

**Avaliação de Técnicas de Recomendação Baseada em
Sessão: Comparação entre Variações do S-KNN**

Cairo Carvalho Campos Faria

Trabalho de Conclusão de Curso
MBA em Inteligência Artificial e Big Data

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Avaliação de Técnicas de
Recomendação Baseada em Sessão:
Comparação entre Variações do S-
KNN

Cairo Carvalho Campos Faria

Cairo Carvalho Campos Faria

Avaliação de Técnicas de Recomendação Baseada em Sessão: Comparação entre Variações do S-KNN

Trabalho de conclusão de curso apresentado ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Prof. Dr. Marcelo Garcia Manzato

USP - São Carlos

2024

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

F224a Faria, Cairo Carvalho Campos
Avaliação de Técnicas de Recomendação Baseada em
Sessão: Comparação entre Variações do S-KNN / Cairo
Carvalho Campos Faria; orientador Marcelo Garcia
Manzato. -- São Carlos, 2024.
42 p.

Trabalho de conclusão de curso (MBA em
Inteligência Artificial e Big Data) -- Instituto de
Ciências Matemáticas e de Computação, Universidade
de São Paulo, 2024.

1. Sistemas de recomendação. 2. Recomendação
baseada em sessão. 3. S-KNN. I. Manzato, Marcelo
Garcia, orient. II. Título.

RESUMO

FARIA, C. C. C. **Avaliação de Técnicas de Recomendação Baseada em Sessão: Comparação entre Variações do S-KNN**. 2024. 52 f. Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

Com o crescimento exponencial da quantidade de informação disponível *on-line*, a personalização das recomendações se tornou um diferencial competitivo fundamental para serviços como *e-commerce*. Os sistemas de recomendação baseados em sessão surgem como solução eficaz para capturar as preferências recentes dos usuários, mesmo em cenários em que não há um histórico longo de interações. Nesse contexto, este trabalho busca avaliar métodos que possam lidar com grandes volumes de dados e fornecer recomendações precisas, oferecendo uma melhor experiência ao usuário. O trabalho tem como objetivo avaliar o desempenho de variações do método S-KNN, incluindo S-SKNN e SF-SKNN, na resolução do problema de recomendação do próximo item em uma base de dados de *e-commerce*. Busca, também, avaliar a influência que critérios de seleção de itens candidatos têm nos resultados. A metodologia envolveu a criação de quatro cenários, variando a quantidade máxima e a forma de seleção de itens candidatos, que poderiam ser escolhidos por popularidade ou aleatoriamente. Após a aplicação dos métodos nessas diferentes situações, foi feita a análise utilizando as métricas de acurácia MRR (*Mean Reciprocal Rank*) e HR (*hit rate*), além de métricas de qualidade, como cobertura de catálogo e média de popularidade. Os resultados indicam que o método SF-SKNN obteve o melhor desempenho geral, conseguindo equilibrar precisão e diversidade nas recomendações. Observou-se, também, melhoras nas métricas ao aumentar o limite de itens candidatos. Quando a seleção desses itens envolveu o viés de popularidade, houve aumento na acurácia assim como diminuição da diversidade de produtos recomendados. Com essas conclusões, este estudo oferece perspectivas pertinentes para o desenvolvimento de sistemas de recomendação em plataformas de *e-commerce*.

Palavras-chave: Sistemas de recomendação; Recomendação baseada em sessão; S-KNN.

ABSTRACT

FARIA, C. C. C. **Evaluation of Session-Based Recommendation Techniques: A Comparison of S-KNN Variations.** 2024. 52 f. Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

With the exponential growth of information available online, personalized recommendations have become a crucial competitive advantage for services such as e-commerce. Session-based recommendation systems have emerged as an effective solution to capture users' recent preferences, even in scenarios where there is no long history of interactions. In this context, this study aims to evaluate methods that can handle large volumes of data and provide accurate recommendations, enhancing the user experience. The goal of this work is to assess the performance of variations of the S-KNN method, including S-SKNN and SF-SKNN, in solving the next-item recommendation problem in an e-commerce dataset. It also seeks to evaluate the influence that candidate item selection criteria have on the results. The methodology involved creating four scenarios, varying the maximum number and selection method of candidate items, which could be chosen by popularity or randomly. After applying the methods in these different situations, an analysis was conducted using accuracy metrics such as MRR (Mean Reciprocal Rank) and HR (Hit Rate), in addition to quality metrics like catalog coverage and average popularity. The results indicate that the SF-SKNN method achieved the best overall performance, successfully balancing precision and diversity in recommendations. It was also observed that increasing the limit of candidate items improved the metrics. When the selection involved a popularity bias, there was an increase in accuracy as well as a reduction in the diversity of recommended products. With these conclusions, this study offers relevant insights for the development of recommendation systems on e-commerce platforms.

Keywords: Recommender systems; Session-based recommendation; S-KNN.

LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplo de Filtragem Colaborativa.....	25
Figura 2 – Gráficos de radar comparando S-KNN, S-SKNN e SF-SKNN em relação as métricas MRR, <i>hit rate</i> (HR), cobertura de catálogo (COV) e popularidade média (POP), com um gráfico para cada um dos quatro cenários.....	35
Figura 2 – Gráficos de radar comparando S-KNN, S-SKNN e SF-SKNN em relação aos resultados obtidos em cada cenário, A B, C e D, com um gráfico para cada métrica.....	36

LISTA DE TABELAS

Tabela 1 – Características dos conjuntos de treinamento e teste.....	31
Tabela 2 – Resultados das métricas MRR, hit rate (HR), cobertura de catálogo (COV) e popularidade média (POP) para os modelos S-KNN, S-SKNN e SF-SKNN nos diferentes cenários testados.....	34

LISTA DE QUADROS

Quadro 1 – Resumo de Métricas de Similaridade.....	23
Quadro 2 – Cenários Avaliados no Experimento.....	32

LISTA DE ABREVIATURAS E SIGLAS

COV – Cobertura de Catálogo

HR–Hit Rate

KNN –K-Nearest Neighbors

MRR–Mean Reciprocal Rank

POP–Popularidade Média

SF-SKNN–Session-based Filtered K-Nearest Neighbors

S-KNN–Session-based K-Nearest Neighbors

SRBS–Sistemas de Recomendação Baseados em Sessão

S-SKNN–Session-based Sequential K-Nearest Neighbors

SUMÁRIO

1 INTRODUÇÃO	20
1.1 Objetivos	20
1.2 Organização do Trabalho	21
2 FUNDAMENTAÇÃO TEÓRICA	22
2.1 Aprendizado de Máquina	22
2.2 Medidas de Similaridade	22
2.3 Pré-Processamento	24
2.4 Sistemas de Recomendação	24
2.4.1 Filtragem Colaborativa	25
2.4.2 Recomendação Baseada em Sessão	25
2.5 Métricas de Avaliação	28
3 METODOLOGIA	29
3.1 Base de Dados	29
3.2 Pré-Processamento	30
3.3 Divisão entre Treinamento e Teste	30
3.4 Seleção de Itens Candidatos	31
3.5 Avaliação	32
4 ANÁLISE DOS RESULTADOS	34
4.1 Métodos de Recomendação	36
4.2 Quantidade de Itens Candidatos	37
4.3 Viés de Popularidade na Escolha de Itens Candidatos	37
4.4 Considerações Finais	38
5 CONCLUSÃO	39
REFERÊNCIAS	41

