**APRENDIZADO DE MÁQUINA PARA ANÁLISE DA QUALIDADE DO VINHO**

Iuri Almeida Pereira

Bacharelado em Ciência da Computação – Universidade Tecnológica Federal do Paraná

LAMIA

*iuripereira.2022@alunos.utfpr.edu.br*

**RESUMO**

O objetivo deste estudo é desenvolver um modelo de aprendizado de máquina capaz de prever vinhos de alta qualidade com base na análise de dados físico-químicos utilizados em sua produção. A proposta visa otimizar a seleção de vinhos, permitindo que sommeliers e apreciadores concentrem seus esforços na avaliação daqueles que atendem aos critérios desejados. O processo de produção de vinho envolve uma ampla gama de fatores físico-químicos que influenciam diretamente em sua qualidade. Tradicionalmente, essa avaliação é realizada por especialistas humanos, mas enfrenta desafios de subjetividade e fadiga sensorial ao longo do tempo. A solução proposta é empregar técnicas de aprendizado de máquina, permitindo a análise objetiva de dados físico-químicos para determinar a qualidade do vinho.

O modelo será desenvolvido com base em dados coletados de vinhos da região de Minho, Portugal, utilizando cinco diferentes algoritmos de classificação, incluindo "Random Forest," "Extra Trees," "Logistic Regression," "K Neighbors Classifier," e "Extreme Gradient Boosting." Os resultados indicam que modelos baseados em árvores de decisão, notavelmente o "Extra Trees," apresentam desempenho superior. O modelo conseguiu atingir os objetivos, prevendo os vinhos de boa e má qualidade. O aprendizado de máquina complementa a expertise humana, reduzindo a carga de trabalho dos sommeliers ao pré-selecionar vinhos que atendem às expectativas de qualidade. Isso não visa substituir a avaliação sensorial, mas sim aprimorá-la, incentivando a produção de vinhos de alta qualidade e garantindo a satisfação dos consumidores.

**Palavras-chave**: Vinho; Aprendizado de máquina; Qualidade; Análise físico-química; Inovação.

***ABSTRACT***

*The aim of this study is to develop a machine learning model capable of predicting high-quality wines based on the analysis of physico-chemical data used in their production. The proposal aims to optimize the selection of wines, allowing sommeliers and enthusiasts to focus their efforts on evaluating those that meet the desired criteria. The wine production process involves a wide range of physico-chemical factors that directly influence its quality. Traditionally, this assessment is carried out by human experts, but it faces challenges of subjectivity and sensory fatigue over time. The proposed solution is to employ machine learning techniques, enabling an objective analysis of physico-chemical data to determine wine quality. The model will be developed based on data collected from wines in the Minho region of Portugal, using five different classification algorithms, including "Random Forest," "Extra Trees," "Logistic Regression," "K Neighbors Classifier," and "Extreme Gradient Boosting." The results indicate that tree-based models, notably "Extra Trees," show superior performance. The model successfully achieved its goals, predicting wines of good and poor quality. Machine learning complements human expertise, reducing the workload of sommeliers by preselecting wines that meet quality expectations. This is not intended to replace sensory evaluation but rather to enhance it, encouraging the production of high-quality wines and ensuring consumer satisfaction.*

***Keywords*** *Wine; Machine Learning; Quality; Physicochemical Analysis; Innovation..*

**INTRODUÇÃO**

A busca incessante por alcançar a qualidade ótima do vinho desempenha um papel crucial tanto na satisfação dos consumidores como na preservação da reputação dos produtores. Afinal, quem deseja degustar uma bebida que, além de não ser prazerosa, possa até prejudicar a saúde? Avaliar e atribuir uma qualidade a um vinho é uma tarefa complexa, que envolve uma apreciação minuciosa de seus sabores, aromas, texturas e nuances. É nesse minucioso processo que a identidade de cada garrafa é forjada.

O uso de aprendizado de máquina na análise da qualidade do vinho visa otimizar a produção e garantir que apenas vinhos de alta qualidade alcancem o mercado. No entanto, vale ressaltar que o objetivo não é eliminar o papel dos sommeliers, cuja expertise é inestimável. Em vez disso, a ideia é tornar o processo mais eficiente, permitindo que os sommeliers concentrem seus esforços na avaliação dos vinhos que atendem às expectativas.

A análise sensorial desempenha um papel fundamental na atribuição de uma nota a um vinho. No entanto, as limitações são evidentes. A fadiga sensorial, que pode surgir após testes repetitivos, e a subjetividade da percepção sensorial de cada indivíduo são desafios que podem afetar a consistência das avaliações. É exatamente aí que a aprendizagem de máquina entra em cena, oferecendo uma solução promissora. Ao pré-selecionar os vinhos que atendem às expectativas por meio de análises objetivas, o aprendizado de máquina pode reduzir a carga de trabalho dos sommeliers e, ao mesmo tempo, incentivar os produtores a aprimorar continuamente a qualidade de seus vinhos.

Em resumo, a fusão de expertise humana com a precisão da tecnologia está pavimentando o caminho para uma indústria vinícola mais eficiente e de maior qualidade. O próximo passo é explorar as metodologias e modelos específicos utilizados nesse processo para compreender melhor como a inteligência artificial e o aprendizado de máquina estão contribuindo para a excelência na produção e avaliação de vinhos.

