

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO DE INFORMÁTICA
ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS E SUAS APLICAÇÕES**

KARINA BAUMGARTEN

**SISTEMA DE RECOMENDAÇÃO PROGRESSIVA PARA
E-MAILS DE LEMBRETE DE CARRINHO ABANDONADO**

MONOGRAFIA

CURITIBA

2021

KARINA BAUMGARTEN ✉

**SISTEMA DE RECOMENDAÇÃO PROGRESSIVA PARA
E-MAILS DE LEMBRETE DE CARRINHO ABANDONADO**

**PROGRESSIVE RECOMMENDATION SYSTEM
FOR ABANDONED CART REMINDER EMAILS**

Monografia apresentada como requisito para obtenção do título de Especialista em Ciência de Dados e Suas Aplicações da Universidade Tecnológica Federal do Paraná (UTFPR).

Orientador: Prof. Dr. Luiz Celso Gomes Jr.

CURITIBA

2021



4.0 Internacional

Esta licença permite remixe, adaptação e criação a partir do trabalho, para fins não comerciais, desde que sejam atribuídos créditos ao(s) autor(es) e que licenciem as novas criações sob termos idênticos. Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.



Ministério da Educação
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
UTFPR - CAMPUS CURITIBA
DIRETORIA-GERAL - CAMPUS CURITIBA
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO - CAMPUS CURITIBA
DEPARTAMENTO DE APOIO DAS ESPECIALIZAÇÕES LATO-SENSU DOS CURSOS
DE INFORMÁTICA - CAMPUS CURITIBA



TERMO DE APROVAÇÃO

SISTEMA DE RECOMENDAÇÃO PROGRESSIVA PARA E-MAILS DE LEMBRETE DE CARRINHO ABANDONADO

por

Karina Baumgarten

Este Trabalho de Conclusão de Curso foi apresentado às 21h30min do dia 28 de julho de 2021 por videoconferência como requisito parcial à obtenção do grau de Especialista em Ciência de Dados e suas Aplicações na Universidade Tecnológica Federal do Paraná - UTFPR - Campus Curitiba. A aluna foi arguida pela Banca de Avaliação abaixo assinados. Após deliberação, a Banca de Avaliação considerou o trabalho aprovado.

Prof. Dr. Luiz Celso Gomes Junior (Presidente/Orientador – DAINF-CT/ UTFPR-CT)

Profa. Dra. Rafaela Moos(Avaliadora 1– SERASA EXPERIAN - DATALAB)

Profa. Dra. Rita Cristina Galarraga Berardi (Avaliadora 2 – DAINF-CT/ UTFPR-CT)

O Termo de Aprovação assinado encontra-se no sistema SEI- Processo nº 23064.028838/2021-51

Referência: Processo nº 23064.028838/2021-51

SEI nº 2189770

RESUMO

BAUMGARTEN, Karina. **Sistema de Recomendação Progressiva para E-mails de Lembrete de Carrinho Abandonado**. 2021. 34 f. Monografia (Especialização em Ciência de Dados e Suas Aplicações) — Universidade Tecnológica Federal do Paraná, Curitiba, 2021.

O abandono de produtos no carrinho de compras é um problema que sites de e-commerce enfrentam ao redor do mundo. Para tentar reverter essa situação e fazer com que seus consumidores finalizem a compra, empresas têm feito uso de sistemas de recomendação para indicar produtos adequados ao perfil dos clientes. Esta monografia apresenta um sistema de recomendação colaborativo para indicação de produtos em e-mails de lembrete de carrinho abandonado em um e-commerce do setor de Beleza. A estratégia central do sistema é utilizar o histórico de compras de produtos para identificar similaridades entre eles. A partir de padrões identificados na análise exploratória, reconhecemos também o papel do valor do produto na recomendação. Portanto além do critério similaridade, o sistema ordena as indicações pelo preço de venda em ordem decrescente, fazendo com que novas recomendações sejam mais baratas do que as anteriores.

Palavras-chave: Sistema de recomendação. Carrinho abandonado. E-mail de lembrete.

ABSTRACT

BAUMGARTEN, Karina. **Progressive Recommendation System for Abandoned Cart Reminder Emails** . 2021. 34 p. Monography (Specialization Degree in Data Science and Its Applications) — Federal University of Technology — Paraná, Curitiba, 2021.

The abandonment of products in the shopping cart is a problem that e-commerce sites face around the world. To try to reverse this situation and make their consumers complete the purchase, companies have been using recommendation systems to indicate products that are suitable for their customers' profile. This monograph presents a collaborative recommendation system for product indication in abandoned cart reminder emails in a Beauty sector e-commerce. The system's central strategy is to use the purchase history of products to identify similarities between them. Based on patterns identified in the exploratory analysis, we also recognized the role of product value in the recommendation. Therefore, in addition to the similarity criterion, the system orders the indications by the sale price in descending order, making new recommendations cheaper than the previous ones.

Keywords: Recommendation system. Cart abandonment. Reminder emails.

LISTA DE ALGORITMOS

Algoritmo 1 – Algoritmo para gerar conjunto de itens recomendados	25
---	----

LISTA DE ILUSTRAÇÕES

Figura 1 – Arquitetura do sistema proposto	22
Figura 2 – Modelo do e-mail de lembrete	23
Gráfico 1 – Ocorrência de abandono e compra por dia do mês	16
Gráfico 2 – Ocorrência de abandono e compra por dia do mês após remoção dos dias 11 e 12	17
Gráfico 3 – Ocorrência de abandono e compra por departamento	17
Gráfico 4 – Ocorrência de abandono e compra por categoria	18
Gráfico 5 – Ocorrência de abandono e compra por dia da semana	19
Gráfico 6 – Distribuição dos preços de abandono de carrinho	19
Gráfico 7 – Distribuição dos preços de abandono de carrinho por departamento	20
Gráfico 8 – Distribuição dos preços de abandono de carrinho por categoria . .	21

LISTA DE TABELAS

Tabela 1 – Modelo de tabela de resultado de itens similares	24
Tabela 2 – Modelo de tabela de apresentação de resultados	30

SUMÁRIO

1	INTRODUÇÃO	9
2	FUNDAMENTOS E TRABALHOS RELACIONADOS	11
2.1	SISTEMAS DE RECOMENDAÇÃO	11
2.1.1	Sistemas de Recomendação Baseados em Conteúdo	11
2.1.2	Sistemas de Recomendação Colaborativos	12
2.1.3	Sistemas de Recomendação Híbridos	12
2.2	CASOS DE UTILIZAÇÃO DE SISTEMAS DE RECOMENDAÇÃO . .	13
3	IMPLEMENTAÇÃO	15
3.1	ANÁLISE EXPLORATÓRIA	15
3.2	ARQUITETURA DO SISTEMA	21
3.3	MODELO PROPOSTO	23
3.3.1	Testes de desempenho	25
3.4	<i>COLDSTART</i>	27
4	PLANEJAMENTO DOS TESTES	28
4.1	TESTE A/B	28
4.2	ANÁLISE DOS RESULTADOS E DISCUSSÃO	29
5	CONSIDERAÇÕES FINAIS	32
	REFERÊNCIAS	33

1 INTRODUÇÃO

Um dos problemas encontrados nos sites de e-commerce são os carrinhos abandonados, ou seja, quando um produto é adicionado ao carrinho, porém a compra não é finalizada. Segundo pesquisa disponibilizada pela (NEOATLAS, 2018), no primeiro semestre de 2018, a média geral de carrinhos abandonados em plataformas de compras online no Brasil foi de 82,3% no período analisado. Para o setor de Beleza esse número foi de 85%. Isso significa que 85% das pessoas que iniciam uma compra não finalizam o pedido.

Mas esse problema não é exclusividade de e-commerces brasileiros. Em pesquisa realizada pela Barilliance (SERRANO, 2021), em 2019 a média global de carrinhos abandonados foi de 77,13%. Especificamente durante a Black Friday e Cyber Monday naquele ano, as taxas de abandono ficaram em 73,78% e 68,94%, respectivamente.

O que leva os usuários a não finalizarem suas compras pode ter inúmeros motivos, como preocupações com a entrega do produto, com a qualidade e desempenho do produto, com a segurança dos dados pessoais e até mesmo com o objetivo de coletar mais informações sobre os itens de interesse. Apesar disso, é importante que as lojas online busquem alternativas para melhorar os processos de modo a atender as necessidades de seus consumidores (MACIEL, 2020).

Uma estratégia utilizada por sites de e-commerce para fazer com que seus conteúdos sejam cada vez mais personalizados de acordo com a preferência dos clientes é o uso de sistemas de recomendação. Eles fazem uso de dados de histórico dos usuários, como atividades dentro do site, compras realizadas e abandonos de carrinhos. A partir disso, é possível determinar quais produtos são mais adequados ao perfil dos clientes e oferecer uma experiência mais individualizada de acordo com os gostos dos usuários (DAOUD; NAQVI; AHMAD, 2014).

