

**Nome da área temática:** Estudos Organizacionais

**Título do trabalho:** APRENDIZAGEM DE ORDENAÇÃO NO COMÉRCIO ELETRÔNICO DE UMA INDÚSTRIA DE BENS PARA O LAR

**Resumo:** A pesquisa aplica técnicas de aprendizado de máquina para a ordenação de resultados de busca que será utilizada no comércio eletrônico de uma empresa, com intuito de melhorar a jornada do usuário após uso do motor de busca. O método de pesquisa adotado é o estudo de caso, baseado em conhecimentos obtidos por meio de uma revisão da literatura nas áreas de Recuperação de Informação e Aprendizagem de Ordenação. Para estimativa de relevância, é utilizado algumas alternativas de acordo com dados reais de usuários que usaram a plataforma. É apresentado um caso prático com uso de técnicas de aprendizado de máquina aplicado a um motor de busca de um comércio eletrônico que pode ser replicado em outras empresas similares ao estudo de caso.

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

**Abstract:** The research aims to apply machine learning techniques for the ordering of search results that will be used in a company's e-commerce, in order to improve the user's journey after using the search engine. The research method adopted is the case study, based on knowledge obtained through a literature review in the areas of Information Retrieval and Learning to Rank. To estimate the relevance of the products, some alternatives are used according to real data from users who used the platform. It is showed a practical case of use with machine learning techniques applied to an e-commerce search engine that can be replicated in other companies similar to the case study.

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

## 1. Introdução

A busca no comércio eletrônico tem crescido muito com o avançar dos anos. No ano de 2020, as vendas na internet representaram 17,8% das vendas totais, 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, pois está bem estabelecido na literatura acadêmica e conhecido pelos varejistas que mais clientes visualizam produtos ordenados nas primeiras posições em comparação com produtos ordenados na parte inferior da página (FERREIRA et al., 2022). É por isso que empresas como a Amazon, investe nesta área há alguns anos utilizando a inteligência artificial na ordenação de produtos do sistema de busca de sua plataforma (SOROKINA; CANTU-PAZ, 2016).

Durante a jornada de busca, é retornada ao usuário uma lista com produtos mais relevantes de acordo com o termo buscado e o que a empresa define como regra para ordenação de produtos. Neste momento, o consumidor compara os produtos e pode selecionar àqueles com mais 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 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 um determinado 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 pode ser um sinal que revela que o índice principal de busca está eficiente.

O objetivo do presente artigo é avaliar as vantagens competitivas no modelo de aprendizado de máquina em um caso real de um mecanismo de busca do comércio eletrônico no segmento de bens para o lar por meio de comparação de resultados entre o método atual da empresa e uma metodologia de solução baseada em um modelo de Learning to Rank (LTR).

A empresa é uma das líderes do comércio eletrônico no segmento de bens para o lar. No cenário atual, a principal ordenação de resultados de busca da empresa, denominado modelo atual, é construída através de atributos de produto recomendados por especialistas da própria empresa conforme a equação (1) abaixo.

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

Por questão de sigilo da empresa, vamos manter A e B para as variáveis.

São aplicadas técnicas de normalização de dados a estes atributos antes dos cálculos para reduzir a redundância dos dados. Após o cálculo, a ordenação de produtos é a classificação decrescente de S (maior para o menor valor calculado). Sendo assim, existem oportunidades para melhorar a ordenação de produtos de busca através de técnicas de aprendizado de máquina.

Este artigo está dividido em 5 sessões. A primeira, contextualiza o tema geral do trabalho, descreve rapidamente o problema que pretende solucionar ou discutir, e aponta o objetivo de sua realização. A segunda seção, apresenta o referencial teórico com conceitos, trabalhos relacionados de aprendizagem de ordenação e a principal métrica para avaliar modelos RI. A terceira sessão, apresenta o desenvolvimento completo da metodologia proposta, explicando todas as etapas necessárias abordadas neste trabalho. A quarta sessão, apresenta os resultados computacionais dos experimentos deste trabalho. A quinta sessão apresenta as considerações finais 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, ou seja, 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 muitos atributos para melhorar a ordenação para atender a necessidade de um usuário. Aprendizagem de ordenação (LTR, do inglês - Learning to Rank) envolve a aplicação de algoritmos de aprendizado de máquina na otimização da estratégia de ordenação e é a técnica fundamental para facilitar melhores serviços de busca de itens (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, incluindo 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 o que melhor performa 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, ou seja, 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 Zhuang et al. (2018) desenvolveu um modelo que otimiza diretamente o Volume Bruto de Mercadorias (GMV, do inglês – Gross Merchandise Volume) com uso de rede neural recorrente para capturar influências relacionadas a ordenação de pedidos na estimativa de probabilidade de compra.

