



PREVISÃO DA CAPACIDADE DE ADSORÇÃO DE GASES LEVES ATRAVÉS DE ANÁLISE EXPLORATÓRIA DE DADOS E REGRESSÃO POR VETORES DE SUPORTE

THAYLANE DA R. BEZERRA^{1*}, LHUCAS T. DE M. DE SOUSA¹, SARAH. A. ALTINO¹

¹Universidade Federal de Uberlândia, Programa de Pós-Graduação em Engenharia Química
Faculdade de Engenharia Química
*e-mail: thaylane.bezerra@gmail.com

RESUMO - Este estudo teve como objetivo extrair conhecimento sobre a adsorção de gases leves por materiais microporosos (zeólitas, MOFs e carvões de ativação) através de análise exploratória e Regressão por Vetores de Suporte (RVS) dos dados reportados em 9 artigos publicados entre os anos 1974 e 2019. Um banco de dados contendo 1086 pontos de dados exibindo os efeitos de 7 variáveis de entrada (área superficial do sólido; temperatura e pressão do experimento; técnica de medida da capacidade de adsorção, polarizabilidade, diâmetro cinético e massa molecular dos gases) sobre a capacidade de adsorção foi construído. Histogramas, gráficos de barra e dispersão foram aplicados (como parte da análise exploratória de dados) para determinar como várias variáveis de entrada relacionam entre si e com a variável de desempenho. Além disso, a RVS foi usada para determinar a importância relativa das variáveis de entrada e as possíveis combinações dessas variáveis levando a algum desempenho (para deduzir heurísticas para alta ou baixa capacidade de adsorção). Constatou-se a partir da análise exploratória dos dados que a pressão, temperatura, polarizabilidade e massa molar do gás foram as variáveis mais significativas que afetaram a capacidade de adsorção. Além disso, combinações de variáveis de entrada que levam a um alto desempenho de adsorção foram reveladas por meio da análise do modelo RSV, as quais podem ser usadas como diretrizes para estudos futuros nesta área.

INTRODUÇÃO

As indústrias químicas e petroquímicas estão cada vez mais conscientes de que o seu desenvolvimento sustentável depende fortemente da aplicação de processos inovadores que utilizem materiais e energia de forma mais eficiente. Os processos de separação e purificação representam em torno de 40 a 60% do capital e dos custos operacionais industriais, respectivamente. A melhoria nesses processos pode reduzir expressivamente os custos, o uso de energia e a geração de resíduo (Pullumbi *et al.*, 2019).

Entre os processos de separação e purificação, encontra-se a adsorção que é uma operação de transferência de massa que ocorre a partir da propriedade que alguns sólidos

possuem de aderir em sua superfície determinadas substâncias presentes em fluidos líquidos ou gasosos. Em relação a vários outros métodos de separação e purificação nas indústrias, a adsorção em colunas tem se destacado devido ao seu baixo custo, a ser um processo descomplicado e ecologicamente correto. Processos baseados em adsorção podem ser uma alternativa atrativa em relação aos processos atuais baseados em destilação criogênica que requerem um alto consumo de energia (Bordonhos *et al.*, 2021; Majd *et al.*, 2022).

Devido a atratividade oferecida pela possível redução no consumo de energia, o desenvolvimento de um método eficaz de separação de gases, como é o caso da adsorção, tem sido incentivado e estudado. Pesquisadores vêm investigando diferentes

tipos de adsorventes como carvão ativado, zeólitas, titanossilicatos e materiais semelhantes à zeólitas para a separação de hidrocarbonetos e gases leves (Golipour *et al.*, 2020). E, a todo tempo, vários sólidos e materiais estão sendo descobertos. Logo, é preciso compreender como as características de um sólido e as condições de processo afetam o fenômeno de adsorção para identificar o melhor candidato para uma aplicação específica. Entretanto, a análise experimental das propriedades de transporte de novos materiais pode ser demorada e estudos precisos requerem equipamentos especializados, além de ser considerado economicamente inviável o teste experimental de muitos materiais. Desta forma, a triagem rápida de materiais via técnicas computacionais seria vantajosa para avaliar o potencial de adsorção de um gás puro, ajudando a concentrar o esforço experimental nos sistemas mais promissores (Tang *et al.*, 2021; Wang *et al.*, 2022; Yuan *et al.*, 2021a).

As ferramentas da análise de dados em conjunto com os modelos de *Machine Learning* (ML), em português Aprendizado de Máquina, têm se mostrando como guias poderosos em diversos campos científicos e tecnológicos para avaliação e descrição de fenômenos e padrões. Essas ferramentas e modelos vêm ganhando destaque em diversos estudos na área de adsorção para triagem rápida de materiais promissores para determinadas tarefas (Cao, 2022a; Li *et al.*, 2022; Zhang; Zheng; He, 2022a).

Assim, este estudo teve como objetivo realizar uma análise detalhada da complexa relação entre os parâmetros potenciais e os desempenhos de adsorção de zeólitas, estruturas metálicas orgânicas e carvões ativados para o desenvolvimento de adsorventes de alta capacidade. Além disso, objetivou-se criar um banco de dados experimentais de adsorção de gases leves: CO₂, N₂, CH₄, C₂H₄ e C₂H₆ fazendo uso de análise exploratória de dados e Regressão por Vetores de Suporte (RVS).

METODOLOGIA

Construção de banco de dados

A base de dados foi construída extraindo os dados de adsorção sobre diversos materiais publicados entre os anos 1974 e 2019. Primeiramente, mais de 30 artigos foram estudados e as variáveis potenciais que poderiam ser críticas no desempenho dos adsorventes foram coletadas. Na segunda etapa, todas as publicações foram reavaliadas a fim de remover aquelas que não continham informações sobre as propriedades selecionadas (por exemplo, propriedades estruturais, etc.) juntamente com os dados de adsorção.

Ao fazer isso, foram eleitas apenas zeólitas, MOFs e carvões ativados como as estruturas mais usualmente investigadas. Devido à natureza das metodologias de mineração de dados, não deve haver nenhum valor exclusivo ou alguns valores para uma variável de entrada no banco de dados usado para modelagem. Isso acontece porque algumas técnicas utilizadas na modelagem são indutivas. Um ou poucos exemplos impossibilitam a previsão ou generalização dos padrões. Portanto, materiais raramente utilizados não puderam ser admitidos no banco de dados, embora pudessem apresentar resultados animadores. Pelo mesmo motivo, a análise de dados ao longo deste trabalho é estabelecida nas faixas de variáveis mais frequentemente acessíveis na literatura.

Por fim, a base de dados criada foi composta por 1086 dados extraídos de 9 publicações, as quais estão listadas na Tabela 1.