Outra estratégia adotada por alguns sites, é o uso de e-mails de lembrete de compra não finalizada. O principal objetivo é lembrar os consumidores dos produtos que chamaram sua atenção e manter o relacionamento com eles para compras futuras (MARQUES; ROLDAN, 2017). Esses e-mails buscam conter personalizações para atrair os clientes para dentro do e-commerce e, eventualmente, a finalização da compra. Também é possível fazer uso de sistemas de recomendação dentro dos e-mails de lembrete. Essa técnica é utilizada por empresas como a Amazon, por exemplo (BORGES;

OLIVEIRA, 2010).

Segundo pesquisa da Barilliance (SERRANO, 2021), a taxa de abertura desses e-mails de lembretes de carrinhos abandonados gira em torno de 40% desde 2016, quando o estudo foi iniciado. Desses e-mails, cerca de 10% acabam gerando compras no e-commerce.

O estudo elaborado e descrito a seguir teve como objetivo desenvolver um sistema de recomendação de produtos para e-mails de lembrete de compra não finalizada. O objetivo principal dessas recomendações de produtos é aumentar as taxas de cliques e compras geradas pelos e-mails de lembrete. O modelo de recomendação usado foi o colaborativo, no qual o principal diferencial da proposta é utilizar o valor do produto na indicação das recomendações, fazendo com que os produtos recomendados fiquem gradativamente mais baratos do que os anteriores, devido à percepção de que o valor de carrinhos abandonados costuma ser superior ao valor das compras finalizadas.

Os dados obtidos foram fornecidos por um e-commerce nacional no setor de Beleza e contém registros de compras e carrinhos abandonados do mês de Fevereiro de 2021.

O restante da monografia está organizada da seguinte forma: o Capítulo 2 apresenta o referencial teórico sobre sistemas de recomendação e trabalhos relacionados; no Capítulo 3 estão os detalhes da implementação do sistema proposto, desde a análise exploratória, organização da arquitetura e testes de desempenho; o Capítulo 4 contém o planejamento de execução dos testes, com análises e discussões que devem ser feitas sobre os resultados obtidos; o Capítulo 5 possui as considerações finais.

2 FUNDAMENTOS E TRABALHOS RELACIONADOS

2.1 SISTEMAS DE RECOMENDAÇÃO

Os sistemas de recomendação são utilizados para oferecer ao usuário um conteúdo que seja de seu interesse. Para isso, os sistemas buscam prever a preferência do usuário sobre os itens, com o objetivo de obter margens de venda maiores e aumentar a satisfação dos usuários (MOTTA et al., 2011).

As principais funções dos sistemas de recomendação são analisar os dados dos usuários e extrair informações úteis para futuras recomendações. Normalmente são buscadas opiniões ou preferências sobre itens em um grupo de usuários e essas informações são utilizadas para apresentar outros itens que possam fazer sentido aos usuários em questão. Dessa forma, sistemas de recomendação utilizam, basicamente, informações sobre as preferências dos usuários e um método para determinar se um item é interessante para um usuário (SOUZA, 2012).

A forma de determinar se um item é interessante para um usuário depende do tipo do sistema de recomendação utilizado. Os tipos mais comuns são: baseado em conteúdo, colaborativo e sistemas híbridos.

2.1.1 Sistemas de Recomendação Baseados em Conteúdo

Os sistemas de recomendação baseados em conteúdo sugerem itens aos usuários que sejam similares aos itens já avaliados anteriormente. Os algoritmos aprendem os perfis de interesse dos usuários com base nas características dos itens bem avaliados. Dessa forma, constroem os perfis dos usuários a partir do perfil dos itens (SOUZA, 2012).

Algumas das vantagens do uso desse tipo de recomendação são: não precisar utilizar dados de outros usuários, ser capaz de recomendar itens novos e não populares, poder fornecer explicações sobre os itens recomendados listando as características do conteúdo que levaram à recomendação (MOTTA et al., 2011).

As principais desvantagens desse modelo seriam: recomendar ao usuário somente itens semelhantes aos já consumidos, podendo levar ao desinteresse nas recomendações por gerar pouca diversidade no conteúdo; e a definição de similaridade

de conteúdo, por cada contexto possuir suas próprias particularidades em relação às características. Nesse segundo caso, se a função de similaridade não estiver bem definida, poderá resultar em recomendações homogêneas, repetitivas ou, até mesmo, nada semelhantes (SANTANA, 2018).

2.1.2 Sistemas de Recomendação Colaborativos

Os sistemas de recomendação colaborativos utilizam o histórico de interações entre itens e usuários para encontrar relacionamentos entre eles. Nesse modelo, são utilizadas correlações entre as avaliações feitas pelo usuário a receber uma recomendação e as avaliações fornecidas por outros usuários. Dessa forma, para sugerir um produto para um usuário, utiliza-se os dados de outros usuários que já tenham avaliado itens de maneira semelhante à dele (MOTTA et al., 2011).

É comum que os perfis dos usuários sejam armazenados como vetores de itens e suas avaliações para eles, sendo que esses vetores tendem a aumentar de tamanho à medida que os usuários interajam com o sistema. Essas avaliações podem ser binárias (gostei/não gostei) ou com valores de acordo com o nível de preferência (SOUZA, 2012).

Esse método é eficiente pelo poder discriminatório, considerando que as recomendações acabam sendo mais relevantes e com maior qualidade conforme o número de avaliações aumenta. Apesar disso, é um método computacionalmente caro devido ao uso da matriz de similaridade, que possui um alto tempo para ser calculada e armazenada para possibilitar respostas rápidas. Outra desvantagem desse método é o fato de requerer que os itens sejam avaliados para entrar no processo, assim sendo, lançamentos acabam não aparecendo nas recomendações (MOTTA et al., 2011).

2.1.3 Sistemas de Recomendação Híbridos

Os sistemas de recomendação híbridos utilizam uma técnica que procura combinar as vantagens e atenuar desvantagem das abordagens colaborativa e baseada em conteúdo. De uma forma geral, os resultados obtidos por esse método são melhores do que os dois separadamente (SANTANA, 2018).

Os melhores resultados podem ser explicados pelos benefícios apresentados

da junção dos dois métodos: bons resultados para usuários incomuns, precisão independentemente do número de usuários, descobertas de novos relacionamentos entre usuários e recomendação de itens diretamente relacionado ao histórico (SANTANA, 2018).

Existem diferentes estratégias de filtragem híbrida, porém a mais simples é baseada na unificação das recomendações geradas por sistemas separados em uma única lista. Já outras abordagens mais complexas podem adicionar também as informações do conteúdo na filtragem colaborativa, por exemplo (SANTANA, 2018).

Na proposta desta monografia, especificamos um sistema de recomendação colaborativo baseado em itens, fundamentado no comportamento de compra dos usuários. No sistema proposto, são encontrados produtos semelhantes a partir das compras realizadas anteriormente. Uma das vantagens dessa escolha foi não precisar analisar a similaridade entre os usuários, mas sim entre os produtos, fazendo com que o algoritmo consuma menos recursos quando comparado com a outra opção, e tenha resultados satisfatórios para o problema descrito, em que a quantidade de usuários é superior a quantidade de produtos.

2.2 CASOS DE UTILIZAÇÃO DE SISTEMAS DE RECOMENDAÇÃO

Sistemas de recomendação são muito utilizadas por empresas de mídia. O YouTube utiliza sistemas híbridos para recomendações formulado como rede neural. O sistema de recomendação é composto por duas etapas: geração de candidatos e ranking. O primeiro utiliza o histórico de visualização dos usuários para identificar usuários com atividades semelhantes. Já o segundo, atribui uma pontuação a cada vídeo usando um conjunto de recursos que descrevem o vídeo e o usuário. Os vídeos com maiores pontuações são apresentados ao usuário (COVINGTON; ADAMS; SARGIN, 2016).

Assim como o YouTube, a Netflix também utiliza sistema de recomendação híbrido em duas etapas, porém através de metatags e do comportamento do usuário. Todo o catálogo da Netflix é organizado utilizando metatags, contendo características de cada título e totalizando em mais de 100 campos. Além disso, o sistema de recomendação mapeia as avaliações e comportamento dos usuários. Com essas informações, são geradas as recomendações personalizadas para cada usuário da plataforma (SIGILIANO;

FAUSTINO, 2016).

