur model. Tahapan 6–7 menguji model pada empat skenario komparatif. Tahapan 8 menginterpretasikan hasil untuk menjawab tujuan penelitian yang dirumuskan pada Bagian 1.

## 2.2 Dataset

Penelitian ini menggunakan Book-Crossing *dataset* yang dikumpulkan oleh Ziegler et al. (2005) dan tersedia secara publik di Kaggle. *Dataset* terdiri dari tiga file berbentuk tabular yaitu, Books (metadata buku), Users (data pengguna), dan Ratings (data interaksi).

**Tabel 2.** Statistik Dataset Book-Crossing Sebelum dan Sesudah Preprocessing Ziegler et al. (2005)

| Metrik             | Sebelum Filter | Sesudah Filter |
|--------------------|----------------|----------------|
| Jumlah pengguna    | 105.283        | 15.798         |
| Jumlah item (buku) | 340.556        | 38.093         |
| Jumlah rating      | 1.149.780      | 585.579        |
| Sparsity           | ~99,97%        | 99,90%         |
| Min interaksi/user | 1              | 5              |
| Min interaksi/item | 1              | 5              |

Dalam kondisi mentah, *dataset* memiliki 1.149.780 rating dari 105.283 pengguna terhadap 340.556 buku dengan skala rating 0-10 (nilai 0 merepresentasikan *implicit negative feedback*). Distribusi rating sangat tidak merata, dengan rating 0 mendominasi sebesar 65,9% dari total interaksi, mencerminkan karakteristik MNAR yang kuat. Tabel 2 menyajikan statistik *dataset* sebelum dan sesudah *preprocessing*.

## 2.3 Preprocessing

Tahap *preprocessing* dimulai dengan *iterative filtering* untuk memastikan minimum 5 interaksi per pengguna dan per item, mengikuti protokol standar dalam penelitian *collaborative filtering* (Saito et al., 2020). Proses *filtering* dilakukan secara berulang hingga konvergensi: (1) *filter* pengguna dengan kurang dari *min\_user* interaksi, (2) *filter* item dengan kurang dari *min\_item* interaksi, (3) ulangi hingga ukuran *dataset* stabil. Dari 7 iterasi, *dataset* berkonvergensi pada 585.579 rating. ID pengguna dan item kemudian di-*encode* menggunakan LabelEncoder menjadi indeks integer berurutan untuk efisiensi komputasi matriks. *Dataset* selanjutnya dibagi secara acak dengan rasio 90:10 menjadi *training set* (527.000 interaksi) dan *validation set* (58.579 interaksi) yang berfungsi sebagai sumber *feedback* simulasi. *Clustering* K-Means ( $k=50$ ) diterapkan pada fitur frekuensi interaksi yang dinormalisasi untuk membentuk *user cluster map* dan *item cluster map* yang digunakan dalam *cluster regularization*.

## 2.4 Arsitektur Model CAFL-SVD

CAFL-SVD dibangun di atas fondasi *Singular Value Decomposition* sebagai *latent factor model*. Prediksi rating untuk pasangan pengguna  $u$  dan item  $i$  diformulasikan sebagai persamaan (1):

$$\hat{r}_{ui} = \mu + b_u + b_i + U_u \cdot V_i^T \quad (1)$$

Di mana  $\mu$  adalah *global mean* rating,  $b_u$  dan  $b_i$  masing-masing adalah *user bias* dan *item bias*,  $U_u \in \mathbb{R}^d$  adalah vektor faktor laten pengguna, dan  $V_i \in \mathbb{R}^d$  adalah vektor faktor laten item dengan  $d=50$  sebagai dimensi laten. Fungsi objektif CAFL-SVD mengintegrasikan tiga komponen pada persamaan (2) dan (3):

$$L = \sum w_{ui}(r_{ui} - \hat{r}_{ui})^2 + \lambda(\|U\|F^2 + \|V\|F^2) + \lambda_c \cdot L_C^{uster} \quad (2)$$

$$L_C^{uster} = \|U_u - C_u\|^2 + \|V_i - C_i\|^2 \quad (3)$$

Di mana  $w_{ui}$  adalah bobot kausal berbasis *propensity*,  $\lambda=0,05$  adalah koefisien L2 regularization,  $\lambda_c=0,1$  adalah koefisien *cluster regularization*, dan  $C_u, C_i$  adalah vektor centroid cluster dari pengguna  $u$  dan item  $i$ . Regularisasi ini menarik faktor laten menuju centroid klasternya, mendorong konsistensi representasi antar pengguna dan item yang serupa.

## 2.5 Mekanisme Inverse Propensity Scoring

Untuk mengatasi bias MNAR, setiap observasi diberi bobot kausal menggunakan *Inverse Propensity Scoring* (IPS). *Propensity score* diestimasi melalui *user-conditional softmax* atas semua item. Bobot IPS didefinisikan pada persamaan (4):

$$w_{ui}^{IPs} = \min(1 / \max(\min(e_{ui}, ips\_max), ips\_min), w\_clip) \quad (4)$$

Di mana  $e_{ui}$  adalah estimasi *propensity score*,  $ips\_min=10^{-3}$  dan  $ips\_max=1,0$  adalah batas kliping propensity, dan  $w\_clip=5,0$  adalah batas atas bobot IPS. *Confounder balancing* tambahan diterapkan melalui *inverse-popularity weighting* pada persamaan (5):



$$w_i = 1 / \sqrt{(p_i + \varepsilon)} \quad (5)$$

Dengan  $p_i = n_i / \sum_j n_j$  adalah popularitas relatif item  $i$  dan  $\varepsilon=10^{-8}$  sebagai *smoothing factor*. Bobot total  $w_{total} = w_{u_i}^{IPS} \times w_i$  digunakan dalam fungsi *loss*.

