



Previsão da Qualidade de Vinhos com Machine Learning

Desvendando os Segredos da Composição Química do Vinho Tinto



por **Rafael Feltrim**

Menção Honrosa: Gustavo Gomes Contiero

Paixão por Vinhos, Visão em Dados

"Minha motivação foi unir o amor por vinhos com o interesse em tecnologia, criando uma IA para avaliar produtos sensoriais."



Amor por Vinhos

Combinando uma paixão pessoal com a análise de dados.



Tecnologia e IA

Explorando o potencial da inteligência artificial.



Previsão Precisa

Desenvolver um modelo de regressão para notas de qualidade.

Objetivo: Desenvolver e validar um modelo de regressão para prever com precisão a nota de qualidade de vinhos tintos, a partir de dados físico-químicos brutos.

Nossos Dados: A Composição Química do Vinho

Baseado no dataset Wine Quality (Red) do UCI Machine Learning Repository (Cortez et al., 2009).

11 Atributos de Entrada Físico-Químicos: Acidez Fixa, Acidez Volátil, Ácido Cítrico, Açúcar Residual, Cloretos, SO2 Livre, SO2 Total, Densidade, pH, Sulfatos, Álcool.

1 Atributo de Saída: Qualidade (escore de 0 a 10, avaliado por especialistas). Sem valores ausentes.

```
dados > winequality-red_teste.csv
You, ontem | 1 author (You)
"fixed acidity";"volatile acidity";"citric acid";"residual s
"sulphates";"alcohol";"quality"
1 9;0.785;0.24;1.7;0.078;10;21;0.99692;3.29;0.67;10;5
2 9;0.785;0.24;1.7;0.078;10;21;0.99692;3.29;0.67;10;5
3 8.5;0.44;0.5;1.9;0.369;15;38;0.99634;3.01;1.1;9.4;5
4 9.9;0.54;0.26;2;0.111;7;60;0.99709;2.94;0.98;10.2;5
5 8.2;0.33;0.39;2.5;0.074;29;48;0.99528;3.32;0.88;12.4;7
6 6.5;0.34;0.27;2.8;0.067;8;44;0.99384;3.21;0.56;12;6
7 7.6;0.5;0.29;2.3;0.086;5;14;0.99502;3.32;0.62;11.5;6
8 9.2;0.36;0.34;1.6;0.062;5;12;0.99667;3.2;0.67;10.5;6
9 7.1;0.59;0.2;2.2;0.078;26;44;0.99522;3.42;0.68;10.8;6
10 9.7;0.42;0.46;2.1;0.074;5;16;0.99649;3.27;0.74;12.3;6
11 7.6;0.36;0.31;1.7;0.079;26;65;0.99716;3.46;0.62;9.5;6
12 7.6;0.36;0.31;1.7;0.079;26;65;0.99716;3.46;0.62;9.5;6
13 6.5;0.61;0.2;2.2;0.095;48;59;0.99541;3.61;0.7;11.5;6
14 6.5;0.88;0.03;5.6;0.079;23;47;0.99572;3.58;0.5;11.2;4
15 7.1;0.66;0.2;2.4;0.052;6;11;0.99318;3.35;0.66;12.7;7
16 5.6;0.915;0.2;2.1;0.041;17;78;0.99346;3.68;0.73;11.4;5
17 8.2;0.35;0.33;2.4;0.076;11;47;0.99599;3.27;0.81;11;6
18 8.2;0.35;0.33;2.4;0.076;11;47;0.99599;3.27;0.81;11;6
19 9.8;0.39;0.43;1.65;0.068;5;11;0.99478;3.19;0.46;11.4;5
20 10.2;0.4;0.4;2.5;0.068;41;54;0.99754;3.38;0.86;10.5;6
21 6.8;0.66;0.07;1.6;0.07;16;61;0.99572;3.29;0.6;9.3;5
22 6.7;0.64;0.23;2.1;0.08;11;119;0.99538;3.36;0.7;10.9;5
23 7;0.43;0.3;2;0.085;6;39;0.99346;3.33;0.46;11.9;6
24 6.6;0.8;0.03;7.8;0.079;6;12;0.9963;3.52;0.5;12.2;5
25 7;0.43;0.3;2;0.085;6;39;0.99346;3.33;0.46;11.9;6
26 6.7;0.64;0.23;2.1;0.08;11;119;0.99538;3.36;0.7;10.9;5
27 8.8;0.955;0.05;1.8;0.075;5;19;0.99616;3.3;0.44;9.6;4
28 9.1;0.4;0.57;4.6;0.08;6;20;0.99652;3.28;0.57;12.5;6
29 6.5;0.885;0.2;3;0.166;6;12;0.99551;3.56;0.51;10.8;5
30 7.2;0.25;0.37;2.5;0.063;11;41;0.99439;3.52;0.8;12.4;7
31 6.4;0.885;0.2;3;0.166;6;12;0.99551;3.56;0.51;10.8;5
32
```

A matriz de correlação abaixo mostra a relação entre as features e a qualidade.