Além de serviços de streaming de filmes e vídeos, comércios eletrônicos, como a Amazon, também costumam utilizar sistemas híbridos para suas recomendações. Além de possuir uma lista de produtos parecidos com o visualizado, existe também a sessão de produtos que foram comprados por outros usuários. Essas recomendações são feitas baseadas nas classificações realizadas por outros usuários dentro de um grupo restrito de conteúdo, ordenadas pela soma de relevância das classificações (ROCHA FERNANDES; LINHARES, 2012).

Apesar de existirem as recomendações dentro da própria plataforma, a Amazon também possui métodos para encaminhar recomendações via e-mail aos consumidores. Esses e-mails podem conter recomendações de produtos novos adicionados ao catálogo, produtos semelhantes aos visualizados anteriormente e, também, caso algum produto que estava indisponível tenha voltado ao estoque (BORGES; OLIVEIRA, 2010).

Semelhante aos citados anteriormente, o sistema de recomendação proposto utiliza o comportamento do usuário para fazer recomendações, porém não utiliza notas atribuídas a vídeos e produtos, mas sim se o produto foi comprado ou não. Além disso, o sistema proposto não utiliza dados específicos dos produtos, como é feito pela Netflix, por exemplo, que faz uso das características dos itens de seu catálogo para identificar similaridades. No sistema proposto, os produtos recomendados são determinados de acordo com o comportamento de compra do usuário, sendo semelhante ao utilizado pela Amazon, que possui uma lista de produtos parecidos em seu site. Além disso, não usamos o histórico de compras do usuário alvo da recomendação, uma vez que o foco é nos produtos abandonados. Apesar das semelhanças, o sistema proposto possui como diferencial a preocupação com o preço do produto, buscando recomendações que fiquem mais baratas conforme os itens são indicados ao consumidor.

3 IMPLEMENTAÇÃO

Para o estudo apresentado foram selecionados dados de carrinhos abandonados e compras realizadas em um site de e-commerce da categoria de Beleza. Os dados obtidos são referentes às ocorrências no mês de Fevereiro do ano de 2021 e totalizam 642.433 registros. Nos dados obtidos existem características dos itens envolvidos, como departamento, categoria e preço, além da data de ocorrência. Para evitar problemas posteriores, registros com algum dado faltante foram removidos da base.

O objetivo do estudo é apresentar um modelo de recomendação de produtos, na tentativa de aumentar a taxa de compras realizadas a partir dos e-mails de lembrete de abandonos de carrinhos no site.

Atualmente, o e-commerce envia e-mails de lembrete contendo o produto abandonado pelo cliente no carrinho de compras e um conjunto de 4 produtos como recomendação. Essa recomendação é feita utilizando os produtos mais vendidos nos últimos 30 dias, ou seja, sem personalização com o que o cliente visitou. Dessa forma, o uso do sistema de recomendação proposto seria a substituição dessa vitrine por uma vitrine personalizada, contendo indicações de produtos que tenham uma maior relação com o produto pelo qual o cliente se interessou.

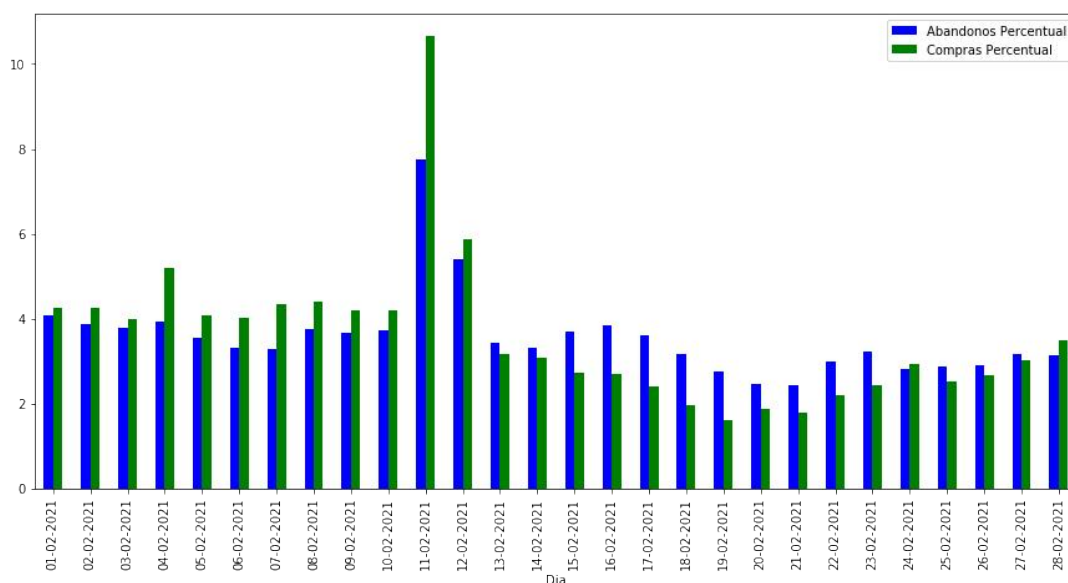
Para o desenvolvimento desse sistema de recomendação, a primeira etapa foi a elaboração da análise exploratória dos dados obtidos.

3.1 ANÁLISE EXPLORATÓRIA

Entre os dados obtidos, 392.999 são de carrinhos abandonados, representando 61,17% dos dados, e 249.434 são de compras realizadas no período. Além disso, 30% das compras foram realizadas por usuários que abandonaram algum item no carrinho.

Analisando as compras e abandonos realizados durante o mês de fevereiro, é possível verificar os picos de ambos nos dias 11 e 12 de fevereiro no gráfico 1. Nessa data específica ocorreu uma promoção relâmpago com descontos em diversos produtos. A ação ocorreu das 20h do dia 11 até as 8h do dia 12.

Devido ao comportamento atípico nos dias 11 e 12, provavelmente impulsionado pelos descontos nos produtos, foram removidos os dados relativos a essas datas. Com a remoção dos dias especificados, a quantidade de dados restante foi de 549.448, sendo

Gráfico 1 – Ocorrência de abandono e compra por dia do mês

Fonte: autoria própria (2021).

341.202 carrinhos abandonados e 208.246 compras. Ou seja, houve uma redução de 14,47% na quantidade total de dados, sendo que houve 13,18% de redução de carrinhos e 16,51% de compras.

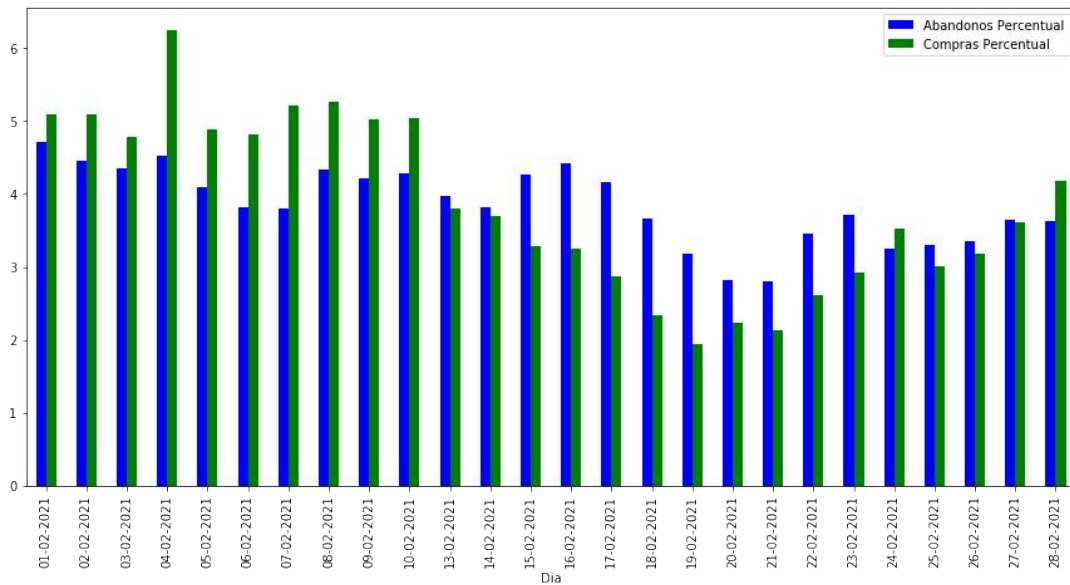
Para a análise dos dias do mês, sem considerar os dias 11 e 12, temos os maiores percentuais de compras realizadas nos dias 4 e 8, e as menores nos dias 19 e 21. Já sobre carrinhos abandonados, o maior percentual ocorreu no dia 1 e o menor no dia 21. Além disso, entre os dias 15 e 23 a porcentagem de carrinhos abandonados foi superior à porcentagem de compras realizadas. Esse comportamento pode ser visualizado no gráfico 2.

Na amostra obtida, os produtos estão divididos em 5 departamentos e 32 categorias, sendo que cada categoria pertence a um departamento. Os departamentos existentes são: Perfumaria; Corpo e Banho; Maquiagem; Cuidados para Pele; Cabelos.

