Capítulo 56
DOI: 10.53934/agronfy-2025-01-56

APRENDIZAGEM DE MÁQUINA NO ESTUDO DA ACEITAÇÃO SENSORIAL: UMA APLICAÇÃO AO CAFÉ

Fernanda Nascimento da Silva[1]; Ana Alice da Silva Xavier Costa[2] ;Valéria Paula Rodrigues Minim[1]*
1 – Universidade Federal de Viçosa; 2 – Universidade Federal da Paraíba

**Autor correspondente (Corresponding author) – Email: vprm@ufv.br*

Resumo: O perfil sociodemográfico e comportamental do consumidor em relação a um produto influencia a tomada de decisão dos consumidores e consequentemente a segmentação de mercado. No entanto, os métodos estatísticos tradicionais não permitem a correlação, ao mesmo tempo, entre os dados de perfil dos consumidores com a nota de aceitação do produto. Sendo assim, a aprendizagem de máquina apresenta-se como uma solução eficiente tanto para avaliação desses dados quanto à sua interpretação, já que essas variáveis têm potencial para afetar a aceitação. Diante disto, este capítulo objetivou apresentar como a aprendizagem de máquina, especificamente a árvore de decisão, pode ser usada para compreender dados de aceitação. Para isto, será apresentado um estudo de caso no qual dois cafés (tradicional e especial) foram avaliados em relação a sua aceitação. Foram também coletados dados de perfil dos consumidores de café. Os dados foram submetidos a análise descritiva e ao algoritmo de árvore de decisão. Com isto, foi possível observar quais as características dos consumidores tiveram maior importância para a nota de aceitação do produto, além de permitir a identificação de grupos de consumidores com notas iguais. Identificar o perfil dos consumidores possibilita entender quem está rejeitando ou aceitando aquele produto, contribuindo para a melhoria do mesmo e seu melhor direcionamento no mercado. Neste estudo de caso foi possível observar que o grau de escolaridade e a idade foram as variáveis de maior importância para a divisão dos dados, e que o algoritmo de árvore de decisão é uma técnica promissora para avaliação de dados sensoriais.

Palavras–chave: árvore de decisão; estudos do consumidor; perfil sociodemográfico; teste de aceitação.

INTRODUÇÃO

Não há dúvidas que a análise sensorial é uma grande aliada da indústria de alimentos por suas excelentes contribuições nas diversas fases do desenvolvimento de produtos, especialmente no que tange aos testes com consumidores. Para um produto ter sucesso no mercado competitivo, ele precisa estar alinhado com o interesse dos consumidores. Além disso, para alcançar esse sucesso, as empresas precisam investir em compreender o comportamento do consumidor frente ao produto. Neste sentido, as características sensoriais são fundamentais, pois o consumidor precisa se sentir atraído pelo produto.

No entanto, embora importantes, as características sensoriais sozinhas não são suficientes para explicar todo o comportamento de consumo, visto que muitas vezes a escolha do alimento ocorre antes do consumo propriamente dito (1). Por este motivo, além das características sensoriais, o perfil do consumidor, suas vivências e informações como marca, embalagens entre outras são exemplos do que se conhece como características não sensoriais e que têm grande impacto sobre as escolhas e a aceitação dos produtos (2).

Em estudos com consumidores é comum, por exemplo, coletar informações relacionadas ao perfil dos participantes juntamente com a resposta do teste de aceitação. Após a coleta dessas informações esses dados são normalmente avaliados por meio de estatísticas tradicionais como análise de variância, agrupamento e análise de correspondência. Contudo, esses métodos estatísticos não permitem correlacionar, ao mesmo tempo, dados de perfil dos consumidores com as notas de aceitação do produto, devido às pressuposições dessas metodologias. Sendo assim, a aprendizagem de máquina pode ser uma alternativa complementar aos métodos tradicionais, permitindo uma análise com todas as variáveis e uma compreensão da sua importância para a nota hedônica (3).

Um exemplo de algoritmo de aprendizagem de máquina supervisionada são as árvores de decisão, utilizadas tanto para problemas de regressão, quanto de classificação (4). Aprendizagem de máquina supervisionada é como ensinar uma máquina usando exemplos, ou seja, dados que possuem entradas e saídas conhecidas, para que ele possa prever ou classificar novos dados com base nesse treinamento (5). O funcionamento básico de uma árvore de decisão envolve a divisão recursiva dos dados em subconjuntos com base nas variáveis que melhor irão separar as notas de aceitação. Isso permite observar quais as características do consumidor são mais importantes para a diferença na aceitação do produto. Assim, é possível entender o perfil do consumidor que mais aceita o produto e o perfil do consumidor que o rejeita, permitindo que a indústria direcione melhor o marketing e atenda às necessidades desses consumidores.

Neste capítulo será abordado o uso do algoritmo de árvore de decisão para regressão no estudo da aceitação sensorial de cafés, com intuito de apresentar uma nova ferramenta de avaliação de dados sensoriais e demonstrar sua aplicação prática, evidenciando como esta pode ser uma ferramenta valiosa para melhor entendimento das variáveis que influenciam na aceitação dos produtos.

MATERIAL E MÉTODOS

A fim de exemplificar como técnicas de aprendizagem de máquina podem ser aplicadas para avaliar dados de aceitação sensorial, será apresentado um estudo de caso em que se objetivou utilizar a técnica de árvore de decisão para compreender como o perfil sociodemográfico influenciou a nota de aceitação sensorial de cafés.

A coleta de dados foi realizada com 120 consumidores de café, recrutados por meio de questionário on-line divulgado em redes sociais e por correio eletrônico para moradores de Viçosa/MG. Todos os participantes selecionados tinham mais que 18 anos, gostavam de café e não tinham restrição de consumo.

Perfil dos participantes

Antes da avaliação sensorial, cada participante preencheu um questionário contendo algumas informações sociodemográficas (idade, gênero e grau de instrução), bem como deviam declarar o quanto eles gostavam de café utilizando a escala hedônica estruturada de 9 pontos, que varia de 9-gostei extremamente a 1-desgostei extremamente.

