



ConBRepro

XIII CONGRESSO BRASILEIRO DE ENGENHARIA DE PRODUÇÃO



IA nas Engenharias

29 nov. a 01 de dezembro 2023

Aprendizagem de ordenação no comércio eletrônico de uma indústria de bens para o lar

Caio Tomy Ono Kaminari

PPGOLD - UFPR

Cassius Tadeu Scarpin

Grupo de Tecnologia Aplicada à Otimização (GTAO) - PPGOLD/PPGMNE - UFPR

Resumo: Esta pesquisa apresenta técnicas de aprendizado de máquina (ML) para a ordenação de resultados de busca utilizadas em comércio eletrônico. O intuito é melhorar a jornada do usuário após uso do sistema de busca da plataforma quando melhorada pelos algoritmos de ML propostos. Os algoritmos apresentados estão enquadrados na literatura na área de Recuperação de Informação e Aprendizagem de Ordenação. Para estimativa de relevância, utilizou-se as alternativas de taxa de clique, taxa de adição no carrinho e taxa de conversão a partir de dados reais de usuários da plataforma. O resultado obtido mostra que um modelo com maior robustez forneceu resultados melhores que o método atual do comércio eletrônico testado quando usado a métrica de avaliação Ganho Cumulativo Descontado Normalizado (NDCG).

Palavras-chave: Comércio eletrônico; busca; aprendizagem de ordenação.

Learning to Rank in e-commerce in a household goods industry

Abstract: This research presents machine learning (ML) techniques for sorting search results used in e-commerce. The objective is to improve the user journey after using the platform's search system when improved by the proposed ML algorithms. The algorithms presented are framed in the literature in the area of Information Retrieval and Learning to Rank. To estimate relevance, we use click-through rate, add to cart rate and conversion rate as alternatives based on real data from users of the platform. The result obtained shows that a model with greater robustness presented better results than the current e-commerce method tested when using the Normalized Discounted Cumulative Gain (NDCG) evaluation metric.

Keywords: E-commerce; search; learning to rank.

1. Introdução

No ano de 2020 as vendas pela internet representaram 17,8% das vendas no setor de varejo em todo o mundo, e a participação do comércio eletrônico deve saltar para 24,5% até 2025, marcando um aumento de 37,6% em apenas cinco anos (RITA; RAMOS, 2022). Os varejistas online enfrentam muitos desafios para aumentar a receita e engajar os clientes na plataforma, pelo fato de o consumidor estar cada vez mais exigente ao ter uma

experiência de compras na internet. Por isso, um dos maiores desafios para as empresas é de como melhorar a ordenação de produtos após algum usuário interagir com o sistema de busca. Está consolidado na literatura acadêmica e é conhecido pelos varejistas que grande parte dos clientes visualizam produtos ordenados nas primeiras posições após acionado o sistema de busca em comparação com produtos ordenados na parte inferior da página (FERREIRA et al., 2022). Isto explica que empresas, como a Amazon, investem nesta área há anos, utilizando a inteligência artificial na ordenação de produtos do sistema de busca em sua plataforma (SOROKINA; CANTU-PAZ, 2016).

Durante a jornada de busca, o usuário recebe uma lista com os produtos mais relevantes de acordo com o termo buscado e a regra de ordenação. Neste momento, o consumidor analisa os produtos e pode selecionar àqueles com maior potencial para finalizar a compra. Nem sempre o que é mais relevante para a empresa é o que o usuário busca, ou seja, existem atributos de produto mais importantes para o usuário do que a regra imposta pela empresa, como por exemplo, preço, marca ou avaliações (ZHUANG et al., 2018). Por consequência, o usuário geralmente não navega por todos os produtos relevantes retornados pelo sistema de busca e pode abandonar a página se houver apenas alguns ou nenhum produto relevante encontrado nos principais resultados. O abandono da pesquisa prejudica ainda mais as plataformas de comércio eletrônico porque o modelo de negócio depende de transações de compra bem-sucedidas em vez de apenas cliques em anúncios (GOSWAMI et al., 2019). Além disso, o cliente típico não tem uma boa ideia do sortimento ou do estilo dos produtos oferecidos e pode não ter em mente todas as variações de um mesmo produto que esteja procurando. O cliente forma uma primeira impressão sobre o sortimento e o estilo da página de busca após consultar os produtos nas primeiras ordenações (FERREIRA et al., 2022). Portanto, quanto mais tempo o usuário interage com os produtos listados, tem-se um indicador que pode revelar que o índice principal (primeiros produtos na lista oferecida) de busca está eficiente.