1 INTRODUÇÃO

Nas últimas duas décadas, tem-se observado um crescimento sem precedentes na quantidade de dados produzidos (CLISSA; LASSNIG; RINALDI, 2023). Diante desse contexto de sobrecarga de informação *on-line*, os sistemas de recomendação se tornam essenciais, oferecendo sugestões personalizadas para usuários em diferentes tipos de serviços, desde plataformas de *streaming* até *e-commerce* (FAYYAZ et al., 2020).

A maior parte dos sistemas de recomendação depende de um longo histórico de interações do usuário e enfrenta dificuldades quando essas informações não estão disponíveis. Esse é, frequentemente, o caso em plataformas de *e-commerce*, nas quais o usuário geralmente não precisa estar logado para navegar. Destaca-se, então, a recomendação baseada em sessão, que foca nos itens visualizados na sessão atual, buscando utilizar as preferências recentes do usuário (WANG et al, 2021).

Múltiplas técnicas podem ser aplicadas na recomendação baseada em sessão, abrangendo desde abordagens mais simples, como métodos heurísticos, até modelos mais avançados, como redes neurais, conforme demonstrado por Ludewig e Jannach (2018). No trabalho, os autores concluíram que abordagens com foco em vizinhança (K-Nearest Neighbors) apresentam resultados comparáveis aos de alternativas mais complexas.

Diante deste cenário, em que a variedade de produtos disponíveis é cada vez maior e não se tem conhecimento prévio das preferências do usuário, o presente trabalho busca analisar soluções para a recomendação baseada em sessão em plataformas de *e-commerce*. Para fazer isso de maneira prática e eficiente, optou-se por utilizar as técnicas baseadas em vizinhança.

1.1 Objetivos

Este projeto tem como objetivo avaliar o desempenho de variações de modelos de recomendação baseados em sessão, com foco em métodos que utilizam vizinhança aplicados ao problema da recomendação do próximo item em uma base dados de *e-commerce*. São analisados os métodos S-KNN (Session-based K-Nearest Neighbors), S-SKNN (Sequential Session-based K-Nearest Neighbors) e SF-SKNN (Sequential Filter Session-based K-Nearest Neighbors), conforme apresentados por Ludewig e Jannach (2018).

Como objetivos específicos, procura-se avaliar o efeito que critérios de seleção de itens candidatos podem ter nos resultados. São observados os efeitos da variação de quantidade e do viés de popularidade na seleção dos itens candidatos.

1.2 Organização do Trabalho

Além do Capítulo 1, contendo a introdução, o trabalho possui outros quatro capítulos. No Capítulo 2, é feita a fundamentação teórica, apresentando conceitos sobre aprendizado de máquina, métricas de similaridade e pré-processamento, além de sistemas de recomendação e suas variações, filtragem colaborativa e sistemas de recomendação baseados em sessão, finalizando com métricas de avaliação. No Capítulo 3, é apresentada a metodologia utilizada no experimento, com destaque para a descrição da base de dados, métodos de pré-processamento, seleção de itens candidatos e técnicas de avaliação. No Capítulo 4, são feitas análises comparativas dos resultados gerados pelo experimento. Por fim, o Capítulo 5 apresenta a conclusão e sugestão para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Aprendizado de Máquina

O *machine learning*, ou aprendizado de máquina, pode ser definido como um conjunto de ferramentas que detecta padrões em dados e os utiliza em previsões e tomadas de decisões. Essa tecnologia tem sido utilizada em diversas áreas, como o *e-commerce*, com a recomendação de produtos para o usuário; em sistemas bancários, detectando possíveis fraudes e tentativas de golpe; e na medicina, auxiliando diagnósticos médicos através dos dados clínicos disponíveis (MURPHY, 2012).

Segundo Murphy (2012), há três principais abordagens no aprendizado de máquina: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço. Cada uma dessas abordagens possui características distintas e pode ser empregada na resolução de diferentes tipos de problemas.

- **Aprendizado supervisionado:** Nesta abordagem, o algoritmo aprende a partir de um conjunto de dados rotulados, denominado conjunto de treinamento, em que cada exemplo de entrada é associado a uma resposta esperada. Essa forma de aprendizado tem como objetivo desenvolver uma função que mapeie o caminho entre entradas e saídas, permitindo a realização de previsões a partir de novos dados de entrada.
- **Aprendizado não supervisionado:** Diferentemente do aprendizado supervisionado, o algoritmo recebe apenas entradas e deve encontrar padrões ocultos nos dados sem o auxílio de um rótulo ou saída pré-definida.
- **Aprendizado por reforço:** Esse formato de aprendizagem envolve algoritmos que buscam maximizar as recompensas ao longo do tempo. Suas ações resultam em um feedback recompensador ou punitivo e seu comportamento é ajustado para otimizar o resultado a longo prazo.

2.2 Medidas de Similaridade

Uma abordagem empregada tanto no aprendizado supervisionado quanto no não supervisionado é a análise de similaridade entre os dados, aplicada para identificar padrões ou agrupar itens semelhantes. Diversas métricas de distância podem ser utilizadas para calcular essa similaridade ou dissimilaridade entre os dados. Segundo Sondur, Chigadani e Nayak (2016), a distância euclidiana, por exemplo, mede a proximidade geométrica entre dois pontos e é mais útil quando os dados estão na mesma escala. A similaridade do cosseno, por sua vez,

avalia a similaridade entre dois vetores com base no ângulo entre eles, sendo apropriada para grandes volumes de dados esparsos, como em análises de texto. A correlação de Pearson é usada para medir o grau da correlação linear entre duas variáveis. Por fim, o coeficiente de Jaccard é ideal para dados binários, medindo a proporção de atributos em comum entre dois conjuntos. No quadro a seguir, temos a comparação entre essas medidas.

Quadro 1 – Resumo de Métricas de Similaridade

Métrica de Similaridade	Descrição	Vantagens	Desvantagens
Distância Euclidiana	Calcula a raiz quadrada da soma das diferenças quadradas entre duas instâncias. Utilizada principalmente para dados numéricos.	Fácil de interpretar; comum em várias aplicações de aprendizado supervisionado e não supervisionado.	Sensível a variações de escala.
Similaridade do Cosseno	Mede o cosseno do ângulo entre dois vetores em um espaço multidimensional.	Ideal para grandes volumes de dados esparsos, como análise de texto.	Não leva em consideração as diferenças de magnitude entre os vetores
Correlação de Pearson	Mede o grau e correlação linear entre duas variáveis. Gera valores entre -1 e 1, sendo 1 uma correlação perfeita, -1 correlação inversa e 0 sem correlação.	Detecta correlações lineares; útil quando os dados têm variação comum.	Não captura relações não lineares.