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 (DPSR, do inglês - Deep Personalized and Semantic Retrieval) para resolver dois problemas conhecidos na pesquisa de comércio eletrônico: como recuperar itens que são semanticamente relevantes, mas não correspondem exatamente aos termos da consulta, e 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 (NDCG, do inglês - Normalized Discounted Cumulative Gain) conforme a equação (2) abaixo:

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

Dado um resultado ordenado de tamanho  $l$ , com  $i$  sendo a posição de algum documento neste resultado, esta métrica soma o ganho baseado na relevância atribuída  $\text{rel}(I_{[i]})$  a cada documento e a divide por um fator de desconto (baseado no log 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

#### 3.1 Estimativa de relevância

No ambiente do comércio eletrônico, é possível estimar a relevância de cada documento a partir de *feedback* dos usuários com os produtos (visualizações, cliques e conversões). Portanto, para este trabalho serão realizadas três alternativas de

estimativa de relevância, considerando um período acumulado de 90 dias, conforme equações (3), (4) e (5) abaixo.

$$\text{Taxa de conversão de pedidos} = \frac{\text{Quantidade de pedidos}}{\text{Número de visualizações do produto}} \quad (3)$$

$$\text{Taxa de adição no carrinho} = \frac{\text{Quantidade de adição no carrinho}}{\text{Número de impressões na busca}} \quad (4)$$

$$\text{Taxa de cliques na busca} = \frac{\text{Quantidade de cliques na busca}}{\text{Número de impressões na busca}} \quad (5)$$

### 3.2 Método de aprendizado de máquina

O método de aprendizagem de máquina escolhido para os experimentos foi o LambdaMART, o qual possui a abordagem listwise. LambdaMART é um modelo de floresta aleatória, 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 com os melhores resultados dentro os modelos LTR aplicados em seu trabalho, que se assemelha com o presente trabalho proposto, por se tratar de dados de comércio eletrônico.

### 3.3 Experimentos

Antes de iniciarmos os experimentos com o método de aprendizagem de máquina escolhido, realizou-se experimentos offline alterando apenas os pesos das variáveis do atual método de ordenação conforme tabela abaixo. O objetivo de realizar esses experimentos antes do método de aprendizado de máquina é para avaliar se necessariamente é preciso rodar um modelo inteligente para atingir uma otimização. Neste caso, os parâmetros de entrada são os pesos das variáveis A usando a equação 2.1, ou seja, altera-se os pesos das variáveis conforme a tabela 1. Por exemplo, para o experimento 5, considera-se 0,5 para a variável A e 0,5 para a variável B. Portanto, o parâmetro de saída será o escore e sempre irá ser alterado de acordo com o experimento adotado.

Tabela 1 - Experimentos com alteração dos pesos das variáveis

Categoria X, Y e Z	Variável A	Variável B
Baseline	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)

Para avaliar os experimentos offline utilizou-se a métrica  $NDCG@k$ , para  $k \in \{10,25,50,75,100\}$  considerando como  $iDCG$  a ordenação da variável dependente escolhida, ou seja, da estimativa de relevância. Nesta variação, computamos as métricas apenas até a posição  $k$ , ao invés de para todo o resultado pois considera-se que as 100 primeiras posições são as mais importantes para o comércio eletrônico de acordo com especialistas da empresa.

Após estes experimentos iniciais, utilizou-se o método LTR escolhido com base nas estimativas de relevância. Para o modelo LambdaMART foram escolhidos 3 experimentos diferentes conforme tabela abaixo. O objetivo de realizar esses experimentos aumentando o número de variáveis é para avaliar se tem impacto positivo no resultado. Neste caso, os parâmetros de entrada serão as próprias variáveis conforme Tabela 2, a qual para cada experimento irão ser alteradas e o parâmetro de saída será a estimativa de relevância.

