

Modelagem e Interpretação de Dados Catalíticos via Machine Learning em Perovskitas ABO_3 Aplicadas à Reforma a Seco do Metano

Lamara Maciel dos Santos^{1*}, Dulce M. A. Melo¹, Rodolfo L. B. A. Medeiros¹, Renata Martins Braga¹, Ângelo A. S. de Oliveira¹, Rebecca A. B. N. Santiago¹, Yuri K.R.O. Silva¹, Joyce Cristine A. da Silva¹, Jennifer E.L. Costa¹

¹Laboratório de Tecnologia Ambiental, Universidade Federal do Rio Grande do Norte, 59078-970, Natal, RN, Brasil.

E-mail: lamaramaciel@gmail.com

Resumo/Abstract

RESUMO – A utilização de técnicas de Machine Learning (ML) tem se mostrado uma ferramenta promissora para modelar e interpretar dados catalíticos em sistemas complexos. Neste trabalho, exploramos a aplicação de diferentes algoritmos de ML para prever a conversão de CH_4 e CO_2 a partir de um banco de dados construído com informações extraídas da literatura sobre perovskitas do tipo ABO_3 aplicadas à reforma a seco do metano. Foram testados oito algoritmos combinados com três conjuntos distintos de variáveis de entrada, totalizando 48 modelos avaliados. O modelo de regressão por vetores de suporte (SVR) com kernel radial (RBF) apresentou o melhor desempenho preditivo e foi selecionado para análise interpretativa com SHAP, permitindo identificar as variáveis mais influentes na conversão dos reagentes. A temperatura de reação foi a principal variável para ambos os alvos, seguida pelo tempo de calcinação e pelos teores de dopantes como Co, Fe e Sr. Os resultados obtidos contribuem para a compreensão dos efeitos operacionais e composicionais na atividade catalítica, oferecendo suporte ao planejamento racional de novos experimentos. O estudo também discute os desafios associados ao uso de bases de dados pequenas e heterogêneas na modelagem preditiva em catálise heterogênea.

Palavras-chave: *Aprendizado de Máquina, Catálise Heterogênea, Perovskitas ABO_3 , Reforma a Seco do Metano, Modelagem Preditiva*

ABSTRACT – Machine learning (ML) techniques have proven to be promising tools for modeling and interpreting catalytic data in complex systems. In this work, we explore the application of various ML algorithms to predict CH_4 and CO_2 conversion using a dataset compiled from literature on ABO_3 -type perovskites applied to dry reforming of methane. Eight algorithms were tested in combination with three distinct input variable sets, totaling 48 models evaluated. The support vector regression (SVR) model with a radial basis function (RBF) kernel achieved the best predictive performance and was selected for SHAP analysis to identify the most influential variables driving reagent conversion. Reaction temperature emerged as the dominant factor for both targets, followed by calcination time and the content of dopants such as Co, Fe, and Sr. The results enhance our understanding of how operational and compositional parameters affect catalytic performance, supporting the rational design of future experiments. The study also discusses the limitations of working with small and heterogeneous datasets in predictive modeling for heterogeneous catalysis.

Keywords: *Machine Learning, Heterogeneous Catalysis, ABO_3 Perovskites, Dry Reforming of Methane, Predictive Modeling*

Introdução

A reforma a seco do metano (DRM) representa uma rota catalítica estratégica para a conversão simultânea de dois gases de efeito estufa, CH_4 e CO_2 , em gás de síntese (H_2 e CO), com aplicações relevantes na produção de combustíveis e insumos químicos (1–3). Entre os catalisadores promissores, destacam-se as perovskitas do tipo $LaNiO_3$, devido à sua estrutura flexível, capacidade redox e boa estabilidade térmica (2,4). A dopagem com diferentes cátions tem sido amplamente utilizada para modular essas propriedades, visando aumentar a dispersão das fases ativas, reduzir a formação de coque e melhorar a resistência à sinterização (4–6).

