

CERVEJA E MACHINE LEARNING: RECOMENDANDO ESTILOS DE CERVEJA

BEER AND MACHINE LEARNING: RECOMMENDING BEER STYLES

CERVEZA Y APRENDIZAJE AUTOMÁTICO: RECOMENDANDO ESTILOS DE CERVEJA

Diogo Costa Pereira¹

RESUMO: Este trabalho teve como objetivo resolver o problema enfrentado por algumas pessoas que não conseguem apreciar cervejas especiais devido à falta de orientação sobre o que escolher. Para alcançar esse objetivo, foi realizado estudos que envolveram a distância euclidiana e o método Term Frequency-Inverse Document Frequency (TF-IDF) para encontrar similaridades entre os estilos, tendo como base um estilo específico. Além disso, explorou-se a vastidão do cenário cervejeiro no Brasil e foi apresentado o guia BJCP, que além de ser o objeto base do estudo, serviu como um recurso direcional para compreender melhor os estilos de cerveja. Por meio de três testes, foram validados os métodos propostos e desenvolvido uma ferramenta capaz de sugerir estilos de cerveja com base em um estilo referencial.

Palavras-chave: Estilos de Cerveja. Inteligência Artificial. Machine Learning. Distância Euclidiana. TF-IDF.

ABSTRACT: This work aimed to solve the problem faced by some people who struggle to appreciate special beers due to a lack of guidance on what to choose. To achieve this goal, studies were conducted involving Euclidean distance and the Term Frequency-Inverse Document Frequency (TF-IDF) method to find similarities between beer styles, based on a specific style. Additionally, the vast Brazilian beer landscape was explored, and the BJCP guide, which served as the study's foundation, was presented as a directional resource for better understanding beer styles. Through three tests, the proposed methods were validated, resulting in the development of a tool capable of suggesting beer styles based on a reference style.

Keywords: Beer Styles. Artificial Intelligence. Machine Learning. Euclidean Distance. TF-IDF.

RESUMEN: Este trabajo tuvo como objetivo resolver el problema que enfrentan algunas personas al no poder apreciar cervezas especiales debido a la falta de orientación sobre qué elegir. Para lograr este objetivo, se realizaron estudios que involucraron la distancia euclidiana y el método de Frecuencia de Término-Inverso de Frecuencia de Documento (TF-IDF) para encontrar similitudes entre los estilos de cerveza, basados en un estilo específico. Además, se exploró la vastedad del panorama cervecero brasileño y se presentó la guía BJCP, que además de ser el objeto base del estudio, sirvió como recurso direcional para comprender mejor los estilos de cerveza. A través de tres pruebas, se validaron los métodos propuestos, lo que resultó en el desarrollo de una herramienta capaz de sugerir estilos de cerveza basados en un estilo de referencia.

Palabras clave: Estilos de Cerveja. Inteligencia Artificial. Aprendizaje Automático, Distancia Euclidiana, TF-IDF.

¹Graduado em Análise e Desenvolvimento de Sistemas (IFTM – Campus Patrocínio), graduado em Gestão Financeira (UNICESUMAR), pós-graduado em Ciência de Dados (Faculdade FOCUS), pós-graduado em Business Intelligence, Big Data e Inteligência Artificial (Faculdade FOCUS), pós-graduado em Tecnologia de Produção Cervejeira (CLARETIANO), pós-graduado em Marketing (UNICESUMAR) e pós-graduado em Consultoria Empresarial: ênfase em R.H (UNICERP).

I. INTRODUÇÃO

Segundo as informações observadas no MAPA (2024), tanto o número de cervejarias registradas quanto o volume de produção de cervejas no Brasil, veem aumentando a cada ano. Apesar deste crescimento, a maioria dos estilos fabricados possuem em comum alguns poucos estilos, sendo o maior volume a família das cervejas do tipo *Lager*.

O Surra de Lúpulo (2024), elaborou uma pesquisa de âmbito nacional no qual desenvolveu *insights* e interpretações dos consumidores de cervejas no país, onde contou com 8.734 respostas. Um dos pontos apresentados no trabalho foi que 61,45% dos consumidores possuem amigos que bebem cervejas especiais e que eles mesmos não bebem. Um total de 32,71% afirmou não gostar da bebida e 23,39% dizem não saber escolher para começar a experimentar.

Diante disso, o presente trabalho possui como objetivo elaborar um estudo sobre como resolver o problema das pessoas que não sabem por onde começar a degustar cervejas especiais, ou seja, cervejas diferentes das convencionais, seja de estilos diferentes e/ou importadas.

Para o desenvolvimento deste projeto de pesquisa, foi inicialmente levantado o referencial teórico, que inclui trabalhos acadêmicos, revistas científicas, livros e outros recursos essenciais com embasamento científico.

Após entendermos sobre a complexidade que envolve as cervejas e os métodos utilizados na tentativa de resolver o problema, foi elaborado uma ferramenta capaz de sugerir estilos de cervejas a partir de um estilo inicial.

2. CONCEITOS IMPORTANTES

Para ater-se às possíveis respostas do objetivo do presente artigo, é de suma importância compreender os principais conceitos relacionados a cerveja e *Machine Learning*.

2.1. CERVEJA

A cerveja possui registros de fabricação datados a mais de 6 mil anos atrás, podendo ser um dos fermentáveis mais antigos do mundo e, sendo uma bebida muito popular, está entre as mais consumidas de todo o mundo, ficando atrás apenas do café, do leite, do chá e da água (PIMENTA, 2020).

Tivemos diversos momentos na história que a cerveja mostrou sua importância, como pode ter sido o real motivo da Revolução Neolítica, alimentava os povos antigos, foi

um produto comercial especial sustentando navios mercantes, proporcionou a troca de cultura nas rotas de comércio, incentivou inovações tecnológicas da Revolução Industrial entre outras (AMORIM, 2022).