O objetivo do presente artigo é avaliar as vantagens competitivas no modelo de aprendizado de máquina (*Maching Learning* – ML) em um caso real de um sistema de busca do comércio eletrônico no segmento de bens para o lar por meio de comparação de resultados entre o método atual da empresa, a ser apresentada, e uma metodologia de solução baseada em um modelo de aprendizagem de ordenação. A empresa é uma das líderes do comércio eletrônico na indústria de bens para o lar. A principal ordenação de resultados do sistema de busca da empresa, denominada método atual, é construída por meio de variáveis recomendadas por especialistas da própria empresa conforme a equação (1). Variáveis podem ser quantidade de vendas, quantidade de visualizações, quantidade de cliques, quantidade de adições ao carrinho, etc.

$$S = 0,4 \times A + 0,6 \times B \quad (1)$$

Por questão de sigilo da empresa, são mantidas A e B para as variáveis. São aplicadas técnicas de normalização de dados a estas variáveis antes dos cálculos para reduzir a redundância dos dados. Após o cálculo individual por produto, a ordenação é a classificação decrescente de S (maior para o menor valor calculado). Sendo assim, existem oportunidades para melhorar a ordenação de produtos buscados por meio de outras técnicas, neste artigo buscou-se aplicar técnicas de aprendizado de máquina.

Este artigo está dividido em 5 sessões. Além desta, na segunda seção apresenta-se o referencial teórico com conceitos, trabalhos relacionados de aprendizagem de ordenação e a principal métrica para avaliar modelos de recuperação de informação. Na terceira sessão apresenta-se o desenvolvimento completo da metodologia proposta, explicando todas as etapas necessárias propostas neste trabalho. A quarta sessão contempla os resultados computacionais dos experimentos deste trabalho. A quinta sessão contém as conclusões deste trabalho e sugestões para trabalhos futuros.

2. Referencial teórico

A Recuperação de Informação (RI) é uma área abrangente da Ciência da Computação que se concentra principalmente em prover aos usuários o acesso fácil às informações de seu interesse. Trata da representação, armazenamento, organização e acesso a itens de informação, como documentos, páginas da internet, catálogos online, registros estruturados e semiestruturados, objetos multimídia, etc. A representação e a organização dos itens de informação devem fornecer aos usuários facilidade de acesso as informações de seu interesse (BAEZA-YATES; RIBEIRO-NETO, 2013). Com a evolução dos modelos de RI e a crescente geração de informação, sistemas de busca modernos levam em consideração inúmeros atributos para melhorar a ordenação para atender a necessidade de um usuário. Aprendizagem de ordenação (*Learning to Rank* - LTR) envolve a aplicação de algoritmos de aprendizado de máquina na otimização da estratégia de ordenação sendo fundamental para ofertar melhores serviços de busca (LI et al., 2020).

O trabalho de Keyhanipour et al. (2016) apresentou um novo algoritmo de LTR, chamado QRC-Rank, o qual é aplicável para conjunto de dados com ou sem informações de cliques de usuários. O QRC-Rank desenvolve um modelo de aprendizagem por reforço com base nessas representações compactas de recursos.

O trabalho de Karmaker Santu et al. (2017) abordou os desafios práticos na aplicação de métodos de LTR na busca do comércio eletrônico. Seu estudo inclui os desafios na aprendizagem de características, obtenção de julgamentos de relevância confiáveis e exploração otimizada de vários sinais de feedback do usuário, como taxas de cliques, taxas de adição ao carrinho, taxas de pedidos e receita. Diversos algoritmos de LTR são utilizados neste trabalho, para avaliar qual destes apresenta melhor desempenho de acordo com as etapas da jornada de compra no comércio eletrônico.

O trabalho de Fonseca (2018) utilizou técnicas de aprendizado de máquina para aprender funções de ordenação de resultados no contexto de comércio eletrônico. Diferentes formas são utilizadas para estimativas de relevância e foram aplicados os métodos LambdaMART, RankNet e SVM em dados reais de empresas de segmentos diferentes.

O trabalho de Hu et al. (2018) usou aprendizagem por reforço para controle da ordenação em cenários de busca de comércio eletrônico. Seu objetivo era desenvolver um método para aprender uma política de ordenação ótima, que maximize as recompensas acumulativas esperadas em uma sessão de pesquisa. Foi reformulado o conceito de processo de decisão de Markov da sessão de pesquisa.

O trabalho de Park et al. (2020) apresentou uma estrutura que aproveita os registros do histórico de cliques dos usuários para compensar as interações de produtos não comprados. Um modelo de aprendizado de máquina chamado P3Stop foi desenvolvido utilizando dados de transações de compra, dados de cliques na listagem de produtos de busca e dados de cliques na listagem de produtos de busca que não resultaram em transações de compras.

O trabalho de Zhang et al. (2020) usou um novo modelo de aprendizado profundo chamado Recuperação Semântica e Personalizada Profunda (*Deep Personalized and Semantic Retrieval* - DPSR) para resolver dois problemas conhecidos na pesquisa de comércio eletrônico: 1) como recuperar itens que são semanticamente relevantes, mas não correspondem exatamente aos termos da consulta, e 2) como recuperar itens que são mais personalizados para usuários diferentes para a mesma consulta de pesquisa.

