

Análise Preditiva da Produção de Gasolina Comum no Brasil via Técnicas de Aprendizado de Máquina

Samantha Vieira Lanzelotti¹

Leonardo Fini²

Marilaine Colnago³

Wallace Casaca⁴

Faculdade de Engenharia e Ciências/UNESP, Rosana, SP

A busca incessante pela inovação ocasionado pelo uso escalonado do Aprendizado de Máquina (do inglês, *Machine Learning* - ML) levou à algoritmos cada vez mais robustos e assertivos em diversos contextos de aplicação, o que inclui problemas de modelagem preditiva no setor de energia, como a geração de energia eólica [4] e a produção de petróleo [2].

Considerando o contexto de aplicação acima, a presente pesquisa propôs o estudo, implementação, e avaliação de três modelos de ML, bem como a utilização de ferramentas de Análise Exploratória de Dados (AED) visando desenvolver uma metodologia computacional integrada, que tenha a capacidade de projetar a produção (em m^3) da *gasolina comum* – a mais usada nos postos de combustíveis – no Estado de São Paulo em horizontes de curto/médio prazo. Nesse contexto, é importante destacar que, no início da Pandemia de Covid-19, houve uma diminuição drástica da demanda internacional por combustíveis fósseis, o que levou à restrições na produção de gasolina pela OPEP a fim de manter os preços praticáveis no mercado. Como resultado da pandemia e das políticas acima, o Brasil tem presenciado uma das maiores crises de combustível de sua história [3].

A Agência Nacional de Petróleo (ANP) disponibiliza um vasto repertório de dados públicos sobre a produção de combustíveis, enquanto a plataforma *Investing* reúne dados relacionados aos preços dos combustíveis. Na presente pesquisa, ambas as fontes foram utilizadas de forma a obter uma coleção de dados integrada, a qual foi usada para alimentar as variáveis empregadas no ajuste e treinamento dos modelos preditivos.

Uma vez coletados os dados dos repositórios, estes passaram por uma etapa de tratamento a fim de detectar e tratar valores faltantes e inconsistentes. Além disso, a etapa de AED foi conduzida de modo a compreender melhor a relação entre as principais variáveis do problema, o que inclui a produção da gasolina. Por exemplo, a fim de mensurar o grau de relação das variáveis, foi calculado na equação (1) o Coeficiente de Correlação de Pearson:

$$\rho = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_1^n (x_i - \bar{x})^2} \sqrt{\sum_1^n (y_i - \bar{y})^2}}, \quad (1)$$

em que n representa o tamanho da amostra, x_i e y_i representam um par de variáveis-alvo, sendo \bar{x} e \bar{y} suas respectivas médias.

Foram também criadas variáveis artificiais a partir dos dados brutos, resultando em 17 variáveis no total. O período de análise da presente pesquisa foi de 01/01/2012 a 01/07/2021.

¹samantha.lanzelotti@unesp.br

²l.fini@unesp.br

³marilaine.colnago@unesp.br

⁴wallace.casaca@unesp.br

Tornando-se em cada modelo as 16 variáveis predictoras, e respeitando a divisão em 85% dos dados para fins de treinamento dos algoritmos e o restante (15%) para teste, foram implementados e treinados três modelos de ML para a tarefa de previsão da produção da gasolina: Florestas Randômicas (RF, do inglês, *Random Forest*), Regressão por Vetores de Suporte (SVR, do inglês, *Support Vector Regression*), e Redes Neurais Artificiais (RNA).

Visando mensurar a assertividade dos modelos, foi calculado o Erro Percentual Médio (MAPE) [1] ao longo do período de teste. Foram obtidos escores do MAPE, de 5,76%, 10,98%, e 12,61% para os modelos preditivos RF, SVR e RNA, respectivamente. A Figura 1 apresenta o plot referente aos últimos dias do conjunto de teste para cada um dos modelos de ML implementados.

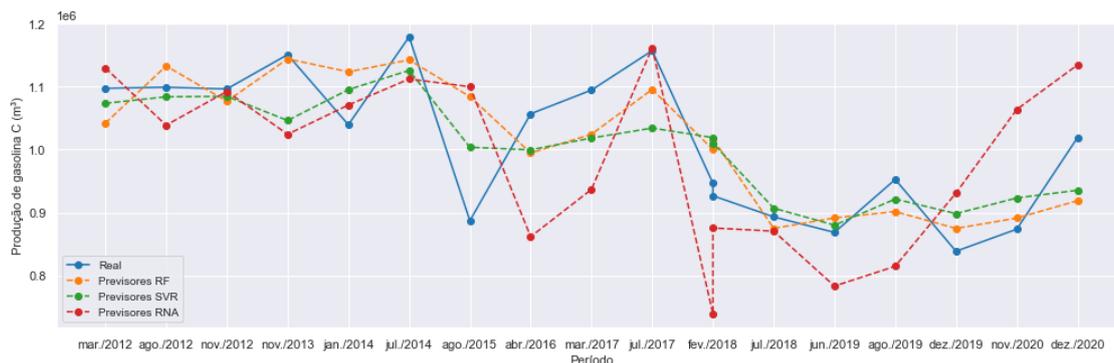


Figura 1: Gráficos da previsão da produção de gasolina comum no Estado de São Paulo.

Averiguou-se que os modelos implementados apresentaram elevados níveis de assertividade na tarefa de previsão da produção da gasolina no Estado de São Paulo. Conclui-se que modelos de ML podem auxiliar atores/gestores na tomada de decisões no setor de petróleo e combustíveis, visto que embora situações inesperadas podem ocorrer, algoritmos baseados em ML tem o potencial de gerar *insights* e prever tendências nas previsões, o que pode ser útil em determinadas circunstâncias.

Agradecimentos

Os autores agradecem à PIBIC/CNPq pelo incentivo e fomento à pesquisa.

Referências

- [1] J. V. Leme, W. Casaca e M. A. Colnago M. and Dias. “Towards Assessing the Electricity Demand in Brazil: Data-Driven Analysis and Ensemble Learning Models”. Em: **Energies** 13.6 (2020), p. 1407. DOI: 10.3390/en13061407.
- [2] W. Liu, W. D. Liu e J. Gu. “Petroleum Production Forecasting Based on Machine Learning”. Em: **Proceedings of the International Conference on Advances in Image Processing**. 2019, pp. 124–128. DOI: 10.1145/3373419.3373421.
- [3] D. G. Oliveira. **Indústria Petrolífera: O Mercado Pós Impactos da COVID-19**. Trabalho de Conclusão de Curso. 2021.
- [4] M. Paula et al. “Predicting Long-Term Wind Speed in Wind Farms of Northeast Brazil: A Comparative Analysis Through Machine Learning Models”. Em: **IEEE Latin America Transactions** 18.11 (2020), pp. 2011–2018. DOI: 10.1109/TLA.2020.9398643.