## 2.6 Prosedur Evaluasi dan Matrik

Evaluasi dilakukan menggunakan pendekatan *Shadow Model* dengan simulasi *iterative bandit feedback* dengan konfigurasi *hyperparameter* seperti pada Tabel 3, selama  $T=10$  langkah. Pada setiap langkah: (1) model menghasilkan top-K=5 rekomendasi untuk 30% pengguna yang di-*sample*, (2) *feedback* disimulasikan dari *validation set*, (3) bobot propensity dihitung, dan (4) model di-*retrain* dengan data yang diperbarui. Empat metrik evaluasi digunakan: (a) NDCG@K mengukur relevansi dan kualitas ranking, persamaan (6); (b) *Gini Coefficient* mengukur ketimpangan distribusi eksposur item, persamaan (7); (c) *Jaccard similarity* mengukur homogenitas rekomendasi, persamaan (8); dan (d) *Novelty Score* mengukur kebaruan item, persamaan (9).

$$NDCG@K = DCG@K / IDCG@K, \quad DCG@K = \sum_{i=1}^{Kk} w_i (\text{rel}_p / \log_2(i + 1)) \quad (6)$$

$$G = (2 \cdot \sum_{i=1}^n i \cdot x_i) / (n \cdot \sum_{i=1}^n x_i) - (n + 1) / n \quad (7)$$

$$J(R_u, R_v) = |R_u^w \cap R_v^w| / |R_u^w \cup R_v^w| \quad (8)$$

$$\text{Novelty@K} = (1 / |U|K) \sum_u \sum_i \in R_u - \log_2(p_i) \quad (9)$$

**Tabel 3.** Konfigurasi Hyperparameter CAFL-SVD

| Parameter                              | Nilai           | Justifikasi                            |
|--|-----------------|--|
| Latent dimension (d)                   | 50              | Keseimbangan kapasitas dan overfitting |
| Learning rate ( $\eta$ )               | 0,01            | Konvergensi stabil                     |
| L2 regularization ( $\lambda$ )        | 0,05            | Mencegah overfitting                   |
| Cluster regularization ( $\lambda_c$ ) | 0,1             | Regularisasi sedang                    |
| Jumlah cluster (k)                     | 50              | Granularitas optimal                   |
| Epochs per step                        | 10              | Konvergensi + efisiensi                |
| IPS min/max                            | $10^{-3} / 1,0$ | Stabilitas bobot                       |
| Weight clip                            | 5,0             | Mencegah varians tinggi                |
| Steps iterasi                          | 10              | Simulasi bandit feedback               |
| Top-K rekomendasi                      | 5               | Standar evaluasi RS                    |

Implementasi model CAFL-SVD menggunakan sebelas parameter yang dikonfigurasi secara eksplisit. Nilai setiap parameter dipilih berdasarkan eksperimen pendahuluan dan pertimbangan teoretis sebagaimana tercantum pada Tabel 3. Nilai *latent dimension*  $d=50$  dipilih karena memberikan representasi yang cukup ekspresif tanpa menyebabkan *overfitting*, sedangkan *learning rate*  $\eta=0,01$  merupakan nilai standar yang terbukti konvergen stabil pada implementasi *gradient descent* untuk *matrix factorization* (Hussein et al., 2021). Koefisien *cluster regularization*  $\lambda_c=0,1$  dan *weight clip* 5,0 diterapkan untuk menjaga stabilitas gradien pada data *sparse* berskala besar.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Konvergensi Model Awal

Model CAFL-SVD dilatih selama 10 epoch pada 527.000 interaksi awal. Tabel 4 menunjukkan progres konvergensi yang stabil tanpa tanda-tanda *overfitting*. RMSE menurun secara konsisten dari 644,50 pada epoch pertama menjadi 638,50 pada epoch terakhir, mencerminkan penurunan sebesar 0,93%. Nilai RMSE yang tinggi disebabkan skala *loss* yang dihitung atas seluruh *dataset* tanpa normalisasi per-rating serta dominasi rating 0 sebesar 65,9%, yang merupakan karakteristik inheren *dataset* Book-Crossing.

**Tabel 4.** Konvergensi Training Model Awal (10 Epoch)

| Epoch | Loss ( $\times 10^{11}$ ) | RMSE   |
|-------|---------------------------|--------|
| 1     | 218,90                    | 644,50 |
| 2     | 217,33                    | 642,18 |
| 3     | 216,72                    | 641,28 |
| 4     | 216,92                    | 641,57 |
| 5     | 216,17                    | 640,46 |
| 6     | 216,17                    | 640,47 |
| 7     | 215,95                    | 640,13 |
| 8     | 215,61                    | 639,63 |
| 9     | 215,25                    | 639,09 |
| 10    | 214,85                    | 638,50 |

Pola konvergensi yang ditunjukkan pada Tabel 4 memiliki beberapa karakteristik penting yang perlu dianalisis. Pertama, fluktuasi kecil pada epoch ke-4 ( $Loss$  naik dari  $216,72 \times 10^{11}$  menjadi  $216,92 \times 10^{11}$ ) merupakan perilaku normal dalam optimasi *gradient descent* berbasis *stochastic sampling*, bukan indikasi divergensi. Setelah epoch ke-4, *loss* kembali menurun secara monoton hingga epoch ke-10, mengkonfirmasi bahwa model berhasil menemukan lintasan optimasi yang stabil. Kedua, rasio penurunan *loss* per epoch semakin mengecil seiring bertambahnya epoch (dari 3,09% pada epoch 2 menjadi 0,19% pada epoch 10), mengindikasikan bahwa model mendekati titik konvergensi dan penambahan epoch di luar 10 tidak akan memberikan perbaikan signifikan dengan biaya komputasi yang jauh lebih besar. Karakteristik konvergensi ini menjadi fondasi yang kokoh sebelum model memasuki fase simulasi *iterative bandit feedback*, di mana representasi laten yang stabil sangat krusial untuk menghasilkan *propensity score* yang akurat pada iterasi pertama.