Comparando os abandonos e compras para o mês e os departamentos existentes, é possível observar que os 2 departamentos que mais tiveram ocorrências foram Perfumaria e Corpo e Banho. O gráfico 3 abaixo apresenta a porcentagem de ocorrência de abandono e compra, separado por departamento. Dessa forma, é possível identificar com maior facilidade as semelhanças e diferenças entre eles.

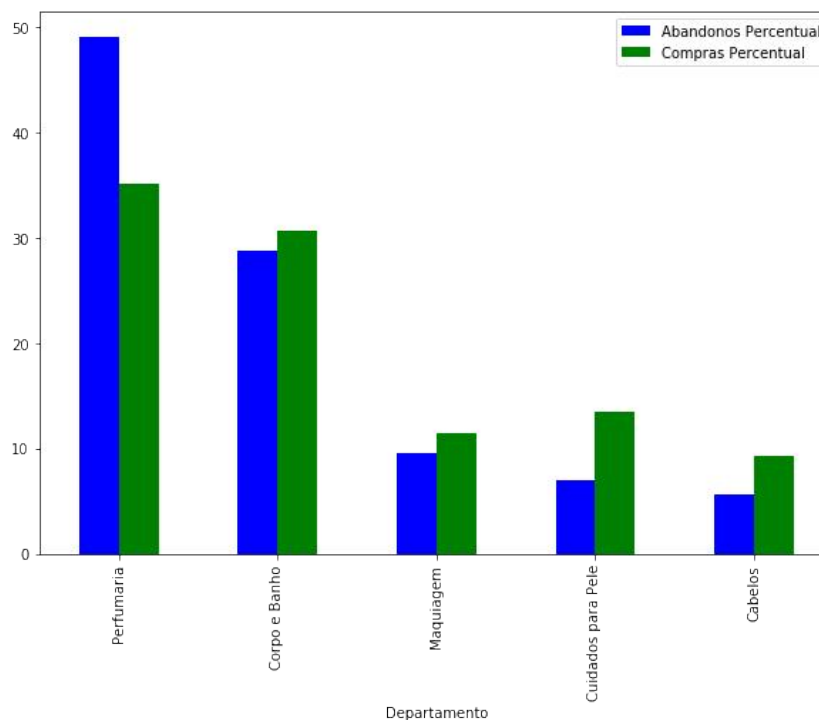
No gráfico 4 é apresentada a distribuição percentual entre as categorias, com remoção das categorias com menos de 1% de ocorrência. A categoria de Feminino, pertencente ao departamento de Perfumaria, é reconhecida tanto como a mais abandonada

Gráfico 2 – Ocorrência de abandono e compra por dia do mês após remoção dos dias 11 e 12



Fonte: autoria própria (2021).

Gráfico 3 – Ocorrência de abandono e compra por departamento

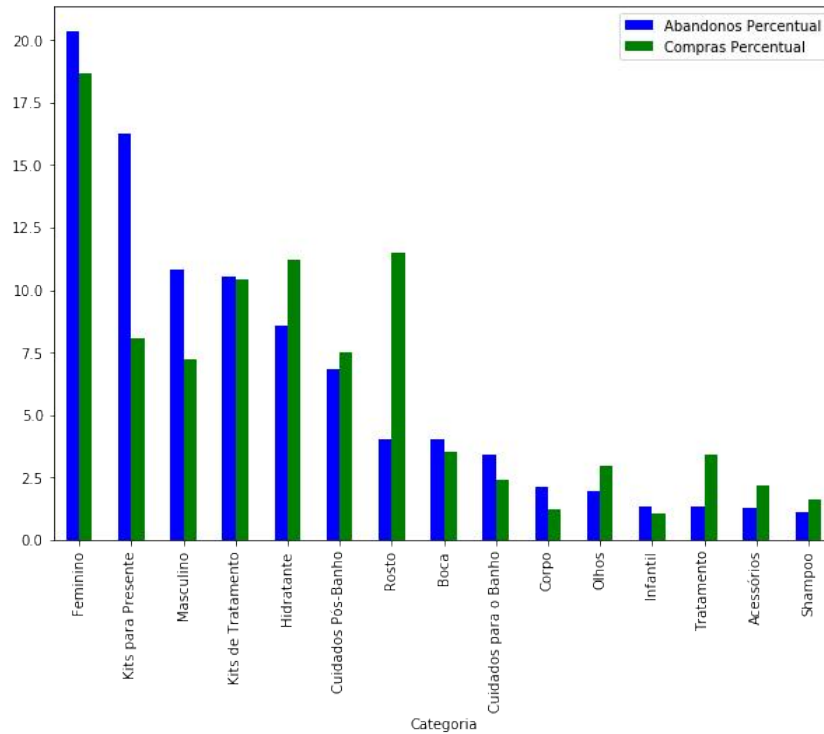


Fonte: autoria própria (2021).

(20%) quanto como a mais comprada (18,66%).

Entre os compradores, as categorias de Rosto (11,47%) e Hidratante (11,22%) são as próximas em percentual de ocorrências, enquanto para os carrinhos abandonados essas mesmas posições são ocupadas por Kits para Presentes (16%) e Masculino (10,84%), respectivamente.

Gráfico 4 – Ocorrência de abandono e compra por categoria



Fonte: autoria própria (2021).

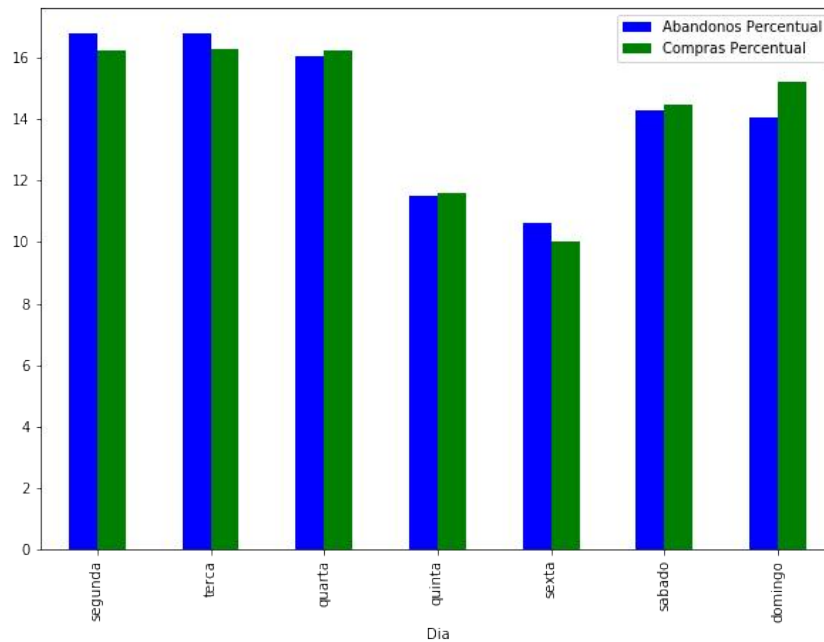
Adicionando a comparação de dias da semana das ocorrências, para abandonos, os maiores números de casos se deram às segundas e terças-feiras (ambos com 16,7%) e o menor às sextas-feiras (10,6%). Seguindo o comportamento observado em abandonos, os dias com maiores ocorrências de compras foram segundas, terças e quartas-feiras (16,9%) e com menor, sextas-feiras (10%). Para melhor visualização, os dados podem ser vistos no gráfico 5.

Outra análise importante de ser feita é em relação ao preço dos itens. No gráfico 6 são apresentados os histogramas de distribuição de preços de compras e abandonos de carrinhos. Em ambos a maior ocorrência dos casos são de itens até 100 reais.

Além do histograma, temos também o valor médio dos produtos comprados ou abandonados separados por departamento (gráfico 7) e categoria (gráfico 8). O gráfico de categorias apresenta apenas as 20 categorias com maiores médias de valores dos itens.

