

ESTIMATIVA DA PRODUTIVIDADE DA VIDEIRA UTILIZANDO MACHINE LEARNING

THAÍS RAYANE GOMES DA SILVA¹, GLAUCO DE SOUZA ROLIM²,
TATIANA FERNANDA CANATA³

¹ Doutoranda em Agronomia (Produção Vegetal), FCAV/UNESP, Jaboticabal-SP. trg.silva@unesp.br

² Docente, Depto. de Ciências Exatas, FCAV/UNESP, Jaboticabal-SP

³ Docente, Depto. de Engenharia, FCAV/UNESP, Jaboticabal-SP

Apresentado no
LIII Congresso Brasileiro de Engenharia Agrícola - CONBEA 2024
6 a 8 de agosto de 2024 – Natal – RN, Brasil

RESUMO: A aplicação de algoritmos na estimativa da produtividade da uva não apenas impulsiona a eficácia das decisões agrícolas, mas também promove a sustentabilidade e a viabilidade econômica dos vinhedos em regiões vitivinícolas do Brasil. A pesquisa abrangeu sete regiões distintas, os dados meteorológicos das regiões foram obtidos a partir do banco de dados da NASA-Power. O banco de dados foi submetido a treinamento por meio do algoritmo Random Forest. A avaliação da performance da estimativa do modelo foi realizada através do coeficiente de determinação ajustado e da raiz do erro quadrático médio. Os dados das diversas regiões foram consolidados para formar um conjunto de dados único, permitindo a extração de respostas gerais por meio da aplicação do modelo. O modelo que incorporou a seleção de atributos mostrou métricas inferiores em comparação com aquele que incorporou todas as atributos disponíveis. Portanto, é importante compreender a relação entre índices climáticos, produtividade e fatores geográficos, evidenciando a complexidade das interações climáticas com a fenologia da uva, e a necessidade de considerar diferentes regiões para entender os impactos das variabilidades climáticas na produção vitivinícola brasileira.

PALAVRAS-CHAVE: variáveis climáticas, regiões vitivinícolas, python

ESTIMATE OF VINE PRODUCTIVITY USING MACHINE LEARNING

ABSTRACT: The application of algorithms to estimate grape yield not only boosts the effectiveness of agricultural decisions, but also promotes the sustainability and economic forecast of vineyards in Brazil's wine-growing regions. The research covers seven distinct regions, and meteorological data for the regions were obtained from NASA-Power database. The database was subjected to training using Random Forest algorithm. The evaluation of the performance of the model estimation was carried out using the adjusted coefficient of determination and the root mean square error. Data from several regions were consolidated to a single data set, allowing the limitation of general responses through application of the model. The model that incorporated a selection of evaluation attributes was inferior compared to the one that incorporated all available attributes. Therefore, it is important to understand the relationship between climatic indices, yield and geographical factors, highlighting the complexity of climatic interactions with grape phenology, and the need to consider different regions to understand the impacts of climate variability on Brazilian wine production.

KEYWORDS: climatic variables, wine regions, python

INTRODUÇÃO: A videira (*Vitis vinifera*) destaca-se como frutífera de proeminência econômica em escala global, ocupando a quarta posição entre as principais frutíferas em produção mundial (AGUIAR et al., 2023; ZANG et al., 2024). Em 2021, foi cultivada em 7 milhões de hectares, resultando em produção total de 78 megatoneladas (ANDRADE et al., 2023), mais de 57% dessa produção foi direcionada para a indústria vinícola. Em todo o mundo, a viticultura e a vinificação representam atividades de extrema relevância, exercendo impactos importantes nas economias locais e regionais. No Brasil, mais de 75 mil ha de vinhedos foram colhidos em 2022, conforme o IBGE (2022). A importância socioeconômica da viticultura no país é destacada por um mercado nacional de vinhos avaliado em R\$ 20,3 bilhões (MELLO e MACHADO, 2020). A qualidade potencial das uvas, aliada aos fatores de produtividade das videiras, desempenham um papel importante no impulsionamento da receita proveniente da produção vinícola (OHANA-LEVI et al., 2020). Um sistema complexo e interativo, composto por práticas climáticas e de manejo, exerce influência significativa sobre o desenvolvimento da videira (COSTA et al., 2015). Este sistema evoluiu para o conceito de *terroir*, definido como “uma área na qual o conhecimento coletivo das interações entre o ambiente físico e biológico identificável e as práticas vitivinícolas aplicadas se desenvolvem, fornecendo características distintivas para os produtos originários desta área” (OIV, 2010). A aplicação de modelos estimativos no cultivo de frutas emerge como uma estratégia promissora para aprimorar a eficácia dos processos de tomada de decisão (ANDRADE et al., 2023). Nesse contexto, buscou-se desenvolver modelos de aprendizado de máquina para a estimativa da produtividade em regiões vitivinícolas de relevância no Brasil.

