

Inteligência Artificial: Vegia

Renata Moura Barreto, estudante, UNIFESP, 2023

Resumo— Este trabalho propõe uma Inteligência Artificial que oferece receitas vegetarianas ou veganas para reduzir o consumo de carne, com a finalidade de solucionar os objetivos doze (Ação contra a mudança global do clima) e treze (Consumo e produção responsáveis) do Objetivo de Desenvolvimento Sustentável criado pela ONU (Organização das Nações Unidas).

I. INTRODUÇÃO

A atmosfera terrestre, composta principalmente por nitrogênio, oxigênio e argônio, atua como um "cobertor natural" ao formar o efeito estufa. Este fenômeno, vital para a vida na Terra, ocorre quando o sol emite radiações, parte das quais é refletida de volta ao espaço, enquanto outra é absorvida pela superfície terrestre, mantendo uma temperatura ideal. Contudo, as atividades humanas têm aumentado a emissão de gases estufa, resultando no aumento da fina camada atmosférica e na retenção excessiva de radiação infravermelha, levando ao fenômeno do aquecimento global. A preocupação acerca disto é tamanha que “tomar medidas urgentes para combater a mudança climática e seus impactos” se tornou um dos 17 objetivos globais traçados pela ONU para o Desenvolvimento Sustentável.

A pecuária de corte bovina, sendo um dos maiores setores do Brasil, contribui substancialmente para as emissões globais de gases de efeito estufa, representando 24% do total em 2017. O problema é agravado pelo desmatamento extenso, especialmente na região amazônica, utilizado para expansão de pastagens, representando uma forte agravante do aquecimento global. O sistema produtivo da pecuária é o maior responsável pela emissão de gases estufa no país, e a carne bovina está no prato de quase todos os brasileiros diariamente, sendo que alguns alimentos já são considerados parte da cultura nacional.

As atividades agropecuárias, como quaisquer outras atividades humanas, são potenciais poluidoras do meio ambiente, portanto devem ter o manejo ambiental como rotina a fim de conservar e preservar os recursos naturais.

Outro fator importante a ser considerado é o volume de água utilizado para se produzir a carne bovina. Os atores da agropecuária sabem a importância que a água tem para suas atividades, mas esse saber não tem se traduzido em manejos, ações, programas, etc., que busquem a eficiência de uso e a gestão do recurso. Além disso, a sustentabilidade do sistema agroalimentar não se limita a internalizar o conceito somente nos sistemas de produção, é preciso também fomentar e praticar o consumo sustentável. Desafios como esses foram pautados no objetivo 12 (Consumo e Produção Responsável) do Objetivo de

Desenvolvimento Sustentável (ODS) criado pela ONU, o que se demonstra de extrema importância para uma melhora do mundo.

Considerando esses empecilhos gerados pela atividade agropecuária e seu consumo, esse trabalho possui o objetivo de apresentar uma maneira de reduzir o consumo de carne a fim de gerar um consumo responsável e de consequentemente ajudar na redução do aquecimento global indiretamente. Para que isso seja possível, este trabalho irá apresentar a implementação de uma Inteligência Artificial (IA) capaz de mostrar diferentes receitas vegetarianas ou veganas a partir de alimentos que o usuário possui. Dessa forma, o desenvolvimento dessa IA e os seus resultados serão o foco principal do trabalho.

Para o desenvolvimento da IA foram implementados dois diferentes algoritmos de IA e um algoritmo que usa técnicas de IA, sendo que todos possuem a finalidade de recomendar receitas. Assim, para a confecção dessa IA, apelidada de Vegia, foram utilizados alguns trabalhos e postagens feitas em sites para programadores como base, sendo eles: a postagem do site “Kaggle” “food_com-vegetarian-analysis” feita por Philip Seif; o trabalho de conclusão de curso intitulado como “TCC_Lorenzo” encontrado no site “GitHub” feita por Lorenzo Benedetti; a postagem “Document Similarity Algorithms Experiment” de Masatoshi Nishimura, encontrada no site “GitHub”; a postagem “TF-recomm” do site “GitHub” feita por Guocong Song; a postagem “hybrid-rs-trainer” do site “GitHub” feita por Felipe Appio.

