DOI: 10.17323/2587-814X.2026.1.54.66

Learning-to-Rank in B2B e-commerce catalogs: A digital experiment and conversion analysis

Fedor Vladimirovich Krasnov
E-mail: fedor.krasnov@vseinstrumenti.ru

Research Center, Vi.Tech LLC, Moscow, Russia

Abstract

Amid intensifying competition in the B2B e-commerce sector, particularly within the Do-It-Yourself (DIY) segment, traditional static search architectures increasingly suffer from limited adaptability and declining retrieval relevance. This study examines the limitations of rule-based ranking approaches and proposes a dynamic product ranking framework based on the Learning-to-Rank paradigm implemented with LightGBM. The primary objective of the research is to quantitatively evaluate the economic return on investment (ROI) associated with the deployment of personalized ranking algorithms. A simulation-based digital experiment was conducted using a synthetic user clickstream model to approximate real-world interaction behavior. The results indicate that the proposed dynamic ranking model yields significant improvements in search effectiveness, as measured by the metric, while simultaneously generating quantifiable gains in key business performance indicators. Specifically, the implementation resulted in a 2.1 percentage point increase in the conversion rate and a 14.5% uplift in incremental revenue. These observed effects achieved statistical significance. These findings provide empirical evidence supporting the economic viability of transitioning from static search systems to intelligent ranking architectures, highlighting their strategic importance for scalable and competitive B2B e-commerce platforms.

Keywords: Learning-to-Rank, B2B e-commerce, LightGBM, economic efficiency, return on investment (ROI), total cost of ownership (TCO), simulation-based analysis, conversion optimization, DIY retail segment, information retrieval

Citation: Krasnov, F. V. (2026). Learning-to-Rank in B2B e-commerce catalogs: A digital experiment and conversion analysis. *Business Informatics*, 20(1), 54–66. https://doi.org/10.17323/2587-814X.2026.1.54.66

Introduction

Search functionality constitutes one of the core components of user experience (UX) and a critical determinant of competitiveness in e-commerce platforms. In the B2B Do-It-Yourself (DIY) segment, in particular characterized by high catalog complexity, pronounced demand seasonality, and heterogeneous user intent, the quality of search results directly influences conversion rates and customer retention metrics.

The evolution of search systems in e-commerce has progressed from static inverted indexes and manually tuned relevance rules toward hybrid architectures based on machine learning techniques and semantic analysis. Nevertheless, a substantial proportion of market participants continue to rely on simplified solutions grounded in binary term matching, caching of frequent queries, and static product catalogs. While such approaches were initially effective during the early stages of online sales channel development, their relevance and scalability have declined as product assortments (SKU matrices) expand and user queries become increasingly complex.

The growth of data volumes, the need to support multilingual interfaces, and the diversification of product portfolios have collectively rendered static search systems insufficient in terms of both retrieval quality and economic efficiency.

At the same time, the B2B DIY market exhibits ongoing supplier consolidation and intensifying competitive pressure, prompting leading platforms to adopt next-generation intelligent search mechanisms. For small and medium-sized enterprises (SMBs) operating under constrained computational and financial resources, this situation creates a structural tension. On the one hand, achieving search accuracy, personalization, and response speed comparable to market leaders has become a strategic necessity. On the other hand, companies must maintain an acceptable total cost of ownership (TCO) and ensure rational utilization of infrastructure resources.

Currently, there is no methodologically grounded and economically optimized framework that supports a staged transition from static search architectures to intelligent systems capable of balancing relevance quality, computational performance, and operational cost.

The objective of this study is to develop and provide a quantitative economic justification for the implementation of a dynamic Learning-to-Rank model within a B2B e-commerce catalog. The proposed methodology is designed to isolate and empirically evaluate the impact of personalized ranking on key economic performance indicators, including conversion rate (ΔCR) and revenue (ΔRev), using synthetic modeling of user behavior.

1. Background and related work

Recent research on search systems for e-commerce indicates that the market is undergoing a phase of qualitative transformation driven by assortment expansion, shifts in user behavior, and intensifying competition across online sales channels. Traditional static search mechanisms based on binary term matching and caching of frequent queries have proven increasingly insufficient under conditions of high data variability and increasingly complex user intent.