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 vão rodar em 3 categorias “estratégicas” da empresa, as quais, por sigilo, são denominadas X, Y e Z. Para a data deste trabalho, a categoria X possui um número total de 41.138 produtos, a categoria Y possui 33.639 produtos e a categoria Z possui 68.836 produtos. Desta forma, serão usados 3 conjuntos de dados de acordo com a respectiva categoria. Foi acordado com a empresa que será necessário um tratamento prévio para eliminar ruídos da base de dados. Sendo assim, para cada conjunto de dados, será filtrado apenas produtos que tiveram vendas nos últimos 30 dias e que possuem, individualmente, pelo menos a média 3 estrelas de avaliação no comércio eletrônico. Depois, após este tratamento inicial, será filtrado em cada conjunto de dados apenas produtos que possuem o número de visualizações de páginas de produto acima da mediana. Este procedimento de tratamento de dados é uma recomendação de especialistas da empresa responsáveis pelas categorias estratégicas definidas. O objetivo principal é obter resultados mais refinados para análise a fim de extrair valor para os produtos mais relevantes.

## 4. Resultados

Para os resultados abaixo, estão em negritos todos os valores superiores ao baseline para cada métrica  $NDCG@k$ . Vale ressaltar que para os experimentos iniciais, ou seja, sem aplicação do modelo LTR, comparou-se a ordenação do score usando a estimativa de relevância como ordenação ideal tanto para o baseline quanto para os demais experimentos. Para os experimentos com modelo inteligente, comparou-se a

ordenação do *output* do modelo (*predicted\_ranking*) também usando a estimativa de relevância como ordenação ideal.

#### 4.1 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
Baseline	0,916642	0,793905	0,802118	0,829422	0,830639
Experimento 1	<b>0,961521</b>	<b>0,883843</b>	<b>0,903808</b>	<b>0,898364</b>	<b>0,895763</b>
Experimento 2	<b>0,923042</b>	<b>0,883843</b>	<b>0,894902</b>	<b>0,877502</b>	<b>0,877540</b>
Experimento 3	<b>0,923042</b>	<b>0,869384</b>	<b>0,889841</b>	<b>0,865857</b>	<b>0,855534</b>
Experimento 4	0,916642	<b>0,837283</b>	<b>0,837120</b>	<b>0,846243</b>	<b>0,852330</b>
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	<b>0,923042</b>	<b>0,93160</b>	<b>0,952499</b>	<b>0,944679</b>	<b>0,942878</b>
LambdaMART 2	<b>0,991657</b>	<b>0,975892</b>	<b>0,969063</b>	<b>0,964859</b>	<b>0,962366</b>
LambdaMART 3	<b>0,961521</b>	<b>0,964185</b>	<b>0,970408</b>	<b>0,965552</b>	<b>0,957995</b>

Fonte: O autor (2023)

Para a categoria X, os experimentos 1, 2 e 3 atingiram resultados superiores em todas as métricas quando comparado ao baseline. O experimento 4, atingiu resultados superiores em 4 das métricas e resultado igual ao baseline para NDCG@10. No entanto, quando observado os resultados dos modelos de aprendizado de máquina LambdaMART, todas as métricas estão com resultados acima do baseline, inclusive 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
Baseline	0,662741	0,768131	0,798057	0,827410	0,794315
Experimento 1	0,628693	0,636021	0,656954	0,690318	0,702195
Experimento 2	<b>0,880078</b>	<b>0,886333</b>	<b>0,892696</b>	<b>0,888230</b>	<b>0,875746</b>
Experimento 3	<b>0,836115</b>	<b>0,847726</b>	<b>0,862035</b>	<b>0,877267</b>	<b>0,850655</b>
Experimento 4	<b>0,748188</b>	<b>0,821468</b>	<b>0,831546</b>	<b>0,836220</b>	<b>0,835394</b>
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	<b>0,994833</b>	<b>0,956889</b>	<b>0,951996</b>	<b>0,943865</b>	<b>0,941483</b>
LambdaMART 2	<b>0,992354</b>	<b>0,978940</b>	<b>0,969820</b>	<b>0,952902</b>	<b>0,956150</b>
LambdaMART 3	<b>0,992354</b>	<b>0,980966</b>	<b>0,954593</b>	<b>0,943823</b>	<b>0,948861</b>

Fonte: O autor (2023)

