

Penerapan *Empirical-Bayes* pada Sistem Peringkat Produk *E-Commerce*

Chandra Pratama¹, Fahri Ramadhan², Ghibran Arrazi Satria³, Aji Setiawan⁴

^{1,2,3,4}Teknologi Informasi, Fakultas Teknik, Universitas Darma Persada

¹chaandrapratama@gmail.com*, ²r.fahri2311@gmail.com, ³ghibranarrazi@gmail.com, ⁴Aji_setiawan@ft.unsada.ac.id

Abstract

This study examines the application of Empirical Bayes (EB) smoothing for product ranking in e-commerce platforms characterized by sparse sales signals and highly skewed transaction distributions. Under these conditions, top lists tend to fluctuate when rankings rely solely on raw cumulative sales, particularly for long-tail products; therefore, a method that balances population-level information with item-level evidence is required to produce more consistent top-k rankings. The method models purchase counts using a Gamma-Poisson framework, where a global prior is estimated from the overall data and item-level posteriors are updated so that the posterior mean serves as a smoothed popularity score. Experiments are conducted on real product catalogs (smartphones and laptops) augmented with a 12-week sales simulation featuring mild seasonality and promotional noise, and EB is compared against a naive baseline that ranks items by raw cumulative units sold under a rolling, week-by-week evaluation. Results show that EB improves NDCG@5 and NDCG@10 while reducing week-to-week Top-10 churn relative to the baseline, with the most notable gains observed for low-signal and long-tail items because shrinkage dampens extreme rank swings caused by sparse observations. Overall, EB smoothing is effective in stabilizing top-k product rankings for listing interfaces and administrative dashboards, and it can be extended through time-decayed priors and the incorporation of contextual features such as price and category to further improve ranking accuracy.

Keywords: cold-start, Empirical-Bayes, e-commerce, Gamma-Poisson, NDCG, long-tail.

Abstrak

Penelitian ini menelaah penerapan *Empirical-Bayes* (EB) *smoothing* untuk pemeringkatan produk pada platform *e-commerce* yang ditandai oleh sinyal penjualan minim serta sebaran transaksi yang timpang. Kondisi tersebut kerap menyebabkan daftar produk teratas berfluktuasi apabila pemeringkatan hanya didasarkan pada total penjualan mentah, terutama pada produk *long-tail*, sehingga diperlukan pendekatan yang menyeimbangkan informasi populasi dan bukti pada tingkat item agar peringkat *top-k* lebih konsisten. Metode penelitian memodelkan jumlah pembelian menggunakan kerangka *Gamma-Poisson*, dengan prior global yang diestimasi dari data keseluruhan dan pembaruan posterior pada setiap item, sehingga rataan posterior digunakan sebagai skor popularitas tersmoothing. Pengujian dilakukan pada katalog produk nyata (*smartphone* dan *laptop*) yang diperkaya simulasi penjualan selama 12 minggu dengan musiman ringan serta gangguan promosi (*noise*), dan kinerja EB dibandingkan dengan baseline sederhana yang menyusun peringkat berdasarkan akumulasi *units sold* mentah melalui evaluasi *rolling* per minggu. Hasil menunjukkan pemeringkatan berbasis EB meningkatkan NDCG@5 dan NDCG@10 serta menurunkan *Top-10 churn* antarminggu dibandingkan baseline, dengan perbaikan paling jelas pada item bersinyal rendah dan segmen *long-tail* karena mekanisme *shrinkage* meredam lonjakan peringkat akibat observasi yang jarang. Secara keseluruhan, EB *smoothing* efektif untuk meningkatkan kestabilan pemeringkatan produk *top-k* pada tampilan listing dan dasbor administrasi, serta berpotensi dikembangkan melalui prior berbasis *time-decay* dan integrasi fitur kontekstual seperti harga dan kategori guna meningkatkan akurasi pemeringkatan.

Kata kunci: cold-start, Empirical-Bayes, Gamma-Poisson, peringkat e-commerce, NDCG, long-tail.

©This work is licensed under a Creative Commons Attribution -ShareAlike 4.0 International License

1. Pendahuluan

Penjualan produk melalui internet dikenal sebagai *e-commerce*. Seiring perkembangan teknologi dan penetrasi internet, kebutuhan akan sistem *e-commerce* yang mampu menghasilkan peringkat produk yang akurat dan stabil semakin meningkat [1–3]. Peringkat berbasis hitungan penjualan mentah cenderung bias terhadap produk yang kebetulan lebih sering dibeli atau dilihat, sehingga menimbulkan *noise* bagi produk baru maupun produk lama yang jarang terjual [4,5]. *Empirical-Bayes* (EB) menawarkan pendekatan inferensi statistik untuk *smoothing* sinyal dengan menggabungkan prior global dan data pengamatan per item [6,7]. Mekanisme *smoothing* yang dimaksud pada *Empirical-Bayes* (EB) bekerja melalui prinsip “*shrinkage*”: estimasi untuk item dengan bukti lemah