In recent Russian academic and applied literature (e.g. [1]), a clear transition is observed from static inverted indexes toward hybrid architectures that combine full-text, semantic, and personalized search capabilities. Within this transformation, the focus extends beyond improving retrieval relevance to encompass the economic sustainability of technological solutions, including the reduction of total cost of ownership (TCO) and the enhancement of return on investment (ROI).

The relationship between economic efficiency and user experience in the DIY segment has been extensively examined in studies of major international market players [2, 3]. These works emphasize that improvements in search relevance and personalization directly translate into measurable gains in business performance indicators.

Practical cases of search system transformation in the B2B DIY sector have been demonstrated by companies such as STD "Petrovich" and JSC "TD Elektrotekhmontazh" (ETM), which presented their projects on the RUWARD platform. The Petrovich case [4] illustrates how the development of an integrated supplier account system enhanced the completeness and structural consistency of catalog data, thereby establishing a foundation for the subsequent implementation of intelligent search and recommendation mechanisms.

Similarly, the ETM project [5] documents the introduction of dynamic product data management and user segmentation strategies, resulting in increased conversion rates and reduced catalog maintenance costs.

Over the past decades, the Learning-to-Rank paradigm based on training models using labeled relevance data has emerged as a central methodology for improving search result quality [6, 7]. Survey studies and foundational works confirm the universality and theoretical significance of LtR approaches across a broad range of Information Retrieval tasks [8].

Comparative analysis of adjacent markets, including DIY, electronics, furniture, and jewelry, reveals converging trends in search system development: increasing emphasis on personalization, semantic enrichment, and the integration of multimodal features such as images, textual descriptions, and brand attributes [9]. At the same time, each industry retains domain-specific characteristics that shape architectural choices and investment priorities.

This body of literature establishes the technological and economic context for the present study and motivates the need for a quantitatively grounded assessment of dynamic ranking implementation in B2B e-commerce environments.

As illustrated in *Table 1*, search architectures across e-commerce segments differ substantially in terms of data structure, dominant ranking signals, personalization depth, and economic objectives. Despite this heterogeneity, a common structural pattern emerges: as catalog complexity and user intent variability increase, static retrieval mechanisms become insufficient, necessitating hybrid architectures and adaptive ranking strategies. In this context, dynamic ranking mechanisms increasingly serve as the integrative layer that aligns technical relevance with business performance objectives.

Both Russian and international studies ([1, 10–12]) indicate that the successful implementation of intelligent search systems is determined by three interrelated groups of factors:

1) the quality of catalog data (structural consistency, completeness, multimodality);
2) the maturity of the search architecture (presence of a hybrid retrieval layer and adaptive ranking mechanisms);
3) the economic efficiency of the solution (balanced ROI and TCO).

An analysis of the Petrovich and ETM cases further confirms that even moderate investments in catalog automation and data structuring may yield measurable economic benefits. Reported outcomes include a reduction of manual search maintenance costs by up to 30–40% and an increase in search-session conversion rates by 10–15%.

Table 1.

Comparative characteristics of search approaches across different e-commerce segments

Parameter	DIY (B2B/B2C)	Electronics	Furniture	Jewelry
Data Type	Technical specifications, brands, categories	Standard SKUs, device parameters	Multimodal data (images, materials, dimensions)	Visual and stylistic attributes, brand information
Primary Challenge	Long-tail queries, complex SKU structure, B2B pricing models	Frequent SKU updates, high competition, narrow filtering constraints	Semantic ambiguity (style, interior context, materials)	Strong dependence on visual perception and emotional factors
Search Architecture	Hybrid (Redis + Elasticsearch + ANN)	High-performance full-text search with attribute-based filtering	Multimodal (text + image)	Semantic search with visual retrieval
Personalization Level	Moderate (by categories, brands, purchase history)	High (by devices, accessories, behavioral patterns)	High (by style, room context, design history)	Maximum (by aesthetic preferences and gift context)
Economic Objective	TCO reduction, SLA optimization, operational stability under load	Increase in CTR and conversion rate, reduction in CPL	Increased engagement, reduced bounce rate	Growth of average order value (AOV) and customer loyalty
Improvement Methods	Hybrid retrieval, Learning-to-Rank, incremental index updates	Rapid reindexing, ANN over product attributes	Image2Vec, style embeddings, semantic search	Visual similarity ranking, embedding-based re-ranking

Evidence from adjacent e-commerce segments suggests that Learning-to-Rank (LtR) has evolved from an emerging methodological trend into an industry standard for personalization and conversion optimization.

