



**INSTITUTO LATINO-AMERICANO DE TECNOLOGIA,
INFRAESTRUTURA E TERRITORIO (ILATIT)
ENGENHARIA QUIMICA**

**ANÁLISE DO DESEMPENHO DE ALGORITMOS DE APRENDIZAGEM DE
MÁQUINA NA PREDIÇÃO DO RENDIMENTO DE BIODIESEL EM REAÇÕES DE
TRANSESTERIFICAÇÃO POR CATÁLISE HETEROGÊNEA**

ABEL DARIO GONZALEZ LEZCANO

Foz do Iguaçu
2024



**INSTITUTO LATINO-AMERICANO DE TECNOLOGIA,
INFRAESTRUTURA E TERRITORIO (ILATIT)
ENGENHARIA QUIMICA**

**ANÁLISE DO DESEMPENHO DE ALGORITMOS DE APRENDIZAGEM DE
MÁQUINA NA PREDIÇÃO DO RENDIMENTO DE BIODIESEL EM REAÇÕES DE
TRANSESTERIFICAÇÃO POR CATÁLISE HETEROGÊNEA**

ABEL DARIO GONZALEZ LEZCANO

Trabalho de Conclusão de Curso apresentado ao Instituto Latino-Americano de Tecnologia, Infraestrutura e Território da Universidade Federal da Integração Latino-Americana, como requisito parcial à obtenção do título de Bacharel em Engenharia Química.

Orientadora: Prof.^a. Dr^a. Andreia Cristina Furtado

Coorientador: Prof. Dr. Luís Antonio Lourenço

Foz do Iguaçu

2024

ABEL DARIO GONZALEZ LEZCANO

ANÁLISE DO DESEMPENHO DE ALGORITMOS DE APRENDIZAGEM DE MÁQUINA NA PREDIÇÃO DO RENDIMENTO DE BIODIESEL EM REAÇÕES DE TRANSESTERIFICAÇÃO POR CATÁLISE HETEROGÊNEA

Trabalho de Conclusão de Curso apresentado ao Instituto Latino-Americano de Tecnologia, Infraestrutura e Território da Universidade Federal da Integração Latino-Americana, como requisito parcial à obtenção do título de Bacharel em Engenharia Química.

Orientadora: Prof.^a. Dr.^a. Andreia Cristina Furtado

Coorientador: Prof. Dr. Luís Antonio Lourenço

BANCA EXAMINADORA

Andreia C. Furtado

Orientadora: Prof.^a. Dra. Andreia Cristina Furtado

UNILA



Documento assinado digitalmente

LUIS ANTONIO LOURENCO

Data: 03/06/2024 07:24:40-0300

CPF: ***.233.099-**

Verifique as assinaturas em <https://v.ufsc.br>

Coorientador: Prof. Dr. Luís Antonio Lourenço

Documento assinado digitalmente



MARLEI ROLING SCARIOT

Data: 31/05/2024 13:12:50-0300

Verifique em <https://validar.iti.gov.br>

Prof.^a. Dra. Marlei Roling Scariot

UNILA



Documento assinado digitalmente

CESAR ADOLFO RODRIGUEZ SOTOMONTE

Data: 27/05/2024 13:45:29-0300

Verifique em <https://validar.iti.gov.br>

Prof. Dr. César Adolfo Rodriguez Sotomonte

UNILA

Foz do Iguaçu, 23 de abril de 2024

AGRADECIMENTOS

Em primeiro lugar, gostaria de expressar minha profunda gratidão à minha mãe Beatriz pelo apoio constante em todos os momentos difíceis ao longo desse caminho acadêmico. Uma mulher batalhadora, ela nos criou com grande força e seu incentivo foi essencial para que eu pudesse alcançar meus objetivos. Estou verdadeiramente agradecido pelos conselhos que ela me deu, e sua presença foi a motivação que me impulsionou em momentos difíceis.

Aos meus irmãos, Griselda, Javier, Mirco, Karen e Fátima, devo um agradecimento especial. Sua presença ao longo da minha vida foi uma fonte constante de apoio e inspiração. Desde os primeiros passos da minha educação até este momento, vocês estiveram ao meu lado, compartilhando alegrias, desafios e conquistas. Obrigado por sempre acreditarem em mim.

Em memória do meu pai, Arcadio, que infelizmente nos deixou, um homem batalhador desde sempre, seus conselhos quando era criança moldaram profundamente quem sou hoje. Mesmo não estando mais fisicamente ao meu lado, lembro todos os dias das suas palavras de sabedoria que continuam na minha mente, inspirando-me a superar obstáculos e perseguir meus objetivos com dedicação. Sua influência é eterna, e sua memória será sempre uma fonte de força e inspiração para mim.

Aos meus professores orientadores, Dra. Andreia Cristina Furtado e Dr. Luís Antonio Lourenço, gostaria de expressar minha profunda gratidão pela oportunidade de participar deste projeto, que não apenas me proporcionou um imenso crescimento pessoal, mas também pelo conhecimento que compartilharam ao longo do curso. Aprendi muito com vocês, e esse aprendizado foi um grande incentivo para o início deste projeto. Agradeço sinceramente pela orientação brindada ao longo deste trabalho e pelo acompanhamento em diversas reuniões, que foram imprescindíveis ao longo desta jornada.

Aos professores da banca, Dr. César Adolfo Rodriguez Sotomonte e Dra. Marlei Roling Scariot, gostaria de expressar minha sincera gratidão por aceitarem o convite para participar como avaliadores. Admiro profundamente o trabalho de vocês

e reconheço a importância da sua presença na banca avaliadora, que contribuiu significativamente para aprimorar meu trabalho.

Não posso deixar de mencionar meus amigos e colegas de curso, Camila Di Rienzo, Matheus Müller e Sara Adrissa Barros, que são pessoas incríveis e se tornaram parte essencial dessa longa jornada. Suas contribuições, apoio e amizade foram importantes ao longo deste caminho acadêmico. Também gostaria de expressar minha profunda gratidão à UNILA por me proporcionar uma oportunidade única de ter uma vaga em uma Universidade Federal com um ensino de qualidade. O apoio e os recursos oferecidos pela instituição foram fundamentais para o meu desenvolvimento acadêmico e pessoal.

Gostaria de expressar minha profunda gratidão à Fundação do Parque Tecnológico da Itaipu por ter sido um apoio fundamental ao longo de todos esses anos de estudo. Agradeço sinceramente pelo incentivo econômico e pelo acompanhamento durante minha formação. O suporte recebido pelo convênio UNILA-AEP COHORTE 2018 foi essencial para o meu crescimento acadêmico e profissional.

Agradeço sinceramente a todos que, de alguma forma, contribuíram para esta conquista!

RESUMO

O crescente aumento dos preços dos combustíveis, as mudanças climáticas e a poluição têm colocado o mundo diante de desafios significativos, evidenciando a escassez de recursos energéticos e seu impacto no desenvolvimento econômico global. Neste cenário, a busca por fontes de energia sustentáveis e renováveis, como o biodiesel, torna-se crucial para mitigar a crise energética e reduzir a dependência dos combustíveis fósseis. A reação de transesterificação é a via mais utilizada para a produção de biodiesel e o rendimento obtido depende vários parâmetros, como a temperatura, tempo de reação, concentração do catalisador e razão estequiométrica entre o álcool e o triglicérido. Os catalisadores heterogêneos são amplamente empregados devido à sua eficiência e capacidade de reutilização. Os métodos convencionais para avaliar estes parâmetros apresentam certas limitações e neste contexto os algoritmos de aprendizagem de máquina mostram-se como uma alternativa. Estes algoritmos realizam previsões a partir de dados fornecidos ao sistema, que podem explicar comportamentos que orientam na tomada de decisões, dessa forma o objetivo deste estudo é utilizar modelos de aprendizagem de máquina para prever o rendimento do biodiesel por transesterificação via catálise heterogênea. Os dados foram obtidos por meio de uma revisão bibliográfica, onde a escolha dos artigos foi restringida ao uso de metanol, óxido de cálcio como catalisador e os parâmetros comumente utilizados. Logo, os dados foram submetidos a análises estatísticas e utilizados para treinar três modelos de aprendizagem de máquina, com o modelo mais preciso foi realizado uma validação comparando a previsão do algoritmo com dois resultados experimentais, além disso com este algoritmo foram realizadas três análises. Os modelos treinados e avaliados foram Regressão Linear, Floresta Aleatória e *Multi-Layer Perceptron (MLP) Regressor*, utilizando o método de *hold-out validation*. O modelo Floresta Aleatória foi o que apresentou o melhor desempenho na previsão do rendimento do biodiesel em relação aos outros modelos, atingindo um valor do erro médio absoluto igual a 6,59 e um R^2 de 0,63. A validação confirmou a precisão dos modelos com erros percentuais inferiores a 10%, e durante as análises dos parâmetros do biodiesel o modelo demonstrou um comportamento típico de acordo com a literatura. Este estudo contribui para o avanço da produção de biodiesel ao fornecer informações sobre a previsão do seu rendimento. É recomendado para pesquisas futuras a utilização de mais dados incluindo outros parâmetros como uma tentativa de melhorar o desempenho do modelo.

Palavras-chave: biodiesel; catálise heterogênea; inteligência artificial; Floresta Aleatória; MLP Regressor.

RESUMEN

El creciente aumento de los precios de los combustibles, los cambios climáticos y la contaminación han puesto al mundo ante desafíos significativos, evidenciando la escasez de recursos energéticos y su impacto en el desarrollo económico global. En este escenario, la búsqueda de fuentes de energía sostenibles y renovables, como el biodiésel, se vuelve crucial para mitigar la crisis energética y reducir la dependencia de los combustibles fósiles. La reacción de transesterificación es la vía más utilizada para la producción de biodiésel y el rendimiento obtenido depende de varios parámetros, como la temperatura, tiempo de reacción, concentración del catalizador y la relación estequiométrica entre el alcohol y el triglicérido. Los catalizadores heterogéneos son ampliamente empleados debido a su eficiencia y capacidad de reutilización. Los métodos convencionales para evaluar estos parámetros presentan ciertas limitaciones y en este contexto los algoritmos de aprendizaje automático se presentan como una alternativa. Estos algoritmos realizan predicciones a partir de datos proporcionados al sistema, que pueden explicar comportamientos que orientan en la toma de decisiones, por lo tanto, el objetivo de este estudio es utilizar modelos de aprendizaje automático para predecir el rendimiento del biodiésel por transesterificación vía catálisis heterogénea. Los datos fueron obtenidos mediante una revisión bibliográfica, donde la elección de los artículos se restringió al uso de metanol, óxido de calcio como catalizador y los parámetros comúnmente utilizados. Posteriormente, los datos fueron sometidos a análisis estadísticos y utilizados para entrenar tres modelos de aprendizaje automático, con el modelo más preciso se realizó una validación comparando la predicción del algoritmo con dos resultados experimentales, además, con este algoritmo se realizaron tres análisis. Los modelos entrenados y evaluados fueron Regresión Lineal, Bosque Aleatorio y *Multi-Layer Perceptron (MLP) Regresor*, utilizando el método de *hold-out validation*. El modelo de Bosque Aleatorio fue el que mostró el mejor desempeño en la predicción del rendimiento del biodiésel en relación con los otros modelos, alcanzando un valor del error medio absoluto igual a 6,59 y un R^2 de 0,63. La validación confirmó la precisión de los modelos con errores porcentuales inferiores al 10%, y durante los análisis de los parámetros del biodiésel, el modelo demostró un comportamiento típico de acuerdo con la literatura. Este estudio contribuye al avance de la producción de biodiésel al proporcionar información sobre la predicción de su rendimiento. Se recomienda para futuras investigaciones la utilización de más datos, incluyendo otros parámetros, como un intento de mejorar el rendimiento del modelo.

Palabras clave: biodiésel; catálisis heterogénea; inteligencia artificial; Bosque Aleatorio; MLP Regresor.

ABSTRACT

The increasing rise in fuel prices, climate change, and pollution have presented the world with significant challenges, highlighting the scarcity of energy resources and their impact on global economic development. In this scenario, the search for sustainable and renewable energy sources, such as biodiesel, becomes crucial to mitigate the energy crisis and reduce dependence on fossil fuels. Transesterification reaction is the most commonly used pathway for biodiesel production, and the yield obtained depends on various parameters such as temperature, reaction time, catalyst concentration, and the stoichiometric ratio between alcohol and triglyceride. Heterogeneous catalysts are widely employed due to their efficiency and reusability. Conventional methods for evaluating these parameters have certain limitations, and in this context, machine learning algorithms emerge as an alternative. These algorithms make predictions based on data provided to the system, which can explain behaviors guiding decision-making. Therefore, the objective of this study is to utilize machine learning models to predict biodiesel yield through heterogeneous catalysis transesterification. Data were obtained through a literature review, where the selection of articles was restricted to the use of methanol, calcium oxide as a catalyst, and commonly used parameters. Subsequently, the data underwent statistical analysis and were used to train three machine learning models. The most accurate model underwent validation by comparing the algorithm's prediction with two experimental results, and three analyses were conducted using this algorithm. The trained and evaluated models were Linear Regression, Random Forest, and Multi-Layer Perceptron (MLP) Regressor, using the hold-out validation method. The Random Forest model showed the best performance in predicting biodiesel yield compared to other models, achieving a mean absolute error value of 6.59 and an R2 of 0.63. Validation confirmed the accuracy of the models with percentage errors below 10%, and during the analysis of biodiesel parameters, the model exhibited a typical behavior according to the literature. This study contributes to the advancement of biodiesel production by providing information on yield prediction. It is recommended for future research to use more data, including other parameters, in an attempt to improve the model's performance.

Keywords: biodiesel; heterogeneous catalysis; artificial intelligence; Random Forest; MLP Regressor.