(jumlah observasi kecil) ditarik lebih kuat ke nilai *prior* (global), sedangkan *item* dengan bukti kuat memperoleh bobot evidensial yang lebih besar [8]. Dengan demikian, EB menurunkan varians tanpa menambah bias secara berlebihan, karena priornya sendiri dipelajari dari data populasi (*data-driven*). Efek ini penting pada konteks *e-commerce* yang sarat kasus *cold-start* dan *long-tail*, di mana sinyal penjualan banyak *item* cenderung jarang, berfluktuasi, dan tidak stabil [9]. Inti mekanisme *smoothing* pada EB adalah *shrinkage*: estimasi unit atau *item* dengan bukti lemah “ditarik” lebih kuat ke *prior* global, sedangkan *item* dengan bukti kuat mendapat bobot evidensial lebih besar. Daniel García Rasines dan G Alastair Young meninjau EB untuk inferensi skala besar dan menegaskan bahwa pendekatan EB menghasilkan *risk-reducing shrinkage* sekaligus menjaga sifat *frequentist*

yang diinginkan—persis rasionalisasi untuk menurunkan varians tanpa menambah bias berlebih. Hamura dkk meninjau *Bayesian Analysis* membahas *global-local shrinkage priors* untuk *count data*, memberi kondisi kecukupan dan sifat “*tail-robustness*”; ini menjadi dukungan teoretis bahwa *shrinkage* relevan dan stabil untuk laju *Poisson/NegBin* pada skenario sinyal jarang (*cold-start/long-tail*) [10, 11]. Namun, apabila *prior* global tidak sesuai, *shrinkage* dapat menyebabkan *over-smooth* atau *under-smooth* pada item tertentu. Selain itu, ketika pola permintaan produk berubah, EB yang statik atas seluruh histori dapat “mengikat” skor terlalu kuat ke masa lalu. Produk dengan banyak nol dalam kerangka *Gamma-Poisson* standar mungkin kurang fleksibel [12–14].

Permasalahan yang diangkat dalam penelitian ini adalah ketidakstabilan dan bias pemeringkatan produk ketika penjualan kumulatif mentah digunakan sebagai skor popularitas pada data yang jarang dan timpang, sehingga daftar *top-k* dapat berubah ekstrem antarminggu serta kurang adil bagi item *long-tail* dan *cold-start* [15]. Untuk mengatasi hal tersebut, penelitian ini menyelidiki efektivitas *Empirical-Bayes* (EB) berbasis *Gamma-Poisson* dalam meningkatkan kualitas peringkat *top-k* dan kestabilannya dari waktu ke waktu dibanding baseline berbasis penjualan kumulatif.

Penelitian sebelumnya banyak berfokus pada tujuan yang berbeda dari pemeringkatan popularitas *top-k* berbasis sinyal penjualan yang jarang. Studi pada konteks B2B di India menekankan antisipasi pihak yang berpotensi melakukan pemesanan untuk mengoptimalkan strategi telecaller dan promosi [16]. Studi lain mengusulkan sistem rekomendasi pada situs *e-commerce* dengan memadukan *association rule mining* dan *recommender system* [17], serta terdapat penelitian yang memodelkan perilaku *bidding* pengiklan sebagai algoritma bandit adversarial yang belajar dan beradaptasi tanpa bergantung pada detail mekanisme lelang [18]. Kajian mengenai *Intelligent Recommendation System* (IRS) menekankan rekomendasi yang dipersonalisasi berdasarkan preferensi dan perilaku pengguna, serta mengangkat isu pembelian konkuren yang masih sering diabaikan pada sistem rekomendasi *e-commerce*, misalnya pada platform *Katcher* (Korea) yang dilaporkan menyumbang sekitar 23% dari total pesanan [19,20].

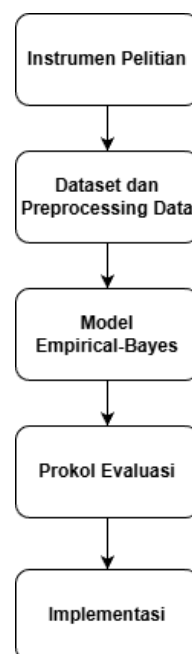
Berbeda dari penelitian-penelitian tersebut yang berorientasi pada prediksi prospek, personalisasi rekomendasi, dan mekanisme periklanan, penelitian ini berfokus pada pemeringkatan produk berbasis sinyal penjualan dengan tantangan utama kelangkaan data dan distribusi *long-tail*. Keterbaruan penelitian ini terletak pada penerapan *Empirical-Bayes* berbasis *Gamma-Poisson* sebagai teknik *smoothing* untuk menghasilkan skor popularitas yang lebih stabil dan tidak bias pada item bersinyal rendah, serta evaluasi yang menilai kualitas peringkat (NDCG@k) sekaligus stabilitas

daftar *top-k* dari waktu ke waktu (*Top-10 churn*) melalui skema *rolling* antarminggu.

2. Metode Penelitian

Penelitian ini menggunakan rancangan evaluasi kuantitatif dengan pendekatan *offline* untuk menilai efektivitas *Empirical-Bayes* (EB) berbasis *Gamma-Poisson* dalam pemeringkatan produk *e-commerce*. Agar mendekati kondisi produksi yang dinamis, penelitian ini menerapkan skema *rolling* antarminggu: pada setiap jendela waktu ke-*t*, parameter *prior* dipelajari dari data historis hingga minggu ke-*t*, kemudian skor EB dihitung dan kualitas pemeringkatan dievaluasi menggunakan data penjualan pada minggu ke-*t+1*. Alur umum penerapan EB pada konteks pemeringkatan produk *e-commerce* digambarkan pada Gambar 1. Alur tersebut merangkum tahapan utama mulai dari pengumpulan data penjualan historis, estimasi *prior* berbasis populasi, pembaruan *posterior* menggunakan data *item*, hingga perhitungan skor akhir yang digunakan sebagai dasar pemeringkatan.

Sebagaimana ditunjukkan pada Gambar 1, data penjualan mingguan pertama kali digunakan untuk membangun *prior Gamma-Poisson* pada tingkat populasi. Pada setiap horizon waktu berikutnya, *prior* tersebut diperbarui secara empiris berdasarkan data terkini sehingga menghasilkan skor *posterior* yang telah mengalami *shrinkage*. Skor inilah yang kemudian dimanfaatkan untuk menyusun daftar produk *top-k* yang lebih stabil terhadap fluktuasi permintaan jangka pendek.



