

## Resumo

A produção de cana-de-açúcar no Brasil registrou um faturamento de US\$18,27 bilhões no ano comercial da safra 2023/2024, com destaque para o estado de São Paulo, líder na produção sucroalcooleira do país. Embora seja uma cultura amplamente difundida na economia, prever a produção da cana-de-açúcar é algo extremamente complexo, principalmente devido aos diversos fatores que influenciam esse resultado, como condições climáticas variáveis, práticas agronômicas específicas e características do solo. Diante desse cenário, este estudo propõe uma análise de algoritmos de machine learning com o objetivo de avaliar os seus desempenhos na previsão da produção de cana-de-açúcar em São Paulo. Para isso, foi criada uma base de dados robusta, composta por informações históricas sobre safras anteriores e variáveis influenciadoras, como clima e solo. Os modelos foram avaliados através de métricas de desempenho, como acurácia, precisão, recall, F1-score, erro médio absoluto e raiz quadrada do erro quadrático médio. Nossos resultados sugerem que a Floresta Aleatória foi o modelo mais eficaz para prever a produção de cana-de-açúcar, embora o Support Vector Machine também tenha obtido um bom desempenho. Este trabalho contribui para a aplicação de métodos de previsão no setor agrícola, oferecendo uma abordagem sistemática para aprimorar a previsão de produção de cana-de-açúcar. Assim, a pesquisa não apenas reforça a importância da análise de dados para uma agricultura mais precisa e eficiente, mas também serve como referência para futuras iniciativas que busquem melhorar a gestão e a sustentabilidade do setor sucroalcooleiro.

### Naive Bayes

Naive Bayes: Cross Validation 10 folds, useKernelEstimator: True

- Acurácia: 53.6199%
- Precisão: Ruim: 0,653 / Boa: 0,449 / Mediana: 0,455 / Ótima: 0,623 / Média ponderada: 0,546
- Recall: Ruim: 0,591 / Boa: 0,465 / Mediana: 0,527 / Ótima: 0,550 / Média ponderada: 0,536
- F1-Score: Ruim: 0,620 / Boa: 0,457 / Mediana: 0,488 / Ótima: 0,584 / Média ponderada: 0,539
- Erro Médio Absoluto (MAE): 0.2821 toneladas/ha
- Raiz Quadrada do Erro Quadrático Médio (RMSE): 0.3908 toneladas<sup>2</sup>/há

Support Vector Machine (SVM): Cross validation 10 folds, Kernel: Puk, BuildCalibrationModels: True

- Acurácia: 61.6354%
- Precisão: Ruim: 0,717 / Boa: 0,516 / Mediana: 0,0,536 / Ótima: 0,685 / Média ponderada: 0,615