II. TRABALHOS RELACIONADOS

A postagem “food_com-vegetarian-analysis” do Philip Seif trata-se de uma análise de receitas vegetarianas e veganas feita a partir de um dataset de receitas. Nessa análise ele verifica dados como a quantidade de receitas vegetarianas e veganas a cada ano, as diferenças das classificações das receitas não vegetarianas e as vegetarianas, se existe uma tendência positiva no número de interações com receitas vegetarianas, entre outros. Dessa forma, algumas pesquisas de dados dessa análise foram implementadas neste trabalho.

O Trabalho de Conclusão de Lorenzo Benedetti consiste em apresentar uma análise sobre sistemas

de recomendação utilizando filtragem colaborativa baseado em memória (KNN) e baseado em modelo (SGD). Com o objetivo de investigar como essa técnica pode ser aplicada para melhorar a experiência do usuário e aumentar a precisão das recomendações, foram utilizados revisão bibliográfica e aplicação de algoritmos. Com isso, a utilização do modelo KNN foi implementada neste trabalho como um dos dois algoritmos citados anteriormente.

Já a postagem “Document Similarity Algorithms Experiment” feita por Masatoshi Nishimura busca encontrar qual algoritmo dentre os algoritmos Jaccard, TF-IDF, Doc2vec, USE e BERT possui melhor desempenho quando usados na similaridade de documentos.

A postagem “TF-recomm” feita por Guocong Song tem como objetivo mostrar que o TensorFlow é uma plataforma que pode ser utilizada em algoritmos de recomendação por oferecer algoritmos avançados de aprendizado por gradiente estocástico (SGD), aceleração para processamento em CPU/GPU e suporte para treinamento distribuído em clusters de computadores. Além disso, o TensorFlow inclui módulos de incorporação, facilitando a implementação de aplicações semelhantes ao word2vec.

A postagem “hybrid-rs-trainer” do Felipe Appio mostra uma forma de algoritmo de recomendação diferente das utilizadas neste trabalho. Ele utiliza o Sistema de Recomendação Híbrido sistemas híbridos que é uma abordagem que possui características ou combinam as abordagens de Filtragem Colaborativa e Baseado em Conteúdo para fazer recomendações de itens para usuários.

III. METODOLOGIA

Como o objetivo deste trabalho é gerar receitas vegetarianas e veganas, para acessar essas receitas foi utilizado o mesmo dataset utilizado na postagem do Philip Seif, porém este trabalho não fez apenas uma análise das receitas ele utilizou as receitas para recomendação. Ademais, esse dataset contém receitas vegetarianas e não vegetarianas, logo, as receitas não vegetarianas foram retiradas para uma melhor análise das receitas vegetarianas e veganas. Além disso, outros tratamentos de dados foram feitos para melhorar a procura dentro do dataset.

Outro fator a pontuar, é que diferentemente dos trabalhos relacionados a este, a IA feita é mais humanizada, ela possui um linguajar mais cotidiano e se chama Vegia, como dito anteriormente. Com isso, o usuário que a usar se sentirá mais acolhido em seu uso. A Vegia perguntará os ingredientes para o usuário e assim, irá recomendar cinco receitas com base nesses ingredientes.

Para o funcionamento da Vegia foram implementadas três diferentes formas de recomendação que serão analisadas e comparadas ao longo do trabalho. Essas implementações são: um algoritmo utiliza a técnica de vetorização TF-IDF para representar os ingredientes de receitas como uma matriz e aplica o algoritmo k-Nearest Neighbors (k-NN) com uma abordagem de “força bruta” e distância de cosseno para encontrar os 5 vizinhos mais próximos com base nos ingredientes fornecidos pelo usuário; um algoritmo que cria matriz TF-IDF utilizando um vetorizador e processamento de linguagem natural (PLN) e calcula a similaridade de Jaccard entre os ingredientes do usuário e os das receitas na matriz TF-IDF; um algoritmo que treina um modelo Word2Vec e usa uma rede neural sequencial para calcular a similaridade entre os ingredientes do usuário e as receitas usando a similaridade de cosseno.

