

Relatório Trabalho de Álgebra

Felipe Steinfeld

November 2025

1 Introdução

O trabalho é um sistema de recomendação de pratos de comida da culinária brasileira, apresentando os similares utilizando o sistema tf-idf. O usuário escolhe um prato e recebe os três pratos mais similares ao escolhido.

A análise tf-idf é feita através da vetorização dos ingredientes dos pratos, assim é possível fazer uma análise matemática e descobrir prato mais semelhantes ao escolhido pelo usuário.

2 Tema

O sistema foi criado para mostrar a utilidade do sistema TF-IDF, e nesse caso como pode ajudar o cliente a encontrar pratos semelhantes ao desejado, focado em pratos da culinária brasileira.

2.1 Projeto

No projeto o usuário vai receber uma lista de pratos típicos brasileiros, cada um com um respectivo número. O usuário deve escolher um dos pratos digitando seu número no console, assim pode ver os 3 pratos mais similares. Após o resultado, o usuário pode escolher a opção de ver uma explicação matemática do que está acontecendo, então deve-se digitar 'Explicação' no console, ou apenas clicar na tecla 'enter' e assim voltando ao menu de opções. De volta ao menu pode-se escolher outro prato ou digitar 'sair' para finalizar o programa.

O resultado de cada pesquisa dá ao usuário os 3 pratos mais similares com os valores do cosseno (similaridade) e o valor do ângulo calculado. Cada resultado de similaridade entre os pratos é classificado de acordo com o valor do cosseno do ângulo entre seus vetores TF-IDF, seguindo os seguintes intervalos:

- Similaridade $\leq 0,33$: **Baixa**
- $0,33 < \text{Similaridade} \leq 0,66$: **Média**
- Similaridade $> 0,66$: **Alta**

3 DataSet

O dataset foi gerado pelo chatgpt. Contém 14 pratos de comida e seus respectivos ingredientes. São duas colunas, 'Prato' e 'Ingredientes', e 14 linhas com dados preenchidos.

Os nomes em 'negrito' pertencem a coluna 'Prato', e os itens que seguem cada tópico fazem parte da coluna 'Ingredientes'.

1. **Feijoada:** feijão preto, carne seca, costelinha, linguiça, arroz, farofa, couve.
2. **Bife a cavalo:** arroz, feijão, bife, ovo, batata frita, salada.
3. **Strogonoff de frango:** frango, creme de leite, ketchup, batata palha, arroz.
4. **Lasanha à bolonhesa:** massa, carne moída, molho de tomate, queijo, presunto.
5. **Moqueca de peixe:** peixe, leite de coco, pimentão, tomate, coentro, cebola.
6. **Escondidinho de carne seca:** carne seca, mandioca, queijo, manteiga.
7. **Feijão tropeiro:** feijão, farinha de mandioca, linguiça, torresmo, ovo.
8. **Galinhada:** arroz, frango, açafrão, cenoura, pimentão, alho.
9. **Bobó de camarão:** camarão, mandioca, leite de coco, azeite de dendê, cebola, coentro.
10. **Carne de panela com batata:** carne, batata, tomate, cebola, alho.
11. **Macarronada com almôndegas:** macarrão, almôndegas, molho de tomate, queijo.
12. **Peixe frito com arroz e salada:** peixe, arroz, salada, tomate, alface, limão.
13. **Virado à paulista:** arroz, tutu de feijão, bisteca, ovo, banana frita.
14. **Picadinho de carne:** carne, cebola, molho de soja, alho, arroz, farofa.

4 Metodologia

O sistema foi feito através da linguagem de programação **Python** com implementação do sistema TF-IDF e as bibliotecas: Pandas, Scikit-learn, Numpy.

5 Importância e Cálculo do TF-IDF

O modelo TF-IDF (*Term Frequency – Inverse Document Frequency*) é uma técnica amplamente utilizada em tarefas de mineração de texto e sistemas de recomendação baseados em conteúdo. No contexto deste trabalho, o TF-IDF é fundamental para representar os ingredientes de cada prato em formato numérico, permitindo medir a similaridade entre receitas com base na presença e relevância dos ingredientes.

5.1 Importância do TF-IDF neste trabalho

Cada prato é descrito por um conjunto de ingredientes, e diferentes receitas podem compartilhar alguns desses termos. O modelo TF-IDF permite quantificar a importância de cada ingrediente dentro de um prato, levando em consideração tanto sua frequência local (no prato) quanto sua raridade global (em todos os pratos do conjunto de dados).

Dessa forma, ingredientes muito comuns — como “arroz” ou “feijão” — recebem um peso menor, enquanto ingredientes mais específicos, como “azeite de dendê” ou “carne seca”, têm peso maior. Essa ponderação é essencial para o cálculo da similaridade, pois faz com que os pratos sejam comparados de maneira mais informativa, destacando as características que realmente os diferenciam.

5.2 Relação com a Similaridade de Cosseno

Após o cálculo do TF-IDF, cada prato é representado como um vetor numérico. A similaridade entre dois pratos é determinada pelo **cosseno do ângulo** entre esses vetores, conforme a Equação:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} \quad (1)$$

onde $A \cdot B$ representa o produto escalar entre os vetores dos pratos, e $\|A\|$ e $\|B\|$ são suas normas. O resultado varia de 0 a 1, indicando o grau de similaridade entre os pratos:

- Valores próximos de 1 indicam alta similaridade (ângulos pequenos, mais perto de 0°);
- Valores próximos de 0 indicam baixa similaridade (ângulos grandes, mais perto de 90°).

Dessa forma, o TF-IDF combinado com a similaridade de cosseno permite recomendar pratos com ingredientes mais semelhantes, reproduzindo uma lógica de sugestão natural em um restaurante de comida caseira.

6 Conclusão

A partir desse trabalho é possível concluir que o modelo TF-IDF é importante e necessário no desenvolvimento de sistemas de recomendação.