MATERIAL E MÉTODOS: A pesquisa englobou as sete regiões vitivinícolas mais importantes no Brasil, cada uma composta por diferentes municípios (Figura 1). Essas regiões incluem a Campanha Gaúcha, com os municípios de Dom Pedrito (Alt 131 m, 95 ha), Rosário do Sul (132 m, 17 ha), Sant’Ana do Livramento (208 m, 48 ha) e Uruguaiana (66 m, 27 ha), com área geográfica de 44.365 km². Além disso, foram consideradas as regiões de Altos Montes, representada por Flores da Cunha (756 m, 4816 ha), a área geográfica delimitada totaliza 174 km²; Farroupilha, compreendendo Bento Gonçalves (690 m, 4474 ha), Caxias do Sul (817 m, 3813 ha) e Farroupilha (770 m, 3796 ha), com área geográfica de 379 km²; e o Vineyards Valley, incluindo Garibaldi (613 m, 2328 ha), com uma área de 72 km². Todas essas localidades são parte do estado do Rio Grande do Sul – RS. Outras regiões estudadas foram os Goethe Grape Valleys, abrangendo Morro da Fumaça (18 m, 54 ha), Orleans (132 m, 17 ha) e Urussanga (49 m, 91 ha) com área delimitada de 459 km², bem como os Highland Wines, que englobam Água Doce (847 m, 52 ha), Arroio Trinta (840 m, 10ha), Caçador (920 m, 387 ha), Curitibaanos (987 m, 7 ha), Fraiburgo (1048 m, 29 ha), Pinheiro Preto (696 m, 413 ha) e Rios das Antas (830 m, 78 ha), com área geográfica delimitada de 19.676 km², localizadas no estado de Santa Catarina – SC. Adicionalmente, a região de Bituruna Wines, situada em Bituruna (900 m, 30 ha), a área delimitada geográfica abrange 1.288 km², pertencente ao estado do Paraná – PR, todas as regiões compõem o Sul do Brasil. As variáveis de entrada consideradas compreenderam a irradiância solar global de ondas curtas (Q_g , em MJ m⁻² dia⁻¹), irradiância solar fotossinteticamente ativa (PAR, em MJ m⁻² dia⁻¹), temperatura do ar a 2 m (T, em °C), temperatura máxima diária a 2 m (T_{max}, em °C), temperatura mínima diária a 2 m (T_{min}, em °C), umidade relativa a 2 m (RH, em %), precipitação (P, em mm dia⁻¹), umidade do solo na zona radicular (SU), ao longo de uma série temporal de 39 anos. As variáveis radiação líquida na superfície do cultivo (R_n, em MJ m⁻² dia⁻¹) e umidade absoluta (AU, em g vapor⁻¹ m⁻³). Os dados diários foram submetidos a procedimentos de homogeneização, filtragem, revisão e pré-processamento, incluindo a remoção de outliers, para todas as variáveis mencionadas. Os registros anuais de produtividade das videiras no período de 1982 a 2021, conforme divulgados pelo Sistema IBGE de Recuperação Automática (Sidra IBGE). As principais bibliotecas em Python utilizadas para o e Random Forest (RF) foi: Pandas 2.0.3, Seaborn 0.13.2 e Numpy 1.26.3.

