



ALGORITMO DE RECOMENDAÇÃO DE RECEITAS A PARTIR DAS COMPRAS REALIZADAS NO MERCADO

Aparecida Vânia de Jesus RA:10407484
Lucas Gomes Porfírio da Silva RA: 10370475
Vanessa Hacklauer de Aguiar RA:10407324
Wagner de Mendonça Trindade RA: 10407917

03 INTRODUÇÃO

04 DATASET

05 ANALISE EXPLORATÓRIA

07 ALGORITMO DE RECOMENDAÇÃO

08 REFERENCIAL TEÓRICO

09 METODOLOGIA

10 BIBLIOTECA

11 AVALIAÇÃO DE DESEMPENHO

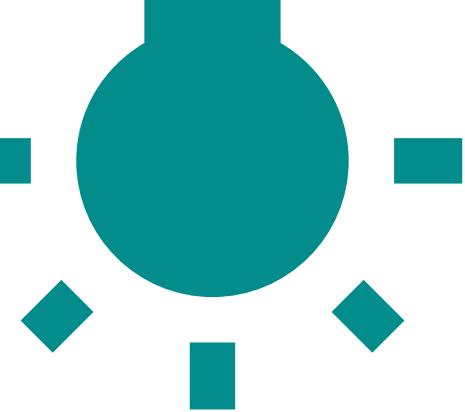
12 PROVA DE CONCEITO

14 CONSOLIDAÇÃO DOS RESULTADOS

18 CONCLUSÃO

19 REFERÊNCIAS

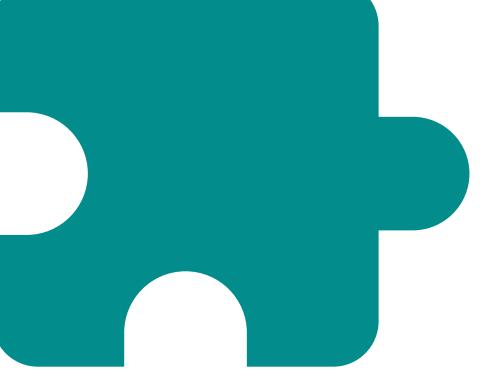




INTRODUÇÃO

Este estudo desenvolveu um algoritmo de recomendação de receitas, que utiliza bibliotecas de mineração de dados e análise estatística para oferecer recomendações personalizadas baseadas nas compras de ingredientes dos usuários. O algoritmo analisa o histórico de compras para identificar as preferências de ingredientes do usuário, compara essas informações com um banco de dados de receitas, e recomenda as três receitas mais relevantes que utilizam esses ingredientes. O processo é projetado para se aprimorar iterativamente, adaptando-se às preferências do usuário à medida que mais dados são coletados.



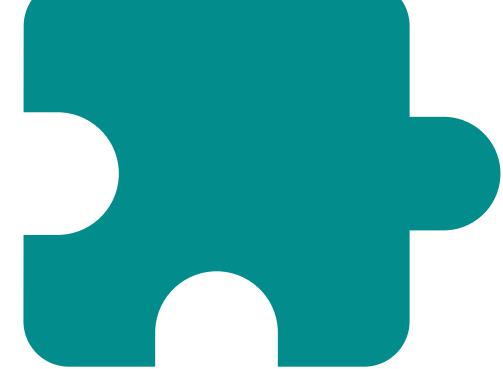


DATASET

Utilizamos dois dataset.

- Dataset Compras composto por 1000 itens que simulam compras em um super mercado.
- Dataset Receitas composto por 50 receitas tradicionais brasileiras, abrangendo saladas, sobremesas, doces, tortas, pratos principais e lanches





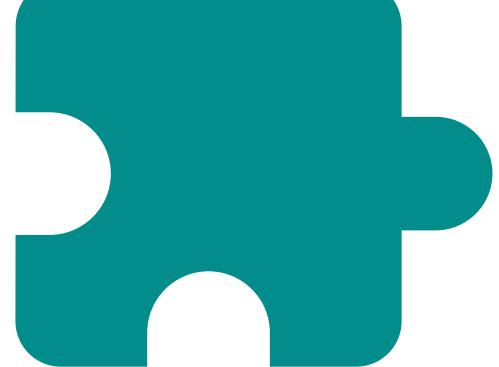
ANÁLISE EXPLORATÓRIA

Base Compras

- Bibliotecas: foi utilizada a biblioteca Pandas Ydata gerando uma analise com o *ProfileReport*
- Overview