O banco de dados foi construído com base em 7 variáveis de entrada agrupadas em quatro: variáveis de operação, técnica de medida da capacidade de adsorção, propriedades adsorventes e propriedades do gás. A lista das variáveis de entrada juntamente com suas faixas é apresentada na Tabela 2. Como a maioria das avaliações de desempenho de adsorção de gases puros são baseadas na capacidade de adsorção (em mmol/g adsorvente), esta variável foi tomada como variável alvo na modelagem.

É importante salientar que, a princípio, buscou-se incluir no banco de dados

informações sobre densidade aparente do sólido, densidade real e porosidade. Contudo, não foi possível coletar 30% de exemplos dotados destas características, razão pela qual tais variáveis foram abandonadas.

Tabela 1: Detalhes sobre os artigos utilizados para a criação do banco de dados.

Referência	Gases	Sólido
Reich (1974)	C ₂ H ₄ , C ₂ H ₆ , CH ₄ e CO ₂	Carvão ativado BPL
Danner; Choi (1979)	C ₂ H ₄ e C ₂ H ₆	Zeólita 13X
Costa <i>et al.</i> (1989)	C ₂ H ₄ , C ₂ H ₆ e CH ₄	Carvão ativado AC-40
Mofarahi; Salehi, (2013)	C ₂ H ₄ e C ₂ H ₆	Zeólita 5A
Zhang <i>et al.</i> (2015)	N ₂ , CH ₄ e CO ₂	MOF MIL-101
Martins <i>et al.</i> (2015)	C ₂ H ₄ e C ₂ H ₆	MOF Cu- BTC
Charalambous <i>et al.</i> (2018)	N ₂ e CO ₂	Zeólita AQSOA- FAM-Z02
Kloutse <i>et al.</i> (2018)	N ₂ , CH ₄ e CO ₂	MOF ZIF- 8 e Zeólita 13 X
Khoramzadeh; Mofarahi; Lee (2019a)	N ₂ e CO ₂	Zeólita 13X, 5A, 4A e Beta

Durante o pré-tratamento dos dados, observou-se que as variáveis: área superficial e volume médio de poros estavam correlacionadas. Tal fato já era esperado, tendo em vista a natureza do método BET (Freeman; Mcleod, 1983), o qual foi aplicado em todas as referências citadas. Como a área superficial foi a informação reportada em todos os artigos, tal variável foi adotada na modelagem subsequente.

Este trabalho visa estimar a capacidade adsorptiva de gases puros leves em mais de um tipo de sólido, utilizando a técnica de aprendizagem de máquina associada a um banco de dados construído a partir de dados coletados da literatura. Trabalhos recentes envolvem bancos de dados gerados por simulação, em geral, calculados por métodos de química teórica. Além disso, os estudos realizados focam na avaliação da adsorção de uma única espécie gasosa (Cao, 2022b; Huang *et al.*, 2022; Meng; Zhong; Wei, 2020; Amar *et al.*, 2022; Raji *et al.*, 2022; Xie; Zhang, 2022; Yan *et al.*, 2021, 2022; Zhang; Zheng; H, 2022b).

Através de dados gerados por simulação, o pesquisador pode coletar uma gama de informações sobre a topografia do sólido. Fato que não está acessível para um banco de dados dotado de informações experimentais. Por esta razão, no banco de dados analisado, trabalhou-se com um artifício para expressão dos efeitos de peneiramento molecular e avaliação da interação energética dos os sítios ativos: foram computadas variáveis (ou *features*) que indicam características físico-químicas dos gases adsorvidos: diâmetro cinético, peso molecular e polarizabilidade.

Tabela 2: Detalhes sobre as variáveis contidas no banco de dados.

Variável de entrada	Intervalos para as variáveis contínuas ou identidades para as variáveis discretas
Área superficial BET [m ² /g]	284,31 - 2560,00
Temperatura [K]	212,7 - 373,15
Pressão [KPa]	0,00 – 6842,00
Método de medida [-]	Gravimétrico e volumétrico
Diâmetro cinético do adsorvato [Å]	3,300 - 4,443
Peso molecular do adsorvato [g/mol]	14,01 - 44,01
Polarizabilidade do adsorvato [cm ³]	1,7403E ⁻²⁴ - 4,46E ⁻²⁴

Com base nas características de exclusão de tamanho molecular do adsorvente, poros estreitos não estão disponíveis para adsorbatos com diâmetros maiores que a largura do poro enquanto que este mesmo poro pode ser acessível para um adsorbato menor (Lashaki *et al.*, 2012). Como resultado, entende-se que o diâmetro cinético é uma variável que está diretamente relacionada ao volume de poros real do adsorvente. Seguindo este mesmo raciocínio, o peso molecular expressa-se como mais uma variável relacionada ao volume acessível do material.

Por outro lado, as interações eletrônicas entre a superfície dos sólidos podem ser explicadas (em parte) com base na polarizabilidade das moléculas. A análise teórica do fenômeno de adsorção mostra que diversos grupos funcionais disponíveis na superfície do sólido são polarizáveis e estão diretamente associadas ao aumento da adsorção de certos gases, em especial, os com maior capacidade de formar dipolos instantâneos (Meek *et al.*, 2012; Schäf *et al.*, 2020).

Detalhes sobre a modelagem e uso de ferramentas computacionais

Para realização da análise exploratória de dados e regressão com modelo de aprendizado de máquina foi empregado o *software* online *Google Colaboratory* em ambiente de programação Python (versão número 3.7.13). Para análise de dados, as principais bibliotecas de uso foram *numpy*¹, *matplotlib*², *seaborn*³ e *pandas*⁴.

Neste trabalho, a biblioteca *scikit-learn*⁵ foi empregada para implementação do algoritmo chamado RVS, o qual pode ser entendido como um algoritmo que constrói hiperplanos em um espaço n-dimensional para regredir dados. O RVS é um modelo muito poderoso e versátil e, por esta razão, é um dos modelos mais populares na área do conhecimento de aprendizado de máquina. O RVS é particularmente adequado para a regressão de conjuntos de dados complexos, porém de pequeno ou médio porte. (Géron, 2019).

Para o sucesso do algoritmo são aplicadas aos dados as funções de *kernel*, as quais distribuem os pontos não linearmente separáveis. Ou seja, essas funções realizam transformações nos pontos, tornando possível sua separação por um hiperplano. As funções de *kernel* podem ser bastante simples (Tharwat, 2019). No caso do presente trabalho, foram estudados os *kerneis* implementados e disponíveis na biblioteca em uso: *linear*, *Radial Basis Function* (RBF), sigmoide e polinomial de grau 3.

Para o uso do modelo de aprendizado de máquina supervisionado, 20% do banco de dados foi separado para o conjunto de teste e 80% para o conjunto de treino. Esta proporção é amplamente aceita pela comunidade científica (Alibakshi, 2018) e foi calibrada via testes preliminares realizados pela construção de curvas de aprendizado⁶.