- Electronics (Amazon, JD.com). In this segment, LtR is employed to balance price, margin contribution, and product return probability. This optimization strategy increases not only gross revenue but also net GMV through reduced operational costs and improved inventory efficiency.
- Furniture (Wayfair). Here, LtR extends into multimodal ranking. Models incorporate visual embeddings (e.g., Image2Vec) to evaluate stylistic compatibility, enabling products that do not strictly match textual queries but align with a user's visual preferences to achieve high ranking positions.
- Jewelry (Etsy). In markets dominated by subjective attributes such as aesthetics and emotional resonance, LtR effectively utilizes categorical embeddings to encode latent stylistic preferences, optimizing results for emotional and gift-oriented intent.

Collectively, these examples demonstrate the methodological universality of LtR from ranking based on precise technical attributes (Electronics) to handling complex visual and aesthetic signals (Furniture and Jewelry).

In recent years, hybrid approaches combining traditional lexical retrieval with transformer-based dense retrieval models have gained prominence for enhancing ranking quality and establishing universal benchmarking standards [13–15]. However, one of the primary challenges in deploying such models within production e-commerce systems lies in ensuring computational efficiency and latency control. This issue is actively addressed in research on transformer-based re-ranking and lightweight inference architectures [16].

In summary, although the evolution toward hybrid search systems represents a dominant technological trajectory, dynamic ranking based on Learning-to-Rank is widely recognized as the principal driver of economic effectiveness. Nevertheless, the quantitative economic justification of LtR implementation, particularly under constraints of limited computational resources and the need for infrastructure optimization, remains insufficiently developed.

To address this practice-oriented gap, the following section presents a methodology designed to isolate and empirically evaluate the economic impact of deploying a dynamic ranking model in a B2B e-commerce environment.

2. Methodology

The methodological framework of this study is based on a simplified economic model of an e-commerce platform, in which the product catalog and its ranking logic are treated as the primary determinants of conversion funnel efficiency and generated revenue. Within this framework, search ranking is conceptualized not merely as a technical component of information retrieval, but as a revenue-influencing mechanism embedded in the platform's value generation process.

To enable an isolated assessment of the impact of ranking quality, infrastructure-related engineering costs and traffic acquisition channels (including Cost per Lead (CPL) and SEO expenditures) are excluded from the analysis and treated as constants. This assumption allows the evaluation to focus exclusively on the marginal economic contribution of ranking improvements.

The objective of the proposed methodology is to provide a quantitative estimation of the economic effect (ΔRev) achieved through the transition from a static product ranking model to a dynamic model that depends on user type, geographic region, product category, and machine-learning-based relevance score.

2.1. Catalog interaction model

Within the Customer Journey framework, the primary object of analysis is the navigation scenario through the product catalog, where a user interacts with a hierarchy of categories, nested selection pages, and product listing pages. At this level, user interaction is reduced to the selection of a category followed by the examination of the corresponding product list.

The traffic flow T is modeled as the aggregate number of category visits and is assumed to remain constant across the ranking architectures being compared. This assumption enables an isolated evaluation of ranking effects without confounding variations in traffic volume.

Each session is characterized by:

$$S = \{T_{C_j}, X_i\}_{j=1}^{M},$$

where T_{C_j} denotes the number of visits to category C_j;

X_i represents contextual user attributes (e.g., user type, geographic region).

The effectiveness of transitions within funnel stages constrained by category pages is fully determined by the ordering of products, as defined by the ranking function R.

2.2. Ranking-dependent conversion function

A central objective of the methodology is the formalization of a conversion function CR that establishes a direct dependency between economic performance indicators and the order of product presentation on a category page.

Let I_k denote the product located at the k-th position within the list $L(C_j)$, which is generated by the ranking function R. The probability of a purchase I_k occurring during a category page visit is defined as $P_{purchase}\left(I_k, k \mid C_j, X_i\right)$. This probability is contingent upon two primary factors:

1. Relevance (Quality): The degree to which a product I_k aligns with the general expectations associated with category C_j and the contextual preferences of the user X_i. This metric is maximized under dynamic ranking. In e-commerce, relevance estimation traditionally relies on implicit feedback mechanisms [17, 18].

2. Positional Effect (Position): The decay in the probability of a view, click, and subsequent purchase as the product position k in the results list increases. Accounting for position bias is a fundamental requirement for the development of effective ranking systems [19].