As métricas de avaliação são importantes para determinar se o modelo de RI está ou não com uma boa performance. Uma das métricas mais utilizadas na área de RI é o Ganho Cumulativo Descontado Normalizado (*Normalized Discounted Cumulative Gain* - NDCG) conforme a equação (2) abaixo:

$$NDCG = \sum \frac{2^{\text{rel}(p_{[i]})} - 1}{\log_2^i + 1} iDCG^{-1} \quad (2)$$

Dado um resultado ordenado de tamanho p , com i sendo a posição de algum documento neste resultado, esta métrica soma o ganho baseado na relevância atribuída $\text{rel}(p_{[i]})$ a cada documento e a divide por um fator de desconto (baseado no logaritmo da posição em que o documento foi apresentado). Este fator é o Ganho Cumulativo Descontado (DCG, do inglês - Discounted Cumulative Gain). O DCG é então dividido pelo DCG ideal ($iDCG$), ou seja, o DCG obtido dos documentos em ordem decrescente de sua relevância real conhecida. Temos assim o NDCG, o DCG normalizado pelo DCG ideal (FONSECA, 2018).

3. Metodologia

A metodologia para a solução do problema real foi baseada em um modelo de LTR de acordo com as implementações das etapas abaixo.

3.1 Estimativa de relevância

No ambiente do comércio eletrônico, é possível estimar a relevância de cada produto a partir de dados reais dos usuários na plataforma. Para este trabalho foram realizadas três alternativas de estimativa de relevância, considerando um período de 90 dias de dados acumulados, conforme equações (3), (4) e (5).

$$TCP = \frac{\#P}{\#V} \quad (3)$$

$$TAC = \frac{\#A}{\#I} \quad (4)$$

$$TCB = \frac{\#C}{\#I} \quad (5)$$

TCP é a taxa de conversão de pedidos.

TAC é a taxa de adição no carrinho.

TCB é a taxa de cliques na busca.

#P é a quantidade de pedidos.

#V é a quantidade de visualizações do produto.

#A é a quantidade de adição o carrinho.

#I é o número de impressões na busca.

#C é o número de cliques na busca.

3.2 Método de aprendizado de máquina

O método de aprendizagem de máquina escolhido para os experimentos foi o LambdaMART que é um modelo de floresta aleatória (*Random Forest* – RF), ou seja, o conjunto de árvores de decisão definem regras para tomada de decisão. O algoritmo criará uma estrutura similar a um fluxograma, com “nós” onde uma condição é verificada, e se atendida o fluxo segue por um ramo, caso contrário, por outro, sempre levando ao próximo nó, até a finalização da árvore. A implementação do método escolhido foi a encontrada no pacote LightGBM com configuração padrão. A justificativa da escolha deste método foi o resultado do trabalho de Karmaker Santu et al. (2017) que trouxe o modelo LambdaMART entre os melhores resultados dentro os modelos LTR aplicados em seu trabalho, que se assemelha com o presente trabalho proposto e por se tratar de dados de comércio eletrônico.

3.3 Experimentos

Antes da inicialização dos experimentos com o método de aprendizagem de máquina escolhido, realizou-se experimentos simulados, alterando apenas os pesos das variáveis do atual método de ordenação conforme equação (1). O objetivo foi avaliar se, necessariamente, era preciso criar um método mais robusto para atingir uma melhoria no NDCG do método atual. Nestas simulações, os parâmetros de entrada foram os pesos das variáveis A e B, ou seja, altera-se os pesos das variáveis conforme a Tabela 1. Por exemplo, para o experimento 5, considerou-se 0,5 para a variável A e 0,5 para a variável B. Portanto, em cada experimento o parâmetro de saída foi a lista ordenada de produtos conforme a ordenação decrescente de S.

Tabela 1 - Experimentos com alteração dos pesos das variáveis

Categoria X, Y e Z	Variável A	Variável B
Resultado Base	0,4	0,6
Experimento 1	0	1
Experimento 2	0,1	0,9
Experimento 3	0,2	0,8
Experimento 4	0,3	0,7
Experimento 5	0,5	0,5
Experimento 6	0,6	0,4
Experimento 7	0,7	0,3
Experimento 8	0,8	0,2
Experimento 9	0,9	0,1
Experimento 10	1	0

Fonte: O autor (2023)

Após estes experimentos simulados, aplicou-se o modelo LTR escolhido em mais 3 experimentos alternando a quantidade de variáveis conforme a Tabela 2. Os dois objetivos foram: 1) atingir melhoria no NDCG do método atual e, 2) avaliar se o número de variáveis traz incremento ao resultado NDCG. Portanto, os parâmetros de entrada foram as variáveis e estimativas de relevância utilizadas no modelo e o parâmetro de saída foi a lista ordenada de produtos conforme o resultado de previsão do modelo (*predicted_ranking*).

Tabela 2 - Experimentos com alteração do número de variáveis

Categoria X, Y e Z	Número de Variáveis	Variáveis
LambdaMART 1	2	A,B
LambdaMART 2	3	A,B,C
LambdaMART 3	11	A,B,C,D,E,F,G,H,I,J,K

Fonte: O autor (2023)

3.4 Conjuntos de dados

Todos os experimentos foram rodados usando conjuntos de dados de 3 categorias estratégicas da empresa. Categorias podem ser guarda-roupas, sofás, escrivaninhas, armários de cozinha, etc. Por sigilo, foram denominadas X, Y e Z. Até a data deste trabalho, a categoria X possuía um número total de 41.138 produtos, a categoria Y possuía 33.639 produtos e a categoria Z possuía 68.836 produtos. Para cada conjunto de dados, foram realizados dois tratamentos prévios: 1) filtrados apenas produtos que tiveram vendas nos últimos 30 dias e que possuem, individualmente, pelo menos a média 3 estrelas de avaliação no comércio eletrônico e, 2) filtrados apenas produtos que possuem o número de visualizações de páginas de produto acima da mediana. Este procedimento de tratamento de dados foi uma recomendação de especialistas da empresa responsáveis pelas categorias estratégicas testadas. O objetivo principal de eliminar produtos sem relevância

no comércio eletrônico da empresa como, por exemplo, produtos novos, produtos com poucas vendas, produtos com baixa avaliação, etc.