O Caminho dos Dados ao Modelo

1

1. Leitura e EDA

Análise inicial e visualização dos dados.

2

2. Engenharia de Atributos

Criação de features não lineares para complexidade.

3

3. Normalização

Padronização dos dados usando Z-score.

4

4. Treinamento do Modelo

Regressão Linear otimizada via BFGS.

5

5. Validação

Avaliação robusta via K-Fold Cross-Validation (5 folds).

6

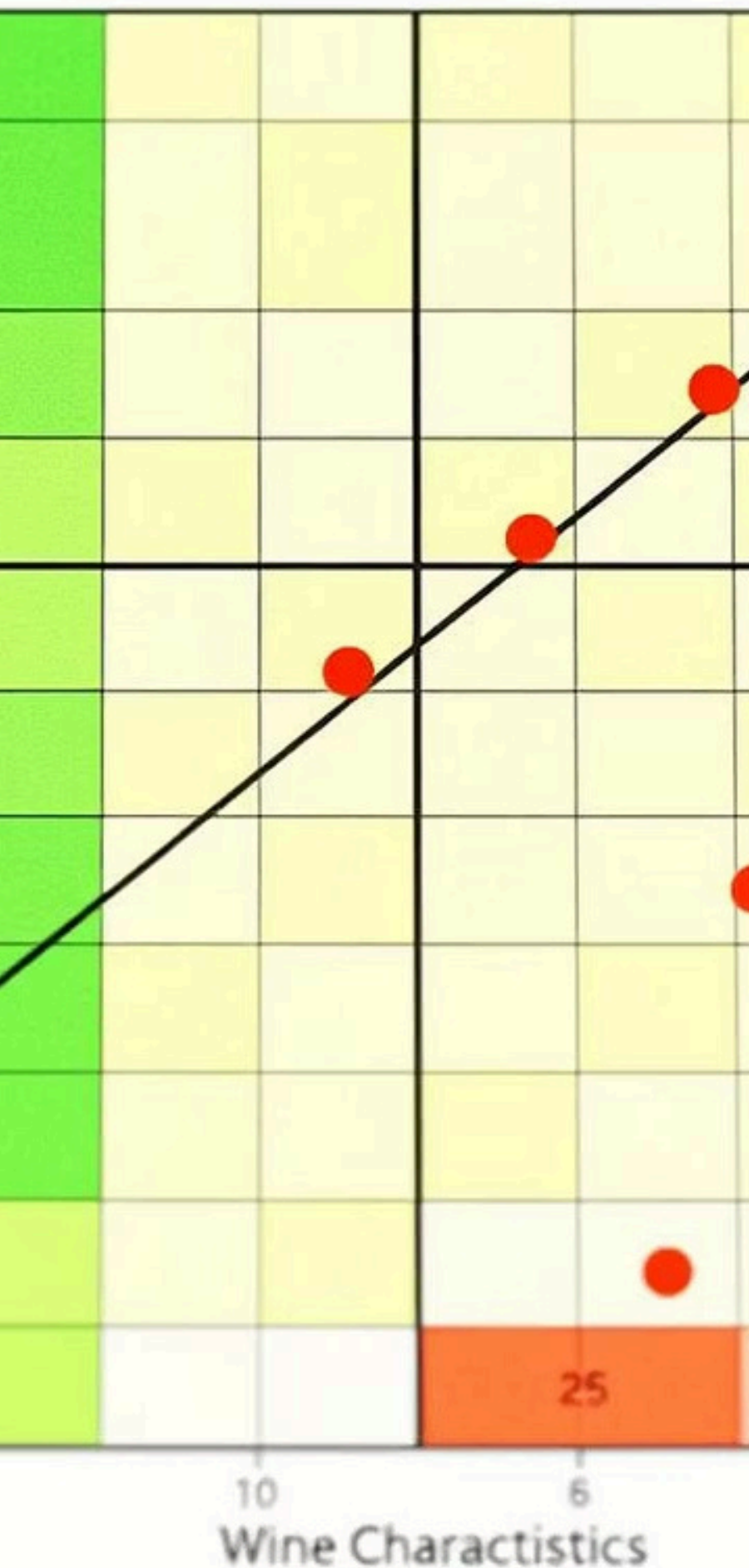
6. Avaliação Externa

Teste em conjunto de dados totalmente separado.

7

7. Armazenamento

Salvamento do modelo e das previsões.



Além do Linear: Criando Novas Perspectivas

Transformamos 11 atributos originais em 99 features, permitindo ao modelo capturar relações não lineares complexas:



Termos Quadráticos (x^2)

Captura relações de curvatura.



Termos Cúbicos (x^3)

Modelagem de tendências mais complexas.



Transformação Logarítmica ($\log(x)$)

Redução de assimetria e escala.