Apesar dos avanços, a correlação entre composição, método de síntese e desempenho catalítico ainda é complexa e de difícil previsão. Nesse contexto, o uso de técnicas de aprendizado de máquina (ML) surge como uma ferramenta valiosa para extrair padrões de dados experimentais heterogêneos e prever o desempenho de materiais catalíticos (7,8). Além de prever resultados, a interpretabilidade de modelos via ferramentas como SHAP (SHapley Additive Explanations) contribui para identificar variáveis-chave e gerar hipóteses fundamentadas para o design racional de catalisadores (9,10).

Este trabalho apresenta uma abordagem de modelagem preditiva baseada em algoritmos de ML aplicados a um banco de dados construído a partir da literatura sobre

perovskitas LaNiO_3 dopadas aplicadas à DRM. O objetivo é avaliar diferentes modelos quanto à sua capacidade de prever a conversão de CH_4 e CO_2 , bem como interpretar os fatores mais influentes no desempenho catalítico.

Experimental

Este trabalho consistiu na construção e avaliação de modelos de aprendizado de máquina para prever a conversão de CH_4 e CO_2 em reações de reforma a seco do metano (DRM), utilizando um banco de dados compilado manualmente a partir da literatura científica. O conjunto de dados reúne informações sobre catalisadores do tipo LaNiO_3 dopados, incluindo composição elementar, método de síntese, condições de calcinação e reação, além dos valores de conversão de reagentes.

O banco foi previamente tratado para remoção de amostras incompletas. Variáveis categóricas (como o método de síntese) foram codificadas via *one-hot encoding*, e os dados numéricos foram normalizados com o método *StandardScaler*, sempre que necessário.

Foram testados oito algoritmos supervisionados de regressão: Random Forest, Gradient Boosting, XGBoost, SVR (RBF), Ridge, Lasso, KNN e MLP. A modelagem considerou três diferentes conjuntos de variáveis de entrada:

- Teste 1 - apenas variáveis contínuas (composição química e condições de síntese/reação);
- Teste 2 - mesmas variáveis do Teste 1 + método de síntese (*one-hot encoded*);
- Teste 3 - apenas La, Ni e condições operacionais (abordagem simplificada).

A avaliação foi realizada por *hold-out* (80% treino, 20% teste), e as métricas utilizadas foram: coeficiente de determinação (R^2), erro absoluto médio (MAE) e raiz do erro quadrático médio (RMSE). Os modelos com melhor desempenho foram interpretados por meio da técnica SHAP (SHapley Additive exPlanations), a fim de identificar as variáveis com maior impacto nas previsões.

Todas as análises foram conduzidas em ambiente Jupyter Notebook, com Python 3.11.3, utilizando as bibliotecas *scikit-learn*, *xgboost*, *shap*, *pandas*, *numpy* e *matplotlib*.

Resultados e Discussão

Análise Descritiva

Antes da etapa de modelagem preditiva, foi realizada uma análise descritiva do banco de dados compilado da literatura, com o objetivo de identificar tendências gerais e potenciais padrões entre os parâmetros de síntese e os valores de conversão reportados. A Figura 1a apresenta a distribuição das amostras em função da temperatura e do

tempo de calcinação. Observa-se uma concentração significativa de experimentos realizados entre 750 e 850 °C, com tempos de 3 a 6 horas, refletindo uma escolha experimental recorrente na literatura. Esses intervalos estão associados à obtenção da fase perovskita desejada, à estabilidade térmica e ao controle da morfologia das partículas (1).

