**Previsão de Avaliação Geral de Cervejas com Base em Características Sensoriais Usando Machine Learning**

Alexandre Polizelli Marino¹\*; Wagner W. Bortoletto²

¹ Rua: Doutor Francisco José Longo, 191 – Apto 113 – Chácara Inglesa; 04140-060 São Paulo, SP, Brasil

**²** PECEGE. Mestre em Engenharia de Produção e Manufatura. Rua Antônio Menardi – Residencial Victor D’Andrea; 13481-344, Limeira, São Paulo, Brasil.

\*autor correspondente: ale\_marino1@hotmail.com

**Previsão de Avaliação Geral de Cervejas com Base em Características Sensoriais Usando Machine Learning**

**Resumo (ou Sumário Executivo)**

O resumo é uma descrição geral do trabalho, apresentando de forma sucinta todas as seções do TCC. **No caso do curso de** **MBA em Gestão de Projetos**, em função das particularidades do trabalho – que pode ser um plano de projeto – **é facultado aos alunos o uso do termo** **Sumário Executivo**. Essa seção deve ser escrita de forma clara e objetiva, informando ao leitor em que consiste o TCC a fim de despertar o seu interesse para a leitura de todo o trabalho. Os termos Resumo (ou Sumário Executivo) e Palavras-chave devem ser grafados em negrito, alinhados à esquerda com apenas a primeira letra da palavra em letra maiúscula. O Resumo ou Sumário Executivo deve conter no máximo 250 palavras, sendo redigido em um único parágrafo, em espaçamento simples e no tempo verbal pretérito perfeito do indicativo (passado). O início da seção deve conter uma curta apresentação sobre a importância/justificativa (contextualização do tema) do trabalho. Em seguida, é necessário apresentar o objetivo geral, que deve ser redigido de maneira sucinta e direta. Ademais, a seção deve conter uma breve descrição da metodologia empregada na pesquisa, abordando os aspectos mais importantes para o entendimento do TCC. É importante destacar os principais resultados obtidos a partir da pesquisa, apresentando relações e/ou considerações que chamem a atenção do avaliador e do público em geral. O final do resumo (ou sumário executivo) deve conter uma conclusão geral (com uma ou duas frases). Deve-se ter cuidado para não repetir a descrição dos resultados nessa seção.

**Palavras-chave:** (inserir até cinco palavras diferentes das contidas no título, separadas por ponto-e-vírgula).

**Atenção:** antes de enviar o arquivo para o Move - Sistema de TCCs, remova todas as instruções originais que estão abaixo do conteúdo dos tópicos.

**Título em inglês** **ou espanhol** (opcional)

**Abstract** ou **Resumen** (opcional)

...................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................

**Keywords** ou **Palabras Clave:** (opcional)

**Introdução**

O mercado de cervejas artesanais tem crescido exponencialmente ao longo dos últimos anos, levando a um aumento na diversidade de estilos, sabores e características sensoriais disponíveis para os consumidores. (MAPA, 2023)

Ao mesmo tempo, plataformas de avaliação de cervejas, como aplicativos e sites especializados, permitem que milhões de usuários compartilhem suas impressões e notas sobre diversos aspectos das bebidas que consomem.‌

A avaliação de qualidade de cervejas tem se tornado um aspecto crucial para a indústria, especialmente considerando o aumento significativo na diversidade de estilos e o crescente número de cervejarias artesanais no Brasil. Em 2022, o crescimento deste mercado no brasil foi de 11,6%, subindo para um total de 1.729 cervejarias registradas e é o terceiro maior produtor de cerveja no mundo, atrás apenas da China e dos Estados Unidos, conforme MAPA (2023).

O aumento de dados disponibilizados por consumidores representa uma oportunidade valiosa para aplicar técnicas de ciência de dados com o objetivo de entender e prever como diferentes características de uma cerveja podem influenciar a avaliação geral que ela receberá.

Estudos anteriores sobre previsão de preferências de consumidores já demonstraram a eficácia de técnicas de machine learning em domínios como recomendação de filmes KNIAZIEVA (2022), e até mesmo previsão de sucesso de músicas, conforme MUKHOPADHYAY (2024). No entanto, a previsão da avaliação geral de cervejas com base em suas características sensoriais ainda é um campo pouco explorado, especialmente considerando a rica diversidade de variáveis sensoriais envolvidas nesse processo de avaliação.