Coeficiente de Jaccard	Mede a similaridade entre conjuntos. Calcula a interseção dividida pela união.	Ideal para dados binários ou categóricos; muito utilizado em sistemas de recomendação baseados em conjuntos.	Não lida bem com dados contínuos; não considera magnitude de diferenças ou semelhanças.
------------------------	--	--	---

Fonte: Elaborado pelo autor

2.3 Pré-Processamento

O pré-processamento de dados é uma etapa fundamental no aprendizado de máquina, pois garante que eles estejam em formato adequado para serem utilizados pelos algoritmos. De acordo com Pang-Ning, Steinbach e Kumar (2006), essa fase envolve uma série de métodos que podem incluir: limpeza de dados, tratando valores ausentes e inconsistências; integração de dados, que combina bases de diferentes fontes; transformação, que pode conter normalização e agregação; redução de dimensionalidade, técnica utilizada para diminuir a complexidade dos dados, removendo variáveis irrelevantes ou redundantes; discretização, que transforma dados contínuos em discretos.

2.4 Sistemas de Recomendação

Os sistemas de recomendação estão diretamente ligados ao aprendizado de máquina. Eles utilizam técnicas dessa área para prever as preferências do usuário com base em interações anteriores, sejam essas interações visualizações, compras ou classificações. O desenvolvimento dessa tecnologia foi impulsionado pela expansão de transações comerciais por meio da internet, como em plataformas de *e-commerce* e serviços de *streaming*. Nesse contexto, a capacidade de recomendar itens relevantes melhora a experiência do cliente, além de contribuir para o aumento de vendas e tempo gasto nas plataformas. Sendo assim, ao utilizar o histórico de interações entre usuários e itens, os sistemas de recomendação se tornam uma importante ferramenta de personalização da experiência do usuário, agregando maior valor as plataformas que oferecem esse serviço (AGGARWAL, 2016).

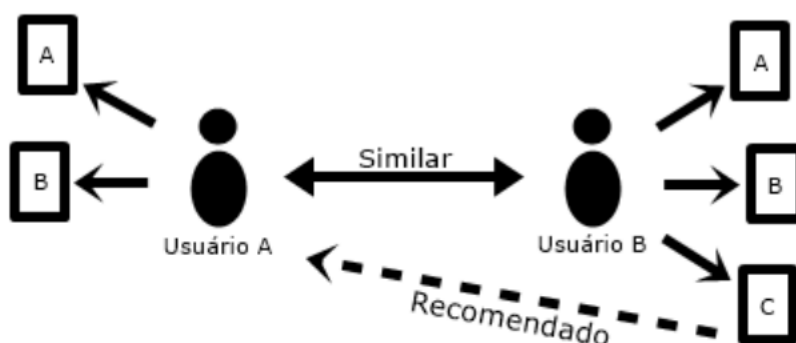
Esses sistemas utilizam diversas técnicas para cumprir sua função de identificar e recomendar itens relevantes para os usuários. De acordo com Ricci et al. (2016), entre as principais técnicas empregadas pelos sistemas de recomendação estão: filtragem colaborativa, que utiliza o comportamento de outros usuários semelhantes para sugerir itens; recomendação

baseada em conteúdo, que utiliza metadados de itens e o perfil do usuário para fazer as sugestões; recomendação por conhecimento, que se baseia em conhecimentos específicos sobre como os itens satisfazem as necessidades do usuário; recomendação híbrida, que combina mais de uma das abordagens para gerar resultados mais precisos.

2.4.1 Filtragem Colaborativa

A filtragem colaborativa não exige compreensão ou conhecimento sobre o conteúdo do item, se baseando apenas na semelhança entre os perfis de usuários (REATEGUI; CAZELLA, 2005). A Figura 1 exemplifica esse conceito, mostrando que, como os usuários A e B acessaram os itens A e B, o item C, acessado pelo usuário B, é recomendado para o usuário A, com base na similaridade entre eles. Segundo Rolim et al. (2017), métricas como distância euclidiana ou similaridade de cosseno podem ser usadas para o cálculo dessa semelhança.

Figura 1 – Exemplo de Filtragem Colaborativa



Fonte: Rolim et al. (2017)

A filtragem colaborativa, apesar de eficiente, tem algumas limitações, como apontado por Rolim et al. (2017). Entre essas limitações, cita-se: problema do *cold start*, ou partida fria, que acontece quando um usuário ou item novo entram no sistema; pontuações esparsas, que podem ocorrer quando o número de usuários for muito pequeno; preferências atípicas ou *outliers*, que implicam na dificuldade de se encontrar usuários similares quando um usuário tem preferências pouco populares.

2.4.2 Recomendação Baseada em Sessão

Além das abordagens já citadas, destacam-se os sistemas de recomendação baseados em sessão (SRBS), que têm crescido em relevância nos últimos anos de acordo com Wang et al

(2021). Segundo os autores, em contraste com métodos como filtragem colaborativa ou baseada em conteúdo, que consideram todo o histórico de usuários ou itens, os SRBS têm como foco o curto prazo e buscam capturar os padrões transicionais que possam ocorrer nesse período. Dessa forma, a abordagem se torna especialmente útil nos casos em que a identificação do usuário não é possível, evento muito comum no *e-commerce*.

A sessão pode ser definida como uma unidade de transação que envolve múltiplos itens consumidos ou interações feitas em um único evento de navegação ou compra. Trata-se, por exemplo da lista de itens que um usuário interagiu em uma plataforma de *e-commerce* durante uma única visita ao site. A dependência entre os itens já conhecidos da sessão é chamada de intra-sessão, enquanto a dependência entre outras sessões anteriores é chamada de inter-sessão (WANG; CAO; WANG, 2018).

Ainda segundo Wang, Cao e Wang (2018), há três principais cenários possíveis para o SRBS: recomendação do próximo item, recomendação da próxima sessão e recomendação do próximo evento/ação. O primeiro caso é o mais utilizado e foca na previsão dos próximos itens que surgirão na sessão atual, considerando a dependência intra-sessão. Já o segundo caso tem como objetivo recomendar um conjunto de itens que possa ocorrer na sessão seguinte, utilizando, principalmente, a dependência inter-sessão. Esses dois cenários têm como foco principal o *e-commerce*, que geralmente possui uma estrutura bem definida de sessões. Esse não é o caso do último cenário, que é pensado, principalmente, para outros tipos de serviço, como plataformas de *streaming*, em que a estrutura das sessões não é tão clara.