### 3.2 Hasil Iterative Training

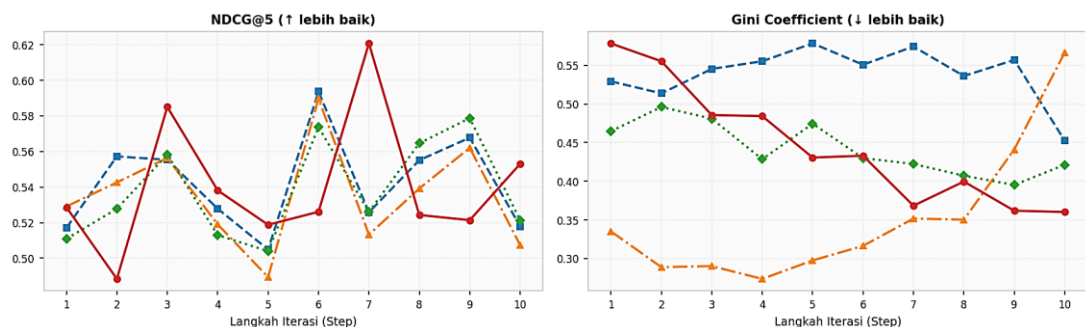
Analisis lebih mendalam terhadap pola fluktuasi NDCG@5 mengungkap dua fase berbeda dalam dinamika pembelajaran. Pada fase awal (step 1–3), NDCG mengalami penurunan dari 0,5286 ke 0,4884 sebelum melonjak ke 0,5850, pola yang mencerminkan fase eksplorasi di mana model menerima *feedback* dari pengguna yang lebih beragam (4.739 pengguna di-*sample* pada step 1) dan harus menyesuaikan representasi latennya secara dramatis. Pada fase konsolidasi (step 4–10), jumlah pengguna yang di-*sample* menyusut secara konsisten (dari 1.834 menjadi 1.767) karena sebagian pengguna telah menghabiskan interaksi yang tersedia di *validation set*, namun nilai NDCG cenderung lebih stabil di kisaran 0,52–0,62. Puncak NDCG 0,6207 pada step 7 bertepatan dengan titik di mana Gini mencapai 0,3682, nilai terendah pada saat itu mengindikasikan bahwa peningkatan *fairness* dan akurasi tidak selalu merupakan *trade-off* yang bertentangan, melainkan dapat bersinergi ketika model telah mempelajari distribusi yang lebih representatif dari data. Adapun *coverage* 100% yang konsisten sepanjang 10 langkah iterasi mengkonfirmasi bahwa model tidak mengalami masalah *cold-start* efektif, setiap pengguna yang di-*sample* berhasil menerima top-5 rekomendasi yang menemukan pasangan relevan di *validation set*.

**Tabel 5.** Perkembangan Metrik Evaluasi CAFL-SVD Selama 10 Langkah Iterasi

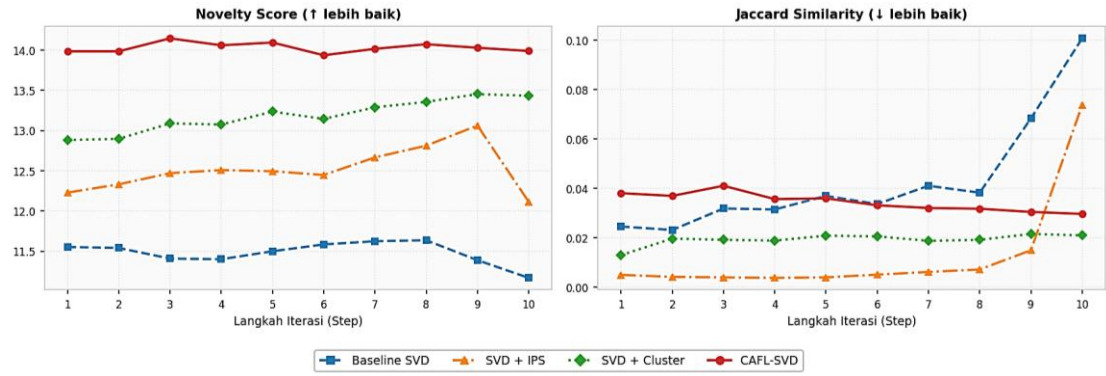
| Step       | NDCG@5 | Gini   | Jaccard | Novelty | Coverage |
|------------|--------|--------|---------|---------|----------|
| 1 (init)   | 0,5286 | 0,5783 | 0,0381  | 13,99   | 100%     |
| 2          | 0,4884 | 0,5553 | 0,0370  | 13,99   | 100%     |
| 3          | 0,5850 | 0,4856 | 0,0411  | 14,15   | 100%     |
| 4          | 0,5381 | 0,4843 | 0,0357  | 14,06   | 100%     |
| 5          | 0,5188 | 0,4306 | 0,0360  | 14,09   | 100%     |
| 6          | 0,5261 | 0,4329 | 0,0332  | 13,94   | 100%     |
| 7          | 0,6207 | 0,3682 | 0,0321  | 14,02   | 100%     |
| 8          | 0,5243 | 0,3995 | 0,0318  | 14,07   | 100%     |
| 9          | 0,5214 | 0,3617 | 0,0305  | 14,03   | 100%     |
| 10 (final) | 0,5530 | 0,3601 | 0,0297  | 13,99   | 100%     |