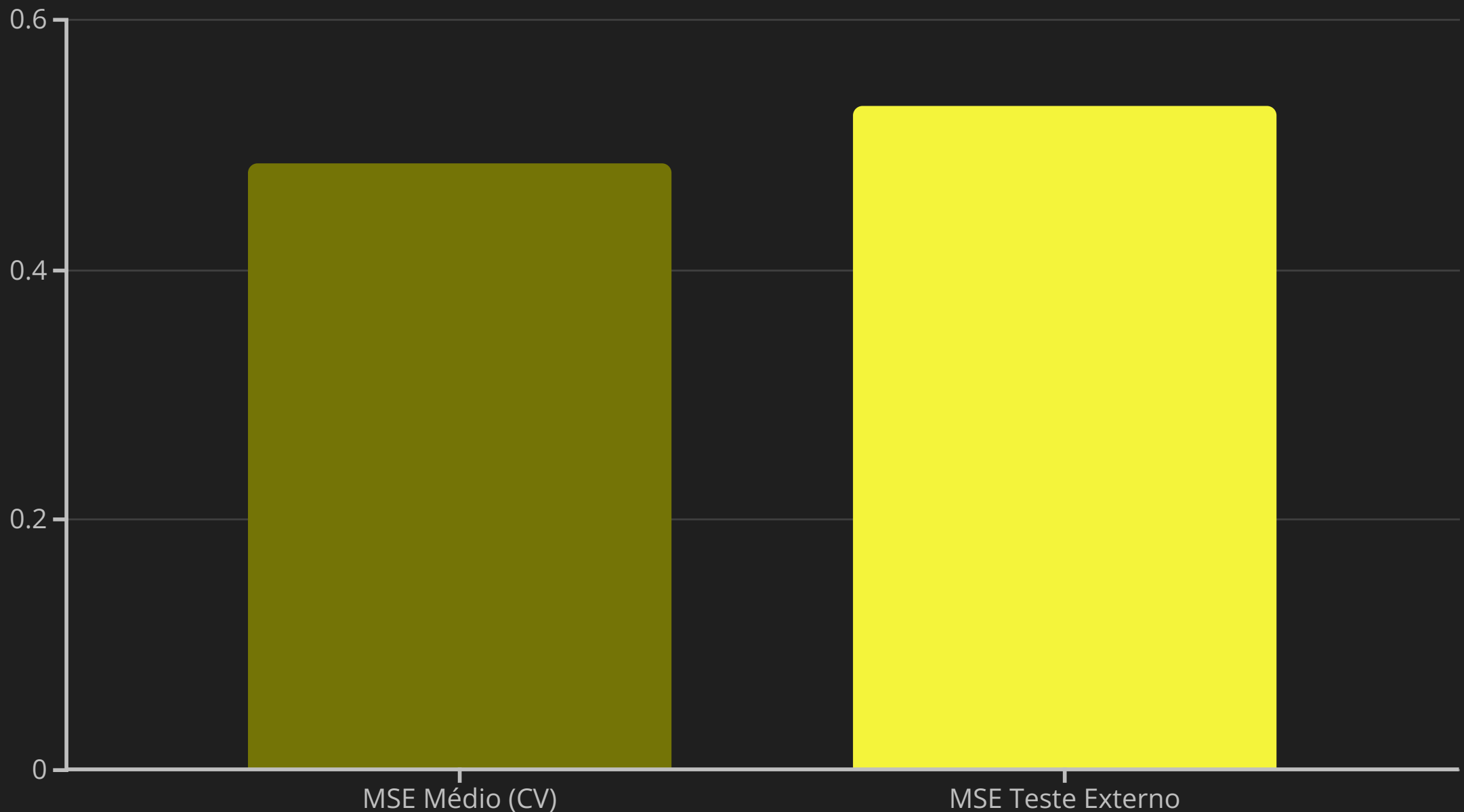


Termos de Interação ($x_i * x_j$)

Descobre como features se influenciam mutuamente.

A Qualidade Preditiva: Nossos Indicadores

Utilizamos o Erro Quadrático Médio (MSE) como métrica principal.



- **MSE Médio (5-Fold Cross-Validation): ~0.4844** Indica a capacidade do modelo de generalizar para dados não vistos durante o treinamento interno.
- **MSE no Conjunto de Teste Externo: ~0.5313** Confirma a boa capacidade de generalização para dados totalmente novos, mostrando que o modelo não está sofrendo de overfitting.

O Futuro: Evoluindo o Modelo

Explorar Outros Modelos

Random Forest, XGBoost, SVR – para capturar mais complexidade e melhorar a precisão.

Adicionar Regularização

Ridge, Lasso – para evitar overfitting, especialmente em modelos mais complexos.

Explicabilidade (XAI)

SHAP ou Permutation Importance – para entender a influência de cada atributo.

Otimização de Hiperparâmetros

Refinar os parâmetros dos modelos para melhor desempenho.

Interface Web

Criar uma aplicação interativa para que usuários possam testar o modelo.

Obrigado!

Rafael Feltrim - Estagiário RPA

[GitHub: @RaFeltrim](#)

[LinkedIn: Rafael Feltrim](#)

[Email: Rafael Feltrim](#)



"A melhor forma de aprender é ensinar. Compartilhe este projeto se ele te ajudou!"

best wayy
is to teach.
bext hwe...
Share his pljec
This helpect
if helpeded!
-you!