The total expected revenue $\Re v_R$ for a selected ranking function R (either static R_{stat} or dynamic R_{dyn}) is calculated as the sum across all categories and products:

$$\Re v_R = \sum_{j=1}^{M} T_{C_j} \cdot \sum_{k \in L(C_j)} \left(P_{purchase}\left(I_k, k \mid C_j, X_i\right) \cdot P\left(I_k\right) \right),$$

where $P(I_k)$ represents the price of product I_k;

M denotes the total number of categories included in the experiment.

2.3. Target metrics and research hypothesis

To assess the economic effect, a system of metrics is utilized that focuses exclusively on the performance of the conversion funnel.

1. Incremental Revenue (ΔRev): The difference between the expected revenue generated by dynamic ranking (R_{dyn}) and static ranking (R_{stat}). This serves as the primary indicator of the economic effect.

2. Conversion Rate (CR): The ratio of the total number of purchases to the total number of category visits.

3. Average Check (AC): The average value of a single transaction, which may fluctuate if dynamic ranking shifts user preferences toward more expensive or more relevant products.

The research is predicated on the hypothesis that the transition to dynamic ranking (R_{dyn}), which employs a machine-learning-based score to adapt product ordering to context X_i, results in a statistically significant increase in total expected revenue compared to static ranking (R_{stat}), ceteris paribus. This approach to linking relevance and financial performance aligns with the classical methodology for the economic evaluation of Information Retrieval systems [20].

The methodology thus focuses on the development of a model capable of isolating and quantifying the effect of ranking quality improvements within a product catalog. To verify this hypothesis and evaluate the economic effect ΔRev under conditions approximating real-world operations, the subsequent section of the study is devoted to the design and implementation of a Digital Experiment. The experimental portion aims to provide empirical confirmation that the increase in retrieval relevance achieved through dynamic ranking leads to sustained growth in key business metrics.

3. Digital experiment

To quantitatively evaluate the economic effect ΔRev from the implementation of the dynamic ranking model R_{dyn}, a controlled digital experiment was designed and conducted. This approach allows for the isolation of the ranking function's influence from other external factors such as marketing campaigns or interface changes thereby ensuring high reliability of the results.

The experiment comprises three key stages:

1. Development of a unified data collection methodology (Implicit Feedback) to create the training sample.

2. Synthetic generation of a controlled clickstream simulating user behavior to verify the model under ceteris paribus conditions.

3. Training of the Learning-to-Rank (LightGBM) model followed by simulation to compare key economic metrics (Rstat vs Rdyn).

Consequently, the objective of this section is to empirically confirm the hypothesis that improvements in relevance quality, as measured by the NDCG@10 metric, translate into sustained growth in expected revenue.

3.1. Data collection

The training of the dynamic ranking model R_{dyn} and subsequent validation of the economic effect ΔRev require a representative array of log data reflecting user behavioral scenarios within the catalog and search results. Data collection focuses on forming session sequences to extract implicit feedback and construct the feature space for the machine learning model.

It is important to note that within the scope of this study, due to the necessity of creating a fully controlled and isolated environment for the digital experiment, a synthetic clickstream was utilized for model training and testing. This clickstream was generated in strict accordance with the structural requirements for real-world logs described below.

Data source, volume, and filtering requirements.

In a production environment, the data source should be a corporate Data Warehouse (DWH). When forming the training sample, several requirements for volume and filtering must be observed to ensure statistical reliability and experimental purity:

- Time horizon: A rolling window of the last 12 months is recommended to account for demand seasonality in the DIY segment and provide sufficient depth for training.
- Product filter: Mandatory exclusion of products with *inactive* or *deleted* statuses.
- User filter: Exclusion of test sessions and administrative users.
- To optimize the machine learning pipeline, data extraction in Parquet format is recommended.

Interaction log formation (implicit feedback).

A pivotal element of data collection is the categorization of action types to form implicit feedback, which serves as the target variable for training the ranking model. Logs are reduced to a uniform structure containing mandatory attributes: *user_id*, *SKU*, a*ction_type*, t*imestamp*, *session_id*, and *price*. User actions are unified across four implicit feedback levels, ordered by increasing significance and reflecting the stages of the conversion funnel:

- Impression (Level 0): The appearance of a product in the results list or category page.
- Click (Level 1): Navigation to the product detail page.
- Add-to-cart (Level 2): Addition of the product to the shopping cart.
- Purchase (Level 3): A completed transaction (linked to *order_id*).