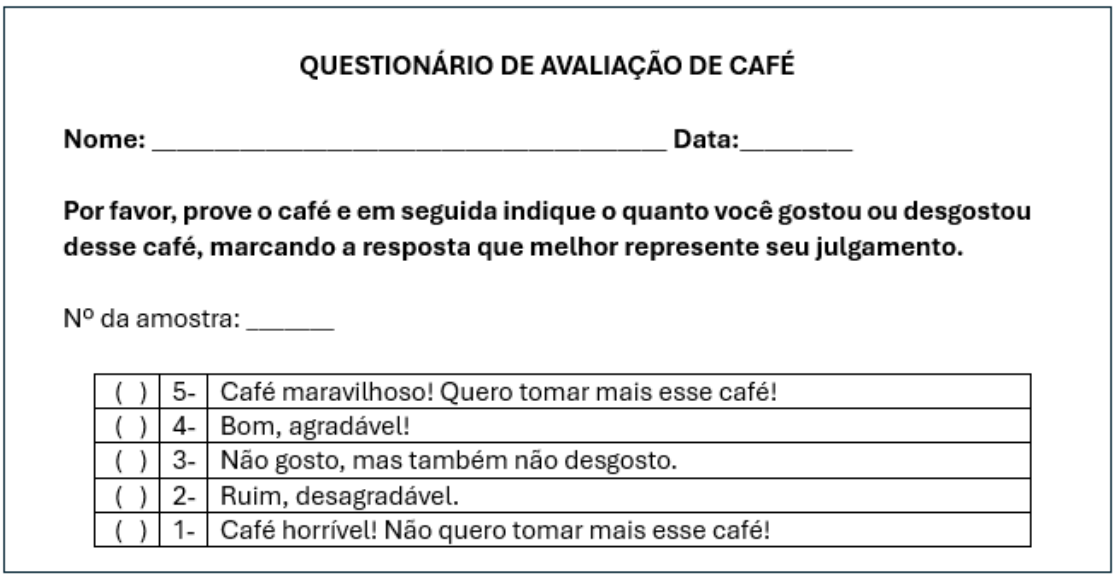
Amostras

Duas amostras de cafés comerciais, um tradicional e um especial, foram selecionadas para esta pesquisa. As amostras foram preparadas por infusão utilizando 10g de pó por 100mL de água mineral entre 92 e 96 °C, em papel filtro. As amostras foram preparadas próximo ao horário da análise e armazenadas em garrafas térmicas padronizadas e herméticas.

Avaliação da aceitabilidade sensorial

Após preencherem o questionário de recrutamento, os consumidores receberam as amostras de café (tradicional e especial) servidas em copos térmicos de poliestireno expandido (EPS) de 50 mL, identificadas com números de três dígitos aleatórios e apresentadas de forma monádica e balanceada. As amostras foram servidas apenas no momento da análise para evitar que esfriassem.

Os consumidores foram orientados a provarem as amostras e avaliarem as mesmas utilizando uma escala hedônica estruturada de 5 pontos, na qual 1 e 2 representavam região de rejeição, 3 representava indiferença e 4 e 5 aceitação, conforme escala criada por Costa (6), ilustrada na Figura 1.



QUESTIONÁRIO DE AVALIAÇÃO DE CAFÉ

Nome: _____ **Data:**_____

Por favor, prove o café e em seguida indique o quanto você gostou ou desgostou desse café, marcando a resposta que melhor represente seu julgamento.

Nº da amostra: _____

( )	5-	Café maravilhoso! Quero tomar mais esse café!
( )	4-	Bom, agradável!
( )	3-	Não gosto, mas também não desgosto.
( )	2-	Ruim, desagradável.
( )	1-	Café horrível! Não quero tomar mais esse café!

Figura 1 – Ficha de avaliação de café
Fonte: Costa (2024)

Análise de dados

Para a análise dos dados é crucial compreender o conjunto de variáveis. Neste estudo de caso, as variáveis independentes incluíram gênero, idade, grau de instrução e o quanto os consumidores gostavam de café. A variável dependente consistiu na nota de aceitação do café, ou seja, à medida que pode ser influenciada pelas variáveis independentes mencionadas anteriormente.

Para obter a matriz com todos os dados, as informações coletadas no questionário como gênero e grau de instrução foram organizadas na planilha como variáveis categóricas. Isso resultou em uma matriz na qual todas as variáveis independentes e a variável dependente foram dispostas em colunas, enquanto os consumidores foram representados nas linhas (Tabela 1).

Para análise e modelagem dos dados foram utilizadas bibliotecas Python. Primeiramente, as bibliotecas Pandas e NumPy foram empregadas para manipulação e processamento dos dados. Em seguida, as bibliotecas Matplotlib e Seaborn foram utilizadas para visualização dos dados em histogramas. Para modelagem foi utilizada a biblioteca scikit-learn. Os passos utilizados estão descritos abaixo e os scripts estão disponíveis para download em https://github.com/anaalicesiler/arvorededecisao

Tabela 1 - Matriz com os dados organizados para avaliação.

Consumidores	Gênero	Idade	Grau de instrução	Grau de gostar de café	Nota de aceitação
1	1	44	3	8	4
2	0	46	4	8	2
…	…	…	…	…	…
120	1	34	7	9	1

Gênero: 0=Feminino, 1=Masculino. Grau de instrução: 0 = Fundamental Incompleto, 1 = Fundamental Completo, 2 = Médio Incompleto, 3 = Médio Completo, 4 = Superior Incompleto, 5 = Superior Completo, 6 = Pós-Graduação Incompleto, 7 = Pós-Graduação Completo.