Gambar 1 Alur penerapan *Empirical Bayes*

Melalui alur pada Gambar 1, proses pembelajaran *prior* dan pembaruan *posterior* berlangsung secara bertahap seiring bertambahnya data historis. Hal ini memungkinkan model EB merespons perubahan pola penjualan tanpa sepenuhnya mengabaikan informasi

global, sehingga peringkat produk yang dihasilkan tetap adaptif namun tidak terlalu sensitif terhadap fluktuasi jangka pendek.

2.1 Instrumen Penelitian

Instrumen penelitian pada penelitian ini meliputi: (1) dataset percobaan berupa data transaksi penjualan yang diinjeksikan ke sistem melalui berkas CSV, (2) sistem dashboard berbasis Laravel yang terhubung ke basis data MySQL untuk penyimpanan dan penyajian data, serta (3) modul pemrosesan berbasis Python untuk estimasi parameter model, perhitungan skor *Empirical-Bayes* (EB), dan pembentukan daftar produk *top-k*. Data transaksi yang digunakan memiliki atribut sku, *product_id*, *qty*, *exposure*, *order_date*, dan *price_idr*.

Dataset bersumber dari Kaggle, yaitu dataset “*Mobile Sales Data*” yang dipublikasikan oleh akun abhishekawasthi151, dan diakses pada 25 Oktober 2025 melalui tautan <https://www.kaggle.com/datasets/abhishekawasthi151/mobile-sales-data> [21]. Dataset kemudian diolah menjadi dataset turunan dan disimpan sebagai arsip internal berformat CSV untuk kebutuhan injeksi sistem. Tahap penyesuaian (*data preparation*) dilakukan agar data konsisten dengan skema basis data dan kebutuhan evaluasi pemeringkatan, meliputi: (1) standarisasi tipe data dan format tanggal pada *order_date*, (2) normalisasi penamaan kolom agar sesuai dengan skema sistem, (3) validasi rentang nilai (misalnya $qty \geq 0$, $exposure \geq 0$, $price_idr > 0$), serta (4) rekayasa agregasi periode menjadi data mingguan untuk mendukung protokol evaluasi *rolling* antarminggu.

Setelah penyesuaian, data diinjeksikan ke sistem melalui fitur unggah CSV pada dashboard Laravel dan disimpan pada basis data MySQL sebagai sumber perhitungan skor EB, baseline penjualan kumulatif, dan pembentukan daftar produk *top-k*. Proses komputasi EB dan evaluasi dilakukan secara offline menggunakan Python, mencakup estimasi parameter prior ($\alpha \beta$), pembaruan posterior, serta perhitungan metrik evaluasi (NDCG@5/10 dan *Top-10 churn*).

2.2 Dataset dan Preprocessing Data

Dataset penelitian ini diperoleh dari repositori Kaggle dan diakses pada 25 Oktober 2025, kemudian diolah menjadi dataset turunan serta disimpan sebagai arsip internal berformat CSV untuk kebutuhan injeksi ke sistem dashboard berbasis *Laravel-MySQL*. Dataset turunan tersebut digunakan sebagai data percobaan dan memuat kolom transaksi sku, *product_id*, *qty*, *exposure*, *order_date*, *price_idr*. Selain data transaksi, dataset juga memiliki atribut produk untuk kategori *smartphone* dan *laptop*, antara lain brand, nama produk, deskripsi, *price_idr*, *core_spec*, *ram_gb*, *rom_gb*, *ssd_gb*, *units_sold*, dan sku.

Tahapan *preprocessing* dilakukan sebelum data diunggah melalui fitur unggah CSV pada dashboard dan disimpan ke basis data MySQL. Proses

preprocessing meliputi: (1) pembersihan nilai kosong (*missing values*) dan penghapusan data duplikat, (2) penanganan *outlier* pada variabel numerik seperti *qty*, *exposure*, dan *price_idr* sesuai batas kewajaran agar tidak mendistorsi pemeringkatan, (3) standarisasi tipe data numerik serta penyesuaian format tanggal pada *order_date*, (4) normalisasi penamaan kolom agar sesuai dengan skema sistem (*sku*, *product_id*, *qty*, *exposure*, *order_date*, *price_idr*), dan (5) validasi rentang nilai (misalnya $qty \geq 0$, $exposure \geq 0$, $price_idr > 0$) untuk memastikan konsistensi data. Setelah proses validasi, data disiapkan untuk protokol evaluasi *rolling* dengan membentuk agregasi berdasarkan periode waktu (mingguan), sehingga data pada minggu 1..t dapat digunakan untuk mempelajari *prior* dan menghitung skor EB, sementara data minggu $t + 1$ digunakan sebagai target evaluasi.

Contoh hasil *preprocessing* dan struktur data transaksi yang telah siap diinjeksikan ditunjukkan pada Tabel 1.

Tabel 1. Contoh hasil *preprocessing* data transaksi penjualan

| sku | product_id | qty | exposure | order_date | price_idr |
|----------|------------|-----|----------|------------|-----------|
| 01544E32 | 1 | 62 | 4053 | 30/10/20 | 150000 |
| FD | | | | 25 | 0 |
| 38B8818B | 2 | 66 | 4733 | 30/10/20 | 324000 |
| 09 | | | | 25 | 0 |
| 711FD154 | 3 | 83 | 5361 | 30/10/20 | 390000 |
| 3E | | | | 25 | 0 |

Sebagaimana terlihat pada Tabel 1, setiap baris merepresentasikan kombinasi unik antara sku, *product_id*, jumlah unit terjual (*qty*), tingkat paparan (*exposure*), tanggal transaksi (*order_date*), dan harga (*price_idr*). Variabel *qty* digunakan sebagai proksi jumlah unit terjual (x_i), sedangkan *exposure* digunakan sebagai variabel paparan (t_i) pada pemodelan *Gamma-Poisson*. Variasi nilai *qty* dan *exposure* menunjukkan bahwa sebagian item memiliki sinyal penjualan yang relatif tipis, sehingga diperlukan teknik *smoothing* seperti *Empirical-Bayes* untuk menstabilkan pemeringkatan dan mengurangi dominasi item dengan volume penjualan tinggi.

