

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Sistemas de recomendação para priorização de tarefas realizadas por representantes de negócio

Bernardo Mota Barbosa

Dissertação de Mestrado do Programa de Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria (MECAI)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Bernardo Mota Barbosa

Sistemas de recomendação para priorização de tarefas realizadas por representantes de negócio

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria.
VERSÃO REVISADA

Área de Concentração: Matemática, Estatística e Computação

Orientador: Prof. Dr. Luis Gustavo Nonato

USP – São Carlos
Julho de 2023

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

M917s Mota Barbosa, Bernardo
Sistemas de recomendação para priorização de
tarefas realizadas por representantes de negócio /
Bernardo Mota Barbosa; orientador Luis Gustavo
Nonato. -- São Carlos, 2023.
47 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Mestrado Profissional em Matemática, Estatística
e Computação Aplicadas à Indústria) -- Instituto de
Ciências Matemáticas e de Computação, Universidade
de São Paulo, 2023.

1. Sistemas de recomendação. 2. Tarefas. 3.
Pontos de venda. 4. Personalização. 5. Vendas. I.
Gustavo Nonato, Luis, orient. II. Título.

Bernardo Mota Barbosa

**Recommender systems for prioritizing tasks performed by
business representatives**

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master – Professional Masters in Mathematics, Statistics and Computing Applied to Industry. *FINAL VERSION*

Concentration Area: Mathematics, Statistics and Computing

Advisor: Prof. Dr. Luis Gustavo Nonato

USP – São Carlos
July 2023

RESUMO

BARBOSA, B. M. **Sistemas de recomendação para priorização de tarefas realizadas por representantes de negócio**. 2023. 47 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

Durante a rotina de atendimento dos vendedores da Ambev aos pontos de vendas dos seus clientes, existem diversas atividades a serem executadas. Dentro de cada visita, o representante de negócio tem o papel de ouvir e resolver as dores de cada cliente, além disso tem a missão de oferecer novos produtos, serviços e também auxiliar o cliente a vender mais. Com uma grande quantidade de possibilidades de ações a serem executadas e pouco tempo disponível, surgiu a necessidade de personalizar cada uma das visitas realizadas aos milhares de clientes da empresa. Para isso, foi criada uma lista de possíveis tarefas a serem executadas dentro de um ponto de venda. Cada uma dessas tarefas tem um público específico de clientes. Após a definição das possíveis tarefas e de seus respectivos públicos surge um problema relevante de priorização, pois o tempo disponível para visitar todos os clientes ainda é pequeno em relação a quantidade de coisas a serem feitas. Por essa razão, neste trabalho, desenvolvemos um sistema de recomendação que visa priorizar as ações realizadas pelos vendedores durante as visitas aos clientes da Ambev.

Palavras-chave: Sistemas de recomendação, tarefas, pontos de venda, personalização, atendimento ao cliente, vendas.

ABSTRACT

BARBOSA, B. M. **Recommender systems for prioritizing tasks performed by business representatives**. 2023. 47 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

During the routine of Ambev's salespeople at their customers' points of sale, there are several activities to be performed. Within each visit, the business representative has the role of listening to and solving the pain of each customer, in addition to offering new products, services and also helping the customer to sell more. With a large number of possibilities for actions to be performed and little time available, the need to customize each of the visits made to the company's thousands of customers arose. For this, a list of possible tasks to be performed within a point of sale was created. Each of these tasks has a specific audience of customers. After defining the possible tasks and their respective audiences, a relevant prioritization problem arises, as the time available to visit all customers is still small in relation to the amount of things to be done. For this reason, in this work, we developed a recommendation system that aims to prioritize the actions taken by salespeople during visits to Ambev customers.

Keywords: Recommendation systems, tasks, points of sale, personalization, customer service, sales.

SUMÁRIO

1	INTRODUÇÃO	11
2	REFERENCIAL TEÓRICO	15
2.1	Sistemas de recomendação	15
2.1.1	<i>Filtro colaborativo</i>	16
2.1.1.1	<i>Método ALS</i>	17
2.1.2	<i>Recomendação baseada em conteúdo</i>	20
3	METODOLOGIA	23
3.1	Base de dados	24
3.2	Atributos das tarefas	25
3.2.1	<i>Critério de elegibilidade</i>	25
3.2.2	<i>Critério de efetividade</i>	26
3.2.3	<i>Racional score primário e secundário</i>	28
3.3	<i>Task pool</i>	29
3.4	Modelagem proposta	31
3.5	Teste de hipótese	35
3.5.1	<i>Hipóteses e teste</i>	35
3.5.2	<i>Seleção da base</i>	35
3.5.2.1	<i>Crítérios de elegibilidade</i>	36
3.5.3	<i>Plano amostral</i>	37
3.5.4	<i>Teste da regressão logística</i>	37
3.5.5	<i>Estatísticas das amostras</i>	39
4	RESULTADOS	41
4.1	Experimento	41
4.2	Piloto	43
5	CONCLUSÃO	45
	REFERÊNCIAS	47

INTRODUÇÃO

Em 2022, a AB InBev, que é a holding global da Ambev, manteve sua posição como a maior cervejaria do mundo. Ela produziu um total de 595 milhões de hectolitros e teve um crescimento significativo, alcançando uma receita anual de 57 bilhões de dólares. Para gerenciar e vender toda essa produção, a Ambev utiliza duas principais ferramentas: o aplicativo BEES, que permite aos clientes fazerem pedidos de forma online sem a necessidade de interação com funcionários, e uma equipe de vendas que conta com mais de mais de 5 mil vendedores apenas no Brasil.

Vale destacar que em 2023 a Ambev está buscando se reinventar no mercado com a diversificação de seus negócios. A empresa, que vem investindo em tecnologia, principalmente na área de dados e agora tem como objetivo se tornar uma plataforma de marcas, vendendo não apenas cerveja, mas também outros produtos para seus clientes, como itens de mercado, leite, carne, bebidas destiladas, entre outros.

Desde seus primeiros anos de história, os vendedores da Ambev sempre tiveram o papel de atender os clientes pessoalmente, principalmente para receber os pedidos, mas também para oferecer novos produtos e realizar várias outras atividades. Segundo [GREGORINI et al. \(2014\)](#), a venda de produtos aos cliente, acontecia através da visita do vendedor até cada ponto de venda. A solicitação do pedido de compra era feita e ao final do dia, depois de todos os clientes atendidos, o vendedor realizava a descarga do que ele vendeu no dia para que os pedidos fossem entregues posteriormente. Em 2019, a Ambev criou o BEES, um marketplace para seus clientes. Com a pandemia de COVID-19 no primeiro trimestre de 2020, a divulgação e adoção do aplicativo foi bastante acelerada, uma vez que o contato físico entre as pessoas foi limitado.

Nesse cenário de digitalização, a força de vendas perdeu uma de suas principais funções, que era a de realizar os pedidos para os clientes, uma vez que muitos donos de pontos de venda agora conseguem fazer isso de maneira autônoma através do BEES.

Considerando essa mudança no processo de vendas, a Ambev optou por direcionar os esforços de sua equipe de representantes de negócio para outras atividades. Com o intuito de tornar esse processo mais inteligente e automatizado, foi lançado no Brasil um projeto chamado Tasks do Bees Force, que tem como objetivo criar tarefas personalizadas para os vendedores que atendem os mais de 1 milhão de pontos de venda da Ambev em todo o país.

O ciclo de uma tarefa é composto pelas seguintes etapas:

- 1- Definição das regras de negócio com as características da tarefa;
- 2- Criação das tarefas;
- 3- Priorização para as visitas;
- 4- Análise de efetividade das tarefas;

Cluster: Marketplace	Tarefa: Engajamento de marketplace	Elegibilidade: 1 - O faturamento de marketplace dos últimos 60 dias é maior que zero. 2 - O faturamento de marketplace dos últimos 30 dias é zero.	Score da tarefa: 80
Texto da tarefa: PDV não comprou SKUs do marketplace nos últimos 30 dias. Auxilie na cobertura.		Efetividade: Verificar dentro da tabela de notas fiscais se ocorreu alguma compra de algum sku de marketplace dentro do prazo de efetividade.	

Figura 1 – Exemplo de tarefa do cluster de Marketplace.

Na Figura 1, podemos observar as principais características de uma tarefa. As tarefas são divididas em clusters, definidos pelo negócio por seu objetivo principal. O cluster de Marketplace tem como objetivo alavancar a venda de produtos de diferentes empresas parceiras da Ambev.

Cada tarefa possui critérios específicos de elegibilidade, efetividade, score e texto, estabelecidos pelos gestores de negócio. Nesse exemplo, tal tarefa pode ser elegível para todos os pontos de venda que tenham realizado alguma compra de marketplace nos últimos 60 dias, mas que não tenham feito nenhuma compra nos últimos 30 dias, indicando um possível desinteresse em continuar comprando. Para priorizar a execução dessa tarefa em relação às outras, ela é atribuída a um score de 0 a 100. O vendedor recebe no aplicativo uma descrição detalhada da tarefa e para que a ação seja contabilizada em sua remuneração, é necessário que o critério de efetividade seja cumprido. Em outras palavras, é preciso que haja uma nota fiscal com alguma compra de marketplace após a visita do vendedor ao ponto de venda.

Das 4 etapas citadas, a terceira é de suma importância para o sucesso do ecossistema de tarefas. Em média os representantes de negócio visitam cada um dos pontos de vendas (PDVs) semanalmente, sendo assim o tempo de visita em cada cliente é limitado e curto. Como no exemplo mostrado, existem dezenas de tipos de tarefas, o que gera milhões de ações todos os meses.

Em um cenário de muitas tarefas e pouco tempo para realizar tudo, é necessário priorizar o que vai ser executado em cada cliente, para isso são definidos o tempo em que cada visita deve durar, o tempo que cada tarefa demora em média e a prioridade de cada tarefa em uma escala de 0 a 100. Sendo assim, o algoritmo de priorização seleciona a maior quantidade de tarefas possível dentro do tempo disponível para uma visita seguindo a ordem do score e algumas outras regras já definidas.

O grande desafio é que tais definições de score são arbitrárias e definidas por conhecimento dos gestores de negócio. Tais atribuições não captam as diferenças entre os mais de 1 milhão de pontos de venda no Brasil, em relação à capacidade de execução de cada uma das tarefas. O que acaba gerando muitas vezes a priorização constante de tarefas que são muito difíceis de serem executadas ou que não são as mais prováveis a serem efetivas dentro da visita a um ponto de venda.

Atualmente já conta-se com quase 200 milhões de tarefas dentro do nosso banco de dados histórico. Toda essa quantidade de informação é valiosa e pode ser utilizada para melhorar o nosso produto cada vez mais.

Estamos com mais de 85 possíveis tarefas distintas que um RN pode receber em uma visita. Diante desse cenário surge a necessidade de priorizarmos melhor as tarefas ao longo das visitas visando uma efetividade maior. Temos PDVs nos quais sobram tarefas mensalmente e outros nos quais insistimos diversas vezes na mesma tarefa que tem pouca chance de ser efetiva, desperdiçando assim o tempo precioso dos nossos RNs.