IV. ANÁLISE EXPERIMENTAL

A. Conjunto de Dados

O dataset está no formato Comma-separated values (CSV) utilizado consiste em um conjunto de receitas que foram retiradas do agregador de receitas on-line “Food.com”. As receitas se encontram em Inglês e suas informações estão divididas em doze colunas, sendo elas: name (nome das receitas), id (identificação das receitas), minutes (tempo de preparo das receitas), contributor_id (identificação do contribuinte), submitted (ano em que a receita foi postada), tags (características da receita), nutrition (tabela nutritiva), n_steps (quantidade de passos da receita), steps (passos da receita), description (descrição da receita, ingredients (ingredientes utilizados) e n_ingredients (quantidade de ingredientes).

B. Configuração do Algoritmo e do Sistema Computacional

A linguagem de programação escolhida foi Python na versão 3.10, para a compilação do código foi utilizado o Google Colaboratory (Colab) no navegador Microsoft Edge e a implementação do código foi feita em Inglês com comentários em Português.

Especificações do Sistema Computacional:

- CPU: Intel(R) Core(TM) i5
- Clock: 2.40GHz
- Núcleos: 4
- trends: 8
- Cache: 8,0 MB
- Memória: 8,0 GB 3200 MHz
- GPU: NVIDIA GeForce MX350
- Memória da GPU Dedicada: 2,0 GB
- Sistema Operacional: Windows 11 Home - 22H2 - x64

As bibliotecas que estão sendo utilizadas são: NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, Time, Gensim.models, TensorFlow.keras.models e TensorFlow.keras.layers.

O algoritmo foi dividido em partes para uma melhor análise, sendo elas e suas explicações:

- Biblioteca: todas as bibliotecas do código estão nessa parte;
- Leitura do dataset: leitura do dataset “recipes.csv”;
- Análise dos Dados: análise dos dados do dataset;
- Manipulando dados: acontece o tratamento dos dados
- Input: interação da Vegia com o usuário e entrada dos ingredientes do usuário;
- Algoritmos de recomendação: essa parte possui as três formas de recomendação e alguns comparativos.

C. Critérios de Análise

Como dito anteriormente, o código possui uma parte de análise apenas do dataset e as informações que foram analisadas são: as cinco primeiras receitas, para verificar quais são os dados que estão na coluna e como eles estão dispostos; os tipos de cada coluna e a quantidade de receitas disponíveis;

A parte da manipulação dos dados também possui algumas análises, como a quantidade de receitas veganas e vegetarianas, o tempo que demora as receitas e a quantidade de receitas após alguns tratamentos.

Para a análise dos algoritmos foram utilizadas diferentes entradas de ingredientes, para saber qual algoritmo conseguiu sugerir mais receitas que obtivessem as entradas e também foi utilizado como parâmetro o tempo decorrido para compilar cada algoritmo.

D. Resultados e Discussões

D.1 Dados

De acordo com o resultado da análise das cinco primeiras receitas (figura 1) é possível notar que as receitas possuem nomes não muito usuais.

	name	id	minutes	contributor_id	submitted	tags	nutrition	n_steps	steps	description	ingredients	n_ingr
0	arriba baked winter squash mexican style	137739	55	47892	2005-09-16	['160-minutes-or-less', 'time-to-make', 'course...']	[51.5, 0.0, 13.0, 0.0, 2.0, 0.0, 4.0]	11	['make a choice and proceed with recipe', 'step...']	autumn is my favorite time of year to cook! th...	['winter squash', 'mexican seasoning', 'mixed ...']	
1	a bit different breakfast pizza	31490	30	26278	2002-05-17	['30-minutes-or-less', 'time-to-make', 'course...']	[173.4, 16.0, 0.0, 17.0, 22.0, 35.0, 1.0]	9	['preheat oven to 425 degrees f', 'press dough...']	this recipe calls for the crust to be prebaked...	['prepared pizza crust', 'sausage patty', 'egg...']	
2	all in the kitchen chili	112140	130	196586	2005-02-25	['time-to-make', 'course', 'preparation', 'meal...']	[269.8, 22.0, 32.0, 48.0, 39.0, 27.0, 5.0]	6	['brown ground beef in a large pot', 'add choppes...']	this modified version of 'mom's' chili was a h...	['ground beef', 'yellow onion', 'sliced tomato...']	
3	alouette potatoes	59389	45	68585	2003-04-14	['30-minutes-or-less', 'time-to-make', 'course...']	[368.1, 17.0, 10.0, 2.0, 14.0, 6.0, 20.0]	11	['place potatoes in a large pot of lightly salt...']	this is a super easy great tasting, make ahead...	['prestable cheese with garlic and herbs', 'h...']	
4	amish tomato ketchup for	44061	190	41706	2002-10-25	['weeknight', 'time-to-make', 'course']	[352.9, 1.0, 337.0, 23.0, 3.0]	5	['mix all ingredients', 'boil for 2 1/2 h...']	my dh's amish mother raised him on	['tomato juice', 'apple cider vinegar']	

Fig. 1. Cinco primeiras receitas.

