



KAJIAN LITERATUR COLD START PROBLEM PADA SISTEM REKOMENDASI E-COMMERCE : TANTANGAN DAN STRATEGI PENANGANANNYA

Syaskiah Salsabila, Hasrullah

STMIK Borneo Internasional Balikpapan

syaskiahsalsabila@gmail.com, Hasrullah@permataabadi.co.id

Abstrak

Perkembangan pesat e-commerce mendorong kebutuhan sistem rekomendasi yang mampu memberikan saran relevan meski dengan keterbatasan data awal. Namun, munculnya pengguna baru, item baru, atau ekspansi ke pasar lintas bahasa menyebabkan munculnya *cold-start problem* yang menurunkan akurasi dan kinerja sistem rekomendasi. Kajian ini menggunakan metode *systematic literature review* terhadap berbagai penelitian tahun 2020-2025 yang membahas strategi penanganan *cold-start* pada sistem rekomendasi e-commerce. Hasil analisis menunjukkan bahwa pendekatan yang digunakan mencakup pemanfaatan *side-information*, *graph-based embedding*, *meta-learning*, *reinforcement learning (LTV-aware)*, serta integrasi *retrieval augmentation* dengan *Large Language Model (LLM)*. Setiap metode memiliki keunggulan dan keterbatasan: *side-information* efektif dengan metadata kaya, *graph-based* unggul pada sistem bertaxonomy, *meta-learning* cepat beradaptasi dengan sedikit data, sedangkan *reinforcement learning* dan *LLM augmentation* mampu mengoptimalkan nilai jangka panjang serta meningkatkan relevansi lintas bahasa. Kajian ini menyimpulkan bahwa tren penelitian terkini bergerak ke arah model hibrida yang menggabungkan berbagai pendekatan untuk menciptakan sistem rekomendasi yang adaptif, efisien, dan berkelanjutan.

Kata kunci : *Cold-start problem, recommender system, e-commerce, meta-learning, graph embedding, reinforcement learning, large language model (LLM).*

Abstract

The rapid growth of e-commerce has driven the need for recommendation systems capable of producing relevant suggestions despite limited initial data. However, the emergence of new users, new items, and cross-lingual market expansion often triggers the *cold-start problem*, reducing recommendation accuracy and overall system performance. This study employs a *systematic literature review* method to analyze research published between 2020 and 2025 that addresses strategies for mitigating the *cold-start problem* in e-commerce recommendation systems. The findings reveal five major approaches: leveraging *side-information*, utilizing *graph-based embeddings*, applying *meta-learning* and *reinforcement learning (LTV-aware)* frameworks, and integrating *retrieval augmentation* with *Large Language Models (LLMs)*. Each method offers unique advantages and limitations. *Side-information* approaches are effective with rich metadata, *graph-based* models excel in taxonomy-rich environments, *meta-learning* enables fast adaptation with minimal data, while *reinforcement learning* and *LLM augmentation* optimize long-term value and cross-lingual relevance. The review concludes that current research trends are

Article History

Received: November 2025

Reviewed: November 2025

Published: November 2025

Plagiarism Checker No 765

Prefix DOI : Prefix DOI :

10.8734/Kohesi.v1i2.365

Copyright : Author

Publish by : Kohesi



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)



moving toward hybrid models that combine these methods to build adaptive, efficient, and sustainable recommendation systems.

Keywords: *Cold-start problem, recommender system, e-commerce, meta-learning, graph embedding, reinforcement learning, large language model (LLM).*

Pendahuluan

Perkembangan pesat perdagangan elektronik membuat katalog produk dan pola perilaku pengguna berubah sangat cepat. Di platform e-commerce modern terdapat tiga kondisi yang kerap memicu masalah pada sistem rekomendasi : (1) kemunculan item baru yang belum memiliki interaksi sama sekali (*new item*), (2) pendaftaran pengguna baru yang belum memberi sinyal preferensi (*new user*), dan (3) ekspansi pasar/platform ke bahasa atau wilayah baru yang membuat label-lokal dan sinyal perilaku kurang tersedia. Kondisi-kondisi inilah yang disebut *cold-start problem*.