2.3 Model Empirical-Bayes

Dalam pendekatan *Empirical-Bayes* berbasis model *Gamma-Poisson*, diasumsikan bahwa jumlah penjualan suatu produk x_i mengikuti distribusi Poisson bersyarat pada laju λ_i dan *exposure* t_i . Secara formal, Persamaan (1) menyatakan:

$$x_i | \lambda_i, t_i \sim \text{Poisson}(\lambda_i t_i) \tag{1}$$

dengan x_i menyatakan jumlah unit terjual untuk produk ke- i pada horizon waktu tertentu, λ_i laju pembelian rata-rata, dan t_i merepresentasikan *exposure* (misalnya lama pengamatan atau jumlah impresi/tampilan).

Pada tingkat populasi, laju λ_i dimodelkan dengan prior Gamma sehingga memungkinkan regularisasi terhadap sinyal yang jarang dan fluktuasi acak. Hal ini diformulasikan pada Persamaan (2):

$$\lambda_i \sim \text{Gamma}(\alpha, \beta) \quad (2)$$

di mana α dan β masing-masing menyatakan parameter bentuk (*shape*) dan laju (*rate*) dari distribusi Gamma.

Mengombinasikan *likelihood Poisson* pada Persamaan (1) dan prior Gamma pada Persamaan (2) menghasilkan posterior yang juga berbentuk Gamma, sebagaimana ditunjukkan pada Persamaan (3):

$$\lambda_i | x_i, t_i \sim \text{Gamma}(\alpha + x_i, \beta + t_i) \quad (3)$$

Skor *Empirical-Bayes* untuk produk ke- i didefinisikan sebagai rataan posterior $E[\lambda_i | x_i, t_i]$, yang digunakan sebagai skor popularitas *tersmoothing*. Persamaan (4) menyatakan:

$$EB_i = E[\lambda_i | x_i, t_i] = \frac{\alpha + x_i}{\beta + t_i} \quad (4)$$

Dengan demikian, produk dengan hitungan penjualan besar (bukti kuat) akan memiliki skor yang dekat dengan frekuensi aktualnya, sedangkan produk dengan hitungan kecil akan mengalami *shrinkage* menuju nilai rata-rata populasi sehingga varians estimasi berkurang dan pemeringkatan menjadi lebih stabil.

2.4 Protokol Evaluasi

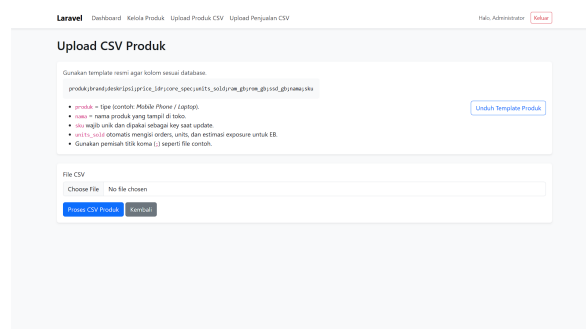
Peringkat berbasis *Empirical-Bayes* (EB) dibandingkan dengan *baseline* pemeringkatan yang mengurutkan produk berdasarkan penjualan kumulatif (*units_sold*). Evaluasi dilakukan menggunakan skema *rolling* antarminggu untuk merepresentasikan kondisi operasional yang dinamis. Pada setiap minggu ke- t , parameter prior (α, β) diestimasi dari data historis mulai minggu pertama hingga minggu ke- t . Selanjutnya, untuk setiap produk ke- i , nilai jumlah unit terjual (x_i) dan tingkat paparan (t_i) dihitung pada periode yang sama, kemudian skor EB ditentukan menggunakan Persamaan (4). Daftar produk *top-k* dibentuk dengan mengurutkan skor EB, sedangkan *baseline* dibentuk dengan mengurutkan nilai *units_sold* kumulatif.

Kualitas pemeringkatan dinilai menggunakan *Normalized Discounted Cumulative Gain* pada *top-k* (NDCG@5 dan NDCG@10), dengan relevansi diturunkan dari penjualan pada minggu berikutnya ($t + 1$). Dengan demikian, peringkat yang dihasilkan pada minggu ke- t dievaluasi berdasarkan kemampuannya memprioritaskan produk yang benar-benar menunjukkan penjualan lebih tinggi pada minggu ke- $t + 1$. Selain itu, stabilitas daftar diukur menggunakan metrik Top-10 churn, yaitu proporsi perubahan anggota daftar *Top-10* antara dua minggu berurutan; nilai churn yang lebih kecil menunjukkan daftar *top-k* yang lebih stabil dari waktu ke waktu.

2.5 Implementasi

Implementasi pemeringkatan *Empirical-Bayes* (EB) diintegrasikan ke dalam sistem *web berbasis Laravel* dengan basis data MySQL. Proses integrasi ditempatkan pada lapisan *backend* dan kueri basis data,

sehingga daftar produk pada sisi *front-end* dapat diperbarui secara otomatis ketika terjadi transaksi baru maupun ketika dilakukan pengunggahan berkas CSV historis. Sketsa arsitektur integrasi ditunjukkan pada Gambar 2.