**1 REFERENCIAL TEÓRICO**

 Para a produção de vinhos de qualidade, é necessária a utilização adequada de metodologias laboratoriais, além dos cuidados realizados no vinhedo e na vinícola. É extremamente importante que os profissionais (enólogos, estudantes e produtores) conheçam as técnicas analíticas básicas empregadas na avaliação da composição físico-química do vinho e os parâmetros utilizados – para comprovar o seu enquadramento – nos padrões de identidade e qualidade estabelecidos pelo Ministério da Agricultura, Pecuária e Abastecimento (Mapa) (RIZZON, L. A. e SALVADOR, M. B. G. 2010).

A análise físico-química poderá contribuir para determinar os aspectos qualitativos (RIZZON, L. A. e SALVADOR, M. B. G. 2010). A determinação das propriedades físico-químicas dos vinhos possibilita identificar ou não sua qualidade (CASTILHOS, M. B. M. DE; DEL BIANCHI, 2011).

A classificação precisa do vinho é fundamental para a segurança alimentar. Garantir que os vinhos estejam dentro dos padrões de qualidade estabelecidos é essencial para evitar riscos à saúde dos consumidores. Problemas como contaminação, adição de substâncias inadequadas ou desequilíbrios químicos podem comprometer a segurança e a integridade do produto final. Portanto, a classificação rigorosa da qualidade do vinho é uma medida preventiva que contribui para a proteção dos consumidores e a conformidade com os regulamentos de segurança alimentar (SKIBSTED, L et al., 2010).

No entanto, o processo tradicional de classificação, baseado em especialistas humanos, apresenta desafios significativos. A subjetividade inerente à percepção sensorial de cada indivíduo e a fadiga sensorial ao longo do tempo são limitações que podem afetar a consistência e a objetividade das avaliações (MIERCZYNSKA-VASILEV, A.; SMITH, P. A., 2023). A automação do processo de classificação, por meio do uso de inteligência artificial (IA) e redes neurais (RN), oferece uma abordagem promissora para superar essas limitações e obter resultados mais confiáveis e padronizados (DI, S.; YANG, Y., 2022).

Ao automatizar o processo de classificação da qualidade do vinho, é possível reduzir a dependência de especialistas humanos altamente treinados, tornando o processo mais eficiente e acessível. Dessa forma, a automatização do processo de classificação da qualidade do vinho contribui para a excelência da indústria vinícola, ao proporcionar maior confiabilidade, padronização e eficiência no processo de avaliação. A aplicação da IA nesse contexto não substitui a expertise humana, mas complementa e aprimora o trabalho dos especialistas, permitindo uma classificação mais precisa, consistente e confiável (BHARDWAJ, P. et al. 2022).

Um computador observa alguns dados, monta um modelo baseado nos dados e usa o modelo como uma hipótese sobre o mundo e um software que pode resolver problemas (RUSSEL, STUART J., 2004). Qualquer componente de um agente pode ser melhorado através do aprendizado de máquina.

Os componentes desses agentes incluem: Um mapeamento direto de condições no estado atual para ações; um meio para deduzir propriedades relevantes do mundo a partir da sequência de percepções; informações sobre o modo como o mundo evolui e sobre os resultados de ações possíveis que o agente pode realizar; informações de utilidade indicando a desejabilidade dos estados do mundo; informações de valores de ações indicando a desejabilidade das ações; metas que descrevem os estados mais desejáveis; um gerador de problema, um crítico, e elemento de aprendizagem para permitir que o sistema melhore (RUSSEL, STUART J., 2004).

Quando a saída for um de um conjunto finito de valores (como ensolarado/nublado/chuvoso ou verdadeiro/falso), o problema da aprendizagem será chamado classificação. Quando for um número (como a temperatura de amanhã, medida como um número inteiro ou real), o problema de aprendizagem será chamado regressão (admitidamente, obscuro) (RUSSEL, STUART J., 2004).

**2 METODOLOGIA**

Nesta seção, descreve-se a metodologia adotada para o experimento de classificação da qualidade do vinho, distinguindo entre "bom" e "ruim". Essa avaliação de qualidade é baseada nas propriedades físico-químicas de vinhos, tanto tintos quanto brancos. Os dados que contêm informações sobre cada vinho foram obtidos a partir do (UCI Machine Learning Repository) e foram originalmente coletados por Paulo Cortez em (CORTEZ, 2009).

A primeira etapa envolveu a definição de objetivos. Isso incluiu uma análise inicial dos dados e a subsequente criação de um modelo de treinamento para classificar vinhos como bons ou ruins. Duas bases de dados foram extraídas: uma contendo 1.599 amostras de vinho tinto e a outra com 4.898 amostras de vinho branco, ambas variantes do "Vinho Verde" produzido na região do Minho, Portugal (VERDES, 2023). Essas amostras foram coletadas no período de maio de 2004 a fevereiro de 2007, utilizando apenas amostras com denominação de origem protegida que foram testadas pela Comissão de Viticultura da Região dos Vinhos Verdes (CVRVV) (VERDES, 2023).

O objetivo geral é determinar se um vinho é de qualidade superior ou inferior, permitindo que os *sommeliers* se concentrem na avaliação das amostras aprovadas. Isso reduziria o desperdício e, ao mesmo tempo, impulsionaria a melhoria da produção de vinhos, incentivando futuros produtores a aprimorarem seus esforços durante a produção.