Oliver (2012) fala que o método mais simples de produzir cerveja engloba a cevada maltada, água, lúpulo e levedura. Mas, o autor ainda enfatiza que para o mestre cervejeiro, esse processo é ainda mais complexo, pois são diversos os tipos de maltes a serem escolhidos, muitos outros cereais torrados, outro leque de cereais não maltados, tipos de açúcares, dezenas de espécies de lúpulos e ainda uma centena de variedades de leveduras, e após, produzir a cerveja exatamente conforme idealizado.

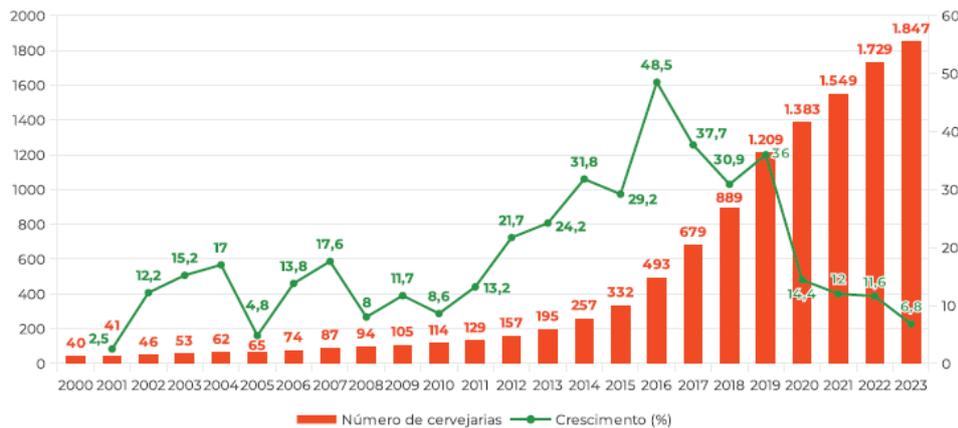
De acordo com a legislação brasileira, a cerveja é uma bebida decorrente de uma fermentação, através da levedura cervejeira, de um mosto de cevada maltada ou de seu respectivo extrato, submetido previamente por meio de um processo de cocção adicionando lúpulo ou seu respectivo extrato. Ela ainda ressalva que parte do malte de cevada (grão/extrato) podem ser substituídos, não totalmente, por adjunto cervejeiro (BRASIL, 2019).

De acordo com a CERVBRASIL (2016), o setor cervejeiro é de fundamental importância no desenvolvimento socioeconômico no Brasil, pois está presente em todas as cidades do país, interligando uma gigantesca rede produtiva que se estende desde o agronegócio até o pequeno varejo.

Quando falamos quantitativamente do setor cervejeiro brasileiro, o país é o 3º maior produtor do mundo, com uma produção anual de 15,1 bilhões de litros, sendo responsável por 2% do PIB, pagando R\$ 49,6 bilhões em impostos no ano e, responsável por mais de 2 milhões de empregos diretos, indiretos e induzidos (SURRA DE LÚPULO, 2024).

O ano de 2023 marcou um importante crescimento na indústria cervejeira do Brasil. O número de cervejarias registradas alcançou um total impressionante de 1.847, representando um acréscimo de 118 novos registros, o equivalente a 6,8% a mais em comparação com o ano anterior. Para colocar em perspectiva, isso significa que existe uma cervejaria para cada 109.952 habitantes no país (MAPA, 2024). No gráfico 1 ilustra claramente esse desenvolvimento.

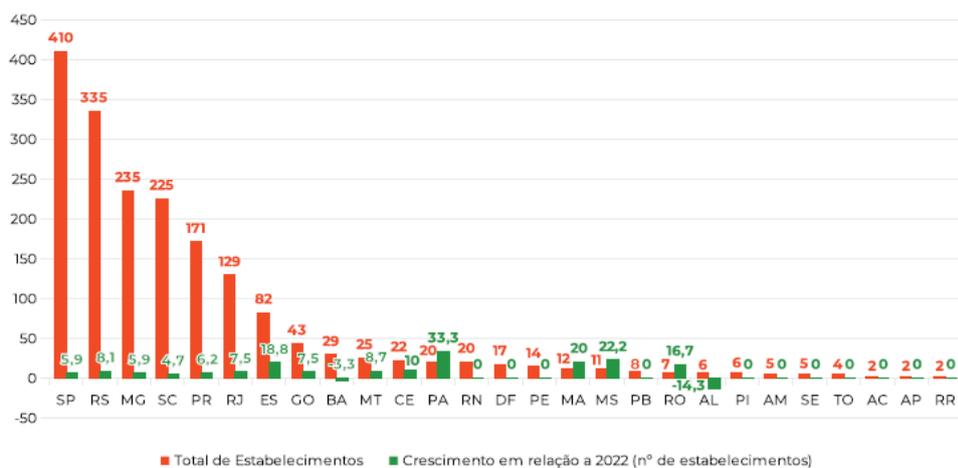
Gráfico 1 - Evolução do número de cervejarias registradas no Brasil.



Fonte: MAPA, 2024.

Ainda segundo o MAPA, 2024, as regiões brasileiras sul e sudeste detêm a maior concentração de cervejaria registradas no país, sendo um total de 1.533 registros, ou seja, 82,99% de todas as indústrias do setor. Na outra ponta, com o menor número de cervejarias registradas, está a região norte com um total de 42 registros, são 2,27% do todo de registros no país (MAPA, 2024). O gráfico 2 demonstra o total de cervejarias registradas por unidade da Federação.

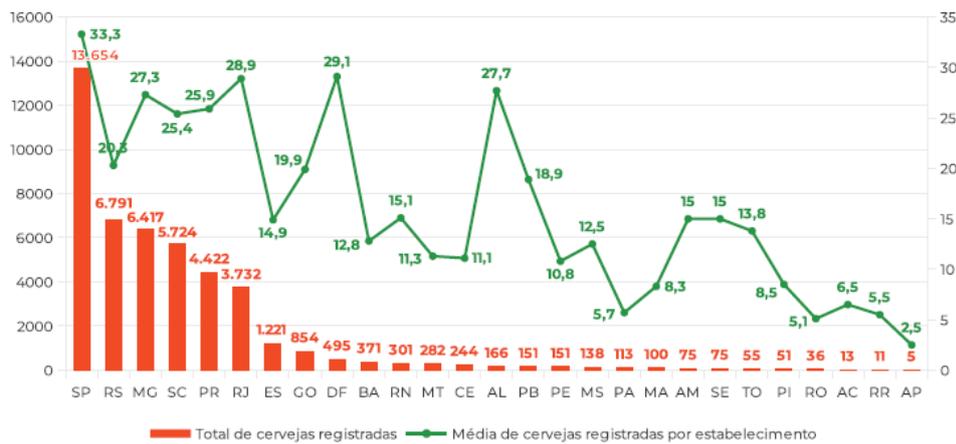
Gráfico 2 - Total de registros de cervejarias por unidade da Federação.