Artigos como o de ARAUJO (2003), procuram entender como as características sensoriais ajudam a traçar um perfil de consumo e este trabalho se propõe a utilizar a regressão logística multinominal para entender como algumas variáveis sensoriais, como textura, gosto, aroma, aparência e se teor alcoólico e tipo de cerveja influenciam na nota geral de uma cerveja do ponto de vista do consumidor.

Dado tamanho e potencial crescimento do mercado, MAPA (2023) a compreensão de como as características sensoriais, como aroma, gosto, aparência e paladar, influenciam a avaliação geral dos consumidores pode fornecer insights valiosos para otimizar a produção e atender melhor às expectativas dos consumidores.

A relevância deste estudo está em demonstrar qual das características é a mais relevante na percepção geral do consumidor ou se há um tipo preferido de cerveja pelos consumidores, além de procurar entender se o teor alcoólico da cerveja possui alguma influência na avaliação final. Os resultados podem beneficiar diretamente cervejarias e distribuidores, que poderiam ajustar suas produções com base nas características mais valorizadas pelos consumidores, aumentando suas chances de sucesso no mercado.

Este projeto busca explorar uma base de dados contendo mais de 1,5 milhões de avaliações de cervejas realizadas por consumidores. O dataset foi disponibilizado pela empresa BeerAdvocate, dos Estados Unidos e contém percepções de mais de 10 anos no que tange à avaliação de cervejas produzidas em mais 5.800 cervejarias diferentes, trazendo uma ampla variedade de observações.

**Metodologia ou Material e Métodos**

O conjunto de dados utilizado contém 1.586.614 linhas e 14 colunas, onde cada linha representa a avaliação do consumidor em relação a cerveja e as colunas representam as variáveis como avaliações sensoriais (aroma, paladar, gosto e aparência), teor alcoólico, nome da cervejaria, estilo de cerveja, data da avaliação entre outros.

As variáveis deste banco de dados, são:

|  |  |  |
| --- | --- | --- |
| Variável (inglês) | Variável (português) | Observação |
| brewery\_id | id\_cervejaria | Identificador da cervejaria |
| brewery\_name | nome\_cervejaria | Nome da cervejaria |
| review\_time | data\_avaliacao | Data da avaliação |
| review\_overall | avaliação\_geral | Avaliação geral para a cerveja |
| review\_aroma | avaliação\_aroma | Avaliação para o aroma da cerveja |
| review\_appearance | avaliação\_aparencia | Avaliação para a aparência da cerveja |
| review\_palate | avaliação\_textura | Avaliação da textura da cerveja |
| review\_taste | avaliação\_gosto | Avaliação do gosto da cerveja |
| review\_profilename | nome\_avaliador | Nome do avaliador |
| beer\_style | estilo\_cerveja | Estilo da cerveja |
| beer\_style | grupo\_estilo\_cerveja | Agrupamento dos estilos de cerveja |
| beer\_name | nome\_cerveja | Nome da cerveja |
| beer\_abv | teor\_acoolico | Teor alcóolico da cerveja |
| beer\_beerid | id\_cerveja | Identificador da cerveja |

Fonte: <https://www.kaggle.com/datasets/thedevastator/1-5-million-beer-reviews-from-beer-advocate/data>.

O desenvolvimento deste projeto evidenciou a importância de uma abordagem meticulosa e sistemática ao se trabalhar com modelos multiclasse, assegurando que cada etapa, desde o pré-processamento até a avaliação do modelo, fosse realizada com rigor e atenção aos detalhes, contribuindo assim para a eficácia do modelo preditivo.

**Pacotes Utilizados**

Foram utilizadas bibliotecas essenciais para o desenvolvimento do modelo de machine learning, como “pandas”, “numpy” e “scikit-learn”. Essas bibliotecas desempenham papéis fundamentais na manipulação e análise de dados, além de fornecerem as ferramentas necessárias para a criação e avaliação dos modelos preditivos.

A biblioteca “pandas”, por exemplo, é utilizada para o carregamento e manipulação do conjunto de dados, permitindo organizar as variáveis de forma clara e acessível. Com o “numpy”, é possível realizar operações matemáticas e manipulações de “arrays” de maneira eficiente, essencial quando se trabalha com grandes volumes de dados. Já o “scikit-learn” é uma biblioteca poderosa para a implementação de algoritmos de machine learning, contendo diversas funcionalidades para realizar a divisão de dados, a construção de modelos e a avaliação de desempenho.