Para busca dos (hiper)parâmetros ótimos do modelo de SVR foi utilizada a biblioteca computacional *Optuna*⁷. Tal biblioteca permite a pesquisa automatizada de (hiper)parâmetros ao implementar o algoritmo Bayesiano como metodologia de otimização (Pravin *et al.*, 2022). Para busca extensiva da região de ótimo em todo o espaço de parâmetros, foi empregado o buscador *GridSearchCV*⁸. Os (hiper)parâmetros que foram avaliados durante a otimização, bem como sua breve descrição, são apresentados na Tabela 3.

A função objetivo de otimização pela rotina implementada com o auxílio da biblioteca *Optuna* trabalhou com a maximização da média dos coeficientes de determinação (R^2) obtidos pelo treinamento do modelo. Inicialmente, realizou-se uma avaliação da influência do número de *k-folds* sobre o coeficiente de determinação e verificou-se não existir influência significativa para valores até 20. Desta forma, durante o treinamento, foi realizado o procedimento de validação cruzada⁹ com *k-folds* fixado em 7 (Zhang; Liu, 2022). Para cada procedimento de busca, 300 passos iterativos foram realizados para a busca do ponto de máximo da função objetivo.

¹ <https://numpy.org/>

² <https://matplotlib.org/>

³ <https://seaborn.pydata.org/>

⁴ <https://pandas.pydata.org/>

⁵ <https://scikit-learn.org/stable/>

⁶ https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.learning_curve.html

⁷ <https://optuna.org/>

⁸ https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

⁹ https://scikit-learn.org/stable/modules/cross_validation.html

Tabela 3: (Hiper)parâmetros, sua descrição breve e intervalo de busca.

(Hiper)parâmetros	Descrição	Kernel aplicado	Intervalo
C	Penalidade do erro	Linear, polinomial, RBF e sigmoid	1- 100
ε	Distância euclidiana	Linear, polinomial, RBF e sigmoid	10^{-2} - 2
γ	Sensibilidade às diferenças nos vetores de erro das características	Polinomial, RBF e sigmoid	10^{-3} - 2
$Coef0$	Termo independente na função do kernel	Polinomial e sigmoid	0 - 4
Tol	Tolerância para o critério de parada	Linear, polinomial, RBF e sigmoid	10^{-6} - 0,1

Por fim, a importância das variáveis do modelo foi avaliada pelo recurso de permutação, a qual é uma técnica de inspeção de modelo que pode ser usada quando os dados são tabulares¹⁰. A importância de uma variável é definida como a diminuição na pontuação de um modelo quando uma única variável é removida aleatoriamente. Este procedimento quebra a relação entre a variável e o alvo. Portanto, a queda na pontuação do modelo é indicativa de quanto o modelo depende da variável (Breiman, 2001).

Outra técnica de interpretabilidade do modelo que é o gráfico de dependência parcial (PDP) foi aplicada. Os PDPs¹¹ possuem uma função regressora que marginaliza o impacto de todos os recursos de entrada no modelo de previsão, exceto para um (ou dois, se necessário) de interesse, apresentando assim análise de sensibilidade local.

RESULTADOS E DISCUSSÃO

Análise básica do banco de dados

A Figura 1 apresenta o número de dados dos tipos de gases adsorvidos tendo por base a Classe do Adsorvente. Foram coletados dados para todos os gases sendo adsorvidos em todas as Classes de Adsorventes, exceto N_2 adsorvido em carvão ativado. De forma geral, nota-se que o banco de dados é bastante heterogêneo em termos das espécies

selecionadas, apresentando 300 dados para CO_2 , 240 para N_2 , 201 para C_2H_6 , 196 para C_2H_4 e 149 para CH_4 . Grande parte dos dados referem-se à adsorção de CO_2 e N_2 em zeólitas, dados que estão amplamente disponíveis na literatura devido ao apelo tecnológico dos processos dos processos de captura de CO_2 , purificação dos gases de chaminé e, ao mesmo tempo, da ampla disponibilidade comercial de materiais zeolíticos (Flanigen, 2001; Khoramzadeh; Mofarahi; Lee, 2019b).

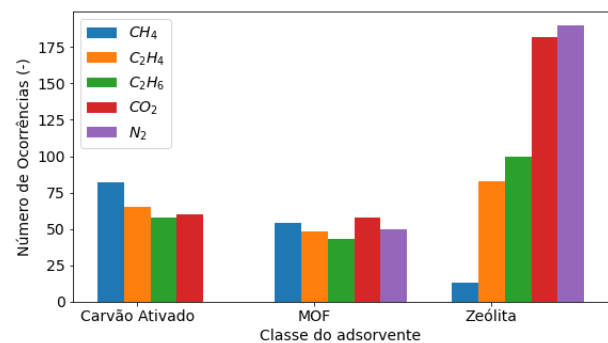
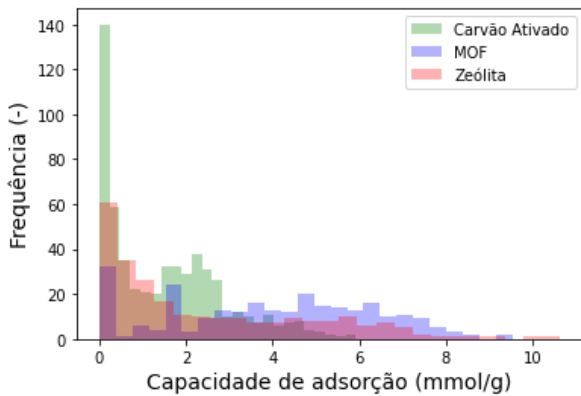


Figura 1: Número de dados listados para cada gás tendo por base a classe do adsorvente.

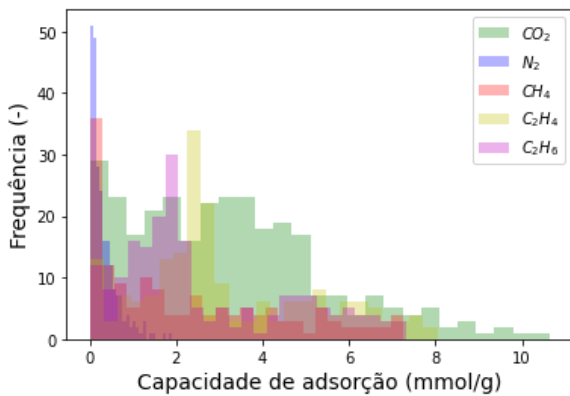
As Figuras 2(a) e 2(b) apresentam histogramas com os valores de frequência de capacidade adsorvida, tendo-se por base o tipo de adsorvente e o gás adsorvido, respectivamente. A análise da Figura mostra que o banco de dados apresenta Classes de Sólidos com desempenhos variados, o que é fundamental quando se deseja aplicar um algoritmo de aprendizado indutivo.