Figura 2 – Conceito de Score 2.0 para priorização de tarefas.

Para lidar com esse desafio de priorização criamos o conceito do Score 2.0, na Figura 2 podemos ver suas três frentes, a de probabilidade de conversão da tarefa, a que pondera o impacto que a tarefa gera dentro do ponto de venda e a terceira que leva em consideração os planos comerciais para tornar um conjunto de tarefas mais relevante por um período de tempo.

Para ilustrar este conceito, podemos dar um exemplo de uma tarefa que consiste na introdução de um produto específico em um ponto de venda que já adquire outros produtos de nosso marketplace. Vamos considerar a venda de produtos como óleo ou arroz, levando em conta as características do ponto de venda (PDV) e do responsável pela negociação (RN). Podemos calcular uma probabilidade de sucesso, ou seja, a probabilidade de que o RN consiga efetivamente realizar essa tarefa no PDV em questão. Essa probabilidade de conversão varia entre 0% e 100%.

Suponhamos que seja relativamente fácil introduzir esse novo produto nesse PDV, resultando em uma probabilidade de conversão de 90%. No entanto, caso o PDV não tenha comprado óleo ou arroz nas semanas seguintes à realização dessa tarefa, podemos concluir que essa ação não gerou um impacto significativo. Medir esse impacto é a segunda frente do Score 2.0.

Por fim, a frente de estratégia comercial relaciona-se ao quanto o time de negócios quer focar nas vendas dos produtos de nosso marketplace em detrimento de outras atividades. Se esse for o foco principal da equipe, a tarefa de introdução do produto deve ser priorizada, mesmo que a probabilidade de conversão ou o impacto imediato não sejam tão altos.

Nosso primeiro foco é o de desenvolver uma espécie de probabilidade de conversão de determinadas tarefas, sendo assim nesse trabalho vamos explicar mais a fundo a metodologia proposta para lidar com tal desafio.

Portanto, o propósito deste trabalho é investigar e propor uma modelagem baseada em sistemas de recomendação que seja capaz de entender o comportamento dos clientes em relação às tarefas e realizar recomendações das tarefas mais prováveis de serem convertidas dentro de cada um dos pontos de vendas.

O presente trabalho segue da seguinte forma: a Seção 2 apresenta uma revisão bibliográfica que tem como objetivo proporcionar ao leitor uma visão geral das contribuições que inspiraram as modelagens realizadas neste estudo. Na Seção 3, serão apresentadas informações acerca do contexto, além da descrição da base de dados e das técnicas utilizadas para a realização dos estudos, bem como as definições pertinentes. Por fim, na Seção 4, serão apresentados os resultados dos experimentos e do estudo piloto conduzido.

REFERENCIAL TEÓRICO

2.1 Sistemas de recomendação

Segundo [ISINKAYE; FOLAJIMI; OJOKOH \(2015\)](#), os sistemas de recomendação lidam com o desafio de extrair informações úteis de uma grande quantidade de dados dinâmicos, informações essas geradas de acordo com as preferências, interesse ou comportamento do usuário. Tais sistemas têm a capacidade de prever se um determinado usuário vai preferir um item ou não com base no seu perfil.

Os sistemas de recomendação se desenvolveram em paralelo com a web. Eles foram inicialmente baseados em filtragem demográfica, baseada em conteúdo e colaborativa. Atualmente, esses sistemas estão incorporando informações sociais. No futuro, eles usarão informações implícitas, locais e pessoais da Internet das coisas ([BOBADILLA et al., 2013](#)).

Desde que surgiu, a teoria e aplicação desses sistemas tem se desenvolvido rapidamente e está cada vez mais presente e valorizado no mundo corporativo. Muitos dos maiores sites de comércio já estão usando sistemas de recomendação para ajudar seus clientes a encontrar produtos para comprar ([SCHAFER; KONSTAN; RIEDL, 1999](#)). Pesquisadores e gerentes reconhecem que os sistemas de recomendação oferecem grandes oportunidades e desafios para negócios, governo, educação e outros domínios ([LU et al., 2015](#)).

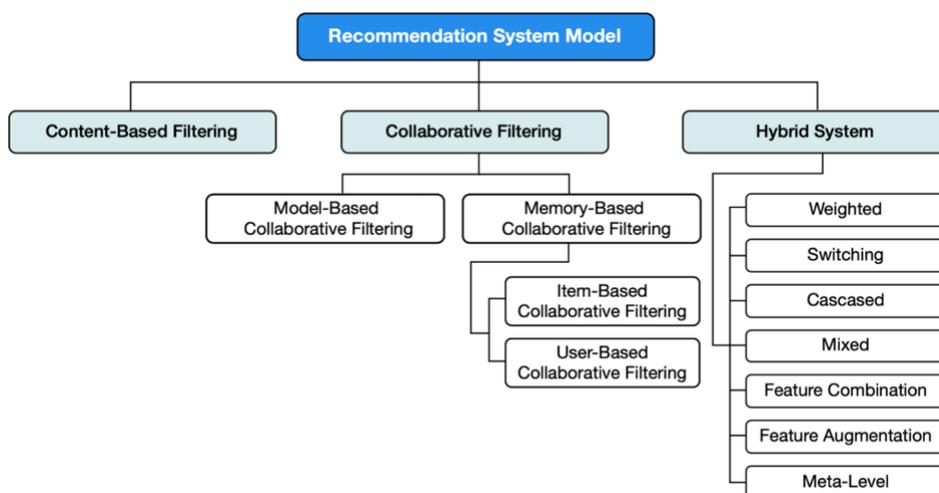


Figura 3 – Visão geral dos modelos de recomendação. Fonte: (KO *et al.*, 2022).

A Figura 3 apresenta a divisão entre as principais abordagens de sistema de recomendação, a recomendação baseada em conteúdo, a de filtro colaborativo e a que une essas duas em uma visão híbrida dos problemas.

Vale destacar que vamos apresentar os métodos de filtro colaborativo e recomendação baseada em conteúdo, apesar de neste trabalho em um primeiro momento só utilizar o filtro colaborativo para realizar a modelagem.

2.1.1 Filtro colaborativo

O filtro colaborativo é um tipo comum de sistema de recomendação, utilizado para filtrar informações usando as interações e dados coletados pelo comportamento de outros usuários. É baseado na ideia de que as pessoas que concordaram nas suas avaliações de certos itens provavelmente vão continuar concordando novamente no futuro.

Na Figura 4, é possível observar a ideia central por trás da abordagem de filtro colaborativo. Um cliente que consome Brahma e Skol é considerado semelhante a outro cliente que, além desses dois produtos, também consome Bohemia. A partir dessa semelhança, é provável que o primeiro cliente possa também se interessar pelo produto Bohemia.

Na prática, as aplicações do filtro colaborativo envolvem um número muito maior de interações entre clientes e produtos, e podem ser aplicadas em diversos cenários. Essa abordagem tem se mostrado bastante eficaz em sistemas de recomendação, ajudando empresas a personalizar a experiência do usuário e aumentar as vendas.

As avaliações dos itens para a modelagem por filtro colaborativo podem ser obtidas por meios implícitos ou explícitos. No caso explícito é solicitado ao usuário que faça uma avaliação do item, em uma escala de 0 a 5, por exemplo. Já no cenário implícito, tal avaliação é inferida a partir de algum comportamento do usuário, como realizar a compra de um produto, assistir

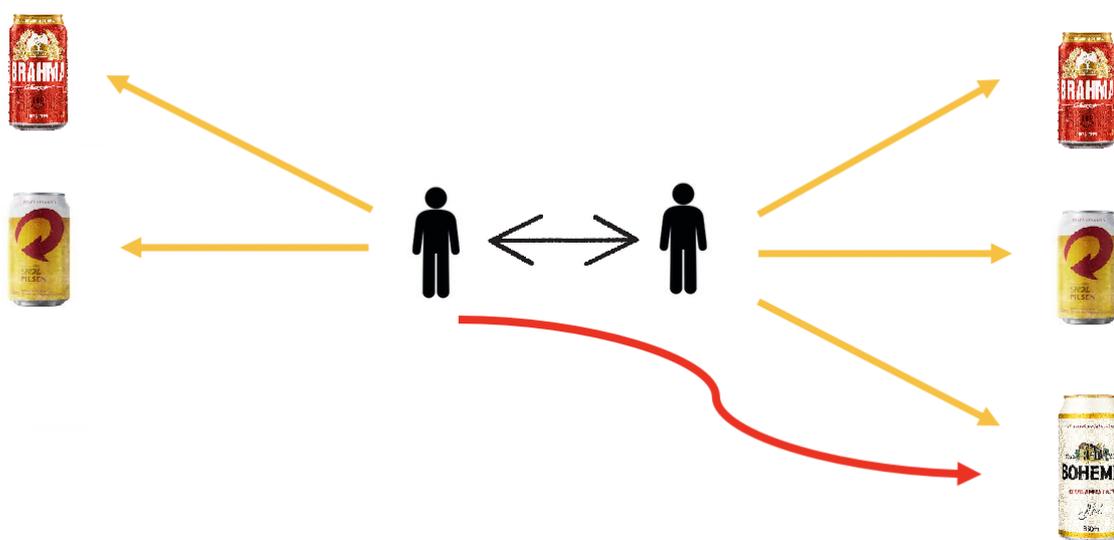


Figura 4 – Exemplo de estrutura do filtro colaborativo.

determinado filme ou série, pesquisar ou colocar determinado item no carrinho, etc.

Como descrito na Figura 3, os algoritmos de filtro colaborativo baseados em memória são divididos em duas vertentes. Os algoritmos baseados no usuário geram previsões com base nas semelhanças entre os usuários, os algoritmos baseados em itens geram previsões com base nas semelhanças entre itens. Além desses algoritmos, existe a abordagem baseada em modelo, como por exemplo a fatoração de matriz usando o algoritmo *Alternating Least Squares* (ALS), que é um tipo de filtragem colaborativa usada para resolver problemas de superajuste em dados esparsos.

2.1.1.1 Método ALS

O método ALS (*Alternating Least Squares*) é uma técnica comum dentro dos sistemas de recomendação de filtro colaborativo baseados em modelo. Esses sistemas são usados para entender as preferências dos usuários em relação a um conjunto de itens, com base no histórico de interações desses usuários com esses itens e nas interações de outros usuários com os mesmos itens.

Tal método é uma abordagem matemática que busca minimizar o erro quadrático médio entre as preferências previstas e as preferências reais dos usuários. Ele funciona alternando entre a otimização de duas matrizes: a matriz de usuários e a matriz de itens. O objetivo é encontrar as representações latentes desses usuários e itens em um espaço vetorial de baixa dimensão.

A matriz de usuários representa as preferências dos usuários para cada item, enquanto a matriz de itens representa as características latentes de cada item. Essas características latentes podem incluir, por exemplo, gênero de um filme, autor de um livro, ou outros atributos relevantes.

