

Aplicação de Modelos de IA para Predição dos Índices de CO₂ na Atmosfera

David F. P. Ferreira,¹ Matheus S. Menezes²
UFERSA, Mossoró, RN

Modelos de predição são perseguidos, estudados e aplicados em contextos acadêmicos e no mundo corporativo. Ao se analisar dados gerados por fontes confiáveis, pode-se encontrar padrões e, em seguida, formular modelos preditivos eficientes. No contexto socioambiental um modelo preditivo pode ajudar a direcionar ações mitigadoras ou nortear para soluções eficazes e eficientes, minimizando ou evitando desastre ambientais. Com base nesse cenário, o objetivo geral desse trabalho é demonstrar a eficiência de modelos de aplicação de IA na predição dos índices de CO₂ na atmosfera em comparação aos modelos de regressão via Método dos Mínimos Quadrados.

A regressão linear é um modelo estatístico que estima a relação linear entre uma variável dependente e uma ou mais variáveis independentes[2]. Em *machine learning*, é um tipo de algoritmo usado para prever um valor numérico com base em dados de entrada. Sendo um método estatístico que estuda a relação entre uma variável dependente (a ser prevista) e uma ou mais variáveis independentes (entradas). Assim, a regressão é usada para tarefas de natureza preditiva, quando se deseja estimar um valor numérico com base em dados históricos.

O oxigênio é uma das fontes primárias da vida na terra. Ele também tem uma função protetora para os seres vivos na Terra, por meio do Ozônio (O₃), que gera a Camada de Ozônio (CO_z), posta na Estratosfera, cerca de 25 a 30km acima da superfície. A CO_z é responsável por filtrar a radiação ultravioleta (raios UV). Porém, alguns gases em excesso, como o Dióxido de Carbono (CO₂), classificado como um dos Gases do Efeito Estufa (GEE), enfraquecem este escudo contra os raios solares. Com enfraquecimento da CO_z, tem-se a elevação de temperatura na atmosfera, por meio do Efeito Estufa (EE), o que pode ocasionar inúmeros problemas, tais como: secas intensas, derretimento das calotas polares, tempestades, entre outros desastres naturais com ligação direta ou indireta a elevação da temperatura e/ou da forte incidência de raios solares nocivos a vida na terra, que, em condições normais, deveriam ser bloqueados pela CO_z. Por tanto, monitorar e prever os níveis de GEE, em especial do CO₂, na atmosfera pode nos colocar na vanguarda da busca por soluções desta problemática. Não obstante, o experimento fora conduzido em um tripé, formados por modelos com base nas teorias do Método dos Mínimos Quadrados (MMQ) e *machine learning* (ML) em conjunto com o *Scikit-Learn* (SKLearn), biblioteca de código aberto para ML.

Com base no que dizem [1][3], entende-se o MMQ como uma técnica de otimização matemática usada para encontrar a linha de melhor ajuste que minimiza a soma das diferenças quadradas entre os pontos de dados observados e seus valores previstos correspondentes. Por isso, este modelo foi escolhido para contrastar os modelos de ML.

Os dados utilizados neste estudo foram obtidos da página oficial, e pública, da NASA[4] em parceria com a Agência Norte-americana de Administração Oceânica e Atmosférica (em inglês National Oceanic and Atmospheric Administration - NOAA), que colhe, compila e disponibiliza em *webpage* específica[5]. Os dados utilizados tinham um total de 779 registros que compreende o período de março de 1958 (primeira amostra coletada e disponibilizada) até janeiro de 2023 (data

¹ david.ferreira@alunos.ufersa.edu.br

² matheus@ufersa.edu.br

do experimento). Os dados são mensurados em partes por milhão por volume (ppmv), 315.7 ppmv como exemplo.

Para experimentação buscou-se encontrar as variáveis a , b , por meio dos dados existentes n (quantidade de dados analisados), x (dados decimais - ano.referência) e y (média do mês - em ppmv). Após isso, dividiu-se os dados para treinamento e teste, nas proporções de 70 e 30, respectivamente. Em seguida, nos modelos MMQ e ML, aplicou-se as fórmulas do MMQ nos casos linear e não lineares (logarítmico e exponencial); sendo que, fora preciso aproximar as funções não lineares da função linear.

Após a obtenção dos resultados de cada modelo preditivo, fez-se uma comparação utilizando o coeficiente de Determinação R^2 , além da métrica de erro RMSE, que estão expostos na Tabela 1. Apresenta-se os valores com quatro casas decimais, destacando em negrito os melhores resultados.

Tabela 1: Resultados obtidos.

	MMQ	ML	Scikit-learn
<i>Caso Linear</i>			
R^2	0.9762	0.9757	0.9762
RMSE	2.2108	4.9417	2.2108
<i>Caso Logarítmico</i>			
R^2	0.9750	0.8342	0.9750
RMSE	2.2378	12.8949	2.2378
<i>Caso Exponencial</i>			
R^2	0.9849	0.8625	0.9849
RMSE	1.7831	0.0324	0.1036

Assim, com base nos dados exposto na **Tabela 1**, pode-se verificar a eficácia do modelo ML, bem como, da biblioteca *SKLearn* quando comparamos estes modelos ao MMQ, um modelo matemático robusto e consolidado. Pode-se verificar, também, que o Caso Exponencial foi o que melhor se ajustou, tendo os melhores parâmetros em todas as métricas.

Agradecimentos

À Universidade Federal Rural do Semi-Árido (UFERSA), pelo apoio e incentivo a iniciação científica; e em especial, ao meu orientador, que sub-escreve, pela paciência e dedicação.

Referências

- [1] C. G. de Almeida. “Cálculo Numérico”. Em: (2019).
- [2] D. C. Montgomery, E. A. Peck e G. G. Vining. **Introduction to linear regression analysis**. John Wiley & Sons, 2021.
- [3] D. C. Montgomery e G. C. Runger. **Applied statistics and probability for engineers**. John wiley & sons, 2010.
- [4] NASA. **Vital Signs: carbon dioxide**. Ed. por S. Callery HOLLY SHAFTEL. Online. Acessado em 21/09/2022, <https://climate.nasa.gov/vital-signs/carbon-dioxide/>. 2022.
- [5] NOAA. **CO2 Measurements: Monthly Measurements from Mauna Loa Observatory**. Online. Acessado em 21/09/2022, https://gml.noaa.gov/webdata/ccgg/trends/co2/co2_mm_mlo.txt. Havaí, 2022.