

Relatório Técnico: Modelação Preditiva de Etanol Hidratado

Autor: Giovanni Silva **Data:** Novembro de 2025 **Contexto:** Mercado Sucroenergético Brasileiro

1. Resumo Executivo

Este projeto visa mitigar a incerteza na precificação do Etanol Hidratado no mercado brasileiro. Utilizando algoritmos de Machine Learning (**Random Forest**), desenvolvemos um modelo capaz de calcular o "Preço Justo" (*Fair Value*) do biocombustível com base em fundamentos de mercado globais (Petróleo, Câmbio e Açúcar). O modelo atingiu uma acurácia superior a **99% (R^2)** em dados de teste e foi operacionalizado através de uma aplicação web interativa.

2. Definição do Problema de Negócio

O setor de biocombustíveis enfrenta alta volatilidade devido à sua dependência de múltiplas variáveis exógenas.

- **O Problema:** Usinas, distribuidoras e traders precisam decidir diariamente entre vender o stock ou armazenar, bem como fixar preços de contratos futuros.
- **A Hipótese:** O preço do Etanol não é aleatório; ele segue uma paridade económica com a Gasolina (derivada do Petróleo Brent) e compete industrialmente com o Açúcar.
- **Objetivo:** Criar um "GPS de Preços" que identifique distorções de mercado (oportunidades de arbitragem).

3. Metodologia (Pipeline de Dados)

3.1 Fontes de Dados

Foram utilizadas fontes públicas e auditáveis, abrangendo um horizonte temporal de 10 anos (2015-2025):

- **Target (Variável Alvo):** Indicador Diário do Etanol Hidratado (Fonte: CEPEA/ESALQ - Posto Paulínia/SP).
- **Features (Variáveis Explicativas):**
 - Petróleo Brent Futuros (Fonte: Yahoo Finance BZ=F).

- Taxa de Câmbio USD/BRL (Fonte: Yahoo Finance BRL=X).
- Açúcar No. 11 Futuros (Fonte: Yahoo Finance SB=F).

3.2 Engenharia de Dados (ETL)

O processo de ingestão e limpeza envolveu desafios técnicos significativos:

- **Tratamento de Localização:** Correção de formatação numérica brasileira (1.000,00) para o padrão computacional (1000.00) nos dados do CEPEA.
- **Sincronização Temporal:** Unificação de bases com calendários diferentes (feriados locais vs. globais) via *Inner Join*.
- **Feature Engineering:**
 - Criação de variáveis de **Sazonalidade** (Mês), cruciais para capturar os ciclos de safra e entressafra da cana-de-açúcar.
 - Análise de correlação para validação de drivers de preço.

4. Modelação Estatística (Machine Learning)

4.1 Seleção do Algoritmo

Inicialmente testou-se a Regressão Linear, que se mostrou insuficiente para capturar a complexidade não-linear do mercado. Optou-se pelo **Random Forest Regressor** (Floresta Aleatória), um algoritmo de *ensemble* robusto a *overfitting* e capaz de modelar interações complexas entre as commodities.

4.2 Métricas de Desempenho

O modelo foi submetido a validação cruzada e teste em dados desconhecidos (20% da amostra):

- **R² Score (Coeficiente de Determinação): 99.6%**
 - *Interpretação:* O modelo consegue explicar 99.6% da variação de preço do etanol apenas observando o Petróleo, Dólar, Açúcar e o Mês do ano.
- **Robustez:** O modelo demonstrou estabilidade mesmo em cenários de alta volatilidade cambial.

4.3 Análise de Importância das Variáveis

O modelo identificou a hierarquia de influência no preço:

1. **Petróleo Brent:** O driver dominante, confirmando a paridade de preços na bomba (Etanol vs Gasolina).
2. **Câmbio (Dólar):** Fator multiplicador que impacta o custo de importação da gasolina.
3. **Açúcar:** Influência secundária ligada ao *mix* de produção das usinas.

5. Produto Final (Deployment)

Para garantir a aplicabilidade prática, o modelo não permaneceu apenas em código (*Notebook*). Foi desenvolvida uma aplicação web em **Streamlit** que permite:

- Simulação de Cenários de Estresse (ex: "E se o Petróleo for a \$100?").
- Visualização da tendência de paridade em tempo real.
- Cálculo instantâneo do prêmio ou desconto do mercado atual face ao preço justo.

6. Conclusão

O projeto demonstrou que é possível antecipar tendências no mercado sucroenergético com alta precisão utilizando dados públicos e Machine Learning. A ferramenta desenvolvida oferece uma vantagem competitiva significativa para a tomada de decisão estratégica em trading e gestão de stocks.