1) A planilha em um arquivo Excel foi lida e convertida para um DataFrame do pandas. Em seguida, foi obtido o resumo estatístico das variáveis numéricas, idade e gostar. A partir do DataFrame, foram preparados os conjuntos de dados X e y para o treinamento do modelo de aprendizado de máquina. Além disso, foi analisado o histograma da distribuição das respostas das variáveis independentes. Também foram visualizados os dados de aceitação em categorias, permitindo o entendimento da distribuição das notas em aceitação, neutralidade e rejeição.

2) Para modelagem foi utilizada a biblioteca scikit-learn empregando o módulo 'DecisionTreeRegressor' para treinamento da árvore de decisão, com validação cruzada e otimização dos hiperparâmetros.
    a) Hiperparâmetros como a profundidade máxima da árvore (max_depth), o número mínimo de amostras necessárias para dividir um nó interno (min_samples_split), e o número mínimo de amostras necessárias para estar em um nó de folha (min_samples_leaf) ajudam a regular a complexidade da árvore. Um ajuste cuidadoso desses parâmetros pode evitar que o modelo se ajuste demais aos dados de treinamento (overfitting) (6).

b) Após a criação do modelo final, foram realizadas predições nos dados de teste e avaliação do desempenho do modelo. A árvore de decisão e a obtenção da importância das características foram realizadas utilizando métodos disponíveis na biblioteca Scikit-learn.

Para este estudo de caso foi utilizado como métrica de avaliação do modelo o erro quadrático médio (Equação 1). Esta é uma medida que quantifica a média dos quadrados das diferenças entre os valores previstos por um modelo estatístico e os valores observados na realidade. Ou seja, quanto menor o erro, melhor a previsão do modelo.

$$\boldsymbol{EQM} = \left(\frac{\mathbf{1}}{\boldsymbol{n}}\right) * \sum(\boldsymbol{real} - \boldsymbol{previsão})^{\mathbf{2}}$$ **Equação 1-** Erro quadrático médio (EQM)

Onde:
n – tamanho da amostra
real – o valor real dos dados
previsão – o valor dos dados previstos

RESULTADOS E DISCUSSÃO

Estatística descritiva – distribuição de frequências

Para facilitar o entendimento dos resultados do teste de aceitação, a análise de frequência do perfil sociodemográfico foi realizada. A Tabela 2 apresenta a estatística descritiva dos dados de idade dos consumidores e o quanto eles declararam gostar de café. Os 120 consumidores tinham em média 35 anos com um desvio padrão de 12,67 anos. Destes, 25% tinham menos de 24 anos, enquanto 75% tinham menos de 44 anos, sugerindo uma distribuição que tende a se concentrar em torno das idades mais jovens, como também pode ser observado na Figura 1. A idade dos participantes é uma característica importante que contribui para melhor entendimento do comportamento do consumidor durante a avaliação e direcionamento mais eficiente de marketing do produto (7).

Tabela 2 – Estatística descritiva das variáveis quantitativas

Variáveis	n	média	DP	min	25%	50%	75%	max
Idade	120	35,76	12,67	18	24	34	44	74
Gostar	120	8,08	0,80	6	8	8	9	9

Idade: anos. Gostar: Nota hedônica variando de 9-gostei extremamente a 1-desgostei extremamente.
DP =desvio padrão.

Em relação ao quanto os consumidores declaram gostar de café, obteve-se uma média de 8,08, a qual indica “gosto muito” na escala de avaliação hedônica utilizada. A pontuação mínima foi de 6, indicando uma variação na aceitação, porém com desvio padrão relativamente baixo de 0,80, tendendo para valores mais altos. Como também pode ser observado na Figura 2, na qual 47,50% dos consumidores declararam gostar muito de café e 32,50% declararam gostar extremamente. O fato de os participantes da pesquisa terem declarado que gostam muito de café torna-se um fator importante já que são eles que consumirão o produto. Sendo assim, as respostas dos consumidores são cruciais para o sucesso do produto no mercado (8).

Na Figura 2 pode ser observado também os histogramas de variação do gênero e do grau de instrução dos 120 consumidores. Sendo a maioria do sexo feminino, 59,17% enquanto o gênero masculino representou 40,83% dos consumidores. Em relação à escolaridade, os consumidores em sua maioria eram estudantes cursando a graduação (37,50%) ou a pós-graduação (27,50%). O que condiz com pesquisas atuais, que demonstram que estudantes consomem café regularmente e com grande frequência (9).

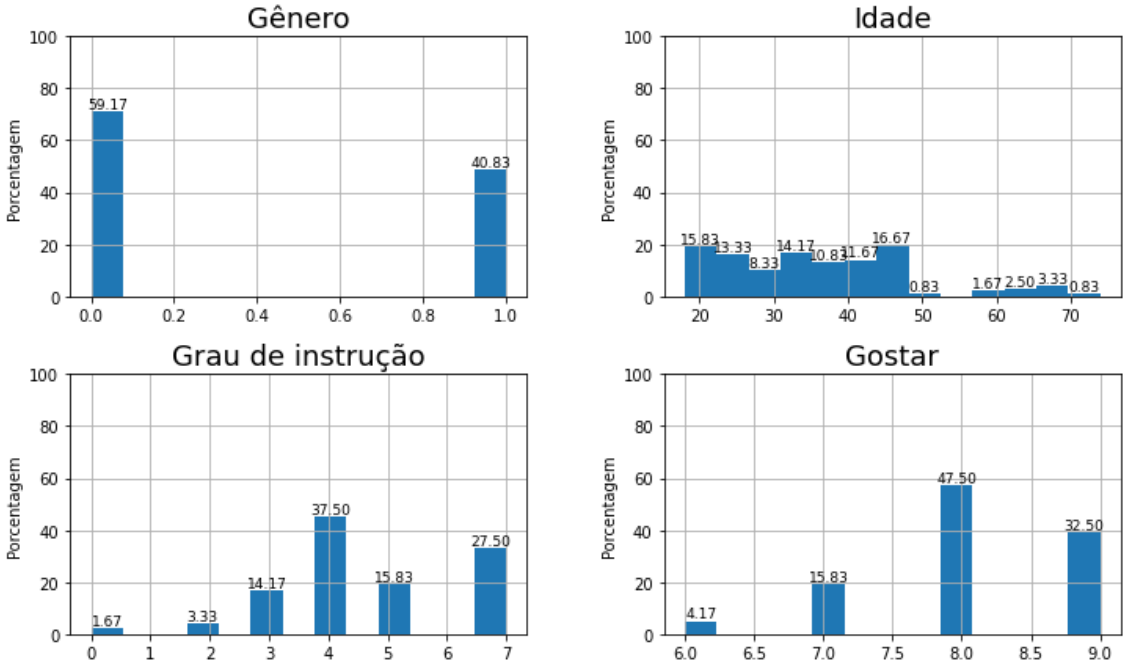


Figura 2 - Histograma da distribuição das variáveis independentes (perfil sociodemográfico e de consumo).