Para a categoria Y, os experimentos 2, 3 e 4 atingiram resultados superiores em todas as métricas quando comparado ao baseline. Quando observados os resultados dos modelos de aprendizado de máquina LambdaMART, todas as métricas estão com resultados acima do baseline e com 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
Baseline	0,730440	0,745971	0,780625	0,762924	0,764631
Experimento 1	<b>0,819593</b>	<b>0,843402</b>	<b>0,878753</b>	<b>0,861782</b>	<b>0,853793</b>
Experimento 2	<b>0,819593</b>	<b>0,829170</b>	<b>0,869004</b>	<b>0,839685</b>	<b>0,836337</b>
Experimento 3	<b>0,807675</b>	<b>0,809320</b>	<b>0,841100</b>	<b>0,822416</b>	<b>0,813709</b>
Experimento 4	<b>0,763099</b>	<b>0,775578</b>	<b>0,794572</b>	<b>0,789274</b>	<b>0,791081</b>
Experimento 5	<b>0,732426</b>	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	<b>0,947858</b>	<b>0,971151</b>	<b>0,980779</b>	<b>0,968771</b>	<b>0,970485</b>
LambdaMART 2	<b>0,964992</b>	<b>0,965872</b>	<b>0,975085</b>	<b>0,969573</b>	<b>0,966901</b>
LambdaMART 3	<b>0,972193</b>	<b>0,974869</b>	<b>0,978344</b>	<b>0,973596</b>	<b>0,970912</b>

Fonte: O autor (2023)

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

#### 4.2 Estimativa de relevância como taxa de cliques 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
Baseline	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	<b>0,831189</b>
Experimento 4	0,938777	0,881092	0,841942	<b>0,835412</b>	0,823461
Experimento 5	0,938777	<b>0,907390</b>	<b>0,846919</b>	0,823830	<b>0,836255</b>
Experimento 6	<b>0,944798</b>	<b>0,926584</b>	<b>0,849583</b>	<b>0,835015</b>	<b>0,845746</b>
Experimento 7	0,921901	<b>0,945777</b>	<b>0,856062</b>	0,823432	<b>0,840195</b>
Experimento 8	<b>0,944798</b>	<b>0,945295</b>	<b>0,880119</b>	0,827697	<b>0,834779</b>
Experimento 9	<b>0,949502</b>	<b>0,945295</b>	<b>0,880119</b>	0,828878	<b>0,837016</b>
Experimento 10	<b>0,948185</b>	<b>0,927336</b>	<b>0,880119</b>	<b>0,833890</b>	<b>0,833803</b>
LambdaMART 1	0,936144	<b>0,915776</b>	<b>0,879373</b>	<b>0,860601</b>	<b>0,862942</b>
LambdaMART 2	<b>0,995296</b>	<b>0,982571</b>	<b>0,909146</b>	<b>0,885721</b>	<b>0,890315</b>
LambdaMART 3	<b>0,977103</b>	<b>0,972075</b>	<b>0,956470</b>	<b>0,936399</b>	<b>0,934199</b>

Fonte: O autor (2023)

Para a categoria X, os experimentos 6 e 10 atingiram resultados superiores em todas as métricas quando comparado ao baseline. 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 baseline para NDCG@75. Quando observados os resultados dos modelos de aprendizado de máquina LambdaMART 2 e 3, todas as métricas estão com resultados acima do baseline a apenas o LambdaMART 1 não

superou o baseline para NDCG@10. Ainda assim, os melhores resultados atingidos fazem parte dos modelos 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
Baseline	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	<b>0,920448</b>	<b>0,873917</b>	<b>0,835492</b>	<b>0,831951</b>	0,815517
Experimento 6	<b>0,920448</b>	<b>0,904624</b>	<b>0,832065</b>	<b>0,834255</b>	<b>0,829330</b>
Experimento 7	<b>0,920448</b>	<b>0,907571</b>	<b>0,834158</b>	<b>0,843154</b>	<b>0,823066</b>
Experimento 8	<b>0,920448</b>	<b>0,907571</b>	<b>0,834538</b>	<b>0,849149</b>	<b>0,829905</b>
Experimento 9	<b>0,920448</b>	<b>0,900127</b>	<b>0,853824</b>	<b>0,853423</b>	<b>0,823642</b>
Experimento 10	<b>0,920448</b>	<b>0,894312</b>	<b>0,856660</b>	<b>0,855392</b>	<b>0,831478</b>
LambdaMART 1	<b>0,946006</b>	<b>0,900127</b>	<b>0,863817</b>	<b>0,867019</b>	<b>0,869072</b>
LambdaMART 2	<b>0,933227</b>	<b>0,909199</b>	<b>0,914023</b>	<b>0,918473</b>	<b>0,916275</b>
LambdaMART 3	<b>0,998132</b>	<b>0,975381</b>	<b>0,959297</b>	<b>0,949222</b>	<b>0,944428</b>