As propriedades físico-químicas usadas para treinar e encontrar o melhor modelo de classificação de qualidade incluem: acidez fixa, acidez volátil, ácido cítrico, açúcar residual, cloretos, dióxido de enxofre livre, dióxido de enxofre total, densidade, pH, sulfatos e teor alcoólico. Há um total de 11 colunas relacionadas às propriedades físico-químicas, e uma décima segunda coluna chamada *quality*, que representa a pontuação atribuída a cada vinho por um sommelier durante a degustação.

Para preparar os dados, foi realizado um processo de normalização, criando dois *dataframes* separados, um para vinhos tintos e outro para vinhos brancos. Foi adicionada uma coluna chamada *color* a cada *dataframe* para indicar o tipo de vinho (tinto ou branco). Posteriormente, os dois *dataframes* foram concatenados, resultando em uma base de dados contendo um total de 6.497 amostras e 13 colunas.

Para classificar os vinhos como "bons" ou "ruins", foi necessário analisar a frequência da variável *quality*. Com base nessa análise, foi criada a coluna *classification*, que recebeu valores binários (0) para vinhos considerados ruins (nota < 6) e (1) para vinhos considerados bons (nota ≥ 6). Esta coluna *classification* foi utilizada como a variável-alvo para o treinamento dos modelos.

No estágio de modelagem de dados, com base na base de dados preparada, as colunas que continham valores não numéricos foram transformadas, uma vez que as variáveis com formato *string* não são adequadas para treinamento. Foram utilizados cinco modelos para o treinamento: *Random Forest*, *Extra Trees*, *Logistic Regression*, *K Neighbors Classifier* e *Extreme Gradient Boosting*. Cada um desses modelos foi empregado para classificar a qualidade dos vinhos com base nas propriedades físico-químicas mencionadas anteriormente. A variável *quality* desempenhou um papel importante ao ajudar a determinar a variável *classification*.

Após o treinamento dos modelos, foi realizada uma avaliação de seu desempenho. A métrica utilizada para verificar a eficácia dos modelos na classificação da qualidade do vinho foi a Acurácia. Com base na acurácia de cada modelo, os resultados foram plotados para permitir uma análise comparativa e identificar o melhor modelo para a classificação de qualidade do vinho.

**3 RESULTADOS E DISCUSSÃO**

Nesta seção, apresenta uma análise do desempenho de cinco modelos de classificação: *Random Forest*, *Extra Trees*, *Logistic Regression*, *K Neighbors Classifier* e *Extreme Gradient* *Boosting*, usados para determinar a qualidade do vinho. Além disso, fornece uma breve explicação de como cada modelo opera.

Ao avaliar qual modelo seria o mais adequado para classificar a qualidade do vinho, procurou otimizar hiperparâmetros específicos sempre que aplicável, a fim de garantir que não houvesse diferenças significativas nos resultados de acurácia. Para os modelos que possuíam hiperparâmetros distintos, buscou a configuração mais eficaz, visando obter os melhores resultados possíveis.

3.1 ARVORE DE DECISÃO

Nesta seção, explicaremos o conceito da árvore de decisão, pois isso ajudará a compreender o funcionamento dos modelos que se baseiam nesse método para alcançar a classificação desejada.

Ao criar uma árvore de decisão, o algoritmo busca a melhor variável para iniciar a construção de um nó. Em seguida, ele determina o ponto ideal para dividir os dados, separando-os em duas partes distintas. Um parâmetro-chave em árvores de decisão é o *n\_estimators*, que representa o número de árvores de decisão a serem criadas e combinadas.

Para realizar a classificação, o algoritmo verifica qual resultado de classe foi o mais frequente e seleciona esse modelo como o melhor. Na Figura 1, apresenta uma parte da árvore de decisão aplicada à análise da qualidade do vinho. Nessa seção específica, podemos observar que a classe predominante é "boa" (1), pois é a mais frequente. As saídas do nó principal são rotuladas como *true* e *false*, indicando que, se a amostra seguir pelo caminho *true*, o resultado será verdadeiro; caso contrário, será falso.