Múltiplas técnicas podem ser aplicadas para recomendações baseadas em sessão. Métodos heurísticos simples, como cadeias de Markov, são fáceis de implementar e têm baixa complexidade computacional. Já métodos baseados em vizinhança (como o S-KNN e suas variantes S-SKNN e SF-SKNN) são amplamente utilizados devido à sua simplicidade e eficácia, considerando a similaridade entre sessões para prever a próxima ação do usuário. Redes neurais recorrentes (RNN), como o GRU4Rec, também se destacam por modelar sequências de ações, podendo assim, identificar as dinâmicas ao longo do tempo. Além desses, existem os métodos de fatoração de matrizes, que buscam modelar as interações entre itens e usuários em uma estrutura de baixa dimensionalidade (LUDEWIG; JANNACH, 2018).

O presente trabalho tem como foco os métodos baseados em vizinhança, que costumam apresentar resultados quase tão bons ou melhores do que métodos mais complexos, como mostra Ludewig e Jannach (2018). Segundo a definição dos autores, as três variações do método KNN são explicadas a seguir.

- O S-KNN (*Session-based K-Nearest Neighbors*) compara a sessão atual com sessões anteriores do conjunto de treinamento para recomendar os prováveis próximos itens que serão escolhidos a seguir pelo usuário. Utiliza medidas de similaridade, como as citadas no Quadro 1, para determinar a pontuação de recomendação para cada item. Conforme descrito na equação (2.1), dada a sessão atual s e o conjunto N_s de sessões prévias, para cada item i é feita a somatória da função de similaridade $sim(s, n)$ entre a sessão atual s e a sessão anterior n . Essa similaridade é multiplicada pela função $1_n(i)$, que retorna 1 quando o item i está em n e 0 caso contrário.

$$score_{SKNN}(i, s) = \sum_{n \in N_s} sim(s, n) \cdot 1_n(i) \quad (2.1)$$

- O S-SKNN (*Sequential Session-based K-Nearest Neighbors*) é uma variação do S-KNN que leva em consideração a sequência dos itens na sessão e dá maior importância aos itens mais recentes. Essa ponderação é feita com o objetivo de melhor refletir a preferência recente do usuário. Na fórmula (2.2), isso se dá através multiplicação pelo peso $w_n(s)$, que por sua vez é definido em (2.3). Na equação, x é a posição em s do item mais recente de s que também está presente em n e $|s|$ é o número total de itens na sessão atual s .

$$score_{S-SKNN}(i, s) = \sum_{n \in N_s} sim(s, n) \cdot w_n(s) \cdot 1_n(i) \quad (2.2)$$

$$w_n(s) = x/|s| \quad (2.3)$$

- O SF-SKNN (*Sequential Filter Session-based K-Nearest Neighbors*) é outra variação que também considera a sequência de itens, porém de maneira mais restritiva. Neste método, só serão considerados itens que apareceram em sessões anteriores logo após o último item da sessão atual. Isso garante que as recomendações sigam uma sequência já observada no passado. Na equação (2.4), isso se dá pela função $1_n(s_{|s|}, i)$, que retorna 1 apenas caso o último item $s_{|s|}$ da sessão atual apareça diretamente antes do item i na sessão anterior n . Caso contrário, a função retorna 0.

$$score_{SF=SKNN}(i, s) = \sum_{n \in N_s} sim(s, n) \cdot 1_n(s_{|s|}, i) \quad (2.4)$$

2.5. Métricas de Avaliação

As métricas de avaliação desempenham um papel crucial na análise de sistemas de recomendação. No presente trabalho, que utiliza o cenário de recomendação de próximo item, são aplicadas as métricas de *hit rate*, ou taxa de acerto, e MRR (*Mean Reciprocal Rank*). O *hit rate* mede a proporção de casos em que o item verdadeiro estava na lista de itens recomendados, sem levar em consideração sua posição nessa lista. Já o MRR verifica essa posição e, em seguida, calcula seu inverso. A métrica, portanto, que varia entre 0 e 1, terá maiores valores quando o item verdadeiro estiver no topo da lista. Sua pontuação final será a média dos resultados (HIDASI et al., 2016).

Além dessas métricas de acurácia, também são aferidos fatores de qualidade, como demonstrado por Jannach e Ludewig (2018). O fator cobertura calcula a porcentagem de itens do catálogo que aparecem ao menos uma vez em listas de recomendação e tem como objetivo refletir a diversidade das recomendações. Já a média de popularidade calcula o quão populares são os itens recomendados, de modo que o item menos popular devolva a pontuação 0 e o item mais popular devolva a pontuação 1.

3 METODOLOGIA

Nesta sessão do trabalho, é explicada a metodologia utilizada para avaliar o desempenho dos modelos baseados em vizinhança no problema de recomendação do próximo item. As variações analisadas foram S-KNN, S-SKNN e SF-SKNN, aplicadas sobre uma base de dados de *e-commerce*. A fim de examinar a influência da quantidade de itens candidatos e do viés de popularidade, foram desenvolvidos quatro cenários distintos. Como métricas de avaliação, utilizou-se MRR e *hit rate* para aferir a acurácia, além de popularidade média e cobertura de catálogo como medidas de qualidade.

3.1 Base de Dados

A base de dados utilizada no experimento foi obtida através da plataforma Kaggle, uma comunidade online de cientistas de dados. Ela contém informações sobre interações de usuários em uma plataforma de *e-commerce* com múltiplas categorias de produtos durante o mês de outubro de 2019. Cada linha da base representa um evento relacionado a produtos e usuários, seja essa interação uma compra, visualização, adição ou remoção do carrinho. O conjunto de dados completo possui 42.448.764 linhas de eventos com 9 atributos (KECHINOV, 2020):

- *event_time*: momento exato em que o evento ocorreu, no formato UTC;
- *event_type*: tipo de evento;
- *product_id*: identificador de cada produto;
- *category_id*: identificador de categoria do produto;
- *category_code*: nome da categoria;
- *brand*: nome da marca do produto;
- *price*: preço do produto;
- *user_id*: identificador de usuário;
- *user_session*: identificador da sessão do usuário.

3.2 Pré-Processamento

Após a obtenção dos dados brutos, foi feita uma preparação para que eles pudessem ser devidamente utilizados pelos algoritmos de recomendação. A primeira etapa envolveu a limpeza e seleção. Foram escolhidas apenas as colunas relevantes para o experimento, sendo elas *event_time*, *event_type*, *product_id*, *category_code*, e *user_session*. Além disso, foram

excluídas todas as linhas que continham valores nulos em qualquer um desses atributos, garantindo, assim, que o conjunto possuísse apenas interações completas.

A seguir, foi realizada a ordenação, primeiro pela coluna *user_session* e depois pela coluna *event_time*, de maneira a assegurar que os eventos dentro de cada sessão sigam a ordem cronológica. Também houve a conversão da coluna *event_time* para o formato data e hora, permitindo, assim, a utilização desse atributo em filtros temporais.