LISTA DE ILUSTRAÇÕES

Figura 1 – Os produtos refinados globais e suas perspectivas de 2017 a 2040	21
Figura 2 – Reação de transesterificação na produção de biodiesel	24
Figura 3 – Reação de esterificação na produção de biodiesel	24
Figura 4 – Processo da produção de Biodiesel por transesterificação	25
Figura 5 – Classificação dos catalisadores.....	26
Figura 6 – Aumento de publicações utilizando AM na pesquisa de Biodiesel	33
Figura 7 – Classificação de tipos de aprendizagem de máquina.....	34
Figura 8 – Estrutura de uma Árvores de Decisão.....	38
Figura 9 – Estrutura do modelo de Floresta Aleatória	39
Figura 10 – Neurônios biológicos transmitindo informações	40
Figura 11 – Estrutura de uma Rede Neural Artificial.....	41
Figura 12 – Método de otimização de hiperparâmetros por Pesquisa Aleatória (<i>Random Search</i>).....	44
Figura 13 – Construção de um Pipeline de Aprendizagem de Máquina	44
Figura 14 – Processo de revisão sistemática para a obtenção de dados.....	48
Figura 15 – Transformação dos dados em <i>dataframe</i> e bibliotecas utilizadas	50
Figura 16 – Etapas da análise exploratória de dados.....	50
Figura 17 – Fontes de óleos encontrados nos dados coletados.....	59
Figura 18 – Histogramas e diagramas de dispersão das variáveis numéricas.....	60
Figura 19 – Boxplots das variáveis numéricas	61
Figura 20 – Boxplots das variáveis numéricas após as transformações	63
Figura 21 – Mapa de calor e correlações de Pearson antes e depois das transformações.....	63
Figura 22 – Comparação do desempenho dos modelos antes e depois das transformações.....	65
Figura 23 – Comparação do desempenho da Floresta Aleatória utilizando hiperparâmetros padrão, com os modelos ajustados.....	66
Figura 24 – Comparação do desempenho do MLP Regressor utilizando hiperparâmetros padrão, com os modelos ajustados.....	67
Figura 25 – Gráfico de dispersão dos valores de rendimentos reais vs. preditos.....	68
Figura 26 – Curvas do rendimento de biodiesel a diferentes temperaturas	71
Figura 27 – Curvas do rendimento de biodiesel a diferentes pesos de catalisador ..	72
Figura 28 – Curvas do rendimento de biodiesel a diferentes razões estequiométricas de metanol e óleo.....	73

LISTA DE TABELAS

Tabela 1 – Vantagens e desvantagens da utilização do Biodiesel	22
Tabela 2 – Dados com label, m features, e n dados	45
Tabela 3 – Modelo da tabela com dados genéricos para o banco de dados	49
Tabela 4 – Grades de busca utilizadas para a Floresta Aleatória	52
Tabela 5 – Grades de busca utilizadas para a MLP Regressor	53
Tabela 6 – Condições de reação utilizadas para a análise de parâmetros	55
Tabela 7 – Artigos acadêmicos selecionados para a coleta de dados	57
Tabela 8 – Estatística descritiva dos dados	58
Tabela 9 – Assimetria e curtose dos dados após as transformações	63
Tabela 10 – Comparação dos modelos ajustados utilizando diferentes métricas	67
Tabela 11 – Validação do modelo com valores experimentais da literatura	70

LISTA DE ABREVIATURAS E SIGLAS

FFAs	Ácidos Graxos Livres
AM	Aprendizagem De Máquina
R ²	Coefficiente De Determinação
MAE	Erro Absoluto Médio
MAPE	Erro Percentual Absoluto
RMSE	Erro Quadrático Médio
MPL Regressor	<i>Multi-Layer Perceptron Regressor</i>
RNA	Redes Neurais Artificiais
RLM	Regressão Linear Múltipla
RLS	Regressão Linear Simples
VSCoDe	Visual Studio Code

SUMÁRIO

1 INTRODUÇÃO	14
2 JUSTIFICATIVA	16
3 OBJETIVOS	18
3.1 OBJETIVO GERAL	18
3.2 OBJETIVOS ESPECIFICOS.....	18
4 FUNDAMENTAÇÃO TEÓRICA	19
4.1 BIOCOMBUSTÍVEIS COMO UMA ALTERNATIVA SUSTENTÁVEL	19
4.2 BIODIESEL	20
4.2.1 Gerações de biodiesel	22
4.2.2 Produção de biodiesel	23
4.2.3 Catalisadores	26
4.3 PARÂMETROS QUE INFLUENCIAM NO RENDIMENTO DA REAÇÃO DE TRANSESTERIFICAÇÃO	28
4.3.1 Temperatura.....	29
4.3.2 Tempo	29
4.3.3 Tipo e quantidade de álcool	30
4.3.4 Tipo e quantidade de catalisador	30
4.4 USO DA INTELIGÊNCIA ARTIFICIAL NA MODELAGEM DA PRODUÇÃO DE BIODIESEL	32
4.5 APRENDIZAGEM DE MÁQUINA.....	33
4.6 ALGORITMOS DE APRENDIZAGEM SUPERVISIONADA	35
4.6.1 Modelos Lineares.....	35
4.6.2 Modelos baseados em árvores	37
4.6.3 Modelos de Redes Neurais Artificiais.....	40
4.7 HIPERPARÂMETROS	42
4.8 PIPELINE DE APRENDIZAGEM DE MÁQUINA.....	44
4.8.1 Obtenção de dados e análise exploratória.....	45
4.8.2 Seleção e treinamento do modelo	46
4.8.3 Avaliação dos modelos	47
5 METODOLOGIA	48
5.1 OBTENÇÃO DOS DADOS	48
5.2 PRÉ-PROCESSAMENTO E ANÁLISE EXPLORATÓRIA DOS DADOS.....	49
5.3 TREINAMENTO DOS MODELOS DE APRENDIZAGEM DE MÁQUINA	51

5.4 AJUSTE E SELEÇÃO DOS MODELOS	52
5.5 VALIDAÇÃO E ANÁLISE DOS PARÂMETROS DA REAÇÃO DE TRANSESTERIFICAÇÃO	54
6 RESULTADOS E DISCUSSÃO	56
6.1 REVISÃO SISTEMÁTICA DE LITERATURA	56
6.2 PRÉ-PROCESSAMENTO DE DADOS E ANÁLISE EXPLORATÓRIA	57
6.3 TREINAMENTO DE MODELOS DE AM	64
6.4 AJUSTE DE HIPERPARÂMETROS E AVALIAÇÃO DOS MODELOS	65
6.5 VALIDAÇÃO E AVALIAÇÃO DOS PARÂMETROS DO BIODIESEL	69
7 CONSIDERAÇÕES FINAIS	75
REFERÊNCIAS BIBLIOGRAFICAS	76
APÊNDICE A – TABELA COM OS DADOS UTILIZADOS PARA O TREINAMENTO DOS MODELOS DE AM	85

1 INTRODUÇÃO

Após a segunda revolução industrial, houve um aumento significativo no uso do petróleo como recurso para a produção de combustíveis, impactando no avanço e no desenvolvimento de novas tecnologias. Entretanto, este aumento no consumo levanta questões sobre a disponibilidade e problemas ambientais devido à queima de combustíveis fósseis. Diante disso, os recursos sustentáveis e energias renováveis, como a biomassa, são alternativas mais adequadas para reduzir as preocupações associadas à crise energética e aos combustíveis fósseis (AL-MUHTASEB et al., 2018).

Anwar et al. (2019) afirmam que o biodiesel é uma alternativa renovável que está se tornando popular devido à sua natureza ecologicamente correta. Entre as opções de matéria-prima para a produção do biocombustível, estão óleo de graxa, óleo de cozinha não processado, óleo animal e óleos derivados de resíduos de culturas não comestíveis (YAASHIKAA; KUMAR; KARISHMA, 2022). O biodiesel apresenta vantagens não apenas em termos de sustentabilidade. Em relação aos motores, este combustível proporciona uma queima mais limpa e uma boa lubrificidade, o que pode prolongar a vida útil do motor. Além disso, destaca-se que o biodiesel possui um poder calorífico semelhante ao do óleo diesel (ROSCHAT et al., 2022).

Existem diferentes métodos para a produção do biodiesel, no entanto, o mais utilizado é o processo de transesterificação, que não só proporciona um alto rendimento, mas também é um dos métodos mais econômicos. Este procedimento se resume a uma reação entre um glicerídeo e um álcool na presença de um catalisador, formando ésteres alquílicos de ácidos graxos, além de um álcool como subproduto (KANWAR GAUR; GOYAL, 2022).

Conforme com Vicente, Martínez e Aracil (2004), o rendimento da reação depende de vários parâmetros, como temperatura, concentração inicial do catalisador e a razão molar entre o álcool e o triglicerídeo. Desta forma, a produtividade do biodiesel é variável e, para otimizar o processo, são empregados os catalisadores heterogêneos, devido à sua eficiência e à facilidade de separação

da mistura de reação, além da capacidade de reutilização, tornando o processo mais econômico (MAHMOUD; EL-MOLLA; IBRAHIM, 2020).

Na literatura, há uma grande variedade de pesquisas avaliando o rendimento da produção de biodiesel, utilizando uma diversidade de matérias-primas e diferentes condições operacionais. Esses dados podem ser utilizados, por exemplo, para o desenvolvimento de modelos de previsão robustos com aplicação de aprendizagem de máquina. De acordo com Wang et al. (2022), a aprendizagem de máquina tem atraído muitos pesquisadores no ramo da bioenergia, pois tem sido uma alternativa confiável para estudar os mecanismos de reação de diferentes processos de conversão de biomassa.

Segundo Umenweke et al. (2022), vários pesquisadores têm utilizado a aprendizagem de máquina para modelar processos químicos complexos, especialmente no campo do biodiesel. As pesquisas mostraram que os modelos baseados em aprendizagem de máquina foram superiores em comparação aos métodos tradicionais. Portanto, o presente trabalho tem como finalidade utilizar algoritmos de aprendizagem de máquina para prever o rendimento de biodiesel por transesterificação via catálise heterogênea, a partir de dados publicados na literatura.

2 JUSTIFICATIVA

Existe um crescimento significativo das demandas energéticas em todo o mundo. Com isso, a urgência da supressão dessas necessidades energéticas tem provocado um aumento considerável no consumo de combustíveis fósseis, o que leva a graves problemas de segurança energética. Esses problemas têm sido agravados devido à crise geopolítica na Europa Oriental, além dos dilemas ambientais causados por esse consumo (PEREIRA, 2022).

A crise ambiental e os problemas de saúde causados pela emissão de poluentes decorrentes da exploração, refino e utilização de combustíveis fósseis têm levado vários países a buscar alternativas e realizar pesquisas direcionadas ao consumo sustentável de energia. No Brasil, em 2024, foi aprovado o projeto de lei "Combustíveis do Futuro", que cria programas nacionais de diesel verde. Este projeto prevê a mistura de biodiesel à gasolina e ao diesel, com o objetivo de atingir uma mistura de 20% até março de 2030 (CÂMARA DOS DEPUTADOS DO BRASIL, 2024).

Esta evolução da proteção dos recursos naturais, requer inovações. De tal maneira, muitos pesquisadores procuram alternativas para resolver todos os problemas associados ao uso do biodiesel, incluindo a exploração de diferentes tipos de matérias-primas, o uso de aditivos e modificações nos motores (SINGH et al., 2020). Outra preocupação é o rendimento da reação de transesterificação, que é o processo mais utilizado na produção de biodiesel. Aspectos fundamentais para alcançar um rendimento satisfatório incluem a otimização da concentração do catalisador, o controle preciso da temperatura durante o processo, o tempo de reação, o tipo de óleo utilizado, entre outros fatores (VICENTE; MARTÍNEZ; ARACIL, 2004).

A medição de todas as propriedades físico-químicas do biodiesel utilizando métodos experimentais é uma tarefa demorada e difícil. Por isso, alternativas são amplamente procuradas, como os modelos matemáticos e estatísticos. No entanto, esses métodos muitas vezes enfrentam limitações, como a capacidade inadequada para capturar sistemas não lineares e o comportamento dinâmico da produção de biodiesel, bem como a disponibilidade limitada de dados

para realizar esses testes estatísticos. Nesse contexto, a inteligência artificial mostra-se promissora como uma alternativa às modelagens tradicionais, permitindo lidar com a natureza não linear, incerta, multivariada e complexa dos sistemas de produção de biodiesel (AGHBASHLO et al., 2021).

Diante disso, este trabalho visa aplicar algoritmos de aprendizagem de máquina, uma subárea da inteligência artificial de modo a identificar por meio de predições, os principais parâmetros que impactam na obtenção de um maior rendimento do biodiesel através do processo de transesterificação via catálise heterogênea. Além disso, este trabalho busca contribuir para futuras otimizações, fornecendo informações que podem orientar na tomada de decisões do processo e direcionar pesquisas subsequentes.

3 OBJETIVOS

3.1 OBJETIVO GERAL

O presente trabalho tem como finalidade utilizar algoritmos de aprendizagem de máquina para a predição do rendimento da produção de biodiesel por transesterificação via catálise heterogênea.

3.2 OBJETIVOS ESPECIFICOS

- Analisar artigos publicados em bases de dados sobre a produção de biodiesel via catalise heterogênea utilizando CaO como catalisador e metanol como álcool, para assim obter um banco de dados do processo de produção de biodiesel;
- Realizar a análise exploratória dos dados utilizando métodos de estatística descritiva;
- Treinar e avaliar algoritmos de aprendizagem de máquina;
- Obter um modelo computacional capaz de realizar predições de rendimento na produção de biodiesel via catalise heterogênea utilizando CaO;
- Realizar uma validação entre os resultados obtidos a partir do modelo preditivo desenvolvido e os resultados experimentais de outros autores;
- Analisar o comportamento dos principais parâmetros da reação de transesterificação utilizando o modelo preditivo.

4 FUNDAMENTAÇÃO TEÓRICA

Neste tópico são apresentadas as principais motivações para a migração energética aos biocombustíveis. O capítulo também trata sobre os conceitos fundamentais da produção de biodiesel e ainda os principais fatores que afetam o rendimento. Além de serem detalhados, conceitos gerais de aprendizagem de máquina.

4.1 BIOCOMBUSTÍVEIS COMO UMA ALTERNATIVA SUSTENTÁVEL

O mundo contemporâneo enfrenta desafios significativos, incluindo o aumento dos preços dos combustíveis, as mudanças climáticas e a poluição. Além disso, a crescente demanda por energia coloca em evidência a escassez de recursos energéticos, o que, por sua vez, impacta o desenvolvimento econômico global.

Os combustíveis podem gerar problemas que impactam de forma social que em grande parte é devido à dependência de fontes não renováveis, resultando em uma distribuição desigual. Um claro exemplo é o Oriente Médio, que detém 63% das reservas globais e é o fornecedor dominante de petróleo, o que torna o sistema insustentável a longo prazo devido a questões de equidade, econômicas, geopolíticas e, sobretudo, ambientais. Se comparado com os recursos de energia renovável, estes são distribuídos de forma mais uniforme no mundo (DEMIRBAS; FATIH DEMIRBAS, 2011).

Visto que os combustíveis fósseis constituem fontes de energia não renováveis, naturalmente surgem alternativas para o desenvolvimento de outros tipos de combustíveis. Por essa razão, com as políticas ambientais cada vez mais voltadas para o estudo dos biocombustíveis visando a redução das emissões, os recentes avanços na engenharia de cultivos e nos processos de fermentação, a produção de bioetanol e biodiesel tornaram-se mais uma vez substitutos viáveis e sustentáveis para os combustíveis derivados do petróleo (LIU et al., 2021).

Os biocombustíveis desempenham um papel fundamental, pois têm a capacidade de substituir os combustíveis derivados do petróleo. Além disso, as fontes de energia à base de biomassa são consideradas potencialmente neutras e sustentáveis. Para a síntese de biocombustíveis, por exemplo, no caso do biodiesel, são comumente utilizados como matéria-prima o óleo de soja, o óleo de palma e outros óleos comestíveis. No entanto, essas fontes são limitadas devido a preocupações com a segurança alimentar. Por outro lado, os óleos de cozinha residuais têm o potencial de substituir esses óleos comestíveis, representando uma alternativa sustentável como matéria-prima, cuja disponibilidade pode ser considerada suficiente (NIGAM; SINGH, 2011).

Sallet e Alvim (2011) afirmam que o Brasil é o terceiro maior produtor mundial de biodiesel, perdendo apenas para Alemanha e França. O uso de energia limpa é significativo, com a maior utilização ocorrendo no setor de transporte. Isso se deve às suas propriedades, como baixa emissão de enxofre, monóxido de carbono e hidrocarbonetos (GURUNATHAN; RAVI, 2015).

A tecnologia dos biocombustíveis é relevante tanto para países industrializados como para países em desenvolvimento, trazendo o conceito de desenvolvimento sustentável com a ideia de interligação e equilíbrio entre o setor econômico, social e ambiental. Os biocombustíveis são uma fonte de energia renovável feita a partir de materiais de base biológica que resultam em produtos como o biogás, bihidrogênio, biodiesel, bioetanol e o bio-carvão. Os custos variam muito conforme a matéria-prima, processo de conversão, escala de produção e região (DEMIRBAS, 2009).