No âmbito de visualização de dados, pacotes como “matplotlib”, “seaborn” e “plotly” foram importados para a visualização dos dados e análise exploratória. Isso foi fundamental para entender melhor a distribuição das variáveis e possíveis correlações entre elas. Visualizações como histogramas, “boxplots” e gráficos de dispersão facilitaram a identificação de padrões e ajudaram na escolha das variáveis mais relevantes para o modelo.

**Pré-processamento dos dados**

Para garantir a robustez e a validade dos resultados, foram adotados cuidados específicos durante o processo de modelagem. Primeiramente, o pré-processamento dos dados se tornou uma etapa fundamental. Tratou-se de garantir que não houvesse valores ausentes ou inconsistentes nas variáveis, uma vez que isso poderia afetar a qualidade do modelo preditivo. Segundo Little e Rubin (2002), a manipulação adequada de dados ausentes é essencial para evitar viés nas análises e nas inferências feitas a partir dos modelos

. A variável beer\_abv apresentou 67.785 observações ausentes, representando um valor considerável, o que exigiu um tratamento que se baseou na atribuição da média do teor alcoólico por estilo de cerveja, assumindo que os mesmos tipos de cerveja possuem teores alcoólicos semelhantes dado controle de linha de produção. Esta abordagem foi adotada para evitar a perda de observações significativas, considerando que o teor alcoólico pode ser característica chave para a predição da nota geral da cerveja.

**Normalização**

No que se refere à escolha do algoritmo, a seleção de um solver apropriado e a aplicação de técnicas de regularização foram consideradas indispensáveis. A regularização, conforme abordado por Zou e Hastie (2005), é uma estratégia que pode mitigar problemas de overfitting, especialmente em cenários com múltiplas variáveis explicativas. O QUE É ESSA REGULARIZAÇÃO?

**Transformação de variáveis explicativas de contínuas para discretas**

A transformação de variáveis preditoras e dependentes em uma escala numérica discreta foi fundamental para o treinamento de modelos de regressão logística ordinal, pois essa abordagem permitiu que as relações entre variáveis fossem capturadas e analisadas de maneira mais eficaz. As variáveis categóricas, quando convertidas em valores numéricos, como variáveis dummy ou utilizando métodos de codificação ordinais, possibilitaram a inclusão adequada de informações sobre a ordem das categorias na modelagem. De acordo com Agresti (2010), a conversão de variáveis categóricas em formas numéricas facilitou a estimativa de probabilidades acumuladas e a interpretação dos parâmetros do modelo. Além disso, McCullagh (1980) destacou que essa transformação foi essencial para garantir que os modelos refletissem corretamente a estrutura ordinal dos dados, permitindo que as transições entre categorias fossem analisadas de maneira significativa. A utilização de escalas numéricas discretas não apenas melhorou a precisão da modelagem, mas também assegurou que os pressupostos estatísticos subjacentes da regressão logística ordinal fossem atendidos, conforme discutido por Long (1997). Assim, a transformação adequada das variáveis explicativas foi uma etapa crucial para garantir a validade e a interpretabilidade dos resultados obtidos por meio de modelos de regressão logística ordinal. O método utilizado neste projeto foi de multiplicação por 2 que fez com que as variáveis mantivessem sua ordem original e perdessem a condição as caracterizava como contínua.

**Separação Treino e Teste**

A divisão dos dados em conjuntos de treino e teste também foi uma prática recomendada. Essa separação é fundamental para garantir que o modelo seja avaliado em dados não vistos, evitando o overfitting. Kohavi (1995) destaca que a validação cruzada é uma técnica eficaz para obter estimativas mais confiáveis da performance do modelo, permitindo uma avaliação mais rigorosa do seu desempenho, por isso os dados foram divididos em treino (80%) e teste (20%) utilizando a função “train\_test\_split” da biblioteca “scikit-learn”. As variáveis preditoras foram normalizadas utilizando a função “StandardScaler”, também da biblioteca “scikit-learn”.