Overview		Alerts (3)	Reproduction
Dataset statistics			
Number of variables	3		
Number of observations	1000		
Missing cells	0		
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	23.6 KIB		
Average record size in memory	24.1 B		
Variable types			
Numeric	2		
Text	1		





ANÁLISE EXPLORATÓRIA

Base Receitas

- Nessa base também temos 3 variáveis com 49 observações
- Overview

Overview		Alerts 2	Reproduction
Dataset statistics			
Number of variables	3		
Number of observations	49		
Missing cells	0		
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	1.3 KiB		
Average record size in memory	26.6 B		





ALGORITMOS DE RECOMENDAÇÃO

O algoritmo de recomendação é uma ferramenta útil e muito utilizada atualmente em sites e aplicativos para indicar produtos que possam atrair os clientes. Os algoritmos de recomendação são multidisciplinares, envolvendo conhecimentos em software, ciência de dados, matemática, estatística, sociologia e psicologia, com as técnicas certas, eles podem aprimorar significativamente a experiência do usuário em plataformas online.



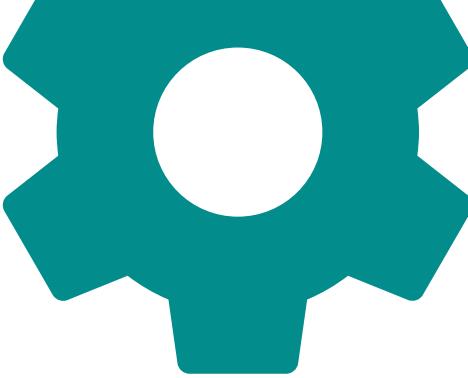
```
    }
    render() {
      return (
        <React.Fragment>
          <div className="py-5">
            <div className="container">
              <Title name="our">
                <div className="row">
                  <ProductConsumer>
                    {(value) =>
                      <div>
                        <h1>Our Products</h1>
                        <ProductList items={value.items} />
                      </div>
                    }
                  </ProductConsumer>
                </div>
              </div>
            </div>
          </React.Fragment>
        )
      )
    }
  }
}

export default App;
```



REFERÊNCIAL TEÓRICO

- PNL
- CountVectorizer



METODOLOGIA

Normalizar e tokenizar os ingredientes das receitas

Normalizamos a coluna de ingredientes no dataset, que contém listas de ingredientes em formato de texto.

Em seguida realizamos o processo de tokenização dessa coluna, ou seja, separar os ingredientes em listas de strings individuais, como converter tudo para minúsculas para garantir a consistência

Preparar o Dataset de Compras/Receitas

No dataset de Compras, transformamos os dados de forma que, todas as colunas chamadas de ‘item..’ foram agregadas em uma só coluna, utilizando a biblioteca Panda e viraram strings dentro de listas.

No dataset de Receitas, como todos os ingredientes já estavam na mesma coluna, foi necessário somente transformá-los em strings dentro de listas.

Converter os dados para o mesmo padrão

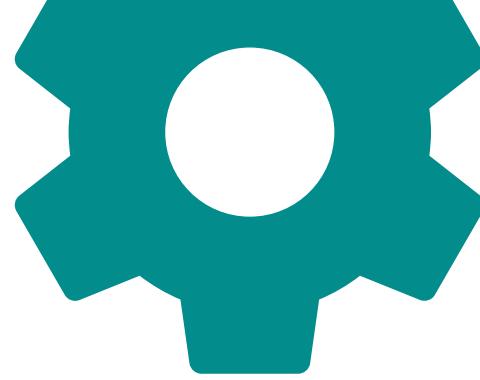
Utilizamos uma função na qual empregamos a biblioteca do python re para remover caracteres especiais dos ingredientes das duas tabelas de dados, como por exemplo traços, espaços extras, parênteses, a palavra e também foi removida, acentos e vírgulas no meio de palavras. Esta função foi aplicada nos dois datasets.

Criamos uma coluna ingredientes_str com os dados normalizados.

Técnica do processo de recomendação

O algoritmo utiliza a similaridade de cosseno para encontrar itens no dataset receitas que são semelhantes a uma entrada do usuário, uma lista de compras do usuário.