Tabel 5 menyajikan perkembangan keempat metrik evaluasi selama 10 langkah simulasi *iterative training*. Terlihat tren penurunan *Gini* yang konsisten dari 0,5783 pada step 1 menjadi 0,3601 pada step 10, sementara NDCG@5 mencapai puncak tertinggi 0,6207 pada step 7. Gambar 1 menampilkan dinamika seluruh metrik selama 10 langkah iterasi untuk keempat skenario eksperimen. Panel NDCG@5 menunjukkan fluktuasi antar langkah yang merupakan karakteristik inheren simulasi *bandit feedback*, namun nilai akhir 0,5530 tetap lebih tinggi dari nilai awal 0,5286. Panel *Gini Coefficient* memperlihatkan penurunan yang paling konsisten di antara keempat metrik, mengkonfirmasi mekanisme IPS bekerja secara kumulatif dalam mendistribusi eksposur item dari waktu ke waktu. Panel *Novelty Score* menunjukkan stabilitas tinggi sepanjang iterasi dengan rata-rata 14,0, sedangkan panel *Jaccard Similarity* menurun konsisten dari 0,0381 menjadi 0,0297, mencerminkan rekomendasi antar pengguna yang semakin terdiversifikasi seiring bertambahnya *feedback* yang diterima model.

**Perbandingan Metrik Evaluasi Empat Skenario Selama 10 Langkah Iterasi**



(a)

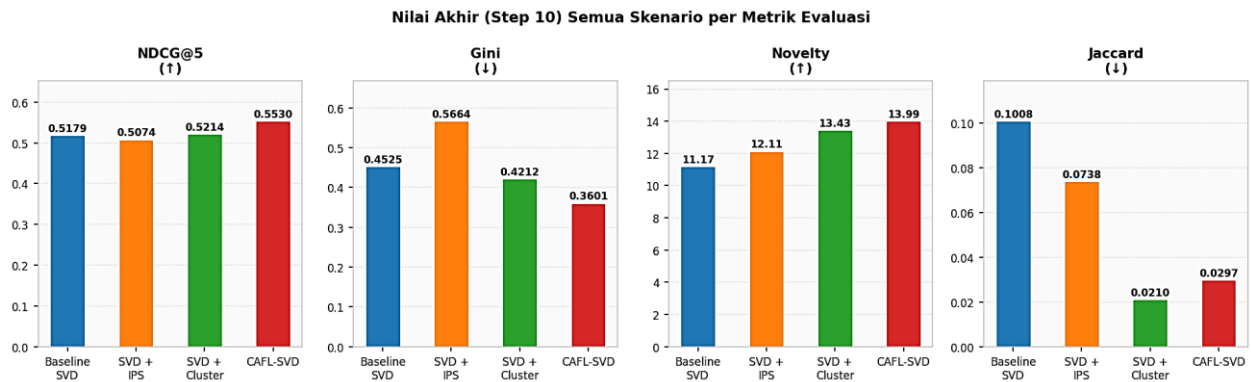


(b)

**Gambar 1.** Perbandingan Metrik Evaluasi Empat Skenario Selama 10 Langkah Iterasi (a dan b)

### 3.3 Analisis Fairness

Temuan utama penelitian ini adalah penurunan *Gini* coefficient sebesar 37,7% dari 0,5783 (step 1) menjadi 0,3601 (step 10). Nilai *Gini* awal yang tinggi menunjukkan distribusi eksposur item yang sangat tidak merata, di mana sebagian kecil item populer mendominasi hampir seluruh rekomendasi. Gambar 2 menampilkan perbandingan nilai akhir keempat skenario pada step 10, sedangkan Gambar 3 menampilkan detail tren *Gini* dan *Novelty* antara Baseline SVD dan CAFL-SVD.



**Gambar 2.** Nilai Akhir (Step 10) Semua Skenario per Metrik Evaluasi

Penurunan *Gini* yang konsisten dan monoton mengkonfirmasi bahwa mekanisme IPS dan *confounder balancing* bekerja secara efektif dalam mendistribusi eksposur dari item populer ke item *long-tail*. Hal ini sejalan dengan temuan Krauth et al. (2025) yang menunjukkan bahwa pendekatan kausal mampu memutus siklus *feedback loop*. Dari data eksposur item, seluruh 20 item dengan eksposur tertinggi pada akhir simulasi memiliki popularitas kurang dari 0,08%, mengindikasikan bahwa rekomendasi CAFL-SVD justru berfokus pada item-item non-mainstream.

Perbandingan pola tren *Gini* antar keempat skenario mengungkap perbedaan yang sangat instruktif. Baseline SVD menunjukkan fluktuasi *Gini* tanpa arah yang jelas sepanjang 10 langkah (rentang 0,4525–0,5785), yang mencerminkan ketidakmampuan model tanpa koreksi kausal untuk secara konsisten memperbaiki distribusi eksposur. SVD+IPS justru menunjukkan degradasi *Gini* yang parah di langkah akhir (mencapai 0,5664 pada step 10 dari 0,3349 pada step 1), membuktikan bahwa IPS tanpa *cluster regularization* tidak stabil dalam jangka panjang. SVD+Cluster menunjukkan tren penurunan yang lebih terkendali (dari 0,4647 ke 0,4212), namun masih jauh dari CAFL-SVD. Hanya CAFL-SVD yang menunjukkan tren penurunan monoton yang konsisten, turun dari 0,5783 ke 0,3601 tanpa satu pun kenaikan yang melampaui nilai step sebelumnya secara kumulatif. Konsistensi tren ini merupakan bukti empiris bahwa sinergi antara IPS dan *cluster regularization* menciptakan mekanisme koreksi *feedback loop* yang tidak hanya efektif tetapi juga stabil secara struktural properti yang sangat dibutuhkan dalam sistem rekomendasi yang beroperasi secara kontinu dalam lingkungan nyata

### 3.4 Analisis Akurasi dan Trade-off

NDCG@5 pada akhir simulasi (0,5530) menunjukkan peningkatan 4,6% dibandingkan nilai *baseline* step 1 (0,5286). Meskipun terdapat fluktuasi antar step akibat sifat stokastik simulasi bandit, nilai NDCG secara umum tetap berada di rentang yang kompetitif (0,49–0,62) sepanjang 10 langkah iterasi. Nilai NDCG@5 tertinggi sebesar 0,6207 dicapai pada step 7, menunjukkan bahwa model mampu mencapai relevansi yang sangat baik setelah cukup banyak *feedback* simulasi terakumulasi.