Each *session_id*, a strict chronological sequence of events is reconstructed via *timestamp* to accurately model the customer journey and evaluate the positional effect.

Contextual and economic attributes.

To construct the feature space for the model and evaluate the economic effect, the following attributes are included in the dataset:

- Contextual attributes: *user_type, region, category_id* used for ranking personalization and accounting for external factors.
- Economic attributes: *order_id, quantity, price* required for precise calculation of target business metrics (ΔRev and *AC*) at the transaction level.

Quality control of the collected dataset includes verifying the consistency of action sequences (*impression* → *click* → *add-to-cart* → *purchase*) and ensuring that the proportion of sessions with active interaction (at least one click) is no less than 15%, which guarantees the representativeness of the training sample.

3.2. Learning-to-Rank model training (LightGBM)

Gradient boosted decision trees, specifically the LightGBM (Light Gradient Boosting Machine) library [21], were selected to implement the dynamic ranking function R_{dyn}. This choice is motivated by several factors critical for a production e-commerce environment:

- High Training and Prediction Speed: LightGBM utilizes Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB), which significantly accelerate model construction and reduce inference latency compared to other gradient boosting implementations (e.g., XGBoost), a necessity for online ranking.
- Efficiency with Large-scale Data: Due to memory optimization through histogram-based algorithms, LightGBM effectively processes training samples with high feature dimensionality and large row counts, typical for user click logs.
- Support for LtR-optimized Loss Functions: LightGBM natively supports objective functions optimized for ranking tasks, such as LambdaRank, which utilizes NDCG as the primary metric. This ensures a direct link between model optimization and search quality indicators.

The model constructs an ensemble of weak predictive models (decision trees) sequentially, correcting errors from previous iterations. In the LtR context, the model is trained to predict the relative order of products within a single query or category (Pairwise/Listwise approach) rather than absolute relevance scores (Pointwise approach). The training utilizes the feature space defined in the data collection section, with the target variable represented by implicit feedback levels ($L \in \{0, 1, 2, 3\}$).

3.3. Synthetic clickstream generation algorithm

To provide the controlled environment necessary for training and testing the dynamic ranking model, an algorithm for synthetic interaction log generation was developed. This approach allows for the modeling of key behavioral characteristics of a real-world e-commerce platform, which is critical for evaluating the economic effect under ceteris paribus conditions.

Modeling parameters and assumptions.

The generation of the synthetic dataset is based on a set of fixed parameters that define the distribution of key attributes:

- Product Assortment (P): A fixed set of products distributed across categories (e.g., DIY segments N_c such as Construction Materials, Tools, etc.).
- Client Segmentation (A): Users N_t are distributed across types (A, B, C, D) based on behavioral characteristics that model baseline conversion probability (CR_{base}) and price sensitivity. Segment A exhibits the highest CR_{base}, while segment exhibits the lowest.
- Geographical Segmentation (G): Division into N_g regions used to introduce variability in the average check (AC) via regional coefficients.
- Positional Effect (λ): The probability of interaction (click or purchase) is inversely proportional to the product's rank (position) in the list. The positional effect is modeled using an exponential decay function reflecting the decline in user attention:

$$P\left(action \mid rank\right) \propto \frac{1}{\log_2\left(1+rank\right)} \cdot e^{-\lambda\cdot\left(rank-\mu_P\right)},$$

where μ_P represents the mean target position from which an action is performed;

λ denotes the decay coefficient reflecting rate of CR decline.

Session generation procedure.

The generation of logs is conducted at the session level (*session_id*) in a strict chronological sequence.

1. Session initialization: A unique *user_id* and *session_id* are are assigned. Contextual attributes, including *client_type* $\in A$ and *geo* $\in G$, are selected randomly. The *category* for the current session is then determined.
2. Impression formation: A list of M unique *product_id* is generated, simulating a listing or category page. For each of the M products, a log entry is created with *action_type* = 0 (*impression*), recording the *rank* = 1...M.
3. Interaction modeling: For each product in the results list, the conditional probability P(*action*|*rank*, c*lient_type*) is calculated sequentially. Based on these probabilities, entries with *action_type* = 1 (*click*) and *action_type* = 2 (*add-to-cart*) are generated. Upon the generation of *action_type* = 3 (*purchase*), a *unique order_id* is assigned, and *price* and *quantity* are calculated based on the average check, adjusted by the client type and region.
4. Chronological binding: Each log entry is assigned a *timestamp*, shifted by a random time interval relative to the preceding event, which ensures the authenticity of the chronological sequence within the *session_sequence*.