Cold-start berakibat langsung pada kualitas rekomendasi awal : model yang bergantung pada data historis (collaborative filtering) akan memberikan rekomendasi yang tidak relevan atau terlalu generik, sehingga menurunkan metrik kunci bisnis seperti konversi, engagement, dan retensi. Selain itu, keputusan alokasi impresi pada tahap awal (mis. menampilkan item baru) merupakan trade-off antara eksplorasi (mencari item bernilai tinggi) dan eksploitasi (memaksimalkan CTR jangka pendek). Oleh karena itu, solusi cold-start perlu dikaji tidak hanya dari sisi akurasi offline tetapi juga dari sisi konsekuensi bisnis jangka-pendek dan jangka-panjang.

Secara garis besar, literatur mengembangkan beberapa strategi utama untuk menanggulangi cold-start :

1. **Memanfaatkan side-information** : metadata item, atribut pengguna, dan sinyal sosial/behavioral;
2. **Struktur relasi** : memanfaatkan hierarki atau graph (taxonomy, knowledge graph, social graph) untuk mentransfer informasi ke entitas baru;
3. **Meta-learning / few-shot** : teknik adaptasi cepat dari sedikit contoh;
4. **Reinforcement learning (LTV-aware)** : optimasi keputusan impresi dengan tujuan jangka panjang;
5. **Retrieval & LLM augmentation** : memperkaya konteks query/produk saat sinyal lokal minim.

Kajian ini merangkum temuan kunci di masing-masing jalur tersebut, membandingkan kelebihan dan keterbatasannya, serta merumuskan gap penelitian yang relevan untuk aplikasi e-commerce.

Menurut penelitian yang dilakukan oleh Yujing Wang dkk. (CSRM-LLM, 2025) yang mengusulkan kerangka CSRM-LLM yang memadukan *multilingual LLM*, tugas pembantu terjemahan, Retrieval-based Query Augmentation (RQA), dan multi-round self-distillation untuk menangani *cold-start relevance matching* pada pasar berbahasa berbeda, penelitian ini menunjukkan bahwa kombinasi LLM + retrieval + distillation mampu meningkatkan relevansi pada hari pertama peluncuran pasar baru dan menghasilkan perbaikan metrik produksi (mis. pengurangan defect ratio dan peningkatan conversion rate), sehingga menegaskan potensi besar pemanfaatan model bahasa besar dan teknik augmentasi dalam skenario lintas-bahasa sampai selesai.



Menurut penelitian yang dilakukan oleh Luo Ji dkk. (RL-LTV, 2021) yang memperkenalkan framework RL-LTV untuk mengoptimalkan *lifetime value* (LTV) item melalui pendekatan reinforcement learning (off-policy actor-critic dengan episodic replay dan LSTM) sehingga sistem dapat memilih alokasi impresi awal yang meningkatkan pendapatan kumulatif multi-periode, studi lapangan pada platform besar menunjukkan peningkatan signifikan pada KPI seperti IPV dan GMV untuk cold-start items tanpa merusak metrik keseluruhan, sehingga menekankan pentingnya mengadopsi perspektif jangka-panjang saat merancang strategi eksplorasi untuk item baru sampai selesai.

Menurut penelitian yang dilakukan oleh Ivan Maksimov dkk. (HGE, 2020) yang memperkenalkan Hierarchical Graph Embedding (HGE) untuk memanfaatkan struktur hirarki item (kategori → subkategori → item) dan menghasilkan embedding hirarkis multilevel untuk memperbaiki representasi item baru, penelitian ini menemukan bahwa HGE meningkatkan HR@k dan precision@k pada dataset yang memiliki taxonomy yang kaya, namun efektivitasnya menurun bila kategori berukuran kecil atau tidak terstruktur, sehingga menunjukkan bahwa memanfaatkan ontology/taxonomy adalah pendekatan efektif untuk cold-start item apabila struktur kategori tersedia dan berkualitas sampai selesai.