**Validação Cruzada**

A validação cruzada é uma técnica essencial para garantir que o modelo preditivo não apenas se ajusta bem aos dados de treino, mas também generaliza adequadamente para dados não vistos. A prática de dividir os dados em conjuntos de treino e teste é crucial para evitar o overfitting e fornecer estimativas mais confiáveis sobre o desempenho do modelo. Kohavi (1995) discute como a validação cruzada pode ser usada para obter uma avaliação rigorosa da performance do modelo, permitindo que o mesmo seja testado em múltiplas subdivisões dos dados. A utilização dessa técnica ajuda a garantir que as conclusões tiradas a partir dos resultados do modelo sejam robustas e confiáveis, refletindo o desempenho esperado em situações do mundo real.

Para validar o desempenho dos modelos e garantir sua generalização, foi utilizada a técnica de validação cruzada k-fold, com k = 5. Isso significa que os dados de treinamento foram divididos em cinco subconjuntos (folds), e o modelo foi treinado e avaliado cinco vezes, utilizando cada fold uma vez como conjunto de teste e os outros quatro como conjunto de treinamento. Ao final, as métricas de avaliação foram calculadas pela média dos resultados obtidos em cada rodada.

**Treinamento de Modelos**

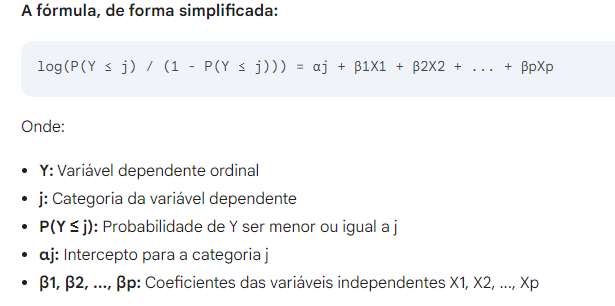
No desenvolvimento de modelos preditivos, a escolha das variáveis e o tratamento adequado dos dados desempenham um papel crucial na eficácia do modelo. Neste contexto, foram utilizadas quatro variáveis sensoriais, a saber, review\_palate, review\_taste, review\_appearance e review\_aroma caracterizadas como variáveis discretas para prever a variável de interesse review\_overall, que se caracteriza como uma variável categórica multiclasse, que atua de forma ordenada, caracterizando a utilização de modelos de regressão logística ordinal.

**Regressão Logística Ordinal**

A regressão logística ordinal é uma extensão da regressão logística que lida com variáveis dependentes categóricas que possuem uma ordem natural. O modelo é particularmente útil quando a variável resposta tem mais de duas categorias ordenadas e calcula a probabilidade cumulativa de uma observação pertencer a uma determinada categoria ou a uma categoria inferior, ou seja, o modelo estima a probabilidade de uma observação pertencer a uma categoria menor ou igual a um determinado valor. Segundo McCullagh (1980), o modelo ordinal trabalha com o pressuposto de que as categorias podem ser ordenadas e que os efeitos das variáveis preditoras são constantes para essas categorias, em um processo conhecido como assunção de “odds” proporcionais.

As variáveis explicativas podem ser contínuas ou categóricas. No caso de variáveis categóricas, como sugerido por Hosmer et al. (2013), estas são geralmente transformadas em variáveis dummy para que possam ser inseridas no modelo, processo que foi feito com as variáveis sensoriais review\_taste, review\_palate, review\_aroma e review\_appearence. Hernan (2010) sugere que essas variáveis dummy são criadas para evitar a multicolinearidade uma vez que elas permitem que os modelos captem a variabilidade entre categorias sem que as variáveis sejam colinearmente dependentes.

A formula da regressão logística ordinal é dada por:



McCullagh (1980) indica que os alfas são chamados de pontos de corte ou interceptos, e definem as fronteiras entre as categorias. Eles representam o logito das probabilidades acumuladas de passar de uma categoria para outra. Cada alfa especifica o ponto em que a probabilidade de pertencer a uma categoria ou inferior se altera. Em termos de interpretação, um maior o valor de alfa implica que é necessário ter um valor maior na combinação das variáveis explicativas para que a probabilidade de estar na categoria superior aumente. Isso sugere que a transição para categorias mais altas é mais difícil, ou seja, requer condições mais favoráveis nas variáveis preditoras para ser alcançada, conforme elencado por Williams, R. (2006).