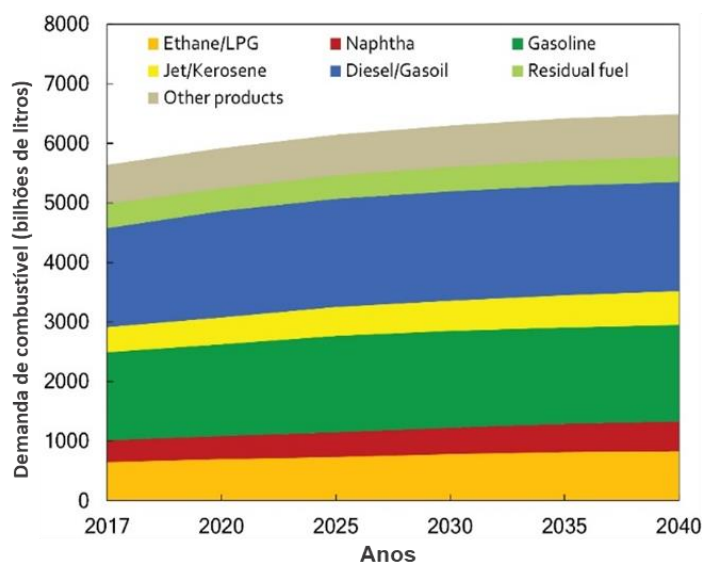
4.2 BIODIESEL

O biodiesel é um dos biocombustíveis mais estudados. Ele é considerado como um éster alquílico de ácido graxo limpo, aromático e biodegradável derivado de óleos vegetais ou gordura animal, e geralmente óleos residuais (XING et al., 2021).

Há vários combustíveis renováveis em desenvolvimento e, de fato, o biodiesel é uma das opções mais promissoras. Isso se deve às suas propriedades ecológicas favoráveis, uma vez que é composto por ésteres alquílicos de ácidos graxos de cadeia longa, preparados a partir da transesterificação de triglicerídeos ou da esterificação de ácidos graxos livres, juntamente com álcool na presença de um catalisador (AGHBASHLO et al., 2021).

Este biocombustível é uma alternativa favorável quando comparado aos produtos derivados do petróleo, como o diesel, que têm uma demanda global muito alta. Existem perspectivas que indicam um aumento ainda maior, conforme com a Figura 1, espera-se que a demanda global por diesel e gásóleo aumente significativamente até 2040.

Figura 1 – Os produtos refinados globais e suas perspectivas de 2017 a 2040



Fonte: adaptado de AGBASHLO et al., 2021

As propriedades similares entre o diesel e o biodiesel incluem o índice de cetano, a viscosidade e o teor de energia. O biodiesel apresenta várias vantagens, como a liberação reduzida de emissões, a ausência de enxofre, propriedades físico-químicas vantajosas e uma combustão mais completa, mas o que o torna ainda mais atraente é a sua produção fácil a partir de matéria-prima renovável (XING et al., 2021).

Na Tabela 1, podem ser observadas algumas vantagens e desvantagens da utilização do biodiesel como o biocombustível que poderia substituir o óleo diesel:

Tabela 1 – Vantagens e desvantagens da utilização do Biodiesel

Vantagens	Desvantagens
- Combustível renovável, proveniente de diversas fontes.	- Emissões de NOx ligeiramente superiores ao diesel convencional.
- Isenção de taxas e tarifas de importação.	- Consumo de combustível ligeiramente superior.
- Baixa toxicidade em comparação ao diesel.	- Dificuldade de partida a frio em climas frios.
- Biodegradável, reduzindo impactos ambientais.	- Ponto de congelamento mais alto.
- Menor emissão de contaminantes.	- Menos estável que o diesel a longo prazo.
- Menor risco para a saúde devido a menores emissões.	- Degradação de juntas e mangueiras.
- Mais rentável, possibilitando produção local.	- Problemas de corrosão em materiais específicos.
- Processo de produção menos moroso.	- Viscosidade superior pode causar problemas no motor.
- Excelentes propriedades lubrificantes.	- Maior depósito de carbono em pistões.
- Contribui para resolução da crise energética.	- Formação de depósitos de goma em longos períodos de uso.
- Não requer modificações nos motores para B20.	- Problemas de bombeamento e atomização.
	- Requer modificações para uso de B100.

Fonte: adaptado de PEREIRA, 2022

Cabe ressaltar que as desvantagens apresentadas na Tabela 1 são significativamente reduzidas quando o biodiesel é misturado com o diesel convencional.

4.2.1 Gerações de biodiesel

De acordo com Singh et al. (2020), o biodiesel é comumente classificado em primeira, segunda, terceira e quarta geração, com base em sua origem e processo de produção:

- A. Biodiesel de primeira geração: nesta categoria, o biodiesel é produzido a partir de matérias-primas comestíveis, como óleo de milho, coco, arroz e outros vegetais.
- B. Biodiesel de segunda geração: aqui, o biodiesel é obtido a partir de matérias-primas não comestíveis, como óleos provenientes de materiais lignocelulósicos, por exemplo, o óleo de linhaça, óleo de jatropa, óleo de karanja, óleo de semente de borracha, etc.

- C. Biodiesel de terceira geração: nesta categoria, o biodiesel é produzido a partir de óleos residuais de gordura animal, óleos alimentares usados e resíduos industriais, bem como óleo de algas.
- D. Biodiesel de quarta geração: a mais recente classificação, o biodiesel de quarta geração, baseia-se em ferramentas biológicas desenvolvidas pelo homem, utilizando técnicas de biologia sintética. Este método concentra-se na obtenção de microrganismos geneticamente modificados para produzir lipídios diretamente a partir de biomassa não alimentar.

4.2.2 Produção de biodiesel

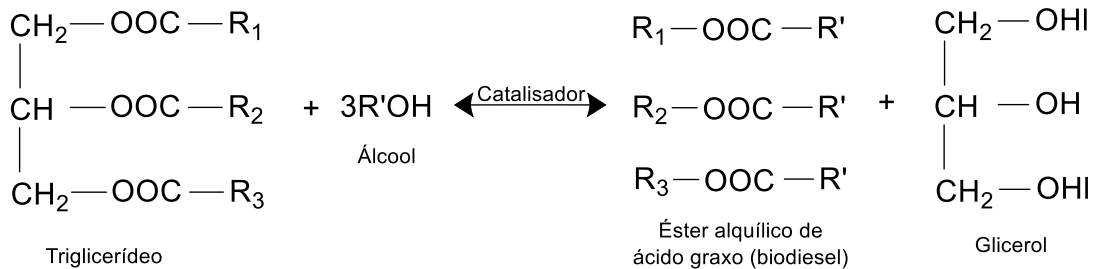
A utilização de óleos vegetais ou gorduras animais como combustíveis demanda a conversão desses materiais em alquil ésteres de ácidos graxos, conhecidos como biodiesel. Essa transformação é necessária devido à elevada viscosidade dessas substâncias não transesterificadas, que pode acarretar sérios problemas operacionais nos motores diesel. Diversos métodos têm sido estudados para mitigar essa alta viscosidade, incluindo a utilização de misturas binárias com petrodiesel, métodos de pirólise, microemulsificação, esterificação e transesterificação (PEREIRA, 2022).

A reação de transesterificação é o principal método utilizado para a produção desse combustível, devido à sua eficiência, simplicidade e custo relativamente baixo. Contudo, sua utilização depende do teor de ácidos graxos livres (FFAs) na matéria-prima. A transesterificação é preferível para produzir biodiesel de alta pureza, enquanto a esterificação pode ser preferível para resíduos com alto teor de FFAs, visando minimizar os custos de pré-tratamento. Em termos comerciais, a transesterificação é de longe o método preferido e mais amplamente utilizado (SABOYA, 2012).

Conforme ilustrado na Figura 2, essa reação envolve a combinação de um óleo ou gordura, seja de origem vegetal ou animal, com um álcool na presença de um catalisador. Como resultado, são obtidos ésteres, que são os componentes do biodiesel, e como subproduto, o glicerol (AMENAGHAWON; EVBARUNEGBE; OBAHIAGBON, 2021). Conforme destacado por Adames et al. (2021), cerca de 10%

da produção de biodiesel corresponde ao glicerol, uma matéria-prima econômica com potencial para a geração de energia renovável, incluindo a produção de gás hidrogênio (H₂) e metano (CH₄).

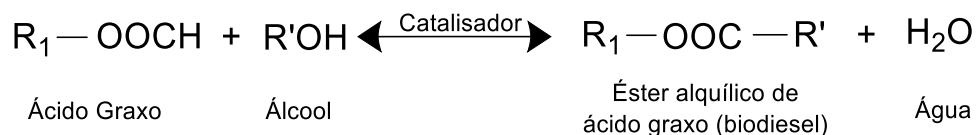
Figura 2 – Reação de transesterificação na produção de biodiesel



Fonte: adaptado de SANDOVAL et al., 2017

Na reação de esterificação, ocorre a reação dos ácidos carboxílicos de cadeia longa (ácidos graxos) com um álcool, produzindo éster e água (Figura 3). Essa abordagem apresenta a vantagem de permitir o uso de matérias-primas residuais. No entanto, requer catalisadores com custo mais elevado, condições de reação mais severas e tem compatibilidade limitada com a indústria existente (SABOYA, 2012).

Figura 3 – Reação de esterificação na produção de biodiesel



Fonte: adaptado de SANDOVAL et al., 2017

Enquanto aos álcoois, segundo Moayedí et al. (2020), o metanol e o etanol são dois álcoois comumente utilizados no processo de transesterificação, embora também existam outros, como o propanol e o butanol. Neste contexto, o metanol é mais amplamente empregado em escala comercial devido ao seu menor preço. É importante ressaltar que os álcoois e os óleos não se misturam completamente, sendo a qualidade da mistura um fator-chave que pode afetar a eficiência da transesterificação.

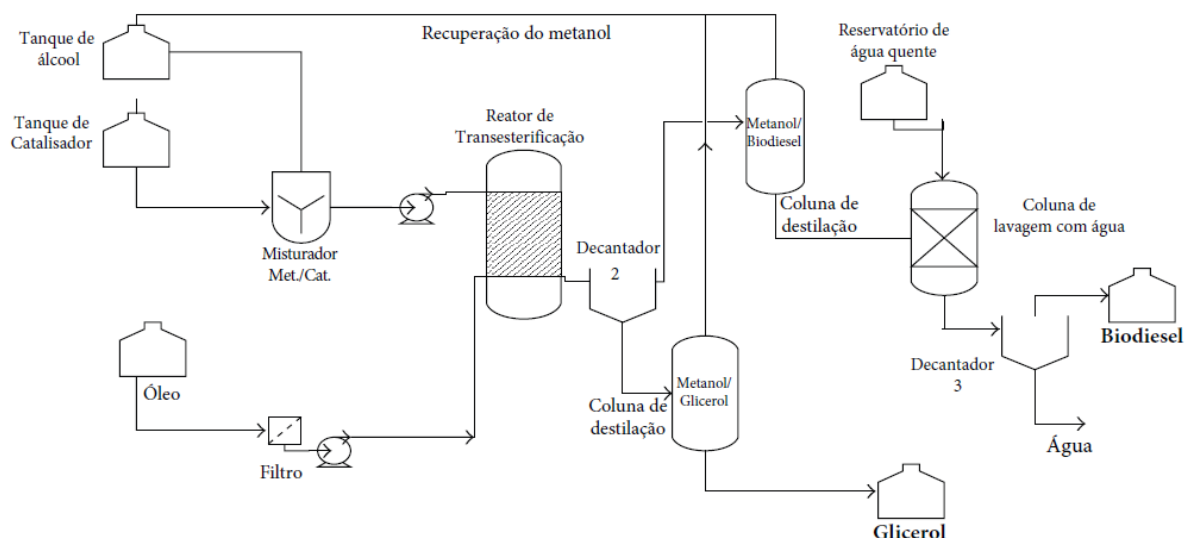
Além disso, cabe ressaltar que o metanol também é especialmente prevalente devido às suas vantagens físicas e químicas, ele é caracterizado por sua

cadeia carbônica curta e polaridade, que faz com que reaja mais rapidamente com os triglicerídeos presentes nos óleos vegetais em comparação com o etanol, apesar de que o etanol tem melhor desempenho na combustão. No entanto, é importante considerar que o uso do metanol acarreta algumas desvantagens significativas, como sua alta toxicidade e sua obtenção geralmente proveniente do petróleo (SABOYA, 2012).

Embora existam muitas vantagens na produção de biodiesel como uma alternativa sustentável, um dos desafios da produção é que vários parâmetros operacionais podem afetar o rendimento do biodiesel durante a reação de transesterificação. Estes incluem o tempo de reação, temperatura, razão molar álcool/óleo e o tipo e quantidade de catalisador utilizado na reação (GUPTA et al., 2021).

Na Figura 4, é apresentado um exemplo do processo de produção de biodiesel por transesterificação, onde o álcool e o catalisador são utilizados para promover a reação química. Nesse processo, ocorre a reação de transesterificação, resultando na formação de biodiesel e glicerol como produtos.

Figura 4 – Processo da produção de Biodiesel por transesterificação



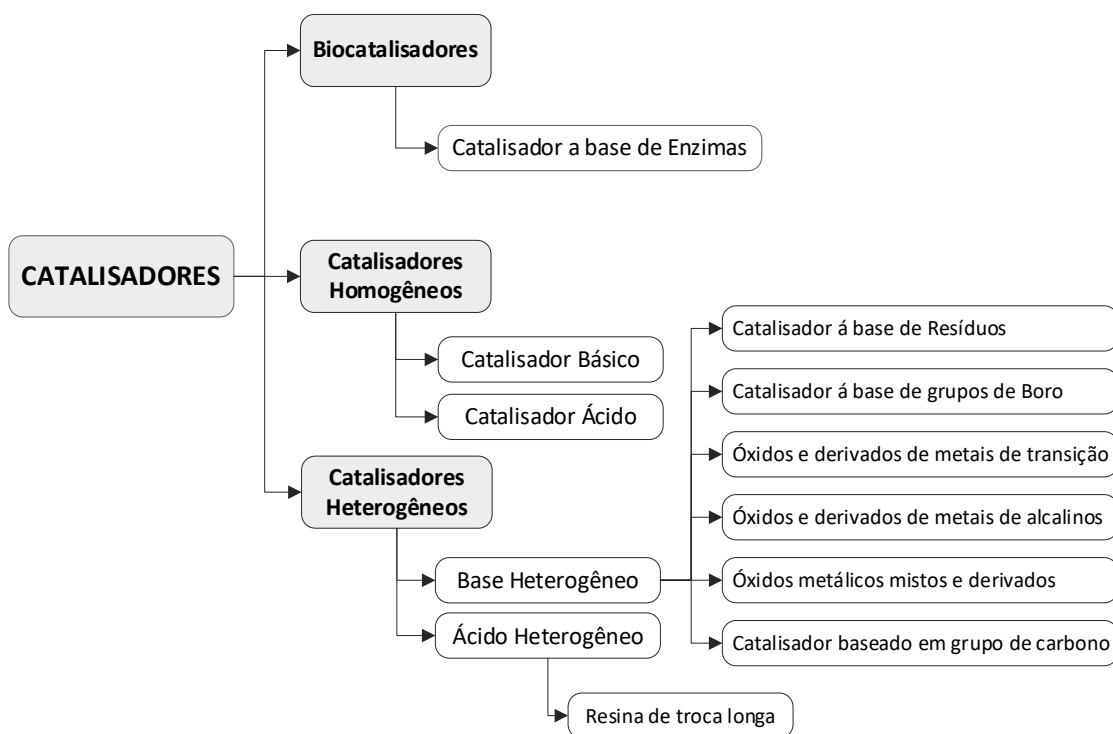
Fonte: adaptado de GNANAPRAKASAM et al., 2013

4.2.3 Catalisadores

Os catalisadores são fundamentais para facilitar a reação de transesterificação, auxiliando no rendimento do biodiesel. Entre os catalisadores convencionais estão os químicos e biológicos. Os catalisadores químicos incluem catalisadores ácidos e básicos, enquanto os catalisadores biológicos envolvem enzimas, como pode ser observado na Figura 5 (PARIDA; SINGH; PRADHAN, 2022).