O método ALS é uma técnica eficaz para sistemas de recomendação, pois permite uma escalabilidade e eficiência computacional, além de apresentar resultados precisos e personalizados para cada usuário. Isso faz com que o ALS seja amplamente utilizado em plataformas de streaming, e-commerce e redes sociais, que dependem de sistemas de recomendação para melhorar a experiência do usuário e aumentar o engajamento e conversão.

Essa abordagem é particularmente útil quando as preferências dos usuários são implícitas, ou seja, quando as classificações são interpretadas como confiança em vez de uma avaliação explícita. Essa técnica permite que as classificações ajam como valores de confiança relacionados à força das preferências do usuário indicadas, em vez de classificações explícitas dadas aos itens.

Segundo [HU; KOREN; VOLINSKY \(2008\)](#),

$$p_{ui} = \begin{cases} 1 & r_{ui} > 0 \\ 0 & r_{ui} = 0 \end{cases}$$

são definidas duas notações, a r_{ui} que mede a confiança e as variáveis binárias p_{ui} , que indicam a preferência do usuário u pelo item i . Se um usuário u consumiu o item i ($r_{ui} > 0$), então temos uma indicação de que u goste de i ($p_{ui} = 1$), caso contrário $p_{ui} = 0$. No geral quanto maior o valor de r_{ui} mais confiança temos de que o usuário goste do item. Vale destacar que o $p_{ui} = 0$ pode indicar que o usuário não gosta de um item mas também pode significar que ele desconhece sua existência ou que não teve a oportunidade de consumi-lo por algum motivo como preço ou disponibilidade.

O método busca estimar a matriz de avaliações R como o produto de duas matrizes de posto menor, X e Y , ou seja, $X * Y^t = R$. A abordagem é iterativa. Durante cada iteração, uma das matrizes de fatores latentes é mantida constante, enquanto a outra é resolvida usando mínimos quadrados. A matriz fator recém-resolvida é então mantida constante enquanto se resolve a outra.

id_usuario/id_produto	1	2	3	4	5	6	7
1	-	-	-	-	-	-	-
2	5	-	-	-	-	-	-
3	-	-	2	-	3	-	-
4	2	-	-	-	-	-	4
5	-	-	-	-	-	5	-
6	4	-	1	-	-	-	-
7	-	-	-	-	-	-	-
8	-	-	4	-	2	-	-

$$=$$

	fl_u_1	fl_u_2	fl_u_3
usuario_1	0,1200	0,6163	0,4034
usuario_2	0,8762	0,1400	1,0259
usuario_3	0,1578	0,0100	0,4821
usuario_4	0,3829	0,5765	0,7254
usuario_5	0,4012	0,4764	1,1222
usuario_6	0,1400	0,4377	0,1249
usuario_7	0,2511	0,5410	0,2788
usuario_8	0,9521	0,0092	1,0100

$$\times$$

	produto_1	produto_2	produto_3	produto_4	produto_5	produto_6	produto_7
fl_p_1	1,4723	1,1756	1,4880	0,9825	0,4592	0,0013	0,8585
fl_p_2	0,3969	0,5923	1,1607	0,9095	1,3877	0,7672	0,1545
fl_p_3	0,5867	0,0302	0,4639	0,2250	0,1551	0,4547	0,5528

$k \times n$

Figura 5 – Exemplo de decomposição de matrizes no método ALS.

Na Figura 5, temos um exemplo da aplicação do método, onde realizamos a decomposição da matriz R de avaliações implícitas dos usuários em relação a um conjunto de produtos em duas matrizes menores de fatores latentes, uma de dimensão $m \times k$ e outra de dimensão $k \times n$, onde m é o número de usuários, n é o número de produtos e k é o número de fatores latentes. Os

fatores latentes representam características ocultas relevantes para a recomendação de produtos e são utilizados para prever as preferências dos usuários. No início do processo, os valores das matrizes de fatores latentes são aleatórios.

id_usuario/id_produto	1	2	3	4	5	6	7
1	-	-	-	-	-	-	-
2	5	-	-	-	-	-	-
3	-	-	2	-	3	-	-
4	2	-	-	-	-	-	4
5	-	-	-	-	-	5	-
6	4	-	1	-	-	-	-
7	-	-	-	-	-	-	-
8	-	-	4	-	2	-	-

$$=$$

	fl_u_1	fl_u_2	fl_u_3
usuario_1	0,1200	0,6163	0,4034
usuario_2	0,8762	0,1400	1,0259
usuario_3	0,1578	0,0100	0,4821
usuario_4	0,3829	0,5765	0,7254
usuario_5	0,4012	0,4764	1,1222
usuario_6	0,1400	0,4377	0,1249
usuario_7	0,2511	0,5410	0,2788
usuario_8	0,9521	0,0092	1,0100

$$\times$$

	produto_1	produto_2	produto_3	produto_4	produto_5	produto_6	produto_7
fl_p_1	1,4723	1,1756	1,4880	0,9825	0,4592	0,0013	0,8585
fl_p_2	0,3969	0,5923	1,1607	0,9095	1,3877	0,7672	0,1545
fl_p_3	0,9867	0,0302	0,4639	0,2250	0,1551	0,4547	0,5528

Fixo *Fixo* *Varia*

Figura 6 – Exemplo de interação da matriz de fatores latentes dos itens.

Na Figura 6 vemos como funciona a primeira iteração do algoritmo onde a matriz de fatores latentes dos usuários permanece fixa, enquanto a matriz de fatores latentes dos produtos é atualizada. Nessa etapa, o método utiliza a matriz de avaliações R e a matriz de fatores latentes dos usuários para atualizar a matriz de fatores latentes dos produtos. A matriz de fatores latentes dos produtos é atualizada de forma a minimizar o erro entre a matriz de avaliações R e o produto das matrizes de fatores latentes.

id_usuario/id_produto	1	2	3	4	5	6	7
1	-	-	-	-	-	-	-
2	5	-	-	-	-	-	-
3	-	-	2	-	3	-	-
4	2	-	-	-	-	-	4
5	-	-	-	-	-	5	-
6	4	-	1	-	-	-	-
7	-	-	-	-	-	-	-
8	-	-	4	-	2	-	-

$$=$$

	fl_u_1	fl_u_2	fl_u_3
usuario_1	0,1200	0,6163	0,4034
usuario_2	0,8762	0,1400	1,0259
usuario_3	0,1578	0,0100	0,4821
usuario_4	0,3829	0,5765	0,7254
usuario_5	0,4012	0,4764	1,1222
usuario_6	0,1400	0,4377	0,1249
usuario_7	0,2511	0,5410	0,2788
usuario_8	0,9521	0,0092	1,0100

$$\times$$

	produto_1	produto_2	produto_3	produto_4	produto_5	produto_6	produto_7
fl_p_1	1,4723	1,1756	1,4880	0,9825	0,4592	0,0013	0,8585
fl_p_2	0,3969	0,5923	1,1607	0,9095	1,3877	0,7672	0,1545
fl_p_3	0,9867	0,0302	0,4639	0,2250	0,1551	0,4547	0,5528

Fixo *Varia* *Fixo*

Figura 7 – Exemplo de interação da matriz de fatores latentes dos usuários.

Na Figura 7, ocorre o contrário: a matriz de fatores latentes dos produtos permanece fixa, enquanto a matriz de fatores latentes dos usuários é atualizada a cada interação. Nesse caso, o método utiliza a matriz de avaliações R e a matriz de fatores latentes dos produtos para atualizar a matriz de fatores latentes dos usuários. A matriz de fatores latentes dos usuários é atualizada de forma a minimizar o erro entre a matriz de avaliações R e o produto das matrizes de fatores latentes.

Essas interações são repetidas diversas vezes até que o erro seja minimizado e a matriz de fatores latentes resultante seja capaz de prever com alta precisão as preferências dos usuários. Um exemplo do comportamento do erro RMSE ao longo dessas iterações é apresentado na Figura 8, ilustrando como o erro diminui à medida que o processo avança.

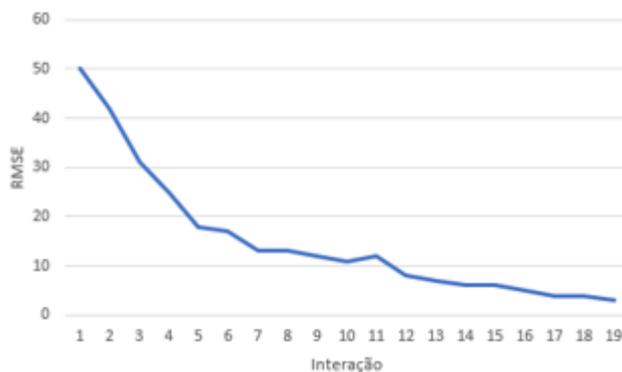


Figura 8 – Exemplo de gráfico do RMSE ao longo das iterações.

Vale destacar que a otimização dos parâmetros no modelo é realizada uma abordagem iterativa que envolve a alternância entre a otimização dos fatores dos usuários e dos itens. Quando um conjunto de fatores é mantido fixo, a função de custo do modelo se torna uma função quadrática, permitindo encontrar o mínimo global de forma eficiente. A cada iteração, os fatores são recalculados usando a matriz de preferências binárias p_{ui} e a matriz de confiança associada aos usuários e itens. Essa alternância entre recalculando os fatores garante que o valor da função de custo seja reduzido a cada passo, levando à convergência do processo de otimização.

Essa abordagem é similar ao gradiente descendente, pois ambos são métodos de otimização iterativos usados para encontrar os mínimos de uma função de custo. No entanto, esse método é mais adequado para esse modelo de fatoração devido à sua estrutura de custo quadrática e à natureza dos dados de feedback implícito. Tal técnica é descrita com mais detalhes em [HU; KOREN; VOLINSKY \(2008\)](#). No artigo, os autores descrevem detalhadamente como o método é utilizado para otimizar os parâmetros do modelo e apresentam resultados experimentais para avaliar a qualidade das recomendações geradas pelo método.

2.1.2 Recomendação baseada em conteúdo

A filtragem baseada em conteúdo utiliza feedbacks implícitos e explícitos para ter a informação de quais itens o usuário gosta e partir disso recomenda ao usuário outros itens que são semelhantes aos avaliados positivamente.

ID	Nome	Culinária	Serviço	Custo
10001	Mike's Pizza	Italiano	Balcão	Baixo
10002	Chris's Cafe	Francês	Mesa	Médio
10003	Jacques Bistro	Francês	Mesa	Alto

Tabela 1 – Exemplo de base de dados de restaurante. Fonte: ([PAZZANI; BILLSUS, 2007](#)).

Na tabela 1 podemos observar um exemplo de base de dados utilizada para ajustar uma recomendação baseada em conteúdo. Temos três restaurantes que apresentam diferentes

culinárias, tipos de serviço e custo. O restaurante Chris's Cafe é bem semelhante ao Jacques Bistro, o que sugere que pessoas que gostam de um tendem a gostar do outro também, pois são itens semelhantes.