Fonte: MAPA, 2024.

Com todo esse número de registros de cervejarias, o mesmo autor ainda enfatiza que é necessário realizar o registro dos produtos que se pretende trabalhar. Com isso, no final de 2023, fechou com um total de 45.648 registros de cervejas (fórmulas) como um todo. Foram 2.817 registros a mais em relação ao ano de 2022. O estado de São Paulo detém o maior número de registros de produtos, um total de 13.654 fórmulas registradas. O gráfico 3 ilustra os registros de produtos por unidade da Federação.

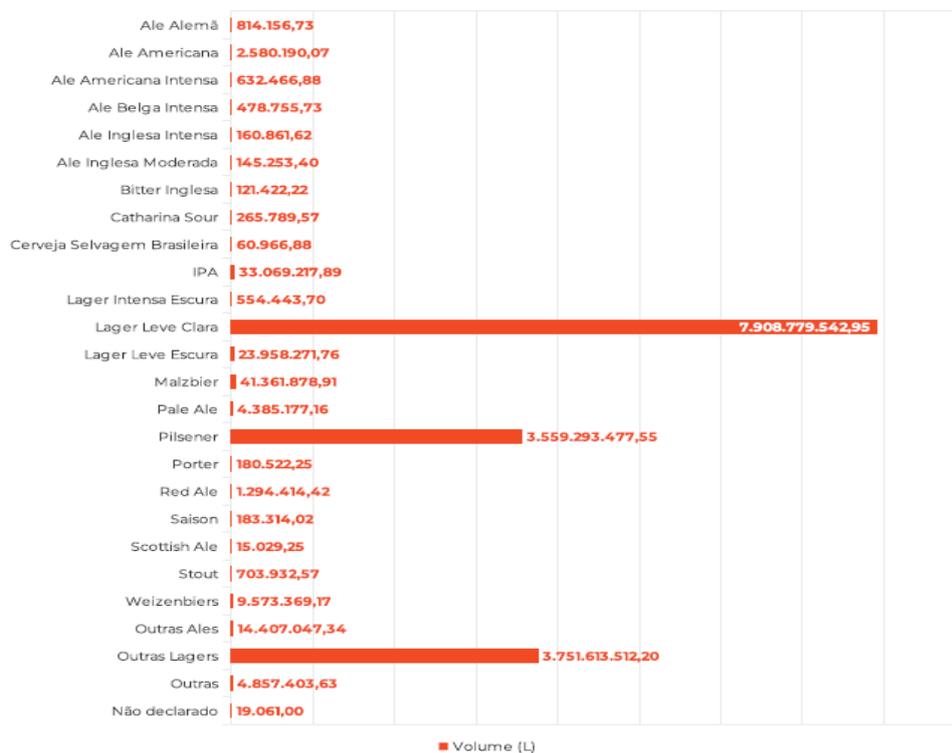
Gráfico 3 - Total de produtos registrados por unidade da Federação.



Fonte: MAPA, 2024.

Anualmente, em janeiro, as cervejarias são obrigadas a apresentar uma declaração de produção para o ano subsequente. Em 2023, foi registrado um impressionante volume de 15.361.344.112,77 litros de cerveja produzidos. Desse total, 29,2% foram atribuídos à categoria Puro Malte, que não contém adição de adjuntos cervejeiros. No que diz respeito ao tipo de fermentação, a esmagadora maioria, 99,5%, foi do tipo Lager, deixando uma pequena margem para os tipos Ale, de fermentação mista e os não declarados (MAPA, 2024). O gráfico 4 ilustra a produção de cervejas, categorizada por diversos estilos da bebida.

Gráfico 4 - Volume de produção de acordo com estilos de cerveja.



Fonte: MAPA, 2024.

Morado (2024), diz que classificar as cervejas pelo tipo de fermentação (Ale, Lager ou Lambic) não tem muita serventia, visto que não é um trabalho fácil, compreende diversos parâmetros que podem indicar essa classificação, como teor alcoólico, cor, ingredientes, origem do estilo, métodos de produção entre outros. Além de disso, outros fatores podem dificultar a classificação de cervejas, como novas tecnologias, tipo de matéria prima disponível, mudanças de paladar dos consumidores e etc.

Existem vários guias que fornecem orientações valiosas para a classificação dos estilos de cerveja. Entre eles, destacam-se o BJCP (Beer Judge Certification Program) e o BA (Brewers Association). Esses guias são ferramentas essenciais para entender a diversidade e a complexidade dos estilos de cerveja. Para os propósitos deste trabalho, vamos nos concentrar principalmente no guia BJCP como nossa referência principal.

2.1.1. *Beer Judge Certification Program (BJCP)*

O guia *Beer Judge Certification Program* ou simplesmente BJCP, é um guia que contempla 34 grupos de cervejas com um total de 123 estilos reconhecidos, além de possuir mais 5 estilos locais, ou seja, estilos promissores para entrarem oficialmente nos grupos de estilos futuramente. De acordo com o BJCP (2021), o guia é um recurso que descreve as principais características e alguns importantes exemplares de cervejas para termos uma ideia do estilo exemplificado. Contudo, ressalva que o guia não é um conjunto de especificações rígidas que devem ser seguidas à risca, mas sim um conjunto de sugestões que permitem uma certa flexibilidade. O guia ainda diz que nem todos estilos possíveis de cervejas estão presentes, sejam por não possuírem ainda um material de pesquisa suficiente, ou são pouco populares ou ainda estilos obscuros demais.