Além disso, as reações que envolvem catalisadores podem ser classificadas em catálise homogênea, onde tanto os catalisadores quanto os reagentes constituem uma só fase. Também se enquadra nessa classificação a catálise heterogênea, onde os catalisadores e os reagentes apresentam mais de uma fase (COSTA, 2011).

Figura 5 – Classificação dos catalisadores



Fonte: adaptado de CHOUHAN e SARMA, 2011

O autor Costa (2011) descreve os tipos de processos catalíticos mais usados para a síntese de biodiesel, da seguinte forma:

A. CATÁLISE HOMOGÊNEA

- I. Catálise Homogênea básica: uma base é empregada como catalisador da reação. Aqui, os catalisadores alcalinos, como hidróxidos e alcóxidos de sódio ou potássio, são amplamente utilizados devido à sua facilidade de manipulação e menor corrosividade em comparação com os catalisadores ácidos.
- II. Catálise Homogênea ácida: na catálise homogênea ácida, os ácidos de Brønsted são os mais comumente empregados, incluindo ácido clorídrico, sulfúrico (anidro) e sulfônicos. Esse método é utilizado na síntese de biodiesel quando a matéria-prima possui um alto teor de ácidos graxos livres e água. Embora a transesterificação em meio ácido resulte em alto rendimento de ésteres monoalquílicos, possui algumas desvantagens. Por exemplo, exige uma alta razão molar álcool/óleo e é um processo mais lento, requerendo períodos de síntese mais longos.

B. CATÁLISE HETEROGÊNEA

A catálise heterogênea oferece vantagens significativas, desde a facilidade de purificação dos ésteres monoalquílicos até a possibilidade de reutilização dos catalisadores e a redução na geração de efluentes. Por outro lado, o uso de catalisadores homogêneos está associado a diversos problemas, como um maior número de etapas, dificuldades na purificação dos produtos, risco de corrosão do reator (em catálise ácida) e saponificação (em catálise básica), além de uma baixa atividade em relação a álcoois de maior massa molar. Além disso, o uso de catalisadores heterogêneos facilita consideravelmente a recuperação e purificação da glicerina.

De acordo com Cardoso et al. (2020), os catalisadores heterogêneos podem ser agrupados em sete grupos: óxidos metálicos, óxidos metálicos mistos e dopados, catalisadores suportados, zeólitas, hidróxidos duplos lamelares, bases orgânicas e resinas aniônicas.

Dada essa classificação, cabe ressaltar que para o biodiesel ou os ésteres alquílicos de ácidos graxos, são comumente utilizados catalisadores homogêneos ácidos, porém tem uma desvantagem que são os problemas

ambientais e de corrosão (MAHMOUD; EL-MOLLA; IBRAHIM, 2020). Por outro lado, os catalisadores homogêneos alcalinos são altamente higroscópicos e absorvem água do ambiente durante o armazenamento, como também formam água quando dissolvidos no reagente álcool, afetando o rendimento (LEUNG; GUO, 2006).

Portanto, pensando na sustentabilidade, os catalisadores heterogêneos apresentam características vantajosas, como a facilidade de separação dos produtos na recuperação e reciclabilidade, resultando em menos resíduos e, conseqüentemente, uma redução do impacto ambiental e dos custos do processo. Isso se deve aos óxidos metálicos, amplamente utilizados como catalisadores (MAHMOUD; EL-MOLLA; IBRAHIM, 2020). Além disso, esses catalisadores não produzem sabões por meio da neutralização de ácidos graxos livres e pela saponificação de triglicerídeos (VICENTE; MARTÍNEZ; ARACIL, 2004).

Entretanto, Leung, Wu e Leung (2010) afirmam que catalisadores enzimáticos também são atraentes para a produção de biodiesel, pois evitam a formação de sabão o que facilita no processo de purificação, porém o tempo de reação é mais longo que em catalisadores químicos, esta característica do ponto de vista comercial, os tornam menos valorizados, além de seu custo elevado.

4.3 PARAMETROS QUE INFLUENCIAM NO RENDIMENTO DA REAÇÃO DE TRANSESTERIFICAÇÃO

Os parâmetros são um conjunto de condições em que o processo de produção de biodiesel é encaminhado. Variáveis como temperatura de reação, concentração inicial de catalisador e razão estequiométrica metanol/óleo vegetal podem influenciar no rendimento da reação. Um dos fatores mais relevantes é justamente a concentração do catalisador e o segundo fator é a temperatura. Os efeitos que ambos causam, têm uma influência negativa no rendimento do biodiesel, devido à reação colateral de saponificação de triglicerídeos que é favorecida em altas concentrações e temperaturas de catalisador (VICENTE; MARTÍNEZ; ARACIL, 2004).

Segundo Pereira (2022), foram identificados quatro parâmetros críticos no processo de síntese do biodiesel. São eles, a temperatura de reação, o tempo de

residência, o tipo e quantidade de álcool e o tipo e quantidade de catalisador. Nas subseções seguintes, cada parâmetro é discutido.

4.3.1 Temperatura

A temperatura de reação desempenha um papel crucial na produção de biodiesel, influenciando não apenas a conversão de glicerídeos, mas também a velocidade da reação. Com o aumento da temperatura, a viscosidade da mistura diminui, o que aumenta a velocidade da reação devido à maior disponibilidade energética para a reação de transesterificação, resultando em uma melhoria na conversão de glicerídeos em ésteres. Em muitos casos, a temperatura mais adequada para o processo está abaixo da temperatura de ebulição do álcool (PEREIRA, 2015).

No caso da utilização de metanol, a temperatura operacional ótima situa-se em um intervalo de 60 °C a 65 °C, podendo ser até temperaturas mais baixas, conforme mencionado por alguns autores. No entanto, a utilização de uma temperatura de reação superior ao ponto de ebulição do álcool não é recomendada devido às perdas substanciais de álcool para a fase vapor. Além disso, uma temperatura de reação excessivamente alta pode propiciar reações indesejadas de saponificação. Embora alguns experimentos possam empregar temperaturas mais altas, é essencial considerar os potenciais efeitos adversos dessas condições térmicas no processo (PEREIRA, 2022).

4.3.2 Tempo

Em geral, a conversão de glicerídeos tende a aumentar com o aumento do tempo de residência no reator. Inicialmente, a reação é lenta devido à agitação fraca e à dispersão limitada do álcool no óleo, resultando em reações apenas na superfície do óleo e dos triglicerídeos. Com o passar do tempo, a reação é acelerada até atingir a conversão de equilíbrio dos glicerídeos. No entanto, observa-se um aumento na conversão de glicerídeos para tempos de residência inferiores a 90 minutos, enquanto tempos de residência mais longos podem levar à perda de produto devido à formação de sabões (LORA; VENTURINI, 2012). Estudos indicam que os tempos de residência ideais para uma reação de transesterificação,

utilizando um catalisador básico, situam-se geralmente entre 1 a 2 horas, embora existam algumas exceções. Em contrapartida, o tempo necessário para reações com catalisadores ácidos pode ser mais prolongado devido à sua baixa reatividade, o tempo irá depender das condições específicas do processo (PEREIRA, 2022).

4.3.3 Tipo e quantidade de álcool

A proporção entre álcool e óleo no abastecimento do reator é um dos aspectos mais sensíveis na produção de biodiesel através da transesterificação, tendo um impacto notável na conversão dos glicerídeos. Segundo os cálculos estequiométricos, cada mol de triglicerídeos reage com 3 mols de álcool, resultando na produção de 1 mol de glicerol e 3 mols de ésteres (SINGH et al., 2021). Para garantir a conclusão da reação, geralmente é introduzido álcool em excesso, promovendo o deslocamento da mesma no sentido desejado. Uma proporção maior de álcool em relação ao óleo favorece uma conversão mais eficiente dos glicerídeos. No entanto, a decisão sobre essa proporção depende do tipo de catalisador e do processo adotado. Como por exemplo, quando se utiliza um catalisador básico, geralmente recomenda-se uma relação estequiométrica álcool/óleo de 6:1 (PEREIRA, 2022).

De acordo com Cardoso et al. (2020), na reação de transesterificação podem ser usados diferentes álcoois, incluindo metanol, etanol, propanol, butanol e álcoois isopropílicos. No entanto, o metanol se destaca em reações de transesterificação, pois é capaz de catalisar a temperatura ambiente para produzir biodiesel e glicerol. Embora o etanol e o propanol também possam atuar como agentes transesterificantes, eles exigem condições reacionais mais severas do que o metanol para atingir a mesma conversão. O metanol além de ser mais econômico que o etanol, possui alta pureza e oferece uma facilidade na separação entre o éster e a glicerina formada, devido à sua cadeia mais curta e maior polaridade, sendo assim o álcool mais predominante utilizado na produção de biodiesel.

4.3.4 Tipo e quantidade de catalisador

Na produção de biodiesel, o tipo e a quantidade de catalisador desempenham papéis cruciais, afetando diretamente o desempenho e a eficiência

do processo. Geralmente, são empregados catalisadores homogêneos ou heterogêneos, podendo ser tanto ácidos quanto básicos dependendo da composição da matéria-prima. Também neste processo podem ser utilizados os catalisadores enzimáticos ou biocatalisadores (DABDOUB; BRONZEL; RAMPIN, 2009).

Os catalisadores homogêneos alcalinos, como NaOH ou KOH, são amplamente utilizados devido à sua eficácia na transesterificação e à rapidez da reação. No entanto, altos teores de ácidos graxos livres (FFAs), como por exemplo em óleos residuais podem tornar esses catalisadores menos eficientes, exigindo a utilização de catalisadores ácidos (PEREIRA, 2022). Embora a transesterificação em meio ácido proporcione um alto rendimento em ésteres monoalquílicos, ela apresenta o inconveniente de exigir uma alta razão estequiométrica de álcool para óleo, além de ser um processo lento, necessitando de longos períodos de síntese (COSTA, 2011).

Apesar de que o processo convencional de produção de biodiesel seja por transesterificação alcalina em meio homogêneo, este método demanda várias etapas de purificação para garantir que os monoésteres produzidos atendam aos padrões recomendados. Por isso, catalisadores heterogêneos emergem como alternativas promissoras, oferecendo vantagens técnicas e ambientais significativas. Estes catalisadores facilitam idealmente a purificação dos monoésteres, permitem a reutilização do catalisador sólido e reduzem a geração de resíduos. Além disso, simplificam consideravelmente a recuperação e purificação da glicerina resultante da alcoólise (RAMOS et al., 2011).

Diversas classes de compostos químicos têm sido propostas como catalisadores potenciais para este processo, incluindo zeólitas, óxidos inorgânicos, sais inorgânicos, compostos de coordenação, resinas trocadoras de íons, líquidos iônicos, ácidos e bases orgânicos, hidróxidos duplos lamelares e óxidos estruturais e carboxilatos lamelares (CORDEIRO et al., 2011). Essa diversidade de opções oferece oportunidades para otimizar o processo de produção de biodiesel e melhorar sua sustentabilidade ambiental.

Em termos de quantidade ótima de catalisador, ela varia conforme a matéria-prima e as condições de reação. De acordo com Athar e Zaidi (2020), os

catalisadores homogêneos alcalinos são preferíveis para óleos com teor de ácidos graxos livres (FFAs) abaixo de 0,5% em peso, sendo crucial a concentração inicial, sendo a quantidade ideal em torno de 1,0% em peso. No entanto, variações menores, como 0,5%, 0,7% ou 0,8%, e até mesmo concentrações mais elevadas, como 6% ou 3,3% de NaOH, são utilizadas. Por outro lado, os catalisadores ácidos, menos sensíveis aos ácidos graxos livres, podem variar sua concentração ideal entre 0,5% e 2% em peso, dependendo do tipo de ácido e das características do óleo.

Uma alternativa são os catalisadores heterogêneos, mas geralmente demandam concentrações mais altas, entre 2% e 10% em peso, além de temperaturas e tempos de reação mais elevados para alcançar rendimentos comparáveis aos catalisadores homogêneos. Dado isso, a escolha cuidadosa do tipo e da quantidade de catalisador é essencial para garantir uma produção eficiente e econômica de biodiesel, considerando as características da matéria-prima e os requisitos de qualidade do produto final (LORA; VENTURINI, 2012).

Entre os catalisadores heterogêneos, o óxido de cálcio (CaO) se destaca na produção de biodiesel devido à sua eficiência na transesterificação, além da capacidade de tolerância aos ácidos graxos do meio reacional (SOARES, 2018). Este catalisador oferece vantagens, desde fácil obtenção a partir de matérias-primas comuns, além de não causar impactos negativos ao médio ambiente. O CaO é considerado como um dos catalisadores heterogêneos potenciais para aplicações industriais, entretanto sua reutilização ainda pode ser um problema, pois exige tratamentos após a sua utilização devido aos resíduos orgânicos em sua superfície (CABELLO; S.; A., 2017).

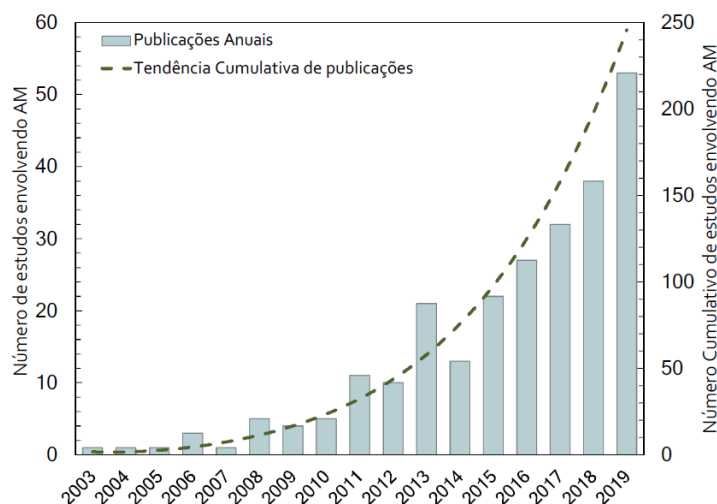
4.4 USO DA INTELIGÊNCIA ARTIFICIAL NA MODELAGEM DA PRODUÇÃO DE BIODIESEL

No que diz respeito à modelagem e otimização do sistema de produção de biodiesel, uma abordagem centrada na aprendizagem de máquina (AM) está surgindo. Espera-se um aumento significativo na adoção dessas técnicas pela indústria, dados os desafios enfrentados devido à complexidade dos mecanismos envolvidos na reação de transesterificação, visto que uma modelagem adequada

permite a obtenção de informações importantes sobre as condições reacionais, resultando em uma melhoria do produto final.

De acordo com Gupta et al. (2021), a medição da produtividade e qualidade do biodiesel enfrenta desafios devido à complexidade dos parâmetros e às interações não lineares. Embora os experimentos físicos tenham sido tradicionalmente empregados para estimar esses parâmetros, sua execução pode ser custosa e demorada. Dado isso, o uso da aprendizagem de máquina é considerado uma alternativa eficiente para analisar grandes conjuntos de dados e otimizar processos, reduzindo a dependência de experimentos físicos extensivos. Por isso, nos últimos anos observa-se um aumento no uso de modelos para realizar predições que envolvem a produção de biodiesel (Figura 6).