¹⁰ https://scikit-learn.org/stable/modules/permutation_importance.html#id2

¹¹ https://scikit-learn.org/stable/modules/generated/sklearn.inspection.plot_partial_dependence.html



(a)



(b)

Figura 2: Frequência dos valores de capacidade adsorvida indicados em termos das (a) classes de adsorventes e (b) dos tipos de gases adsorvidos disponíveis no banco de dados.

A Figura 3 apresenta uma sequência de gráficos de dispersão e histogramas para as variáveis Temperatura (K), Pressão (kPa) e Área BET média (m^2/g), tendo por base o método empregado para obtenção das medidas de adsorção. Nota-se que o banco de dados idealizado apresenta, em sua maioria, dados coletados a pressões baixas e temperaturas próximas à ambiente. Tal constatação pode estar relacionada ao fato de baixas pressões e temperaturas ambientes serem comumente adotadas na etapa de adsorção ou dessorção de diversos sistemas de colunas como PSA (*Pressure Swing Adsorption*), VPSA (*Vacuum Pressure Swing Adsorption*) e TSA (*Temperature Swing Adsorption*) (Danaci *et al.*, 2021; Speight, 2019). Prevalentemente, também são observados dados contendo áreas BET médias inferiores à $1000 \text{ m}^2/\text{g}$, característica esta referente a maioria dos

adsorventes sintéticos (Karimi *et al.*, 2019). Não foi possível observar qualquer tendência entre as variáveis ilustradas e os métodos empregados nas medidas experimentais.

Na sequência, realizou-se uma análise descritiva de todas as variáveis contínuas de entrada e variável alvo, submetendo-as a análise de média, desvio padrão, mínimo e máximo, com o objetivo de adquirir uma percepção inicial do conjunto de dados. Os dados dessa avaliação são apresentados na Tabela 4.

Ao observar a correlação de Pearson entre o conjunto de variáveis de entrada do banco de dados utilizado para previsão da capacidade de adsorção dos gases leves, Figura 4, nenhuma delas apresentou uma forte correlação entre si. A ausência de uma forte correlação entre as variáveis de entrada favorece para que todas elas contribuam individualmente para a construção do modelo preditivo (Yuan *et al.*, 2021b).

Observa-se ainda pela Figura 4 que as variáveis com maior correlação linear com a capacidade de adsorção são: pressão, temperatura, polarizabilidade e massa molar do gás.

Aplicação do algoritmo RVS

O algoritmo de RVS foi utilizado para prever a capacidade de adsorção dos gases leves com base no conjunto das variáveis de entrada apresentadas na Tabela 2. Quatro modelos foram gerados com a aplicação dos seguintes *kernel*s: linear, polinomial de grau 3, RBF e sigmoide.

Para a avaliação dos (hiper)parâmetros ótimos para cada kernel por meio da biblioteca Optuna, considerou-se o número de *k-folds* igual a 7. Esse valor foi arbitrado visto que se observou por meio de avaliação preliminar que esse parâmetro não impactou significativamente no ajuste do modelo de previsão da capacidade de adsorção de gases leves, conforme pode-se observar pela Figura 5 e Figura 6.

A Tabela 5 apresenta os resultados dos (hiper)parâmetros ótimos do modelo de SVR para cada *kernel* avaliado.

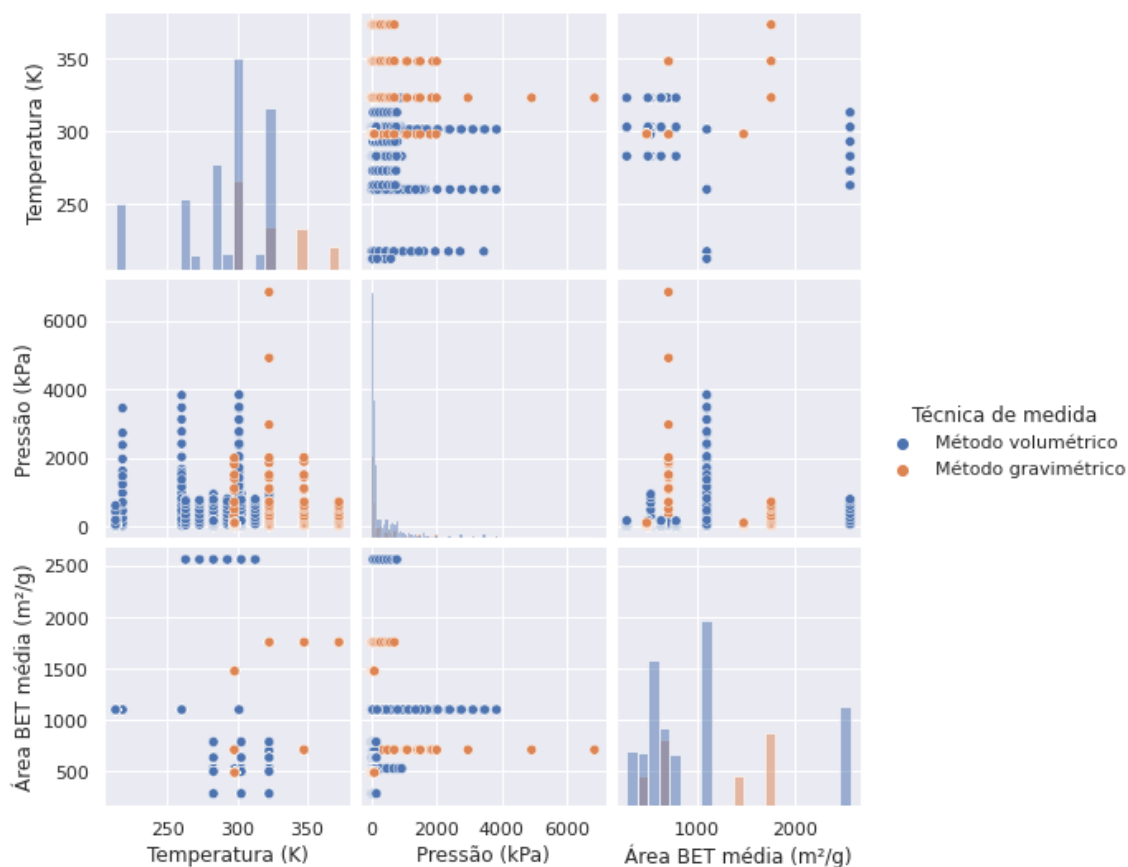


Figura 3: Gráfico de pares para as variáveis Temperatura (K), Pressão (kPa) e Área BET média (m^2/g) contidas no banco de dados analisado tendo por discriminante a técnica de medida empregada no experimento de adsorção.

Tabela 4: Análise descritiva das variáveis do banco de dados.