Como a base de dados escolhida é muito extensa, contendo mais de 42 milhões de linhas, foi selecionada uma amostra incluindo o primeiro milhão de linhas. Isso foi feito com objetivo de reduzir a carga computacional, tempo de processamento e ainda assim manter a representatividade dos dados. Para garantir a integridade de todas as sessões, assegurou-se que o ponto de corte não pudesse ocorrer no meio de uma sessão.

Após a limpeza e seleção da amostra, foi realizado o mapeamento de itens e sessões, substituindo os identificadores anteriores por índices inteiros e sequenciais. Isso reduz a complexidade das operações computacionais e facilita o processamento dos dados.

Na etapa seguinte, foi criada uma lista contendo todas as sessões da amostra em formato de tupla. Cada tupla, por sua vez, constituída por dois elementos, o identificador da sessão e a lista de itens que foram vistos nessa sessão, em ordem cronológica. Considerou-se apenas itens com *event_type* de visualização e sessões com mais de um item visualizado. Essas restrições ajudam a uniformizar os dados e garantir que exista um histórico suficiente para fazer a projeção do próximo item.

Além dessa lista de sessões, foi criado, também, um dicionário contendo, para cada item, todas as sessões em que ele aparece. Isso foi feito com o objetivo de acelerar o processamento durante as avaliações.

3.3 Divisão entre Treinamento e Teste

Com os dados devidamente preparados para o processamento, foi feita a divisão entre o conjunto de treinamento e o conjunto de teste. Para que as regras de seleção de itens candidatos pudessem ser aplicadas corretamente, primeiro, garantiu-se que as sessões referentes ao primeiro dia do mês estivessem no conjunto de treinamento. As demais sessões foram, então, divididas aleatoriamente, sendo 80% usadas para teste e 20% para treinamento, percentuais tipicamente utilizados no aprendizado de máquina, conforme descrito por Murphy (2012). Ainda segundo o autor, essa divisão é fundamental para que o modelo treinado em uma parte do conjunto possa ser validado utilizando dados não vistos durante o treinamento. Informações sobre os conjuntos gerados podem ser vistos na Tabela 1.

Tabela 1 – Características dos conjuntos de treinamento e teste

Dataset	Treinamento	Teste	Total
Eventos	723.846	175.442	899.288
Sessões	119.538	28.783	148.321
Itens	31.842	17.370	34.107
Eventos por Sessão	6,1	6,1	6,1
Itens Únicos por Sessão	3,6	3,6	3,6

Fonte: Elaborado pelo autor.

3.4 Seleção de Itens Candidatos

No presente trabalho, foram utilizadas regras para a filtragem e seleção de itens candidatos, restringindo o conjunto de itens a um número menor e mais relevante. Isso foi feito para evitar que todos os itens fossem percorridos em cada iteração das avaliações de pontuação, o que tornaria o tempo de processamento inviável. Os critérios utilizados para a seleção foram janela de tempo, categoria dos produtos e fixação de uma quantidade máxima de itens para o conjunto de candidatos.

O filtro de janela de tempo garante que apenas itens visualizados dentro de um intervalo de tempo específico, anterior à sessão em análise, sejam considerados como candidatos. Esse período foi fixado em 24 horas, com o objetivo de capturar as preferências mais recentes, além de limitar a quantidade de produtos considerados. Devido a esse critério, todas as sessões que ocorreram no primeiro dia da base de dados foram colocadas exclusivamente no conjunto de treinamento. Isso foi feito pois, como não seria possível aplicar o filtro de maneira adequada nesse subconjunto, a qualidade dos resultados poderia ser comprometida.

O segundo filtro aplicado utilizou o atributo categoria. Considerou-se como candidatos apenas itens que pertencem a alguma das categorias já visualizadas até então na sessão em análise. Isso foi feito para assegurar que as recomendações estejam alinhadas aos tipos de produto que o usuário demonstrou interesse recentemente. Por exemplo, caso o usuário tenha visualizado durante a sessão apenas eletrodomésticos, não serão recomendados itens de outras categorias, como vestuário.

O último critério aplicado foi a limitação da quantidade de itens candidatos, visando acelerar o processamento computacional. Essa limitação foi implementada de diferentes maneiras, com variações tanto no número máximo de itens quanto nos parâmetros utilizados para fazer a seleção. Ao todo, foram gerados quatro cenários distintos, que buscam avaliar a influência que a quantidade máxima e o viés de popularidade na escolha dos candidatos têm na qualidade dos resultados.

Os valores n , que representam a quantidade máxima de itens no conjunto de candidatos, foram estabelecidos em 100 e 200. Esses valores foram escolhidos para limitar a quantidade de itens de forma que o tempo de processamento continuasse viável. Para a escolha desses itens, foram aplicados dois diferentes critérios de seleção na lista gerada após os filtros de tempo e de categoria: popularidade e aleatoriedade. O primeiro escolheu os n itens que mais apareceram na lista, enquanto o segundo selecionou n itens de forma aleatória. Caso a lista já possuísse uma quantidade menor ou igual a n , ela seria considerada por completo em ambos os critérios. Os cenários gerados por essas variações estão resumidos no Quadro 2.

Quadro 2 – Cenários Avaliados no Experimento

Cenário	Quantidade de Itens Candidatos	Critério de Escolha de Itens Candidatos
Cenário A	100	Popularidade
Cenário B	200	Popularidade
Cenário C	100	Aleatoriedade
Cenário D	200	Aleatoriedade

Fonte: Elaborado pelo autor.

3.5 Avaliação

Dado que o objetivo do presente trabalho é avaliar o desempenho dos algoritmos na recomendação do próximo item, removeu-se o último item visualizado das sessões de teste, criando, assim, sessões de teste parciais para as quais o próximo item precisa ser previsto. Esse item removido é, depois, utilizado para aferir a acurácia das recomendações.

Entre as diversas medidas de similaridade que poderiam ser aplicadas nas técnicas de recomendação, optou-se pelo uso da Similaridade de Jaccard. Como demonstrado por Pang-Ning, Steinbach e Kumar (2006), esse método funciona bem no tipo de cenário em questão, uma vez que foca nos atributos em comum entre os dois conjuntos, calculando a razão entre intersecção e união desses elementos.

Com a métrica de similaridade definida, foi possível calcular uma matriz de similaridade entre todas as sessões de treinamento e todas as sessões de teste parciais. Isso aumentou a eficiência do processo de recomendação, uma vez que os coeficientes de similaridade estavam, previamente, processados.