Figura 6 – Aumento de publicações utilizando AM na pesquisa de Biodiesel



Fonte: adaptado de AGHBASHLO et al., 2021

4.5 APRENDIZAGEM DE MÁQUINA

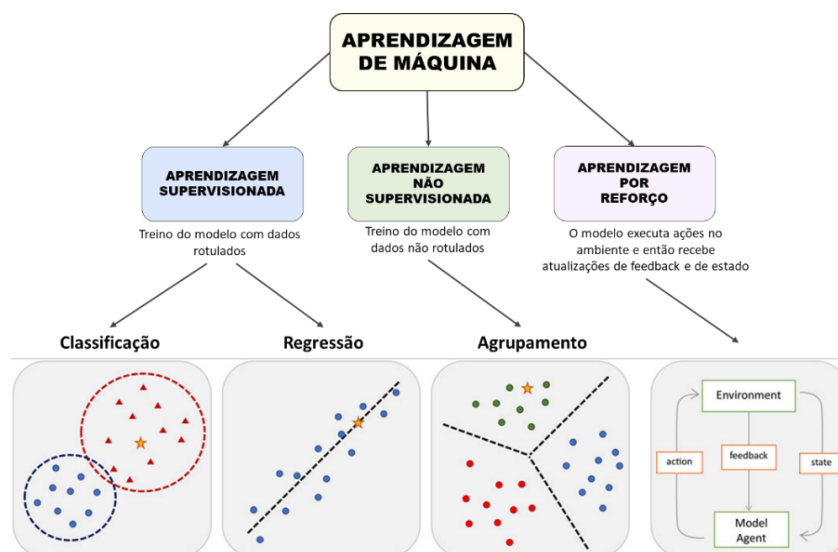
Com a disponibilidade de grandes volumes de dados, as empresas estão cada vez mais interessadas em explorar essas observações documentadas para obter vantagem competitiva. A ciência de dados é um conjunto de princípios e técnicas que orientam na extração de informações desses dados. Uma das ferramentas mais utilizadas dentro desta ciência é a aprendizagem de máquina

(AM), uma das principais forças que impulsionam à indústria 4.0 (PROVOST; FAWCETT, 2013).

A AM procura desenvolver métodos eficazes para reconhecer padrões, analisando um conjunto de dados de treinamento (SIMEONE, 2018). Estes dados servem como uma experiência para elaborar uma função ou hipótese que descreve o problema (CARVALHO et al., 2011). De acordo com Alpaydin (2020), para o reconhecimento de padrões, a AM usa teorias estatísticas para a construção de modelos matemáticos, pois sua tarefa principal é a inferência amostral.

Na Figura 7, podemos observar que os métodos de aprendizagem de máquina são categorizados com base na abordagem de aprendizagem, adotada pelo algoritmo para fazer as predições. Entre as mais comuns estão a aprendizagem supervisionada, não supervisionada (BISHOP, 2006), além de métodos como aprendizagem por reforço (SCHWEIDTMANN et al., 2021).

Figura 7 – Classificação de tipos de aprendizagem de máquina



Fonte: adaptado de PENG et al., 2021

A) Aprendizagem de máquina supervisionada: nesta categoria o algoritmo é treinado com um conjunto de dados rotulados e estruturados, ou seja, as variáveis de saída são conhecidas e estão associadas às variáveis de entrada, podendo ser estas contínuas, categóricas ou binárias (KOTSIANTIS et al., 2007). Há duas principais tarefas que este tipo de aprendizagem executa, que são os problemas de classificação e regressão (PENG et al., 2021).

I. Regressão: a saída é uma variável contínua, tendo como objetivo entender e modelar a relação entre uma variável dependente (ou resposta) e uma ou mais variáveis independentes (ou preditoras). Por exemplo, regressão pode ser aplicada para prever propriedades físicas e químicas de substâncias, como densidade, viscosidade, ponto de ebulição, solubilidade, entre outras.

II. Classificação: a saída é uma variável discreta correspondente a uma categoria de classificação. Por exemplo: Classificar imagens em categorias pré-definidas. Os modelos são treinados com base em características visuais das imagens, como formas, cores e texturas, para identificar e classificar objetos nas imagens.

B) Aprendizagem de máquina não supervisionada: a aprendizagem não supervisionada explora dados não rotulados, onde não há saídas associadas com as entradas para descobrir estruturas ocultas. Este tipo de aprendizagem é amplamente utilizado para agrupamentos (*clustering*), extração de características e detecção de anomalias (SCHWEIDTMANN et al., 2021).

C) Aprendizagem de máquina por reforço: a aprendizagem por reforço é uma forma na qual um agente aprende a tomar decisões em um ambiente dinâmico, por meio de um processo de tentativa e erro para determinar a melhor ação. Durante tal processo, o agente descobre quais ações rendem melhor resultado a partir de recompensas e penalidades em forma de feedbacks, sendo usado em áreas como robótica e em controle de processos industriais (MAHESH, 2019)(SUTTON; BARTO, 2018).

4.6 ALGORITMOS DE APRENDIZAGEM SUPERVISIONADA

Existem vários modelos de aprendizagem de máquina supervisionada, cada um com suas próprias características e aplicações. Nesta seção serão apresentados alguns dos modelos mais comuns:

4.6.1 Modelos Lineares

Os modelos lineares são amplamente utilizados e fazem a predição das saídas usando uma função linear das variáveis de entrada (MÜLLER; GUIDO,

2016). Os modelos lineares se tornam atraentes para diversas aplicações, pois são escaláveis para grandes conjuntos de dados, além da simplicidade e compreensibilidade dos resultados. De acordo com Ritvo et al. (2013), a regressão é um dos métodos mais intuitivos que emprega métodos estatísticos para modelar relações entre as variáveis dependentes e independentes.

A regressão linear é uma técnica amplamente utilizada na aprendizagem de máquina supervisionada para prever valores contínuos. Sua abordagem é simples, se assume que há uma relação linear entre a variável de entrada e a variável de saída. Este modelo é conhecido como Regressão Linear Simples (RLS) quando envolve apenas uma variável de entrada e de Regressão Linear Múltipla (RLM), quando duas ou mais variáveis de entrada estão presentes. Esta relação linear é expressa como uma equação de primeira ordem (Equação 4.1).

$$y = w_1x_1, \dots, w_nx_n + b \quad (4.1)$$

As variáveis x_i denotam as entradas (valores das variáveis independentes, ou preditoras), enquanto w_i e b são os parâmetros do modelo que são ajustados durante o treinamento. A variável y representa a predição do modelo (variável dependente ou *target*). Este procedimento visa encontrar os valores ideais para w_i e b , que minimizam a soma dos quadrados dos resíduos, isto é, a diferença entre o valor predito e o valor real (MÜLLER; GUIDO, 2016).

Neste método, assume-se que as variáveis de entrada e saída têm uma relação linear. No entanto, nem sempre essa relação é linear, dessa forma, é imprescindível recorrer a métodos estatísticos que possam tornar o modelo mais robusto e minimizar viés (JUNQUEIRA et al., 2018). Outros modelos lineares que se destacam são os modelos de regressão regularizada, como a Regressão Ridge e Lasso. Esses modelos utilizam técnicas para reduzir a complexidade e evitar *overfitting*, um problema comum na regressão linear simples (BHATTACHARYYA, 2018).

4.6.2 Modelos baseados em árvores

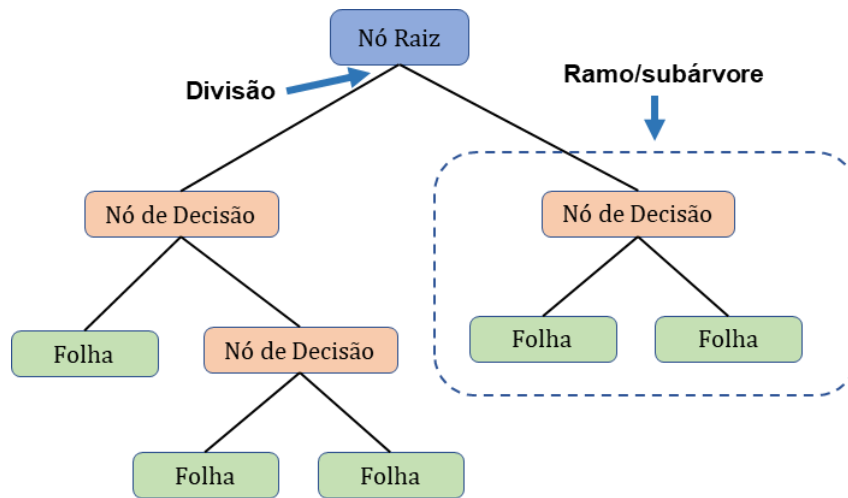
Os modelos baseados em árvores são uma classe de algoritmo de AM que representam decisões em uma estrutura semelhante a uma árvore e que lembram um fluxograma. Portanto, possuem certa facilidade de entendimento e podem ser utilizados tanto para fins de classificação quanto de regressão, além disso, ao contrário dos modelos lineares, eles mapeiam muito bem as relações não lineares das entradas e saídas (JAIN, 2018). Entre os principais modelos baseados em árvores estão as Árvores de Decisão e Floresta Aleatória.

A. Árvores de Decisão

A modelagem por Árvores de Decisão é uma abordagem eficaz no campo da AM. Na Figura 8, podemos observar que ela é composta por um nó raiz, nós de decisão e folhas, formando uma estrutura hierárquica semelhante a uma árvore, onde começa com um nó raiz e termina com uma decisão tomada pelas folhas. O nó raiz representa o início da árvore e avalia a variável que melhor divide os dados. Os nós intermediários são avaliados com base em diferentes variáveis, mas não produzem diretamente as previsões. Por outro lado, as folhas são os finais da árvore, onde as previsões são feitas para uma categoria ou valor numérico (SAINI, 2024).

No algoritmo das Árvores de Decisão, se utiliza a lógica de "verdadeiro" ou "falso", onde cada decisão leva ao próximo nó dependendo da condição avaliada. Essas árvores são construídas de maneira iterativa, avaliando diferentes atributos em cada etapa do processo e selecionando aquele que melhor divide os dados em cada nó (ZORNOZA, 2020).

Figura 8 – Estrutura de uma Árvore de Decisão



Fonte: adaptado de SAINI, 2024

De acordo com Jain (2018), se tanto as variáveis de entrada quanto as de resposta possuem um comportamento linear, os modelos de Regressão Linear tendem a ter uma capacidade de predição superior em comparação aos modelos que utilizam nós de decisão, como as Árvores de Decisão ou Redes Neurais Artificiais. Porém, tratando-se de variáveis que se correlacionam de forma mais complexa e tendendo a modelos não-lineares, as Árvores de Decisão podem ser mais adequadas para a análise.

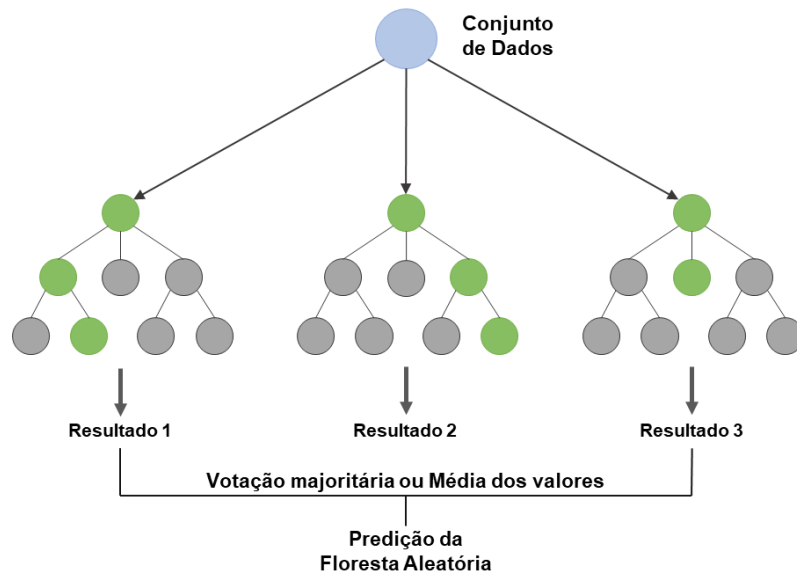
B. Floresta Aleatória

A Floresta Aleatória é outro modelo baseado em Árvores de Decisão utilizados tanto para problemas de regressão como também para classificação. Na engenharia química existem estudos desse algoritmo, por exemplo, para estimar o número de cetano a partir da composição do biodiesel (MIRABOUTALEBI; KAZEMI; BAHRAMI, 2016) e na predição das propriedades do biodiesel, otimizando o perfil de ácidos graxos (SUVARNA GAWALI, 2022).

Este algoritmo utiliza um conjunto de Árvores de Decisão independentes, combinando as predições de cada uma para obter um resultado final. Para problemas de regressão, a decisão final é determinada pela média das decisões individuais de cada árvore, por outro lado, para problemas de classificação a decisão

final será a variável categórica mais frequente entre as Árvores de Decisão formadas como pode ser visto na Figura 9.

Figura 9 – Estrutura do modelo de Floresta Aleatória



Fonte: adaptado de YEHOOSHUA, 2023

Na aprendizagem de máquina supervisionada a Floresta Aleatória é amplamente utilizada, pois ele é um método multivariado não paramétrico, o que significa que os dados da amostra não precisam seguir uma distribuição normal e várias variáveis podem ser trabalhadas simultaneamente. Isso possibilita melhorar a precisão da classificação ou regressão e determinar a importância das variáveis (POLIKAR, 2012).

A Floresta Aleatória usa uma técnica de estatística de reamostragem chamada de *bootstrap*, que serve para criar diferentes conjuntos de dados de treinamento para cada árvore, selecionando aleatoriamente exemplos do conjunto de dados original com substituição, isso significa que o mesmo exemplo pode ser selecionado mais de uma vez (LEE; ULLAH; WANG, 2019).

Segundo Jain (2018), a abordagem da Floresta Aleatória também contribui para contornar alguns problemas frequentes encontrados no processo de tratamento de dados, tais como dados faltantes ou valores discrepantes (*outliers*), além de outras vantagens como a redução do *overfitting* nas Árvores de Decisão, o que torna o modelo mais preciso e flexível para problemas de regressão ou

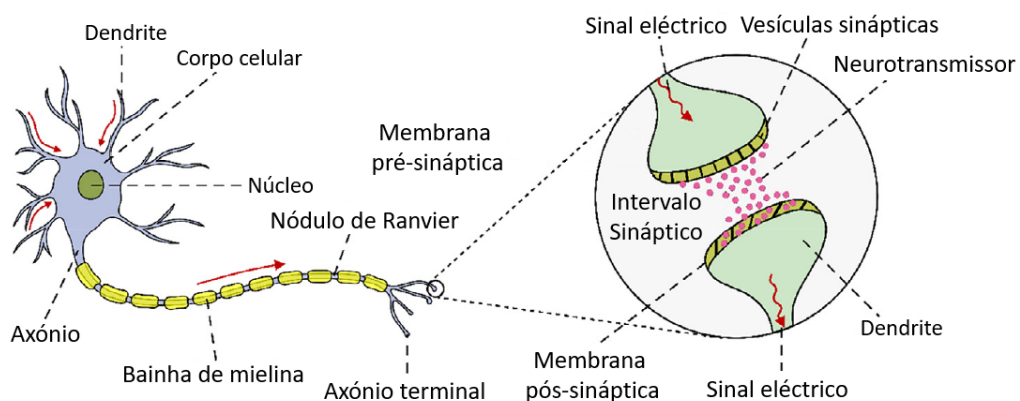
classificação. Em contrapartida, o método requer um alto recurso computacional quando há a necessidade da construção de muitas Árvore de Decisão.