4. Resultados

Conforme recomendação de especialistas da empresa, a qualidade do sistema de busca do comércio eletrônico foi baseada na ordenação das 100 primeiras posições dos produtos listados do principal termo buscado da categoria estratégica. Sendo assim, em todos os experimentos utilizou-se a métrica $NDCG@k$, para $k \in \{10,25,50,75,100\}$ e considerando como $iDCG$ a ordenação da estimativa de relevância escolhida. Para os resultados a seguir, em destaque negrito, estão todos os valores superiores ao valor referência (resultado base).

4.1 Estimativa de relevância como taxa de conversão de pedidos

A seguir temos os resultados consolidados dos experimentos das 3 categorias usando a estimativa de relevância como taxa de conversão de pedidos.

Tabela 3 - Experimentos da categoria x usando taxa de conversão de pedidos

Categoria X	NDCG@10	NDCG@25	NDCG@50	NDCG@75	NDCG@100
Resultado Base	0,916642	0,793905	0,802118	0,829422	0,830639
Experimento 1	0,961521	0,883843	0,903808	0,898364	0,895763
Experimento 2	0,923042	0,883843	0,894902	0,877502	0,877540
Experimento 3	0,923042	0,869384	0,889841	0,865857	0,855534
Experimento 4	0,916642	0,837283	0,837120	0,846243	0,852330
Experimento 5	0,874963	0,779446	0,784305	0,819623	0,811387
Experimento 6	0,868563	0,777407	0,747961	0,792603	0,803047
Experimento 7	0,868563	0,772367	0,726744	0,776645	0,788058
Experimento 8	0,851877	0,727133	0,717837	0,762890	0,762955
Experimento 9	0,791605	0,713568	0,700117	0,751080	0,755832
Experimento 10	0,780062	0,685544	0,687493	0,735987	0,738196
LambdaMART 1	0,923042	0,93160	0,952499	0,944679	0,942878
LambdaMART 2	0,991657	0,975892	0,969063	0,964859	0,962366
LambdaMART 3	0,961521	0,964185	0,970408	0,965552	0,957995

Fonte: O autor (2023)

Para a categoria X, na Tabela 3, os experimentos 1, 2 e 3 atingiram resultados superiores em todas as métricas quando comparado ao resultado base. O experimento 4, atingiu resultados superiores em 4 das métricas e resultado igual ao resultado base para $NDCG@10$. No entanto, quando observados os resultados dos experimentos do modelo de aprendizado de máquina LambdaMART, todas as métricas estão com resultados acima do resultado base, inclusive sendo os melhores resultados atingidos.

Tabela 4 - Experimentos da categoria y usando taxa de conversão de pedidos

Categoria Y	NDCG@10	NDCG@25	NDCG@50	NDCG@75	NDCG@100
Resultado Base	0,662741	0,768131	0,798057	0,827410	0,794315
Experimento 1	0,628693	0,636021	0,656954	0,690318	0,702195
Experimento 2	0,880078	0,886333	0,892696	0,888230	0,875746
Experimento 3	0,836115	0,847726	0,862035	0,877267	0,850655
Experimento 4	0,748188	0,821468	0,831546	0,836220	0,835394
Experimento 5	0,650565	0,745301	0,775315	0,809437	0,793750
Experimento 6	0,650565	0,719043	0,744654	0,783657	0,776255
Experimento 7	0,640869	0,671357	0,709479	0,745953	0,738808
Experimento 8	0,640869	0,648527	0,691724	0,732045	0,737409
Experimento 9	0,628693	0,642975	0,661468	0,718136	0,711943
Experimento 10	0,628693	0,636021	0,656954	0,690318	0,702195
LambdaMART 1	0,994833	0,956889	0,951996	0,943865	0,941483
LambdaMART 2	0,992354	0,978940	0,969820	0,952902	0,956150

LambdaMART 3	0,992354	0,980966	0,954593	0,943823	0,948861
--------------	-----------------	-----------------	-----------------	-----------------	-----------------

Fonte: O autor (2023)

Para a categoria Y, na Tabela 4, os experimentos 2, 3 e 4 atingiram resultados superiores em todas as métricas quando comparado ao resultado base. Quando observados os resultados dos experimentos do modelo de aprendizado de máquina LambdaMART, todas as métricas estão com resultados superiores, sendo os melhores resultados atingidos.