Variável	Média	Desvio padrão	Mínimo	Máximo
Área BET média (m^2/g)	1041,34	671,64	284,31	2560,00
Temperatura (K)	297,68	33,48	212,70	373,15
Pressão (kPa)	321,60	587,05	0,00	6842,00
Diâmetro cinético (A)	3,805	0,419	3,300	4,443
Polarizabilidade (cm^3)	3,14E-24	1,02E-24	1,74E-24	4,46E-24
Massa molar do adsorvato (g/mol)	28,08	11,60	14,01	44,01
Capacidade de adsorção (mol/kg)	2,308	2,166	0,000	10,640

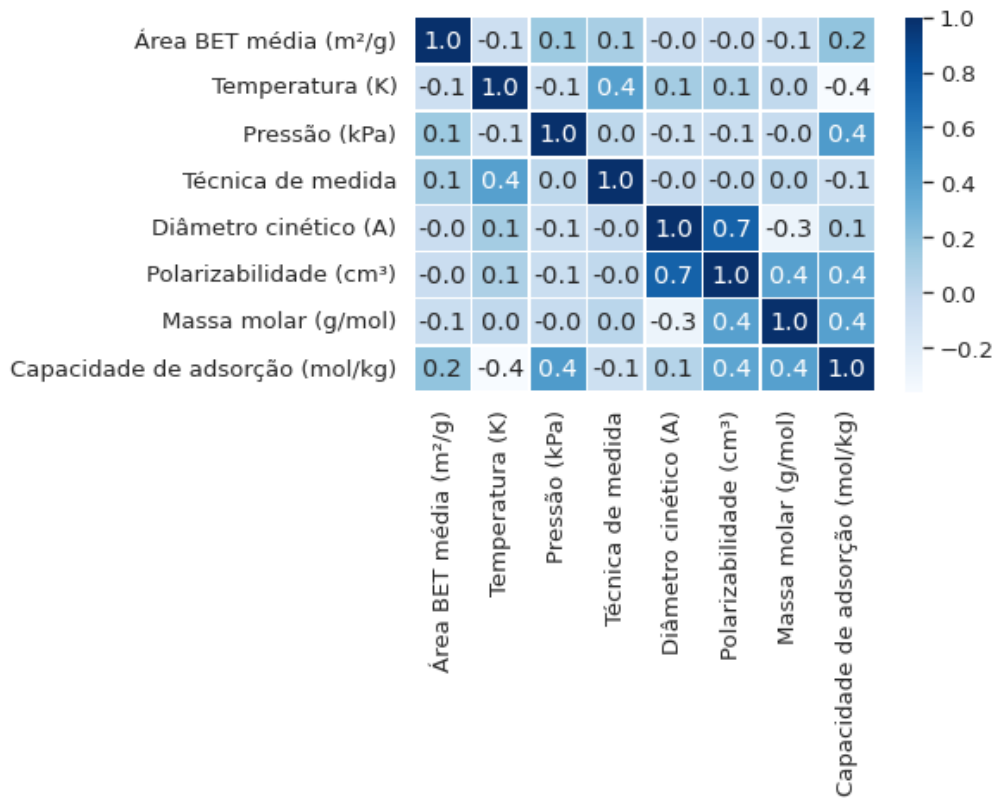


Figura 4: Matriz de correlação de Pearson para todas as variáveis incluídas no banco de dados.

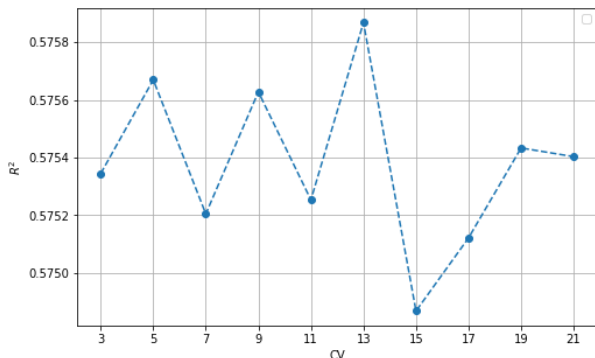


Figura 5: Variação do coeficiente de determinação (R^2) do modelo SVR com kernel linear ao variar o número de k -folds

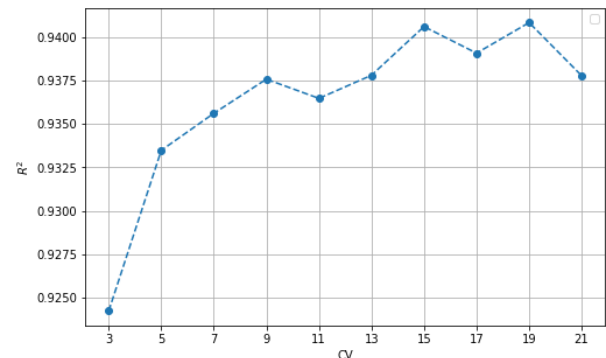


Figura 6: Variação do coeficiente de determinação (R^2) do modelo SVR com kernel RBF ao variar o número de k -folds

Tabela 5: (Hiper)parâmetros ótimos para os *kerneis* estudados.

(Hiper)parâmetros	Linear	Polinomial de grau 3	RBF	Sigmoid
C	12,75	124,56	20,95	23,52
ϵ	0,436	0,026	0,013	0,629
γ	-	0,167	0,802	0,001
Coef0	-	0,002	-	0,741
Tol	0,083	0,026	0,093	0,029

Uma vez que os modelos SVR foram desenvolvidos e seus hiperparâmetros otimizados, eles foram avaliados por sua precisão de previsão no conjunto de dados

tanto de treinamento quanto de teste. As métricas utilizadas para essa avaliação foram o coeficiente de determinação (R^2), o Erro Quadrático Médio (MAE - do inglês, *Mean*

Absolute Error) e a Raiz do Erro Quadrático Médio (RMSE - do inglês, *Root Mean Squared Error*) que são apresentadas na

Tabela 6. Por meio desses dados, pode-se observar que entre os *kerneis* avaliados para previsão da capacidade de adsorção de gases leves, o RBF foi o que apresentou melhor performance, seguido do polinomial de grau 3. Em diversos trabalhos, a função *kernel* RBF mostrou melhor desempenho quando comparada a outros (Kooh *et al.*, 2022; Ramedani *et al.*, 2014; Rohmah *et al.*, 2021), além disso Ramedani *et al.* (2014) afirmam que comparada a outras funções de *kernel*, a função RBF é compactada e suportada, tornando-a adequada para restringir o processo

de treinamento computacional e melhorar a eficiência de generalização. Os mesmos autores afirmam ainda que o RBF supera o modelo polinomial em termos de precisão da previsão, corroborando com o resultado aqui encontrado.

Salienta-se ainda, pelos resultados apresentados na Tabela 6, que o modelo SVR com *kernel* RBF desempenhou de modo satisfatório tanto na etapa de treinamento quanto de teste, não apresentando desta forma sub ou sobreajuste. O uso do *kernel* polinomial, apenas do ajuste satisfatório para o treinamento, apresentou sobreajuste.