O coeficiente beta representa o efeito multiplicativo das variáveis explicativas nas “odds” cumulativas, conforme elencado por McCullagh (1980). Um beta positivo aumenta a chance de a resposta estar em uma categoria superior, enquanto um beta negativo aumenta a chance de a resposta estar em uma categoria inferior.

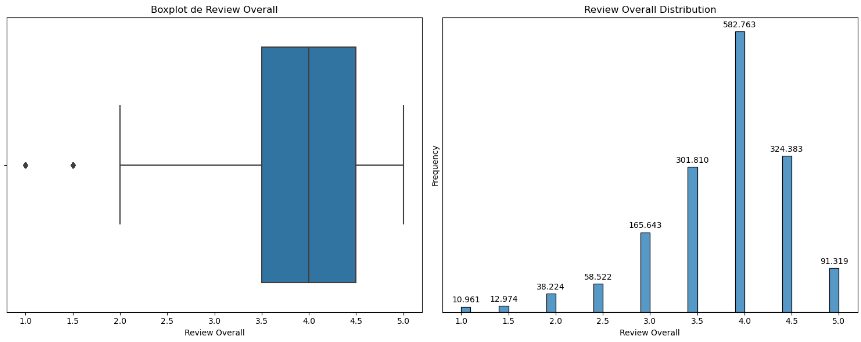
**Métricas de Avaliação**

Finalmente, a interpretação dos resultados requer atenção especial. Menard (2002) enfatiza a importância de compreender os coeficientes do modelo, pois eles fornecem insights sobre a relação entre as variáveis preditoras e a variável de resposta. O uso de métricas de avaliação adequadas, como precisão e F1-score, é essencial para mensurar o desempenho do modelo de forma eficaz, conforme discutido por Sokolova e Lapalme (2009).

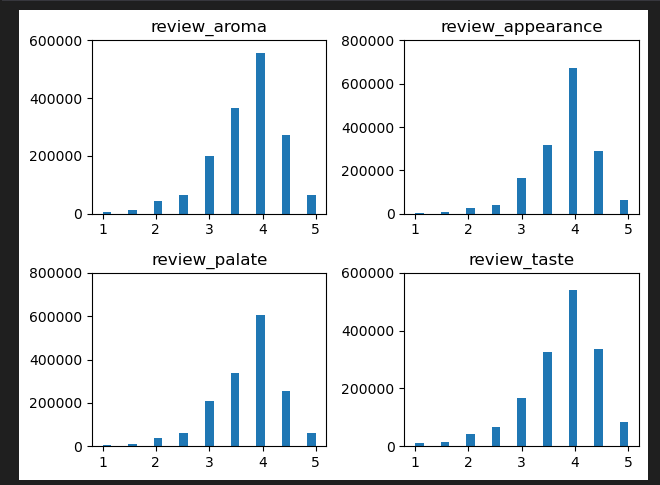
**Resultados e Discussão**

**Análise Exploratória de Dados**

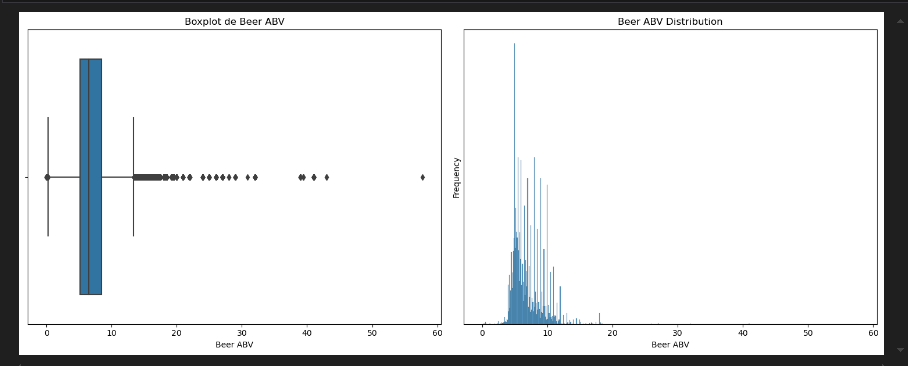
A análise uni variada das variáveis revelou algumas características importantes. Para a variável resposta “review\_overall”, que mede a nota geral da avaliação do consumidor, tida neste projeto como variável resposta, foram gerados histogramas e “boxplots”, que mostraram uma distribuição que, embora próxima da normal, apresentava uma cauda alongada à esquerda. Isso indica que a maioria das avaliações tende a ser positiva, mas há uma proporção menor de notas baixas.