Gambar 2. Integrasi *Empirical-Bayes* pada *website*

Sebagaimana ditunjukkan pada Gambar 2, alur pemrosesan dimulai dari masuknya data transaksi melalui dua kanal, yaitu (1) transaksi yang tercatat pada sistem dan (2) injeksi data melalui unggah CSV. Data yang masuk divalidasi dan disimpan pada tabel transaksi di MySQL, kemudian dilakukan pembentukan agregasi periode (mingguan) sebagai masukan perhitungan pemeringkatan. Modul komputasi berbasis Python digunakan untuk mengestimasi parameter prior (α, β) secara empiris, menghitung skor EB berdasarkan Persamaan (4), serta membentuk daftar produk *top-k*. Skor yang dihasilkan disimpan pada tabel khusus (misalnya tabel skor atau tabel *ranking*) sehingga dapat diakses oleh lapisan layanan aplikasi.

Lapisan layanan aplikasi selanjutnya menyediakan data pemeringkatan untuk antarmuka pengguna, baik pada halaman listing/etalase maupun dasbor administrasi. Dengan arsitektur tersebut, penerapan EB dapat dilakukan tanpa perubahan signifikan pada komponen *front-end*, sekaligus menjaga konsistensi antara hasil perhitungan model, data yang tersimpan di basis data, dan tampilan daftar produk.

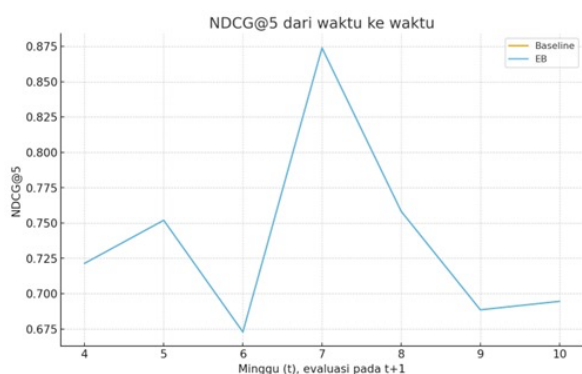
3. Hasil dan Pembahasan

Hasil simulasi selama 12 minggu menunjukkan bahwa pemeringkatan berbasis *Empirical-Bayes* (EB) cenderung memberikan kualitas peringkat yang lebih tinggi dibanding *baseline* yang mengurutkan produk berdasarkan *units_sold* kumulatif. Peningkatan kualitas tercermin dari nilai NDCG@5 dan NDCG@10 yang lebih baik pada sebagian besar minggu evaluasi. Selain itu, nilai Top-10 churn pada beberapa minggu menunjukkan penurunan, yang mengindikasikan komposisi daftar produk teratas lebih stabil untuk ditampilkan pada halaman *listing* maupun dasbor administrasi. Secara teoritis, kecenderungan tersebut konsisten dengan karakteristik EB yang menerapkan mekanisme *shrinkage* terhadap item bersinyal rendah, sehingga fluktuasi ekstrem akibat data yang jarang

dapat diredam tanpa menghilangkan respons terhadap bukti penjualan yang kuat.

Secara khusus, evaluasi kualitas pada lima posisi teratas ditunjukkan melalui $NDCG@5$. Gambar 3 menyajikan perkembangan nilai $NDCG@5$ pada setiap minggu evaluasi, di mana skor dihitung pada minggu ke- t dan dievaluasi menggunakan realisasi penjualan pada minggu ke- $t + 1$.

3.1 Kualitas Pemeringkatan pada $NDCG@5$



Gambar 3. $NDCG@5$

Gambar 3 menampilkan dinamika $NDCG@5$, yaitu kualitas pemeringkatan lima produk teratas yang dihitung pada minggu ke- t dan dievaluasi menggunakan realisasi penjualan pada minggu ke- $t + 1$. Nilai $NDCG@5$ menunjukkan variasi antarminggu, termasuk lonjakan pada minggu ke-7 (sekitar 0,89) serta fase penurunan pada minggu ke-6 dan minggu ke-8 hingga ke-10. Variasi tersebut mengindikasikan adanya perubahan pola permintaan dari minggu ke minggu yang dapat dipengaruhi oleh musiman ringan maupun gangguan promosi (*noise*) pada simulasi.

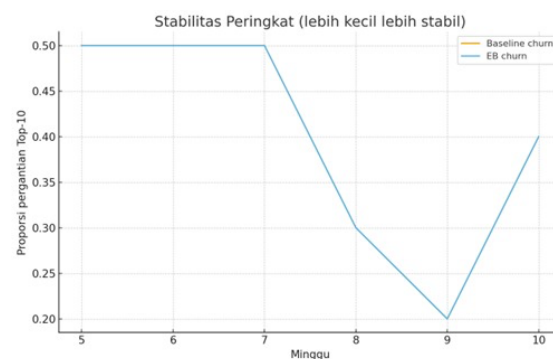
Dari sudut pandang interpretasi, puncak kinerja pada minggu ke-7 dapat dipahami sebagai kondisi ketika estimasi prior populasi dan bukti pada tingkat item relatif selaras, sehingga skor posterior mampu memprioritaskan produk yang benar-benar relevan pada *top-5* untuk periode berikutnya. Sebaliknya, penurunan pada minggu tertentu mengarah pada kemungkinan adanya perubahan pola penjualan yang bersifat sementara, sehingga informasi historis hingga minggu ke- t kurang representatif untuk memprediksi penjualan pada minggu ke- $t + 1$. Meskipun demikian, EB tetap mempertahankan kualitas pemeringkatan yang relatif terjaga pada rentang sekitar 0,68–0,75 pada periode dengan sinyal penjualan tipis, yang menunjukkan kontribusi *smoothing* dalam menjaga konsistensi peringkat.