Além disso, a partir do resultado da segunda parte da análise de dados (figura 2) é visível que o

dataset possui doze colunas, 231637 receitas e que a maioria dos tipos dos dados das colunas são objetos.

```
RangeIndex: 231637 entries, 0 to 231636
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   name                   231636 non-null object
1   id                     231637 non-null int64
2   minutes                231637 non-null int64
3   contributor_id         231637 non-null int64
4   submitted              231637 non-null object
5   tags                   231637 non-null object
6   nutrition              231637 non-null object
7   n_steps                231637 non-null int64
8   steps                  231637 non-null object
9   description            226658 non-null object
10  ingredients            231637 non-null object
11  n_ingredients          231637 non-null int64
dtypes: int64(5), object(7)
memory usage: 21.2+ MB
None
```

Fig. 2. Resultado da segunda parte da análise de dados.

Após a retirada das receitas não vegetarianas foi obtido que o dataset possui 35651 receitas não vegetarianas (figura 3).

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 231637 entries, 0 to 231636
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   name                   231636 non-null object
1   id                     231637 non-null int64
2   minutes                231637 non-null int64
3   contributor_id         231637 non-null int64
4   submitted              231637 non-null object
5   tags                   231637 non-null object
6   nutrition              231637 non-null object
7   n_steps                231637 non-null int64
8   steps                  231637 non-null object
9   description            226658 non-null object
10  ingredients            231637 non-null object
11  n_ingredients          231637 non-null int64
dtypes: int64(5), object(7)
memory usage: 21.2+ MB
None
```

Fig. 3. Receitas vegetarianas e veganas.

Com a geração de um gráfico de dispersão do tempo de confecção da receita (figura 4), é notório que existem receitas que possuem mais de um dia (1440 minutos) de duração.

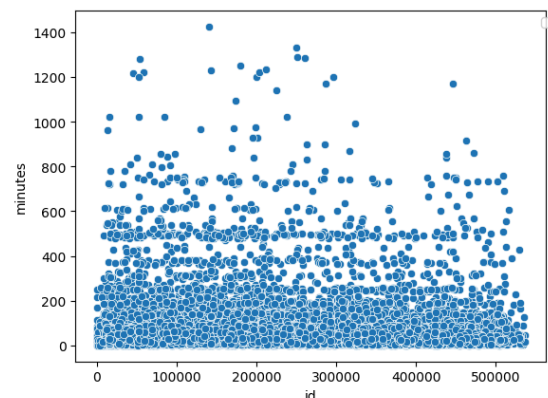


Fig. 4. Gráfico de dispersão.

Com a retirada das receitas que demoram mais de 1 dia para serem feitas, sobraram 35323 receitas (figura 5).

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 35323 entries, 0 to 231633
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   name                   35323 non-null  object
1   id                     35323 non-null  int64
2   minutes                35323 non-null  int64
3   contributor_id         35323 non-null  int64
4   submitted              35323 non-null  object
5   tags                   35323 non-null  object
6   nutrition               35323 non-null  object
7   n_steps                35323 non-null  int64
8   steps                  35323 non-null  object
9   description             34696 non-null  object
10  ingredients             35323 non-null  object
11  n_ingredients           35323 non-null  int64
dtypes: int64(5), object(7)
memory usage: 3.5+ MB
None
```

Fig. 5. Receitas que demoram menos de 1 dia.

D.2 Algoritmos de Recomendação

Para todas as entradas testadas o algoritmo que utiliza KNN e o algoritmo que usa o PLN e a similaridade de Jaccard obtiveram as mesmas receitas recomendadas, porém o tempo de compilação do algoritmo que usa o PLN teve o tempo de execução maior em todos os casos.