Segundo PAZZANI; BILLSUS(2007), o banco de dados representado na tabela 1 pode por exemplo ser usado como base para um site que recomenda restaurantes. Obviamente, um site normalmente tem mais informações do que as mostradas na tabela 1, como uma descrição de texto do restaurante, uma avaliação do restaurante ou até mesmo um menu. Esses podem ser facilmente armazenados como campos adicionais no banco de dados e utilizados na modelagem.

A abundância de informações disponíveis na internet, em combinação com sua natureza dinâmica e heterogênea, gerou uma dificuldade crescente em encontrar o que queremos quando precisamos de uma maneira rápida. Como consequência, o papel da modelagem do usuário e do acesso personalizado à informação está se tornando crucial, os usuários precisam de um suporte personalizado para filtrar grandes quantidades de informações disponíveis, de acordo com seus interesses e gostos (LOPS; GEMMIS; SEMERARO, 2011). Tal cenário tem tornado essencial o uso e o avanço das técnicas de recomendação baseada em conteúdo.

METODOLOGIA

Esta seção descreve o conjunto de dados, detalha o contexto das tarefas estudadas e descreve o método utilizado para coletar evidências e avaliar a abordagem proposta.

O objetivo deste trabalho é propor uma modelagem de sistemas de recomendação que seja capaz de realizar uma predição para ranquear cada uma das tarefas em cada um dos PDVs, o intuito é posicionar as tarefas mais propensas a serem efetivas no topo do ranking, permitindo que enviemos primeiro para as visitas aquelas tarefas com mais chance de serem convertidas, otimizando assim o tempo dos nossos RNs.

Desenvolvemos nossa modelagem com base em validações utilizando dados históricos. Após a construção e o alcance de métricas satisfatórias, implementamos a priorização personalizada utilizando o ranking gerado pelo sistema de recomendação em um piloto realizado com dois grupos de RNs na prática. O objetivo foi validar a hipótese de que nossa modelagem realmente proporciona melhores resultados na execução das tarefas por parte dos RNs, em comparação com a metodologia de priorização atual.

É importante ressaltar que não podemos simplesmente utilizar o percentual de efetividade das tarefas históricas. Isso ocorre porque a cada semana introduzimos novas tarefas e, muitas vezes, algumas delas não têm uma quantidade relevante de realizações em visitas a determinados PDVs, o que prejudica o cálculo de um bom percentual de conversão histórica.

Para solucionar esse desafio, desenvolvemos tal modelo preditivo que busca estimar o comportamento da probabilidade de conversão das tarefas, levando em conta as características do comportamento de cada PDV em relação a efetivação das tarefas.

3.1 Base de dados

O conjunto de dados é composto por 6.483.598 observações de tarefas, provenientes de um total de 200.709 pontos de vendas. Cada tarefa foi gerada com base em um critério de elegibilidade e está associada a um dos 7 clusters definidos pelo time de negócios. Ao todo, a base contém 79 tipos distintos de tarefas. Vale ressaltar que este conjunto é uma amostra selecionada de uma base ainda maior, que inclui mais de 200 milhões de tarefas provenientes de mais de 1 milhão de clientes.

As tarefas começaram a ser geradas em 24/02/2022, nos primeiros meses foram realizados alguns testes e análises que promoveram grandes mudanças. Sendo assim, a base analisada só considera tarefas de visitas que foram realizadas entre 01/08/2022 e 31/01/2023.

Cluster	Número de tarefas	Percentual de tarefas
Portfólio	2170947	33,44%
Volume	32170	00,49%
Execução	1819683	28,06%
Nível de serviço	270305	04,17%
Digitalização BEES	1274986	19,66%
Serviços	273690	04,22%
Marketplace	641817	09,97%

Tabela 2 – Distribuição de tarefas por cluster.

A distribuição das tarefas pelos 7 clusters pode ser observada na Tabela 2. É interessante notar que o cluster de Portfólio é responsável pela maior quantidade de tarefas, totalizando 2.170.947 (33,44%) ações. Em seguida, o cluster de Execução corresponde a 28,06% das tarefas, enquanto os clusters de Volume, Nível de Serviço e Serviços apresentam um percentual relativamente baixo em relação aos demais. Por outro lado, os clusters de Digitalização BEES e Marketplace somados correspondem a quase 30% do total de tarefas, o que sugere que esses também são áreas de grande relevância para a atuação da companhia.

3.2 Atributos das tarefas

O objetivo dessa seção é explicar todo o contexto das tarefas, suas características e atributos.

3.2.1 Critério de elegibilidade

O critério de elegibilidade é uma função com retorno *True* ou *False*. Essa função compreende um conjunto de uma ou mais métricas que devem ser satisfeitas para retornar *True*, o que significa que a tarefa pode aparecer para o representante de negócio (RN) realizar no PDV.

Essa análise é crucial para garantir a precisão das tarefas geradas para cada ponto de venda. Se alguma tarefa estiver sendo gerada incorretamente para um PDV, é necessário revisar os critérios de elegibilidade para corrigir o problema.

Tarefa	Critério de elegibilidade 1: Temos o produto no estoque?	Critério de elegibilidade 2: Volume de compras do mês = 0?	Check critérios elegibilidade
Venda da cerveja Spaten 600ml	OK	OK	OK
Venda da cerveja Spaten 600ml	OK	NOK	NOK
Venda da cerveja Spaten 600ml	NOK	OK	NOK

Tabela 3 – Exemplificação do check de critérios de elegibilidade.

Na tabela 3 podemos ver que para a tarefa de venda da cerveja Spaten ser elegível ao ponto de venda, é necessário que os dois critérios de elegibilidade sejam atendidos.

3.2.2 Critério de efetividade

A efetivação das tarefas é a etapa final do ciclo de vida de cada uma delas, na qual o objetivo principal da tarefa é alcançado. É nessa fase que se verifica se as ações propostas foram realizadas de forma bem-sucedida. Por exemplo, consideremos uma tarefa que tem como propósito orientar o RN a incentivar a venda de produtos de marketplace para um determinado cliente. A efetivação dessa tarefa ocorrerá quando for possível verificar, por meio da tabela de notas fiscais, se houve efetivamente uma compra de algum SKU de marketplace por parte do PDV dentro do prazo estabelecido para a tarefa. Dessa maneira, uma tarefa que atende ao critério dentro do prazo estipulado é considerada efetiva (*True*), caso contrário não efetiva (*False*).

Tarefa	Data de completude da tarefa	Prazo de efetividade (dias)	Data máxima para efetividade	Critério de efetividade 1: Reconhecimento de imagem = "OK"?	Critério de efetividade 2: Reconhecimento de imagem realizado entre 5/12/2021 e 19/12/2021?	Check critérios de efetividade
Negociação de ponto extra Brahma 269ml	05/12/2021	14	19/12/2021	OK	OK	OK
Negociação de ponto extra Brahma 269ml	05/12/2021	14	19/12/2021	OK	NOK	NOK
Negociação de ponto extra Brahma 269ml	05/12/2021	14	19/12/2021	NOK	OK	NOK

Tabela 4 – Exemplificação do check de critérios de efetividade.

Cabe destacar que não é obrigatório uma tarefa apresentar critério de efetividade. Mas é uma ferramenta importante para metrificar as entregas dos vendedores e se certificar que a tarefa ocorreu. Podemos ver alguns exemplos de como a efetividade funciona na Tabela 4.

A Tabela 5 mostra exemplos dos critérios de efetivação para algumas das tarefas do projeto. Para a tarefa 1_1_3 o critério de efetivação é verificar se há uma nota fiscal com a compra de algum SKU de NAB. Isso indica que a tarefa foi cumprida, pois o objetivo era auxiliar o ponto de venda (PDV) a realizar a cobertura de produtos NAB, vale destacar que NAB é a categoria de produtos não alcoólicos, como refrigerante.

Já para a tarefa 3_1_1 de texto "Execute e tire uma foto do cartazete de Menu de Cerveja precificado.", temos que o critério de efetivação envolve o reconhecimento de imagem, ou seja, é necessário que a foto tirada do cartazete de Menu de Cerveja esteja de acordo com o modelo pré-estabelecido. Assim, é possível validar se a tarefa foi realizada corretamente.

Para a tarefa 5_1_1, o critério de efetivação é baseado no faturamento gerado através do aplicativo BEES. É necessário que haja um registro de faturamento maior que zero para validar que a tarefa foi concluída com sucesso, já que o intuito é realizar a instalação e a primeira compra pelo app.

Esses exemplos ilustram como os critérios de efetivação são utilizados para verificar se as tarefas foram devidamente cumpridas. Cada tarefa possui seu próprio critério, que pode envolver a verificação de notas fiscais, fotos, respostas de pesquisas ou outros dados relevantes. Esses critérios são fundamentais para garantir que as atividades propostas tenham sido realizadas corretamente e que os objetivos de cada tarefa tenham sido alcançados.

id_task_pool	texto_da_tarefa	critério_de_efetividade
1_1_3	PDV não comprador de NAB. Auxilie na cobertura de qualquer SKU de NAB.	Possuir qualquer nota fiscal contendo algum SKU de NAB.
3_1_1	Execute e tire uma foto do cartazete de Menu de Cerveja precificado.	Foto com reconhecimento de imagem do Menu de Cerveja.
4_2_5	Responda a pesquisa para mapear as geladeiras do PDV.	Resposta da pesquisa.
5_1_1	Apresente o app BEES ao cliente e ajude-o na instalação e primeira compra.	Faturamento Bees > 0
7_2_5	Auxilie o PDV a comprar [N] SKUs distintos de MARKETPLACE.	PDV realizar a compra de [N] SKUs de Marketplace

Tabela 5 – Exemplos de critério de efetividade por tarefa

3.2.3 Racional score primário e secundário

- Score primário: o score primário se divide em mínimo e máximo como forma de parametrizar a relevância entre os clusters primários de acordo com a estratégia comercial. É um input arbitrário que determina como ordenar os clusters.
- Score secundário: dentro de um cluster primário, existem clusters secundários e variações de tarefas que devem ser ordenadas de acordo com suas regras de score secundário. Cada cluster primário pode ter suas próprias regras de organização, mas todos variam entre 0 e 100%.
- Score final: o score final de cada tarefa é composto por um cálculo entre score primário e secundário de modo a organizar todas as tarefas de todos os clusters numa mesma lista ranqueada. O cálculo é descrito como:

A escala de score secundário é fixa entre 0 e 100, sendo assim $Score_{SecundarioMin} = 0$ e $Score_{SecundarioMax} = 100$, portanto, temos:

$$\frac{Score_{Primario} - Score_{PrimarioMin}}{Score_{PrimarioMax} - Score_{PrimarioMin}} = \frac{Score_{Secundario} - Score_{SecundarioMin}}{Score_{SecundarioMax} - Score_{SecundarioMin}} =$$

$$\frac{Score_{Secundario} - 0}{100 - 0} = \frac{Score_{Secundario}}{100}$$

Vale destacar que o $Score_{PrimarioMin}$, $Score_{PrimarioMax}$ e $Score_{Secundario}$ são definidos pelo time de negócios para cada tarefa, sendo assim basta substituir os valores na fórmula acima para projetar o $Score_{Secundario}$ na escala do $Score_{PrimarioMin}$, $Score_{PrimarioMax}$. Os intervalos de score primário tem como objetivo comparar a importância de cada cluster e a do score secundário é comparar quais tarefas são mais relevantes dentro do mesmo cluster. Tal metodologia foi criada pelo time de Tasks do Bees Force.