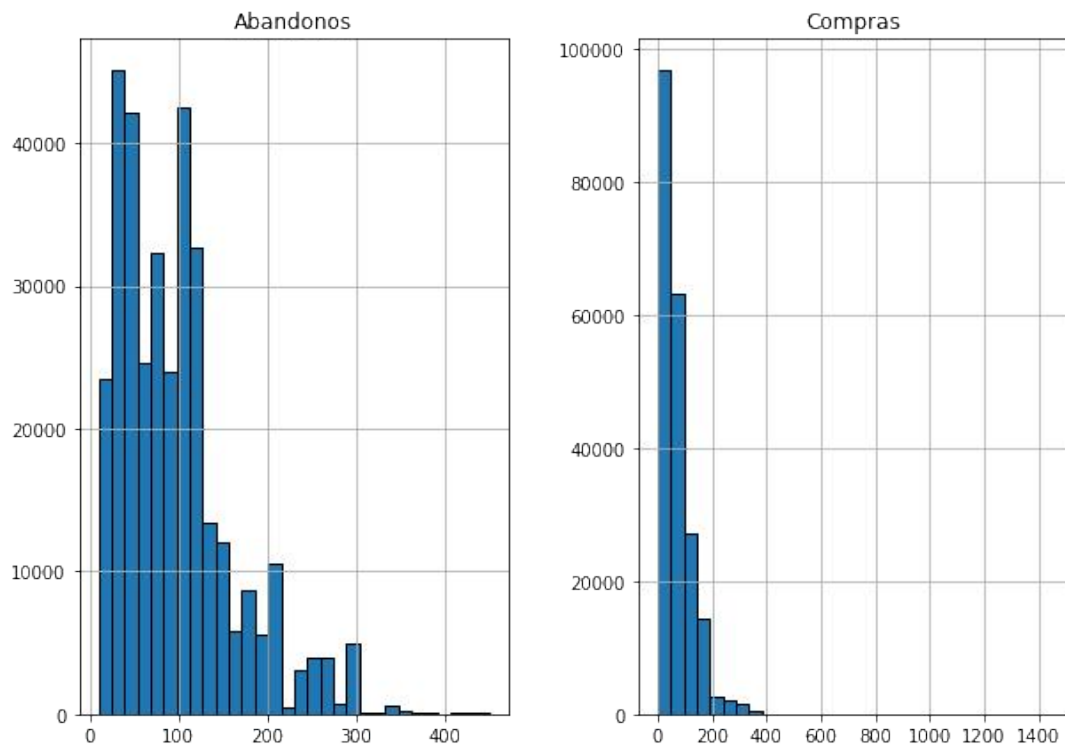
Como visto anteriormente, o departamento de Perfumaria é o que possui maior ocorrência tanto em compras quanto em abandonos. Além disso, também possui a média de preços mais alta. A média do valor dos itens abandonados do departamento de Perfumaria é de R\$137,60, enquanto a média em compras é de R\$106,36.

Gráfico 5 – Ocorrência de abandono e compra por dia da semana



Fonte: autoria própria (2021).

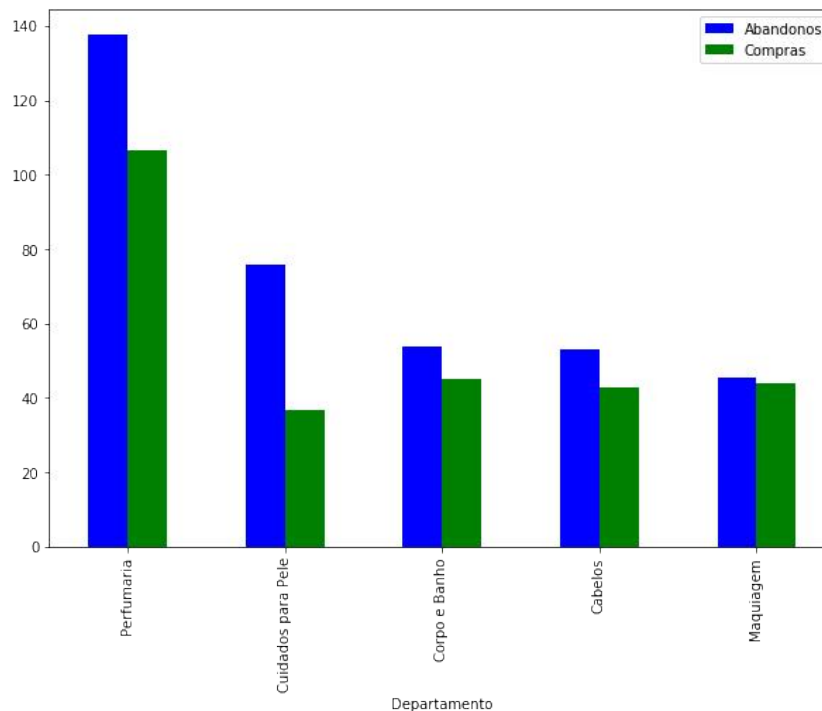
Gráfico 6 – Distribuição dos preços de abandono de carrinho



Fonte: autoria própria (2021).

Por outro lado, a menor média de valores está no departamento de Maquiagem. Para itens abandonados no carrinho, o valor médio dos produtos é de R\$45,52 e para os itens comprados desse departamento, o valor é de R\$43,69.

Já em categorias, é possível notar que Kits para Presente possui a maior média

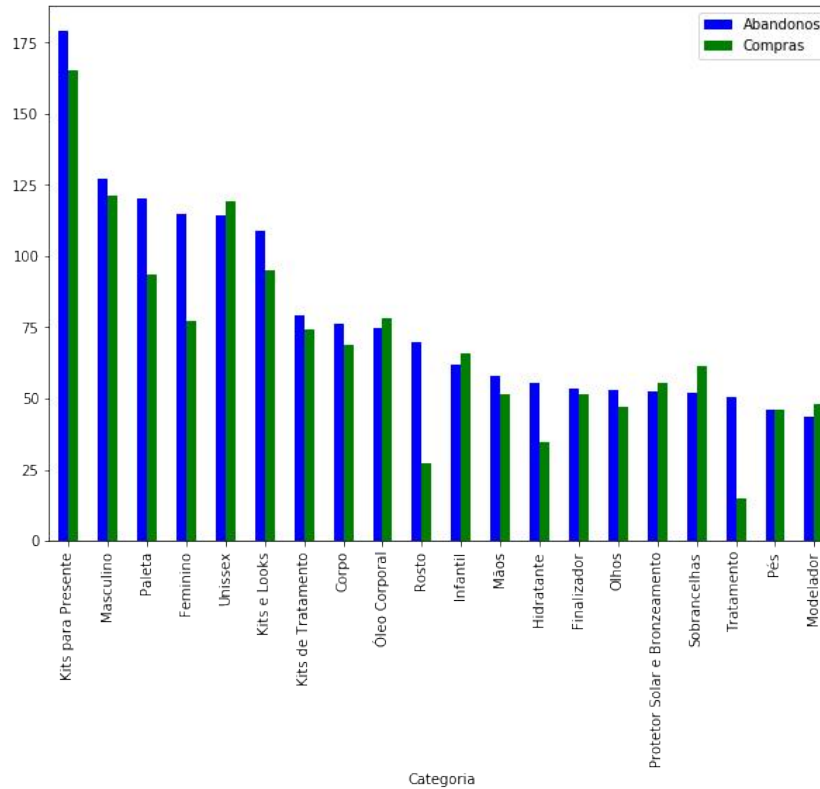
Gráfico 7 – Distribuição dos preços de abandono de carrinho por departamento

Fonte: autoria própria (2021).

de valor, sendo R\$179,34 para abandonos e R\$165,41 para compras. A maior diferença de médias de preço em categorias está em Rosto (R\$42,45), Feminino (R\$37,83), Tratamento (R\$35,89) e Paleta (R\$26,55). No gráfico 8 essas são as categorias que apresentam maiores diferenças visíveis. Além disso, é possível notar que, das 20 categorias apresentadas, somente 6 apresentam médias de preços de compras acima das médias de preços de abandonos.

Após finalizada a análise exploratória dos dados, foi definido que seriam utilizados para a criação do sistema de recomendação somente os dados de abandono e compras de itens do departamento de Perfumaria, por ser o departamento com maior quantidade de ocorrências e por ter médias de valores maiores. Além de ter produtos mais caros, existe uma diferença significativa entre a média de preço de produtos abandonados e vendidos, isso leva a acreditar que um dos motivos pelo abandono dos produtos possa ser o preço do item.

Por acreditar que essa premissa possa ser verdadeira, o sistema de recomendação proposto também leva em consideração o preço dos produtos, buscando indicar itens que sejam mais baratos. Com isso, espera-se que o consumidor que abandonou devido ao preço do produto, finalize a compra com um item semelhante mas de preço inferior.

Gráfico 8 – Distribuição dos preços de abandono de carrinho por categoria

Fonte: autoria própria (2021).