Fonte: O autor (2023)

Para a categoria Y, a partir do experimento 6 até o experimento 10, os resultados são superiores em todas as métricas quando comparado ao baseline. O experimento 5 não superou o baseline apenas para NDCG@100. Os experimentos dos modelos de aprendizado de máquina LambdaMART, superaram todas as métricas o baseline 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
Baseline	0,947916	0,859072	0,853583	0,852916	0,869296
Experimento 1	0,947916	<b>0,870475</b>	<b>0,857941</b>	<b>0,855679</b>	0,851703
Experimento 2	0,947916	<b>0,870475</b>	<b>0,857267</b>	<b>0,860706</b>	0,852742
Experimento 3	0,947916	0,858844	<b>0,857941</b>	<b>0,864605</b>	0,862807
Experimento 4	0,947916	0,846412	0,850987	<b>0,858533</b>	<b>0,869339</b>
Experimento 5	0,947916	<b>0,870400</b>	<b>0,855887</b>	<b>0,862887</b>	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	<b>0,978271</b>	<b>0,929263</b>	<b>0,905473</b>	<b>0,901133</b>	<b>0,900261</b>
LambdaMART 2	<b>0,975573</b>	<b>0,945217</b>	<b>0,926983</b>	<b>0,899738</b>	<b>0,902848</b>
LambdaMART 3	<b>0,969645</b>	<b>0,985381</b>	<b>0,955286</b>	<b>0,952825</b>	<b>0,950675</b>

Fonte: O autor (2023)

Para a categoria Z, nenhum experimento de 1 a 10 superou todos os valores das métricas do baseline, sendo que os melhores foram os experimentos 1, 2 e 5 que tiveram 3 métricas maiores que o baseline, enquanto os experimentos 3 e 4 tiveram apenas 2 métricas maiores. No entanto, novamente os experimentos dos modelos de aprendizado de máquina LambdaMART, superaram todas as métricas do baseline com os melhores resultados atingidos.

### 4.3 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
Baseline	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	<b>0,766304</b>	0,753713	<b>0,761441</b>
Experimento 3	0,937541	0,877166	<b>0,773051</b>	0,762506	<b>0,765895</b>
Experimento 4	0,963985	0,890570	<b>0,780675</b>	0,754609	<b>0,760571</b>
Experimento 5	0,963985	<b>0,918223</b>	<b>0,772173</b>	0,760810	0,757580
Experimento 6	<b>0,968771</b>	<b>0,918223</b>	<b>0,766607</b>	0,787990	<b>0,758980</b>
Experimento 7	<b>0,995215</b>	<b>0,937613</b>	0,763096	0,788411	0,756457
Experimento 8	<b>1,000000</b>	<b>0,933355</b>	0,754595	0,782211	0,757857
Experimento 9	<b>0,995215</b>	<b>0,933355</b>	0,753717	0,781573	0,755587
Experimento 10	<b>0,995215</b>	<b>0,919107</b>	0,752839	0,779240	0,751133
LambdaMART 1	<b>0,990430</b>	<b>0,966953</b>	<b>0,824655</b>	<b>0,830175</b>	<b>0,843912</b>
LambdaMART 2	<b>0,990430</b>	<b>0,930745</b>	<b>0,881697</b>	<b>0,861438</b>	<b>0,861658</b>
LambdaMART 3	0,906312	0,859503	<b>0,781428</b>	<b>0,792701</b>	<b>0,842355</b>

Fonte: O autor (2023)

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

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

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

Fonte: O autor (2023)

Para a categoria Y, os experimentos de 6 a 10 tiveram melhores resultados quando comparados ao baseline, superando 4 métricas. O experimento 5 superou o baseline em apenas 3 métricas, seguido dos experimentos 1 com apenas 2 métricas acima e os experimentos 2, 3 e 4 com apenas 1 métrica vencedora. Os modelos de aprendizado de máquina LambdaMART 1 e 2, tiveram todos resultados superiores ao baseline e o LambdaMART 3 só não superou o NDCG@10. Ainda assim, os maiores resultados foram atingidos pelos modelos inteligentes.