A lista de compras fornecida pelo usuário é transformada em um vetor utilizando o mesmo método usado para vetorizar os dados das receitas, converte a string de texto em uma representação numérica que pode ser comparada com as receitas vetorizadas.



BIBLIOTECA

Pandas

é usada para realizar a manipulação e análise de dados em Python de maneira eficiente. Ela nos permite realizar uma série de operações essenciais, como a limpeza de dados, seleção de colunas específicas, e agregação de informações, sendo instrumental no pré-processamento de dados.

NumPy

Oferece suporte na manipulação de arrays e matrizes multidimensionais, além de prover um vasto conjunto de operações matemáticas. Isso é particularmente útil para realizar cálculos numéricos complexos e operações de alta performance, fundamentais no tratamento eficaz dos dados.

Matplotlib

Para a visualização de dados, utilizamos a biblioteca Matplotlib para gerar gráficos, como demonstrado no código de distribuição dos valores de similaridade de cosseno. Isso nos permite visualizar a distribuição e entender melhor os resultados do nosso algoritmo de recomendação.

Pyplot

É um módulo do pacote Matplotlib, que fornece uma interface que imita a do MATLAB, facilitando a criação de visualizações gráficas como gráficos de linhas, barras, histogramas, nuvens de pontos, de maneira simples, amplamente utilizada para visualização de dados e análise exploratória.

Scikit-Learn

O uso do `CountVectorizer` e da função `cosine_similarity`, da biblioteca Scikit-Learn, oferece ferramentas para pré-processamento de dados, aprendizado de máquina, e métricas de avaliação, como a similaridade de cosseno, essencial para calcular a proximidade entre os vetores de itens e recomendar os mais relevantes.



AVALIAÇÃO DE DESEMPENHO

Alguns métodos de avaliação de desempenho como Acurácia, F1-Score e outras mais utilizadas não podem ser aplicadas neste algoritmo devido a especificidade dos dados.

- Cobertura de Catalogo
- Diversidade
- Novidade
- Serenidade





PROVA DE CONCEITO

Obtivemos sucesso com o algoritmo de recomendação, ele realizou a recomendação de receitas para as 1000 entradas.

Devido à falta de uma coluna de avaliação na base de dados, os algoritmos de recomendação que trabalham com separação de dados entre treino e teste não puderam ser testados nem aplicados, pois os algoritmos que utilizam este tipo de abordagem precisam de um item que eles chamam de ‘avaliação da recomendações’ ou da ‘nota de relevância’ da receita que seria dada pelo usuário.

Como não temos estes dados, foi empregado o método de funções, uma para definir as recomendações através da similaridade do cosseno e outra para escolher os 3 resultados mais relevantes e para imprimir os resultados para o usuário.





CONSOLIDAÇÃO DOS RESULTADOS

Resultado geral, obtido sem a definição de uma métrica de aceitação, podemos verificar que **as recomendações estão corretas** e que as receitas realmente tem em sua lista de ingredientes os mesmo ingredientes das compras de referência:

Recomendações para a lista de compras 1:

- Nhoque à bolonhesa
- Massa de Pizza
- Strogonoff de Carne

Recomendações para a lista de compras 2:

- Danoninho Caseiro
- Bobó de Camarão
- Moqueca de Peixe

Recomendações para a lista de compras 3:

- Danoninho Caseiro
- Massa de Panqueca
- Nhoque à bolonhesa





CONSOLIDAÇÃO DOS RESULTADOS

Devido a falta de um rating, utilizaremos métricas mais específicas como Precision@k, Recall@k e a cobertura do catálogo.

A métrica Precision@k é utilizada para avaliar como as receitas recomendadas são relevantes para o usuário. Uma recomendação é relevante se contém pelo menos metade dos ingredientes de compras do usuário.

A Precisão média foi de 92,77%.

Um resultado excelente, que indica que a maioria das recomendações que o seu sistema faz nos top-k, assumindo k como 3 - já que o algoritmo recomenda as 3 receitas, são realmente relevantes para os usuários. Isso sugere que o sistema é muito eficaz em identificar e sugerir receitas aos usuários, baseando-se na lista de compras.