Gênero: 0=Feminino, 1=Masculino. Grau de instrução: 0 = Fundamental Incompleto, 1 = Fundamental Completo, 2 = Médio Incompleto, 3 = Médio Completo, 4 = Superior Incompleto, 5 = Superior Completo, 6 = Pós-Graduação Incompleto, 7 = Pós-Graduação Completo. Idade e gostar: Frequência quantitativa.

Influência das características sociodemográficas na aceitação sensorial do café

A Figura 3 apresenta a variação dos consumidores em relação à aceitação, indiferença e rejeição de ambos os tipos de café. Conforme observado, o café tradicional foi aceito por 56,67% dos consumidores, enquanto 44,33% rejeitaram ou foram indiferentes a este café. Para o mesmo grupo de consumidores, o café especial teve uma aceitação ligeiramente menor, aceito por 52,5% dos consumidores.

As médias hedônicas foram muito próximas, conforme demonstrado na Tabela 3 (em uma escala de cinco pontos). Esse estudo de caso demonstra que a média sozinha pode não representar de forma ideal o conjunto de dados quando se trata de aceitação pelos consumidores, visto que ela pode ser facilmente influenciada por *outliers.* Sendo assim, entender o perfil dos grupos de consumidores é importante para a observação de quais indivíduos estão aceitando ou rejeitando o produto (10).

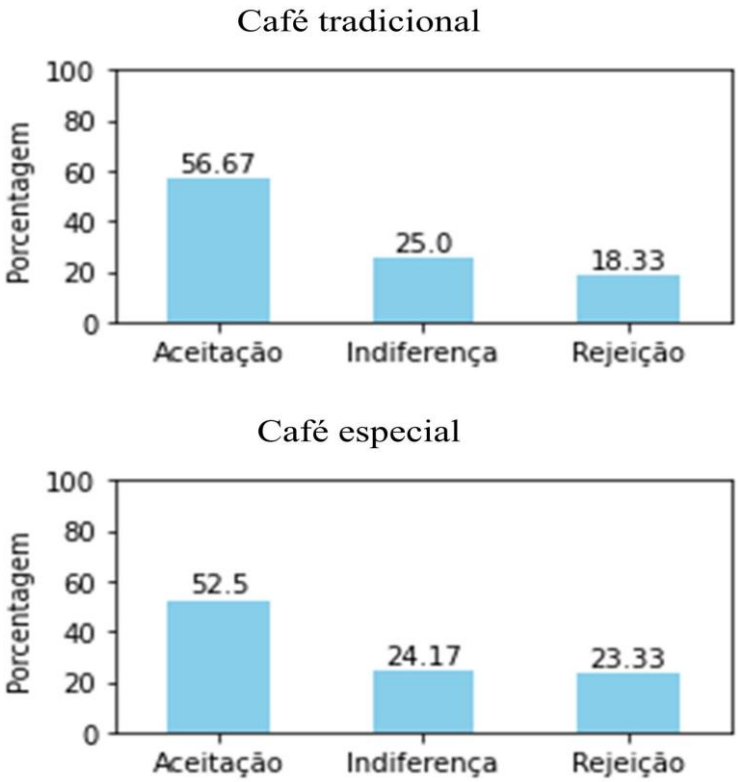


Figura 3 - Histograma de Frequência da porcentagem de aceitação, indiferença e rejeição dos consumidores para os cafés

Aceitação: notas 4 a 5. Indiferença: nota 3. Rejeição: notas 1 a 2.

Tabela 3 – Média e desvio padrão da aceitação sensorial para os cafés (escala de cinco pontos)

Cafés	Nota Média	Desvio padrão
Tradicional	3,5	1,06
Especial	3,44	1,10

A análise estatística descritiva dos dados auxilia na interpretação da análise da árvore de decisão, sendo de suma importância para entendimento do comportamento dos dados. Além disso, ela permite uma melhor compreensão dos resultados obtidos por análises estatísticas mais robustas e compreensão do comportamento de algoritmos como os de aprendizagem de máquina.

Aplicação da aprendizagem de máquina na avaliação do café - Árvore de decisão

Para interpretar a árvore de decisão deve-se observar o caminho da raiz até as folhas. Cada nó na árvore representa uma decisão baseada em uma variável preditora, que irá dividir os dados em ramos que irão representar os resultados baseados nessa decisão (11). O nó raiz é o nó superior da árvore, onde começa a primeira divisão dos dados. Os nós internos são nós intermediários que representam decisões baseadas em uma variável, possuindo ramos que levam a outros nós (internos ou folhas). As Folhas são os nós finais que não se dividem mais, representando a saída ou a previsão do modelo. Nesse caso, o

sistema aprende com os dados, entende suas relações, realiza os cálculos a fim de entender quais são os nós mais importantes e ajusta as condições, as quais irão determinar os caminhos que a árvore irá percorrer.

No estudo de caso aqui apresentado, a árvore de decisão possibilitou a visualização de forma mais intuitiva de como o perfil dos consumidores influenciou na nota de aceitação do café. A Figura 4 apresenta a árvore de decisão para a aceitação do café tradicional em relação às variáveis preditoras testadas (idade, grau de instrução, gênero e o quanto os consumidores declararam gostar de café).