Tabela 6. Desempenho dos modelos nas etapas de treino e teste.

Kernel	R ² de treino	R ² de teste	MAE de treino	MAE de teste	RMSE de treino	RMSE de teste
Linear	0,551	0,503	2,112	2,270	1,453	1,507
Polinomial	0,883	0,809	0,552	0,871	0,743	0,933
RBF	0,959	0,942	0,192	0,262	0,438	0,512
Sigmoid	0,555	0,507	2,094	2,252	1,447	1,501

Análise de importâncias

Para o modelo de melhor performance, SVR com *kernel* RBF, a análise de importância por permutação foi utilizada para determinar os efeitos do conjunto de variáveis de entrada na variável alvo que é a capacidade de adsorção. A Figura 7 apresenta a influência de cada atributo na variável alvo para o *kernel* RBF. As variáveis de maior importância ao utilizar o *kernel* RBF foram Pressão, Área BET média e Temperatura. Destaca-se ainda que a Técnica de Medida foi a variável com menor influência na capacidade de adsorção para esse modelo, indicando que existe baixo viés na forma como os dados foram obtidos experimentalmente.

Pressão, Área BET média e Temperatura média aparecerem entre os mais influentes do modelo, são reflexo do peneiramento molecular que ocorre nos sólidos microporosos, para os quais a difusão de Knudsen e de superfície são fenômenos dominantes (Gilron; Soffer, 2002).

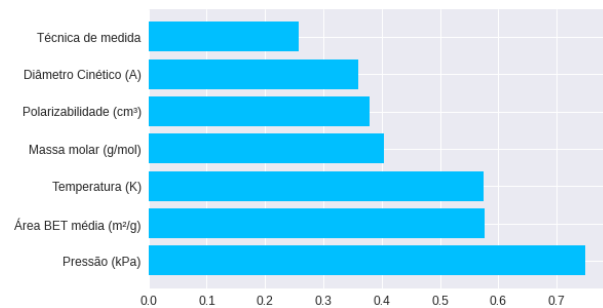


Figura 7: Análise de importâncias para a regressão aplicando-se o modelo SVR com *kernel* RBF.

Apesar de auxiliarem no sucesso da capacidade preditiva do modelo, os parâmetros Massa Molar, Polarizabilidade e Diâmetro Cinético se apresentaram como de menor importância. Então, visando-se apresentar o efeito marginal das variáveis no valor predito do modelo de aprendizado de máquina, a Figura 8 apresenta o gráfico de dependência parcial (PDP) das variáveis sobre a variável alvo. Por conveniência, excluiu-se a Técnica de Medida, variável menos importante.

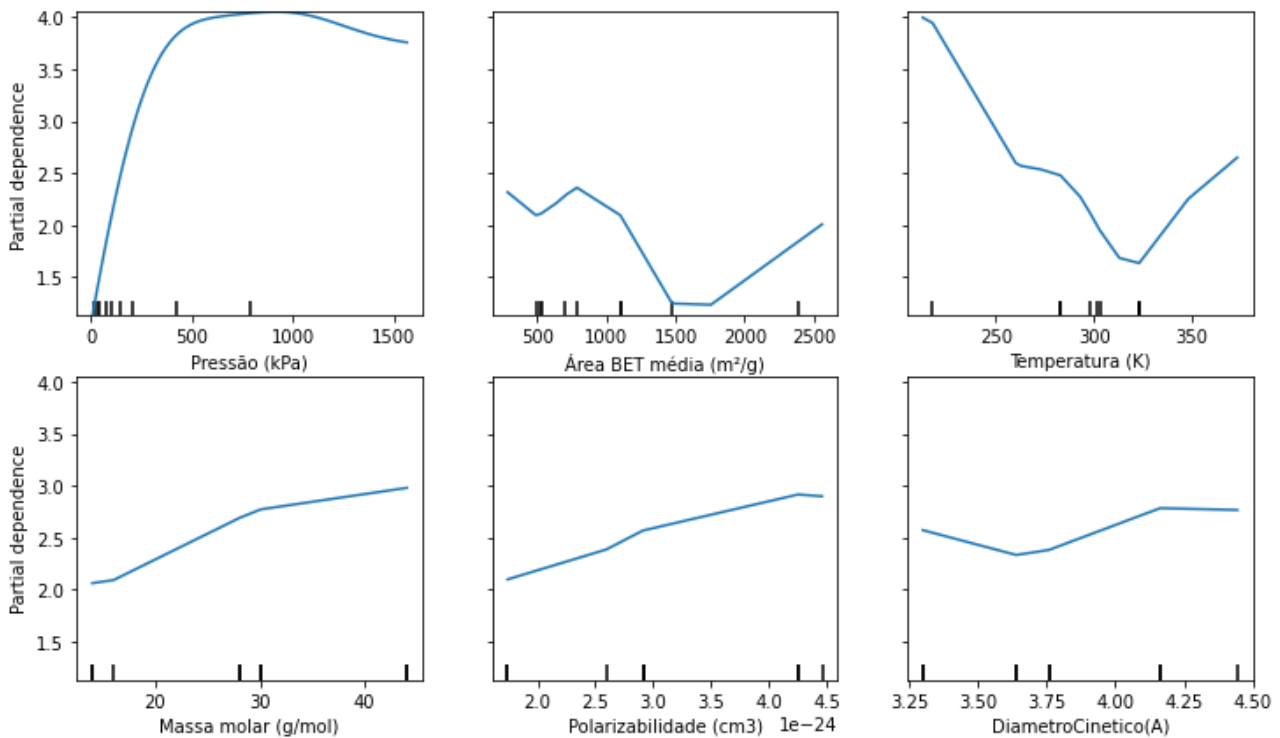


Figura 8: Gráficos de dependência parcial (PDP) para análise de como as variáveis afetam individualmente a capacidade de adsorção de gases nos sólidos estudados no banco de dados.

A Figura 8 mostra que a dependência parcial da variável Capacidade de Adsorção (alvo) mantém-se crescendo com o aumento da Massa Molar, da Polarizabilidade e do Diâmetro cinético. Contudo, apesar de crescente, o aumento de uma variável com relação à outra é suave. O efeito mais expressivo sobre o aumento da Capacidade de Adsorção ocorre com o aumento da Pressão e com a diminuição da Temperatura. Efeitos esperados devido ao fato das moléculas de adsorbato terem facilidade de acesso aos poros com o aumento de pressão e do fenômeno ser exotérmico. No caso da Área BET média, o PDP revela que a capacidade de adsorção pode se manter decrescente quando os valores de Área caem até valores em torno de 1600 m²/g.