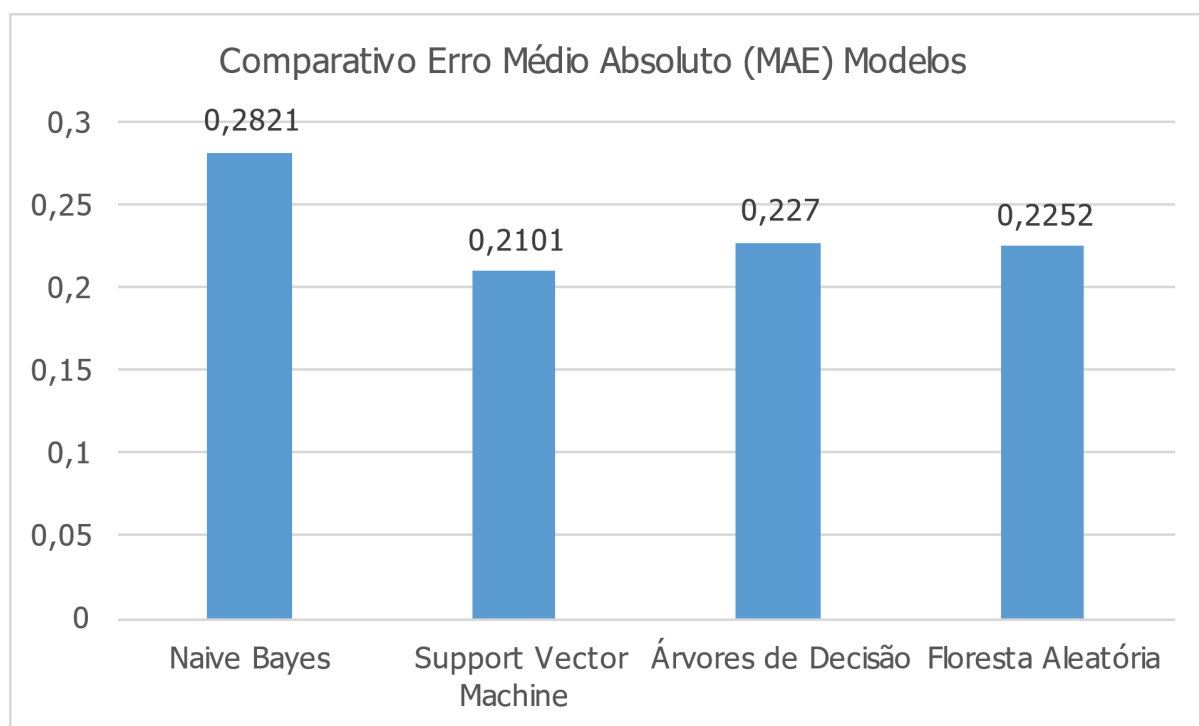
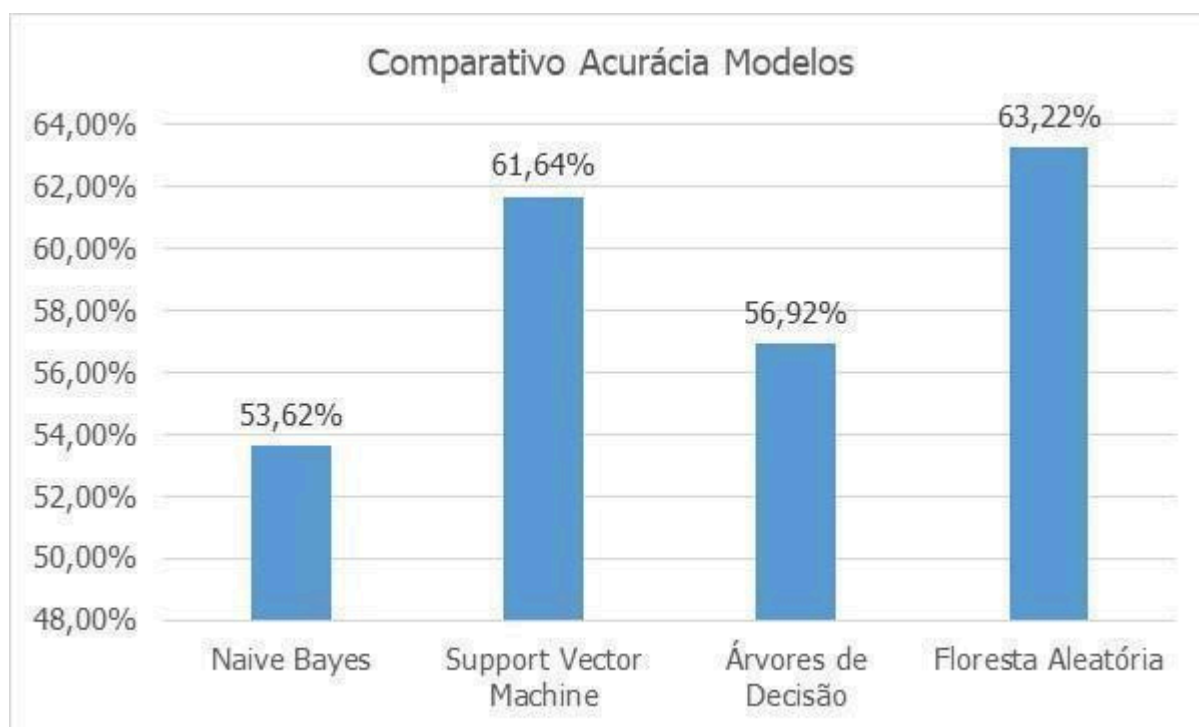
- *Recall: Ruim: 0,733 / Boa: 0,533 / Mediana: 0,507 / Ótima: 0,688 / Média ponderada: 0,616*
- *F1-Score: Ruim: 0,725 / Boa: 0,525 / Mediana: 0,521 / Ótima: 0,686 / Média ponderada: 0,616*
- *Erro Médio Absoluto (MAE): 0.2101 toneladas/ha*
- *Raiz Quadrada do Erro Quadrático Médio (MSE): 0.3835 toneladas<sup>2</sup>/há*

#### *Árvores de Decisão: Cross-validation 10 folds*

- *Acurácia: 56.9166%*
- *Precisão: Ruim:0,665 / Boa:0,459 / Mediana:0,446 / Ótima:0,682 / Média Ponderada:0,562*
- *Recall: Ruim:0,733 / Boa:0,477 / Mediana:0,379 / Ótima:0,691 / Média Ponderada:0,569*
- *F1-Score: Ruim:0,698 / Boa:0,468 / Mediana:0,410 / Ótima:0,686 / Média Ponderada:0,564*
- *Erro Médio Absoluto (MAE): 0.227 toneladas/ha*
- *Raiz Quadrada do Erro Quadrático Médio (MSE): 0.416 toneladas<sup>2</sup>/há*

#### *Floresta aleatória:Cross-Validation 10 folds, NumTree:1000, MaxDepth10*

- *Acurácia: 63.2191%*
- *Precisão: Ruim:0,737 / Boa:0,523 / Mediana:0,551 / Ótima:0,695 / Média Ponderada:0,628*
- *Recall: Ruim:0,753 / Boa:0,521 / Mediana:0,504 / Ótima:0,755 / Média Ponderada:0,632*
- *F1-Score: Ruim:0,745 / Boa:0,522 / Mediana:0,526 / Ótima:0,724 / Média Ponderada:0,630*
- *Erro Médio Absoluto (MAE): 0.2252 toneladas/ha*
- *Raiz Quadrada do Erro Quadrático Médio (MSE): 0.3595 toneladas<sup>2</sup>/há*



Comparativo Raiz Quadrada do Erro Quadrático Médio (RMSE) Modelos