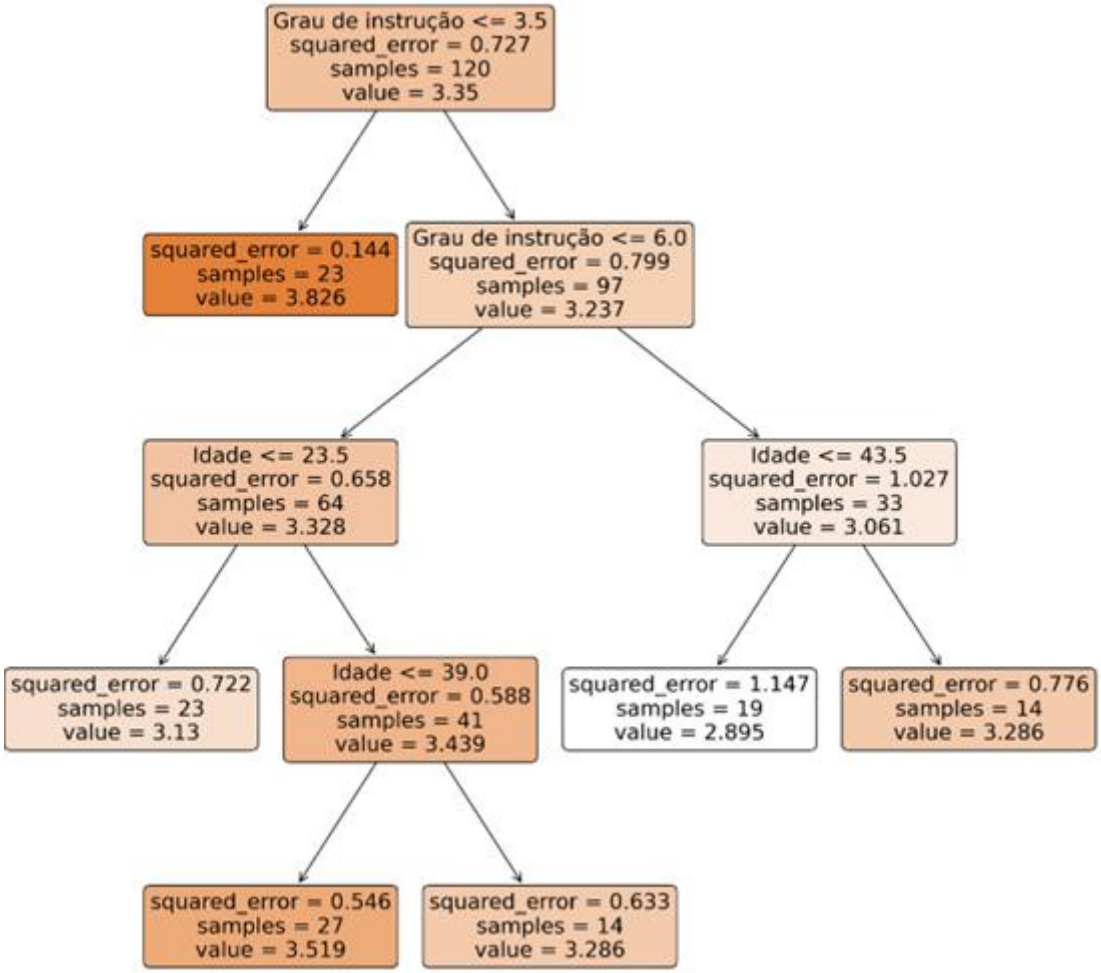


Figura 4 - Árvore de decisão para aceitação do café tradicional em relação ao perfil dos consumidores

Squared_erro: erro quadrático (critério de divisão para criar a árvore de decisão). Sample: quantidade de consumidores. Value: Nota de aceitação dos consumidores.

O primeiro destaque da Figura 4 é que o grau de instrução aparece no topo da árvore, representando o nó raiz. Isto significa que esta foi a variável que mais contribuiu para a redução da entropia dos dados de aceitação. A entropia é uma medida de incerteza ou impureza em um conjunto de dados. No contexto de árvores de decisão, a entropia quantifica a desordem ou a imprevisibilidade das classificações. Portanto, ao selecionar a variável que mais reduz a entropia, a árvore de decisão escolhe a variável que melhor

organiza os dados, criando divisões mais puras e melhorando a capacidade preditiva do modelo (12).

Após o nó raiz, a árvore de decisão faz uma divisão dos consumidores (Figura 4). No lado esquerdo, encontramos um nó final (folha) onde a condição atendida é grau de instrução menor ou igual a 3,5, contendo 23 consumidores. Esses consumidores, cujas notas de aceitação são influenciadas pelo grau de instrução inferior a 3,5 (ensino médio completo) têm uma nota hedônica média de 3,8.

Para uma melhor compreensão da segmentação dos consumidores pela árvore, é importante observar os nós finais (folhas), pois estes refletem todas as condições de decisão percorridas desde a raiz. A idade, por exemplo, demonstra ser uma variável importante para a nota de aceitação em cinco grupos de consumidores, identificáveis nos nós finais (folhas) da árvore de decisão (Figura 4).

Por exemplo, no último nó final (folha) do lado direito da árvore, observa- se um segmento de 27 consumidores que deram uma nota hedônica média de 3,5 para o café tradicional. Ao percorrer a árvore, notamos que estes consumidores têm um grau de instrução entre 3,5 e 6,0 (superior incompleto a pós-graduação incompleto), e têm idades entre 23,5 e 39 anos. Para o restante da árvore, o processo é semelhante: observar os nós finais (folhas) e analisar a trajetória que estes seguiram conforme as condições de decisão (Figura 4). Quando o galho vai em direção à direita da árvore quer dizer que atendeu a condição, ou seja, aquele grupo de consumidores tem características menor ou igual ao valor determinado para aquela variável. Enquanto o lado esquerdo da árvore implica em que a condição não foi atendida, ou seja, aquele grupo de consumidores possui características maiores do que o valor determinado para aquela variável. O algoritmo de árvore de decisão escolhe o valor específico de uma variável dividindo os dados de modo a minimizar o erro na previsão das avaliações sensoriais dos cafés. Isto ajuda a entender como diferentes características dos consumidores estão associadas às suas avaliações sensoriais dos cafés.