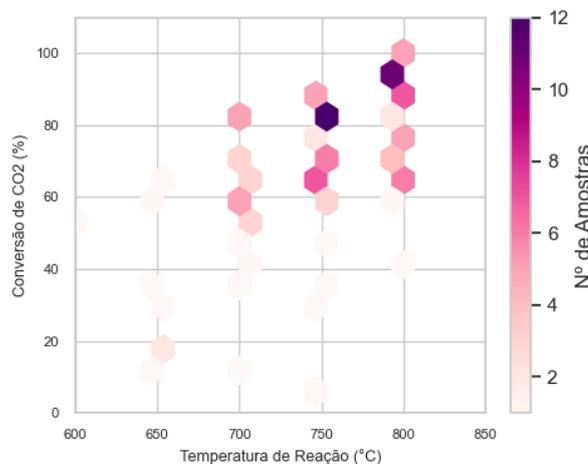
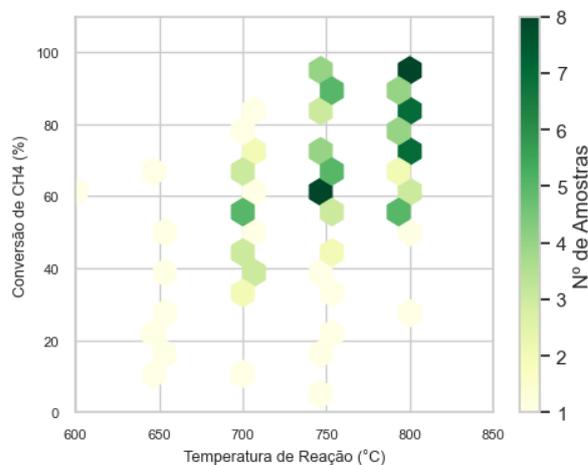
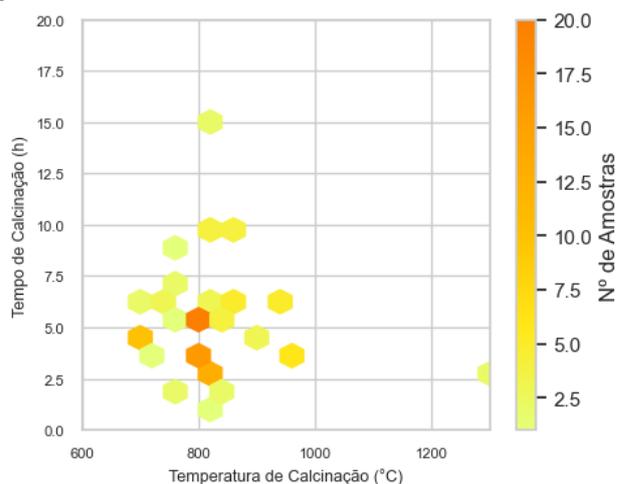


Figura 1. Gráficos hexbin representando a densidade de amostras no banco de dados em função das variáveis mais relevantes. (a) Temperatura de calcinação versus tempo de calcinação; (b) Temperatura de reação versus conversão de CH₄; (c) Temperatura de reação versus conversão de CO₂. Regiões com maior densidade de dados são indicadas por cores mais escuras.

A Figura 1b mostra a relação entre temperatura de reação e conversão de CH₄, evidenciando um aumento da conversão com a elevação da temperatura. A maior densidade de amostras encontra-se entre 750 e 800 °C, faixa na qual também se observam os maiores valores de conversão de CH₄. Esse comportamento é consistente com a natureza endotérmica da reação de DRM, favorecida em temperaturas elevadas (2–4). A Figura 1c, por sua vez, explora a conversão de CO₂ em função da temperatura de reação. Novamente, a faixa de 750–800 °C se destaca como a região de maior densidade de pontos e de melhor desempenho catalítico, com diversas amostras atingindo conversões superiores a 80%. Esses resultados corroboram achados da literatura sobre o bom desempenho de catalisadores à base de LaNiO₃ operando nesse intervalo (5,6). Tais observações antecipam a importância da temperatura de reação, posteriormente confirmada pelas análises com SHAP.

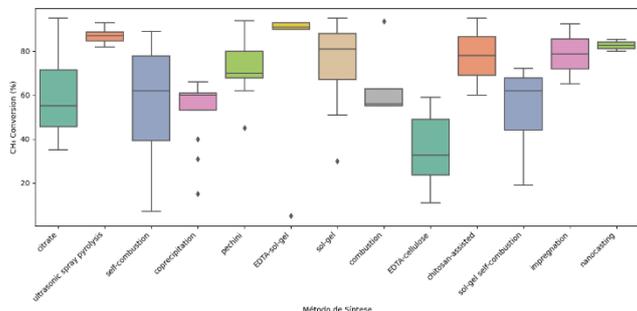


Figura 2. Distribuição da conversão de CH₄ (%) para diferentes métodos de síntese de perovskitas aplicadas à DRM. As caixas representam os quartis (Q1–Q3), a linha horizontal marca a mediana e os pontos fora dos “bigodes” correspondem a outliers.