Para a entrada: cheese, corn, beans, carrot, water, butter o resultado das receitas recomendadas pelo KNN e pelo que utiliza PNN foram:

1. simply corn: possui três ingredientes das entradas mais 2 ingredientes que não estão na entrada;
2. kittencal s sweet cooked corn removed from the cob: possui 4 ingredientes mais um ingrediente que não está na entrada;
3. crook pot corn on the cob: possui apenas 2 ingredientes da entrada;
4. best glazed carrots: possui 3 ingredientes da entrada;
5. sweet corn for the freezer: possui 4 ingredientes da entrada mais 1 ingrediente que não está na entrada.

Além disso, o tempo de execução deles foi de 1.16 segundos e 2.51 segundos, respectivamente. Já o algoritmo que utiliza Rede Neural, teve o 55.02 segundos de tempo de execução e como receitas recomendadas:

1. mushroom burgers: possui apenas 1 ingrediente da entrada mais 10 ingredientes que não estão na entrada;
2. olive and cheese wonton bites: possui apenas 1 ingrediente da entrada mais 8 ingredientes que não estão na entrada;

3. monterey jack queso fiesta corn: possui 2 ingredientes da entrada mais 8 ingredientes que não estão na entrada;

4. vegetarian shepard s pie: possui 4 ingredientes da entrada e 12 ingredientes que não estão na entrada;

5. vegetarian cabbage rolls: possui 3 ingredientes da entrada e 10 ingredientes que não estão na entrada.

Para a entrada: “chocolate, banana, milk, coffee, lemon” o resultado das receitas recomendadas pelo KNN e pelo que utiliza PNN foram:

1. sweet banana snack: possui 2 ingredientes da entrada mais 1 ingrediente que não está na entrada;
2. mexican hacienda coffee: possui 2 ingredientes da entrada mais 5 que não estão na entrada;
3. frozen banana cappuccino: possui 3 ingredientes da entrada mais 2 que não estão na entrada;
4. chocolate milk and banana smoothie: possui 2 ingredientes da entrada mais 1 ingrediente que não está na entrada;
5. italian mocha bicerin: possui 1 ingrediente que está na lista e 1 ingrediente que não está.

Além disso, o tempo de execução deles foi de 0.62 segundos e 1.29 segundos, respectivamente. Já o algoritmo que utiliza Rede Neural, teve o 31.16 segundos de tempo de execução e como receitas recomendadas:

1. white choc brandy sauce: possui 1 ingrediente que está na entrada e 2 que não estão;
2. super quick exotic rose ice cream: não possui nenhum ingrediente de entrada e 8 ingredientes que não estão;
3. ruth s coffee ruth s chris steak house recipe: possui 1 ingrediente que está na entrada e 4 que não estão;
4. crook pot mocha hot spot: possui 3 ingredientes que estão na entrada e 3 que não estão;
5. double peppermint bark: possui 1 ingrediente que está na entrada e 3 que não estão;

Para a entrada: “water, salt, oil, apple” o resultado das receitas recomendadas pelo KNN e pelo que utiliza PNN foram:

1. super healthy no fat smoothie: possui 2 ingredientes que estão na entrada e 1 que não está;
2. apple cider sorbet: possui 2 ingredientes que estão na entrada e 1 que não está;
3. apple bites: possui 1 ingrediente que está na entrada e 1 que não está;

4. sore and strep throat away gargle: possui 2 ingredientes que estão na entrada;

5. apple spiced rice: possui 3 ingredientes que estão na entrada e 4 que não estão;

Além disso, o tempo de execução deles foi de 1.16 segundos e 2.48 segundos, respectivamente. Já o algoritmo que utiliza Rede Neural, teve o 31.23 segundos de tempo de execução e como receitas recomendadas:

1. toasted sunflower seeds: possui 2 ingredientes que estão na entrada e 1 que não está;

2. oat and cashew waffles: possui 3 ingredientes que estão na entrada e 2 que não estão;

3. mum s pizza base: possui 3 ingredientes que estão na entrada e 1 que não está;

4. vegan graham cracker crust for vegan cheesecake: possui 3 ingredientes que estão na entrada e 2 que não estão;

5. pol roti coconut flatbread: possui 3 ingredientes que estão na entrada e 2 que não estão;

V. CONCLUSÃO

Infere-se, portanto, que segundos as análises feitas, o algoritmo de recomendação que utiliza KNN obteve as mesmas receitas do algoritmo que utiliza o PLN e a similaridade de Jaccard. No entanto, o que usa o KNN teve um tempo de execução menor em todos os casos, o que foi uma surpresa, porque o algoritmo do KNN foi mais complexo de implementar por não ser tão comumente usado como um algoritmo de recomendação, logo foi esperado que eles obtivessem um tempo de execução parecido. Ademais, em todos os testes eles obtiveram boas sugestões de receitas, pois as receitas recomendadas utilizavam bastante ingredientes da entrada se comparado com a quantidade de ingredientes na entrada.