3.2 ARQUITETURA DO SISTEMA

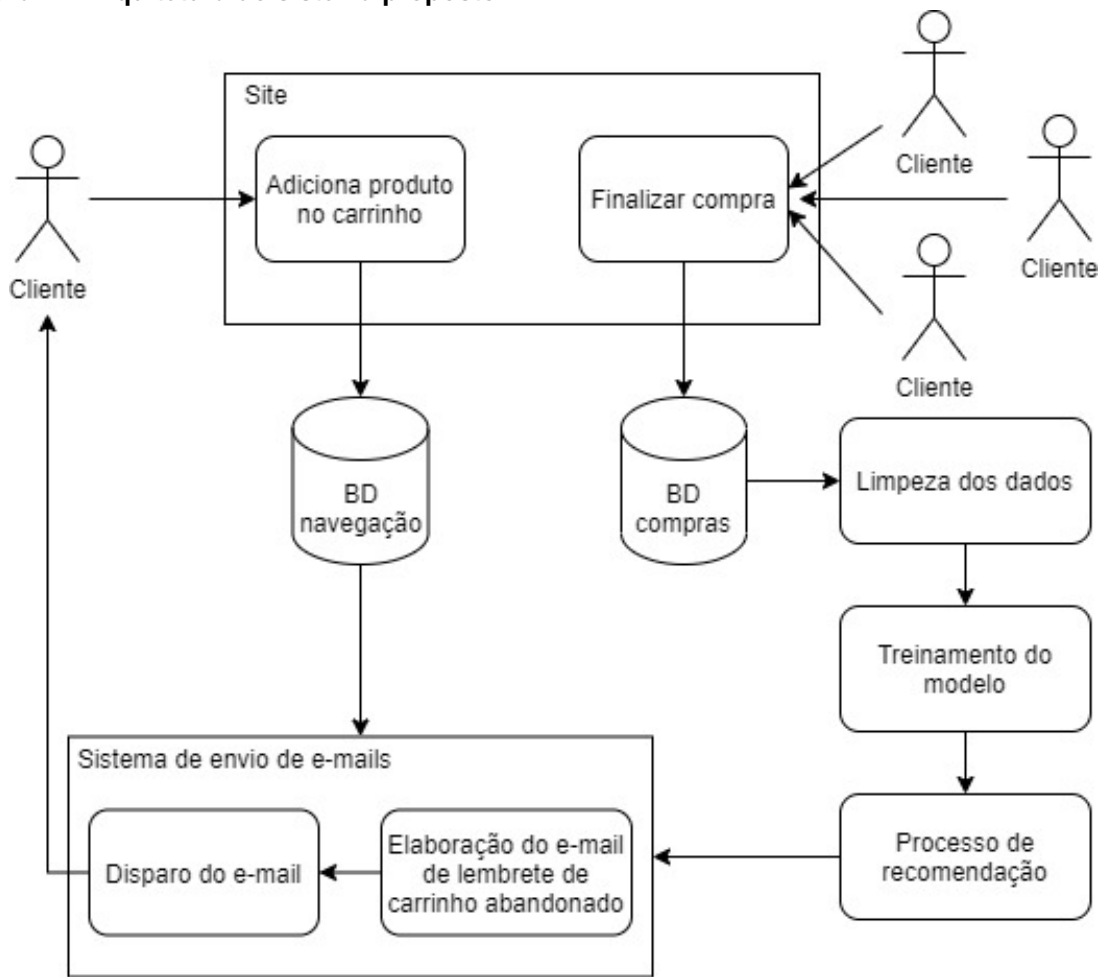
A arquitetura do sistema proposto pode ser visualizada na figura 1 abaixo. Ela se dá a partir de ações dos clientes como adicionar um produto no carrinho ou finalizar a compra. Cada uma das ações é armazenada em bases de dados separados, uma contendo as navegações realizadas pelos clientes e outra contendo todas as compras finalizadas.

Atualmente, assim que um cliente adiciona um produto no carrinho, mas não finaliza a compra, ele recebe e-mails de lembrete do produto adicionado e uma recomendação de quatro produtos mais vendidos do mesmo departamento. Com o sistema de recomendação proposto, as recomendações que são encaminhadas ao cliente junto com o produto visualizado, são determinadas através do modelo, que busca encontrar itens similares ao abandonado.

Na figura 2 é possível visualizar um modelo do e-mail de lembrete, contendo a posição do item visualizado e das quatro recomendações na vitrine.

Para chegar até as recomendações, alguns passos precisam ser executados.

Figura 1 – Arquitetura do sistema proposto



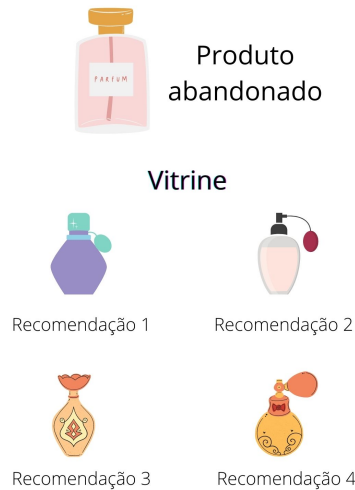
Fonte: autoria própria (2021).

Primeiramente foi realizada uma limpeza dos dados contidos na base de dados de compras finalizadas, para utilizá-los como entrada de treinamento do modelo de recomendação. Foram removidos os dados que não seriam utilizados posteriormente, ficando somente com dados de ID do cliente e item comprado. Em caso de dados faltantes, os registros foram removidos para evitar problemas com o treinamento do modelo. Além disso, devido à quantidade de produtos existentes, foram filtradas somente compras de itens do departamento de Perfumaria.

Após o treinamento do modelo (detalhes na Seção 3.3), o processo de recomendação é iniciado, gerando uma tabela com as 16 recomendações com maior grau de similaridade para cada produto do departamento de Perfumaria disponível no site. O número de 16 recomendações foi determinado pois o e-commerce envia quatro e-mails de recomendação para cada carrinho abandonado. Como cada e-mail contém quatro produtos recomendados, para que não haja repetição de recomendação, 16

Figura 2 – Modelo do e-mail de lembrete

BANNER DO SITE



Fonte: autoria própria (2021).

recomendações são necessárias para cada produto de Perfumaria disponível.

Essa tabela de recomendações é ordenada de forma decrescente pelo valor do produto, para que cada recomendação seja mais barata do que a anterior, devido ao comportamento identificado na análise exploratória, em que os abandonos tendem a ter preços maiores do que as compras.

A tabela final é disponibilizada ao sistema de envio de e-mails, assim como a base de dados de navegação dos clientes. Dentro do sistema de envio de e-mails, para todos os usuários que abandonaram um produto no carrinho sem a finalização da compra, são elaborados e enviados e-mails de lembrete, contendo o produto visualizado e um conjunto de 4 recomendações de produtos que possam ser de interesse do cliente.

3.3 MODELO PROPOSTO

O modelo de recomendação proposto visa encontrar produtos similares dentro do catálogo de produtos existente no site, utilizando a filtragem colaborativa. Para isso, utiliza como base as compras realizadas por outros clientes.

Após o procedimento de limpeza de dados realizado na base de dados de compras finalizadas, foi construída uma matriz M_{ij} onde cada linha i representa um ID

de cliente e cada coluna j , um produto do departamento de Perfumaria do site. A matriz contém o valor 1 na posição M_{ij} quando o cliente da linha i já comprou o produto da coluna j e 0 em caso contrário.

Com a matriz binária completamente preenchida, foram aplicadas técnicas para identificar a maior similaridade entre os produtos. As técnicas utilizadas foram correlação de Pearson, distância euclidiana e distância de Jaccard. O principal objetivo de utilizar 3 técnicas diferentes foi identificar qual traria um resultado mais satisfatório para a empresa.

A distância euclidiana calcula a distância entre dois pontos, sendo que suas posições são determinadas em relação às suas coordenadas em eixos cartesianos. A correlação de Pearson avalia se existe relação linear entre as variáveis. Jaccard é uma taxa entre o número de coincidências e o número total de itens, sem considerar as ausências conjuntas. As três técnicas podem ser aplicadas para dados binários, como o caso do modelo proposto necessita (MEYER, 2002).

Depois de aplicadas as técnicas sobre a matriz binária, os resultados foram ordenados pelo índice de similaridade e utilizados para gerar tabelas menores, uma para cada técnica, contendo somente os 16 itens mais similares para cada produto de Perfumaria do site. Essas tabelas de resultados dos itens similares foram ordenadas de forma decrescente pelo valor do item, visando que as recomendações fiquem mais baratas conforme o cliente receba os e-mails de lembretes. Na tabela 1 é possível ver a organização da tabela de produtos similares a ser utilizada para a recomendação de produtos.

Tabela 1 – Modelo de tabela de resultado de itens similares

Item abandonado	Rec. 1	Rec. 2	Rec. 3	Rec. 4	Rec. 5	Rec. 6	...	Rec. 16
11494	22636	76533	76681	82475	75906	76630	...	77314
19685	81850	70899	77200	74465	70670	74391	...	75411
19734	22249	22230	71876	22248	74446	71640	...	72315
19741	81690	81850	77509	81405	73856	80971	...	75474
70671	76682	73609	73614	81850	74030	25475	...	71403
...
81541	81850	81532	21372	74463	81114	77117	...	81113

Fonte: autoria própria (2021).