CONCLUSÕES

Este trabalho envolveu o estudo capacidade adsorptiva no equilíbrio de gases leves usando diferentes tipos de zeólitas, carvões ativados e MOFs como adsorvente. Os dados experimentais coletados da literatura foram usados para treinar modelos com algoritmo SVR utilizando diferentes tipos de kernels: linear, polinomial de grau 3, RBF e sigmoide. De acordo com os resultados do

coeficiente de determinação nas etapas de treino e teste, o *kernel* RBF foi o que se mostrou com melhor desempenho. Desta forma, pode-se afirmar que o modelo SVR se mostrou promissor como ferramenta para desenvolvimento de adsorventes de alta capacidade por meio de uma avaliação prévia entre as condições de adsorção (pressão e temperatura), área superficial do adsorvente e propriedades físico-química dos gases (massa molar, diâmetro cinético e polarizabilidade), reduzindo assim a quantidade de materiais a serem analisados experimentalmente.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), Edital 001/2021 – Demanda Universal.

REFERÊNCIAS

ALIBAKSHI, A. (2018), Strategies to Develop Robust Neural Network Models:

- Prediction of Flash Point as a Case Study, *Analytica Chimica Acta*, v. 1026, p. 69–76.
- BORDONHOS, M.; LOURENÇO, M.; GOMES, J. R. B.; FERREIRA, P.; PINTO, M. L. (2021). Exploring Periodic Mesoporous Organosilicas for Ethane-Ethylene Adsorption-Separation. *Microporous and Mesoporous Materials*, v. 317, p. 110975.
- BREIMAN, L. (2001), *Random Forests*. *Machine Learning*, v. 45, n. 1, p. 5–32.
- CAO, L. (2022), Recent Advances in the Application of Machine-Learning Algorithms to Predict Adsorption Energies, *Trends in Chemistry*, v. 4, n. 4, p. 347–360
- CHARALAMBOUS, C.; SANTORI, G.; VILARRASA-GARCIA, E.; BASTOS-NETO, M.; CAVALCANTE JR., C. L., BRANDANI, S. (2018), Pure and Binary Adsorption of Carbon Dioxide and Nitrogen on AQSOA FAM Z02, *Journal of Chemical & Engineering Data*, v. 63, n. 3, p. 661–670.
- COSTA, E.; CALLEJA, G.; MARRON, C.; JIMENEZ, A.; PAU, J. (1989), Equilibrium Adsorption of Methane, Ethane, Ethylene, and Propylene and their Mixtures on Activated Carbon, *Journal of Chemical & Engineering Data*, v. 34, n. 2, p. 156–160.
- DANACI, D.; WEBLEY, P. A.; PETIT, C. (2021), Guidelines for Techno-Economic Analysis of Adsorption Processes, *Frontiers in Chemical Engineering*, v. 2, p. 20.
- DANNER, R.; CHOI, E. C. (1979), Mixture Adsorption Equilibria of Ethane and Ethylene on 13X Molecular Sieves, *Industrial & Engineering Chemistry Fundamentals*, v. 18, n. 3, p. 300–300.
- FLANIGEN, E. M. (2001), Zeolites and Molecular Sieves: An Historical Perspective, *Studies in Surface Science and Catalysis*, v. 58, p. 11–35.
- FREEMAN, J. J.; MCLEOD, A. I. (1983), Nitrogen BET Surface Area Measurement as a Fingerprint Method for the Estimation of Pore Volume in Active Carbons, *Fuel*, v. 62, n. 9, p. 1090–1091.
- GÉRON, A. (2019), *Mãos à Obra: Aprendizado de Máquina com Scikit-Learn & TensorFlow*, 1. Ed, Alta Books, São Paulo.
- GILRON, J.; SOFFER, A. (2002), Knudsen Diffusion in Microporous Carbon Membranes with Molecular Sieving Character, *Journal of Membrane Science*, v. 209, n. 2, p. 339–352.
- GOLIPOUR, H.; MOKHTARANI, B.; MAFI, M.; MORADI, A.; GODINI, H. R. (2020), Experimental Measurement for Adsorption of Ethylene and Ethane Gases on Copper-Exchanged Zeolites 13X and 5A, *Journal of Chemical & Engineering Data*, v. 65, n. 8, p. 3920–3932.
- HUANG, M.; XU, H.; YU, H.; ZHANG, H.; MICHEAL, M.; YUAN, X.; WU, H. (2022), Fast Prediction of Methane Adsorption in Shale Nanopores Using Kinetic Theory and Machine Learning Algorithm, *Chemical Engineering Journal*, v. 446, p. 137221.
- KARIMI, S.; TAVAKKOLI YARAKI, M.; KARRI, R. R. (2019), A Comprehensive Review of the Adsorption Mechanisms and Factors Influencing the Adsorption Process from the Perspective of Bioethanol Dehydration, *Renewable and Sustainable Energy Reviews*, v. 107, p. 535–553.
- KHORAMZADEH, E.; MOFARAHI, M.; LEE, C. (2019), Equilibrium Adsorption Study of CO₂ and N₂ on Synthesized Zeolites 13X, 4A, 5A, and Beta, *Journal of Chemical & Engineering Data*, v. 64, n. 12, p. 5648–5664.
- KLOUTSE, F. A.; HOURRI, A.; NATARAJAN, S.; BENARD, P.; CHAHINE, R. (2018), Experimental Benchmark Data of CH₄, CO₂ and N₂ Binary and Ternary Mixtures Adsorption on MOF-5, *Separation and Purification Technology*, v. 197, p. 228–236.
- KOOH, M. R. R.; THOTAGAMUGE, R.; CHAU, Y. C.; MAHADI, A. H.; LIM, C. M. (2022), Machine Learning Approaches to Predict Adsorption Capacity of *Azolla Pinnata* in the Removal of Methylene Blue, *Journal of*