Em segunda instância, a Rede Neural obteve poucos resultados satisfatórios. Como o esperado de um algoritmo desse, o seu tempo de execução foi considerável em todos os casos testados, porém ele foi aproximadamente 10 vezes maior que o tempo do outros algoritmos. Mais, nas entradas “cheese, corn, beans, carrot, water, butter” e “chocolate, banana, milk, coffee, lemon” seus resultados foram inferiores aos dos demais algoritmos, na maioria das receitas sugeridas havia apenas um ingrediente da entrada. Entretanto, no teste em que as entradas eram “water, salt, oil, apple” ele obteve resultados melhores que os outros, em 4 receitas sugeridas utilizavam 3 dos 4 ingredientes da entrada e poucos ingredientes que não estavam.

Nesse contexto, o algoritmo que melhor se adequaria a IA Vegia seria o KNN, por obter

resultados de receita melhores que os demais algoritmos e por ser mais veloz no tempo de execução.

VI. REFERÊNCIAS BIBLIOGRÁFICAS

[1] F. M. Silva, “Pecuária bovina de corte brasileira: sua contribuição para o aquecimento global nos últimos 20 anos e o desrespeito ao art. 170, VI da constituição Federal”, 2019.

[2] “Consumo e produção responsáveis”, Embrapa.br. [Online]. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/184241/1/ODS-12-Consumo-e-producao-responsaveis.pdf>. [Acesso em: 01-dez-2023].

[3] T. T. de Almeida Silva, “PECUÁRIA BOVINA DE CORTE BRASILEIRA: SUA CONTRIBUIÇÃO PARA O AQUECIMENTO GLOBAL NOS ÚLTIMOS 20 ANOS E O DESRESPEITO AOS PRINCÍPIOS AMBIENTAIS CONSTITUCIONAIS”, Cidp.pt. [Online]. Disponível em: https://www.cidp.pt/revistas/rjlb/2020/1/2020_01_1285_1317.pdf. [Acesso em: 01-dez-2023].

[4] M. Santana, “Deep Learning para Sistemas de Recomendação (Parte 1) — Introdução”, Data Hackers, 06-dez-2018. [Online]. Disponível em: <https://medium.com/data-hackers/deep-learning-par-a-sistemas-de-recomenda%C3%A7%C3%A3o-part-e-1-introdu%C3%A7%C3%A3o-b19a896c471e>. [Acesso em: 01-dez-2023].

[5] M. Valmor, “Utilização de Redes Neurais em Sistemas de Recomendação”, Micreiros.com. [Online]. Disponível em: <https://micreiros.com/utilizacao-de-redes-neurais-e-m-sistemas-de-recomendacao/>. [Acesso em: 01-dez-2023].

[6] L. Benedetti, “TCC_Lorenzo”. [Online]. Disponível em: https://github.com/LorenzoSbenedetti/TCC_Lorenzo/tree/TCC_Lorenzo. [Acesso em: 03-dez-2023].

[7] M. Nishimura, “document_similarity_algorithms_experiments”. [Online]. Disponível em: https://github.com/massanishi/document_similarity_algorithms_experiments. [Acesso em: 03-dez-2023].

[8] G. Song, “TF-recomm”. [Online]. Disponível em: <https://github.com/songgc/TF-recomm>. [Acesso em: 04-dez-2023].

[9] A. Felipe “hybrid-rs-trainer”. [Online]. Disponível em: <https://github.com/SeniorSA/hybrid-rs-trainer>. [Acesso em: 4-dez-2023].

[10] S. Philip, “Food_com-vegetarian-analysis”, Disponível em: <https://www.kaggle.com/code/seifip/food-com-vegetarian-analysis/notebook>. [Acesso em: 30-nov-2023].