Tabela 5 - Experimentos da categoria z usando taxa de conversão de pedidos

Categoria Z	NDCG@10	NDCG@25	NDCG@50	NDCG@75	NDCG@100
Resultado Base	0,730440	0,745971	0,780625	0,762924	0,764631
Experimento 1	0,819593	0,843402	0,878753	0,861782	0,853793
Experimento 2	0,819593	0,829170	0,869004	0,839685	0,836337
Experimento 3	0,807675	0,809320	0,841100	0,822416	0,813709
Experimento 4	0,763099	0,775578	0,794572	0,789274	0,791081
Experimento 5	0,732426	0,726121	0,726538	0,724451	0,736524
Experimento 6	0,675931	0,675271	0,688036	0,693403	0,712622
Experimento 7	0,675931	0,652085	0,666355	0,667498	0,694559
Experimento 8	0,675931	0,633376	0,630233	0,645645	0,663655
Experimento 9	0,631355	0,614667	0,603007	0,626585	0,645532
Experimento 10	0,598696	0,591480	0,590411	0,599845	0,625638
LambdaMART 1	0,947858	0,971151	0,980779	0,968771	0,970485
LambdaMART 2	0,964992	0,965872	0,975085	0,969573	0,966901
LambdaMART 3	0,972193	0,974869	0,978344	0,973596	0,970912

Fonte: O autor (2023)

Para a categoria Z, na Tabela 5, os experimentos 1, 2, 3 e 4 atingiram resultados superiores em todas as métricas quando comparado ao resultado base e o experimento 5 apenas para NDCG@10. Quando observados os resultados dos experimentos do modelo de aprendizado de máquina LambdaMART, todas as métricas estão com resultados superiores para o NDCG, sendo os melhores resultados atingidos.

4.2 Estimativa de relevância como taxa de cliques na busca

A seguir temos os resultados consolidados dos experimentos das 3 categorias usando a estimativa de relevância como taxa de clique na busca.

Tabela 6 - Experimentos da categoria x usando taxa de cliques na busca

Categoria X	NDCG@10	NDCG@25	NDCG@50	NDCG@75	NDCG@100
Resultado Base	0,938777	0,904079	0,843456	0,829396	0,828368
Experimento 1	0,859362	0,848844	0,822841	0,820247	0,819256
Experimento 2	0,886963	0,850079	0,823592	0,820243	0,825907
Experimento 3	0,909860	0,869755	0,839278	0,818211	0,831189
Experimento 4	0,938777	0,881092	0,841942	0,835412	0,823461
Experimento 5	0,938777	0,907390	0,846919	0,823830	0,836255
Experimento 6	0,944798	0,926584	0,849583	0,835015	0,845746
Experimento 7	0,921901	0,945777	0,856062	0,823432	0,840195
Experimento 8	0,944798	0,945295	0,880119	0,827697	0,834779
Experimento 9	0,949502	0,945295	0,880119	0,828878	0,837016
Experimento 10	0,948185	0,927336	0,880119	0,833890	0,833803
LambdaMART 1	0,936144	0,915776	0,879373	0,860601	0,862942
LambdaMART 2	0,995296	0,982571	0,909146	0,885721	0,890315
LambdaMART 3	0,977103	0,972075	0,956470	0,936399	0,934199

Fonte: O autor (2023)

Para a categoria X, na Tabela 6, os experimentos 6 e 10 atingiram resultados superiores em todas as métricas quando comparado ao resultado base. Os experimentos 3 e 4, atingiram resultados superiores em apenas uma das métricas avaliadas, enquanto os experimentos 5 e 7 em três das métricas avaliadas. Os experimentos 8 e 9 só não atingiram resultados maiores ao resultado base para NDCG@75. Quando observados os resultados dos experimentos do modelo de aprendizado de máquina LambdaMART 2 e 3, todas as métricas estão com resultados superiores e apenas o LambdaMART 1 não superou o resultado base para NDCG@10. Ainda assim, os melhores resultados atingidos são com o modelo de LTR.

Tabela 7 - Experimentos da categoria y usando taxa de cliques na busca

Categoria Y	NDCG@10	NDCG@25	NDCG@50	NDCG@75	NDCG@100
Resultado Base	0,906712	0,864185	0,802310	0,825021	0,818966
Experimento 1	0,827680	0,758418	0,770673	0,810522	0,791084
Experimento 2	0,827680	0,796568	0,772175	0,818050	0,798400
Experimento 3	0,814901	0,796568	0,784358	0,812724	0,810218
Experimento 4	0,857983	0,837666	0,801550	0,823155	0,816249
Experimento 5	0,920448	0,873917	0,835492	0,831951	0,815517
Experimento 6	0,920448	0,904624	0,832065	0,834255	0,829330
Experimento 7	0,920448	0,907571	0,834158	0,843154	0,823066
Experimento 8	0,920448	0,907571	0,834538	0,849149	0,829905
Experimento 9	0,920448	0,900127	0,853824	0,853423	0,823642
Experimento 10	0,920448	0,894312	0,856660	0,855392	0,831478
LambdaMART 1	0,946006	0,900127	0,863817	0,867019	0,869072
LambdaMART 2	0,933227	0,909199	0,914023	0,918473	0,916275
LambdaMART 3	0,998132	0,975381	0,959297	0,949222	0,944428

Fonte: O autor (2023)

Para a categoria Y, na Tabela 7, a partir do experimento 6 até o experimento 10, os resultados são superiores em todas as métricas. O experimento 5 não superou apenas para NDCG@100. Os experimentos do modelo de aprendizado de máquina LambdaMART, superaram todas as métricas e os melhores resultados atingidos fazem parte do LambdaMART 3.