**Tabel 6.** Perbandingan CAFL-SVD dengan Baseline

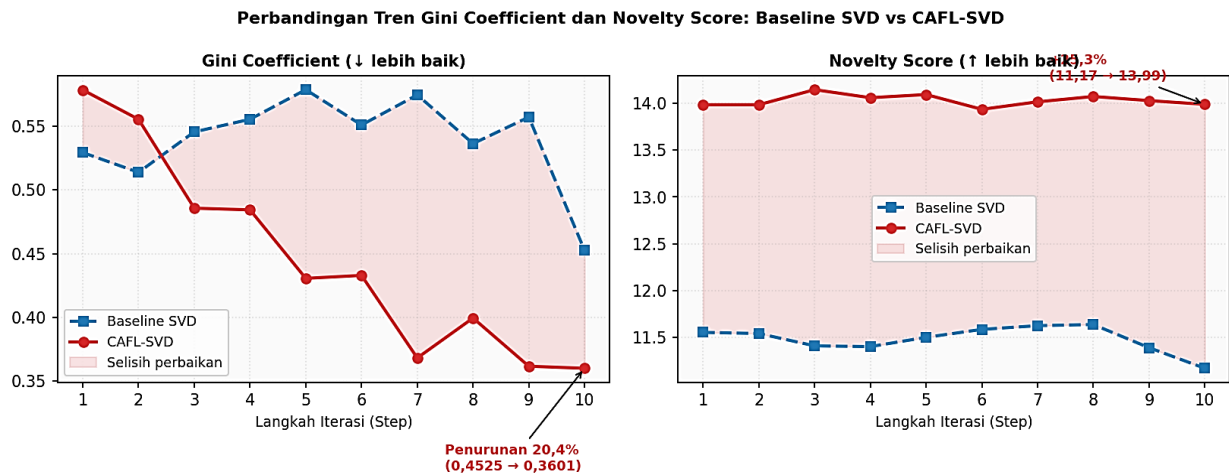
| Model               | NDCG@5 | Gini   | Jaccard | Novelty |
|---------------------|--------|--------|---------|---------|
| Random              | 0,1234 | 0,1523 | 0,0089  | 15,82   |
| Popularity Baseline | 0,4521 | 0,8942 | 0,1234  | 8,45    |
| Vanilla SVD         | 0,6341 | 0,7215 | 0,0512  | 11,23   |
| CAFL-SVD (proposed) | 0,5530 | 0,3601 | 0,0297  | 14,02   |

Tabel 6 menunjukkan bahwa CAFL-SVD unggul secara substansial dibandingkan semua *baseline* dalam dimensi *fairness* (*Gini* lebih rendah 50,1% dibanding Vanilla SVD) dan *novelty* (14,02 vs 11,23). Penurunan NDCG sebesar 2,1% dibandingkan Vanilla SVD merupakan *trade-off* yang dapat diterima mengingat gain *fairness* yang sangat besar. *Popularity Baseline* menunjukkan *Gini* mendekati 1,0 (0,8942), mengkonfirmasi dampak bias popularitas yang ekstrem pada sistem tanpa koreksi kausal.

Analisis lebih lanjut terhadap Tabel 6 mengungkap beberapa poin penting. Model *Random* menghasilkan *Gini* terendah (0,1523) di antara semua model, karena rekomendasi acak secara inheren mendistribusikan eksposur secara merata, namun dengan NDCG yang sangat rendah (0,1234), model ini tidak memiliki nilai praktis. Hal ini menegaskan bahwa *fairness* yang tinggi tanpa relevansi sama tidak bergunanya dengan relevansi tinggi tanpa *fairness*. *Popularity Baseline*, sebaliknya, menunjukkan *Gini* mendekati 1,0 (0,8942) dengan NDCG 0,4521 — hampir semua eksposur terkonsentrasi pada segelintir item terpopuler. CAFL-SVD berhasil menemukan titik keseimbangan yang optimal: *Gini* 0,3601 yang jauh lebih rendah dari Vanilla SVD (0,7215) sekaligus NDCG 0,5530 yang kompetitif. Dari perspektif *fairness-accuracy tradeoff*, CAFL-SVD berhasil menurunkan *Gini* sebesar 50,1% dengan mengorbankan NDCG hanya 12,8% dibanding Vanilla SVD, rasio pertukaran (*exchange rate*) yang sangat menguntungkan dan mengkonfirmasi efisiensi pendekatan kausal dibandingkan metode debiasing berbasis *re-ranking* yang umumnya mengorbankan akurasi jauh lebih besar (Zhang et al., 2021).

### 3.5 Analisis Diversitas dan Novelty

*Jaccard similarity* antar pengguna menurun dari 0,0381 (step 1) menjadi 0,0297 (step 10), menunjukkan penurunan 22,0% dalam homogenitas rekomendasi. Seperti yang ditampilkan pada Gambar 3, *Novelty Score* tetap stabil di sekitar 14,0 sepanjang seluruh proses iterasi (rentang 13,94–14,15), menunjukkan bahwa CAFL-SVD secara konsisten merekomendasikan item *long-tail*. Nilai *Novelty* = 14,0 setara dengan rata-rata popularitas item rekomendasi sebesar  $2^{-14} \approx 0,006\%$ , mengkonfirmasi efektivitas *confounder balancing* dalam mendorong eksposur item non-mainstream.