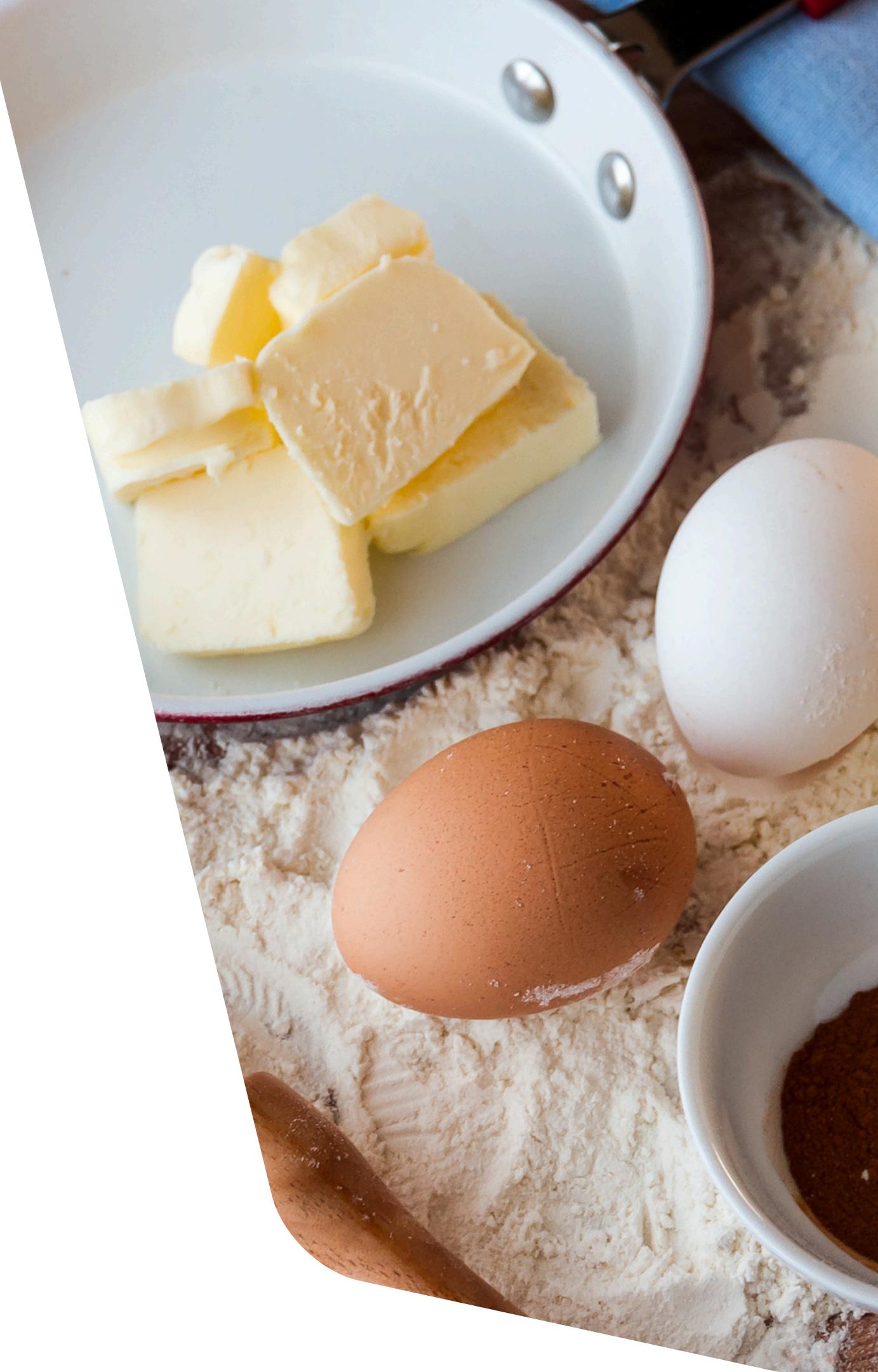


CONSOLIDAÇÃO DOS RESULTADOS

A métrica Recall@k é utilizada para medir quantos dos ingredientes comprados foram cobertos pelas receitas recomendadas. Calcula o recall das recomendações, ou seja, **quão bem as recomendações cobrem os ingredientes que o usuário realmente comprou.**

Recall é o número de itens relevantes e recomendados dividido pelo total de relevantes., com 32,13% o recall foi relativamente baixo.

Isso significa que, embora as recomendações que seu sistema faz são relevantes, ele está capturando apenas uma pequena fração das receitas que seriam consideradas relevantes pelos usuários, ou seja, muitas receitas potencialmente relevantes não estão sendo recomendadas.