4.6.3 Modelos de Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA) são modelos de aprendizagem de máquina inspirados na estrutura e função do cérebro humano, que consiste em muitos neurônios interconectados responsáveis por processar e transmitir informações. De forma similar, as redes neurais artificiais são compostas por camadas de neurônios artificiais, que têm a capacidade de aprender com os dados e desempenhar uma variedade de tarefas. Essas redes são reconhecidas por sua versatilidade, sendo aplicadas em uma ampla variedade de contextos, que vão desde tarefas de regressão e classificação até problemas que envolvem agrupamento e detecção de anomalias (KUYUCU, 2023).

Para entender o modelo, é crucial compreender como os neurônios biológicos transmitem informações. Os neurônios são responsáveis por receber e transmitir impulsos em diferentes áreas do corpo humano, operando em conjunto, formando redes complexas que desempenham várias funções no cérebro e no sistema nervoso. Como observado na Figura 10, os dendritos recebem os sinais, enquanto o axônio os transmite para outros neurônios, a informação é recebida por meio da sinapse, e essa interconexão é essencial para o funcionamento do sistema nervoso (PREMANAND S, 2022).

Figura 10 – Neurônios biológicos transmitindo informações

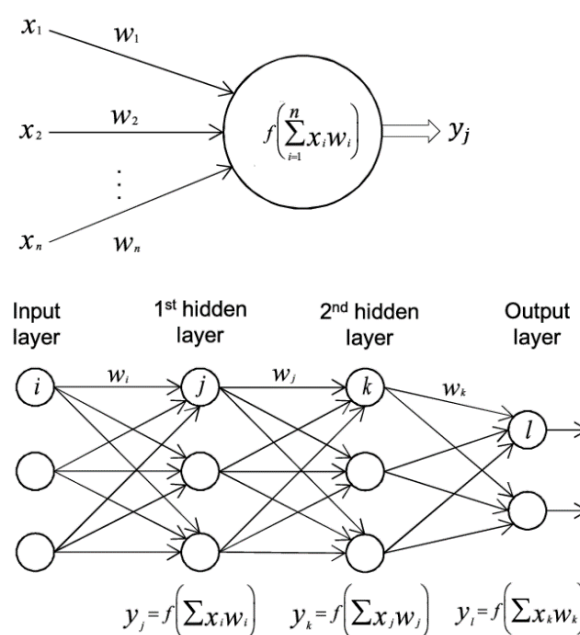


Fonte: adaptado de AGHBASHLO et al., 2021

De forma similar, os neurônios artificiais são as unidades básicas das redes neurais artificiais, representando funções matemáticas capazes de receber entradas, realizar cálculos e produzir uma saída. Cada neurônio é caracterizado por um conjunto de pesos e um viés, o que determina como as entradas são combinadas e transformadas (Figura 11). Os pesos e as tendências são parâmetros de aprendizagem da rede neural. Além disso, cada neurônio possui uma função de ativação que determina a saída com base na entrada. Esta função pode apresentar um comportamento linear ou não linear, e entre as funções comuns de ativação estão o sigmoide, tangente hiperbólica, ReLU, softmax, que são utilizadas de acordo com a natureza do problema (KUYUCU, 2023).

Pereira (2022) no seu trabalho explica que os neurônios em uma rede neural recebem uma sequência de informações (x_i) ponderadas por fatores de peso (w_i). Conforme observado na Figura 11, a estrutura da rede é composta por camadas, sendo a primeira a camada de entrada (*input layer*), seguida por camadas intermediárias chamadas camada de oculta (*hidden layers*) e, por fim, a camada de saída (*output layer*). A informação é transmitida entre neurônios, propagando-se da camada de entrada para as camadas intermediárias e, por fim, para a camada de saída.

Figura 11 – Estrutura de uma Rede Neural Artificial



Fonte: adaptado de VIEIRA; PINAYA; MECHELLI, 2017

A camada de entrada atua como um espaço reservado para os dados, sem realizar cálculos, na camada oculta ocorre a aprendizagem da rede e a maioria dos cálculos. Além disso, ela pode ser composta por várias subcamadas, cada uma com diferentes funções de ativação, adaptadas de acordo com a natureza do problema. Porém, a inclusão de mais camadas pode aumentar a complexidade do modelo. Por fim, a última camada da rede (*output layer*) é responsável por produzir a saída final. Ela pode conter um ou mais neurônios, dependendo do problema ou da saída desejada, o que possibilita a utilização de diferentes tipos de neurônios e funções de ativação (KUYUCU, 2023).

A função de perda é uma medida de quão bem a saída da rede neural corresponde à saída desejada. Para redes neurais artificiais, minimizar a função de perda ajustando seus pesos é conhecido como o processo de aprendizagem. Para minimizar a função de perda, as redes neurais artificiais utilizam o algoritmo de otimização de gradiente descendente (SMANIOTTO, 2023). Este algoritmo atualiza os pesos do modelo usando uma taxa de aprendizagem que determina o produto entre o tamanho da atualização e o gradiente da função de perda. O gradiente é calculado usando a técnica chamada *backpropagation*, que utiliza a regra da cadeia para calcular as derivadas da função de perda em relação aos pesos. Este processo é repetido até que a rede converge para um conjunto de pesos ótimos que minimizem a função de perda (FRANÇA, 2019).

4.7 HIPERPARÂMETROS

Na aprendizagem de máquina existem dois tipos de valores que influenciam o comportamento de um modelo, conhecidos como parâmetros internos e parâmetros de orientação. Uma convenção de nomenclatura para diferenciar entre os dois tipos de valores é chamar os parâmetros internos simplesmente de parâmetros e os parâmetros de orientação de hiperparâmetros. Os parâmetros são variáveis internas de um modelo que são aprendidas a partir dos dados durante o processo de treinamento, eles definem a representação do modelo dos padrões subjacentes nos dados. O processo de treinamento, tem como objetivo encontrar valores ideais para estes parâmetros, que minimizem a discrepância entre as previsões do modelo e os resultados reais (MEHTA, 2023).

Cada modelo possui seu conjunto de parâmetros. Na Regressão Linear os parâmetros são os coeficientes dos preditores, em uma rede neural, os parâmetros são os pesos e desvios dos nós e nos modelos baseados em árvores são os pontos de divisão e os critérios de divisão em cada nó. Na aprendizagem de máquina, os hiperparâmetros são parâmetros cujos valores são definidos antes do início do processo de aprendizagem, esses parâmetros não são aprendidos a partir dos dados e devem ser predefinidos, eles são responsáveis em auxiliar no controle do processo de aprendizagem e podem influenciar significativamente o desempenho do modelo (LÍRIO COSTA, 2024).

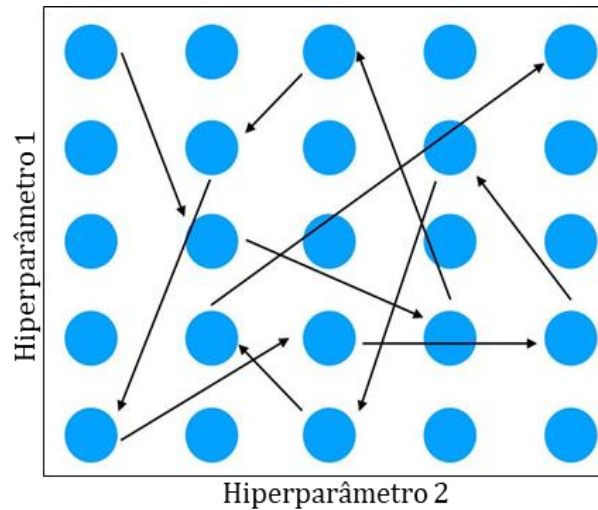
Nos modelos de RNA, os hiperparâmetros podem incluir a taxa de aprendizagem, o número de camadas na rede ou o número de nós em cada camada. Nos modelos de Árvores de Decisão, a profundidade máxima da árvore é um dos hiperparâmetros mais importantes, enquanto que no modelo de Floresta Aleatória, o número de estimadores, ou seja, o número de árvores na floresta, deve ser avaliado. Os melhores valores para hiperparâmetros muitas vezes não podem ser determinados antecipadamente e devem ser encontrados por tentativa e erro (GODOY, 2018).

No trabalho de Lee (2019), é descrita a classificação de dois métodos principais para otimização de hiperparâmetros conhecidos como Pesquisa Aleatória e Pesquisa em Grade. Sabe-se que na aprendizagem de máquina, diferentes modelos são testados e os hiperparâmetros são ajustados para obter melhores previsões. A escolha do melhor modelo e hiperparâmetros são desafios que devem ser resolvidos para melhorias nas previsões. Os hiperparâmetros são ajustados escolhendo os valores ideais dos parâmetros para maior precisão, porém este processo pode ser difícil e demorado.

A Pesquisa em Grade (*Grid Search*), também conhecida como pesquisa exaustiva, examina todas as combinações possíveis de hiperparâmetros especificados. Isso significa testar cada conjunto de valores de hiperparâmetros para determinar a melhor configuração. Por outro lado, a Pesquisa Aleatória (*Random Search*) emprega combinações aleatórias de hiperparâmetros, conforme sugerido pelo próprio nome (Figura 12). Ao contrário da pesquisa em grade, não são testados

todos os valores dos parâmetros. Em vez disso, os parâmetros são amostrados com base em um número fixo de iterações fornecido pelo usuário.

Figura 12 – Método de otimização de hiperparâmetros por Pesquisa Aleatória (*Random Search*)

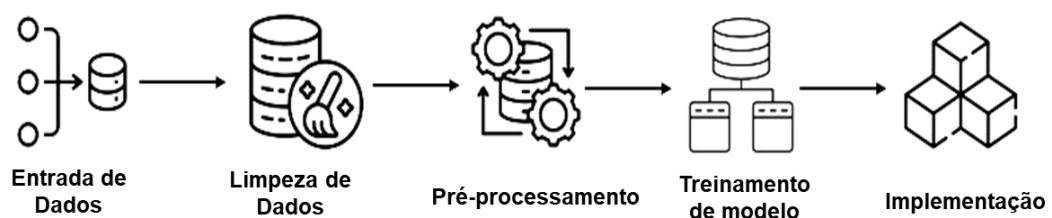


Fonte: adaptado de FABIEN, 2019

4.8 PIPELINE DE APRENDIZAGEM DE MÁQUINA

Um pipeline de AM é uma sequência de etapas ordenadas e interconectadas que são aplicadas a dados para desenvolver e implementar modelos de machine learning. Esse pipeline abrange desde a preparação inicial dos dados até a implantação e monitoramento do modelo em um ambiente de produção. Na Figura 13 pode ser observada uma visão geral das etapas comuns em um pipeline de aprendizagem de máquina (ARRAIS, 2022).

Figura 13 – Construção de um Pipeline de Aprendizagem de Máquina



Fonte: adaptado de PRASAD, 2023

4.8.1 Obtenção de dados e análise exploratória

Após a formulação de um problema, segundo Pereira (2022) a primeira etapa de um fluxo de trabalho de aprendizagem de máquina é a obtenção de dados, envolvendo a coleta e o armazenamento dos dados em sua forma bruta. Em seguida, segue-se a etapa de *feature engineering*, na qual os dados brutos passam por vários tipos de processamento para extrair informações valiosas. Esta etapa consiste em preparar os dados para análise. Isso pode incluir a limpeza dos dados para remover valores ausentes ou inconsistentes (*outliers*), normalização e transformações para garantir que todas as variáveis tenham a mesma escala (MCKINNEY, 2012).

Esta etapa inicial é considerada fundamental para garantir a consistência dos dados desde o início do processo de análise já que os dados podem ser obtidos de diversas fontes, como bancos de dados, sensores, arquivos de texto, imagens ou vídeos, pois dados mal preparados podem levar a resultados imprecisos ou enviesados (HASTIE et al., 2009). Portanto, o pré-processamento e a análise cuidadosa dos dados são etapas essenciais para garantir um bom desempenho.

Um conjunto de dados é organizado em uma tabela onde cada coluna corresponde a uma variável preditora (*feature*) e cada linha corresponde a um registro. A coluna que corresponde à variável de resposta (neste trabalho é o rendimento) é chamada de rótulo ou *label*. Na Tabela 2, é mostrada a organização do conjunto de dados.

Tabela 2 – Dados com label, m features, e n dados

Feature 1	Feature 2	. . .	Feature n	Label
X_{11}	X_{12}	. . .	X_{1m}	y_1
X_{21}	X_{22}	. . .	X_{2m}	y_2
.
.
.
X_{n1}	X_{n2}	. . .	X_{nm}	y_n

Fonte: elaborado pelo Autor, 2024

4.8.2 Seleção e treinamento do modelo

Após a preparação dos dados, a etapa subsequente envolve a escolha de um algoritmo de aprendizagem de máquina mais adequado para o problema em questão. Conforme explicado por James et al. (2013), essa seleção é influenciada pela natureza dos dados e pela finalidade da tarefa, que pode variar entre classificação, regressão, agrupamento (*clustering*) ou por reforço. Há uma variedade de algoritmos disponíveis para diferentes finalidades, sendo importante optar inicialmente por modelos mais simples, para verificar o quão longe é possível chegar e assim entender quais se adaptam melhor ao problema e aos dados disponíveis.

Uma vez escolhido, o algoritmo é treinado com o conjunto de dados de treino. Segundo Géron (2019), durante o treinamento o modelo ajusta seus parâmetros para minimizar uma função de perda ou maximizar uma função de recompensa, dependendo do tipo de problema. O objetivo é encontrar os parâmetros que melhor descrevem os padrões nos dados de treinamento, de modo que o modelo possa fazer previsões precisas ou tomar decisões corretas quando apresentado a novos dados.

O algoritmo pode ser treinado usando o método da divisão treino-teste (*hold-out validation*) ou validação cruzada (*cross validation*). O primeiro método tem uma abordagem mais simples. Nesta técnica, uma parte dos dados é reservada para treinar o modelo, enquanto a outra parte é usada para avaliar seu desempenho. Esta divisão é geralmente feita em uma proporção, como 70-30 ou 80-20, dependendo do tamanho do conjunto de dados e da complexidade do problema e é amplamente adotada devido à sua simplicidade e eficácia (MÜLLER; GUIDO, 2016).

Há diferentes tipos de validação cruzada, porém a técnica mais comum é conhecida como *k-fold cross validation*, que consiste em dividir o conjunto de dados em várias amostras, chamadas de *folds*. Este método envolve treinar o modelo em subconjuntos distintos e avaliar seu desempenho de forma individual. Ao final, os resultados são combinados para obter uma estimativa mais confiável do desempenho do algoritmo, sendo especialmente útil em conjuntos de dados limitados (HASTIE et al., 2009).

4.8.3 Avaliação dos modelos

Após o treinamento do modelo é necessário que ele passe por uma avaliação, para logo ajustá-lo conforme necessário. Nesta etapa, são utilizados os resultados da avaliação e os ajustes são realizados visando melhorar o desempenho do modelo (MÜLLER; GUIDO, 2016). Para a avaliação são utilizadas algumas métricas como o Erro Percentual Absoluto Médio (MAPE), Erro Quadrático Médio (RMSE), Erro Absoluto Médio (MAE) e o Coeficiente de Determinação (R^2) para medir o desempenho do modelo (RASCHKA; MIRJALILI, 2019).