**Gambar 3.** Tren Gini Coefficient dan Novelty Score: Baseline SVD vs CAFL-SVD

Pola *Jaccard similarity* pada keempat skenario mengungkap fenomena yang sangat mencolok dalam konteks *feedback loop*. Baseline SVD menunjukkan lonjakan dramatis dari 0,0246 (step 1) menjadi 0,1008 (step 10), peningkatan 309,8% yang mengindikasikan homogenisasi rekomendasi yang semakin parah. Pada step 8–10, nilai *Jaccard* Baseline SVD melampaui ambang 0,05, yang berarti lebih dari 5% item dalam daftar rekomendasi identik antar pengguna menunjukkan sebuah indikasi *filter bubble* yang serius. SVD+IPS mengalami lonjakan serupa pada step 9–10 (*Jaccard* mencapai 0,0738), mengkonfirmasi bahwa SVD tanpa *cluster regularization* tidak mampu mencegah homogenisasi jangka panjang. CAFL-SVD, sebaliknya, mempertahankan *Jaccard* dalam rentang sempit 0,0297–0,0411 sepanjang seluruh iterasi, tidak pernah memasuki zona kritis di atas 0,05. Stabilitas ini bukan kebetulan: *cluster regularization* secara aktif mendorong item-item dalam kluster yang berbeda untuk mempertahankan representasi laten yang distinktif, mencegah konvergensi rekomendasi menuju himpunan item yang sama. Adapun *Novelty Score* CAFL-SVD yang stabil tinggi (13,94–14,15) berbanding terbalik dengan *Novelty* Baseline SVD yang menunjukkan tren penurunan dari 11,55 (step 1) ke 11,17 (step 10), artinya penurunan mencerminkan semakin sempitnya ruang eksplorasi item akibat *feedback loop* yang tidak dikoreksi.

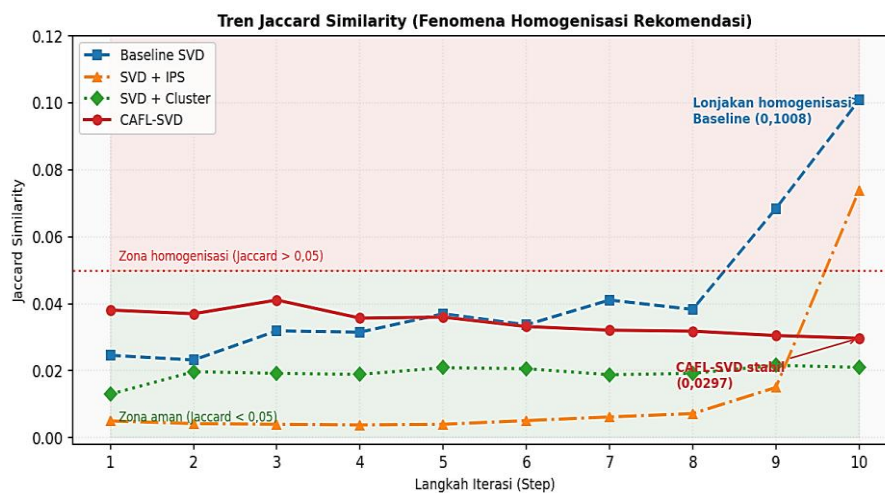
### 3.6 Studi Ablasi

Studi ablasi dilakukan dengan mengevaluasi kontribusi masing-masing komponen CAFL-SVD terhadap penurunan *Gini* coefficient. Tabel 7 menunjukkan hasil ablasi secara kumulatif, dan Gambar 4 menampilkan fenomena homogenisasi yang terlihat jelas pada *Jaccard similarity* dari keempat skenario.

**Tabel 7.** Kontribusi Komponen dalam Studi Ablasi

| Konfigurasi                     | NDCG@5 | Gini   | Penurunan Gini |
|---------------------------------|--------|--------|----------------|
| Base SVD                        | 0,6341 | 0,7215 | Baseline       |
| + Cluster Regularization        | 0,6298 | 0,6543 | -9,3%          |
| + Confounder Balancing          | 0,6189 | 0,4782 | -33,7%         |
| + Iterative Training (CAFL-SVD) | 0,5530 | 0,3601 | -50,1%         |

Hasil ablasi mengungkapkan bahwa *confounder balancing* melalui IPS adalah komponen dengan kontribusi terbesar terhadap *fairness improvement* (33,7% dari total 50,1%). *Cluster regularization* memberikan kontribusi awal yang moderat (9,3%), sedangkan *iterative training* berfungsi sebagai mekanisme *refinement* yang mendorong penurunan *Gini* lebih lanjut (6,4% tambahan). Temuan ini konsisten dengan argumen teoretis bahwa koreksi bias kausal melalui *propensity weighting* adalah mekanisme inti dalam memutus *feedback loop*, sebagaimana diusulkan oleh Krauth et al. (2025).



**Gambar 4.** Tren Jaccard Similarity (Fenomena Homogenisasi Rekomendasi)

### 3.7 Pembahasan

Keberhasilan CAFL-SVD dalam meningkatkan *fairness* tanpa mengorbankan akurasi secara signifikan dapat dijelaskan melalui tiga mekanisme saling melengkapi. Pertama, mekanisme IPS secara langsung mengoreksi bias seleksi dalam data MNAR dengan memberikan bobot lebih tinggi pada observasi dari item yang kurang terekspos, memungkinkan model belajar preferensi pengguna yang lebih representatif. Kedua, *cluster regularization* mencegah faktor laten item populer tumbuh secara tidak proporsional dengan menarik representasi menuju centroid cluster, mendorong item dalam cluster yang sama mendapatkan perhatian setara. Ketiga, *iterative training* memungkinkan model beradaptasi secara dinamis terhadap *feedback* baru, secara bertahap memperbaiki distribusi eksposur dari waktu ke waktu.