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

Categoria Z	NDCG@10	NDCG@25	NDCG@50	NDCG@75	NDCG@100
Baseline	0,942502	0,874800	0,808283	0,788636	0,792665
Experimento 1	<b>0,987487</b>	<b>0,931695</b>	<b>0,866094</b>	<b>0,802014</b>	<b>0,810976</b>
Experimento 2	<b>0,987487</b>	<b>0,931695</b>	<b>0,833828</b>	<b>0,807527</b>	<b>0,810976</b>
Experimento 3	0,942502	<b>0,895356</b>	<b>0,838013</b>	0,784740	<b>0,809335</b>
Experimento 4	0,942502	<b>0,895356</b>	0,808283	<b>0,793056</b>	0,788365
Experimento 5	0,938723	0,874800	0,792379	0,770411	<b>0,796408</b>
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	<b>0,886457</b>	<b>0,879909</b>	<b>0,827746</b>	<b>0,839462</b>
LambdaMART 2	0,913697	<b>0,886737</b>	<b>0,925321</b>	<b>0,919726</b>	<b>0,914397</b>
LambdaMART 3	<b>0,974974</b>	<b>0,942113</b>	<b>0,915392</b>	<b>0,874294</b>	<b>0,868269</b>

Fonte: O autor (2023)

Para a categoria Z, os experimentos 1 e 2 tiveram resultados melhores em todas as métricas quando comparados ao baseline. O experimento 3 superou o baseline em apenas 3 métricas, seguido do experimento 4 com 2 acima e o experimento 5 com apenas 1 métrica vencedora. Os resultados do modelo LambdaMART 3 venceu todas as métricas do baseline e os demais modelos inteligentes só não tiveram melhores resultados que o baseline em apenas 1 métrica. Ainda assim, o modelo 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 experimentos

Conforme a Tabela 12 abaixo, temos o consolidado de todos os experimentos do presente trabalho, o qual cada métrica teve a chance de superar o baseline até 9 vezes.

Tabela 12 - Resultados consolidados por experimentos

Consolidado	NDCG@10	NDCG@25	NDCG@50	NDCG@75	NDCG@100	> baseline
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 o modelo inteligente atingiu uma performance melhor que o baseline de 96% para LambdaMART 1, 98% para

LambdaMART 2 e 93% para LambdaMART 3 mostrando muito mais eficiência quando comparado aos experimentos sem inteligência. A maior performance do experimento alternando os pesos veio do experimento 2 que obteve 58% dos resultados melhores que o baseline.

## 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. Com base no referencial teórico que abordou o tema de *Learning to Rank*, foi construído um programa computacional a partir do método escolhido. O crescente avanço da inteligência artificial e desafios no ambiente corporativo somando-se a crescente pesquisas acadêmicas do tema proposto, justificou-se o estudo deste trabalho.

Nesse sentido, foi elaborada uma metodologia para a solução do problema real. A implementação abordou o passo a passo da seleção e construção do modelo LTR explicando sobre as estimativas de relevâncias definidas, o método inteligente escolhido, os tipos de experimentos para validação de resultados e os conjuntos de dados aplicados ao estudo de caso.

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

Como dito anteriormente, a proposta deste trabalho foi a elaboração de um modelo inteligente para o problema real de um sistema de busca no comércio eletrônico de uma indústria de bens para o lar, através de técnicas de aprendizado de máquina. Sendo assim, quando aplicada uma técnica de LTR, pode-se concluir que os resultados são bem superiores. Conforme a tabela de experimentos consolidado, o modelo inteligente nos 3 experimentos testados obteve uma performance maior que 90%, com destaque ao modelo 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, desbancando o mito de que os modelos de aprendizado de máquina só melhoram com o aumento de número de variáveis.

Portanto, espera-se com esse trabalho, contribuir principalmente para a aplicação prática no estudo de caso. Foi possível concluir que a aplicação de uma técnica de aprendizado de máquina 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 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 *feedback* de usuários como taxa de conversão de pedidos, taxa de cliques na busca e taxa de adição no carrinho e recomenda-se usar outros métodos de relevância para estudos futuros como, por exemplo, estimativas de relevância que trabalhar com 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; SONDHI, 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.