BJCP (2021) fala que o guia se utiliza de sessões como formato padrão para descrever cada um dos estilos apresentados, são eles:

- **Impressão Geral:** é descrito a individualidade de um estilo de cerveja, demonstrando suas características distintas.
- **Aparência, Aroma, Sabor, Sensação na Boca:** se concentram as percepções sensoriais, sendo os pilares que formam e definem o estilo da cerveja.
- **Comentários:** detalhes interessantes ou algumas notas adicionais sobre o estilo que não afeta sua avaliação sensorial.
- **História:** um breve descritivo histórico sobre o estilo.

- **Ingredientes:** não se trata de uma receita para produzir o estilo, e sim os principais ingredientes ou até processos típicos e comuns que englobam o caráter característico do estilo.
- **Comparação de Estilos:** demonstra o estilo, talvez nem tão conhecido, comparado a alguns outros para facilitar a sua compreensão. Nem todos os estilos possuem esta sessão descrita.
- **Instruções para Inscrição:** destaca as informações essenciais para que os juízes avaliem as inscrições em uma competição.
- **Estatísticas Vitais:** são diretrizes, que possuem uma faixa superior e outra inferior, que auxiliam no julgamento ou identificações. Complementando, Morado (2017), fala destas características vitais, sendo:
 - **OG:** *original gravity* ou densidade original, é um indicador que mede a densidade do mosto em relação a da água. Demonstra a quantidade de açúcares fermentáveis e não fermentáveis presentes no mosto cervejeiro antes de passar pelo processo de fermentação.
 - **FG:** *final gravity* ou densidade final, utiliza o mesmo princípio da OG, mas é medido após o processo de fermentação, quando as leveduras consomem parte dos açúcares presentes no mosto.
 - **IBU:** *International Bitterness Unit* é uma medida internacional utilizada para medir o amargor de uma cerveja.
 - **SRM:** *Standard Reference Method* é uma das escalas de cores amplamente utilizadas na indústria cervejeira para medição da cor dos maltes e das cervejas.
 - **ABV:** *alcohol by volume* ou álcool por volume é um padrão de medida para expressar o teor alcoólico de bebidas. Possui uma relação direta com a OG e FG, pois, a partir destas, é possível realizar o cálculo de teor alcoólico.
 - **Exemplos Comerciais:** demonstra exemplares comerciais atuais que representam o estilo descrito no período da publicação do guia.
 - **Atributos de Estilo:** são TAGs que representam os atributos ou informações do estilo.

2.2. MACHINE LEARNING

Cerri e Carvalho (2017) mencionam que a Inteligência Artificial (IA) vem sendo cada vez mais utilizada no intuito de resolver problemas em diversos setores, como

economia, pesquisas, gestão pública entre outras. Enfatizam também que uma área de grande destaque, quando falamos de IA, é a área de Aprendizado de Máquina.

O Aprendizado de Máquina ou *Machine Learning*, é uma área de pesquisa da Inteligência Artificial com intuito de produzir softwares de computador, ou em outros dispositivos, com a habilidade de aprender um determinado trabalho por si só. É uma área de estudo multidisciplinar que engloba probabilidade, estatística, teoria da complexidade computacional, filosofia, psicologia, teoria da informação, neurobiologia, a própria inteligência artificial entre outras (CERRI e CARVALHO, 2017).

Algumas técnicas de Machine Learning podem utilizar de modelos matemáticos por meio de dados e/ou organização de dados, e outros podem utilizar modelos matemáticos de alta complexidade. Alguns métodos são: redes neurais artificiais (RNAs), árvore de decisão, regressão logística, *random forests*, rede bayesiana, *support vector machine* (SVM), *deep learning* entre outros (PAIXÃO, 2022).

Neste trabalho será desenvolvido um algoritmo com um método utilizando o cálculo de distância euclidiana e um outro algoritmo utilizando a técnica *Term Frequency-Inverse Document Frequency* (TF-IDF).

2.2.1. Distância euclidiana

Cruz (2022) diz que a distância euclidiana é a medida de separação entre dois pontos ou elementos, onde $x = (x_1, x_2, \dots, x_n)$ e $y = (y_1, y_2, \dots, y_n)$, em um ambiente com n dimensões. Essa distância é calculada através da raiz quadrada da somatória das diferenças entre as coordenadas ao quadrado, representada pela equação 1.

Equação 1 - Cálculo para distância euclidiana.

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

Fonte: CRUZ, 2022.

Apesar da sua efetividade em vários casos, ela pode não satisfazer para propósitos estatísticos, pois cada coordenada colabora igualmente para o cálculo da distância. No momento em que as coordenadas possuem variações em grandes magnitudes, algumas coordenadas terão peso proporcionalmente maior na medição e, nesse caso, o ideal é normalizar os dados para reduzir essas grandes entre elas (CRUZ, 2022).

Heaton (2010) diz que a normalização consiste em mapear um intervalo numérico para outro, como na faixa de -1 e 1, tirando assim as discrepâncias extrema entre os valores. Ele diz que a equação 2 pode ser utilizada para tal fim.

Equação 2 - Cálculo para normalização de valores.

$$f(x) = \left(\left(\frac{x - \min}{\max - \min} \right) \cdot (\text{high} - \text{low}) \right) + \text{low}$$

Fonte: HEATON, 2010.

Onde:

- x = valor a ser normalizado;
- \min = valor mínimo que x pode alcançar dentro de sua faixa atual;
- \max = valor máximo que x pode alcançar dentro de sua faixa atual;
- low = a faixa menor do valor a ser aceito para o novo mapeamento, podendo ser -1 ou 0 ;
- high = a faixa máxima do valor a ser aceito para o novo mapeamento, podendo ser 0 ou 1 ;

2.2.2. Term frequency-inverse document frequency (TF-IDF)

A mineração de textos é um procedimento para extrair informações e conhecimentos relevantes ou até padrões ainda desconhecidos a partir de um documento textual não estruturado, ou seja, um texto que não pode ser lido ou interpretado por um sistema de computador sem algum tratamento adicional. A principal função da área de mineração de textos engloba transformar um texto em um vetor número, em outras palavras, transformar em uma maneira que propicie uma análise matemática, podendo a categorizar documentos e analisar frequências de um determinado termo dentro de um texto entre outros (KIDO, JUNIOR e MORIGUCHI, 2014).

