

## **DESENVOLVIMENTO DE UM MODELO DE APRENDIZAGEM DE MÁQUINA PARA PREDIÇÃO DE CONSTANTES CINÉTICAS DE INTERESSE AMBIENTAL**

**Gabriel Longo Lopes<sup>1</sup>, Bruno Ramos<sup>1</sup>, Flávio Olímpio Sanches-Neto<sup>2</sup>, Antonio Carlos Silva Costa Teixeira<sup>1</sup>**

<sup>1</sup> *Escola Politécnica da Universidade de São Paulo, São Paulo, Brasil*  
*(gabriel.longo@usp.br)*

<sup>2</sup> *Instituto de Química, Universidade de Brasília, Distrito Federal, Brasil*

A água é um recurso escasso e de extrema importância à vida. Apesar disso, parte considerável das reservas de água doce do mundo já contém diversos poluentes de origem antrópica. Nos corpos d'água, os contaminantes são degradados por rotas distintas: alguns de origem orgânica não sofrem destruição por vias bióticas, mas somente através de fotólise. Assim, é importante estudar sua cinética de destruição por estas rotas. Uma das formas de fotólise é por via indireta, quando o fóton é absorvido por uma outra substância, gerando uma espécie reativa que interage com o poluente, degradando-o. A determinação experimental da cinética dessas rotas é complexa. Dessa forma, alternativas computacionais são exploradas, como o uso de inteligência artificial para prevê-las. Este trabalho apresenta um algoritmo de aprendizagem de máquina (*machine learning*) que estima a constante de destruição de poluentes orgânicos por meio de uma dessas espécies reativas, o *oxigênio singlete* ( ${}^1\text{O}_2$ ). Para isso, foi desenvolvida uma base de dados com valores experimentais presentes na literatura e três modelos de aprendizagem (*Redes Neurais, Floresta Aleatória e XGBoost*) com dois tipos de descritores moleculares (*molecular fingerprints*): MACCS e Morgan. A precisão dos algoritmos foi avaliada por meio do coeficiente de determinação ( $R^2$ ) e do erro quadrático médio (RMSE). Além disso, foi realizada a análise SHAP (explanação composta dos valores Shapley), do modelo para verificar a consistência físico-química das previsões. Todos os métodos estudados apresentaram previsões satisfatórias, com  $R^2 > 0,63$  e  $\text{RMSE} < 0,0096$ . O que apresentou melhor resultado utilizou redes neurais com descritores MACCS, com  $R^2 = 0,9033$  e  $\text{RMSE} = 0,0025$ . A análise SHAP dos modelos demonstra que eles preveem as constantes cinéticas coerentemente com os mecanismos de reação descritos na literatura: quanto maior o número de carbonos aromáticos no poluente, maior é a velocidade de reação; por outro lado, a presença de grupos receptores de elétrons na substância a diminui. Assim, conclui-se que os métodos empregados de aprendizagem de máquina realizam boas previsões com significado físico.

**Palavras-chave:** Aprendizagem de máquina; Constante cinética; Oxigênio Singlete.