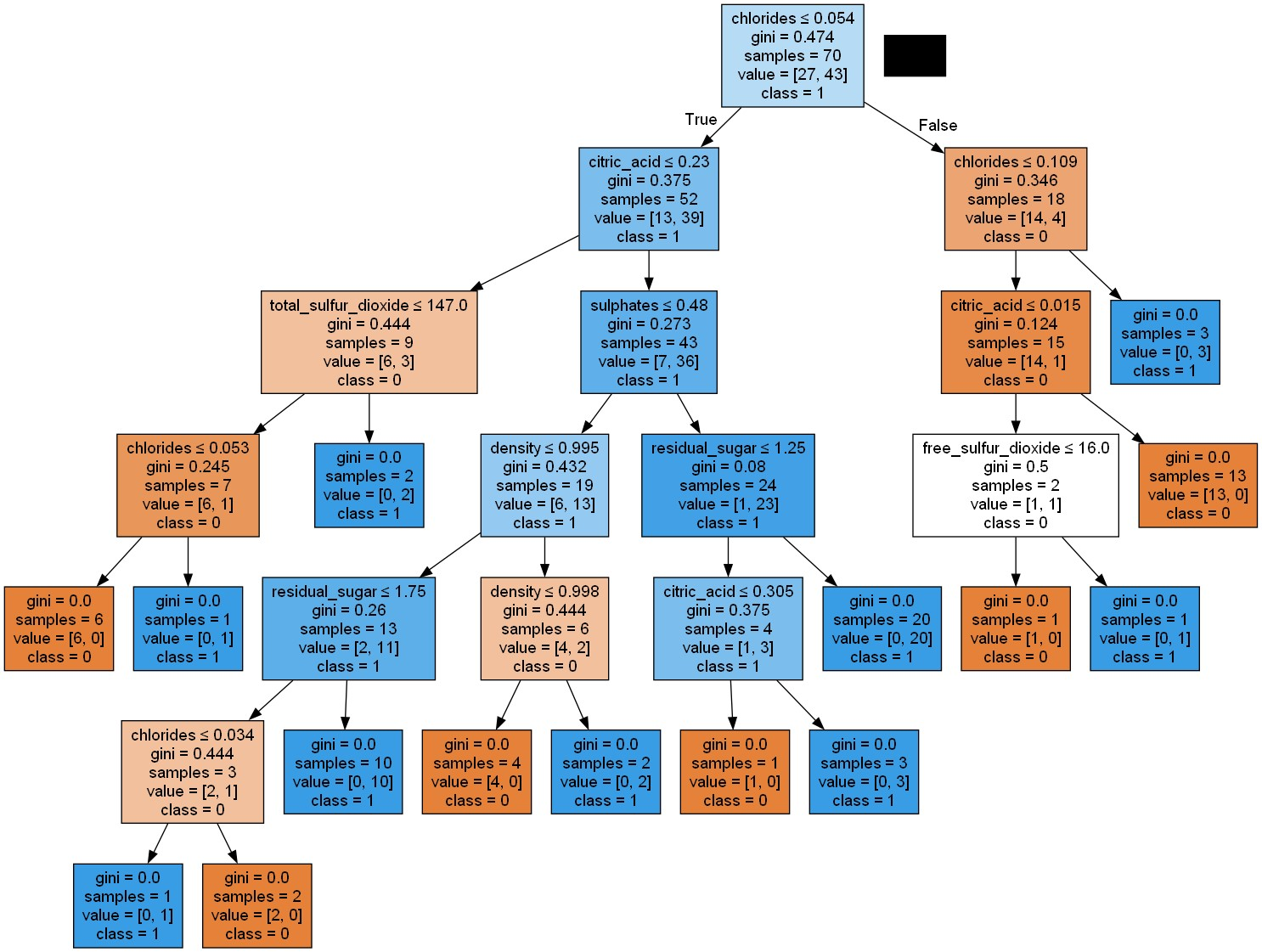


Figura 1: Árvore de Decisão do modelo de treinamento.

3.1.1 *Random Forest*

O algoritmo *Random Forest* é uma extensão das árvores de decisão que introduz um elemento de aleatoriedade na escolha das variáveis principais. Ao invés de selecionar apenas uma variável, o *Random Forest* escolhe aleatoriamente duas ou mais variáveis e realiza cálculos com base nessas amostras selecionadas para determinar qual delas será utilizada no primeiro nó. Esse processo é repetido para os demais nós da árvore, com a condição de que variáveis previamente selecionadas não podem ser escolhidas novamente, como ilustrado na Figura 2.

O *Random Forest* se destaca quando lidamos com grandes conjuntos de dados e buscamos construir um modelo robusto e preciso. Na aplicação desse modelo para prever a qualidade do vinho, foi utilizado um valor de *n\_estimators* igual a 100, resultando em uma acurácia de 0,8143. Isso significa que o modelo alcançou uma boa precisão na classificação da qualidade do vinho.

3.1.2 *Extra Trees*

O *Extra Trees* também opera com árvores de decisão, mas apresenta uma distinção em relação ao *Random Forest* que, em muitos casos, pode torná-lo uma escolha superior ou igualmente eficaz. A eficácia depende do cenário, já que, quando se trata de conjuntos de dados pequenos e bem organizados, o *Extra Trees* pode ser uma opção mais ágil e eficiente.

O processo do *Extra Trees* é, em grande parte, similar ao do *Random Forest*, com a diferença significativa de introduzir mais aleatoriedade. Após a escolha aleatória das variáveis candidatas para o primeiro nó, os dados presentes em cada uma dessas variáveis também são separados de maneira aleatória. Diferentemente do *Random Forest*, o *Extra Trees* não busca escolher a melhor separação possível, como ilustrado na Figura 2.