A Figura 2 apresenta a distribuição da conversão de CH₄ (%) em função do método de síntese empregado para as perovskitas, evidenciando a forte correlação entre a rota sintética e o desempenho catalítico. Os métodos nanocasting e EDTA-sol-gel destacam-se, exibindo altas medianas de conversão (>85%) e baixa dispersão dos dados, o que indica excelente reprodutibilidade e eficiência catalítica. Rotas como ultrasonic spray pyrolysis, sol-gel convencional e chitosan-assisted também apresentam desempenho relevante, com distribuições concentradas e medianas acima de 75%. Em contraste, os métodos EDTA-cellulose e coprecipitation mostraram menor atividade (medianas <60%) e maior variabilidade, sugerindo sensibilidade a

parâmetros de síntese (ex.: temperatura de calcinação, homogeneidade dos precursores). Os resultados corroboram que rotas que favorecem alta área superficial e dispersão metálica (ex.: nanocasting) são críticas para otimizar a atividade catalítica na reforma a seco do metano (1,2).

Desempenho dos Modelos de Machine Learning

O desempenho de oito algoritmos supervisionados de regressão (SVR, RF, XGBoost, GB, Ridge, Lasso, KNN e MLP) foi avaliado com base em três diferentes conjuntos de variáveis de entrada. A Tabela 1 apresenta as principais métricas (R², MAE e RMSE) dos modelos com melhor desempenho para a predição das conversões de CH₄ e CO₂.

Tabela 1. Desempenho dos modelos preditivos

Target	Teste	Modelo	R ² Treino	R ² Teste	MAE Teste
CH ₄ Conversion	1	SVR (RBF)	0.830	0.535	6.75
CH ₄ Conversion	1	GB	0.940	0.383	8.57
CH ₄ Conversion	2	SVR (RBF)	0.932	0.446	8.12
CH ₄ Conversion	3	Ridge	0.678	0.354	9.43
CO ₂ Conversion	2	GB	0.974	0.497	7.18
CO ₂ Conversion	2	SVR (RBF)	0.902	0.452	7.94
CO ₂ Conversion	3	GB	0.973	0.416	7.36
CO ₂ Conversion	3	Ridge	0.574	0.348	8.79

Nota: Os testes referem-se aos diferentes conjuntos de entrada definidos na seção experimental.

O algoritmo SVR com kernel RBF apresentou os melhores resultados globais, com R² de teste de 0,54 para CH₄ (Teste 1) e 0,45 para CO₂ (Teste 2). Apesar de moderados, esses valores indicam uma capacidade de generalização satisfatória, especialmente considerando a heterogeneidade do banco de dados construído a partir da literatura — que reúne perovskitas ABO₃ com diferentes composições no sítio B, métodos de síntese e condições operacionais.

Modelos baseados em Gradient Boosting também se destacaram, com desempenho competitivo, embora apresentem tendência ao sobreajuste (overfitting), como evidenciado pelos altos R² de treino e queda no conjunto de teste. O modelo linear Ridge mostrou resultados consistentes no Teste 3, sugerindo que até abordagens mais simples podem capturar parcialmente as tendências, quando as variáveis são bem definidas.

De forma geral, os resultados demonstram que, mesmo com um banco de dados pequeno e diverso, é possível construir modelos preditivos úteis para orientar o desenho de catalisadores.

Interpretação dos Modelos com SHAP

As Figuras 3 e 4 apresentam os gráficos SHAP (beeswarm) gerados para os modelos SVR treinados com o conjunto de variáveis contínuas (Teste 1), visando prever as conversões de CH₄ e CO₂ na reforma a seco do metano (DRM). Essa abordagem permite avaliar a contribuição individual de cada variável nas previsões do modelo, fornecendo insights quantitativos sobre a influência relativa dos parâmetros operacionais e composicionais no desempenho catalítico.