Na Figura 4 observa-se o erro quadrático médio que é uma medida de desempenho do modelo para a predição de novos dados. Quanto menor este valor, melhor o desempenho do modelo em termos de predição (13). Como pode ser observado, os retângulos com menor erro quadrático estão representados por cores mais fortes na árvore, o que significa que naquele nó para aquele grupo de consumidores o modelo está conseguindo prever melhor a nota de aceitação.

Porém, neste estudo de caso, a árvore de decisão está sendo utilizada apenas para a visualização da influência das características de perfil dos consumidores na nota de aceitação do café. Caso o intuito fosse fazer predições, seria interessante que houvesse mais dados para o treinamento do modelo (14). Fazer predições refere-se à capacidade do modelo de estimar a nota de aceitação de novos consumidores com base em seus perfis. Por exemplo, supondo que um analista tenha acesso ao perfil (idade, grau de instrução etc.) de um consumidor, o modelo poderia prever a sua nota de aceitação para o café mesmo sem a sua participação na análise sensorial daquela amostra. No entanto, para que essas predições sejam precisas, o modelo precisaria ter sido treinado com uma quantidade suficiente de dados representativos.

A Figura 5 resume a importância das variáveis testadas. No caso do café tradicional apenas o grau de instrução e a idade influenciaram na nota hedônica, conforme já demonstrado na Figura 3. O grau de instrução dos consumidores foi a variável de maior importância para a divisão dos dados, representando 71,86% da importância total. Isso indica que o nível educacional dos participantes tem um papel preponderante na determinação de suas preferências em relação ao café tradicional.

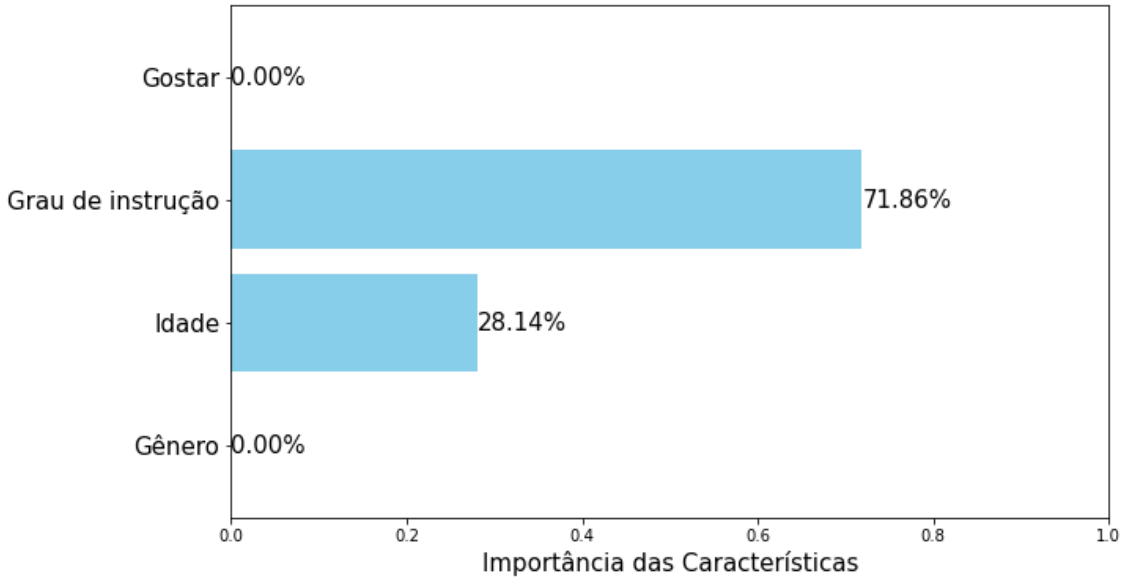


Figura 5 - Histograma da importância das variáveis independentes (perfil sociodemográfico e de consumo) para aceitação do café tradicional

As Figuras 6 e 7 representam como as variáveis influenciaram a nota hedônica do café especial. Neste caso, todas as variáveis estudadas impactaram a nota atribuída ao café, com maior importância para a variável idade, com 74,71% da importância total, seguida de quanto os consumidores declararam gostar de café, grau de instrução e gênero (Figura 7).

Podemos observar vários nós finais (folhas) na árvore de decisão (Figura 6), podendo tirar conclusões sobre diferentes grupos de consumidores. Diante disso, torna-se interessante observar o grupo de consumidores com menor nota e compreender o seu perfil. No centro da árvore (retângulo branco) é possível observar um grupo de 10 consumidores que deram nota média de 2,4. Ao percorrer a árvore é possível concluir que esses são influenciados apenas pela idade, variando entre 32,5 e 36,5 anos. A Figura 5 também apresenta um retângulo laranja mais escuro com um grupo de 9 consumidores que apresentaram a maior média hedônica (4,4). Ao percorrer a árvore é possível observar que esses são consumidores que gostam extremamente de café, com idade entre 21,5 e 27,5 anos.