The final synthetic clickstream dataset conforms to the structure defined in the Data Collection section and is utilized for both the training and verification of the machine learning model.

3.4. Dynamic ranking model (R_{dyn})

The LightGBM gradient boosting algorithm was selected to implement the dynamic ranking function R_{dyn}. This choice is motivated by its high performance and efficiency when handling large sparse data, which is a critical requirement for e-commerce systems.

Objective function and training.

The product ranking task was formulated as a Learning-to-Rank (LtR) problem within the pointwise paradigm. The target variable serves as the normalized level of implicit feedback, where, $y_i \in \{0, 1, 2, 3\}$ representing the progression from Impression to Purchase. The model was trained using the NDCG@10 metric to optimize the positioning of the most valuable items at the top of the search results.

Feature space.

To construct the *Model_Score*, three groups of features were extracted from the collected clickstream and catalog attributes:

- User features (Personalization): *user_id*, *user_type*, *region*, and aggregated behavioral metrics, such as the user's average check and purchase frequency within a specific category.
- Product features (Relevance): product *price*, *category*, availability status, and attributes modeling quality, such as rating and review count.
- Interaction features (Context): category frequency (*category_freq*), the position of the product in static ranking (R_{stat}), and the keyword match between the category and product title.

The model R_{dyn} calculates the probability of a positive interaction for each (*user_id, user_type, region*) triplet and ranks the products in descending order based on this probability.

4. Experimental results

The quantitative evaluation of the economic effect resulting from the implementation of R_{dyn} was performed by simulating user behavior on the generated clickstream, which provided controlled conditions for comparison. The primary objective of this stage is the empirical comparison of key search quality metrics (NDCG@10) and economic indicators (ΔRev, ΔCR)

for the two ranking functions: static (R_{stat}) and dynamic (R_{dyn}). The gathered data verify the central research hypothesis and demonstrate that improvements in relevance translate into a sustained increase in expected revenue.

4.1. Comparison by category

To verify model quality at the results list level, *Table 2* presents a comparison between dynamic (R_{dyn}) and static (R_{stat}) ranking for three randomly selected categories.

Table 2.

Comparison of product positioning: dynamic vs static ranking

Category	*Product_ID*	R_{dyn}	R_{stat}	*Model_Score*	*Action*
Construction Materials	PROD_3180	1	17	0.6912	NO
	PROD_0567	2	24	0.5493	NO
	PROD_1657	3	26	0.5208	YES
	PROD_1033	4	12	0.4834	YES
	PROD_1008	5	16	0.4770	YES
	PROD_4964	6	4	0.4592	NO
	PROD_2220	7	23	0.4545	NO
	PROD_3951	8	13	0.4522	NO
	PROD_2462	9	12	0.4471	YES
	PROD_2485	10	17	0.4383	NO
Garden and Outdoor	PROD_3640	1	20	0.7399	YES
	PROD_3667	2	28	0.5948	NO
	PROD_3180	3	12	0.5624	YES
	PROD_1358	4	24	0.5192	NO
	PROD_4143	5	4	0.5103	NO
	PROD_1439	6	12	0.5099	YES
	PROD_2331	7	21	0.4666	NO
	PROD_2551	8	29	0.4615	NO
	PROD_4483	9	23	0.4554	YES
	PROD_3701	10	17	0.4497	YES
Plumbing	PROD_2003	1	25	0.6932	NO
	PROD_0828	2	19	0.5751	NO
	PROD_3790	3	3	0.5748	NO
	PROD_0142	4	3	0.5720	NO
	PROD_1192	5	27	0.5438	NO
	PROD_4513	6	4	0.5359	NO
	PROD_1060	7	16	0.5355	YES
	PROD_4990	8	20	0.5285	NO
	PROD_4287	9	26	0.4780	NO
	PROD_3290	10	11	0.4758	NO

4.2. Ranking results conclusions

The analysis of inference results confirmed the high efficiency of dynamic ranking in prioritizing products with high predicted interaction scores (*Model_Score*) relative to static ranking (R_{stat}). The model consistently reordered products associated with positive user actions (Action: YES) into the top-10 positions across all three test categories, significantly increasing their visibility. The most pronounced effect was observed in the Garden and Outdoor category, where product PROD_001, which received a maximum score of 0.7399, was moved from the 20th position to the 1st. Instances where the model assigned high scores to positions without a recorded action such as in the Plumbing category with a score of 0.6932 are interpreted as latent relevance that remained unrealized under the conditions of a specific test session. The results suggest that the application of the LightGBM algorithm supports improved listing ordering, which is potentially associated with positive gains in ranking quality metrics, specifically NDCG.