Após o cálculo da matriz, deu-se início à aplicação dos modelos de recomendação baseados em sessão. Para cada sessão de teste parcial, foi gerada uma lista de itens candidatos

seguindo os critérios e variações descritos no tópico 3.2. Em seguida, utilizou-se os métodos S-KNN, S-SKNN e SF-SKNN, conforme detalhados no tópico 2.4.2, para calcular a pontuação de cada item dessa lista, com o auxílio da matriz de similaridade. Ao final, os itens foram ordenados de forma decrescente e aqueles com as 10 maiores pontuações foram incluídos na lista de recomendados.

Definidas as listas de itens recomendados, foram aplicadas as métricas de acurácia e qualidade. Para cada uma dessas listas, o *hit rate* e o MRR foram calculados utilizando o último item visualizado na sessão de teste como referência. Além disso, foram computadas as médias de popularidade e cobertura de catálogo. Ao final dos experimentos, as médias dessas métricas foram agregadas para cada um dos métodos, possibilitando a comparação do desempenho em todos os cenários analisados.

4 ANÁLISE DOS RESULTADOS

Neste capítulo, são discutidos os resultados obtidos a partir do experimento. Ao todo, foram utilizados três métodos de recomendação, S-KNN, S-SKNN e SF-SKNN. Esses métodos foram aplicados em quatro cenários diferentes, com variações de quantidade e critério de escolha dos itens candidatos. Para a avaliação dos resultados, foram aplicados quatro critérios, MRR, *hit rate* (HR), cobertura de catálogo (COV) e popularidade média (POP).

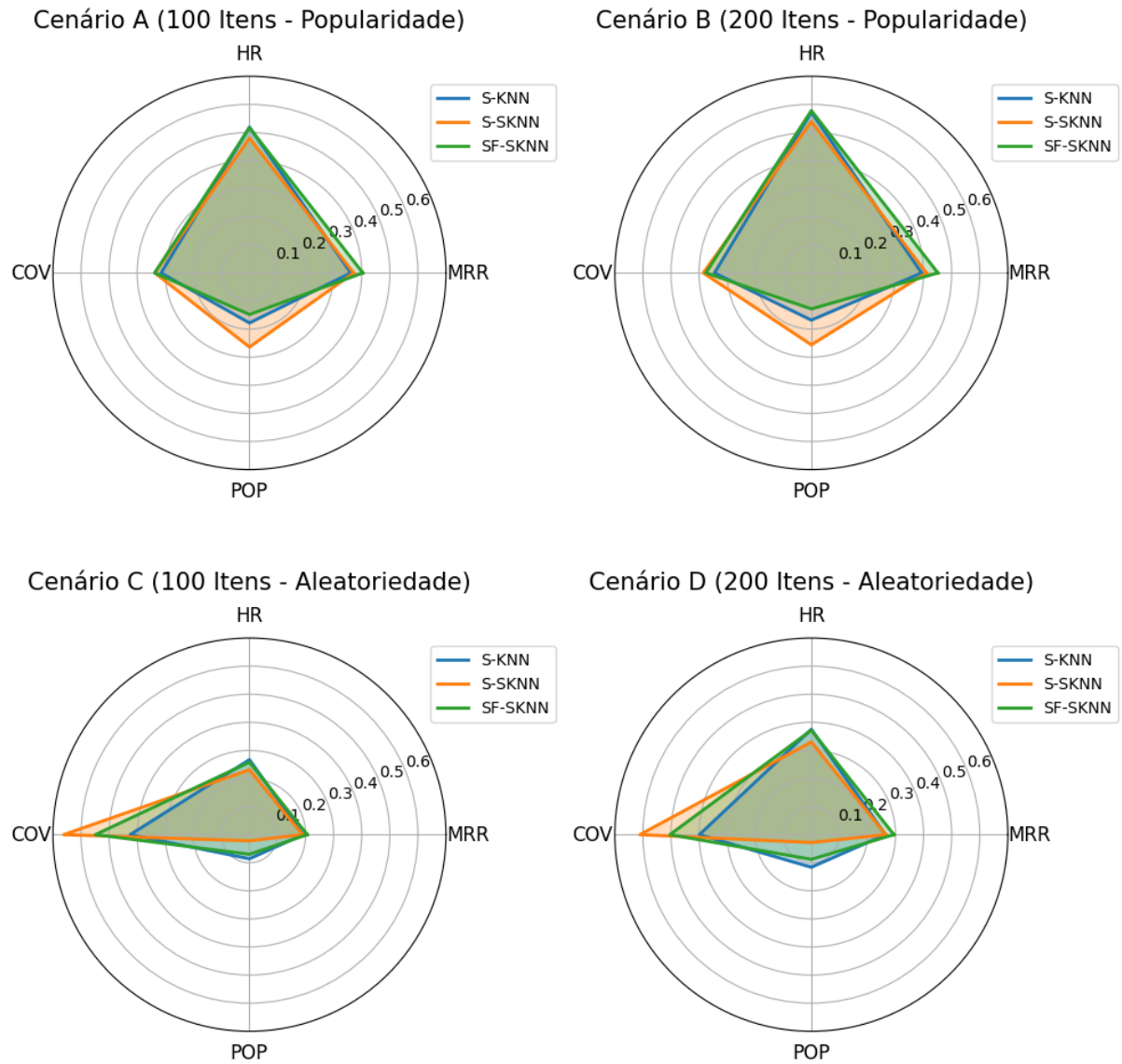
Todos os resultados estão presentes na Tabela 2. A fim de facilitar a interpretação e visualização, esses dados também foram transformados em gráficos de radar, vistos nas Figura 2 e 3. A Figura 2 traz um gráfico para cada cenário, assim os modelos podem ser comparados em relação às suas métricas. Já a Figura 3 exibe um gráfico para cada métrica, permitindo a comparação dos modelos em relação aos resultados obtidos em cada cenário.

Tabela 2 – Resultados das métricas MRR, *hit rate* (HR), cobertura de catálogo (COV) e popularidade média (POP) para os modelos S-KNN, S-SKNN e SF-SKNN nos diferentes cenários testados.