Uma técnica estatística, de acordo com Kido, Junior e Moriguchi (2014), é o TF-IDF, na qual se utiliza de um processo de mineração de texto e possui como fundamental função descobrir palavras importantes em um texto não estruturado ou semi estruturado. Esse processo basicamente adiciona um valor a cada termo, que pondera sua frequência no texto e em todos dos documentos da base, demonstrando sua importância.

Zhang, Yoshida e Tang (2011), de acordo com seus estudos, complementam que o TF-IDF surgiu da modelagem linguística de expressões em que um documento definido pode ser dividido em duas categorias, as que possuem um termo relevante e as que não possuem relevância em relação ao tema de um documento. Diante disso, os autores apresentam a seguinte fórmula para o cálculo do TF-IDF.

Equação 3 - Cálculo para o peso do TF-IDF.

$$w_{ij} = tf_{ij} * \log \left(\frac{N}{df_i} \right)$$

Fonte: KIDO, JUNIOR e MORIGUCHI, 2014.

Onde:

- w_{ij} = peso TF-IDF para o termo i no documento j ;
- tf_{ij} = frequência do termo i no documento j (Term Frequency);

- N = número total de documentos considerados para análise;
- df_i = número de documentos que contêm o termo i (Document Frequency);

3. RESULTADOS E DISCUSSÕES

Para a apuração dos resultados propostos neste trabalho, foi desenvolvido um software para a plataforma Desktop na linguagem de programação Python. Por já existir diversos algoritmos prontos para uso, as famosas bibliotecas, foi utilizado a *Scikit-Learn*, *Pandas*, *Numpy* e a *SciPy* para o desenvolvimento da parte de *machine learning* do software.

O primeiro teste a ser implementado foi com a utilização do cálculo da distância euclidiana para medir as distâncias entre um estilo base em relação aos outros, utilizando as características vitais dos estilos de cerveja apresentadas pelo guia BJCP.

Para o segundo teste, foi utilizado a técnica do TF-IDF para interpretar os atributos como aroma, aparência, sabor e TAGs dos estilos de cerveja do guia BJCP, para então classificar os estilos próximos por meio de um estilo base apresentado.

Como foi utilizado no primeiro teste uma técnica quantitativa para calcular a similaridade entre os estilos de cerveja e, no segundo teste foi utilizado uma técnica qualitativa para analisar as palavras que resumem os estilos de cerveja e calcular a semelhança entre eles, no terceiro teste, por tratamos de dois termos diferentes para comparar os estilos, foi realizado a localização da intersecção dos estilos de cerveja através de um estilo base, ou seja, por meio de um estilo base para ambas as técnicas utilizadas, pegamos apenas os estilos que estão presentes no resultado das duas técnicas como retorno do algoritmo desenvolvido.

3.1. Teste 1: cálculo de similaridade através da distância euclidiana entre os estilos de cerveja

Nesta etapa foi utilizado os valores das características vitais dos estilos de cerveja apresentados no guia de estilos do BJCP. Como exemplo dos dados, na tabela 1 pode ser observado alguns destes.

Tabela 1 - Alguns estilos do guia BJCP com suas características vitais.

SIGLA	NOME DO ESTILO	IBU MÍNIMO	IBU MÁXIMO	SRM MÍNIMO	SRM MÁXIMO	OG MÍNIMO	OG MÁXIMO	FG MÍNIMO	FG MÁXIMO	ABV MÍNIMO	ABV MÁXIMO
1A	AMERICAN LIGHT LAGER	8	12	2	3	1.028	1.040	0.998	1.008	2.8	4.2
2A	INTERNATIONAL PALE LAGER	18	25	2	6	1.042	1.050	1.008	1.012	4.5	6.0
3A	CZECH PALE LAGER	20	35	3	6	1.028	1.044	1.008	1.014	3.0	4.1
4A	MUNICH HELLES	16	22	3	6	1.044	1.048	1.006	1.012	4.7	5.4
5A	GERMAN LEICHTBIER	15	28	1.5	4	1.026	1.034	1.006	1.010	2.4	3.6
6A	MÄRZEN	18	24	8	17	1.054	1.060	1.010	1.014	5.6	6.3
7A	VIENNA LAGER	18	30	9	15	1.048	1.055	1.010	1.014	4.7	5.5
8A	MUNICH DUNKEL	18	28	17	28	1.048	1.056	1.010	1.016	4.5	5.6
9A	DOPPELBOCK	16	26	6	25	1.072	1.112	1.016	1.024	7.0	10.0
10A	WEISSBIER	8	15	2	6	1.044	1.053	1.008	1.014	4.3	5.6
11A	ORDINARY BITTER	25	35	8	14	1.030	1.039	1.007	1.011	3.2	3.8
12A	BRITISH GOLDEN ALE	20	45	2	5	1.038	1.053	1.006	1.012	3.8	5.0

Fonte: BJCP, 2021.

Inicialmente foi criado uma estrutura orientada a objetos para representar os estilos de cerveja, tendo como atributos a sigla do estilo, seu nome e suas características vitais. Em seguida todos os dados das características vitais foram normalizados para a escala de -1 a 1, para que seja tratado as discrepâncias de valores. Na etapa seguinte foi realizado os cálculos da distância de cada estilo de cerveja entre todos os outros estilos. Ressalvo que 18 estilos, por não possuírem os dados das características vitais explicitas no guia, não entraram neste teste. Estes podem ser observados na tabela 2.

Tabela 2 - Estilos de cerveja sem as características vitais presentes no guia BJCP.