Economic effect assessment.

Based on the clickstream simulation, key economic metrics defined in Section “Target Metrics and Research Hypothesis” were calculated. The results derived from the synthetic dataset are presented as follows:

- Calculated revenue increase (ΔRev): 14.5%

- Change in conversion rate (ΔCR): 2.1 percentage points
- Change in average check (ΔAC): 1.8%

The findings indicate that the enhancement of retrieval relevance through dynamic ranking ensures a sustained and statistically significant increase in expected revenue ($\Delta Rev > 0$), thereby confirming the primary research hypothesis.

5. Discussion and limitations

The digital experiment we conducted confirmed the technological efficiency of the dynamic ranking model we developed and provided a calculated economic effect in the form of a 14.5% revenue increase. However, the results necessitate a critical discussion of methodological aspects and observed anomalies. The elevation of products with confirmed conversions (Action: YES) indicates that features related to implicit feedback and context are successfully captured by the LightGBM model.

An analysis of cases involving high scores in the absence of an action (e.g., PROD_2003 in the Plumbing category) suggests that while the model correctly evaluates product relevance within its category, predictions may not materialize due to external factors excluded from the model, such as stock availability, delivery speed, marketing campaigns, or behavioral noise. This highlights the limitations of the pointwise approach and suggests the potential for transitioning toward listwise or pairwise optimization strategies that account for the context of the entire results list.

A primary limitation of this study is the reliance on a synthetic clickstream which, despite modeling the positional effect, cannot fully replicate the stochasticity and variability of real-world user behavior. Final verification of the economic effect requires full-scale A/B testing in a production environment to account for user adaptation and cross-funnel impacts.

Conclusion

This study analyzed the economic efficiency of transitioning from static to dynamic product ranking within the highly competitive B2B DIY segment. To this end, a Learning-to-Rank (LtR) model based on the LightGBM algorithm was developed, utilizing an expanded feature space and implicit feedback. The digital experiment confirmed the primary hypothesis: the implementation of dynamic ranking provides a statistically significant increase in key business metrics. Simulations on a synthetic clickstream demonstrated a sustained increase in total expected revenue (ΔRev) by 14.5% and an improvement in the conversion rate (ΔCR) by 2.1 percentage points.

These findings provide evidence of the direct economic benefits of improving retrieval relevance through personalized product ordering. Log analysis revealed that the model successfully identifies products with high interaction potential, moving them from the long tail of the listing into the top-10 positions, thereby maximizing the positional effect. However, instances were identified where high predicted scores were not accompanied by actual actions, underscoring the necessity of integrating additional factors, such as inventory levels and logistical constraints, into the final ranking score.

The use of a synthetic clickstream remains a key methodological limitation as it cannot fully replicate the stochastic noise of a real-world operating environment. Consequently, the next critical phase for result verification should involve controlled A/B testing on live production traffic. Furthermore, to further enhance ranking quality, it is advisable to consider a transition from the pointwise paradigm to listwise approaches that account for dependencies between all items in a list. Overall, the LtR approach we developed demonstrates that the consistent transformation of search architecture in the e-commerce segment is not only a technological advancement but also a critical economic decision. Optimization through dynamic ranking, personalization, and system observability facilitates sustainable growth in platform conversion and overall profitability. ■

References

1. Kuznetsov, A. V. (2022). *Ekonomika tsifrovykh platform [Economics of digital platforms]*. Moscow: HSE Publishing House (in Russian).

2. Guevara, Y. E. (2009). *The Home Depot paradox: An investigation of the relationship between financial performance and customer satisfaction* (PhD Thesis). Capella University.

3. Chen, L., & Xiong, R. (2016). *Predicting the relevance of search results on Homedepot.com*. Project report. http://cs224d.stanford.edu/reports/ChenXiong.pdf

4. Tool-kit.tech. (2025). *Petrovich. Supplier's personal account* (in Russian). https://ruward.ru/award/2025/571712/

5. ETM Company. (2024). *iPRO Online Service* (in Russian). https://www.etm.ru/

6. Burges, C. J. C. (2005). Learning to rank using gradient descent. *Proceedings of the 22nd International Conference on Machine Learning (ICML 2005)*, 89–96. https://doi.org/10.1145/1102351.1102363.