Menurut penelitian yang dilakukan oleh Shilong Liu dkk. (AMeLU, 2023) yang menggabungkan Model-Agnostic Meta-Learning (MAML) dengan mekanisme attention dalam model AMeLU untuk memungkinkan adaptasi cepat terhadap pengguna atau item baru sekaligus menangkap diversitas minat pengguna melalui attention, evaluasi pada MovieLens, BookCrossing, dan dataset aplikasi menunjukkan bahwa AMeLU mengungguli baseline meta-learners dan model konvensional di scenario cold-start, sehingga kombinasi meta-learning + attention menjadi solusi yang efektif untuk kasus *new user* dengan sedikit sinyal awal sampai selesai.

Menurut penelitian yang dilakukan oleh Jake Snell dkk. (Prototypical Networks dan literatur few-shot) yang memperkenalkan gagasan representasi prototipe kelas dalam few-shot learning, konsep ini diadaptasi dalam rekomendasi untuk membentuk prototipe preferensi pengguna atau kluster item dari sedikit contoh sehingga model dapat menggeneralisasi ke pengguna baru dengan cepat; penerapan prinsip-prinsip few-shot dan prototypical approaches pada domain rekomendasi memperkuat fondasi teoritis bagi metode-metode meta/few-shot yang kini banyak dipakai untuk menangani cold-start sampai selesai.

Metode Penelitian

Penelitian ini menggunakan pendekatan **kajian literatur sistematis (systematic literature review)** untuk menganalisis berbagai studi yang membahas *cold-start problem* pada sistem rekomendasi e-commerce serta strategi yang digunakan untuk mengatasinya. Pendekatan ini dipilih karena mampu memberikan pemahaman komprehensif terhadap perkembangan penelitian di bidang tersebut.

1. Identifikasi Masalah dan Pertanyaan Penelitian

Tahap awal dilakukan dengan mengidentifikasi isu utama, yaitu permasalahan *cold-start* yang memengaruhi kualitas rekomendasi pada sistem e-commerce. Berdasarkan isu tersebut, dirumuskan beberapa pertanyaan penelitian sebagai berikut :



1. Apa penyebab utama terjadinya *cold-start problem* dalam sistem rekomendasi e-commerce?
2. Strategi dan pendekatan apa yang telah dikembangkan untuk mengatasi masalah tersebut?
3. Bagaimana kelebihan, kelemahan, serta peluang pengembangan dari setiap strategi yang telah diusulkan?

2. Pengumpulan Data Literatur

Proses pengumpulan literatur dilakukan melalui penelusuran artikel ilmiah dari basis data akademik seperti **IEEE Xplore**, **ScienceDirect**, **SpringerLink**, dan **Google Scholar**. Kata kunci yang digunakan antara lain "*cold-start problem*," "*recommender system*," "*e-commerce*," "*meta-learning*," "*graph embedding*," dan "*reinforcement learning*." Literatur yang dikaji dibatasi pada publikasi antara tahun 2020 hingga 2025, untuk memastikan relevansi terhadap perkembangan teknologi terkini.

3. Kriteria Inklusi dan Eksklusi

Untuk menjaga kualitas sumber, penelitian ini menetapkan kriteria seleksi sebagai berikut :

1. Kriteria inklusi :

- a. Artikel berbahasa Inggris atau Indonesia.
- b. Penelitian yang secara eksplisit membahas *cold-start problem* pada sistem rekomendasi e-commerce.
- c. Artikel yang mengusulkan atau mengevaluasi pendekatan penanganan *cold-start*.

2. Kriteria eksklusi :

- a. Artikel non-ilmiah seperti blog, laporan industri, atau opini editorial.
- b. Penelitian yang hanya membahas sistem rekomendasi secara umum tanpa konteks *cold-start*.

4. Analisis dan Klasifikasi Literatur

Seluruh literatur yang memenuhi kriteria kemudian dianalisis dan diklasifikasikan berdasarkan pendekatan yang digunakan, meliputi :

1. Pendekatan berbasis **side-information** (metadata item dan atribut pengguna).
2. Pendekatan berbasis **struktur relasi/graph** (taxonomy, knowledge graph, social graph).
3. Pendekatan **meta-learning** atau **few-shot learning**.
4. Pendekatan **reinforcement learning** (LTV-aware).
5. Pendekatan **retrieval dan LLM augmentation**.