CONSOLIDAÇÃO DOS RESULTADOS

A métrica Cobertura de catálogo é utilizada para indicar a porcentagem de receitas diferentes recomendadas pelo sistema em relação ao total de receitas disponíveis no catálogo.

A cobertura do catálogo é o número de itens únicos recomendados dividido pelo tamanho do catálogo. Com 66,67%, a cobertura do catálogo é relativamente boa, indicando que o sistema explora dois terços do catálogo de receitas disponíveis ao fazer recomendações.

Isso é positivo porque sugere que o sistema não está apenas se concentrando em um pequeno conjunto de receitas populares ou frequentes, mas está oferecendo uma gama razoavelmente ampla de opções.

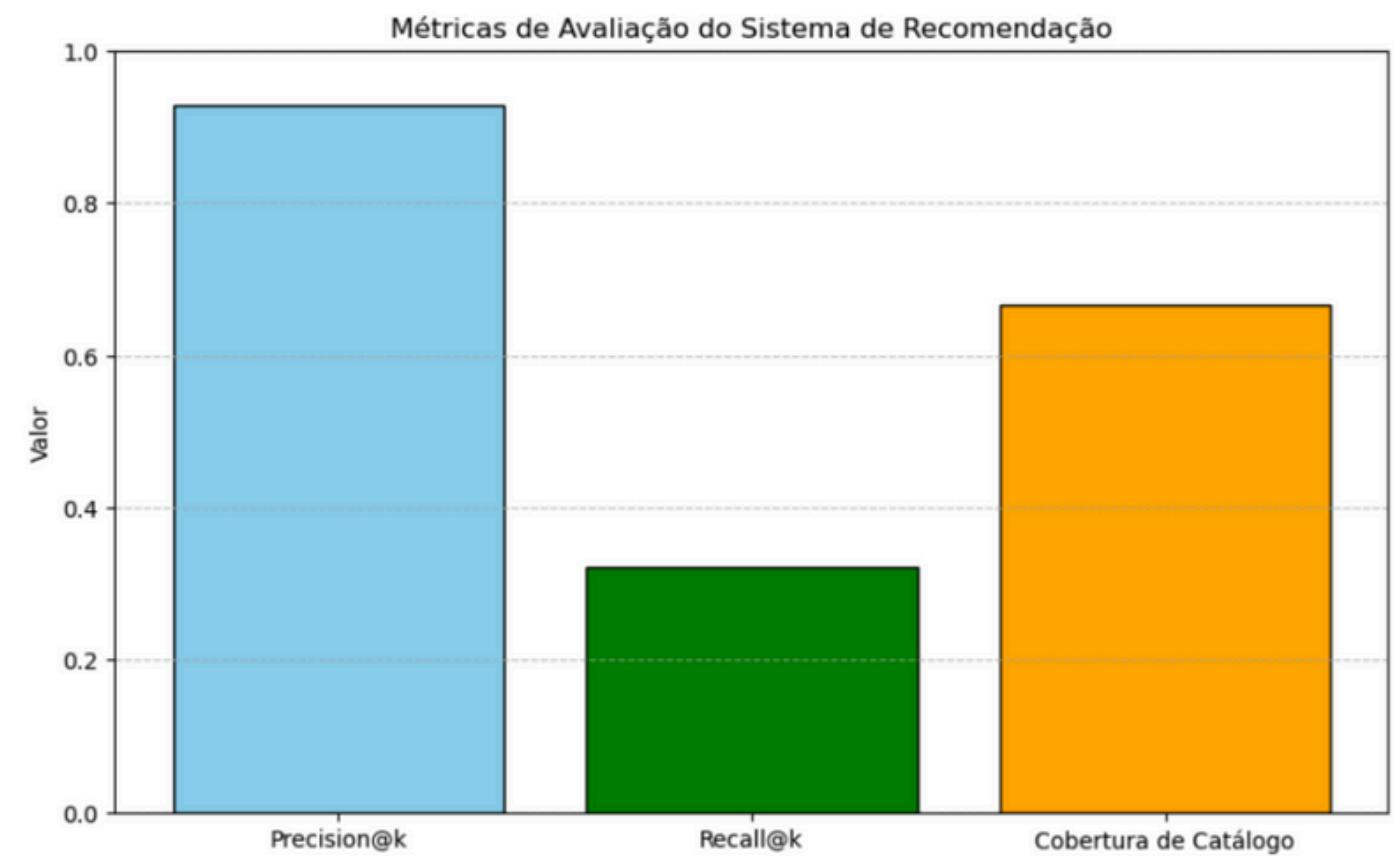




CONSOLIDAÇÃO DOS RESULTADOS

Um gráfico comparativo serve para que tenhamos uma visão geral das métricas e possamos compará-las de forma mais efetiva.