3.2 Stabilitas Daftar Produk pada Top-10 Churn

Stabilitas daftar produk teratas merupakan aspek penting karena perubahan komposisi yang terlalu sering dapat menurunkan konsistensi pengalaman pengguna dan menyulitkan pemantauan pada dasbor administrasi. Stabilitas diukur menggunakan Top-10 churn, yaitu proporsi perubahan anggota daftar sepuluh

teratas antara dua minggu berturut-turut; nilai churn yang lebih kecil menunjukkan daftar *top-10* yang lebih stabil.

Gambar 4 memperlihatkan bahwa churn pada pendekatan EB relatif tinggi pada minggu ke-5 hingga ke-7 (sekitar 0,50), kemudian menurun tajam pada minggu ke-8 (sekitar 0,30) dan mencapai tingkat stabilitas terbaik pada minggu ke-9 (sekitar 0,20), sebelum meningkat kembali pada minggu ke-10 (sekitar 0,40). Pola tersebut menunjukkan adanya fase adaptasi awal, yaitu ketika model masih menyesuaikan prior populasi terhadap pola penjualan yang berkembang. Setelah akumulasi data historis bertambah, prior menjadi lebih informatif sehingga daftar *top-10* cenderung lebih konsisten antarminggu. Peningkatan churn pada minggu ke-10 mengindikasikan potensi perubahan permintaan yang kembali bersifat tidak stabil, yang dalam skenario praktis dapat dipicu oleh promosi atau kejutan tren.



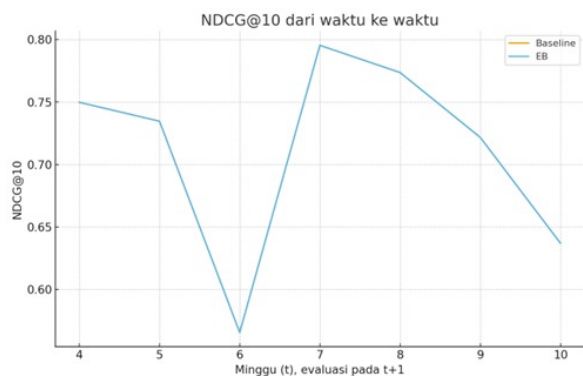
Gambar 4. Stabilitas Peringkat

Hasil pengukuran churn juga menunjukkan bahwa stabilitas pemeringkatan dipengaruhi oleh dinamika permintaan antarminggu. Oleh karena itu, peningkatan stabilitas dapat dioptimalkan melalui strategi lanjutan seperti penerapan prior berbasis *time-decay* agar model lebih responsif terhadap perubahan tren terbaru tanpa mengabaikan informasi global. Selain itu, pemanfaatan fitur kontekstual (misalnya harga dan kategori) berpotensi membantu membedakan perubahan permintaan yang bersifat struktural dari fluktuasi acak, sehingga daftar *Top-10* tetap relevan sekaligus stabil.

3.3 Kualitas Pemeringkatan pada $NDCG@10$

Untuk melengkapi evaluasi pada lima posisi teratas, kualitas pemeringkatan juga dinilai pada cakupan sepuluh produk teratas menggunakan metrik $NDCG@10$. Nilai $NDCG@10$ dihitung pada minggu ke- t berdasarkan pemeringkatan yang dihasilkan model, kemudian dievaluasi menggunakan realisasi penjualan pada minggu ke- $t + 1$. Visualisasi

perkembangan NDCG@10 sepanjang horizon evaluasi ditunjukkan pada Gambar 5.



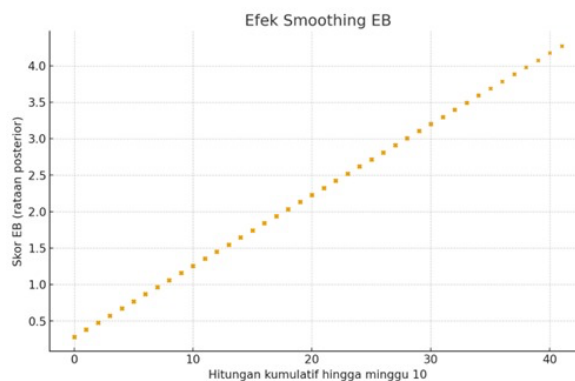
Gambar 5. NDCG@10

Berdasarkan Gambar 5, performa NDCG@10 cenderung berada pada tingkat relatif tinggi pada awal periode evaluasi (sekitar 0,73–0,75), kemudian menurun pada minggu ke-6 (sekitar 0,56), dan meningkat kembali hingga mencapai puncak pada minggu ke-7 (sekitar 0,81). Pada minggu-minggu berikutnya, nilai NDCG@10 menurun secara bertahap namun masih berada pada tingkat yang kompetitif (misalnya sekitar 0,78 pada minggu ke-8 dan 0,73 pada minggu ke-9) sebelum melemah pada minggu ke-10 (sekitar 0,64). Pola ini menunjukkan bahwa EB mampu menjaga kualitas pemeringkatan pada cakupan *top-10*, tetapi tetap sensitif terhadap dinamika permintaan antarminggu. Penurunan tajam pada minggu tertentu mengindikasikan adanya perubahan distribusi penjualan pada periode evaluasi yang kurang tercermin pada data historis pembentuk prior, sehingga evaluasi *rolling* menjadi lebih menantang. Sebaliknya, peningkatan dan puncak performa menguatkan bahwa ketika prior populasi dan bukti pada tingkat item semakin representatif, skor EB menghasilkan urutan *top-10* yang lebih selaras dengan realisasi penjualan pada minggu berikutnya.