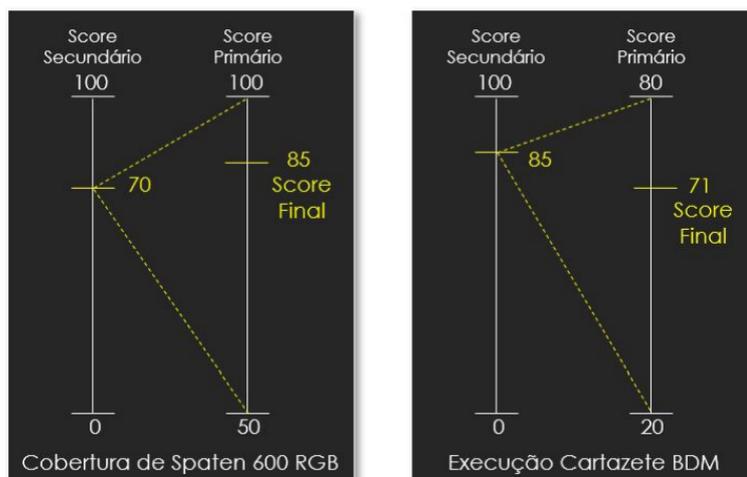


Figura 9 – Exemplo de cálculo do score final.

Na Figura 9, exemplificamos o cálculo do score final para as tarefas de cobertura da cerveja Spaten e execução de cartazete BDM. No contexto da cobertura, o cluster correspondente a essa tarefa possui um score primário variando entre 50 e 100, além de um score secundário de 70 dentro do próprio cluster. Combinando esses valores, chegamos a um score final de 85.

É importante ressaltar que, ao mencionarmos cobertura na Ambev, estamos nos referindo à venda de um novo produto que o cliente ainda não adquire regularmente.

3.3 *Task pool*

O *Task Pool* é um conjunto de tarefas que um RN pode receber na visita a cada um dos PDVs. Cada tarefa dentro desse *pool* tem um conjunto de critérios de elegibilidade para ser gerada e de efetividade que devem ser atendidos para que a tarefa seja considerada válida. Essas tarefas são baseadas em objetivos específicos de negócio, como a promoção de produtos, o aumento das vendas ou a fidelização de clientes.

Cada tarefa do *pool* é pontuada com um score, que é uma medida do seu valor para o negócio. Quanto maior o score, mais priorizada será a tarefa. Além disso, cada tarefa é acompanhada de um texto descritivo que explica o que precisa ser feito para executar a tarefa com sucesso.

As tarefas do *Task Pool* são uma ferramenta poderosa para aumentar a eficácia do RN, ajudando-os a priorizar suas atividades e maximizar o valor que geram para o negócio.

id_task_pool	score_final	critério_elegibilidade	critério_efetividade	texto	duracao
1_5_1	72,5	([SKU] ter sido colocado no carrinho entre 3 a 10 dias atrás) & (PDV não pode pertencer as franquias nacionais) & (faturamento L90D > 0) & (PDV não estar inadimplente (CDDs)) & (PDV não estar bloqueado) & (Volume [BRAND PACK SKU] L30D = 0) & ([SKU] ser crítico ou foco* Geo) & ([SKU] não ter nenhum pedido pendente)	Possuir qualquer nota fiscal contendo [SKU].	[SKU] adicionado ao carrinho recentemente. Ajude na venda.	210
5_1_1	95	(PDV sem cadastro no Bees)	Faturamento Bees > 0	Apresente o app BEES ao cliente e ajude-o na instalação e primeira compra.	90
7_1_1	90	(PDV não pode pertencer as franquias nacionais) & (Quantidade de SKUs de Marketplace L30D = 0) & (PDV não estar inadimplente (CDDs)) & (Faturamento L90D > 0) & (PDV não estar bloqueado)	PDV ter realizado compra de qualquer SKU de marketplace.	PDV não comprador de MARKETPLACE. Ajude na cobertura de qualquer SKU de MARKETPLACE.	180

Figura 10 – Exemplo de tarefas do *task pool*.

Na Figura 10, apresentamos três exemplos das mais de 80 tarefas disponíveis em nosso *task pool* atual. Cada tarefa é identificada por um *id_task_pool* único e possui um score final de 0 a 100, que é calculado a partir de um score primário e secundário e é utilizado para ranquear as tarefas.

Como descrito nas seções 3.2.1 e 3.2.2, cada tarefa possui critérios de elegibilidade e efetividade definidos. Além disso, cada tarefa tem uma duração em segundos e conta com um texto personalizado, como no exemplo da tarefa com o *id_task_pool* 1_5_1, que pode gerar uma tarefa para cada produto (SKU) que atenda aos critérios, por exemplo: "Budweiser 600 ml adicionado ao carrinho recentemente. Ajude na venda". Para efetivar a tarefa, basta que o RN venda o SKU mencionado para o PDV.

Suponhamos que um PDV tenha apenas essas três tarefas disponíveis para serem realizadas, e que tenha um tempo de visita de no máximo 5 minutos. Nesse caso, podemos ordenar as tarefas por score final e priorizar para a visita apenas as tarefas 5_1_1 e 7_1_1, dado que são as de maior score e cabem dentro do tempo disponível. A soma da duração dessas duas tarefas é de 4,5 minutos, sendo assim, não haveria tempo suficiente para realizar a tarefa 1_5_1, que demora cerca de 3,5 minutos.



Figura 11 – Visão geral de todos os clusters de tarefas

Na Figura 11, podemos observar os 7 clusters primários de tarefas. É importante ressaltar que esses clusters representam categorias definidas pelo time de negócios para segmentar as tarefas em diferentes assuntos. Eles são numerados de 1 a 7, respectivamente.

Além disso, os clusters secundários também recebem numeração. Por exemplo, dentro do cluster primário de Marketplace, temos o cluster secundário de Adoção. Portanto, quando mencionamos a tarefa de Marketplace 7_1_1, estamos nos referindo a uma tarefa do cluster primário 7 e do cluster secundário de adoção, que possui o índice 1. O terceiro número varia de 1 a N, representando uma das várias tarefas presentes nesse cluster secundário.

3.4 Modelagem proposta

Utilizamos um algoritmo de filtro colaborativo chamado *Alternating Least Squares* (ALS) que utiliza fatoração de matriz para resolver problemas em dados esparsos. Nossa base de treino utilizou as tarefas geradas para os meses de agosto à dezembro de 2022. E a validação foi realizada em tarefas geradas para o mês de janeiro de 2023.

A escolha do método ALS para a resolução do problema em questão foi baseada em suas notáveis vantagens. O ALS é um algoritmo eficiente e escalável, especialmente em ambientes

distribuídos, permitindo a paralelização do processo de otimização. Além disso, sua abordagem iterativa de otimização alternada entre usuários e itens oferece uma maneira elegante de lidar com a esparsidade dos dados e a falta de informações em alguns casos. A simplicidade do ALS torna sua implementação relativamente fácil, e sua capacidade de lidar com grandes conjuntos de dados o tornou uma escolha ideal para o nosso desafio de modelagem.

Para realizar a modelagem utilizamos a biblioteca *Machine Learning Library* (MLlib) disponível no *pyspark*. Vale destacar que tal modelagem foi feita por UNB. No contexto da Ambev, UNB significa "Unidade de Negócios". As UNBs são responsáveis por gerenciar e operar negócios específicos dentro da empresa, como marcas ou linhas de produtos específicos, em determinadas regiões geográficas. Cada UNB tem seu próprio conjunto de metas e objetivos de negócios, e é liderada por uma equipe de gestão dedicada. A Ambev possui cerca de 250 UNBs em todo o Brasil para gerenciar seus negócios de bebidas alcoólicas e não alcoólicas em diferentes mercados e regiões geográficas. Sendo assim cada UNB terá um modelo específico para prever o seu contexto.

Essa abordagem de um modelo por UNB foi adotada após constatarmos que a redução do contexto do filtro colaborativo apresentou resultados superiores ao uso de apenas um modelo para toda a população. Isso se deve ao fato de que cada PDV pertencente à mesma UNB compartilha características e comportamentos semelhantes. Ressaltamos que a decisão de ter essa granularidade na modelagem se deu pelo fato de que uma redução excessiva do contexto não proporcionou bons resultados nos testes quando criamos por exemplo um modelo por UNB e canal, observamos que algumas combinações de UNB e canal possuem poucas observações, tornando a modelagem inviável. Vale destacar que canal é referente ao segmento do PDV, que pode ser mercado (ASR), bar (ROTA), sub distribuidora (SUB), etc.

unb_pdv	id_task_pool	efetivas	total_tarefas	perc_efetivas	variavel_final
120006_10028	1_4_1	4	4	100%	4,00
970_99999	6_2_7	0	0	0%	0,00
970_99972	1_4_1	9	12	75%	6,75
...
970_99994	3_9_2	2	3	66%	1,32
120006_10028	3_7_1	6	8	75%	4,50
970_99954	5_2_4	5	15	30%	0

Tabela 6 – Dados de entrada para o modelo ALS.

Os dados de entrada dessa modelagem tem o formato apresentado na Tabela 6. Para cada cliente temos um identificador único chamado de `unb_pdv`, e para cada tarefa existente no *pool* temos o total de vezes que a mesma foi efetiva no período, o total de tarefas priorizadas no período e realizando a divisão dessas colunas temos o percentual de conversão.

Portanto, por exemplo o unb_pdv 970_99972, efetivou a tarefa de id_task_pool 1_4_1 relacionada a venda de um produto 9 vezes durante o período dos dados de treinamento, mas essa tarefa subiu 12 vezes para visitas do RN a esse PDV, portanto seu percentual de efetivas, ou seja, sua conversão nessa tarefa é de 75%.

Para calcular a variável final que é utilizada como input da matriz de confiança do filtro colaborativo utilizamos a fórmula apresentada abaixo:

$$\text{Varivel final} = \text{Total de efetivas} * \frac{\text{Total de efetivas}}{\text{Total de priorizadas}} \quad (3.1)$$

É realizada a multiplicação do total de tarefas efetivas pela variável percentual de efetividade, com o objetivo de aumentar a confiabilidade de tarefas que tenham uma quantidade maior de tarefas efetivas no período e ponderar por uma taxa de conversão, dado que não podemos ter alta confiança de que uma tarefa de baixa conversão será efetivada.