Vemos neste gráfico que a precisão está realmente acima da média e a distribuição e varredura das receitas está boa, poderia melhorar, no entanto ainda temos a limitação gerada pela falta de rating



CONCLUSÃO

Este trabalho teve como objetivo desenvolver um algoritmo de recomendação de receitas a partir dos itens comprados pelos usuários, visando reduzir o desperdício de alimentos e melhorar a nutrição. Neste sentido escolhemos fazer um algoritmo de recomendação de receitas a partir dos itens que o usuário comprou para reduzir o desperdício de comida e melhorar a nutrição das pessoas.

Para desenvolver este algoritmo utilizamos algumas bibliotecas e ferramentas de software livre para preparar os dados, através destas ferramentas o objetivo foi alcançado e o software foi desenvolvido com sucesso, tendo uma assertividade na relevância das recomendações superior a noventa por cento.

Contudo, enfrentamos desafios significativos, como a ausência de uma API para acesso aos dados das receitas e avaliações, o que limitou a utilização de algoritmos de recomendação mais avançados e impactou na avaliação de desempenho do nosso sistema. A falta de um rating nos dados iniciais nos obrigou a adotar métricas menos convencionais para validar a eficácia do sistema.

Concluímos analisando os resultados que, como melhorias em trabalhos futuros, o sistema precisa de ajustes para melhorar o recall, talvez aumentando o número de recomendações ou diminuindo os critérios de similaridade para inclusão de itens na lista de sugestões. Essa abordagem pode ajudar a equilibrar melhor a precisão e o recall, aumentando a abrangência das recomendações sem comprometer significativamente a relevância das mesmas.

REFERÊNCIAS

- **Git hub:** https://github.com/VaniaJesus/Proj_APLICADO-3
- ALEXANDER, N. Catered to your future self: Netflix’s “predictive personalization” and the mathematization of taste. In: MCDONALD, K.; SMITH - ROWSEY, D. (ed.). The Netflix effect:technology and entertainment in the 21st Century. Nova York: Bloomsbury Academic, 2016. p. 81-98.
- Algoritmo de recomendação: o que é e como ele contribui para as estratégias personalizadas. meio&mensagem. Disponível em: Algoritmo de recomendação: o que é e como eles funcionam? (meioemensagem.com.br). Acesso em: 20 de Março de 2024.
- Cesta básica nacional. Universidade Federal de Mato Grosso do Sul. Disponível em: [https://cpcs.ufms.br/custo-cesta-basica-setembro-2023/#:~:text=A%20cesta%20b%C3%A1sica%20de%20alimentos,%2C%20banha%2F%C3%B3leo%20e%20manteig a](https://cpcs.ufms.br/custo-cesta-basica-setembro-2023/#:~:text=A%20cesta%20b%C3%A1sica%20de%20alimentos,%2C%20banha%2F%C3%B3leo%20e%20manteiga). Acesso em: 04 de Março de 2024.
- Domine a Programação em Python com a Biblioteca Pandas: um Guia Completo. Awari. Disponível em: <https://awari.com.br/domine-a-programacao-em-python-com-a-biblioteca-pandas-um-guia-completo/#:~:text=0%20Pandas%20%C3%A9%20uma%20biblioteca%20pod%20que%20oferece%20uma%20variedade,partir%20dos%20conjuntos%20de%20dados> Acesso em: 08 de Março de 2024.
- LANDOW, G. Hipertexto: la convergencia de la teoría crítica contemporánea y la tecnología. Barcelona: Paidós, 1992.
- McKinney, W. (2012). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media.
- O que são Redes Neurais. Ibm. Disponível em:O que são Redes Neurais? | IBM. Acessado em: 08 de Março de 2024 34
- ODS. Nações Unidas Brasil. Disponível em: <https://brasil.un.org/pt-br/sdgs>. Acessado em: 08 de Março de 2024
- Receitas, Site Tudo Gostoso. Disponível em: <https://www.tudogostoso.com.br/>. Acesso em: 04 de Março de 2024.
- VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media.