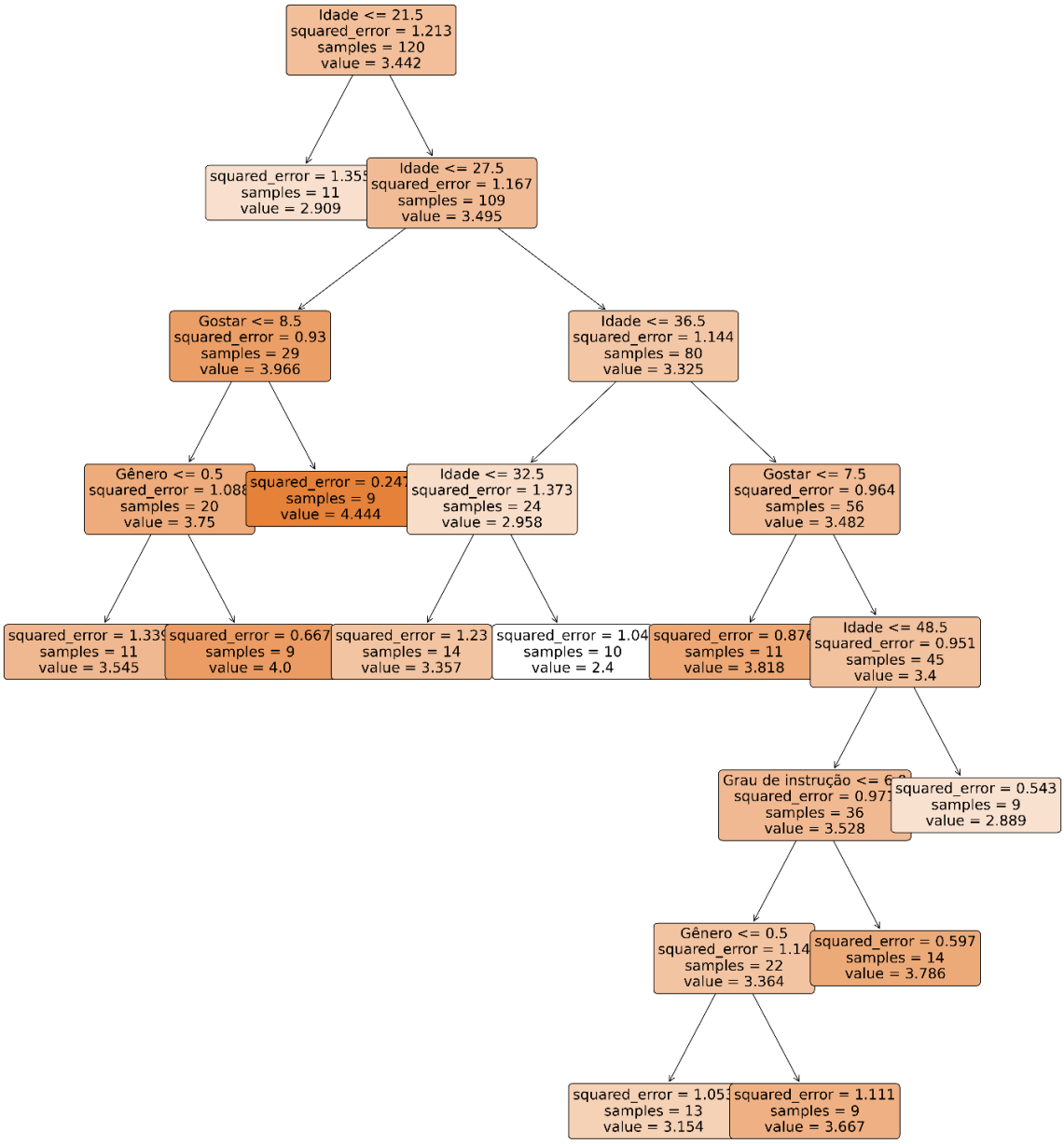


Figura 6 - Árvore de decisão para amostra de café especial em relação ao perfil dos consumidores.

Squared_erro: erro quadrático (critério de divisão para criar a árvore de decisão). Sample: quantidade de consumidores. Value: Nota de aceitação dos consumidores.

A árvore de decisão para o café especial (Figura 6) apresentou um número maior de divisões na árvore do que a do café tradicional (Figura 4), o que pode dificultar a visualização gráfica, porém permite o melhor entendimento de como as variáveis estão interagindo e influenciando na decisão. Quanto mais a variável aparece na árvore de decisão, mais importante ela é para a previsão (16).

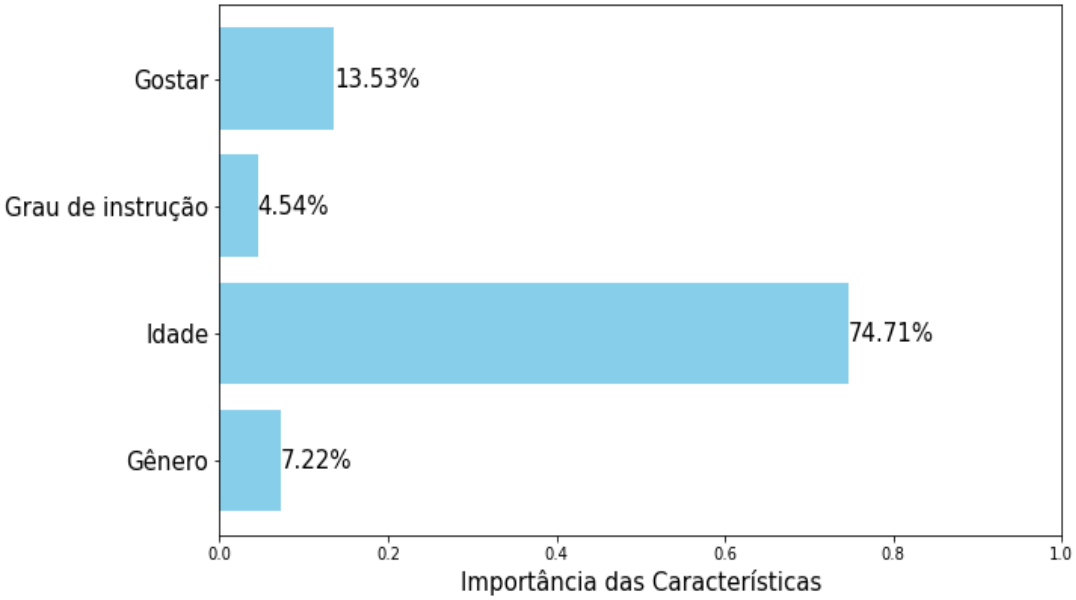


Figura 7 - Histograma da importância das variáveis independentes (perfil sociodemográfico e de consumo).