Setiap pendekatan dibandingkan dari aspek metodologi, dataset yang digunakan, hasil penelitian, serta kelebihan dan keterbatasannya.



Metode	Core Idea	Dataset Tipikal	Metrik Utama	Kekuatan	Keterbatasan
Side-Information (content / social)	Menggunakan metadata item (judul, kategori, brand) dan atribut pengguna sebagai pengganti interaksi awal.	MovieLens, Amazon, dataset e-commerce industri	NDCG@K, HR@K, MRR	Mudah diimplementasikan jika metadata berkualitas; baik untuk item dengan banyak atribut.	Sensitif terhadap kualitas metadata; perlu ekstraksi fitur; isu privasi pada data sosial.
Graph / Hierarchical (GNN, HGE)	Memanfaatkan relasi antar-entitas (kategori, taxonomy, knowledge graph) untuk mentransfer informasi ke entitas baru.	Amazon, RetailRocket, dataset dengan struktur kategori	HR@K, Precision@K	Menangkap relasi semantik antar-item; efektif untuk cold-item dengan taxonomy.	Memerlukan taxonomy berkualitas; komputasi tinggi; performa turun pada kategori kecil.
Meta-Learning / Few-Shot (MAML, AMeLU)	Melatih model agar cepat beradaptasi dari sedikit contoh (<i>support set</i> kecil).	MovieLens-1M, BookCrossing	MAE, RMSE, NDCG@K, HR@K	Adaptasi cepat untuk <i>new user/item</i> ; efektif dengan sedikit sinyal awal.	Desain <i>meta-task</i> sensitif; tuning dan komputasi kompleks.
Reinforcement Learning (RL-LTV)	Mengoptimalkan impresi dengan tujuan jangka panjang (<i>maximize LTV</i>), bukan sekadar CTR.	Data produksi (Taobao-like)	GMV, IPV, NDCG@K, reward-based metrics	Fokus pada nilai bisnis jangka panjang; cocok untuk eksplorasi strategi rekomendasi.	Kompleksitas tinggi; risiko UX jika eksplorasi tidak dikontrol; sulit diinterpretasi.
Retrieval + LLM Augmentation (RQA, CSRM-LLM)	Memperkaya konteks query/item menggunakan <i>retrieval + multilingual LLM</i> ; distillation ke model produksi.	Cross-market e-commerce corpora	Micro-F1, EM@1, Defect@K	Sangat efektif untuk pasar baru/lintas bahasa; meningkatkan relevansi rekomendasi.	Biaya komputasi tinggi; pipeline kompleks; butuh data terkurasi.
Frequent-Pattern / ARM (DFIS)	Mengekstraksi <i>frequent itemsets</i> per klaster dan mengisi matriks user-item untuk mengurangi <i>sparsity</i> .	MovieLens	Precision, Recall, F1, PLPV	Meningkatkan densitas data dan <i>precision</i> ; memperbaiki <i>cold-start CF</i> .	Mahal untuk skala besar; sensitif terhadap hasil klusterisasi.
Zero-input (ZeroMat)	Memodelkan distribusi rating (Zipf) dan <i>Probabilistic Matrix Factorization</i> untuk estimasi tanpa observasi awal.	MovieLens (small / 1M)	MAE, fairness metrics	Memberi baseline analitis tanpa data; cocok untuk studi teoritis.	Asumsi distribusi tidak selalu berlaku; terbatas pada <i>explicit ratings</i> .

Tabel 1. Ringkasan Pendekatan Penelitian Cold-Start Problem pada Sistem Rekomendasi E-Commerce

5. Sintesis dan Penarikan Kesimpulan

Hasil analisis digunakan untuk menyusun sintesis temuan penelitian terdahulu, merumuskan tren utama dalam penanganan *cold-start problem*, serta mengidentifikasi celah (gap) penelitian yang masih terbuka untuk dikaji lebih lanjut pada konteks sistem rekomendasi e-commerce.