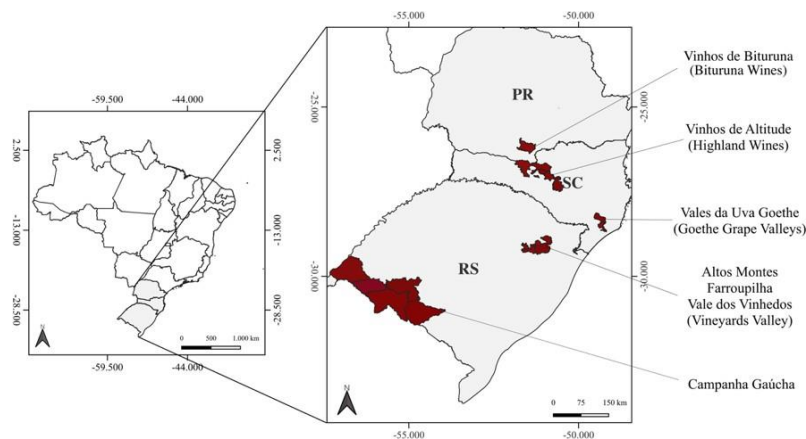


FIGURA 1. Mapa das principais regiões vitivinícolas do Brasil utilizadas no trabalho

RESULTADOS E DISCUSSÃO: Os dados provenientes de todas as regiões foram consolidados em um único conjunto de dados, com o propósito de extrair uma resposta geral por meio da aplicação de modelos. No contexto do modelo de RF, foram configurados os parâmetros de `max_depth`, variando de 10 a 50 em intervalos de 10 unidades. Dentre essas opções, o modelo foi treinado utilizando o valor de 30 para `max_depth`. De forma semelhante, utilizando o método `GridSearchCV` da biblioteca `scikit-learn`, foi feito para todos os parâmetros. No modelo de Random Forest (RF) procedeu-se à seleção de variáveis, dentro do conjunto de dados que continham todas as regiões. Foram selecionadas, com base em seus respectivos critérios de importância, as variáveis que melhor se ajustaram ao modelo. Na análise comparativa do modelo foi obtido resultados significativos nas métricas de ajuste e erro (Figura 2). O modelo de RF destacou-se alcançando R^2 adj de 0,89 e RMSE de 1.543,22 kg ha⁻¹ sem seleção de variáveis, enquanto com seleção de variáveis apresentou R^2 adj de 0,69 e RMSE de 2.644,85 kg ha⁻¹. Esses resultados sugerem uma capacidade superior do RF em ajustar-se aos dados, mesmo com a redução de variáveis.