A partir desses resultados, podem ser realizados alguns ajustes buscando diminuir o erro de generalização. Por exemplo, pode-se otimizar os hiperparâmetros, selecionar as variáveis mais relevantes ou até escolher outro algoritmo (BIGOTO; ELEN, 2020). Após isso, os dados são novamente treinados, sendo o processo repetido algumas vezes até encontrar o melhor arranjo dos hiperparâmetros (BRINK; RICHARDS; FETHEROLF, 2016).

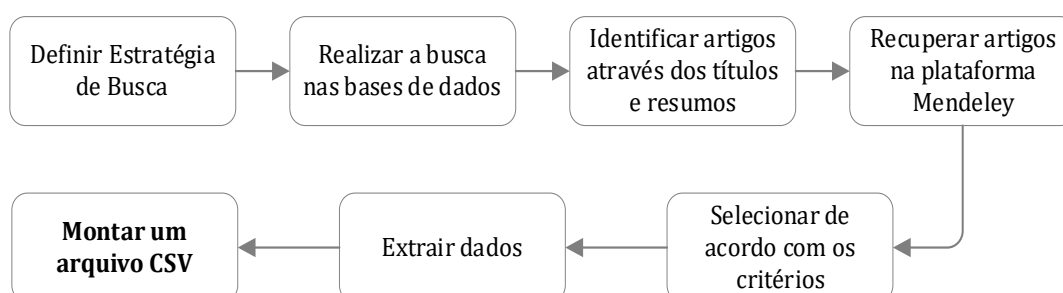
5 METODOLOGIA

Neste tópico é descrita a metodologia empregada no trabalho; apresentando as características do conjunto de dados, as etapas de pré-processamento, a análise descritiva dos dados, a avaliação de desempenho dos modelos de aprendizagem de máquina e por último a validação e análise dos parâmetros da reação de transesterificação, por meio de um modelo treinado.

5.1 OBTENÇÃO DOS DADOS

A obtenção de dados foi feita a partir de uma pesquisa bibliográfica exploratória, utilizando um modelo sistemático através da busca no Portal de Periódicos Capes e pelos bancos de dados Science Direct, SCOPUS e SciELO. O processo de revisão sistemática utilizado propõe algumas avaliações, no momento da obtenção de dados (Figura 14).

Figura 14 – Processo de revisão sistemática para a obtenção de dados



Fonte: elaborado pelo Autor, 2024

Inicialmente, foram utilizadas palavras-chave nas bases de dados mencionadas para a busca de artigos. As palavras-chaves selecionadas foram: "*optimization*" and "*biodiesel*", "*biodiesel*" and "*heterogeneous catalysis*", "*biodiesel yield*" and "*Transesterification Biodiesel*". Os trabalhos de pesquisa realizados no campo da produção de biodiesel foram identificados através dos títulos e resumos e gerenciados utilizando o software Mendeley.

Para a delimitação da busca, foram considerados apenas os artigos que utilizavam catalisadores heterogêneos com presença de CaO (Óxido de Calcio) na sua composição e metanol como álcool. Além disso, os trabalhos abordados continham informações das variáveis no processo de produção, sendo estas: o

tempo, a temperatura, o óleo utilizado, a proporção de óleo e álcool, o catalisador, peso do catalisador e o rendimento obtido. Estas variáveis são parâmetros que influenciam diretamente o rendimento final na produção de biodiesel.

Após a seleção, um arquivo CSV foi criado contendo uma tabela na qual cada linha representa os experimentos dos artigos acadêmicos selecionados, levando em consideração os critérios mencionados. As colunas incluem os parâmetros (*features*) e o rendimento (*label*), como demonstrado na Tabela 3.

Tabela 3 – Modelo da tabela com dados genéricos para o banco de dados

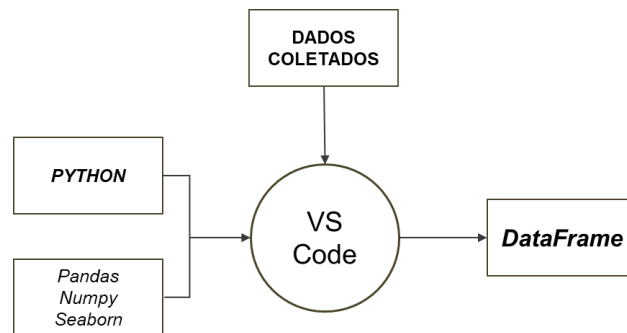
FEATURES						LABEL
Óleo Utilizado	Cat.	Peso de Cat. (%m/m)	RE (met./óleo)	Temperatura (°C)	Tempo (min)	Rendimento (%)
Óleo de Fritura	CaO_puro	5	3	50	60	75
Azeite de Oliva	CaO_não puro	4	6	65	80	68

Fonte: elaborado pelo Autor, 2024

A divisão de CaO puro e não puro se deu pela composição do catalisador, os puros foram classificados em aqueles catalisadores que possuem uma alta quantidade de óxido de cálcio e com menores quantidades de impurezas e os não puros aqueles que tem maior quantidade de impurezas.

5.2 PRÉ-PROCESSAMENTO E ANÁLISE EXPLORATÓRIA DOS DADOS

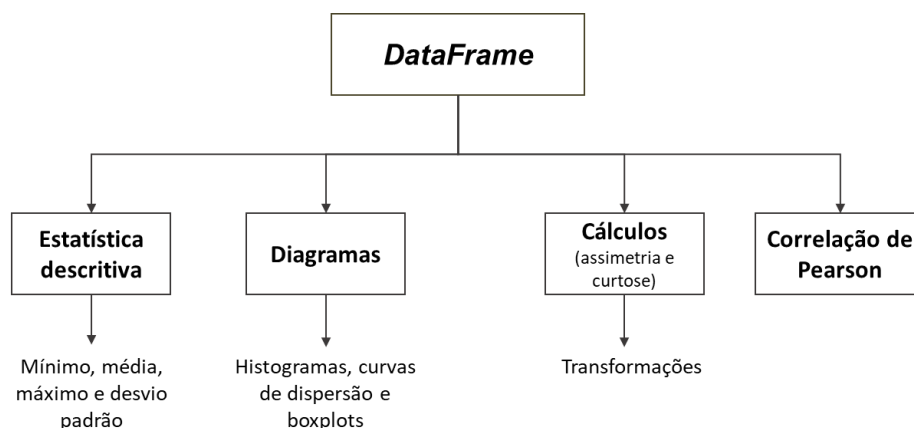
Para esta etapa e as seguintes, foi essencial o uso do editor de código-fonte Visual Studio Code (VSCode), utilizando a linguagem de programação Python. Dentro do editor, foram importadas inicialmente bibliotecas como: *pandas*, *numpy* e *seaborn*. Logo, os dados CSV preparados na etapa anterior foram carregados no ambiente do VSCode e transformado em um DataFrame usando a biblioteca *pandas*, como pode ser visto na Figura 15.

Figura 15 – Transformação dos dados em *dataframe* e bibliotecas utilizadas

Fonte: elaborado pelo Autor, 2024

Depois disso, foi iniciada a análise exploratória dos dados, começando pela verificação de dados nulos no banco de dados. Em seguida, foram aplicados métodos de estatística descritiva utilizando método *describe()* da biblioteca pandas. Entre as medidas calculadas, incluíram-se o mínimo, a média, o máximo e o desvio padrão para analisar os dados, conforme a Figura 16. Logo, foram construídos os diagramas de distribuição das colunas numéricas, tanto em forma de histogramas com as curvas de frequência relativa utilizando *Seaborn*, assim como também os diagramas de dispersão.

Foi realizado o cálculo da assimetria e curtose para analisar a normalidade da distribuição normal. Após isso, foram aplicadas transformações como logaritmo (log) e normalização (*min_max_scaler*). Posteriormente, foi realizada novamente a elaboração dos diagramas a partir desses dados transformados e o cálculo da assimetria e curtose.

Figura 16 – Etapas da análise exploratória de dados

Fonte: elaborado pelo Autor, 2024

Também foi necessário verificar as correlações entre os dados utilizando a correlação linear de Pearson. Estas etapas são cruciais para obter informações relevantes sobre quais variáveis têm maior influência no rendimento, o que pode ser importante na seleção dos modelos preditivos.

5.3 TREINAMENTO DOS MODELOS DE APRENDIZAGEM DE MÁQUINA

Segundo as características dos dados, foram escolhidos os modelos de Regressão Linear, Floresta Aleatória e o MPL Regressor (*Multi-layer Perceptron Regressor*) da biblioteca *scikit-learn* para o treinamento. Nesta etapa, optou-se em usar Regressão Linear como linha de base, pois é comum usar um modelo simples como ponto de partida para análise, pois permite avaliar o desempenho de modelos mais complexos (PRAKASH, 2023).

Conforme discutido por Agrawal e Dhawane (2024), a Floresta Aleatória é um modelo eficaz para lidar com conjuntos de dados multidimensionais, bem como com variáveis categóricas e numéricas. Outra vantagem, é sua capacidade de resistir a dados que não seguem uma distribuição normal, além de funcionar bem na presença de outliers. De acordo com James et al. (2013), se as variáveis se correlacionam de maneira mais complexa e tendendo a modelos não-lineares, as Árvores de Decisão poderão melhor suprir as demandas.

No estudo de biocombustíveis, diversas pesquisas têm utilizado técnicas de aprendizagem de máquina, em particular as Redes Neurais Artificiais (RNA), como abordagem científica para a produção de biodiesel, demonstrando eficácia. Sua habilidade na modelagem é principalmente atribuída à capacidade de lidar com conjuntos de dados não lineares. Os modelos baseados em RNA permitem a representação de sistemas multivariados dinâmicos sem a necessidade de compreensão dos princípios subjacentes envolvendo processos complexos, como os sistemas de produção de biodiesel (PEREIRA, 2022).

Para o treinamento dos modelos, foi utilizado tanto o conjunto de dados brutos e também os dados transformados, isto com o objetivo de comparar e avaliar se os modelos tiveram melhorias com as transformações realizadas. Optou-se pelo método de *hold-out validation* para o treinamento, dividindo os dados em 75% para

treinamento e 25% para teste. Na etapa de avaliação, utilizou-se a métrica MAE para avaliar o desempenho dos modelos. O erro absoluto médio (MAE) mede o quão perto as previsões estão dos resultados eventuais. O erro absoluto é a diferença entre o valor predito e o valor “verdadeiro”. O MAE é a média de todos os erros absolutos. É calculado como:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_0| \quad (5.1)$$

Onde n é o número de amostras, y_i o valor real e y_0 o valor predito.

5.4 AJUSTE E SELEÇÃO DOS MODELOS

Para aprimorar o desempenho dos modelos selecionados, utilizou-se a técnica de seleção de hiperparâmetros *RandomizedSearchCV*, uma técnica de otimização de hiperparâmetros da biblioteca *scikit-learn*. Os modelos escolhidos para a otimização de hiperparâmetros foram Floresta Aleatória e MLP Regressor. Foram realizadas várias rodadas de ajuste utilizando duas grades que definem o espaço de hiperparâmetros para a pesquisa, uma mais restrita e outra mais abrangente. Em cada rodada de ajuste, variou-se o número de iterações e armazenados os melhores hiperparâmetros obtidos.

Para ajustar os modelos de Floresta Aleatória, foram conduzidas quatro rodadas na primeira grade, com diferentes números de iterações: 10, 50, 500 e 1000, respectivamente. Na segunda grade, foram realizadas sete rodadas, utilizando iterações nos valores de 10, 50, 100, 300, 500, 1000 e 2880. As grades utilizadas na Floresta Aleatória podem ser observadas na Tabela 4.

Tabela 4 – Grades de busca utilizadas para a Floresta Aleatória

GRADE 1	GRADE 2
n_estimators: [10, 50, 100, 200] max_depth: [None, 10, 20, 30] min_samples_split: [2, 5, 10] min_samples_leaf: [1, 2, 4], random_state: [42]	n_estimators: [10, 50, 100, 200, 500] max_depth: [None, 10, 20, 30, 40, 50] Min_samples_split: [2, 5, 10, 15], min_samples_leaf: [1, 2, 4, 6], max_features: ['auto', 'sqrt', 'log2'] bootstrap: [True, False] random_state: [42]

Fonte: elaborado pelo Autor, 2024

Para o modelo MLP Regressor, foram realizadas cinco rodadas na primeira grade, com iterações variando entre 10, 50, 100, 500 e 1000. Na segunda grade, foram realizadas seis rodadas, com iterações nos valores de 10, 50, 100, 500, 1000 e 5000. As grades utilizadas no MPL Regressor podem ser observadas na Tabela 5.

Tabela 5 – Grades de busca utilizadas para a MLP Regressor

GRADE 1	GRADE 2
hidden_layer_sizes: [(50,50,50),(50,100,50), (100,)]	hidden_layer_sizes: [(50,50,50),(50,100,50),(100,),(150,),(200,)]
activation: ["identity", "logistic", "tanh", "relu"]	activation: ["identity", "logistic", "tanh", "relu"]
solver: ["lbfgs", "sgd", "adam"]	solver: ["lbfgs", "sgd", "adam"]
alpha: [0.0001, 0.05]	alpha: [0.0001, 0.001, 0.01, 0.05, 0.1, 0.5]
learning_rate: ["constant", "invscaling", "adaptive"]	learning_rate: ["constant", "invscaling", "adaptive"]
max_iter: [100, 200, 300]	learning_rate_init: [0.001, 0.01, 0.1]
random_state: [42]	power_t: [0.5, 0.6, 0.7, 0.8, 0.9]
	max_iter: [100, 200, 300, 400, 500]
	shuffle: [True, False]
	random_state: [42]
	tol: [0.0001, 0.001, 0.01]
	momentum: [0.9, 0.8, 0.7, 0.6, 0.5]
	nesterovs_momentum: [True, False]
	beta_1: [0.9, 0.8, 0.7, 0.6, 0.5]
	beta_2: [0.999, 0.99, 0.9, 0.8, 0.7]

Fonte: elaborado pelo Autor, 2024

Cada conjunto de hiperparâmetros obtidos nas rodadas, foram utilizados para treinar o modelo, cujas previsões foram avaliadas com a métrica MAE. O conjunto com a melhor pontuação MAE para cada modelo foi escolhido. Finalmente, a Regressão Linear, a Floresta Aleatória e o MLP Regressor, foram comparados utilizando diferentes métricas para escolher o modelo que apresenta melhor desempenho. Essa comparação foi realizada através da Raiz Quadrada do Erro Médio (RMSE), o Coeficiente de Determinação (R^2) e o Erro Percentual Absoluto Médio (MAPE). Nesta última etapa, também foi realizada a comparação com o modelo de linha base que é a Regressão Linear.

O RMSE mede as diferenças entre os valores previstos por um modelo e os valores realmente observados e ele é definido por:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_0)^2} \quad (5.2)$$

Onde n é o número de amostras, y_i o valor real e y_0 o valor predito.

O MAPE fornece uma medida da precisão média das previsões em termos percentuais, sendo útil para interpretar a magnitude dos erros em relação aos valores reais, sendo definido por:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - y_0}{y_i} \right| \quad (5.3)$$

Onde n é o número de amostras, y_i o valor real e y_0 o valor predito.

O R^2 é o índice mais importante para verificar a precisão do resultado previsto de um algoritmo de regressão, cujo intervalo é de $[0, 1]$, ele é definido por:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_0)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (5.4)$$

Onde n é o número de amostras, y_i o valor real e y_0 o valor predito.