Hasil

Berdasarkan hasil kajian terhadap berbagai literatur yang terpilih, ditemukan bahwa pendekatan penanganan *cold-start problem* dapat dikelompokkan ke dalam enam kategori utama: **side-information**, **graph-based**, **meta-learning**, **reinforcement learning (RL-LTV)**, **retrieval** dan **LLM augmentation**, serta **frequent-pattern mining (ARM/DFIS)**.

Pendekatan berbasis **side-information** mengandalkan metadata pengguna dan item untuk membangun representasi awal. Studi pada dataset seperti *Amazon Product Review* dan *MovieLens* menunjukkan peningkatan performa *NDCG@K* dan *Hit Ratio* bila kualitas metadata terjaga.

Metode **graph-based**, seperti *Hierarchical Graph Embedding (HGE)*, memanfaatkan relasi antar-item dan kategori untuk mentransfer informasi ke entitas baru. Hasil eksperimen menunjukkan efektivitas tinggi dalam mengatasi *cold-item problem* pada sistem dengan taxonomy kaya.

Pendekatan **meta-learning** seperti *AMeLU* dan *Prototypical Network* terbukti mampu mempercepat adaptasi model terhadap pengguna baru yang memiliki sedikit interaksi, dengan performa tinggi pada metrik *RMSE* dan *HR@K*.

Sementara itu, pendekatan **reinforcement learning (RL-LTV)** berfokus pada optimasi jangka panjang, misalnya dengan memaksimalkan *Gross Merchandise Value (GMV)* menggunakan metode *actor-critic* dan *episodic training*.

Pendekatan **retrieval-based** dan **LLM augmentation**, contohnya *CSRM-LLM (Cold Start Recommendation Model with Large Language Model)*, memperlihatkan performa terbaik pada skenario lintas bahasa dan pasar baru, meski menuntut sumber daya komputasi yang besar.



Metode tambahan seperti **frequent-pattern mining (DFIS)** dan **zero-input (ZeroMat)** digunakan sebagai baseline analitis untuk menilai kinerja algoritma dalam kondisi data ekstrem.

Pembahasan

Hasil kajian literatur menunjukkan bahwa tidak ada satu pendekatan tunggal yang mampu menyelesaikan seluruh variasi *cold-start problem* secara komprehensif. Pendekatan berbasis **side-information** memiliki keunggulan pada kemudahan implementasi dan interpretasi hasil, namun rentan terhadap ketidaksempurnaan data atribut dan isu privasi.

Metode **graph-based (GNN/HGE)** terbukti kuat dalam menangkap relasi semantik antar-item, tetapi sangat bergantung pada kelengkapan dan konsistensi taxonomy. Ketika struktur kategori terbatas, performanya menurun signifikan.

Pendekatan **meta-learning** memberikan fleksibilitas tinggi dengan kemampuan adaptasi cepat, menjadikannya solusi ideal untuk platform baru atau produk baru. Namun, desain *meta-task* dan kebutuhan komputasi menjadi hambatan utama.

Reinforcement learning (LTV-aware) merupakan inovasi penting karena mempertimbangkan dampak jangka panjang terhadap metrik bisnis seperti GMV dan *customer lifetime value*. Walaupun menjanjikan, metode ini memerlukan integrasi sistem yang kompleks dan pengawasan eksplorasi agar tidak mengganggu pengalaman pengguna.

Terakhir, integrasi **retrieval dan LLM augmentation** seperti pada *CSRM-LLM* merupakan tren paling mutakhir yang menggabungkan kekuatan pemahaman bahasa alami dengan pembelajaran rekomendasi. Pendekatan ini efektif untuk sistem lintas bahasa dan pasar baru, namun menimbulkan tantangan dari sisi biaya, infrastruktur, serta kebutuhan *fine-tuning* lanjutan.

Berdasarkan keseluruhan analisis, arah penelitian masa depan perlu difokuskan pada **pengembangan model hibrida** yang memadukan *graph learning*, *meta-learning*, dan *retrieval-augmented LLM*, dengan tujuan menciptakan sistem rekomendasi yang adaptif, efisien, serta mampu mempertahankan performa tinggi dalam kondisi data minim.