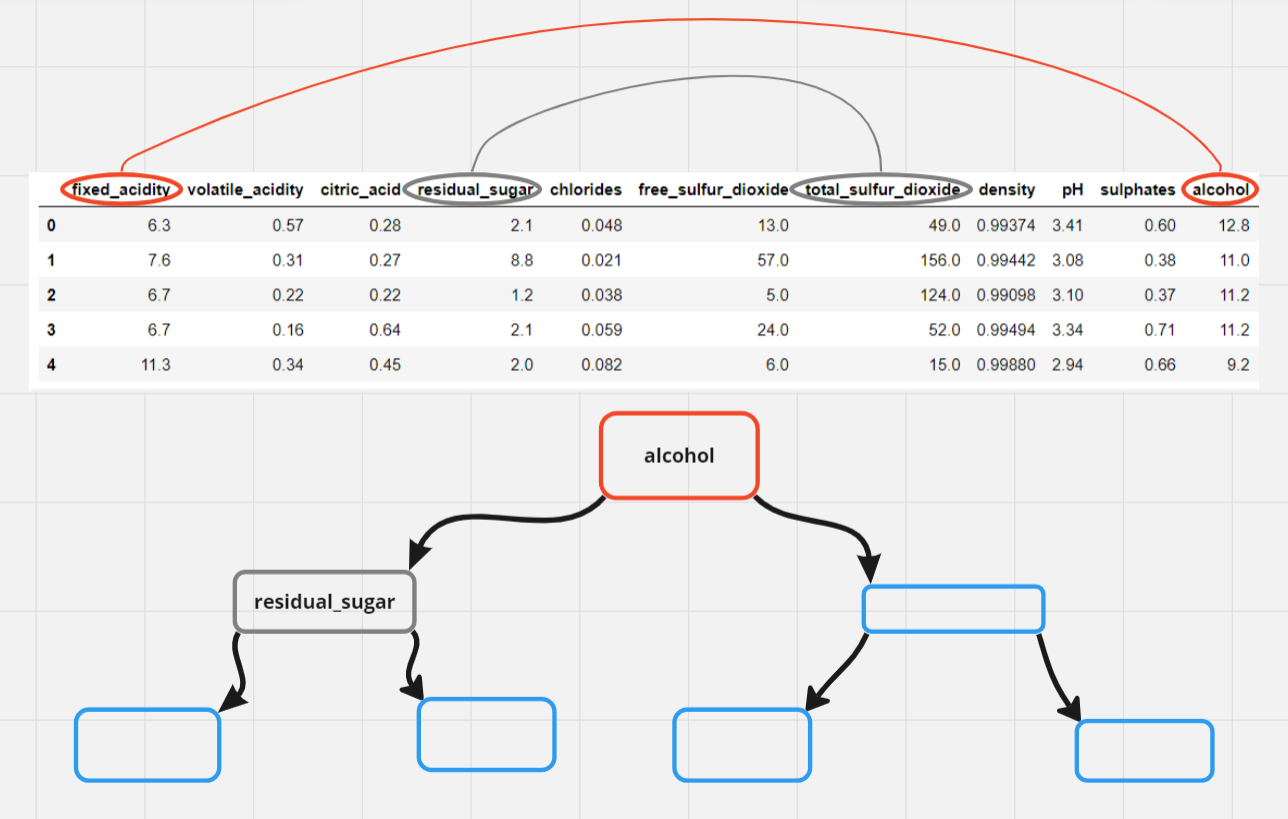


Figura 2: Ilustração do funcionamento de uma árvore de decisão aleatória;

Na aplicação do modelo *Extra Trees* para prever a qualidade do vinho, utilizou-se um valor de *n\_estimators* igual a 100, resultando em uma acurácia de 0,8261. Isso indica que o modelo alcançou um alto grau de precisão na classificação da qualidade do vinho.

3.1.3 *Extreme Gradient Boosting*

Este algoritmo também pertence à categoria baseada em árvores de decisão e é notável por sua capacidade de manipular eficazmente uma ampla gama de tipos de dados, demonstrando robustez em diversas situações.

O termo *Gradient Boosting* (aumento de gradiente) se refere ao uso do algoritmo *Gradient Descent* (descida de gradiente) para minimizar a perda (*loss*) à medida que novos modelos são adicionados. Essa abordagem é notável por sua extrema flexibilidade e, assim como os dois algoritmos anteriores, oferece a possibilidade de ajustar hiperparâmetros de acordo com o problema em questão, o que pode resultar em desempenho otimizado.

Ao aplicar esse modelo à previsão da qualidade do vinho, obteve-se uma acurácia de 0,7994, indicando um bom desempenho na tarefa de classificação da qualidade do vinho.

3.2 *LOGISTIC REGRESSION*

Dando um passo adiante em relação aos modelos baseados em árvores de decisão, nesta seção exploraremos um modelo distinto.

A regressão logística difere dos modelos de árvore, pois é construída por meio da aplicação de uma transformação, chamada de função logística ou *sigmoid*, à regressão linear. A principal finalidade da regressão linear é encontrar a melhor linha de ajuste para um conjunto de dados, determinando a combinação dos coeficientes b0 (coeficiente linear) e b1 (coeficiente angular) (y = b0 + b1 \* x1) que minimiza os erros de predição. Quanto mais os pontos de dados se alinharem à linha, melhor será o resultado. A Figura 3 ilustra a aplicação desses conceitos usando duas variáveis da análise de qualidade do vinho. Observando a figura, fica evidente que, em termos de classificação de dados, a "regressão logística" pode não ser a opção mais adequada, já que muitos pontos de dados estão dispersos em relação às linhas de classificação.

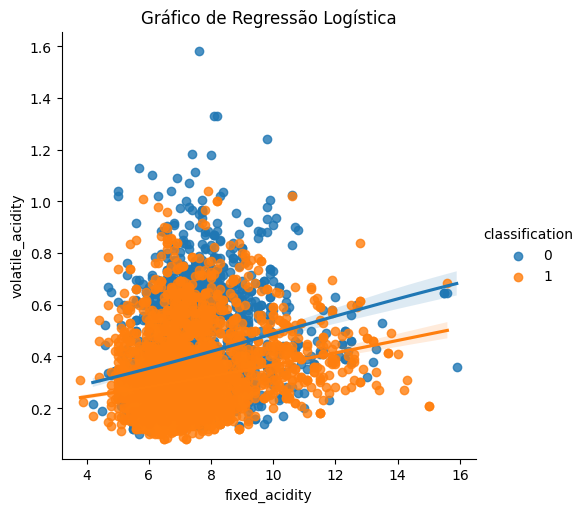


Figura 3: Exemplo de regressão logística aplicado no modelo de treinamento.