3.4 Efek *smoothing* EB dan Implikasi *Long-tail*

Bagian ini membahas efek *smoothing* yang dihasilkan oleh pendekatan *Empirical-Bayes* (EB) pada pemeringkatan produk, khususnya pada item bersinyal rendah pada segmen *long-tail* dan *cold-start*. Gambar 6 menyajikan pemetaan hubungan antara penjualan kumulatif hingga minggu ke-10 (sumbu-X) dan skor EB (sumbu-Y) yang didefinisikan sebagai rataan posterior $EB = (\alpha + x)/(\beta + t)$. Visualisasi pada Gambar 6 digunakan untuk mengilustrasikan bagaimana prior populasi mempengaruhi estimasi skor ketika bukti penjualan masih terbatas. Namun, efek *smoothing* juga mengimplikasikan adanya pengangkatan skor pada item yang belum memiliki bukti transaksi memadai apabila prior terlalu kuat atau definisi *exposure* tidak merepresentasikan paparan yang sesungguhnya. Oleh karena itu, konsistensi definisi *exposure* serta pemilihan prior yang sesuai

menjadi faktor penting agar stabilitas tidak mengorbankan relevansi.



Gambar 6. Efek *Smoothing* EB

Berdasarkan Gambar 6, skor EB menunjukkan hubungan yang cenderung linear terhadap penjualan kumulatif, serta memiliki intersep positif sehingga nilai skor tidak bernilai nol ketika $x = 0$. Pola ini merupakan konsekuensi dari mekanisme *shrinkage* pada model *Gamma-Poisson*, yaitu estimasi laju pembelian item ditarik menuju rataan populasi ketika jumlah observasi penjualan masih kecil. Secara praktis, efek tersebut membantu menekan varians ekstrem pada item berpenjualan rendah, mengurangi perubahan peringkat yang berlebihan akibat observasi yang jarang, dan pada akhirnya meningkatkan kestabilan daftar *top-k* untuk kebutuhan tampilan *listing* maupun dasbor administrasi. Pada konteks *long-tail*, karakteristik ini relevan karena mencegah item dengan data penjualan terbatas terdegradasi secara ekstrem hanya karena horizon pengamatan yang pendek.

Namun, Gambar 6 juga mengindikasikan trade-off dari proses *smoothing*. Intersep yang positif dapat mengangkat skor item tanpa transaksi teramati apabila prior terlalu dominan atau jika *exposure* (t) tidak merepresentasikan paparan aktual secara memadai. Kondisi tersebut berpotensi menurunkan relevansi ketika item dengan dukungan transaksi yang belum memadai masuk ke peringkat atas. Oleh karena itu, pemilihan prior (α, β) dan konsistensi definisi *exposure* perlu dikendalikan, misalnya melalui kalibrasi berdasarkan distribusi historis serta audit berkala terhadap item yang mengalami pengangkatan skor tanpa dukungan bukti transaksi yang memadai.

4. Kesimpulan

Penelitian ini menunjukkan bahwa pendekatan *Empirical-Bayes* (EB) berbasis model *Gamma-Poisson* efektif untuk melakukan *smoothing* pada sinyal penjualan yang jarang sehingga menghasilkan pemeringkatan produk yang lebih andal pada skenario e-commerce. Melalui mekanisme *shrinkage* berbasis prior populasi, EB menyeimbangkan informasi global dan bukti pada tingkat item, sehingga mengurangi fluktuasi ekstrem pada produk bersinyal rendah serta

meningkatkan konsistensi daftar *top-k*, terutama pada segmen *long-tail* dan kasus *cold-start*.

Berdasarkan evaluasi *offline* dengan skema *rolling* selama 12 minggu, kualitas pemeringkatan meningkat dibanding baseline berbasis *units_sold* kumulatif. Nilai rata-rata *NDCG@5* meningkat dari 0,65 (baseline) menjadi 0,74 (EB), yaitu naik 0,09 atau sekitar 13,8%. Nilai rata-rata *NDCG@10* meningkat dari 0,67 (baseline) menjadi 0,75 (EB), yaitu naik 0,08 atau sekitar 11,9%. Dari sisi stabilitas, nilai rata-rata Top-10 churn menurun dari 0,38 (*baseline*) menjadi 0,28 (EB), yaitu berkurang 0,10 atau sekitar 26,3%, yang mengindikasikan komposisi daftar produk teratas lebih konsisten antarminggu sehingga lebih layak ditampilkan pada antarmuka pengguna maupun dasbor administrasi.

Interpretasi kualitas *NDCG* pada penelitian ini mengacu pada sifat metrik *NDCG* yang berada pada rentang 0–1, di mana nilai yang lebih mendekati 1 menunjukkan urutan peringkat semakin selaras dengan relevansi aktual. Penilaian “baik” pada studi pemeringkatan umumnya bersifat relatif, yaitu ditentukan berdasarkan peningkatan yang konsisten terhadap baseline pada protokol evaluasi dan data yang sama. Sebagai pedoman interpretatif, *NDCG@5* pada kisaran $\geq 0,70$ dapat dipandang menunjukkan kualitas pemeringkatan yang baik dalam konteks eksperimen ini, sedangkan peningkatan terhadap baseline menjadi tolok ukur utama keberhasilan.

Meskipun demikian, penelitian ini memiliki keterbatasan karena evaluasi bersifat *offline* dan bergantung pada skenario data simulasi/terkontrol. Penelitian lanjutan disarankan untuk memperluas periode evaluasi, menguji beberapa konfigurasi *exposure* dan tingkat *sparsity*, menerapkan prior berbasis *time-decay*, serta mengintegrasikan fitur kontekstual (misalnya harga dan kategori) untuk meningkatkan adaptivitas pemeringkatan terhadap perubahan tren.

Daftar Rujukan

[1] D. Roy dan M. Dutta, “A systematic review and research perspective on recommender systems,” *J Big Data*, vol. 9, no. 1, Des 2022, doi: 10.1186/s40537-022-00592-5.