Tabela 8 - Experimentos da categoria z usando taxa de cliques na busca

Categoria Z	NDCG@10	NDCG@25	NDCG@50	NDCG@75	NDCG@100
Resultado Base	0,947916	0,859072	0,853583	0,852916	0,869296
Experimento 1	0,947916	0,870475	0,857941	0,855679	0,851703
Experimento 2	0,947916	0,870475	0,857267	0,860706	0,852742
Experimento 3	0,947916	0,858844	0,857941	0,864605	0,862807
Experimento 4	0,947916	0,846412	0,850987	0,858533	0,869339
Experimento 5	0,947916	0,870400	0,855887	0,862887	0,863720
Experimento 6	0,941989	0,841749	0,839155	0,843277	0,865401
Experimento 7	0,941989	0,840418	0,845363	0,844401	0,854655
Experimento 8	0,941989	0,839086	0,847375	0,840244	0,847129
Experimento 9	0,914332	0,835203	0,839536	0,831788	0,827729
Experimento 10	0,917031	0,848566	0,836559	0,829603	0,834349
LambdaMART 1	0,978271	0,929263	0,905473	0,901133	0,900261
LambdaMART 2	0,975573	0,945217	0,926983	0,899738	0,902848
LambdaMART 3	0,969645	0,985381	0,955286	0,952825	0,950675

Fonte: O autor (2023)

Para a categoria Z, na Tabela 8, nenhum experimento de 1 a 10 superou todos os valores das métricas. Os melhores foram os experimentos 1, 2 e 5 que tiveram 3 métricas maiores que o resultado base, enquanto os experimentos 3 e 4 tiveram apenas 2 métricas maiores.

No entanto, novamente os experimentos do modelo de aprendizado de máquina LambdaMART, superaram todas as métricas com os melhores resultados atingidos.

4.3 Estimativa de relevância como taxa de adição no carrinho

A seguir temos os resultados consolidados dos experimentos das 3 categorias usando a estimativa de relevância como taxa de adição no carrinho.

Tabela 9 - Experimentos da categoria x usando taxa de adição no carrinho

Categoria X	NDCG@10	NDCG@25	NDCG@50	NDCG@75	NDCG@100
Resultado Base	0,963985	0,903975	0,766015	0,789049	0,758641
Experimento 1	0,937541	0,878009	0,760146	0,759913	0,757517
Experimento 2	0,937541	0,878009	0,766304	0,753713	0,761441
Experimento 3	0,937541	0,877166	0,773051	0,762506	0,765895
Experimento 4	0,963985	0,890570	0,780675	0,754609	0,760571
Experimento 5	0,963985	0,918223	0,772173	0,760810	0,757580
Experimento 6	0,968771	0,918223	0,766607	0,787990	0,758980
Experimento 7	0,995215	0,937613	0,763096	0,788411	0,756457
Experimento 8	1,000000	0,933355	0,754595	0,782211	0,757857
Experimento 9	0,995215	0,933355	0,753717	0,781573	0,755587
Experimento 10	0,995215	0,919107	0,752839	0,779240	0,751133
LambdaMART 1	0,990430	0,966953	0,824655	0,830175	0,843912
LambdaMART 2	0,990430	0,930745	0,881697	0,861438	0,861658
LambdaMART 3	0,906312	0,859503	0,781428	0,792701	0,842355

Fonte: O autor (2023)

Para a categoria X, na Tabela 9, o experimento 6 foi o que teve melhor performance comparado ao resultado base, superando 4 métricas, enquanto os experimentos 2, 3, 4, 5, 7, 8, 9, 10 só superaram 2 métricas. Os experimentos do modelo de aprendizado de máquina LambdaMART 1 e 2, superaram todas as métricas, enquanto o LambdaMART 3 superou apenas 3 métricas. Com exceção do NDCG@10, os melhores resultados das métricas fazem parte do modelo proposto.

Tabela 10 - Experimentos da categoria y usando taxa de adição no carrinho

Categoria Y	NDCG@10	NDCG@25	NDCG@50	NDCG@75	NDCG@100
Resultado Base	0,951521	0,903708	0,777906	0,781836	0,769378
Experimento 1	0,908831	0,819795	0,790448	0,742192	0,798518
Experimento 2	0,908831	0,819795	0,802869	0,764949	0,755237
Experimento 3	0,908831	0,819795	0,815291	0,766665	0,756725
Experimento 4	0,945732	0,881182	0,815291	0,775967	0,761564
Experimento 5	0,951521	0,906803	0,788288	0,791138	0,769378
Experimento 6	0,951521	0,909897	0,792366	0,792854	0,775704
Experimento 7	0,951521	0,912992	0,800709	0,800439	0,780166
Experimento 8	0,951521	0,912992	0,800709	0,800439	0,780166
Experimento 9	0,951521	0,912992	0,802748	0,800439	0,772352
Experimento 10	0,951521	0,912992	0,792366	0,800439	0,773839
LambdaMART 1	0,994211	0,971285	0,861085	0,830199	0,852481
LambdaMART 2	0,994211	0,971285	0,900509	0,906572	0,881935
LambdaMART 3	0,945732	0,920045	0,879744	0,843346	0,846155