Para fazer a validação de qual técnica teve melhor desempenho em determinar a similaridade entre os itens, foi solicitado à três cientistas de dados da empresa que escolhessem qual das técnicas trouxe resultados mais satisfatórios. A avaliação feita

por eles foi qualitativa a partir de uma amostra com 20 produtos. Para cada item, eles analisaram qualitativamente quais foram as recomendações geradas em cada um dos casos para identificar qual continha a recomendação que mais fazia sentido para a empresa. Após a avaliação, a técnica escolhida foi correlação de Pearson, que trouxe resultados que mais se adequavam ao que era esperado, quando comparados com as demais técnicas.

Com a devida técnica escolhida, a tabela de similaridade é encaminhada ao sistema de envio de e-mails. Como os e-mails são montados e disparados nesse sistema, a manipulação da tabela pode ser feita de forma eficaz.

No momento da elaboração do e-mail de lembrete, é consultada a tabela de similaridade para a geração da vitrine de produtos recomendados. O conjunto de itens de recomendação é composto por quatro produtos e é dado a partir da entrada do item abandonado e do número da sequência do e-mail de alerta. O número da sequência do e-mail de alerta é responsável pelo índice do primeiro produto a ser recomendado, e os demais produtos são os três próximos na sequência da tabela de similaridade.

O conjunto de produtos recomendados (C_{rec}) é dado conforme o Algoritmo 1 a seguir, onde *item* representa o produto abandonado, *sequencia* o número de sequência do e-mail de alerta, M a matriz de itens similares.

Algoritmo 1 – Algoritmo para gerar conjunto de itens recomendados

Requer: $item \geq 0$ e $sequencia \geq 1$
Garantir: $sequencia \leq 4$ {Máximo valor}
 1 $j \leftarrow 4 \times (sequencia - 1)$
 2 **para** $i = 1$ até 4 **faz**
 3 $C_{rec}[i] \leftarrow M[item, j + i]$
 4 **finaliza para**

Fonte: autoria própria (2021).

O número da sequência de e-mail é considerado para a geração das recomendações pois, atualmente, o e-commerce em questão utiliza uma sequência de quatro e-mails de lembretes. Dessa forma, a cada e-mail enviado, as recomendações são diferentes e, devido à ordenação por preço, com preços mais baixos.

3.3.1 Testes de desempenho

Foram realizados medições do tempo de execução do modelo até a geração da tabela final com as recomendações que deveriam ser encaminhadas ao sistema de

envio de e-mails. O resultado apresentado foi calculado a partir da média do tempo de 5 execuções com dados de amostragem, utilizando somente os dados de produtos do departamento de Perfumaria.

Para a execução do processo de criação da tabela no formato necessário, preenchida com 1 quando os itens foram comprados pelos clientes, a média do tempo de execução foi de 1,61 segundos. Para a aplicação das métricas de similaridade sobre a matriz, as médias foram as seguintes: correlação de Pearson com 12,62 segundos; distância euclidiana com 8,22 segundos; e distância de Jaccard com 29,73 segundos. Para garantir que as recomendações contemplam futuras alterações de produtos no catálogo, esse processo será executado periodicamente, portanto o tempo de execução é um valor importante de ser considerado.

Para a geração da tabela de recomendação ordenada por preço, como foram geradas para validação, em cada métrica houve uma média diferente. Para a matriz gerada pela correlação de Pearson, a média de execução foi de 1,73 segundos; para a matriz de distância euclidiana, foi de 1,19 segundos; e, assim como em Pearson, para matriz de distância de Jaccard a média foi de 1,17 segundos. Apesar de não terem médias muito diferentes, como a correlação de Pearson foi a métrica escolhida, é importante que a média de execução tenha sido parecida com as demais.

Como somente dados do departamento de Perfumaria foram utilizados para os testes de desempenho, é importante ressaltar que a média de tempo teria sido superior caso fossem utilizados todos os produtos do e-commerce. Da base inicial de compras fornecida pela empresa, as compras de produtos do departamento de Perfumaria representam 35% do total (86.842 compras), sendo dividido em 337 dos 1.993 produtos, representando 17% dos itens comprados.

Por possuir quase seis vezes mais produtos do que o utilizado na amostragem tratada, o tempo de processamento seria muito maior. Para termos de comparação, enquanto a matriz de produtos do departamento de Perfumaria foi criada em 1,61 segundos, a matriz de todos os produtos do site levou 33,62 segundos para ser criada. Para a aplicação dos algoritmos de similaridade essa diferença ficaria muito maior devido ao tamanho da matriz de produtos, considerando que o número total é de 1.993. Apesar disso, essa alteração no tamanho da matriz tem grande impacto para a utilização do modelo por ser um processo off-line, fazendo com que somente a etapa de treinamento e processo de recomendação sejam impactadas e demorem mais para

executar.

3.4 *COLDSTART*

Coldstart é um problema que acontece quando não existe histórico sobre um produto (um produto novo no catálogo, por exemplo) e, dessa forma, não teria como fazer uma recomendação desse produto utilizando como base o que outros clientes também compraram além desse novo item (SANTANA, 2018).

Para o caso estudado, *coldstart* poderia ser considerado um problema somente para novos produtos no site, principalmente porque a recomendação seria enviada por e-mail após um abandono do novo produto no carrinho. Ou seja, não existe obrigatoriedade de compra para que a pessoa receba a recomendação por e-mail. Assim, até que houvesse um histórico de compras suficiente para gerar a recomendação de produtos para o novo item, não haveria similaridade com outros produtos por compras realizadas por outros clientes.

Para resolver esse problema, produtos novos no catálogo poderiam utilizar como recomendação, por exemplo, os produtos mais vendidos do mesmo departamento, semelhante à forma que ocorre atualmente. Dessa forma, ainda haveria a sessão de produtos recomendados no e-mail de lembrete, apesar de não ser customizada com a similaridade entre os produtos.

4 PLANEJAMENTO DOS TESTES

Para garantir que a utilização da vitrine gerada pelo sistema de recomendação realmente cause um impacto positivo na taxa de compras, é importante a utilização de testes antes de adotar o modelo gerado.

Para o e-commerce em questão, seria necessário um teste de pelo menos 1 mês para obter resultados satisfatórios para classificar o impacto da alteração da vitrine. Devido à esse tempo, não foi possível fazer a implementação e análise dos resultados até o momento de conclusão dessa monografia. Portanto, a seguir será descrito o planejamento do teste a ser executado e como será feita a coleta e análise dos resultados, assim como os resultados esperados após a finalização do teste.

4.1 TESTE A/B

Os testes A/B são experimentos controlados onde os usuários são divididos de maneira aleatória em subgrupos e expostos a diferentes experiências para analisar suas respostas. Nesses testes são utilizadas métricas para ter uma visão mais objetiva dos resultados (NEVES; DUDUCHI, 2019).

Para o teste do uso da vitrine, o tamanho dos subgrupos serão iguais e as métricas utilizadas serão taxas de abertura, clique e conversão do e-mail enviado. Além dessas, será necessária a métricas de valor médio da compra, que contém o valor médio gasto nas compras realizadas no período.

A taxa de abertura se refere à visualização das informações contidas no e-mail; clique se refere a quando o usuário acessa os links disponíveis para obter mais informações; conversão seria quando o usuário finaliza da compra após clicar no link do e-mail enviado.

Por se tratar de uma alteração no conteúdo do e-mail, é esperado que a maior diferença esteja na taxa de cliques e conversão, e até mesmo no valor médio de compras, considerando que é possível que uma recomendação agrade o cliente a ponto de adicionar o produto à sua compra final juntamente com o produto abandonado.

Com a ajuda do sistema de envio de e-mails, é possível definir links personalizados que ajudem a mensurar cliques e identificar tráfego no site, conhecido por UTM (*Urchin Tracking Module*) (ANDREWS, 2014). Dessa forma, serão utilizadas UTMs

diferentes para cada subgrupo, uma para a vitrine antiga e outra para a vitrine nova. Além disso, cada item dentro dos e-mails irá conter UTMs específicas para ser possível rastrear o que levou o cliente até o site.

Ou seja, nos e-mails de lembrete recebidos pelos clientes, cada item irá conter um link que nos dirá, ao final dos testes, se o e-mail recebido continha uma vitrine nova ou antiga, e de qual produto do e-mail partiu o clique do cliente. Os dois subgrupos receberão e-mails utilizando o mesmo modelo, conforme apresentado anteriormente na figura 2, a alteração se dará especificamente nos produtos apresentados na vitrine, indicando produtos mais vendidos (vitrine antiga) ou produtos de maior similaridade ao produto abandonado (vitrine nova).