Vale destacar que só consideramos para a matriz de confiança o valor da variável final para tarefas com pelo menos 65% de conversão dentro do pdv, valores abaixo disso foram zerados, com o intuito de ensinar ao modelo que temos baixa confiança na conversão de tais tarefas. Tal decisão foi inspirada na literatura e tomada após vários testes que foram realizados e que indicaram melhorias nos resultados.

unb_pdv	id_task_pool	rating	ranking_prediction
230_20817	4_2_1	1.13	1
230_20817	5_2_6	1.08	2
230_20817	5_2_2	0.95	3
...
230_20817	7_2_1	0.2	79
230_20817	3_3_5	0.15	80
230_20817	4_4_3	0.05	81

Tabela 7 – Tabela com resultados finais do modelo ALS.

Após todo o processo de fatoração de matrizes, como podemos ver na Tabela 7 o resultado final dessa modelagem traz um ranking de 1 a N para cada tarefa, sendo N a quantidade distinta de tarefas existentes no *pool*. Sendo assim, realiza a predição do ranking de cada tarefa para cada um dos PDVs. Podemos interpretar em tal resultado que os primeiros rankings são as tarefas mais prováveis a serem efetivas dentro do PDV.

O ranking é atribuído com base no *rating* calculado para cada tarefa. O *rating* é ordenado de forma decrescente em cada PDV, resultando em uma classificação final. Na análise dos resultados da modelagem para essas UNBs, decidimos utilizar as predições para priorizar as tarefas de acordo com o ranking. Atualmente, o score final varia de 0 a 100, onde os scores 99 e 100 são reservados para tarefas muito relevantes ao negócio, que são incluídas manualmente.

Dessa forma, distribuimos os scores entre os rankings a partir de 98. Por exemplo, uma tarefa com ranking 1 terá score final de 98, e a cada aumento no ranking, o score final é reduzido em uma unidade.

$$\text{Score final} = \max(99 - \text{ranking}, 0) \quad (3.2)$$

Assim, as tarefas com maior probabilidade de efetividade serão priorizadas nas visitas, seguindo o ranking previsto pelo modelo. O objetivo é garantir que as tarefas mais prováveis de serem efetivas sejam realizadas primeiro, a fim de aumentar sua chance de sucesso e para que permaneçam no histórico por mais tempo, a fim de serem executadas durante o mês.

3.5 Teste de hipótese

Para avaliar com rigor se a modelagem apresenta o resultado esperado decidimos por realizar um teste de hipótese. O teste de hipóteses é o processo utilizado para avaliar a força de uma evidência observada em uma amostra. O método serve para entender o quão confiável é extrapolar os resultados obtidos em uma amostra de estudo para toda a população da qual a amostra foi extraída. O investigador formula uma hipótese específica, avalia os dados da amostra e usa esses dados para decidir se eles suportam as hipóteses definidas.

3.5.1 Hipóteses e teste

Temos duas hipóteses para o teste, a primeira é chamada de hipótese nula (H_0), tal hipótese é confrontada pela que chamamos de hipótese alternativa (H_1):

- H_0 : Efetividade média dos RNs que recebem priorização personalizada é igual a efetividade média dos RNs que não recebem priorização personalizada.
- H_1 : Efetividade média dos RNs que recebem priorização personalizada é maior que a efetividade média dos RNs que não recebem priorização personalizada.

Para avaliar tais hipóteses utilizamos o teste da regressão logística. A ideia foi modelar a variável resposta efetividade pela variável do grupo ao qual pertence o PDV, sendo ele do grupo teste ou do controle e calcular o p-valor associado a variável explicativa para entender o impacto. A divisão dos grupos será discutida na próxima seção.

3.5.2 Seleção da base

Conforme mencionado na seção anterior a ideia do teste de hipóteses aplicado foi de separar duas amostras da nossa população seguindo alguns critérios de elegibilidade. Tivemos o grupo de teste ao qual foi aplicada a priorização personalizada de tarefas e o grupo de controle que é bem semelhante ao de teste mas que não sofreu nenhuma alteração.

Sendo assim, ao fim do estudo comparamos o desempenho da efetividade nos dois grupos para entender se a priorização apresentou o efeito positivo esperado na validação das tarefas. Vale destacar que a duração do teste foi de um mês e ocorreu durante março de 2023.

3.5.2.1 Critérios de elegibilidade

A modelagem foi feita por UNB com os meses de setembro de 2022 à fevereiro de 2023 para prever março. Com o objetivo de selecionar uma boa amostra para o estudo, definimos alguns critérios de elegibilidade tanto para a modelagem quanto para os RNs que foram selecionados.

Buscando ter um grupo ativo dentro do contexto de tarefas, realizamos alguns filtros. Para estarem elegíveis ao estudo os RNs, no período selecionado, deveriam:

- Atender de 70 à 250 PDVs;
- Ter efetividade média de 30% à 50%;
- Ter no mínimo 500 tarefas efetivas.

Pensando na modelagem outros critérios foram definidos:

- Foram considerados para modelagem PDVs com pelo menos 5 tarefas efetivas no período;
- Foram considerados para modelagem tarefas que tenham sido efetivadas pelo menos 60 vezes dentro do UNB no período;
- Foram considerados para modelagem UNBs que tenham no mínimo 9000 tarefas efetivas no período.

Dado os critérios acima ficaram de fora da modelagem apenas 10 (4%) das 249 UNBs para as quais geramos tarefas. Sendo elas: 46, 613673, 613770, 639257, 984507, 987280, 988677, 990582, 5000044 e 6655447. Além disso, não consideramos cerca de 1,5% das tarefas após remover as com poucas observações dentro de cada UNB.

Vale destacar que nenhuma comunicação foi realizada em relação ao teste à população de RNs, pois isso poderia enviesar o comportamento dos grupos em detrimento das conclusões do estudo.

3.5.3 Plano amostral

Para realizar o teste da regressão, segundo HSIEH; BLOCH; LARSEN (1998) a amostra necessária para obter as conclusões do teste foi calculada com a seguinte fórmula:

$$n = \frac{\left\{ Z_{1-\alpha/2} [P(1-P)/B]^{1/2} + Z_{1-\beta} [P1(1-P1) + P2(1-P2)(1-B)/B]^{1/2} \right\}^2}{[(P1 - P2)^2(1-B)]} \quad (3.3)$$

Sendo,

- p1: efetividade média esperada para o grupo controle.
- p2: efetividade média esperada para o grupo de teste.
- B: proporção da amostra com priorização personalizada.
- alfa: probabilidade do erro tipo I.
- poder: um menos a probabilidade de cometer o erro tipo II.

Definimos p1 como 50%, p2 como 52%, estimando uma melhora na efetividade ao aplicarmos a priorização personalizada. Com B = 0.5 para dividirmos a amostra igualmente nos dois grupos e os usuais alpha = 0.05 e poder = 0.9 para fixar baixas probabilidades de erros do teste. Dado os parâmetros definidos, a amostra calculada foi de 26.254 tarefas. Sendo 13.127 para cada grupo.

Com o objetivo de proporcionar ao RN a experiência completa de ter todos os PDVs de sua carteira dentro da priorização personalizada, decidimos separar os RNs em grupos de teste e controle, ou seja, todos os PDVs de um determinado RN vão pertencer a apenas um grupo.

Considerando boa margem de segurança, foram escolhidos 60 RNs, 30 para cada grupo em uma amostragem aleatória estratificada.

3.5.4 Teste da regressão logística

Nesta seção, descreveremos o teste de regressão logística utilizado para investigar se a priorização personalizada aplicada ao grupo de teste está de fato associado ao aumento da efetividade das tarefas. O objetivo principal deste teste é determinar se a variável regressora binária (denotada por X) relacionada ao do grupo teste ou controle tem algum efeito significativo na variável de resposta binária (denotada por Y), que indica se a tarefa foi efetiva ($Y = 1$) ou não ($Y = 0$).

Para esta análise, utilizamos o teste de hipótese da regressão com base na teoria dos modelos lineares generalizados (GLM) com função de ligação logit. O modelo de regressão utilizado foi formulado da seguinte maneira:

$$\text{logit}(Y) = \beta_0 + \beta_1 X + \varepsilon \quad (3.4)$$

Onde:

- Y é a variável de resposta binária, representando a efetividade da tarefa no PDV (1 se efetiva, 0 se não efetiva).
- X é a variável regressora binária que indica se o PDV pertence ao grupo teste ou controle (1 se teste, 0 se controle).
- β_0 é o intercepto do modelo, representando o sucesso da tarefa no grupo controle (quando $X = 0$).
- β_1 é o coeficiente de regressão associado à variável binária X , representando a diferença esperada na probabilidade de sucesso da tarefa entre o grupo teste e o grupo controle.
- ε é o termo de erro, que captura a variabilidade não explicada pelo modelo.

O teste de hipótese que realizamos tem o objetivo de verificar se o coeficiente β_1 é igual a zero ou não. A hipótese nula (H_0) afirma que não há diferença significativa entre a efetividade da tarefa nos grupos teste e controle, ou seja, $\beta_1 = 0$. A hipótese alternativa (H_1), por sua vez, indica que existe uma diferença significativa entre os grupos, e, portanto, $\beta_1 \neq 0$.

Para realizar o teste de hipótese, utilizamos o teste t sobre o coeficiente β_1 , levando em conta o erro padrão e o tamanho da amostra. Com base na significância estatística do teste (usualmente adotando um nível de significância α de 0,05), podemos decidir entre rejeitar a hipótese nula em favor da hipótese alternativa ou falhar em rejeitar a hipótese nula.

Os resultados do teste de regressão nos fornecem insights importantes sobre a relação entre o tratamento aplicado (representado por X) e a efetividade da tarefa (representada por Y). Se o teste indicar que β_1 é estatisticamente diferente de zero, podemos inferir que o tratamento teve um impacto significativo na efetividade da tarefa. Por outro lado, se β_1 não for estatisticamente diferente de zero, não podemos afirmar que o tratamento teve efeito significativo.

Assim, o teste de regressão nos permite compreender melhor a relação entre as variáveis binárias envolvidas, bem como determinar a eficácia do tratamento aplicado em relação à priorização em questão.

3.5.5 Estatísticas das amostras

Após aplicarmos todos os critérios de elegibilidade, nossa base de RNs foi reduzida a 2616. Dessa forma, procedemos com uma amostragem estratificada, considerando a principal geografia e o canal de atendimento do RN. É importante mencionar que temos 7 geos no Brasil, que correspondem a regiões definidas pela Ambev para dividir o território brasileiro. Essas regiões são similares às divisões geográficas do país e incluem: Minas (MG), Nordeste (NE), Norte (NO), Centro-Oeste (CO), Rio de Janeiro e Espírito Santo (RJ/ES), São Paulo (SP) e Sul (SUL).