Fonte: O autor (2023)

Para a categoria Y, na Tabela 10, os experimentos de 6 a 10 tiveram melhores resultados quando comparados ao resultado base, superando 4 métricas. O experimento 5 superou o em apenas 3 métricas, seguido dos experimentos 1 com apenas 2 métricas e os experimentos 2, 3 e 4 com apenas 1 métrica vencedora. Os experimentos do modelo de

aprendizado de máquina LambdaMART 1 e 2, tiveram todos resultados superiores e o LambdaMART 3 só não superou o NDCG@10. Ainda assim, os maiores resultados foram atingidos pelo modelo proposto.

Tabela 11 - Experimentos da categoria z usando taxa de adição no carrinho

Categoria Z	NDCG@10	NDCG@25	NDCG@50	NDCG@75	NDCG@100
Resultado Base	0,942502	0,874800	0,808283	0,788636	0,792665
Experimento 1	0,987487	0,931695	0,866094	0,802014	0,810976
Experimento 2	0,987487	0,931695	0,833828	0,807527	0,810976
Experimento 3	0,942502	0,895356	0,838013	0,784740	0,809335
Experimento 4	0,942502	0,895356	0,808283	0,793056	0,788365
Experimento 5	0,938723	0,874800	0,792379	0,770411	0,796408
Experimento 6	0,932150	0,868163	0,779234	0,763855	0,764737
Experimento 7	0,932150	0,861526	0,777965	0,758484	0,763829
Experimento 8	0,932150	0,856409	0,776996	0,754014	0,759353
Experimento 9	0,932150	0,853091	0,770953	0,730277	0,735658
Experimento 10	0,932150	0,853091	0,770953	0,742846	0,746852
LambdaMART 1	0,927194	0,886457	0,879909	0,827746	0,839462
LambdaMART 2	0,913697	0,886737	0,925321	0,919726	0,914397
LambdaMART 3	0,974974	0,942113	0,915392	0,874294	0,868269

Fonte: O autor (2023)

Para a categoria Z, na Tabela 11, os experimentos 1 e 2 tiveram resultados melhores em todas as métricas quando comparados ao resultado base. O experimento 3 superou em apenas 3 métricas, seguido do experimento 4 com 2 métricas e o experimento 5 com apenas 1 métrica vencedora. Os resultados dos experimentos do modelo LambdaMART 3 venceu todas as métricas e os demais só não tiveram melhores resultados em apenas 1 métrica. Ainda assim, o LambdaMART 2 teve maiores resultados para as métricas NDCG@50, NDCG@75 e NDCG@100, seguido do LambdaMART 3 que teve o maior resultado para NDCG@25. Os experimentos 1 e 2 tiveram o maior resultado para NDCG@10.

4.4 Resultados consolidados por experimento

A seguir temos o consolidado de todos os experimentos deste trabalho, o qual cada métrica teve a chance de superar o resultado base até 9 vezes.

Tabela 12 - Resultados consolidados por experimento

Consolidado	NDCG@10	NDCG@25	NDCG@50	NDCG@75	NDCG@100	> Resultado Base
Experimento 1	3	4	5	4	4	44%
Experimento 2	4	5	7	5	5	58%
Experimento 3	3	4	7	4	6	53%
Experimento 4	2	4	5	6	5	49%
Experimento 5	1	5	5	3	2	36%
Experimento 6	3	4	4	3	4	40%
Experimento 7	2	4	3	2	3	31%
Experimento 8	3	4	3	2	3	33%
Experimento 9	3	4	3	2	3	33%
Experimento 10	3	4	3	3	3	36%
LambdaMART 1	7	9	9	9	9	96%
LambdaMART 2	8	9	9	9	9	98%
LambdaMART 3	7	8	9	9	9	93%

Fonte: O autor (2023)

Com esses resultados consolidados, pode-se concluir que os experimentos do modelo proposto atingiram performance melhores que o resultado base de 96% para LambdaMART 1, 98% para LambdaMART 2 e 93% para LambdaMART 3 mostrando muito mais eficiência quando comparado aos experimentos com a regra atual. A melhor performance dos experimentos alternando os pesos da metodologia atual foi obtido com o experimento 2 que obteve 58% dos resultados melhores que o resultado base.

5. Conclusões

O presente trabalho abordou um estudo de caso com uso de técnicas de aprendizado de máquina para ordenação inteligente de resultados de busca com dados reais de um comércio eletrônico do segmento de bens para o lar. A partir da literatura sobre *Learning to Rank* (LTR) foi construído um programa computacional do método escolhido. A justificativa do trabalho é o crescente avanço da inteligência artificial e desafios no ambiente corporativo somando-se a crescente pesquisas acadêmicas do tema proposto.