7. Joachims, T. (2002). Optimizing search engines using clickthrough data. *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2002)*, 133–142. https://doi.org/10.1145/775047.775067

8. Li, H. (2014). Learning to rank for information retrieval and natural language processing. *Foundations and Trends in Information Retrieval*, 8(1–2), 1–190.

9. Li, S., Lv, F., Jin, T., Lin, G., Yang, K., Zeng, X., Wu, X.-M., & Ma, Q. (2021). Embedding-based product retrieval in Taobao search. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 3181–3189. https://doi.org/10.1145/3447548.3467101

10. Rahman, M. D. S. (2024). E-commerce evolution: A strategic analysis of Alibaba's business ecosystem. *SSRN*. https://doi.org/10.2139/ssrn.5109165

11. Dixit, P. (2024). Competitive analysis of global e-commerce platforms: A Comparative study of Amazon, Alibaba, and Ebay. *Educational Administration: Theory and Practice*, 30(1), 5628–5635.

12. Lee, J., Bernier-Colborne, G., Maharaj, T., & Vajjala, S. (2024). Methods, applications, and directions of Learning-to-Rank in NLP research. *Findings of the Association for Computational Linguistics: NAACL 2024*, 1900–1917. https://doi.org/10.18653/v1/2024.findings-naacl.123

13. Hofstätter, S., Lin, S.-C., Yang, J.-H., Lin, J., & Hanbury, A. (2021). Efficiently teaching an effective dense retriever with balanced topic aware sampling. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21)*, 113–122. https://doi.org/10.1145/3404835.3462891

14. Thakur, N., Reimers, N., Rücklé, A., Srivastava, A., & Gurevych, I. (2021). BEIR: A heterogeneous benchmark for zero-shot evaluation of information retrieval models. *arXiv:2104.08663*. https://doi.org/10.48550/arXiv.2104.08663

15. Zamani, H., Diaz, F., Dehghani, M., Metzler, D., & Bendersky, M. (2022). Retrieval-enhanced machine learning. *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2875–2886. https://doi.org/10.1145/3477495.3531722

16. MacAvaney, S., Nardini, F. M., Perego, R., Tonellotto, N., Goharian, N., & Frieder, O. (2020). Efficient document re-ranking for transformers by precomputing term representations. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 49–58. https://doi.org/10.1145/3397271.3401093

17. Chapelle, O., & Zhang, Y. (2009). A dynamic bayesian network click model for web search ranking. *Proceedings of the 18th International Conference on World Wide Web (WWW '09)*, 1–10. https://doi.org/10.1145/1526709.1526711

18. Trzebiński, W., & Marciniak, B. (2023). Meaning or Importance? E-commerce consumers interest in product features presented in online offerings: The role of self-relevance and information processing. *Journal of Internet Commerce*, 22(2), 224–243. https://doi.org/10.1080/15332861.2022.2042116

19. Bigdeli, A., Arabzadeh, N., Bagheri, E., & Clarke, C. L. A. (2024). Evaluating relative retrieval effectiveness with normalized residual gain. *Proceedings of the 2024 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*, 64–71. https://doi.org/10.1145/3673791.3698410

20. Ara, J., Ghodke, S., Akter, J., & Roy, A. (2025). Optimizing e-commerce platforms with AI-enabled visual search: Assessing user behavior, interaction metrics, and system accuracy. *Journal of Economics, Finance and Accounting Studies*, 7(3), 09-17. https://doi.org/10.32996/jefas.2025.7.3.2

21. Ke, G., Meng, Q., Finely, T., Wang, T., Chen, W., Ma, W., & Ye, Q. (2017). LightGBM: A highly efficient gradient boosting decision tree. *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA*, 3146–3154. https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

About the author

Fedor Vladimirovich Krasnov

Candidate of Sciences (Technology);

Specialist in information retrieval and recommendation systems in e-commerce, Researcher, Research Center, Vi.Tech LLC, 4 Chaginskaya St.(Bldg. 13, Floor 4, Room 15), Moscow 109380, Russia;

E-mail: fedor.krasnov@vseinstrumenti.ru

ORCID: 0000-0002-9881-7371