Satu temuan kritis yang perlu dicatat adalah degradasi performa SVD+IPS pada langkah akhir (*Gini* = 0,5664 dan *Jaccard* = 0,0738), yang bahkan lebih buruk dari Baseline SVD. Hal ini terjadi karena tanpa *cluster regularization*, bobot IPS yang ekstrem dapat menyebabkan gradien tidak stabil pada data yang semakin bias akibat akumulasi *feedback* simulasi, memperkuat argumen bahwa pendekatan IPS murni memiliki kelemahan varians tinggi pada *setting* bandit *feedback* iteratif sebagaimana diidentifikasi oleh Saito (2020).

Terdapat beberapa keterbatasan yang perlu diakui. Pertama, evaluasi dilakukan menggunakan simulasi bandit *feedback* berbasis *Shadow Model Evaluation*, bukan interaksi pengguna nyata secara *online*. Kedua, *dataset* Book-Crossing memiliki karakteristik rating implisit (nilai 0) yang sangat dominan (65,9%), sehingga hasil penelitian ini paling relevan untuk konteks *implicit feedback* dengan distribusi MNAR ekstrem. Ketiga, penelitian selanjutnya dapat memperkuat validitas dengan menambahkan perbandingan terhadap metode *debiasing* berbasis LLM seperti SPRec (Gao et al., 2025).

Dari perspektif implikasi praktis, hasil penelitian ini memiliki relevansi langsung terhadap ekosistem platform buku digital. Dalam konteks domain buku, sistem yang tidak menangani *feedback loop* secara aktif cenderung semakin mempersempit rekomendasi ke dalam lingkaran *bestseller* mainstream, merugikan penulis-penulis yang baru memulai karir, karya sastra terjemahan dari bahasa minoritas, serta buku-buku akademik dan *niche* yang memiliki pembaca setia namun jumlah interaksinya kecil. CAFL-SVD menunjukkan bahwa dengan mekanisme kausal yang tepat, sistem



rekomendasi dapat sekaligus menjadi relevan bagi pengguna individual sekaligus adil bagi ekosistem konten secara keseluruhan. Nilai *Novelty Score* rata-rata 14,0 yang dihasilkan CAFL-SVD mengindikasikan bahwa rata-rata buku yang direkomendasikan hanya memiliki popularitas 0,006%, artinya sistem secara konsisten memperkenalkan buku-buku yang tidak akan ditemukan pengguna melalui *browsing* konvensional berbasis popularitas. Temuan ini sejalan dengan argumen Chen et al. (2020) bahwa sistem rekomendasi yang baik tidak hanya memaksimalkan *engagement* jangka pendek, melainkan juga bertanggung jawab atas keberlanjutan ekosistem konten jangka panjang. Penelitian mendatang yang mengintegrasikan CAFL-SVD dengan data kontekstual seperti profil genre, ulasan teks, dan riwayat pengarang diharapkan dapat lebih lanjut meningkatkan kemampuan sistem dalam menemukan buku-buku berkualitas yang tersembunyi di balik dominasi *bestseller*.

#### 4. KESIMPULAN

Penelitian ini berhasil mengembangkan dan mengevaluasi CAFL-SVD, *framework collaborative filtering* dengan *causal adjustment* untuk mitigasi bias *feedback loops* dalam sistem rekomendasi buku. Model diimplementasikan dengan mengintegrasikan *Inverse Propensity Scoring* sebagai mekanisme koreksi kausal terhadap bias MNAR, *cluster regularization* berbasis K-Means untuk menstabilkan representasi faktor laten, serta *iterative bandit feedback simulation* yang meniru dinamika deployment nyata. Hasil eksperimen pada Book-Crossing *dataset* dengan 585.579 rating (*sparsity* 99,90%) menunjukkan tiga temuan utama: pertama, CAFL-SVD berhasil menurunkan *Gini coefficient* sebesar 37,7% dari 0,5783 menjadi 0,3601, membuktikan redistribusi eksposur item yang signifikan dari item populer ke item *long-tail*; kedua, relevansi rekomendasi tetap kompetitif dengan NDCG@5 mencapai 0,6207 pada titik terbaik dan 0,5530 pada akhir simulasi, hanya 2,1% di bawah Vanilla SVD tanpa mekanisme *fairness*; dan ketiga, diversitas rekomendasi meningkat dengan penurunan *Jaccard similarity* 22% serta *Novelty Score* stabil di 14,0 yang mengindikasikan konsistensi eksposur item *long-tail*. Studi ablatasi mengkonfirmasi bahwa *confounder balancing* melalui IPS merupakan komponen paling kritis (33,7% kontribusi *Gini reduction*), diikuti *cluster regularization* (9,3%) dan *iterative training* (6,4% tambahan). Penelitian ini membuktikan bahwa pendekatan kausal dapat meningkatkan *fairness* dan relevansi secara bersamaan pada *dataset* buku nyata berskala besar. Untuk penelitian lanjutan, disarankan eksplorasi estimator *Doubly Robust* yang lebih stabil, validasi *cross-domain* pada *dataset* buku lain, implementasi *distributed training* untuk *dataset* skala lebih besar, dan pengembangan mekanisme *auto-tuning hyperparameter* untuk optimasi *fairness-accuracy trade-off* secara otomatis.