Diante dessa constatação, ao aplicar o treinamento com o modelo, alcançou-se uma acurácia de 0,7066. Mesmo que essa taxa esteja na faixa dos 70%, não é um resultado ideal, o que reforça a conclusão de que, para bases de dados que requerem classificação, a regressão logística pode não ser a escolha mais apropriada.

3.3 *“K NEIGHBORS CLASSIFIER”*

O *K-nearest neighbors* (KNN), também conhecido como K-vizinhos mais próximos, é um algoritmo que busca classificar cada amostra de um conjunto de dados ao avaliar sua proximidade em relação aos vizinhos mais próximos. Se os vizinhos mais próximos pertencerem à mesma classe, a amostra em questão será classificada nessa categoria.

Para compreender o funcionamento, imagine que está recebendo uma nova amostra de vinho. Na Figura 4, as amostras de vinho estão representadas, e supondo que a nova amostra esteja localizada próxima à posição 4 na dimensão *Fixed Acidity* e próxima à posição 8 na dimensão *Alcohol*. Ao analisar essa posição, percebe-se que os vizinhos mais próximos são vinhos de qualidade inferior, o que levaria à classificação dessa nova amostra como pertencente à categoria de qualidade inferior. As áreas destacadas em azul claro no mapa indicam onde uma amostra pode ser considerada de boa qualidade, enquanto as áreas em vermelho claro sugerem que a amostra pode ser de qualidade inferior.

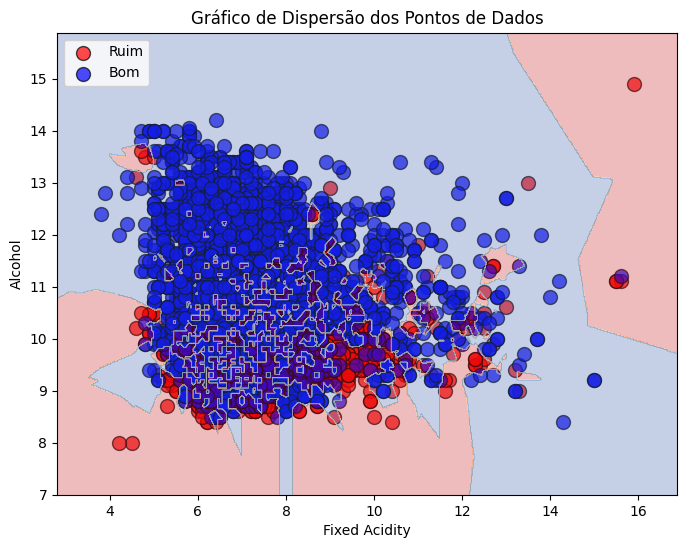


Figura 4: Ilustração para detecção de vizinhos. Modelo KNN.

No entanto, ao aplicar esse modelo à base de dados do projeto em questão, o resultado da acurácia foi de 0,6641, o que ficou consideravelmente abaixo das expectativas. Esse desempenho mais baixo indica que o KNN pode não ser a escolha ideal para o treinamento deste projeto.

3.4 VARIÁVEIS DE MAIOR IMPORTÂNCIA

A fim de identificar as principais influências em cada amostra de vinho, aplicou-se um algoritmo que utilizou o modelo treinado, no caso, o *Extra Trees*, devido à sua melhor acurácia. Esse algoritmo permitiu avaliar a relevância de cada variável no modelo.

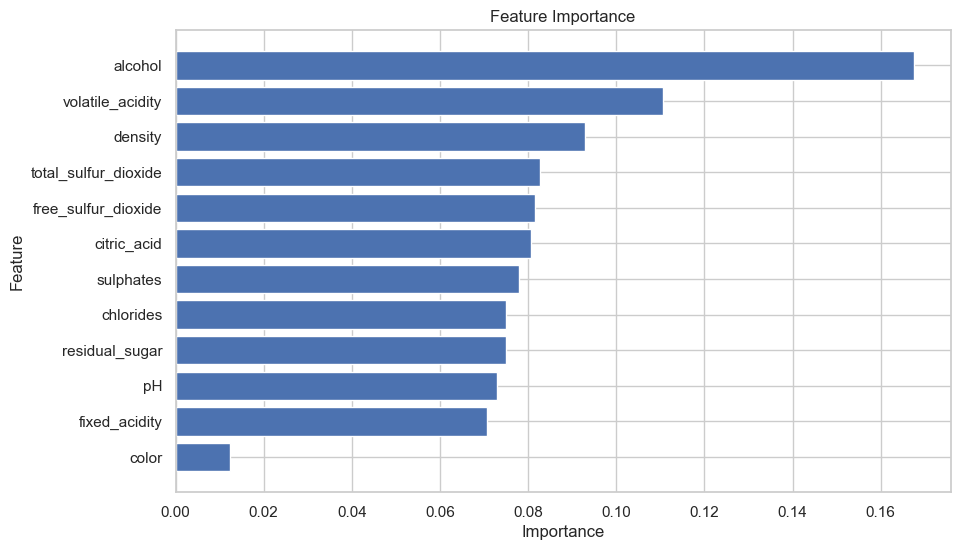


Figura 5: Importância de cada variável em relação a cada amostra.