Na conversão de CH₄ (Figura 3), a temperatura de reação foi a variável mais influente, com valores mais altos associados a impactos positivos na predição, comportamento consistente com a natureza endotérmica da reforma a seco do metano, que favorece a conversão em temperaturas elevadas (1,2). O tempo de calcinação também apresentou contribuição significativa, sugerindo que períodos mais longos favorecem a formação de fases cristalinas e melhor dispersão do metal ativo (3,11).

Entre os elementos químicos, os dopantes Fe, Co e Sr demonstraram influência positiva moderada. Cobalto, em particular, tem sido associado à maior dispersão de Ni e à formação de fases mais ativas (4). Já o teor de Ni apresentou impacto mais neutro ou levemente negativo, possivelmente devido a efeitos de sobrecarga metálica ou sinterização em concentrações elevadas (5).

Na predição da conversão de CO₂ (Figura 4), a temperatura de reação novamente se destacou como a principal variável, indicando que temperaturas mais elevadas também favorecem a ativação e conversão do CO₂. O tempo de calcinação seguiu relevante, ainda que com menor intensidade que no modelo de CH₄. Os dopantes Co e Fe voltaram a se destacar positivamente, possivelmente por atuarem como centros redox ou por promoverem maior basicidade superficial, fator importante para a adsorção de CO₂ (6,7,13). Elementos como Sr e Ti também apresentaram impacto positivo moderado, contribuindo para a estabilidade estrutural e distribuição de fases ativas.

De modo geral, os resultados indicam que a temperatura de reação é o principal fator operacional para ambas as conversões. No entanto, a influência de dopantes metálicos é mais pronunciada para CH₄ do que para CO₂, reforçando que a ativação de metano exige combinações mais específicas de sítios ativos. Esses achados corroboram a importância de uma abordagem integrativa que considere, simultaneamente, variáveis operacionais e composicionais para o desenvolvimento racional de catalisadores perovskíticos para reforma a seco do metano.

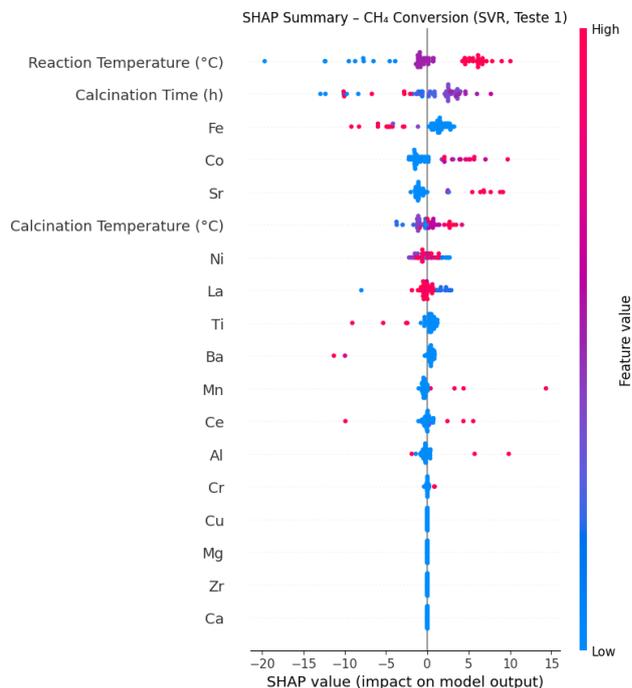


Figura 3. Gráfico SHAP (beeswarm) para o modelo SVR aplicado à conversão de CH₄. O eixo horizontal representa a magnitude da contribuição de cada variável na predição. Cores indicam os valores das variáveis: azul (baixo) a rosa/vermelho (alto).