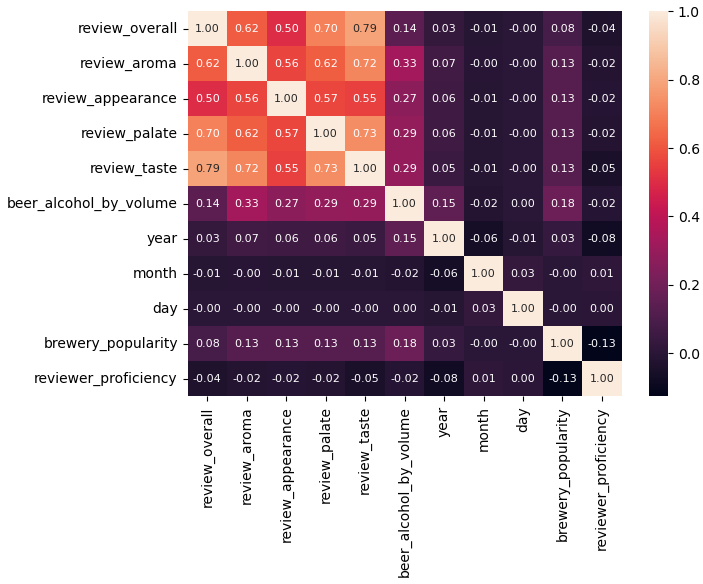
Para as variáveis sensoriais (review\_aroma, review\_taste, review\_palate, review\_appearance), foram observadas distribuições semelhantes, conforme figura abaixo:



Por outro lado, a variável beer\_abv mostrou uma distribuição bem alongada à direita, mostrando que o teor alcoólico tem um range maior do que as outras variáveis quantitativas.



Na análise multivariada, foi gerado um mapa de calor para analisar as correlações entre as variáveis numéricas. As variáveis review\_palate e review\_taste apresentaram as maiores correlações com a variável resposta review\_overall, com coeficientes de 0,79 e 0,72, respectivamente. Isso sugere que o paladar e o gosto da cerveja desempenham um papel crítico na percepção geral do consumidor.



**Conclusão(ões) ou Considerações Finais**

O título da seção Conclusão(ões) ou Considerações Finais deve ser alinhado à esquerda e grafado em negrito. Fica a critério do aluno e do orientador a escolha de qual termo melhor se adequa ao trabalho. Esta seção deve conter frases curtas, apresentando as conclusões e inferências elaboradas a partir da discussão dos resultados. É importante que estas frases não sejam construídas como meras reproduções dos resultados, respondendo aos objetivos propostos no trabalho. Os autores não devem, em hipótese alguma, mencionar, citar ou reproduzir resultados de outros estudos na(s) Conclusão(ões) ou Considerações Finais do TCC. Por fim, salienta-se que essa seção não deve conter tabelas ou figuras, sendo redigida de forma sucinta.

**Atenção:** antes de enviar o arquivo para o Move - Sistema de TCCs, remova todas as instruções originais que estão abaixo do conteúdo dos tópicos.

**Agradecimento** (opcional, 1 parágrafo, bem sucinto)

O título da seção Agradecimentos deve ser alinhado à esquerda e grafado em negrito, com a primeira letra da palavra grafada em letra maiúscula. Trata-se de uma seção opcional, de no máximo três linhas, na qual o autor agradece àqueles que contribuíram de maneira relevante para o desenvolvimento do trabalho e elaboração do TCC, mas que não tiveram o envolvimento intelectual necessário à atribuição de coautoria. O autor deve se abster da menção ou citação dos nomes das empresas, instituições ou pessoas que permitiram ou contribuíram com o desenvolvimento do trabalho, a menos que esteja documentalmente autorizado a fazê-lo.

**Atenção:** antes de enviar o arquivo para o Move - Sistema de TCCs, remova todas as instruções originais que estão abaixo do conteúdo dos tópicos.