Outro gráfico que mereceu atenção é um que é gerado por um cálculo do algoritmo, que revela o impacto de cada variável em cada amostra de vinho. A análise da Figura 6 revela o seguinte: a variável mais crucial no modelo é o *alcohol*, demonstrando que um teor alcoólico mais elevado está associado a uma maior probabilidade de a amostra ser classificada como boa. A segunda variável mais influente é a *volatile\_acidity*, na qual um valor mais alto está associado a uma menor probabilidade de a amostra ser considerada boa. A terceira variável importante é a *density*, que possui uma relação semelhante à anterior, ou seja, quanto maior a densidade, menor a probabilidade da amostra ser de boa qualidade. As variáveis *sulphates*, *free\_súlfur\_dioxide*, *citric\_acid* e *residual\_sugar* indicam que maiores valores estão correlacionados com uma maior probabilidade da amostra ser considerada boa.

Essa análise fornece informações valiosas sobre as variáveis que mais influenciam a qualidade do vinho, auxiliando na compreensão dos fatores-chave que determinam se uma amostra de vinho será classificada como boa ou ruim.

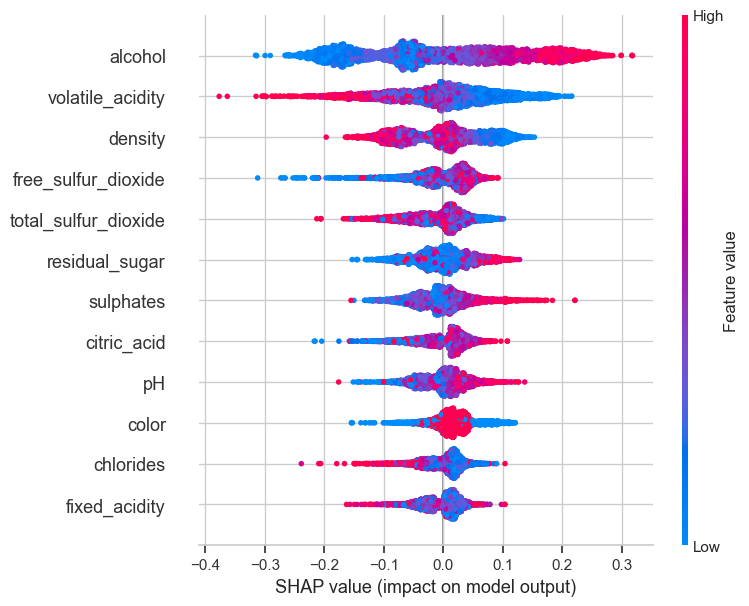


Figura 6: Impacto de cada variável diminuindo ou aumentando dentro de cada amostra.

3.5 COMPARAÇÃO DOS RESULTADOS

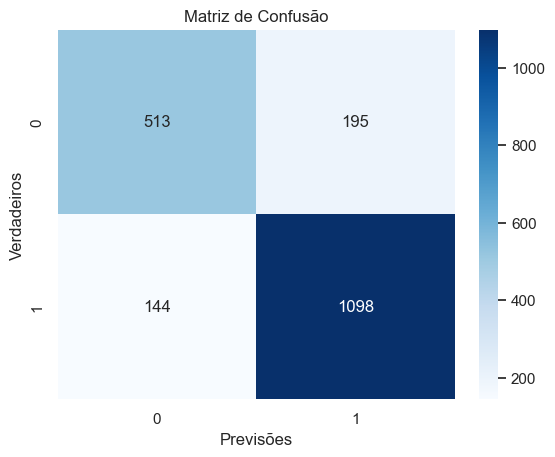
Após uma análise detalhada dos resultados, chegou-se à conclusão de que o modelo "Extra Trees" obteve a melhor acurácia, superando todos os outros modelos. No entanto, é importante destacar que os modelos de árvore de decisão também apresentaram acurácias bastante próximas. No contexto da classificação de vinhos em categorias de "bom" ou "ruim," os modelos de árvore de decisão, e em particular o "Extra Trees," se destacaram como as escolhas mais eficazes para o treinamento.

A Tabela 1 resume o desempenho dos modelos testados:

Tabela 1: Resultado da acurácia de cada modelo treinado

|  |  |
| --- | --- |
| **Modelos** | **Acurácia** |
| Extra Trees | 0.8262 |
| Random Forest | 0.8144 |
| Extreme Gradient Boosting | 0.7995 |
| Logistic Regression | 0.7067 |
| K Neighbors Classifier | 0.6641 |

Para uma avaliação mais abrangente da qualidade do modelo, foi gerada uma matriz de confusão, que fornece informações sobre Verdadeiros Positivos, Falsos Positivos, Falsos Negativos e Verdadeiros Negativos. Nesse contexto de análise da qualidade do vinho, a matriz de confusão retorna para Verdadeiros Vinhos Ruins, Falsos Vinhos Ruins, Falsos Vinhos Bons e Verdadeiros Vinhos Bons.