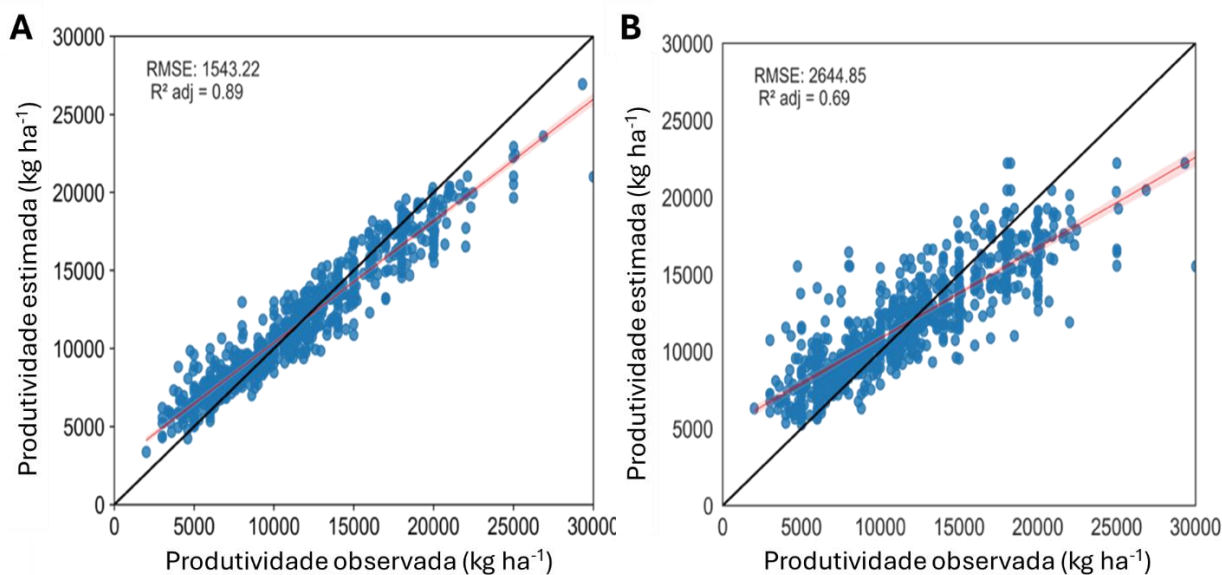


FIGURA 2. Desempenho de modelos para todas as regiões. A) Random Forest (RF) sem seleção de variáveis e B) com seleção de variáveis

CONCLUSÕES: Diversos atributos exerceram influência, embora amplamente variáveis, sobre os componentes de desempenho dos modelos preditivos. As condições climáticas influenciaram de forma significativa na fenologia da videira e, conseqüentemente na produtividade. Esta influência ocorreu de forma diferenciada em cada região produtora do país.

AGRADECIMENTOS: Ao Group of Agrometeorological Studies (GAS) da Faculdade de Ciências Agrícolas e Veterinárias (FCAV) da UNESP por sua colaboração neste estudo.

REFERÊNCIAS

- AGUIAR, A.C.; HIGUCHI, M.T.; YAMASHITA, F.; ROBERTO, S.R. SO₂-generating pads and packaging materials for postharvest conservation of table grapes: A review. *Horticulturae*, v. 9, p. 724–738, 2023. <https://doi.org/10.3390/horticulturae9060724>
- ANDRADE, C.B.; MOURA-BUENO, J.M.; COMIN, J.J.; BRUNETTO, G. Grape yield prediction models: approaching different machine learning algorithms. *Horticulturae*, v. 9, p. 1294, 2023. <https://doi.org/10.3390/horticulturae9121294>
- COSTA, R.; FRAGA, H.; MALHEIRO, A.C.; SANTOS, J.A. Application of crop modelling to Portuguese viticulture: implementation and added-values for strategic planning. *Ciênc. Têc. Vitiv.*, v. 30, p. 29–42, 2015. <http://dx.doi.org/10.1051/ctv/20153001029>
- IBGE, 2022. *Systematic Survey of Agricultural Production* [Online]. Instituto Brasileiro de Geografia e Estatística. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9201-levantamento-sistematico-da-producao-agricola.html>. Acesso: 5 de janeiro de 2024.
- MELLO, L.M.R.; MACHADO, C.A.E.; Brazilian Winemaking: Panorama 2019. Embrapa Grape & Wine, 2020, Brazil.
- OHANA-LEVI, N.; MUNITZ, S.; BEN-GAL, A.; SCHWARTZ, A.; PEETERS, A.; NETZER, Y. Multiseasonal grapevine water consumption – Drivers and forecasting. *Agricultural and Forest Meteorology*, v. 280, p. 107796, 2020. <https://doi.org/10.1016/j.agrformet.2019.107796>
- ZANG, R.; BIAN, Z.; WU, P.; LIU, Y.; LI, B.; XIONG, J.; ZHANG, Y.; ZHU, B. Nondestructive prediction of fruit detachment force for investigating postharvest grape abscission. *Postharvest Biol. Technol.* v. 209, p. 112691, 2024. DOI: <https://doi.org/10.1016/j.postharvbio.2023.112691>