[2] B. Divya, M. Amritha, N. Nikitha, G. Jahnavi, dan C. Akshitha, “E – COMMERCE PRODUCT RECOMMENDATION SYSTEM USING COLLABORATIVE FILTERING,” *International Journal of Engineering Technology Research & Management*, vol. 9, no. 06, Jun 2025, [Daring]. Tersedia pada: <https://www.ijetrm.com/IJETRM>

[3] P. M. Alamdari, N. J. Navimipour, M. Hosseinzadeh, A. A. Safaei, dan A. Darwesh, “A Systematic Study on the Recommender Systems in the E-Commerce,” *IEEE Access*, vol. 8, hlm. 115694–115716, 2020, doi: 10.1109/ACCESS.2020.3002803.

[4] S. Sameena, G. Javali, N. Srilakshmi, M. Jhansi, dan S. S. Sk, “Personalized product recommendation system for e-commerce platforms,” *ITM Web of Conferences*, vol. 74, hlm. 03012, 2025, doi: 10.1051/itmconf/20257403012.

[5] E. Mena-Maldonado, Y. Ren, dan M. Sanderson, “Popularity Bias in False-Positive Metrics for Recommender Systems Evaluation,” 2021. [Daring]. Tersedia pada: <https://github.com/elikary/tois2021>.

[6] B. Efron, “Empirical Bayes: Concepts and Methods,” 2021.

[7] R. Koenker dan J. Gu, “Empirical Bayes for the Reluctant Frequentist,” Apr 2024, [Daring]. Tersedia pada: <http://arxiv.org/abs/2404.03422>

[8] N. Ignatiadis, S. Saha, D. L. Sun, dan O. Muralidharan, “Empirical Bayes Mean Estimation With Nonparametric Errors Via Order Statistic Regression on Replicated Data,” *J Am Stat Assoc*, vol. 118, no. 542, hlm. 987–999, 2023, doi: 10.1080/01621459.2021.1967164.

[9] G. L. Britten *dkk.*, “Evaluating the Benefits of Bayesian Hierarchical Methods for Analyzing Heterogeneous Environmental Datasets: A Case Study of Marine Organic Carbon Fluxes,” *Front Environ Sci*, vol. 9, Mar 2021, doi: 10.3389/fenvs.2021.491636.

[10] D. G. Rasines dan G. A. Young, “Empirical Bayes and Selective Inference,” 1 Oktober 2022, *Springer*. doi: 10.1007/s41745-022-00286-0.

[11] Y. Hamura, K. Irie, dan S. Sugasawa, “On Global-Local Shrinkage Priors for Count Data*,” *Bayesian Anal*, vol. 17, no. 2, hlm. 545–564, Jun 2022, doi: 10.1214/21-BA1263.

[12] D. Gerard dan M. Stephens, “Empirical Bayes shrinkage and false discovery rate estimation, allowing for unwanted variation,” *Biostatistics*, vol. 21, no. 1, hlm. 15–32, Jan 2020, doi: 10.1093/biostatistics/kxy029.

[13] C.-H. Yang, H. Doss, dan B. C. Vemuri, “An Empirical Bayes Approach to Shrinkage Estimation on the Manifold of Symmetric Positive-Definite Matrices,” Okt 2020, [Daring]. Tersedia pada: <http://arxiv.org/abs/2007.02153>

[14] P. Septianingsih dan I. Y. Wulansari, “Small Area Estimation Using Empirical Bayes Poisson Gamma on Adolescent Fertility Rate in Indonesia,” *Indonesian Journal of Statistics and Its Applications*, vol. 7, no. 2, hlm. 114–129, Des 2023, doi: 10.29244/ijsa.v7i2p114-129.

[15] A. Iqbal, L. A. Al-Essa, M. Y. Shad, F. S. Alduais, M. F. Yassen, dan M. A. Raza, “E-Bayesian Estimation of Hierarchical Poisson-Gamma Model on the Basis of Restricted and Unrestricted Parameter Spaces,” *Complexity*, vol. 2023, 2023, doi: 10.1155/2023/8767200.

[16] T. S. De, P. Singh, dan A. Patel, “A Machine learning and Empirical Bayesian Approach for Predictive Buying in B2B E-commerce,” dalam *ACM International Conference Proceeding Series*, Association for Computing Machinery, Jan 2024, hlm. 17–24. doi: 10.1145/3647750.3647754.

[17] N. Padhy, S. Suman, T. S. Priyadarshini, dan S. Mallick, “A Recommendation System for E-Commerce Products Using Collaborative Filtering Approaches †,” *Engineering Proceedings*, vol. 67, no. 1, 2024, doi: 10.3390/engproc2024067050.

[18] M. Chen, S. Nabi, dan M. Siniscalchi, “Advancing Ad Auction Realism: Practical Insights & Modeling Implications,” Apr 2024, [Daring]. Tersedia pada: <http://arxiv.org/abs/2307.11732>

[19] P. Chinnasamy, “Transforming E-Commerce with Intelligent Recommendation Systems: A Review of Current Trends in Machine Learning and Deep Learning,” 17 Maret 2025, *Prof.Dr. Iskender AKKURT*. doi: 10.22399/ijcesen.1183.

[20] M. Park dan J. Oh, “Enhancing E-Commerce Recommendation Systems with Multiple Item Purchase Data: A Bidirectional Encoder Representations from Transformers-Based Approach,” *Applied Sciences (Switzerland)*, vol. 14, no. 16, Agu 2024, doi: 10.3390/app14167255.

- [21] Abhishekawasthi151, "Mobile Sales Data," Kaggle. [Daring]. Tersedia pada: <https://www.kaggle.com/datasets/abhishekawasthi151/mobile-sales-data>