- the Taiwan Institute of Chemical Engineers, v. 132, p. 104134.
- LASHAKI, M. J.; FAYAZ, M.; NIKNADDAF, S.; HASHISHO, Z. (2012), Effect of the Adsorbate Kinetic Diameter on the Accuracy of the Dubinin-Radushkevich Equation for Modeling Adsorption of Organic Vapors on Activated Carbon. *Journal of Hazardous Materials*, v. 241–242, p. 154–163.
- LI, W.; SUHAYB, M. K.; THANGAVELU, L.; MARHOON, H. A.; PUSTOKHINA, I.; ALQSAIR, U. F.; EL-SHAFAY, A. S.; ALASHWALH, M. (2022), Implementation of AdaBoost and Genetic Algorithm Machine Learning Models in Prediction of Adsorption Capacity of Nanocomposite Materials, *Journal of Molecular Liquids*, v. 350, p. 118527.
- MAJD, M. M.; KORDZADEH-KERMANI, V.; GHALANDARI, V.; ASKARI, A.; SILLANPÄÄ, M. (2022), Adsorption Isotherm Models: A Comprehensive and Systematic Review (2010–2020), *Science of The Total Environment*, v. 812, p. 151334.
- MARTINS, V. F. D.; RIBEIRO, A. M.; FERREIRA, A.; LEE, U.; HWANG, Y. K.; CHANG, J.; LOUREIRO, J. M.; RODRIGUES, A. E. (2015), Ethane/Ethylene Separation on a Copper Benzene-1,3,5-tricarboxylate MOF, *Separation and Purification Technology*, v. 149, p. 445–456.
- MEEK, S. T.; TEICH-MCGOLDRICK, S. L.; PERRY IV, J. J.; GREATHOUSE, J. A.; ALLENDORF, M. D. (2012), Effects of Polarizability on the Adsorption of Noble Gases at Low Pressures in Monohalogenated Isorecticular Metal–Organic Frameworks, *The Journal of Physical Chemistry C*, v. 116, n. 37, p. 19765–19772.
- MENG, M.; ZHONG, R.; WEI, Z. (2020), Prediction of Methane Adsorption in Shale: Classical Models and Machine Learning Based Models. *Fuel*, v. 278, p. 118358.
- AMAR, M. N.; LARESTANI, A.; LV, Q.; ZHOU, T. HEMMATI-SARAPARDEH, A. (2020), Modeling of Methane Adsorption Capacity in Shale Gas Formations Using White-Box Supervised Machine Learning Techniques, *Journal of Petroleum Science and Engineering*, v. 208, p. 109226.
- PRAVIN, P. S.; TAN, J. Z. M.; YAP, K. S.; WU, Z. (2022), Hyperparameter Optimization Strategies for Machine Learning-Based Stochastic Energy Efficient Scheduling in Cyber-Physical Production Systems, *Digital Chemical Engineering*, p. 100047.
- PULLUMBI, P.; BRANDANI, F.; BRANDANI, S. (2019), Gas Separation by Adsorption: Technological Drivers and Opportunities for Improvement, *Current Opinion in Chemical Engineering*, v. 24, p. 131–142.
- RAJI, M.; DASHTI, A.; ALIVAND, M. S.; ASGHARI, M. (2022), Novel Prosperous Computational Estimations for Greenhouse Gas Adsorptive Control by Zeolites Using Machine Learning Methods, *Journal of Environmental Management*, v. 307, p. 114478
- RAMEDANI, Z.; OMID, M.; KEYHANI, A.; SHAMSHIRBAND, S.; KHOSHNEVISAN, B. (2014), Potential of Radial Basis Function Based Support Vector Regression for Global Solar Radiation Prediction, *Renewable and Sustainable Energy Reviews*, v. 39, 1005-1011.
- REICH, R. (1974), Adsorption on Activated Carbon of Methane, Ethane, and Ethylene Gases and Their Mixtures and Carbon Dioxide at 212 K, 260 K, and 301 K and up to Thirty-five Atmospheres, Thesis-Atlanta, GA: Georgia Institute of Technology.
- ROHMAH, M. F.; PUTRA, I. K. G. D.; HARTATI, R. S.; ARDIANTORO, L. (2021), Comparison Four Kernels of SVR to Predict Consumer Price Index, *Journal of Physics: Conference Series*. Anais IOP Publishing Ltd.
- SCHÄF, O.; TORTET, L.; SIMON-MASSERON, A.; PATARIN, J.; DEFOUR, S.; BLANC, R.; COSTE, C.; ZEREGA, Y. (2020), Importance of

- PCDD/F Molecules' Polarizability and Steric Hindrance on their Adsorption onto Zeolites in a Standard EN1948-1 Sampling Device for Incinerator Emission Monitoring, *Chemosphere*, v. 259, p. 127457.
- SPEIGHT, J. G. (2019), Hydrogen Production. In: *Heavy Oil Recovery and Upgrading*, Elsevier, p. 657–697.
- TANG, H.; XU, Q.; WANG, M.; JIANG, J. (2021), Rapid Screening of Metal-Organic Frameworks for Propane/Propylene Separation by Synergizing Molecular Simulation and Machine Learning, *ACS Applied Materials & Interfaces*, v. 13, n. 45, p. 53454–53467.
- THARWAT, A. (2019), Parameter Investigation of Support Vector Machine Classifier with Kernel Functions, *Knowledge and Information Systems*, v. 61, n. 3, p. 1269–1302.
- WANG, Z.; ZHOU, Y.; ZHOU, T.; SUNDMACHER, K. (2022), Identification of Optimal Metal-Organic Frameworks by Machine Learning: Structure Decomposition, Feature Integration, and Predictive Modeling, *Computers & Chemical Engineering*, v. 160, p. 107739.
- XIE, J.; ZHANG, L. (2022), Machine Learning and Symbolic Regression for Adsorption of Atmospheric Molecules on Low-Dimensional TiO₂. *Applied Surface Science*, v. 597, p. 153728.
- YAN, Y.; ZHANG, L.; LI, S.; LIANG, H.; QIAO, Z. (2021), Adsorption Behavior of Metal-Organic Frameworks: From Single Simulation, High-Throughput Computational Screening to Machine Learning, *Computational Materials Science*, v. 193, p. 110383, jun. 2021.
- YAN, Y.; SHI, Z.; LI, H.; LI, L.; YANG, X.; LI, S.; LIANG, H.; QIAO, Z. (2022), Machine Learning and in-Silico Screening of Metal-Organic Frameworks for O₂/N₂ Dynamic Adsorption and Separation, *Chemical Engineering Journal*, v. 427, p. 131604.
- YUAN, Q.; LONGO, M.; THORNTON, A. W.; MCKEOWN, N. B.; COMESAÑA-GÁNDARA, B.; JANSEN, J. C.; JELFS, K. E. (2021a), Imputation of Missing Gas Permeability Data for Polymer Membranes Using Machine Learning, *Journal of Membrane Science*, v. 627, p. 119207.
- YUAN, X.; SUVARNA, M.; LOW, S.; DISSANAYAKE, P. D.; LEE, K. B.; LI, J.; WANG, X.; OK, Y. S. (2021b), Applied Machine Learning for Prediction of CO₂ Adsorption on Biomass Waste-Derived Porous Carbons, *Environmental Science and Technology*, v. 55, n. 17, p. 11925–11936.
- ZHANG, X.; LIU, C.-A. (2022), Model Averaging Prediction by K-fold Cross-Validation, *Journal of Econometrics*. In Press.
- ZHANG, X.; ZHENG, Q.; HE, H. (2022), Machine-Learning-Based Prediction of Hydrogen Adsorption Capacity at Varied Temperatures and Pressures for MOFs Adsorbents, *Journal of the Taiwan Institute of Chemical Engineers*, v. 138, p. 104479.
- ZHANG, Y.; SU, W., SUN, Y.; LIU, J.; LIU, X.; WANG, X. (2015), Adsorption Equilibrium of N₂, CH₄, and CO₂ on MIL-101, *Journal of Chemical & Engineering Data*, v. 60, n. 10, p. 2951–2957.