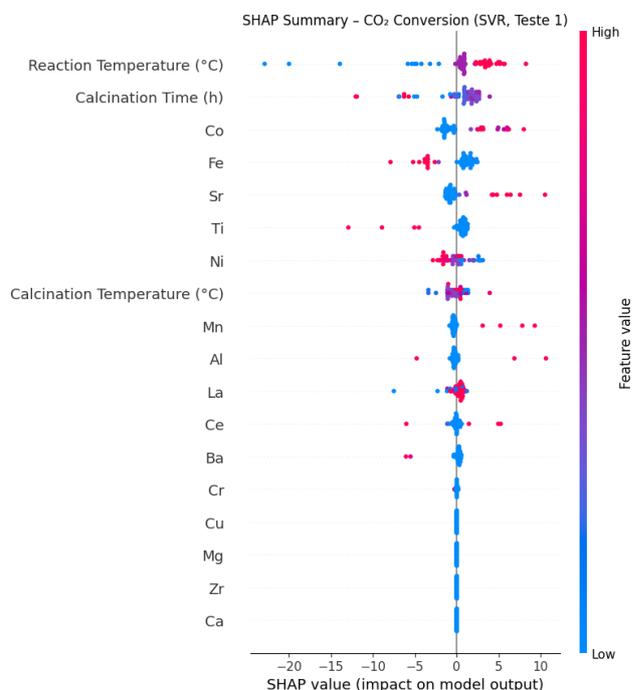


Figura 4. Gráfico SHAP (beeswarm) para o modelo SVR aplicado à conversão de CO₂. O eixo horizontal representa a magnitude da contribuição de cada variável na predição. Cores indicam os valores das variáveis: azul (baixo) a rosa/vermelho (alto).

Conclusões

Neste trabalho, foi explorada a aplicação de algoritmos de aprendizado de máquina para prever a conversão de CH₄ e CO₂ em catalisadores do tipo perovskita ABO₃ aplicados à reforma a seco do metano (DRM), utilizando um banco de dados construído exclusivamente a partir de informações experimentais extraídas da literatura. Foram avaliados oito algoritmos supervisionados combinados com três diferentes conjuntos de variáveis de entrada, totalizando 48 modelos preditivos.

O modelo SVR com kernel RBF apresentou os melhores desempenhos preditivos para os dois alvos, especialmente com o conjunto de variáveis contínuas (Teste 1), atingindo valores de R² de teste de até 0,54 para CH₄ e 0,46 para CO₂. As análises interpretativas com SHAP mostraram que a temperatura de reação foi a variável mais determinante para ambos os alvos, com impacto positivo consistente. Outras variáveis relevantes incluíram o tempo de calcinação e a presença de dopantes como Co, Fe e Sr, que contribuíram positivamente para as conversões, especialmente para CH₄.

As visualizações exploratórias e os gráficos SHAP reforçaram tendências compatíveis com a literatura, oferecendo suporte para hipóteses químicas sobre a influência dos parâmetros composicionais e operacionais. Apesar das limitações associadas ao tamanho e à heterogeneidade do banco de dados, os resultados obtidos demonstram o potencial da modelagem preditiva como ferramenta complementar para a racionalização do design de catalisadores.

Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), ao Fundo Nacional de Desenvolvimento Científico e Tecnológico (FNDCT) e ao Ministério da Ciência, Tecnologia e Inovações (MCTI) pelo apoio financeiro concedido por meio do edital SisH₂.

Referências

1. Q. Wang, P. Su, S. Yuan, Z. Wang, F. Yu, *Int. J. Hydrogen Energy*, 2024, **61**, 4439–4445.
2. A. Bhaskaran, S.A. Singh, P. Da Costa, S. Roy, *Int. J. Hydrogen Energy*, 2024, **61**, 4439–4445.
3. A. Shahnazi, S. Firoozi, *J. CO₂ Util.*, 2021, **47**, 101455.
4. Z. Du et al., *J. CO₂ Util.*, 2023, **67**, 102317.
5. M. Wei, X. Shi, *Methane*, 2024, **3**, 10006.
6. S. Azeem et al., *Int. J. Chem. Eng.*, 2022, Article ID 3139696.
7. Roh et al., *Catal. Today*, 2023, **418**, 113829.
8. Vellayappan et al., *Chem. Eng. J.*, 2024, **468**, 143647.
9. Yoon et al., *Appl. Catal. B*, 2023, **321**, 123456.
10. Wang et al., *AIChE J.*, 2024, **70**, e17901.
11. Favacho et al., *Int. J. Hydrogen Energy*, 2024, **78**, 1391–1428.
12. Favacho et al., *Catal. Today*, 2025, **448**, 115167.
13. Azevedo et al., *J. Energy Inst.*, 2025, **120**, 102123.