

Performance preditiva do algoritmo *Random Forest* na previsão da produção de cana-de-açúcar.

Bruno Rafael de Souza Viana, Alan Rodrigo Panosso, Hugo Guiné Pinto Ferreira, Lucas de Oliveiras Gonçalves, Gener Tadeu Pereira.

Palavras Chave: *inteligência artificial, data mining, agricultura 4.0.*

Introdução

A produção agrícola é uma complexa interação entre as variáveis ambientais, os atributos do solo e a dinâmica de nutrientes no sistema solo-planta-atmosfera, assim novos estudos sobre o potencial produtivo da cultura em solos brasileiros são necessários. Alguns estudos são encontrados na literatura utilizando técnicas de mineração de dados, como algoritmos de *random forest* (floresta aleatória - RF) para estimar a produtividade da cana-de-açúcar mostrando o potencial dessas ferramentas^[1,2].

Objetivo

O objetivo do trabalho foi investigar a performance preditiva do algoritmo *Random Forest* na previsão espaço-temporal do rendimento da cana-de-açúcar.

Material e Métodos

Para o estudo foram utilizados dados de duas unidades produtoras de cana-de-açúcar, com um total de aproximadamente 55 mil ha. Foram levadas em consideração as variáveis de manejo que mais interligadas com a produtividade: **nº cortes**) é o número de vezes que o canavial foi mecanicamente colhido; **idade**) tempo que a cultura ficou no campo na colheita do canavial; **variedade**) foram consideradas as principais variedades plantadas nas áreas e **mês**) mês do ano em que fora realizada a operação de colheita. Foram tomadas uma observação a cada 3 hectares, para a determinação da produção da cultura (TCH) e amostragem de solo nas profundidades de 0,00 a 0,25 m e de 0,25 a 0,50 m e os atributos químicos. Para a validação do modelo RF, foi dividido os dados no treinamento e conjuntos de testes, onde ¾ de dados foi usado para treinamento e ¼ para teste, resultando em 1624 e 542 observações, respectivamente. Durante a etapa de treinamento, foi utilizada o processo de validação cruzada para o ajuste dos hiperparâmetros do modelo. Em sequência, as métricas: coeficiente de determinação (R^2) e raiz do erro médio quadrático (RMSE) foram estimadas para o conjunto de testes.

Resultados e Discussão

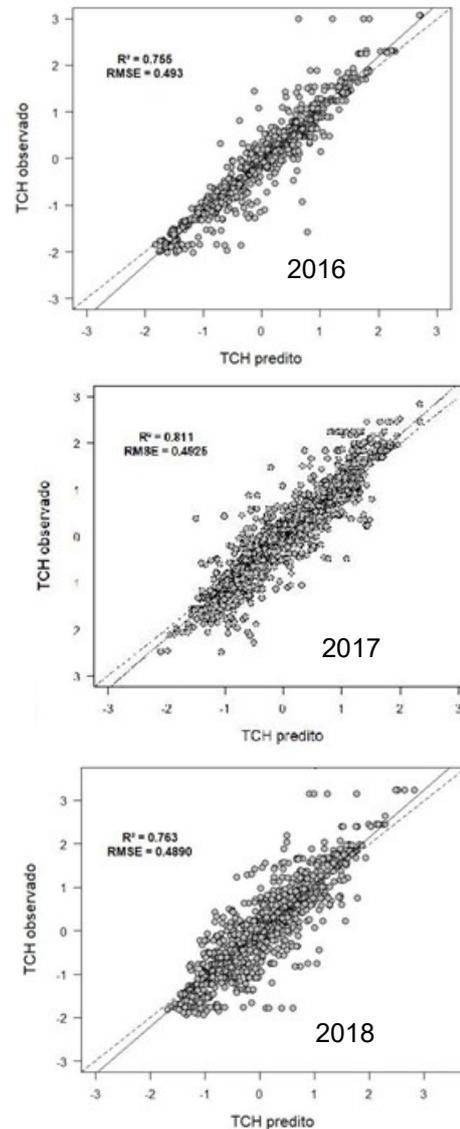


Figura 1. Performance preditiva do RF para os anos avaliados (2016 a 2018).

Conclusões

Os resultados indicam que os modelos preditivos a apresentaram precisão satisfatórias, evidenciando pelos valores de R^2 maiores que 76% e RMSE inferiores a 0,45, promissora para suporte à tomada de decisão dentro das usinas e unidade produtoras.

¹ Everingham, Y. L.; Smyth, C. W.; Inman-Bamber, N. G. Ensemble data mining approaches to forecast regional sugarcane crop production. *Agricultural and Forest Meteorology*, v. 149, n. 3-4, p. 689-696. doi: 10.1016/j.agrformet.2008.10.018, 2009.

² Wolfert S., Ge L., Verdouw C.; Bogaardt M.J. (2017) Big data in smart farming, a review, *Agricultural Systems*, 153, pp. 69-80.