Optamos por extratificar por essa variável para garantir a representatividade de todo o país na amostra. Além disso, devido à variedade de tipos de estabelecimentos, segmentamos os PDVs em categorias como ASR, FER, HIPERSEG, LC, ROTA, SUB, TRAD, VIP e OUTROS. Por exemplo, o canal SUB corresponde a sub distribuidoras, enquanto VIP representa bares mais premium, e ASR inclui mercados de pequeno porte.

Essa abordagem nos permite formar dois grupos similares em termos de canais de atendimento e regiões, garantindo assim uma representação abrangente e confiável da população.

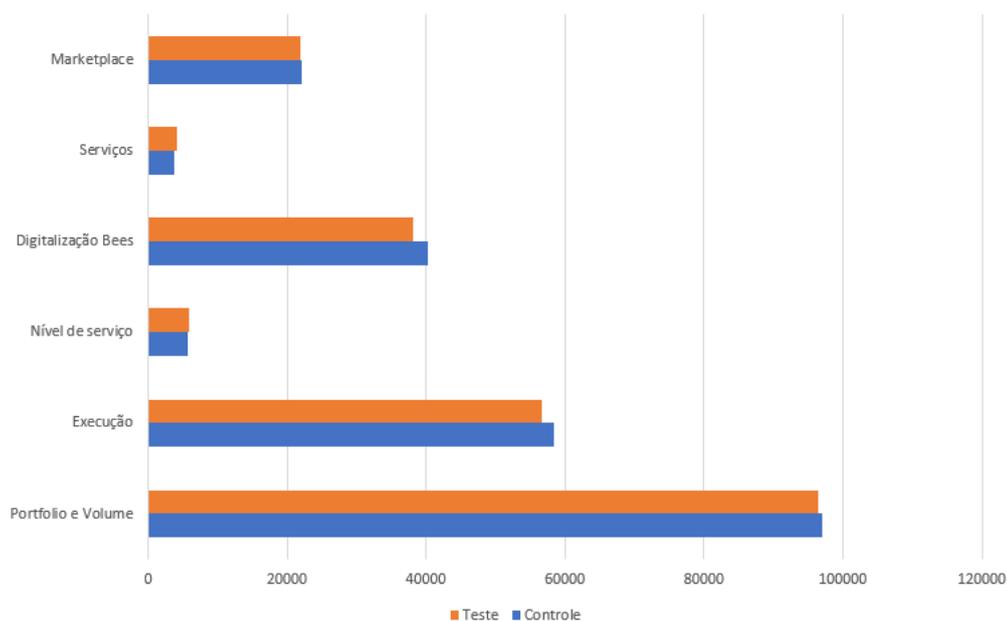


Figura 12 – Gráfico do total de tarefas por cluster em cada grupo selecionado.

Tal seleção resultou em 60 RNs divididos em 46 UNBs das cerca de 250 UNBs existentes, sendo que destas, 26 são revendas. Os grupos apresentaram comportamento semelhante em relação a efetividade, tendo o de controle 97.639 tarefas efetivas no período e o de teste 97.093. Com um percentual de efetividade de 38.88% e 38.40%, respectivamente.

Os 30 RNs do grupo de teste atendiam cerca de 4961 PDVs, já os de controle aproximadamente 4780. Nas Figuras 12, 13 e 14, podemos ver que as distribuições entre clusters de

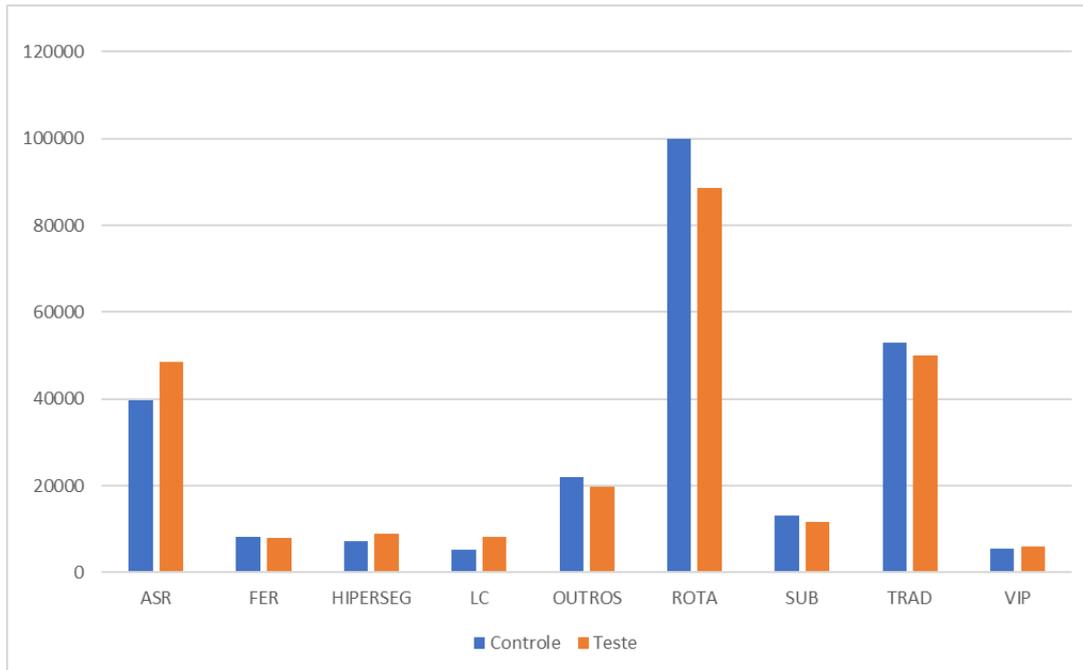


Figura 13 – Gráfico do total de PDVs por canal em cada grupo selecionado.

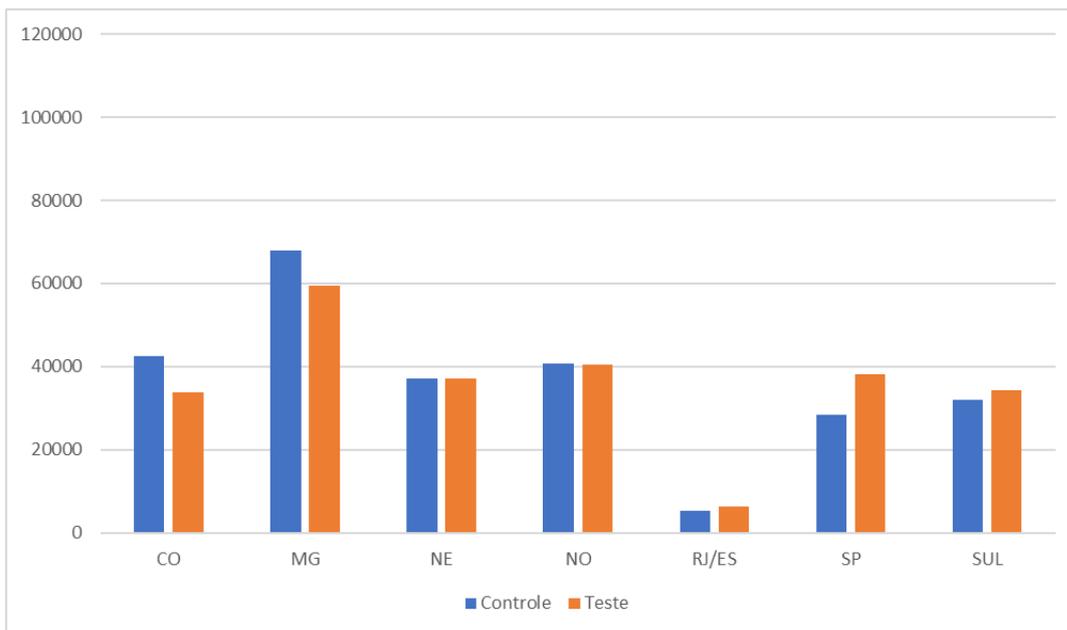


Figura 14 – Gráfico do total de PDVs por geo em cada grupo selecionado.

tarefas, geo e canais dos PDVs também era semelhante entre os grupos. Sendo assim, temos fortes evidências de que as amostras foram selecionadas de forma satisfatória para o estudo.

RESULTADOS

4.1 Experimento

Realizamos um estudo onde selecionamos 25 UNBs dentre as cerca de 250 disponíveis. A seleção foi feita devido ao alto custo computacional que seria necessário para trabalhar com toda a população de clientes. Essas 25 UNBs apresentam um total de 200.709 pontos de venda, e consideramos um total de 6.483.598 tarefas para o estudo. Essas tarefas foram usadas para treinar o modelo e realizar previsões para os dados de janeiro de 2023. Ajustamos a modelagem para todos os tipos de tarefas presentes em nosso conjunto de dados. O objetivo deste experimento foi validar o modelo em dados históricos, a fim de assegurar a qualidade da modelagem antes de realizar o piloto com a amostra de 60 RNs presentes em 46 UNBs do Brasil.

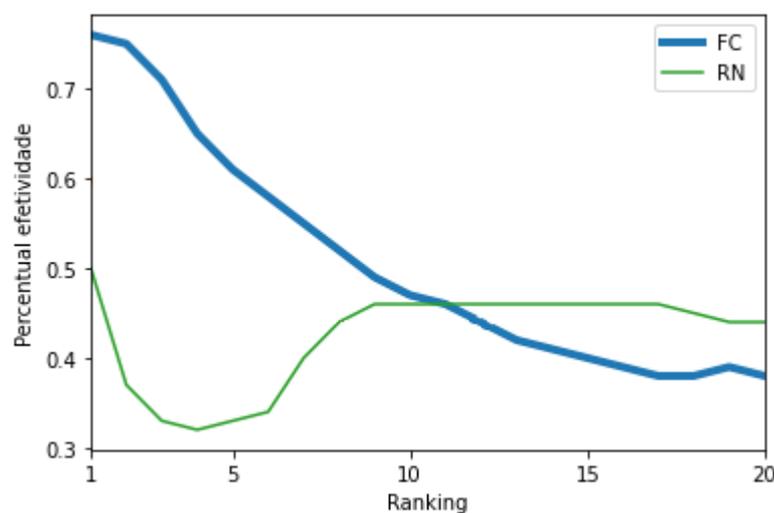


Figura 15 – Gráfico da efetividade em relação ao ranking da predição.

Na Figura 15 podemos ver o percentual de efetividade das tarefas em cada uma das posições do ranking predito para o filtro colaborativo ajustado em comparação com a priorização

realizada pela regra de negócio usada atualmente. Para o modelo as tarefas que estão nos rankings de 1 à 5 apresentam efetividade média que varia entre 60% e 80%, já a regra de negócio entre 30% e 50%. Vale destacar que o cálculo foi realizado com base em todas as tarefas priorizadas após realizar alguns filtros, como o de pegar apenas tarefas criadas pelo nosso *pool* que o RN tenha interagido, ou seja, desconsiderou tarefas que subiram para visitas em que o RN não interagiu com as tarefas.