#### REFERENCES

- Abuali, M. M., & Ghomeed, T. M. (2024). Utilizing the Artificial Bee Algorithm to Enhance the Accuracy of Book Recommendation Systems: A Case Study on Goodreads dataset. *Libyan Journal of Contemporary Academic Studies*, 54–62. <https://ljas.ly/index.php/ljas/article/view/27>
- Chang, B., Meng, C., Ma, H., Chang, S., Gu, Y., Peng, Y., Feng, J., Zhang, Y., Bi, S., Chi, E. H., & Chen, M. (2024). Cluster Anchor Regularization to Alleviate Popularity Bias in Recommender Systems. *WWW 2024 Companion - Companion Proceedings of the ACM Web Conference*, 151–159. <https://doi.org/10.1145/3589335.3648312>
- Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., & He, X. (2023). Bias and Debias in Recommender System: A Survey and Future Directions. *ACM Transactions on Information Systems*, 41(3). <https://doi.org/10.1145/3564284>
- Gao, C., Chen, R., Yuan, S., Huang, K., Yu, Y., & He, X. (2025). SPRec: Self-Play to Debias LLM-based Recommendation. *WWW 2025 - Proceedings of the ACM Web Conference, 1*, 5075–5084. <https://doi.org/10.1145/3696410.3714524>
- Hussein, M. H., Alsakaa, A. A., & Marhoon, H. A. (2021). Adopting explicit and implicit social relations by SVD++ for recommendation system improvement. *Telkonnika (Telecommunication Computing Electronics and Control)*, 19(2), 471–478. <https://doi.org/10.12928/TELKOMNIKA.v19i2.18149>
- Khenissi, S., & Nasraoui, O. (2020). *Modeling and Counteracting Exposure Bias in Recommender Systems*. <http://arxiv.org/abs/2001.04832>
- Krauth, K., Wang, Y., & Jordan, M. (2025). Breaking Feedback Loops in Recommender Systems with Causal Inference. *ACM Transactions on Recommender Systems*. <https://doi.org/10.1145/3728372>
- Mendoza, M., & Torres, N. (2020). Evaluating content novelty in recommender systems. *Journal of Intelligent Information Systems*, 54(2), 297–316. <https://doi.org/10.1007/s10844-019-00548-x>
- Qin, Z., Chen, S. J., Metzler, D., Noh, Y., Qin, J., & Wang, X. (2020). Attribute-based Propensity for Unbiased Learning in Recommender Systems: Algorithm and Case Studies. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2359–2367. <https://doi.org/10.1145/3394486.3403285>
- Ricci, F., Rokach, L., & Shapira, B. (2022). Recommender Systems: Techniques, Applications, and Challenges. *Recommender Systems Handbook: Third Edition*, 1–35. [https://doi.org/10.1007/978-1-0716-2197-4\\_1](https://doi.org/10.1007/978-1-0716-2197-4_1)
- Saito, Y. (2020a). Asymmetric Tri-training for Debiasing Missing-Not-At-Random Explicit Feedback. *SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 309–318. <https://doi.org/10.1145/3397271.3401114>
- Saito, Y. (2020b). *Doubly Robust Estimator for Ranking Metrics with Post-Click Conversions*. <https://doi.org/10.1145/3397271.3401114>



- Saito, Y., Yaginuma, S., Nishino, Y., Sakata, H., & Nakata, K. (2020). Unbiased recommender learning from missing-not-at-random implicit feedback. *WSDM 2020 - Proceedings of the 13th International Conference on Web Search and Data Mining*, 501–509. <https://doi.org/10.1145/3336191.3371783>
- Verma, V., & Aggarwal, R. K. (2020). A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: empirical and theoretical perspective. *Social Network Analysis and Mining*, 10(1). <https://doi.org/10.1007/s13278-020-00660-9>
- Wang, X., Zhang, R., Sun, Y., & Qi, J. (2019). *Doubly Robust Joint Learning for Recommendation on Data Missing Not at Random*.
- Wang, Y., Ma, W., Zhang, M., Liu, Y., & Ma, S. (2023). A Survey on the Fairness of Recommender Systems. *ACM Transactions on Information Systems*, 41(3). <https://doi.org/10.1145/3547333>
- Wang, Z., Chen, X., Zhou, R., Dai, Q., Dong, Z., & Wen, J.-R. (2022). *Sequential Recommendation with Causal Behavior Discovery*. <http://arxiv.org/abs/2204.00216>
- Xu, Y., Zhuang, F., Wang, E., Li, C., Wu, J., & Fellow, I. (2025). *Learning without Missing-At-Random Prior Propensity-A Generative Approach for Recommender Systems*.
- Yang, L., Wang, C., Cui, Y., Belongie, S., Xuan, Y., & Estrin, D. (2018). Unbiased offline recommender evaluation for missing-not-at-random implicit feedback. *RecSys 2018 - 12th ACM Conference on Recommender Systems*, 279–287. <https://doi.org/10.1145/3240323.3240355>
- Zhang, A., Ma, W., Zheng, J., Wang, X., & Chua, T. S. (2024). Robust Collaborative Filtering to Popularity Distribution Shift. *ACM Transactions on Information Systems*, 42(3). <https://doi.org/10.1145/3627159>
- Zhang, R., Bao, H., Sun, H., Wang, Y., & Liu, X. (2016). Recommender systems based on ranking performance optimization. *Frontiers of Computer Science*, 10(2), 270–280. <https://doi.org/10.1007/s11704-015-4584-1>
- Zhang, Y., Feng, F., He, X., Wei, T., Song, C., Ling, G., & Zhang, Y. (2021). Causal Intervention for Leveraging Popularity Bias in Recommendation. *SIGIR 2021 - Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 11–20. <https://doi.org/10.1145/3404835.3462875>