Esses resultados demonstram a importância de analisar detalhadamente os perfis dos consumidores em estudos de aceitação de produtos. A segmentação dos consumidores, mostrada pela árvore de decisão, oferece informações valiosas para melhor atender às preferências dos diferentes grupos, aumentando a satisfação do consumidor e o sucesso do produto no mercado.

Considerações em relação ao uso da técnica de árvore de decisão

A partir da técnica de árvore de decisão foi possível observar as variáveis que tiveram maior importância para as diferenças na nota de aceitação dos cafés. Além disso, foi possível observar os grupos de consumidores formados com notas semelhantes e quais eram os perfis desses consumidores. O uso do algoritmo de aprendizagem de máquina possibilitou a segmentação dos consumidores de forma mais visual e interpretativa, permitindo uma identificação mais precisa sobre os fatores de maior influência sobre a aceitação.

Observar as características sociodemográficas dos participantes da pesquisa permite que a indústria tome decisões mais assertivas de melhorias das características sensoriais ou de marketing para o produto, pois direciona a comunicação para seu público-alvo. Neste estudo de caso, foi possível observar que o café tradicional teve a maior aceitação por grupos com menor escolaridade, e que os mais jovens, especialmente os menores de 21 anos, e os com idade entre 32,5 e 36,5 tenderam a rejeitar o café especial.

Apesar dos dados deste estudo serem pouco representativos e ambas as amostras terem nota hedônica média próxima (3,5), a utilização da técnica de árvore de decisão possibilitou entender a relação do perfil dos consumidores com a nota de aceitação. Sendo assim, explorar esta alternativa para avaliação de dados com maior número de consumidores é uma oportunidade de gerar mais *insights* para a indústria de alimentos, em relação ao desenvolvimento e melhoria de produtos e direcionamento de marketing.

CONCLUSÕES

O modelo de aprendizagem de máquina denominado de árvore de decisão possibilita a observação das variáveis de maior importância para a decisão na nota de aceitação dos produtos. Neste trabalho foi possível observar o perfil dos consumidores e como ele influenciou a avaliação dos cafés. Além de observar a árvore de decisão, é fundamental relacioná-la com a importância das variáveis, além da estatística descritiva. Neste estudo há uma limitação em relação ao número de consumidores, pois se sabe que o ideal é que haja maior representatividade dos consumidores para a utilização mais eficiente da técnica. No entanto, o intuito deste trabalho foi demonstrar a possibilidade de avaliação dos dados sensoriais por meio de aprendizagem de máquina, uma técnica promissora que tem grande potencial para auxiliar na segmentação de mercado, desenvolvimento e melhorias do produto.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001; do Conselho Nacional do Desenvolvimento Científico e Tecnológico – CNPq e da Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG.

REFERÊNCIAS

1. Saldaña E, Martins MM, Behrens JH, Valentin D, Selani MM, Contreras-Castillo CJ. Looking at non-sensory factors underlying consumers' perception of smoked bacon. Meat Science. 2020;163:108072.

2. Minim VPR. Análise sensorial: estudos com consumidores. 4ª ed. Viçosa, MG: Ed. UFV; 2018.

3. Bishop CM. Model-based machine learning. Philosophical transactions Series A, Mathematical, physical, and engineering sciences. 2013;371.

4. Jena M, Dehuri S. DecisionTree for Classification and Regression: A State-of-the Art Review. Informatica (Slovenia). 2020;44.

5. Syed AN. An overview of supervised machine learning. International Research Journal of Modernization in Engineering Technology and Science. 2024;06(03).

6. Costa AASX. “EMOCOFFEE”: um novo instrumento para avaliação da resposta emocional e afetiva do consumidor ao café. [Doutorado]. Viçosa: Universidade Federal de Viçosa; 2024.

7. Trivedi UB, Bhatt MB, Srivastava P. Prevent Overfitting Problem in Machine Learning: A Case Focus on Linear Regression and Logistics Regression. Innovations in Information and Communication Technologies (IICT-2020); 2021.

8. Vukasović T. Understanding the Consumers' Personal Characteristics as the Starting Point for Targeted Marketing. Management Science. 2020;15:29-41.

9. Tugba A. Food Product Development: From the Consumers Aspect. E3S Web of Conferences. 2021.

10. Nair D, Shahabuddin Shaikh T. Attitude and Perception towards Coffee Consumption Effects: Insights from College Students in Mumbai. Pelita Perkebunan (a Coffee and Cocoa Research Journal). 2023.

11. Lei Z, Yin D, Mitra S, Zhang H. Swayed by the reviews: Disentangling the effects of average ratings and individual reviews in online word-of-mouth. Production and Operations Management. 2022;31:2393-2411.

12. Gupta B, Rawat A, Jain A, Arora A, Dhami N. Analysis of Various Decision Tree Algorithms for Classification in Data Mining. International Journal of Computer Applications. 2017;163:15-9.

13. Zighed DA, Ritschard G, Marcellin S. Asymmetric and Sample Size Sensitive Entropy Measures for Supervised Learning. Advances in Intelligent Information Systems. 2010.

14. Singh A, Singh P, Singh DK. Application of decision tree for data mining in sales prediction: A case study. International Journal of Computer Applications. 2018;183:1-5.

15. Abu Zohair LM. Prediction of Student's performance by modelling small dataset size. International Journal of Educational Technology in Higher Education. 2019;16.

16. Saputri AE. Analysis of Decision Tree Algorithm to Predict Heart Disease. Journal of Physics: Conference Series. 2023.

17. Kazemitabar SJ, Amini AA, Bloniarz A, Talwalkar A, editors. Variable Importance Using Decision Trees. Neural Information Processing Systems; 2017.