Além disso, podemos ver que o percentual de validação diminui ao longo do aumento do ranking do filtro colaborativo, sendo assim temos evidências de que o modelo é capaz de captar bem a probabilidade de determinada tarefa ser efetiva dado seu ranking dentro do PDV e que é superior a priorização por regra de negócio que até o ranking 10 tem desempenho bem inferior a modelagem.

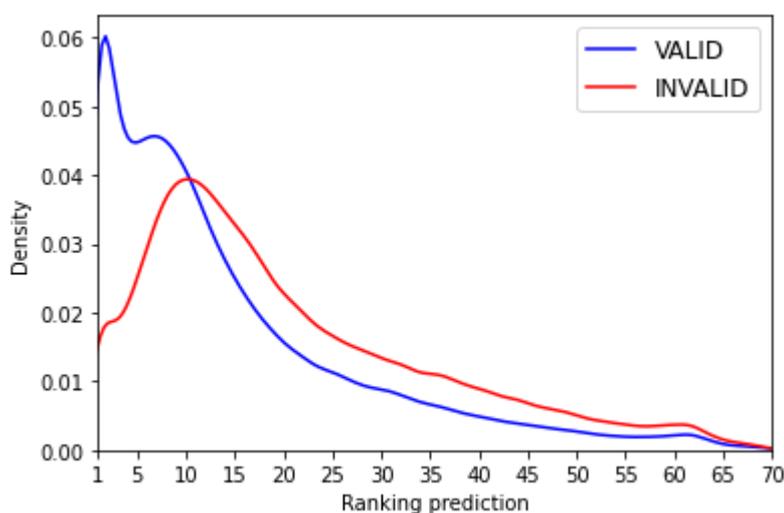


Figura 16 – Gráfico da distribuição em relação ao ranking e a efetivação das tarefas.

Na Figura 16 vemos o mesmo comportamento mas agora analisando a quantidade de tarefas válidas e não válidas. Até o ranking 10 vemos uma presença maior de tarefas válidas e após esse corte vemos que o total de tarefas não válidas é maior. Vale destacar que a quantidade diminui para ambos os resultados de validação com o aumento do ranking o que sugere que os últimos rankings normalmente são tarefas não elegíveis aos PDVs, o que é outro indício de que o modelo captou bem a probabilidade de efetivação de uma tarefa dado o contexto do PDV.

4.2 Piloto

Nesta seção, apresentaremos os resultados da abordagem proposta para modelar a probabilidade de conversão, que é um dos pilares da nova forma de priorização de tarefas denominada Score 2.0. Os resultados do piloto realizado em março de 2023 com o grupo de teste do filtro colaborativo e o grupo controle foram analisados e comparados entre si, levando em consideração a efetividade geral e os resultados do teste de hipótese.

No total, tivemos no período 16.252 tarefas válidas para o filtro colaborativo e 14.070 para o grupo controle. A tabela abaixo apresenta a efetividade geral de cada abordagem:

Abordagem	Efetividade Geral
Filtro Colaborativo	54,34%
Grupo Controle	52,61%

Tabela 8 – Comparação da efetividade geral das duas abordagens.

Os resultados da tabela 8 indicam que a efetividade geral do filtro colaborativo foi superior à do grupo controle.

Foi realizado um teste de hipótese da regressão logística para avaliar a significância estatística das duas abordagens. O resultado mostrou que a hipótese nula foi rejeitada, indicando que a abordagem foi estatisticamente significativa. O p-valor para o filtro colaborativo foi de $1.674733e-07$ e o coeficiente para o teste de hipótese da regressão logística foi de 0,08.

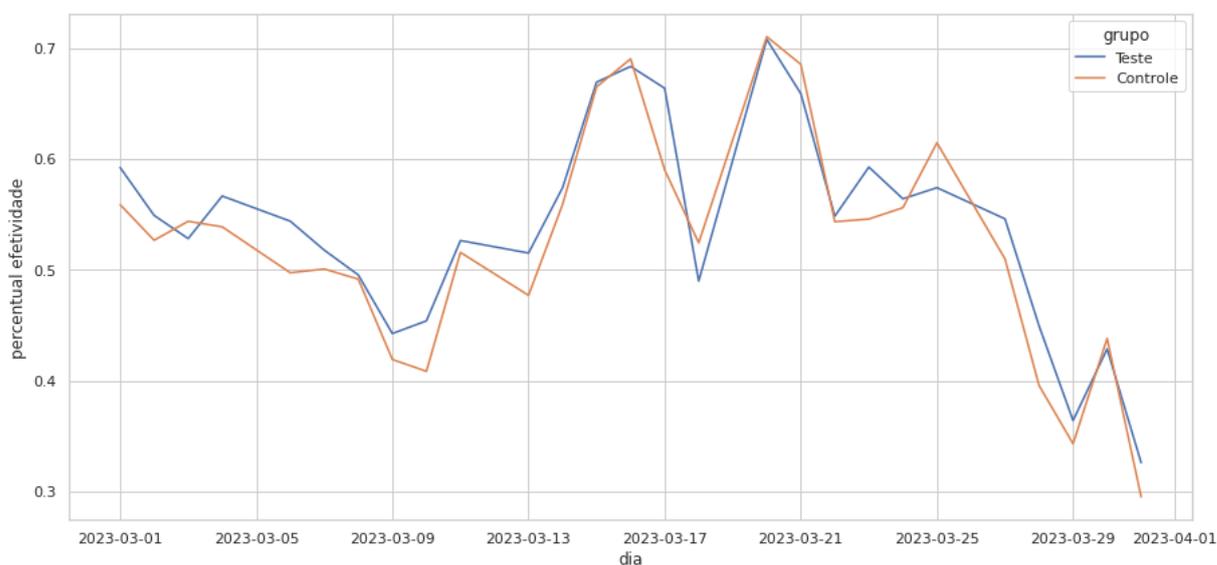


Figura 17 – Comparação do percentual de efetividade dos dois grupos por dia.

Na Figura 17 temos evidências de que em geral a modelagem proposta performou melhor do que a regra de negócio principalmente nas primeiras duas semanas. Essa diferença pode ser explicada pelo fato de que, nesse período, o modelo prioriza as tarefas com maior probabilidade de conversão. Consequentemente, é natural esperar uma taxa de conversão mais elevada nas primeiras semanas. Vale destacar também o fato de que essas tarefas ficam mais tempo no histórico para serem realizadas pelo RN ao longo do mês, o histórico de tarefas permite que o RN consulte e execute uma tarefa que não foi realizada em uma visita passada naquele mês, o que contribui para o aumento geral das conversões dessas tarefas que sobem nas duas primeiras semanas.

É importante destacar que os resultados do grupo de teste, embora superiores aos do grupo controle, não atingiram o mesmo nível de excelência apresentado na Figura 15. Isso se deve à consideração de um fator crucial: a meta do RN. Cada RN possui uma meta específica para a efetivação de tarefas, que influencia diretamente sua remuneração. Embora as análises históricas indiquem uma conversão significativamente alta, na prática, o RN se esforça para alcançar sua meta, que geralmente varia entre 30% e 50%. Por esse motivo, é compreensível que, mesmo com tarefas mais fáceis de serem convertidas ocorrendo nas visitas, a taxa de conversão se assemelhe ao grupo de controle. O RN não tem muitos incentivos para realizar muitas mais tarefas do que o estabelecido por sua meta percentual. Como resultado, a diferença tende a ser semelhante.

Em resumo, os resultados indicam que uso do filtro colaborativo trouxe um ganho de efetividade geral em relação ao grupo controle. Esses resultados podem contribuir para aprimorar a efetividade das visitas aos PDVs, otimizando o tempo dos representantes de vendas e melhorando assim os resultados de vendas e satisfação dos clientes.

CONCLUSÃO

A Ambev está empenhada em inovar seus processos de vendas para manter sua liderança e competitividade no mercado. Com o lançamento do aplicativo BEES e o projeto *Tasks do Bees Force*, a empresa conseguiu facilitar a vida dos clientes e automatizar parte do processo de vendas. Entretanto, a definição de score para priorização das tarefas apresentava limitações, o que levou à criação do conceito do Score 2.0. Esse modelo utiliza um filtro colaborativo para prever a chance de conversão das tarefas e definir o ranking de cada tarefa para cada ponto de venda, priorizando as tarefas com maior probabilidade de sucesso. O estudo realizado mostrou que o modelo performou melhor do que a regra de negócio, especialmente nas primeiras semanas, e portanto pode contribuir para otimizar o tempo dos representantes de negócio, melhorando os resultados de vendas e a satisfação dos clientes. Assim, a Ambev continua investindo em tecnologia e inovação para se tornar uma plataforma de marcas e aprimorar sua eficiência no mercado.

REFERÊNCIAS

- BOBADILLA, J.; ORTEGA, F.; HERNANDO, A.; GUTIÉRREZ, A. Recommender systems survey. **Knowledge-based systems**, Elsevier, v. 46, p. 109–132, 2013. Citado na página 15.
- GREGORINI, G. d. S. *et al.* Estratégia competitiva no mercado de bebidas: estudo de caso na companhia de bebidas das américas–ambev. 2014. Citado na página 11.
- HSIEH, F. Y.; BLOCH, D. A.; LARSEN, M. D. A simple method of sample size calculation for linear and logistic regression. **Statistics in medicine**, Wiley Online Library, v. 17, n. 14, p. 1623–1634, 1998. Citado na página 37.
- HU, Y.; KOREN, Y.; VOLINSKY, C. Collaborative filtering for implicit feedback datasets. In: IEEE. **2008 Eighth IEEE international conference on data mining**. [S.l.], 2008. p. 263–272. Citado nas páginas 18 e 20.
- ISINKAYE, F. O.; FOLAJIMI, Y. O.; OJOKOH, B. A. Recommendation systems: Principles, methods and evaluation. **Egyptian informatics journal**, Elsevier, v. 16, n. 3, p. 261–273, 2015. Citado na página 15.
- KO, H.; LEE, S.; PARK, Y.; CHOI, A. A survey of recommendation systems: recommendation models, techniques, and application fields. **Electronics**, MDPI, v. 11, n. 1, p. 141, 2022. Citado na página 16.
- LOPS, P.; GEMMIS, M. d.; SEMERARO, G. Content-based recommender systems: State of the art and trends. **Recommender systems handbook**, Springer, p. 73–105, 2011. Citado na página 21.
- LU, J.; WU, D.; MAO, M.; WANG, W.; ZHANG, G. Recommender system application developments: a survey. **Decision Support Systems**, Elsevier, v. 74, p. 12–32, 2015. Citado na página 15.
- PAZZANI, M. J.; BILLSUS, D. Content-based recommendation systems. In: **The adaptive web**. [S.l.]: Springer, 2007. p. 325–341. Citado nas páginas 20 e 21.
- SCHAFER, J. B.; KONSTAN, J.; RIEDL, J. Recommender systems in e-commerce. In: **Proceedings of the 1st ACM conference on Electronic commerce**. [S.l.: s.n.], 1999. p. 158–166. Citado na página 15.