**Referências**

LITTLE, R. J. A.; RUBIN, D. B. Statistical Analysis with Missing Data. Wiley, 2002.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The Elements of Statistical Learning. Springer, 2009.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. International Joint Conference on Artificial Intelligence, 1995, p. 1137-1143.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), v. 67, n. 2, p. 301-320, 2005.

MENARD, S. Applied Logistic Regression Analysis. Sage Publications, 2002.

SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. Information Processing and Management, v. 45, n. 4, p. 427-437, 2009.

Ministério da Agricultura, Pecuária e Abastecimento [MAPA], 2022. Setor cervejeiro segue crescendo a cada ano, aponta anuário. Disponível em: https://www.gov.br/agricultura/pt-br/assuntos/noticias/setor-cervejeiro-segue-crescendo-a-cada-ano-aponta-anuario..Acesso em: 17 de junho de 2024

CARVALHO,. Impacto do aumento do número de cervejarias artesanais sobre o mercado de cervejas comerciais nacional. Repositorio.ufu.br, 2022

Descobrir LIMA (2023) - https://repositorio.ufu.br/bitstream/123456789/39812/1/CiênciaDadosAprendizado.pdf

ARAÚJO, F. B.; SILVA, P. H. A.; MINIM, V. P. R. Perfil sensorial e composição físico-química de cervejas provenientes de dois segmentos do mercado brasileiro. Ciência e Tecnologia de Alimentos, v. 23, n. 2, p. 121–128, ago. 2003.

KNIAZIEVA, Y. Guide to Movie Recommendation Systems Using Machine Learning. Disponível em: <https://labelyourdata.com/articles/movie-recommendation-with-machine-learning>.

O'BRIEN, R. M. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, v. 41, n. 5, p. 673-690, 2007.

HARRELL, F. E., Jr. Ordinal logistic regression. Em: **Springer Series in Statistics**. Cham: Springer International Publishing, 2015. p. 311–325.

MENCIONAR ESTUDO DO CORAÇÃO COMO UM PROBLEMA PARECIDO E JUSTIFICAR ESCOLHA DO

BENDER, R.; GROUVEN, U. Ordinal logistic regression in medical research. Journal of the Royal College of Physicians of London, v. 31, n. 5, p. 546–551, 1997.

PARECE TER O MODUS OPERANDI PARA MODELOS DE REGRESSÃO LOGISTICA ORDINAL

FERNANDES, A. A. T. et al. Read this paper if you want to learn logistic regression. **Revista de Sociologia e Política**, v. 28, n. 74, p. 006, 2020.

Disponível em: <https://www.sagepub.com/sites/default/files/upm-binaries/36660\_8.pdf>. Acesso em: 3 out. 2024b.

GARAVAGLIA, S.; DUN, A. S.; HILL, B. M. A smart guide to dummy variables: Four applications and a macro. Disponível em: <https://stats.oarc.ucla.edu/wp-content/uploads/2016/02/p046.pdf>. Acesso em: 3 out. 2024.

ARTIGOS INDICADOS PELO ORIENTADOR ACHO QUE NÃO SÃO O MEU CASO POIS HÁ ORDEM

McCullagh, P. (1980). Regression models for ordinal data. Journal of the Royal Statistical Society: Series B (Methodological), 42(2), 109-142.

Agresti, A. (2010). Analysis of Ordinal Categorical Data. John Wiley & Sons.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression. Wiley.

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. The Stata Journal, 6(1), 58-82.

Hernán, M. A., & Robins, J. M. (2010). Causal Inference: What If. Chapman & Hall/CRC.

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. The Stata Journal, 6(1), 58-82.

Long, J. S. (1997). Regression Models for Categorical and Limited Dependent Variables. SAGE Publications.

artigos para regressão logistica ordinal

**Apêndice ou Anexo** (opcional)

Os apêndices são textos e/ou documentos que foram elaborados pelos autores e que são importantes para complementar a argumentação do trabalho. Anexos são textos ou documentos que ilustram o trabalho, mas que não foram elaborados pelos autores. Apêndices deverão seguir as mesmas normas de formatação do restante do texto, inclusive para as figuras e tabelas. O TCC deverá conter no máximo 30 páginas, incluindo o(s) Apêndice(s) e/ou Anexo(s).

**Atenção:** antes de enviar o arquivo para o Move - Sistema de TCCs, remova todas as instruções originais que estão abaixo do conteúdo dos tópicos.