Cenários	Modelos	MRR	HR	COV	POP
Cenário A	S-KNN	0,361	0,519	0,315	0,178
Cenário A	S-SKNN	0,371	0,480	0,337	0,264
Cenário A	SF-SKNN	0,404	0,517	0,338	0,148
Cenário B	S-KNN	0,393	0,572	0,345	0,168
Cenário B	S-SKNN	0,410	0,539	0,385	0,256
Cenário B	SF-SKNN	0,451	0,578	0,376	0,128
Canário C	S-KNN	0,191	0,267	0,425	0,086
Canário C	S-SKNN	0,189	0,231	0,662	0,022
Canário C	SF-SKNN	0,206	0,258	0,549	0,070
Cenário D	S-KNN	0,264	0,375	0,399	0,116
Cenário D	S-SKNN	0,265	0,330	0,611	0,028
Cenário D	SF-SKNN	0,294	0,372	0,504	0,088

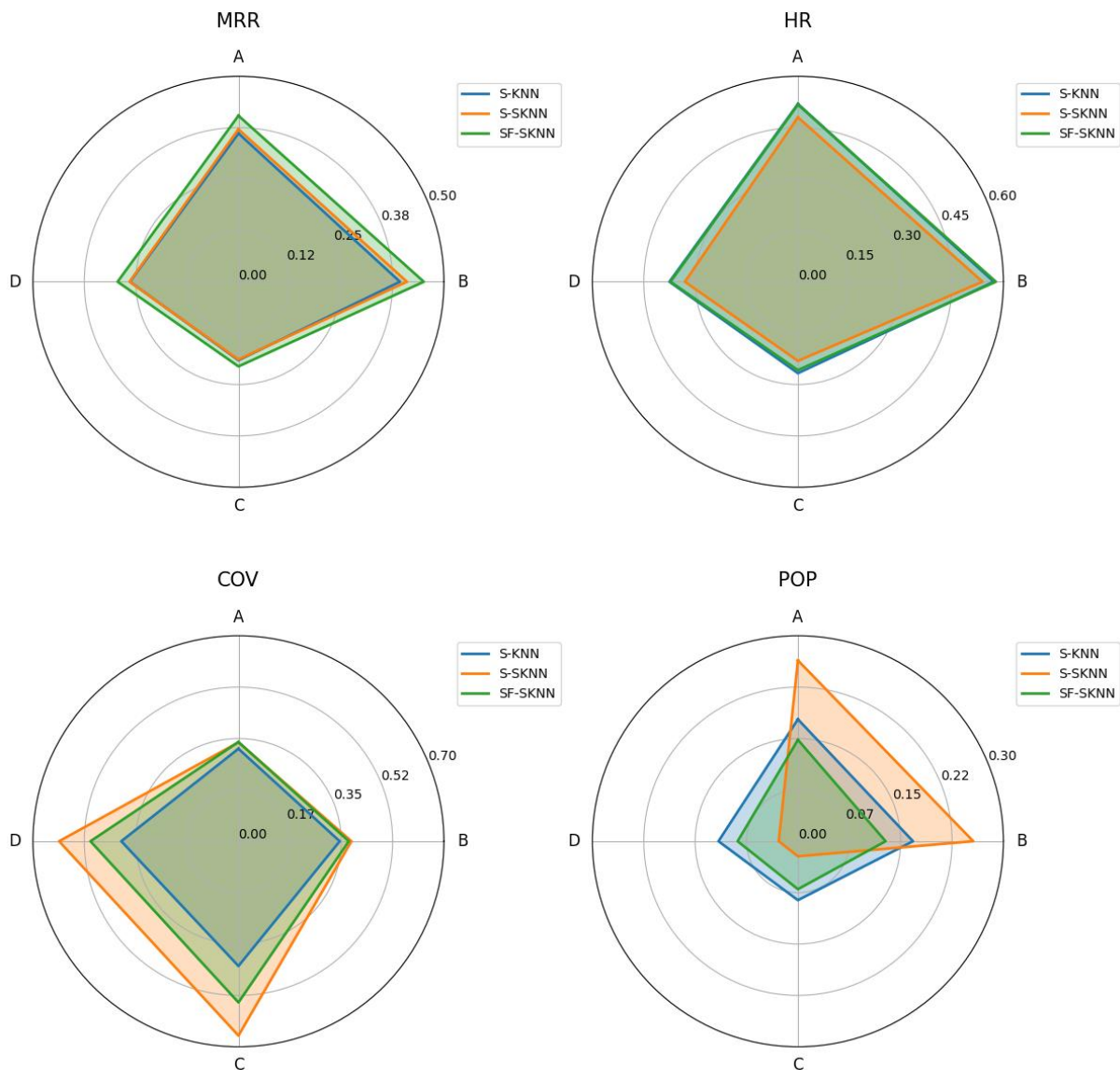
Fonte: Elaborado pelo autor.

Figura 2 – Gráficos de radar comparando S-KNN, S-SKNN e SF-SKNN em relação as métricas MRR, *hit rate* (HR), cobertura de catálogo (COV) e popularidade média (POP), com um gráfico para cada um dos quatro cenários.



Fonte: Elaborado pelo autor.

Figura 3 – Gráficos de radar comparando S-KNN, S-SKNN e SF-SKNN em relação aos resultados obtidos em cada cenário, A B, C e D, com um gráfico para cada métrica.



Fonte: Elaborado pelo autor.

4.1 Métodos de Recomendação

Em todos os cenários, o método SF-SKNN obteve os melhores resultados de acordo com a métrica MRR, indicando que o modelo foi mais eficaz em posicionar o item verdadeiro mais próximo do topo na lista de recomendados. Já segundo a métrica HR, o SF-SKNN ficou na segunda colocação em 3 dos cenários, atrás do S-KNN. Isso significa que, apesar de não ser tão preciso na ordenação dos itens, figurando em terceiro em 3 cenários, o S-KNN obteve mais sucesso em garantir que o item verdadeiro estivesse na lista recomendada. O método S-SKNN se manteve na última posição em todos os cenários na métrica HR, mas ficou em segundo em 3 dos cenários segundo a métrica MRR. Isso demonstra que, quando comparado ao método S-

KNN, o peso aplicado ao coeficiente de similaridade ajudou a posicionar o item verdadeiro mais próximo do topo da lista, mas fez com que ele, muitas vezes, não aparecesse entre os recomendados.

Em relação as medidas de qualidade, o S-KNN apresentou, em todos os cenários, o menor índice de cobertura de catálogo (COV), recomendando uma menor variedade de itens do que os demais métodos. O modelo S-SKNN se destacou nessa métrica, alcançando a primeira posição em 3 dos cenários, enquanto o SF-SKNN obteve um resultado intermediário. No que diz respeito a média de popularidade (POP), foi possível observar que os pesos aplicados pelo método S-SKNN resultaram na recomendação dos itens mais populares quando a escolha de itens candidatos teve viés de popularidade, cenários A e B, e itens menos populares quando a escolha foi aleatória, cenários C e D. Com exceção desses últimos casos, o SF-SKNN recomendou os produtos menos populares, enquanto o S-KNN teve um maior viés de popularidade.

4.2 Quantidade de Itens Candidatos

Ao comparar os cenários com até 100 itens candidatos (A e C) com os cenários com até 200 itens candidatos (B e D), nota-se que todas as medidas de acurácia apresentaram resultados superiores com a maior quantidade de itens. Isso sugere que o aumento no número de candidatos proporciona mais opções relevantes e aumenta a chance de o item verdadeiro estar presente e bem posicionado na lista de recomendação.

