

Utilização de Inteligência Artificial e Aprendizagem de Máquina para Previsão do Preço do Etanol no Estado de São Paulo

Thomaz A. F. Assis¹

FCT/Unesp, Presidente Prudente, SP

Marilaine Colnago², Wallace Correa de Oliveira Casaca³

IBILCE/Unesp, São José do Rio Preto, SP

O etanol, um biocombustível derivado da cana-de-açúcar, é essencial na matriz energética brasileira, especialmente em São Paulo, líder nacional na produção e consumo desse biocombustível, abrangendo até 45% da fabricação e 60% do consumo no país [4]. Desde a crise do petróleo na década de 1970, o Brasil tem investido no etanol como alternativa aos combustíveis fósseis, reduzindo emissões de gases de efeito estufa, garantindo segurança energética e impulsionando o desenvolvimento econômico regional. Dada sua relevância, é crucial desenvolver modelos matemáticos e técnicas computacionais para otimizar a produção e comercialização, abrangendo desde o cultivo da cana até o controle de qualidade e consumo final do combustível.

Uma técnica computacional amplamente utilizada nos dias atuais é o Aprendizagem de Máquina (AM), um ramo da Inteligência Artificial (IA) que se concentra no desenvolvimento de algoritmos que permitam aos computadores aprender a partir de dados e experiências prévias.

Um dos algoritmos de AM mais populares para tarefas de regressão é o Florestas Aleatórias (RF, do inglês *Random Forest*), uma técnica que combina múltiplos modelos de árvores de decisão para obter previsões mais precisas e estáveis. Assim, cada árvore é treinada de forma independente em uma amostra aleatória do conjunto de dados de treinamento, o que significa que diferentes árvores podem ser treinadas a partir de diferentes subconjuntos de dados e características. A previsão final é então determinada por meio de votação majoritária (no caso de classificação) ou média (no caso de regressão) das previsões individuais de cada árvore [1].

Dado o contexto acima, o presente trabalho propõe estimar o preço futuro do etanol no estado de São Paulo a partir da análise das séries temporais do valor médio transicionado no período de 09 de maio de 2004 a 26 de novembro de 2023. Para essa tarefa, foram empregados algoritmos de AM supervisionados, em particular, a técnica computacional RF, construindo assim um modelo preditivo para a previsão do preço do etanol, combinando diferentes variáveis como o preço de revenda do etanol (R\$), safra (variável binária), preço da saca de açúcar (R\$), preço de revenda do petróleo (R\$), preço do ATR - Açúcares Totais Recuperáveis (R\$), inflação (%), preço de revenda da gasolina (R\$), cotação do dólar (R\$) e preço do açúcar na bolsa de Nova York (U\$).

Previamente à aplicação da modelagem preditiva, foi necessário realizar o pré-processamento dos dados, conduzindo o tratamento de valores inconsistentes (*data cleaning*), além da etapa de Análise Exploratória dos Dados (AED) a fim de avaliar o comportamento das variáveis em estudo. A Figura 1 apresenta a série temporal do preço do etanol, na qual se verifica um pico entre 2021 até meados de 2022, período pós-pandemia COVID-19, fato que aumenta a dificuldade de modelos de AM capturarem com sucesso tal comportamento, uma vez que a série apresenta alta volatilidade [2].

¹thomaz.augusto@unesp.br

²marilaine.colnago@unesp.br

³wallace.casaca@unesp.br

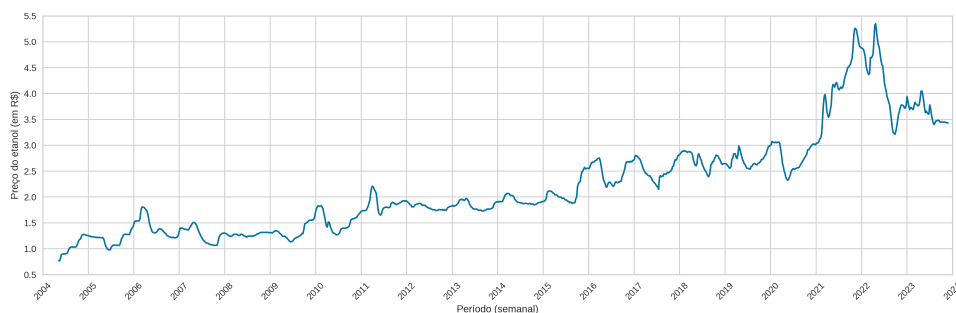


Figura 1: Preço do etanol no estado de São Paulo, de maio de 2004 a novembro de 2023. Fonte: Autoral.

Na sequência, de posse dos dados, foi realizado o treinamento e a prototipação do modelo preditivo definitivo. Para essa tarefa, foram utilizados 5% das instâncias da base para teste, e 95% para treinamento, visto que tal proporção de divisão da base foi necessária em razão da grande variação nos preços de combustíveis em períodos curtos, sobretudo na pandemia.

Para validação dos resultados obtidos, foram adotadas as métricas estatísticas de avaliação, incluindo o Erro Médio Quadrático (MSE), Erro Médio Absoluto (MAE) e Erro Percentual Médio Absoluto (MAPE) [3]. Os escores obtidos encontram-se sumarizados na Tabela 1. Na métrica MSE, o valor atingido é significativamente próximo a zero, apontando um bom estimador. Para a métrica MAE, o modelo preditivo RF estimou satisfatoriamente o preço, visto que tal variável *target* apresenta alta oscilação. Finalmente, a métrica MAPE apresenta um erro próximo ao desejável, se mantendo próxima do valor de referência de 10%.

Tabela 1: Métricas de validação computadas para os resultados produzidos pelo modelo RF.

| Métricas de validação | | |
|-----------------------|------|--------|
| MSE | MAE | MAPE |
| 0,27 | 0,42 | 11,84% |

Esta pesquisa objetivou a criação de um protótipo de modelo preditivo para o preço do etanol no Estado de São Paulo. Esses resultados, de precisão satisfatória, têm o potencial de beneficiar a indústria do etanol, contribuindo para uma melhor previsibilidade no setor. Como desenvolvimentos futuros, a proposta é dar sequência com novos testes envolvendo outras técnicas de AM, bem como reajustar os hiperparâmetros a fim de melhorar as métricas adotadas.

Referências

- [1] T. Hastie, R. Tibshirani e J. H. Friedman. **The elements of statistical learning: data mining, inference, and prediction**. Vol. 2. Springer, 2009.
- [2] F. A. de Oliveira, C. N. Nobre e L. E. Zárate. “Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index—Case study of PETR4, Petrobras, Brazil”. Em: **Expert systems with applications** 40.18 (2013), pp. 7596–7606.
- [3] M. Paula, W. Casaca, M. Colnago, José R da Silva, Kleber Oliveira, Mauricio A Dias e R. Negri. “Predicting Energy Generation in Large Wind Farms: A Data-Driven Study with Open Data and Machine Learning”. Em: **Inventions** 8.5 (2023), p. 126.
- [4] Fundação SEADE. **São Paulo lidera produção de etanol no país**. Online. Disponível em: <https://www.seade.gov.br/sao-paulo-lidera-producao-de-etanol-no-pais>. Acessado em: 22/03/2024. 2021.