

Modelagem Preditiva da Bioenergia na Agroindústria Paulista: uma Abordagem via Florestas Aleatórias

Lívia G. S. Souza¹

Instituto de Química/UNESP, Araraquara, SP

Marilaine Colnago²

IBILCE/UNESP, São José do Rio Preto, SP

A bioenergia proveniente principalmente da cana-de-açúcar desempenha um papel estratégico na matriz energética brasileira, destacando-se como uma alternativa sustentável aos combustíveis fósseis. No contexto paulista, essa fonte de energia assume uma relevância ainda maior, visto que o Estado de São Paulo lidera a produção nacional de cana-de-açúcar, responsável por aproximadamente 50% da área plantada no país [1]. Atualmente, cerca de 80% da matéria-prima utilizada para a geração de bioeletricidade no Brasil vem do bagaço da cana, contribuindo para uma produção total superior a 14 mil MW no estado, o que equivale à capacidade instalada da usina hidrelétrica de Itaipu [2]. Esse potencial energético permite não apenas a diversificação da matriz elétrica, mas também a redução da emissão de gases de efeito estufa, consolidando a bioeletricidade da cana como um pilar essencial na transição para uma economia de baixo carbono. Diante da significativa relevância econômica e social da bioenergia, o estudo de modelos matemáticos e métodos computacionais voltados à análise da sua produção torna-se fundamental para embasar decisões estratégicas e eficientes.

Dessa forma, este trabalho pretende explorar o método de aprendizagem Floresta Aleatória (sigla RF, do inglês, *Random Forest*) para estimar a futura produção de bioenergia advinda da cana-de-açúcar no estado de São Paulo. Para tal, foi necessária a integração de dados de diversas fontes em um único modelo de predição, oferecendo uma visão mais completa e precisa do processo de produção de bioenergia, com implicações diretas para a sustentabilidade e a segurança energética em escala global.

O modelo RF é um algoritmo de aprendizagem de máquina supervisionado que cria aleatoriamente árvores de decisão agrupadas, em que é possível ver uma floresta que combina os resultados e chega a uma conclusão final a partir das variáveis escolhidas [3]. Para sua aplicação, considerou-se dados mensais, de janeiro de 2013 a janeiro de 2023, e a fase de teste, foi realizada para o período de um ano, conforme evidenciado na Figura 1. As variáveis utilizadas encontram-se na Tabela 1.

Tabela 1: Variáveis utilizadas, as unidades de medida e fonte.

Variável	Unidade de Medida	Fonte
Bioeletricidade	MWh (megawatt-hora)	Observatório da Cana
Períodos de Safra	-	Instituto Brasileiro de Geografia e Estatística
Condições Hídricas	-	Companhia Nacional de Abastecimento
Área Colhida	Mil hectares	Observatório da Cana
Preço Médio	R\$/Tonelada	CONSECANA
Produção	Mil toneladas	Observatório da Cana

¹livia.souto@unesp.br

²marilaine.colnago@unesp.br

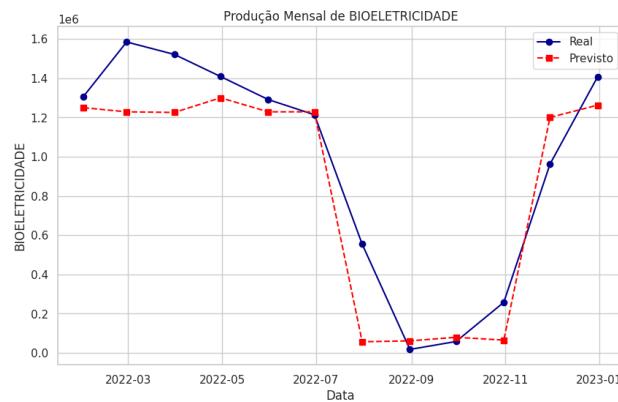


Figura 1: Predição da Produção Mensal de Bioeletricidade. Fonte: autores.

Os valores previstos indicam que a produção real apresenta variações sazonais marcantes, com picos e quedas acentuadas. O modelo captura bem a tendência geral, mas suaviza oscilações abruptas, resultando em discrepâncias em certos períodos. A validação estatística revelou um erro médio absoluto (MAE) de 169.527,11 MWh, evidenciando a sensibilidade do modelo a valores discrepantes. O erro médio percentual absoluto (MAPE) foi de 45%, refletindo a alta variabilidade da geração de bioeletricidade devido a fatores externos não diretamente modelados, como políticas energéticas, flutuações na eficiência industrial e variações climáticas não capturadas integralmente nas condições hídricas. Além disso, a natureza não linear e complexa do processo produtivo pode exigir abordagens híbridas ou a inclusão de novas variáveis para aprimorar a acurácia do modelo.

Em suma, o estudo reforça a importância da bioeletricidade e destaca desafios que precisam ser superados para otimizar sua produção. O uso de algoritmos de aprendizado de máquina mostrou-se promissor para prever tendências futuras, mas melhorias metodológicas, como a incorporação de variáveis adicionais, ajustes nos parâmetros do modelo ou o uso de abordagens híbridas que combinem métodos baseados em aprendizado de máquina com modelos físicos ou econométricos, são necessárias para alcançar maior precisão nas previsões e auxiliar na tomada de decisão estratégica no setor energético.

Agradecimentos

Agradecemos à FAPESP pelo fomento à pesquisa (Nº Processo: 2024/19995-7).

Referências

- [1] CONAB.Companhia Nacional de Abastecimento. **Boletim da safra da cana-de-açúcar**. Online. Acessado em 05/02/2025, <https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar>.
- [2] Secretaria de Infraestrutura e Meio Ambiente. **Potencial da Biomassa no Estado**. Online. Acessado em: 28/05/2024. Disponível em: <https://www.infraestruturameioambiente.sp.gov.br>.
- [3] S. N. Ludovico. “Previsão de indicadores diários de preços no mercado futuro de commodities agrícolas utilizando aprendizagem de máquina.” Dissertação de mestrado. Universidade Federal de Alfenas, Alfenas, 2020.