A aumento da lista de candidatos também resultou em maior variedade e menor média de popularidade quando a escolha desses candidatos foi feita com viés de popularidade. Isso acontece, pois, ao expandir a lista, itens menos populares são incluídos e podem ser recomendados. Já quando não houve esse viés, observou-se o contrário, diminuição na cobertura de catálogo e aumento na média de popularidade. Nesse caso, ao aumentar a lista de candidatos, os novos itens inclusos podem ser mais populares, visto que eles tendem a aparecer com mais frequência, o que explicaria essas variações.

4.3 Viés de Popularidade na Escolha de Itens Candidatos

A utilização do viés de popularidade na escolha dos itens candidatos teve, como consequência, recomendações mais precisas, como pode ser observado nas métricas MRR e HR. Isso ocorre porque itens mais populares têm probabilidade maior de serem relevantes para mais usuários, o que aumenta as chances de o item verdadeiro estar entre os recomendados. Como consequência, entretanto, há redução na diversidade de produtos presentes na

recomendação, assim como aumento na média de popularidade, fatores que podem limitar a capacidade do sistema de recomendar novos produtos.

4.4 Considerações Finais

A análise dos resultados permite concluir que o SF-SKNN obteve o desempenho superior aos outros dois métodos. Na métrica MRR, posicionou melhor o item verdadeiro na lista de recomendados em todos os cenários e, apesar de não ter ficado em primeiro na métrica HR em todas as instâncias, sempre obteve resultado próximo ao maior observado. Quanto a cobertura de catálogo, apresentou resultado satisfatório, ficando atrás somente do S-SKNN em 3 cenários. Foi, também, o modelo que recomendou menos os produtos mais populares, com exceção de dois casos do S-SKNN já discutidos.

Conclui-se, também, que o aumento na quantidade de itens candidatos influenciou positivamente as métricas de acurácia, efeito também observado no uso do viés de popularidade na escolha dos itens. Essa escolha, porém, resultou em recomendações menos diversas, com maiores médias de popularidade.

5 CONCLUSÃO

Conforme os objetivos estabelecidos, este trabalho conseguiu avaliar o desempenho das técnicas S-KNN e suas variações S-SKNN e SF-SKNN na resolução do problema de recomendação do próximo item em uma base de e-commerce. Concluiu-se que, nas condições analisadas, o método SF-SKNN apresentou os melhores resultados, alcançando um equilíbrio entre acurácia e diversidade. Se manteve em primeiro na métrica MRR e figurou entre primeiro e segundo nas métricas de taxa de acerto e cobertura de catálogo. O modelo clássico S-KNN ofereceu a menor variedade de recomendações, tendendo aos itens mais populares, porém foi mais eficaz em garantir que o item verdadeiro estivesse na lista de recomendados. O modelo S-SKNN, por outro lado, teve as menores taxas de acerto, mas alcançou resultado intermediário na métrica MRR.

Observou-se, também, que o aumento na quantidade de itens na lista de candidatos leva a resultados mais fidedignos. O mesmo efeito pode ser notado com o uso do viés de popularidade na escolha de itens candidatos. Esse viés, porém, também implica em recomendações menos diversas.

Tais descobertas podem auxiliar plataformas de e-commerce na escolha de técnicas de recomendação mais adequadas para seu modelo de negócio, buscando maximizar a quantidade de vendas e melhorar a experiência do usuário.

Entre as limitações do trabalho, cita-se o tamanho da base de dados utilizada, que precisou ser restringida a 1 milhão de linhas. Em aplicações reais, o uso de bases maiores seria necessário, assim como uma maior velocidade de processamento. Além disso, foram utilizados diversos filtros para a seleção de itens candidatos, o que condicionou as recomendações que poderiam ser feitas.

Dessa forma, como recomendação para futuros trabalhos, sugere-se a aplicação de métodos diferentes para a seleção de itens candidatos, de forma que seja possível analisar suas influências nos resultados. Além disso, também é recomendado o uso de outras técnicas de recomendação baseada em sessão que possam complementar as abordagens utilizadas neste trabalho, ampliando, assim, o entendimento dessa área de estudo.

REFERÊNCIAS

- AGGARWAL, Charu C. **Recommender Systems: The Textbook**. Cham: Springer International Publishing, 2016. ISBN 978-3-319-29657-9.
- CLISSA, Luca; LASSNIG, Mario; RINALDI, Lorenzo. **How Big is Big Data? A Comprehensive Survey of Data Production, Storage, and Streaming in Science and Industry**. *Frontiers in Big Data*, v. 6, 2023.
- FAYYAZ, Zeshan; EBRAHIMIAN, Mahsa; NAWARA, Dina; IBRAHIM, Ahmed; KASHEF, Rasha. **Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities**. *Applied Sciences*, v. 10, n. 21, p. 7748, 2020.
- HIDASI, Balázs; QUADRANA, Massimo; KARATZOGLOU, Alexandros; TIKK, Domonkos. **Parallel Recurrent Neural Network Architectures for Feature-rich Session-based Recommendations**. 2016.
- KECHINOV, Maksim. **Ecommerce Behavior Data from Multi Category Store**. Kaggle, 2020. Disponível em: <https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store?resource=download&select=2019-Oct.csv>. Acesso em: 25 jun. 2024.
- LUDEWIG, Malte; JANNACH, Dietmar. **Evaluation of Session-based Recommendation Algorithms**. TU Dortmund; AAU Klagenfurt, 2018.
- MURPHY, Kevin P. **Machine Learning: A Probabilistic Perspective**. Cambridge, Massachusetts: MIT Press, 2012. ISBN 978-0-262-01802-9.
- REATEGUI, E. B.; CAZELLA, S. C. **Sistemas de Recomendação**. In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL, XXV CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 2005.
- RICCI, Francesco; ROKACH, Lior; SHAPIRA, Bracha (Ed.). **Recommender Systems Handbook**. 2. ed. New York: Springer, 2015. ISBN 978-1-4899-7636-9.
- ROLIM, Vitor; FERREIRA, Rafael; COSTA, Evandro de B.; PINHEIRO, Anderson. **Um Estudo Sobre Sistemas de Recomendação de Recursos Educacionais**. In: WORKSHOP DE COMPUTAÇÃO NA EDUCAÇÃO, 2017, Fortaleza.
- SONDUR, Sridhar Dilip; CHIGADANI, Amit P.; NAYAK, Shantharam. **Similarity Measures for Recommender Systems: A Comparative Study**. *Journal for Research*, v. 2, n. 3, p. 76-79, maio 2016.
- TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introduction to Data Mining**. Harlow: Pearson Education Limited, 2014.
- WANG, Shoujin; CAO, Longbing; WANG, Yan; SHENG, Quan Z.; ORGUN, Mehmet A.; LIAN, Defu. **A Survey on Session-based Recommender Systems**. *ACM Comput. Surv.*, v. 9, n. 4, p. 39, 2021.

WANG, Shoujin; CAO, Longbing; WANG, Yan. **A Survey on Session-based Recommender Systems**. University of Technology Sydney; Macquarie University, 2018.