Kesimpulan

Kajian ini menegaskan bahwa *cold-start problem* merupakan tantangan utama dalam pengembangan sistem rekomendasi e-commerce, terutama ketika berhadapan dengan pengguna atau item baru yang minim interaksi. Berbagai pendekatan telah diusulkan untuk menanganinya, mulai dari pemanfaatan *side-information* hingga penerapan *reinforcement learning* dan *large language models (LLM)*.

Pendekatan berbasis metadata tetap relevan untuk sistem dengan deskripsi produk lengkap, sementara *graph-based* dan *meta-learning* memberikan solusi yang lebih adaptif terhadap konteks baru. Pendekatan *reinforcement learning (LTV-aware)* menghadirkan paradigma baru dengan mengoptimalkan nilai bisnis jangka panjang, sedangkan integrasi *retrieval augmentation* dan *multilingual LLM* memungkinkan peningkatan performa pada pasar lintas bahasa.



Namun, setiap pendekatan memiliki keterbatasan dalam aspek efisiensi, interpretabilitas, dan biaya komputasi. Oleh karena itu, penelitian masa depan disarankan untuk mengembangkan strategi integratif berbasis *multi-modal learning* dan *transfer learning* yang dapat menyeimbangkan akurasi, kecepatan adaptasi, serta keberlanjutan sistem.

Selain itu, diperlukan evaluasi melalui *online A/B testing* agar hasil akademik dapat diterapkan secara langsung pada konteks industri, sekaligus memperkuat keterkaitan antara riset akademis dan implementasi praktis dalam ekosistem e-commerce modern.

Daftar Pustaka

- Al Ghifari, N. T., Sitohang, B., & Saptawati, G. A. P. (2021). *Addressing Cold Start New User in Recommender System Based on Hybrid Approach: A Review and Bibliometric Analysis*. IT Journal Research and Development (ITJRD), 6(1).
- Ji, L., Qin, Q., Han, B., & Yang, H. (2021). *Reinforcement Learning to Optimize Lifetime Value in Cold-Start Recommendation*. Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21). ACM.
- Liu, S., Liu, Y., Zhang, X., Xu, C., He, J., & Qi, Y. (2023). *Improving the Performance of Cold-Start Recommendation by Fusion of Attention Network and Meta-Learning*. Electronics, 12(2), 376.
- Maksimov, I., Rivera-Castro, R., & Burnaev, E. (2020). *Addressing Cold Start in Recommender Systems with Hierarchical Graph Neural Networks*. arXiv preprint arXiv:2009.03455.
- Panteli, A., & Boutsinas, B. (2023). *Addressing the Cold-Start Problem in Recommender Systems Based on Frequent Patterns*. Algorithms, 16(4), 182.
- Snell, J., Swersky, K., & Zemel, R. (2017). *Prototypical Networks for Few-Shot Learning*. Advances in Neural Information Processing Systems (NeurIPS).
- Wang, H. (2021). *ZeroMat: Solving Cold-start Problem of Recommender System with No Input Data*. arXiv preprint arXiv:2112.03084.
- Wang, Y., Chen, Y., Li, H., Xu, C., Luo, Y., Mao, X., Li, C., Du, L., Ma, C., Jiang, Q., Wang, Y., Gao, F., Mo, W., Wen, P., Kumar, S., Park, T., Song, Y., Rajaram, V., Cheng, T., Durgia, S., Kolari, P., & Rajaram, V. (2025). *CSRM-LLM: Embracing Multilingual LLMs for Cold-Start Relevance Matching in Emerging E-Commerce Markets*. Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM '25). ACM.
- Zhang, Y., Maekawa, T., & Hara, T. (2021). *Using Social Media Background to Improve Cold-Start Recommendation Deep Models*. arXiv preprint arXiv:2106.02256.
- Zhou, G., Zhu, X., Song, C., Fan, Y., Zhu, H., Ma, X., Yan, Y., Jin, J., Li, H., & Gai, K. (2018). *Deep Interest Network for Click-Through Rate Prediction*. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18), 1059-1068.