SIGLA	NOME DO ESTILO
	HISTORICAL BEER: KELLERBIER
28A	BRETT BEER
28B	MIXED-FERMENTATION SOUR BEER
28C	WILD SPECIALTY BEER
29A	FRUIT BEER
29B	FRUIT AND SPICE BEER
29C	SPECIALTY FRUIT BEER
30A	SPICE, HERB, OR VEGETABLE BEER
30B	AUTUMN SEASONAL BEER
30C	WINTER SEASONAL BEER
30D	SPECIALTY SPICE BEER
31A	ALTERNATIVE GRAIN BEER
31B	ALTERNATIVE SUGAR BEER
32A	CLASSIC STYLE SMOKED BEER
32B	SPECIALTY SMOKED BEER
34A	COMMERCIAL SPECIALTY BEER
34B	MIXED-STYLE BEER
34C	EXPERIMENTAL BEER

Fonte: BJCP, 2021.

Após toda a estruturação de cálculos do software, foi criado um campo onde se informa a sigla ou o nome de um estilo base e, o programa retorna os 10 estilos mais

próximos de acordo com a distância calculada. Depois de alguns testes utilizando diversos estilos como base, na tabela 3 pode-se observar alguns destes resultados.

Tabela 3 - Resultados do teste 1.

ESTILO BASE	ESTILOS SIMILARES	DISTÂNCIA CALCULADA
1A-AMERICAN LIGHT LAGER	23A-BERLINER WEISSE	0.44
	HISTORICAL BEER: LICHTENHAINER	0.53
	5A-GERMAN LEICHTBIER	0.75
	1B-AMERICAN LAGER	0.81
	X4-CATHARINA SOUR	0.82
	23G-GOSE	0.84
	HISTORICAL BEER: PIWO GRODZISKIE	0.92
	1C-CREAM ALE	0.99
	3A-CZECH PALE LAGER	1.00
	23D-LAMBIC	1.03
7B-ALTBIER	19A-AMERICAN AMBER ALE	0.35
	19B-CALIFORNIA COMMON	0.36
	11C-STRONG BITTER	0.37
	11B-BEST BITTER	0.38
	3C-CZECH AMBER LAGER	0.56
	24B-BELGIAN PALE ALE	0.56
	7A-VIENNA LAGER	0.56
	18B-AMERICAN PALE ALE	0.57
	13B-BRITISH BROWN ALE	0.60
	HISTORICAL BEER: KENTUCKY COMMON	0.65
16A-SWEET STOUT	16B-OATMEAL STOUT	0.70
	20A-AMERICAN PORTER	0.81
	15C-IRISH EXTRA STOUT	0.90
	16C-TROPICAL STOUT	0.96
	15B-IRISH STOUT	1.01
	19C-AMERICAN BROWN ALE	1.03
	HISTORICAL BEER: PRE-PROHIBITION PORTER	1.04
	3D-CZECH DARK LAGER	1.05
	20B-AMERICAN STOUT	1.06
	8B-SCHWARZBIER	1.08

Fonte: Elaborado pelo autor.

Quando analisado as características vitais entre o estilo base e os seus mais próximos (com menor valor), de acordo com a distância euclidiana, o algoritmo obteve bons resultados, soube calcular a similaridade dos estilos com suas intensidades de acordo com os atributos analisados. Ressalvo que este método não leva em consideração aos fatores qualitativos entre os estilos, como características condimentadas, amadeiradas, defumadas entre outras.

3.2. Teste 2: cálculo de similaridade através da técnica TF-IDF entre os estilos de cerveja

Para este teste, foi utilizado uma abordagem qualitativa para encontra um certo grau de similaridade entre um estilo base e os demais estilos. Aqui também foi utilizado a

abordagem orientada a objetos para melhor representar os estilos. A classe criada possui os atributos sigla e nome do estilo para as devidas identificações, e os atributos aroma, aparência, sabor e TAGs que são analisados pelo algoritmo. Diferente do teste anterior, este contempla todos os estilos presente no guia BJCP. Na tabela 4 pode ser observado alguns exemplos de entrada ao algoritmo criado.

Tabela 4 - Alguns estilos de cerveja com suas TAGs.

SIGLA	NOME DO ESTILO	AROMA	APARÊNCIA	SABOR	TAGs
1A	AMERICAN LIGHT LAGER	Low malt aroma optional, but may be perceived as grainy, sweet, or corn-like, if present. Light spicy, floral, or herbal hop aroma optional. While a clean fermentation profile is desirable, a light amount of yeast character is not a fault.	Very pale straw to pale yellow color. White, frothy head seldom persists. Very clear.	Relatively neutral palate with a crisp, dry finish and a low to very low grainy or corn-like flavor that might be perceived as sweetness due to the low bitterness. Low floral, spicy, or herbal hop flavor optional, but is rarely strong enough to detect. Low to very low bitterness. Balance may vary from slightly malty to slightly bitter, but is usually close to even. High carbonation may accentuate the crispness of the dry finish. Clean fermentation profile.	session-strength, pale-color, bottom-fermented, lagered, north-america, traditional-style, pale lager-family, balanced
2A	INTERNATIONAL PALE LAGER	Low to medium-low grainy-malty or slightly corny-sweet malt aroma. Very low to medium spicy, floral, or herbal hop aroma. Clean fermentation profile.	Pale straw to gold color. White, frothy head may not be long lasting. Very clear.	Low to moderate levels of grainy-malt flavor, medium-low to medium bitterness, with a crisp, dry, well-attenuated finish. The grain character can be somewhat neutral, or show a light bready-crackery quality. Moderate corny or malty sweetness optional. Medium floral, spicy, or herbal hop flavor optional. Balance may vary from slightly malty to slightly bitter, but is usually relatively close to even. Neutral aftertaste with light malt and sometimes hop flavors.	standard-strength, pale-color, bottom-fermented, lagered, traditional-style, pale-lager-family, balanced
3A	CZECH PALE LAGER	Light to moderate bready-rich malt combined with light to moderate spicy or herbal hop bouquet; the balance between the malt and hops may vary. Faint hint of caramel is acceptable. Light (but never intrusive) diacetyl and light, fruity esters are optional. No sulfur.	Light yellow to deep gold color. Brilliant to very clear, with a long-lasting, creamy white head.	Medium-low to medium bready-rich malt flavor with a rounded, hoppy finish. Low to medium-high spicy or herbal hop flavor. Bitterness is prominent but never harsh. Flavorful and refreshing. Low diacetyl or fruity esters are optional, but should never be overbearing.	session-strength, pale-color, bottom-fermented, lagered, central-europe, traditional-style, pale lager-family, bitter, hoppy

Fonte: BJCP, 2021.