  
Figura 7: Matriz de Confusão

Analisando a Figura 7, observamos que o modelo identificou corretamente 513 vinhos de má qualidade, mas erroneamente classificou 195 deles como bons. Em relação à categoria "bom", houve 144 classificações errôneas e 1098 classificações corretas.

Além disso, foi gerada a curva ROC (*Receiver Operating Characteristic*) para avaliar o desempenho do modelo. A interpretação da curva ROC envolve a análise da sensibilidade (capacidade de identificar corretamente resultados positivos) e da especificidade (capacidade de identificar corretamente resultados negativos). Quanto maiores os valores de sensibilidade e especificidade (mais próximos de 100%), melhor o desempenho do teste.

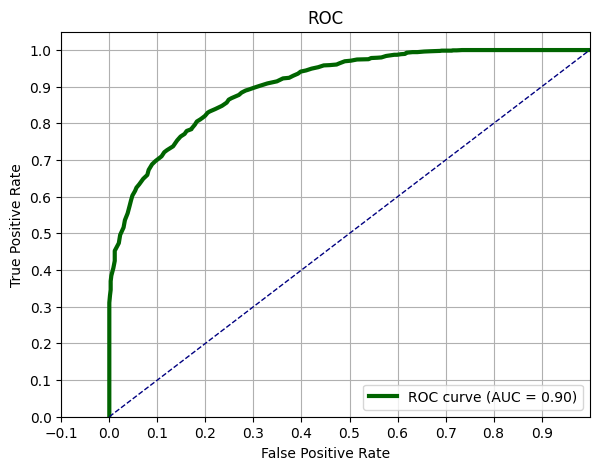
****

Figura 8: Curva ROC

A Figura 8 ilustra a curva ROC do modelo. A área sob a curva ROC é igual a 0.90, o que indica que o modelo apresenta um desempenho muito bom, uma vez que a sensibilidade e a especificidade estão próximas de 100%.

**CONSIDERAÇÕES FINAIS**

O objetivo de desenvolver um modelo de aprendizado de máquina capaz de prever vinhos de alta qualidade com base na análise dos dados dos componentes físico-químicos é fundamental para aprimorar a seleção de vinhos. Isso proporciona aos sommeliers e apreciadores a oportunidade de concentrar seus esforços na avaliação daqueles que atendem aos critérios desejados.

O modelo alcançou um marco significativo em relação a esse objetivo, pois conseguiu prever, na maioria dos casos, se um vinho é de boa ou má qualidade.

No entanto, é importante destacar que qualquer empreendimento, incluindo este, tem potencial para melhorias contínuas. O segredo reside na dedicação ao estudo, pesquisa e busca constante de aprimoramentos. Ao consolidar esses fatores, é possível descobrir novas técnicas de treinamento e, com o aperfeiçoamento dessas técnicas, o modelo pode alcançar resultados ainda mais robustos e precisos no futuro. O aprendizado de máquina é uma jornada em constante evolução, e o potencial de crescimento é ilimitado, desde que o compromisso com a excelência seja mantido.

**REFERÊNCIAS**

BHARDWAJ, P. et al. A machine learning application in wine quality prediction. Machine learning with applications, v. 8, n. 100261, p. 100261, 2022.

CASTILHOS, M. B. M. DE; DEL BIANCHI, V. L. CARACTERIZAÇÃO FÍSICO-QUÍMICA E SENSORIAL DE VINHOS “WHITE”S DA REGIÃO NOROESTE DE SÃO PAULO. **HOLOS - ISSN 1807-1600**, v. 4, p. 148–158, 2011.

CORTEZ, P. et al. Modeling wine preferences by data mining from physicochemical properties. **Decision support systems**, v. 47, n. 4, p. 547–553, 2009.

DI, S.; YANG, Y. Prediction of Red Wine quality using one-dimensional convolutional neural networks. 2022. Disponível em: <http://arxiv.org/abs/2208.14008>.

MEDEIROS, E. C. et al. Análise comparativa de algoritmos de aprendizagem profunda na classificação da qualidade do vinho. CONTRIBUCIONES A LAS CIENCIAS SOCIALES, v. 16, n. 8, p. 9383–9401, 2023.

MIERCZYNSKA-VASILEV, A.; SMITH, P. A. Current state of knowledge and challenges in wine clarification: Knowledge and challenges in wine clarification. Australian journal of grape and wine research, v. 21, p. 615–626, 2015.

RIZZON, L. A.; SALVADOR, M. B. G. **Metodologia para Analise de Vinho**. 1a ed. Brasília: Embrapa, 2010.

RUSSELL, Stuart J.; NORVIG, Pedro. Inteligência Artificial: Uma Abordagem Moderna. Grupo GEN, 2022. E-book. ISBN 9788595159495. Disponível em: https://integrada.minhabiblioteca.com.br/#/books/9788595159495/. Acesso em: 22 set. 2023.

SKIBSTED, L.; RISBO, J.; ANDERSEN, M. (EDS.). Chemical deterioration and physical instability of food and beverages. Boca Raton, FL, USA: CRC Press, 2010.

**UCI machine learning repository**. Disponível em: <https://archive.ics.uci.edu/dataset/186/wine+quality>. Acesso em: 22 set. 2023.

VERDES, C.-C. D. V. D. R. D. V. (2023). Vinho Verde, like no other wine in the world. Sprache Wohlen. https://www.vinhoverde.pt/en/history-of-vinho-verde