

# Aplicação de Inteligência Artificial na Previsão de Produção de Açúcar no Estado de São Paulo

Lucas A. B. Matos<sup>1</sup>

<sup>1</sup>Instituto de Química/UNESP, Araraquara, SP

Marilaine Colnago<sup>2</sup>

<sup>2</sup>IBILCE/UNESP, São José do Rio Preto, SP

A cana-de-açúcar desempenha um papel essencial na economia brasileira. O país é o maior produtor do mundo, sendo que aproximadamente 50% da área plantada está localizada no Estado de São Paulo. O estado é também o maior produtor de açúcar do Brasil, respondendo por cerca de 60% da produção nacional e cerca de 30% da produção mundial [2]. Diante de um cenário de tamanha importância a nível econômico e social, o estudo de modelos matemáticos e de métodos computacionais que buscam analisar a produção de açúcar tem se tornado essencial para respaldar decisões importantes.

Dessa forma, o presente trabalho busca empregar técnicas de Análise Exploratória de Dados (AED) e Aprendizado de Máquina (AM) para estimar a produção futura de açúcar advinda da cana-de-açúcar no estado de São Paulo, viabilizando ferramentas de suporte para tomada de decisões, garantindo a competitividade e a sustentabilidade desse setor.

Antes da execução e aplicação do modelo preditivo, procedeu-se à fase de pré-processamento dos dados, envolvendo a identificação e tratamento de valores nulos ou inconsistentes (*data cleaning*). Além disso, realizou-se uma análise exploratória dos dados para compreender o comportamento das variáveis em estudo. Essa etapa incluiu a identificação de padrões e a utilização de ferramentas gráficas e estatísticas para explorar relações entre as variáveis, proporcionando uma compreensão mais profunda do conjunto de dados [1].

Em relação ao modelo preditivo adotado, utilizou-se Florestas Aleatórias (RF, do inglês *Random Forest*). Aplicou-se a engenharia de recursos, que consiste na criação de novas variáveis a partir das já existentes: inicialmente a base de dados possuía 7 variáveis (*features*) apresentadas com frequência mensal e, após a inserção de variáveis artificiais criadas a partir das originais, a base analisada passou a ter 14 variáveis. Também foi aplicada a ferramenta *Random Search* para a seleção dos melhores hiperparâmetros deste modelo. Ambas técnicas possuem como intuito melhorar a assertividade desse algoritmos [3]. Na predição, foi adotada uma proporção de 80% dos dados para treinamento (referente ao período de 31/01/2012 à 31/12/2020) e os outros 20% para teste (de 31/01/2021 à 31/03/2023).

Os valores previstos pelo método podem ser observados a partir da Figura 1, que compara com os valores reais, utilizados para teste. Para melhor validação dos resultados, as métricas estatísticas MAE, MAPE e RMSE também foram calculadas e são apresentadas na Tabela 1 (para mais detalhes sobre as métricas, ver [3]). A partir do resultado visual (qualitativo) e dos valores obtidos pelas métricas (quantitativos), fica evidente que o método teve um bom comportamento, com MAPE próximo de 5%.

---

<sup>1</sup>lucas.b.matos@unesp.br

<sup>2</sup>marilaine.colnago@unesp.br

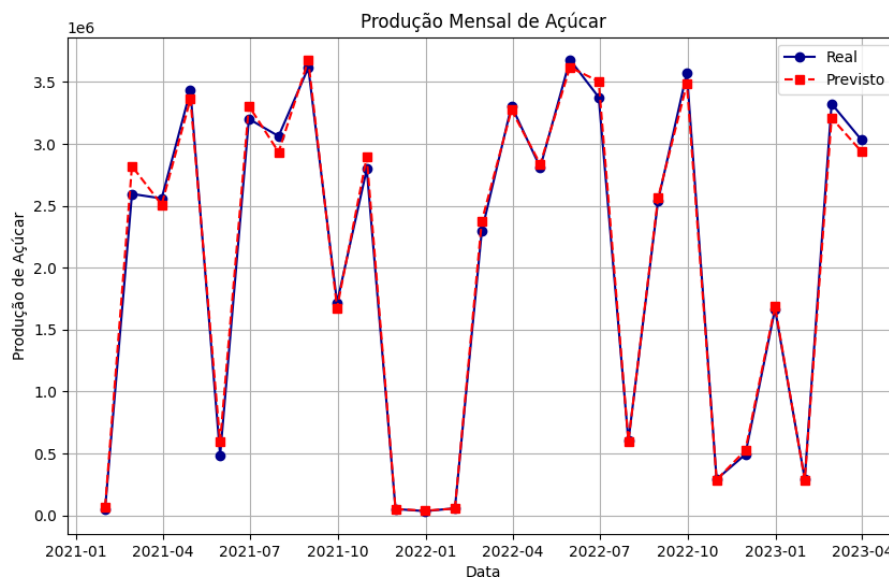


Figura 1: Gráfico de previsão de produção de açúcar realizada pelo modelo RF. Fonte: dos autores.

Tabela 1: Métricas de validação do modelo RF.

| MAE       | MAPE  | RMSE      |
|-----------|-------|-----------|
| 60.588,90 | 5,02% | 80.178,55 |

A pesquisa apresentada revelou que o modelo preditivo desenvolvido foi capaz de estimar de maneira altamente satisfatória a produção de açúcar no estado de São Paulo. Esse resultado destaca a eficácia dos modelos de Aprendizado de Máquina na previsão da produção açucareira, indicando seu potencial para auxiliar em decisões estratégicas de agências governamentais, atender às necessidades dos consumidores locais e oferecer *insights* valiosos para setores específicos, como a indústria da agricultura de energia ou o mercado de subprodutos da cana-de-açúcar.

## Agradecimentos

Agradecemos à FAPESP pelo fomento à pesquisa (N° Processo: 2023/05265-4).

## Referências

- [1] G. E. de A. P. Batista. “Pré-processamento de dados em aprendizado de máquina supervisionado”. Tese de doutorado. Universidade de São Paulo, 2003.
- [2] CONAB. **Companhia Nacional de Abastecimento. Boletim da safra da cana-de-açúcar**. Online. Acessado em 05/02/2024, <https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar>.
- [3] M. Paula et al. “Predicting Energy Generation in Large Wind Farms: A Data-Driven Study with Open Data and Machine Learning”. Em: **Inventions** 8.5 (2023), p. 126. DOI: 10.3390/inventions8050126.