Na próxima etapa, após todos os dados já estarem no algoritmo, foi realizado o pré-processamento dos dados para retirar as palavras que, para análise de textos, não contribuem para o significado da frase, como: “o”, “a”, “e”, “em”, “com” entre outras. Logo em seguida foi criado um *DataFrame* contendo uma coluna com todos os atributos a serem analisados e, em seguida, esta coluna foi direcionada ao algoritmo TF-IDF.

Em um total de 4.94 segundos o software fez todas as análises de cada estilo de cerveja em relação a todos os outros. Na tabela 5 pode-se observar alguns resultados apresentados.

Tabela 5 - Resultados do teste 2.

ESTILO BASE	ESTILOS SIMILARES	SIMILARIDADE
X2-ARGENTINE IPA	12C-ENGLISH IPA	33%
	17D-ENGLISH BARLEY WINE	31%
	22A-DOUBLE IPA	31%
	25B-SAISON	31%
9A-DOPPELBOCK	6C-DUNKLES BOCK	44%
	10C-WEIZENBOCK	43%
	9B-EISBOCK	42%
	2C-INTERNATIONAL DARK LAGER	41%
26C-BELGIAN TRIPEL	25C-BELGIAN GOLDEN STRONG ALE	55%
	25A-BELGIAN BLOAND ALE	46%
	26B-BELGIAN DUBBEL	46%
	25B-SAISON	34%
22A-DOUBLE IPA	21A-AMERICAN IPA	67%
	21B-SPECIALITY IPA: RYE IPA	52%
	22B-AMERICAN STRONG ALE	50%
	21B-SPECIALITY IPA: BRUT IPA	46%
1B-AMERICAN LAGER	1A-AMERICAN LIGHT LAGER	93%
	2A-INTERNATIONAL PALE LAGER	54%
	1C-CREAM ALE	44%
	HISTORICAL BEER: PRE-PROHIBITION LAGER	39%

Fonte: Elaborado pelo autor.

Conforme demonstrado na tabela acima, alguns estilos exibiram na faixa de 30% de similaridade. Embora seja um percentual modesto, o desempenho do algoritmo é considerado satisfatório. Por exemplo, ao analisar o estilo base X2-ARGENTINE IPA em comparação com o guia BJCP, observa-se que os estilos correspondentes compartilham características similares ao estilo base.

De forma geral, os resultados são aceitáveis quanto ao objetivo do trabalho, uma vez que o software consegue apresentar ótimas opções de recomendações baseadas em um estilo base, levando em consideração a parte qualitativa descritas nos atributos analisados do guia BJCP.

3.3. Teste 3: intersecção dos resultados entre as técnicas do teste 1 e 2

Neste terceiro teste, buscou-se aprimorar ainda mais as sugestões de estilos de cerveja com base em um estilo inicial. Para isso, foi considerado a intersecção dos estilos presentes nos testes 1 e 2, ou seja, identificou-se os estilos que estão presentes em ambos os testes, a partir de uma base comum. Os resultados dessa análise podem ser observados na tabela 6.

Tabela 6 - Resultados do teste 3.

ESTILO BASE	ESTILOS SIMILARES
1B-AMERICAN LAGER	1C-CREAM ALE
5B-KÖLSCH	2A-INTERNATIONAL PALE LAGER
	4A-MUNICH HELLES
	18A-BLONDE ALE
12C-ENGLISH IPA	11C-STRONG BITTER
	X2-ARGENTINE IPA
18A-BLOND ALE	1D-AMERICAN WHEAT BEER
	1C-CREAM ALE
	X1-PAMPAS GOLDEN ALE
20C-IMPERIAL STOUT	16D-FOREIGN EXTRA STOUT
	22C-AMERICAN BARLEYWINE

Fonte: Elaborado pelo autor.

O teste conseguiu sugerir estilos de forma concisa, tanto com base nas características vitais quanto nas descrições dos estilos. É importante lembrar que nem todos os estilos estão presentes no teste, especialmente porque o teste 1 não abrange todos eles e alguns estilos podem não ter intersecção entre os testes.

CONSIDERAÇÕES FINAIS

É nítido que o consumo de cerveja é muito apreciado no Brasil, uma vez que podemos observar o crescente aumento do registro de fábricas e volume da bebida no mercado brasileiro. Outro ponto chave é que a cerveja é um fator muito importante para a economia brasileira, além de representas 2% do PIB e pagar bilhões em impostos anuais, o número de empregos envolvendo toda a cadeia da bebida é bastante grande.

Foi observado também que existem muitos tipos de cervejas e uma grande diversidade de fatores que as diferenciam, tendo um único estilo, uma gama enorme de como desenvolvê-lo. Com isso, vimos também que existem alguns guias que servem de direcionamento para identificarmos esses estilos, e que um deles bastante conhecido é o BJCP.

Ao analisar a produção, foi visto que o maior volume está com as cervejas do tipo *Lager*, ou seja, são poucos estilos para uma dimensão grande de estilos existentes. Quando observado somente para os estilos Lager Leve Clara e a Pilsener, representa 74,65% de toda a produção. E quando comparado esses com o guia BJCP, são grandes as chances de pertencerem a um único estilo.

Foi observado também, de acordo com o Retrato dos consumidores de cervejas de 2023 da equipe Surra de Lúpulo (2024), que existem uma gama grande de pessoas que não sabem o que beber e/ou não gostam da bebida. Mas, para os que não gostam, vai uma pergunta: Experimentaram uma cerveja leve e suave ou já partiram para os estilos mais radicais como uma IPA?

Já a tecnologia, ela vem evoluindo muito nos últimos anos, principalmente quando falado em inteligência artificial. Quando observado os algoritmos de *Machine Learning*, pôde-se ver um grande potencial para se trabalhar em diversos ramos e, um deles é o da cerveja.