Para interpretar o tráfego no site, uma outra ferramenta é utilizada, o Google Analytics. Ele compila toda a navegação do e-commerce, sendo possível rastrear de onde elas tiveram início. Dessa forma, através dos links personalizados com UTMs e da ferramenta para interpretação do tráfego, é possível descobrir como foi o desempenho do teste.

4.2 ANÁLISE DOS RESULTADOS E DISCUSSÃO

Após o período de 1 mês do início dos testes, serão extraídos do Google Analytics os resultados das métricas definidas (aberturas, cliques, conversão e receita média) originadas dos e-mails de lembrete. Devido às UTMs criadas anteriormente, será possível identificar exatamente quais ações que foram ocasionadas por quais e-mails.

Nos resultados obtidos, deverão ser analisadas as ações de cliques e conversão, que são os principais objetivos da inclusão de uma vitrine personalizada nos e-mails. Para as duas análises, deverão ser separados de grupos A e B, e também qual dos itens do e-mail levou ao clique ou conversão.

Os resultados podem ser apresentados em formato de tabela, conforme a tabela 2 a seguir, para melhor visualização. Como cada e-mail contém o produto abandonado e mais 4 itens recomendados, é importante que na visualização tenha as taxas tanto para o produto abandonado quanto para cada um dos itens recomendados enviados por e-mail. Com isso, é possível ter a maior visibilidade sobre qual dos 4 itens recomendados obteve uma taxa maior de clique, conversão e receita.

Tabela 2 – Modelo de tabela de apresentação de resultados

Nro email - Vitrine	Item Abandonado			Item Recomendado 1			...
	Clique	Conversão	Receita	Clique	Conversão	Receita	
1 - antiga	x%	y%	R\$	z%	w%	R\$...
1 - nova	x%	y%	R\$	z%	w%	R\$...
2 - antiga	x%	y%	R\$	z%	w%	R\$...
2 - nova	x%	y%	R\$	z%	w%	R\$...
3 - antiga	x%	y%	R\$	z%	w%	R\$...
3 - nova	x%	y%	R\$	z%	w%	R\$...
4 - antiga	x%	y%	R\$	z%	w%	R\$...
4 - nova	x%	y%	R\$	z%	w%	R\$...

Fonte: autoria própria (2021).

Entre os resultados esperados, não são esperadas grandes alterações de taxa de aberturas dos e-mails, por isso, essa métrica não será utilizada para a tomada de decisão. Além disso, também não é esperado que exista grandes alterações nas taxas de cliques e conversões dos itens abandonados, considerando que não há diferença desse conteúdo entre os grupos.

A maior diferença esperada é nas taxas de cliques das recomendações. Espera-se que com a nova vitrine trazendo produtos similares ao abandonado, o cliente seja atraído por esses itens e levado ao site para descobrir mais detalhes. Essa ação deve causar uma diferença significativa nas taxas de cliques das recomendações quando comparadas com o grupo que recebeu a vitrine antiga.

Ainda seguindo a mesma lógica apresentada pela taxa de cliques, também espera-se um aumento na conversão de recomendações. Por mais que o cliente realmente tenha desistido de finalizar a compra do produto abandonado, alguma sugestão da nova vitrine pode gerar interesse. Como um dos objetivos da nova vitrine é recomendar produtos com valores cada vez menores, caso o motivo da não-conversão do produto abandonado tenha sido o valor, recomendações de produtos similares, porém mais baratos, podem provocar um aumento de interesse por parte do cliente, levando à finalização da compra.

Além disso, o valor médio por compra é uma métrica que pode ser alterada com a nova vitrine. Incluir uma vitrine que traga produtos semelhantes ao abandonado no carrinho aumenta a possibilidade de que o cliente adicione à sua compra itens recomendados, além do produto abandonado. Dessa forma, em vez do valor da compra ser de somente o produto abandonado, terá acrescido também os valores dos itens recomendados. Com isso, o valor médio por compra pode acabar aumentando.

Da mesma forma, se o motivo para o cliente ter abandonado o produto no carrinho foi o preço, ele tem a possibilidade de comprar uma das recomendações de valor mais baixo. Assim, o valor médio por compra também é alterada, refletindo esse comportamento do consumidor ao preferir produtos de preços mais baixos.

Apesar desses resultados serem previamente esperados para os testes, outros comportamentos podem ser identificados. Por isso é necessária a existência do teste A/B durante o período determinado de 1 mês.

5 CONSIDERAÇÕES FINAIS

Um dos problemas que atinge e-commerces do mundo inteiro são os abandonos de carrinhos, ou seja, quando produtos são adicionados ao carrinho de compras e a compra não é finalizada. Na tentativa de recuperar esses consumidores, alguns desses e-commerces fazem uso de e-mails de lembrete, podendo conter também recomendações de produtos além do produto abandonado.

Na tentativa de melhorar as taxas de cliques e finalizações de compras ocasionadas por esses e-mails, foi desenvolvido um sistema de recomendações colaborativo para que os produtos indicados nessas comunicações atraíssem mais a atenção dos clientes do que simplesmente indicar produtos mais vendidos no site.

Apesar de não ter sido possível chegar a resultados concretos sobre a alteração das recomendações devido à necessidade de execuções de testes A/B, espera-se que ao final dos testes as taxas de cliques e conversões aumentem sobre os produtos recomendados. Além disso, também é esperado que haja alguma diferença na receita média das compras realizadas no período, devido à troca do produto abandonado por algum produto recomendado ou adição de outros itens ao que foi abandonado anteriormente.

O detalhamento e análise dos resultados dos testes A/B pode ser elaborada em trabalhos futuros, assim como a implementação do sistema de recomendação proposto para todos os itens do e-commerce.

REFERÊNCIAS

- ANDREWS, A. **As cinco melhores práticas para criação de tags UTM**. 2014. Disponível em: <https://imasters.com.br/analytics/as-cinco-melhores-praticas-para-criacao-de-tags-utm/?trace=1519021197&source=single>. Acesso em: 31 mai. 2021.
- BORGES, D. M.; OLIVEIRA, F. L. **Análise e comparação dos sistemas de recomendação de produtos existentes em três empresas de comércio eletrônico (ce): saraiva, submarino e amazon**. 2010.
- COVINGTON, P.; ADAMS, J.; SARGIN, E. **Deep neural networks for youtube recommendations**. 2016.
- DAOUD, M.; NAQVI, S.; AHMAD, A. Opinion Observer: Recommendation System on ECommerce Website. **International Journal of Computer Applications**, Citeseer, v. 105, n. 14, p. 37–42, 2014.
- MACIEL, L. M. S. **Comportamento de não-compra: motivações para o abandono do carrinho de compras online**. 2020.
- MARQUES, B. T.; ROLDAN, L. B. **E-mal Marketing - Um Estudo de Caso da Empresa Alfa**. 2017.
- MEYER, A. d. S. **Comparação de coeficientes de similaridade usados em análises de agrupamento com dados de marcadores moleculares dominantes**. 2002.
- MOTTA, C. et al. **Sistemas de recomendação**. 2011.
- NEOATLAS. **NEOATLAS: E-commerce Radar 1º Semestre de 2018**. 2018. Disponível em: <https://pt.scribd.com/document/387551835/Radar-E-commerce-1%5C%C2%5CBA-Semestre-2018>. Acesso em: 12 mai. 2021.
- NEVES, F.; DUDUCHI, M. **Testes A/B como uma das estratégias para inovação em produtos digitais: uma análise bibliométrica**. 2019.
- ROCHA FERNANDES, A. M. da; LINHARES, B. L. **Utilização de Técnicas de Sistemas de Recomendação para Aprimoramento de um E-commerce do Tipo B2b**. 2012.
- SANTANA, M. **Deep Learning para Sistemas de Recomendação: (Parte 1)—Introdução**. 2018. Disponível em: <https://medium.com/data-hackers/deep-learning-para-sistemas-de-recomenda%5C%C3%5CA7%5C%C3%5CA3o-parte-1-introdu%5C%C3%5CA7%5C%C3%5CA3o-b19a896c471e>. Acesso em: 12 abr. 2021.
- SERRANO, S. **Complete List of Cart Abandonment Statistics: 2006-2021**. 2021. Disponível em:

<https://www.barilliance.com/cart-abandonment-rate-statistics/#tab-con-7>.
Acesso em: 31 mai. 2021.

SIGILIANO, D.; FAUSTINO, E. NETFLIX: Sistemas de Recomendação Inteligentes
[http://dx. doi. org/10.15601/1983-7631/rt. v9n16p13-26](http://dx.doi.org/10.15601/1983-7631/rt.v9n16p13-26). **Revista Tecer**, v. 9, n. 16,
2016.

SOUZA, B. **Modelos de Fatoração Matricial para Recomendação de Vídeos**. 2012.