Ao analisar os resultados dos experimentos, com base na métrica NDCG@k, concluiu-se a necessidade de construção de um modelo com maior robustez. Os resultados apresentados revelam que somente ajustes de parâmetros nos pesos das variáveis do método atual, não trouxe resultados melhores. Dos 10 experimentos testados, apenas os experimentos 2 e 3 ficaram acima de 50% com 58% e 53%, respectivamente.

Como apresentado anteriormente, a proposta foi a elaboração de um modelo com maior robustez para o problema real de um sistema de busca no comércio eletrônico de uma indústria de bens para o lar, por meio de técnicas de aprendizado de máquina. Sendo assim, após aplicada uma técnica de LTR, pode-se concluir que os resultados são superiores. Conforme a Tabela 12 de resultados consolidados por experimento, o modelo proposto nos 3 experimentos testados obteve uma performance maior que 90%, com destaque ao LambdaMART 2 muito próximo do 100%.

Não obstante, vale ressaltar que o aumento do número de variáveis, não necessariamente aumenta a performance do modelo LTR. Em alguns casos até pioraram as métricas apresentadas. Por isso, vale sempre ter cautela na implementação de novas variáveis no modelo, devendo fazer os testes estatísticos adequados sobre as variáveis que explicam o fenômeno para serem utilizadas de entrada nos modelos de aprendizado de máquina. Portanto, foi possível concluir que a aplicação de uma técnica LTR fornece resultados melhores (maior NDCG@k) em relação ao método atual da empresa.

5.1 Sugestões para trabalhos futuros

Ao que tange o campo de estudo LTR, existem diversos modelos sendo desenvolvidos e abordados na literatura acadêmica como o apresentado neste trabalho. Este trabalho se limitou ao uso do modelo LambdaMART e comparar com outros modelos de LTR pode ser uma alternativa para avaliar se existe um modelo que performa melhor no estudo de caso. Além disso, uma outra possibilidade é aplicar na prática testes online, ou seja, testes A/B no comércio eletrônico para dar mais embasamento nos resultados apresentados e se de fato houve melhoria na receita da empresa.

O modelo apresentado também se limitou a aplicação de estimativas de relevância baseada em avaliação de usuários como taxa de conversão de pedidos, taxa de cliques na busca e taxa de adição no carrinho. Recomenda-se usar outros métodos de relevância para estudos futuros como, por exemplo, estimativas de relevância que considera o viés da posição do produto na página de listagem de busca.

Por fim, este trabalho também se restringiu a um segmento específico de um comércio eletrônico, uma indústria de bens para o lar com apenas 3 categorias estratégicas sendo

testadas, e pode ser interessante aplicar o modelo LambdaMART para diferentes segmentos do comércio eletrônico e avaliar se os resultados também serão positivos.

Referências

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. **Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca**. Bookman Editora, 2013.

FERREIRA, Kris J.; PARTHASARATHY, Sunanda; SEKAR, Shreyas. Learning to rank an assortment of products. **Management Science**, v. 68, n. 3, p. 1828-1848, 2022.

FONSECA, Roberto Cidade. **Learning to rank para busca em Comércio Eletrônico**. Dissertação – Universidade Federal do Amazonas. 2018.

GOSWAMI, Anjan; ZHAI, Chengxiang; MOHAPATRA, Prasant. Learning to Diversify for E-commerce Search with Multi-Armed Bandit. In: **eCOM@ SIGIR**. 2019.

HU, Yujing et al. Reinforcement learning to rank in e-commerce search engine: Formalization, analysis, and application. In: **Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining**. 2018. p. 368-377.

KARMAKER SANTU, Shubhra Kanti; SONDHAI, Parikshit; ZHAI, ChengXiang. On application of learning to rank for e-commerce search. In: **Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval**. 2017. p. 475-484.

KEYHANIPOUR, Amir Hosein et al. Learning to rank with click-through features in a reinforcement learning framework. **International Journal of Web Information Systems**, 2016.

LI, Pengcheng et al. Improving multi-scenario learning to rank in e-commerce by exploiting task relationships in the label space. In: **Proceedings of the 29th ACM International Conference on Information & Knowledge Management**. 2020. p. 2605-2612.

PARK, Chanyoung et al. Click-aware purchase prediction with push at the top. **Information Sciences**, v. 521, p. 350-364, 2020.

RITA, Paulo; RAMOS, Ricardo F. Global research trends in consumer behavior and sustainability in E-Commerce: A bibliometric analysis of the knowledge structure. **Sustainability**, v. 14, n. 15, p. 9455, 2022.

SOROKINA, Daria; CANTU-PAZ, Erick. Amazon search: The joy of ranking products. In: **Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval**. 2016. p. 459-460.

ZHANG, Han et al. Towards personalized and semantic retrieval: An end-to-end solution for e-commerce search via embedding learning. In: **Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval**. 2020. p. 2407-2416.

ZHUANG, Tao; OU, Wenwu; WANG, Zhirong. Globally optimized mutual influence aware ranking in e-commerce search. **arXiv preprint arXiv:1805.08524**, 2018.