Para resolver o problema do número de pessoas que não sabem qual estilo de cerveja degustar, o presente trabalho apresentou uma solução que pode servir de orientação para elas. Foram elaborados 3 testes envolvendo formas diferentes para tratar o problema e, em todos estes, é apresentado um estilo base e o algoritmo retorna os mais similares a este.

Para o primeiro teste, foi realizado uma tratativa quantitativa para buscar a menor distância entre os estilos utilizando o cálculo da distância euclidiana aplicado as características vitais entre os estilos. Já no segundo teste, observou a parte qualitativa dos estilos para buscar as melhores semelhanças, como: aroma, aparência, sabor e TAGs. No terceiro teste, buscou-se unir os outros testes em busca de resultados mais concisos.

Em todos os testes, quando analisado o guia BJCP, os resultados foram aceitáveis para os estilos apresentados a partir de um estilo base de análise para o algoritmo.

Após todos os testes aplicados neste trabalho, como forma de uma solução mais abrangente para resolver o objetivo deste trabalho, foi desenvolvido um aplicativo para dispositivos móveis, para a plataforma Android, com o intuito de ser um guia para quem busca se aventurar entre os estilos de cerveja. O aplicativo se chama Beer Hunter Brasil, disponível na Google Play Store.

Como proposta para trabalhos futuros, deixo a sugestão de desenvolver uma ferramenta capaz de sugerir harmonizações entre cervejas e comidas de acordo com o perfil de quem está consumindo. Pode se utilizar, para tal fim, os estudos envolvendo sistemas de recomendação, como: regras de associação, filtragem colaborativa ou filtragem de conteúdo.

REFERÊNCIAS BIBLIOGRÁFICAS

AMORIM, Bia (Org.). **Guia da sommelieria de cervejas**. 1. Ed. Porto Alegre, RS: Editora Krater, 2022. 480 p. ISBN: 978-65-99165-15-3

BJCP – BEER JUDGE CERTIFICATION PROGRAM. 2021 **Style guidelines**: beer style guidelines. 2021. Disponível em: https://www.bjcp.org/download/2021_Guidelines_Beer.pdf. Acesso: 11 de junho de 2024.

BRASIL. Decreto nº 9.902, de 8 de julho de 2019. Altera o Anexo ao Decreto nº 6.871, de 4 de junho de 2009, que regulamenta a LEI Nº 8.918, de 14 de julho de 1994, que dispõe sobre a padronização, a classificação, o registro, a inspeção, a produção e a fiscalização de bebidas. **Diário Oficial da União**: seção 1, Brasília, DF, n. 130, p. 7, 9 jul. 2019.

CERRI, Ricardo; CARVALHO, André Carlos de Leon Ferreira de. **Aprendizado de máquina**: breve introdução e aplicações. *Cadernos de Ciência & Tecnologia*, Brasília, v. 34, n. 3, p. 297-313, set/dez. 2017.

CERVBRASIL. **Anuário 2016**. Disponível em: [CervBrasil-Anuario2016_WEB.pdf](#). Acesso em: 08 de junho de 2024.

CRUZ, Yan Américo Lobianchi da. **Classificação de imagens multiespectrais da mão utilizando análise por componentes principais e KNN**. 2022. Trabalho de conclusão de curso (Graduação em Engenharia de Computação). Pontifícia Universidade Católica de Goiás, Goiânia/GO, 2022.

HEATON, Jeff. **Programming neural networks with Encog3 in Java**. Heaton Research, Inc. WordsRU.com, 2010. Disponível em: <https://github.com/jeffheaton/jeffheaton-book-code/raw/master/ebooks/EncogJavaEdi.zip>. Acesso em: 11 de junho de 2024.

KIDO, Guilherme Sakaji; JUNIOR, Sylvio Barbon; Moriguchi, Stella Naomi. 2014. **Comparação entre TF-IDF e LSI para pesagem de termos em micro-blog**. Disponível em: (PDF) [Comparação entre TF-IDF e LSI para pesagem de termos em micro-blog \(researchgate.net\)](#). Acesso em: 12 de junho de 2024.

418

MAPA – Ministério da Agricultura, Pecuária e Abastecimento. **Anuário da cerveja 2024**: ano de referência 2023. Secretaria de Defesa Agropecuária. Brasília: MAPA/SDA, 2024. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/noticias/com-1-847-cervejarias-registradas-no-brasil-setor-cresce-6-8-em-2023/anuario-da-cerveja>. Acesso em: 08 de junho de 2024.

MORADO, Ronaldo. **Larousse da cerveja**: a história e as curiosidades de uma das bebidas mais populares do mundo. São Paulo: Alaúde Editorial, 2017.

OLIVER, Garret. **A mesa do mestre-cervejeiro**: descobrindo os prazeres das cervejas e das comidas verdadeiras. Tradução de Anthony Cleaver. São Paulo: Editora Senac São Paulo 2012.

PAIXÃO, Gabriela Miana de Mattos et al. (2022). **Machine Learning na medicina**: revisão e aplicabilidade. *Arq Bras Cardiol*. 2022; 118(1):95-102.

PIMENTA, Larissa Bicalho. *et al.* **A história e o processo da produção de cerveja**: uma revisão. *Cadernos de Ciência & Tecnologia*, Brasília, v. 37, n. 3, e26715, 2020. DOI: [10.35977/0104-1096.cct2020.v37.26715](https://doi.org/10.35977/0104-1096.cct2020.v37.26715)

SURRA DE LÚPULO. **Retrato dos consumidores de cervejas 2023**. Disponível em: <https://surradelupulo.com.br/retrato-dos-consumidores-de-cervejas/>. 2024. Acesso em: 08 de junho de 2024.

ZHANG, Wen; YOSHIDA, Taketoshi; TANG, Xijin. 2011. **A comparative study of TF*IDF, LSI and multi-words for text classification**. Elsevier. Expert Systems With Applications. Volume 38, Issue 3, March 2011, Pages 2758-2765. Disponível em: <https://doi.org/10.1016/j.eswa.2010.08.066>. Acesso em: 12 de junho de 2024.