
Cinética de degradação de compostos orgânicos através de modelagem molecular, inteligência artificial e experimentação

Souza Teles, Eduardo¹

Instituto Federal de Goiás, Câmpus Valparaíso de Goiás

Olimpio Sanches Neto, Flávio²

Instituto Federal de Goiás, Câmpus Valparaíso de Goiás

A contaminação de ecossistemas aquáticos por poluentes orgânicos tem se consolidado como uma preocupação ambiental de escala global, uma vez que tais substâncias podem comprometer a qualidade da água, afetar a biodiversidade e trazer riscos diretos à saúde humana. Entre os parâmetros mais relevantes para compreender o impacto e a persistência desses compostos no meio ambiente, destacam-se as constantes de velocidade cinética, que fornecem informações cruciais sobre a taxa de degradação e transformação química dessas moléculas. Entretanto, a determinação experimental dessas constantes é um processo frequentemente demorado, custoso e, em alguns casos, tecnicamente desafiador, o que limita a disponibilidade de dados confiáveis em larga escala. O estado da arte aponta para o uso de abordagens computacionais como alternativa viável para suprir essa lacuna. Modelos de

predição baseados em métodos estatísticos e, mais recentemente, em técnicas de Aprendizado de Máquina (Machine Learning), têm ganhado destaque por possibilitar análises rápidas e de menor custo, com potencial de alcançar elevada acurácia. Contudo, ainda persiste a necessidade de desenvolver ferramentas robustas, capazes de lidar com a diversidade estrutural dos poluentes e fornecer estimativas consistentes. Neste trabalho, propôs-se a aplicação de algoritmos de Machine Learning para prever as constantes cinéticas de poluentes orgânicos a partir de descritores moleculares extraídos de suas estruturas químicas. Foram desenvolvidos e comparados dois modelos de regressão supervisionada: Random Forest (RF) e XGBoost (XGB). A análise de desempenho demonstrou a viabilidade da abordagem. O modelo Random Forest apresentou resultados mais satisfatórios, alcançando um coeficiente de determinação (R^2) de 0,632 e uma Raiz do Erro Quadrático Médio (RMSE) de 0,167. Esses valores indicam que o modelo foi capaz de explicar aproximadamente 63% da variância dos dados, configurando uma correlação moderada, mas estatisticamente relevante. Conclui-se, portanto, que a utilização de Machine Learning representa uma estratégia eficaz e promissora para a estimativa da reatividade de poluentes orgânicos, fornecendo uma metodologia de baixo custo, com maior rapidez e aplicabilidade em larga escala. Tal avanço pode acelerar estudos de risco ambiental e oferecer subsídios científicos importantes para políticas públicas de saneamento, controle de poluição e preservação dos recursos hídricos.

Palavras-chave

Python; Propriedades Físico-Químicas; Poluentes Orgânicos; Análise de Dados; Machine Learning.

Agradecimentos

O presente trabalho foi realizado com apoio do Instituto Federal de Goiás. Agradecemos ao CNPq pela bolsa concedida.