De acordo com Cai et al. (2020), normalmente quanto menor os valores de MAE, RMSE e MAPE os resultados do ajuste são melhores. Já para o R^2 , quando o resultado é igual a 1, o modelo de regressão dá previsões sem qualquer erro. Em geral, quanto maior o valor de R^2 , melhor será o resultado do ajuste.

5.5 VALIDAÇÃO E ANÁLISE DOS PARÂMETROS DA REAÇÃO DE TRANSESTERIFICAÇÃO

Com a escolha do modelo, foram realizados dois experimentos de validação, nos quais foram utilizados dados experimentais sobre a produção de biodiesel. Um dos artigos empregou o CaO puro como catalisador (VAZ, 2015), enquanto o outro utilizou um CaO não puro (TANG et al., 2013), no entanto, ambos os estudos tiveram o óleo de soja como matéria-prima. Esses dados experimentais, foram então comparados com as previsões feitas pelo modelo treinado, com o objetivo de avaliar sua capacidade de generalização.

Por outro lado, foram conduzidas três análises para avaliar diferentes parâmetros do biodiesel, utilizando o modelo com maior precisão. As condições de reação utilizadas e os parâmetros analisados para cada análise são descritos na Tabela 6.

Tabela 6 – Condições de reação utilizadas para a análise de parâmetros

Temperaturas = 55°C e 65°C	
Análise 1 Influência da temperatura	<i>Condições de reação:</i> Fonte: Óleo de cozinha, Tipo de catalisador: CaO não puro, RE met./óleo: 9:1, Peso cat.: 5% m/m
Pesos de cat. = 2% m/m e 5% m/m	
Análise 2 Influência do peso de catalisador	<i>Condições de reação:</i> Fonte: Óleo de cozinha, Tipo de catalisador: CaO não puro, Temperatura: 65°C, RE met./óleo: 9:1
RE metanol/óleo = 3:1 e 6:1	
Análise 3 Influência da RE metanol/óleo	<i>Condições de reação:</i> Fonte: Óleo de cozinha, Peso cat.: 2% m/m, Temperatura: 65°C, Tipo de catalisador: CaO não puro

Fonte: elaborado pelo Autor, 2024

6 RESULTADOS E DISCUSSÃO

Nesta seção serão apresentadas os principais resultados e discussões deste trabalho.

6.1 REVISÃO SISTEMÁTICA DE LITERATURA

Utilizando a estratégia de busca descrita na metodologia, foram identificados 25 artigos acadêmicos relacionados à produção de biodiesel por catálise heterogênea. A partir da análise destes trabalhos, foram coletados 297 conjuntos de dados, publicados em 13 artigos diferentes. Esses conjuntos de dados continham, sempre que possível, informações sobre o tipo de óleo utilizado, temperatura, tempo, álcool empregado, proporção molar entre o álcool e o óleo, catalisador utilizado, peso do catalisador e rendimento da formação de biodiesel.

Considerando que a coleta dos dados dependeu da publicação pelos autores dos trabalhos encontrados na literatura, é importante ressaltar que não há na literatura um padrão uniforme. Ou seja, nem todos os trabalhos apresentam os mesmos dados reacionais. Dessa forma, os artigos selecionados deveriam incluir informações sobre o tipo de catalisador utilizado, indicando se era CaO puro ou não puro, e o álcool utilizado deveria ser o metanol. Experimentos nos quais foram utilizados o etanol como álcool foram desconsiderados, uma vez que a quantidade de dados encontrados foi consideravelmente menor em comparação ao metanol.

Após esse processo de seleção, o número de dados foi reduzido para 205, provenientes de 10 artigos acadêmicos, como pode ser verificado na Tabela 7. Essa redução reflete a necessidade de garantir uma maior consistência nos dados analisados. Por exemplo, observou-se que o uso do metanol como álcool para esta reação é bastante comum, havendo uma quantidade limitada de dados envolvendo outros tipos de álcoois.

No que diz respeito aos catalisadores, foi observada uma gama mais ampla de possibilidades, sendo o CaO um dos mais estudados. Portanto, os modelos construídos com base nesses dados são restritos ao uso do metanol como álcool e ao CaO como catalisador. Essa seleção mais rigorosa pode limitar a

generalização dos modelos, mas contribui para a obtenção de resultados mais confiáveis, o que é crucial para o desenvolvimento da pesquisa.

Tabela 7 – Artigos acadêmicos selecionados para a coleta de dados

Autores	Revista	Tipo de óleo	Tipo de catalisador	Dados coletados
Bedir; Doğan, 2021	Fuel	Óleo de Girassol	CaO puro	17
Liu et al., 2021	Fuel	Óleo de Girassol	CaO puro	14
Gohain et al., 2020	Renewable Energy	Óleo de cozinha	CaO puro	14
Huang et al., 2021	Fuel	Óleo de Soja	CaO puro	25
Laskar et al., 2020	Renewable Energy	Óleo de Soja	CaO não puro	29
Kataria; Mohapatra; Kundu, 2019	Journal of the Energy Institute	Óleo de cozinha	CaO não puro	20
Khatibi; Khorasheh; Larimi, 2021	Renewable Energy	Óleo de Canola	CaO não puro	15
Basumatary et al., 2021	Fuel	Óleo de Canola	CaO não puro	12
Lani; Ngadi; Inuwa, 2020	Renewable Energy	Óleo de cozinha	CaO não puro ¹	20
Lani; Ngadi; Inuwa, 2020	Renewable Energy	Óleo de cozinha	CaO não puro ²	20
Abdul Mutalib et al., 2020	Bioenergy Research	Óleo de Palma	CaO não puro	19

Fonte: elaborado pelo Autor, 2024

Nota: O CaO não puro¹ e CaO não puro² referem-se a que os catalisadores foram produzidos por via seca e úmida, respectivamente.

6.2 PRÉ-PROCESSAMENTO DE DADOS E ANÁLISE EXPLORATÓRIA

Na etapa do pré-processamento, foi verificada a ausência de dados nulos e a imputação não foi necessária. Essa ausência é devido aos critérios de seleção definidos na etapa de obtenção de dados. Dessa forma, foi possível garantir a qualidade dos dados desde sua origem, eliminando a possibilidade de lidar com quaisquer inconsistências ou lacunas, o que contribuiu na robustez e precisão das análises realizadas nas etapas subsequentes da pesquisa.

Foram avaliados um total de 205 dados que atendiam aos critérios estabelecidos, onde a fonte do óleo e o tipo de catalisador são variáveis categóricas. Foram identificados 70 dados com a utilização de CaO puros, enquanto os demais foram classificados como CaO não puros. Na Tabela 8, encontram-se as estatísticas descritivas dos dados utilizados no estudo, permitindo uma compreensão detalhada da distribuição dos mesmos.

Tabela 8 – Estatística descritiva dos dados

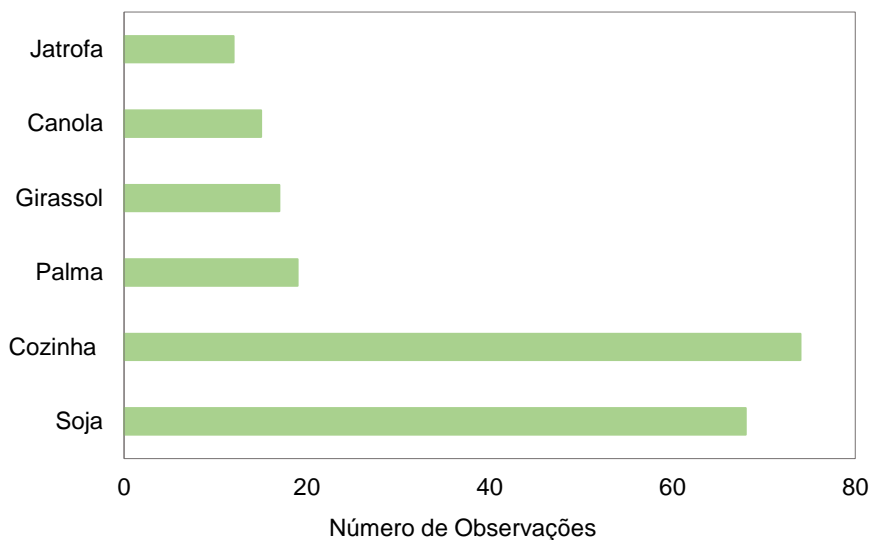
Parâmetros	Mínimo	Média	Máximo	Desvio Padrão
Fonte de óleo	**	**	**	**
Tipo de catalisador	**	**	**	**
Peso do Cat. (m/m)	1	4,89	15	2,37
RE metanol/óleo	3	12,20	25	5,44
Temperatura (°C)	25	57,89	100	15,81
Tempo (min)	10	136,51	300	67,75
Rendimento (%)	4	76,79	98	21,32

Fonte: elaborado pelo Autor, 2024

Nota: os símbolos ** referem- se às variáveis categóricas.

A partir da tabela, verifica-se que o peso do catalisador, expresso em porcentagem em massa (m/m) em relação ao óleo, variou de 1 a 15, com uma média de 4,89 e um desvio padrão de 2,37. A razão estequiométrica entre metanol e óleo na reação (RE metanol/óleo) apresentou os seguintes valores: 3; 12,20; 25 e 5,44, correspondentes aos valores mínimos, médios, máximos e desvios padrão, respectivamente. Quanto à temperatura de reação, variou de 25 a 100 °C, com uma média de 57,89 e um desvio padrão de 15,81. O tempo de reação, por sua vez, variou de 10 a 300 minutos, com uma média de 136,51 e um desvio padrão de 67,75. Por fim, o rendimento do processo de produção de biodiesel, expresso em porcentagem, apresentou valores mínimos, médios, máximos e desvios padrão de 4, 76,79, 98 e 21,32, respectivamente. Esses dados proporcionam uma visão abrangente dos parâmetros envolvidos no processo, incluindo suas distribuições e variabilidade.

Dentre os óleos utilizados na análise, encontram-se o óleo de soja, óleo de girassol, óleo de canola e o óleo de palma, tradicionalmente conhecido como óleo de dendê, todos eles utilizados na produção de biodiesel de primeira geração. Além desses, também foram analisados o óleo de pinhão-mansão (jatrofa) e o óleo de cozinha residual, utilizados na produção de biodiesel de segunda e terceira geração, respectivamente. Dos dados coletados, foi observado que o óleo de soja teve 68 ocorrências, enquanto o óleo de cozinha residual apresentou 74 ocorrências, indicando uma frequência maior deste último, conforme observado na Figura 17.

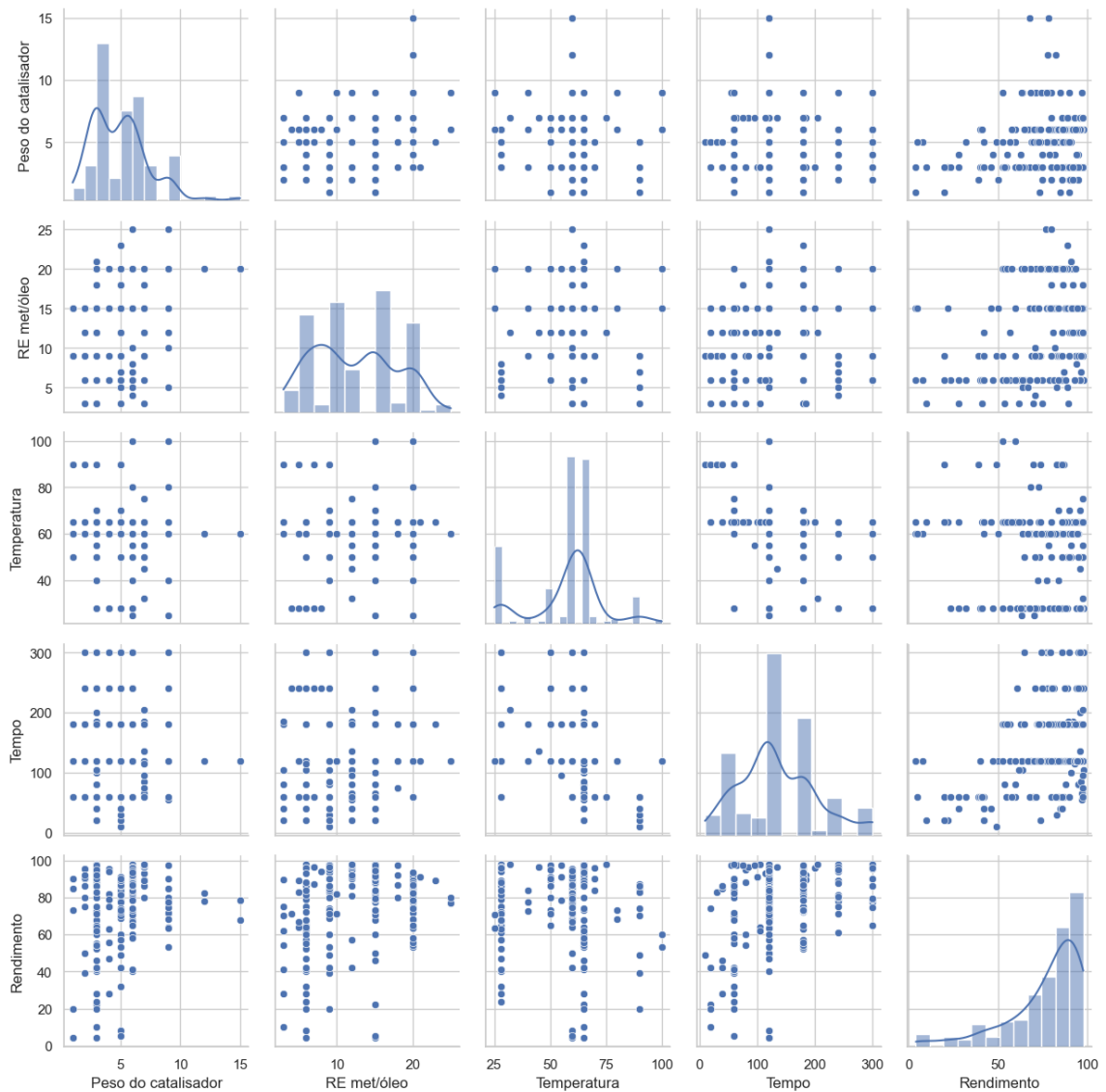
Figura 17 – Fontes de óleos encontrados nos dados coletados.

Fonte: elaborado pelo Autor, 2024

Os histogramas das variáveis numéricas ilustrada na Figura 18, oferece uma visão abrangente da distribuição dos dados. A análise revela que os dados não seguem uma distribuição normal, dada as características assimétricas dos histogramas. No entanto, destaca-se que o tempo e a temperatura apresentam distribuições mais próximas da normalidade, como evidenciado pelas curvas de frequência relativa.

Com relação aos gráficos de dispersão, observa-se uma estratificação dos dados no gráfico, em vez de uma distribuição aleatória. Essa estratificação pode ser atribuída à natureza dos experimentos descritos nos artigos, nos quais os dados foram coletados. Por exemplo, muitos dos experimentos relatados compartilham condições de reação semelhantes. É comum encontrar que a razão estequiométrica utiliza valores específicos, como 3, 6, 9 e 12, enquanto para o peso do catalisador, são frequentemente encontrados valores como 1, 3, 5 e 6% m/m. Esse padrão de valores recorrentes pode resultar na estratificação dos dados nos gráficos de dispersão, ao invés de uma distribuição mais aleatória.

Figura 18 – Histogramas e diagramas de dispersão das variáveis numéricas

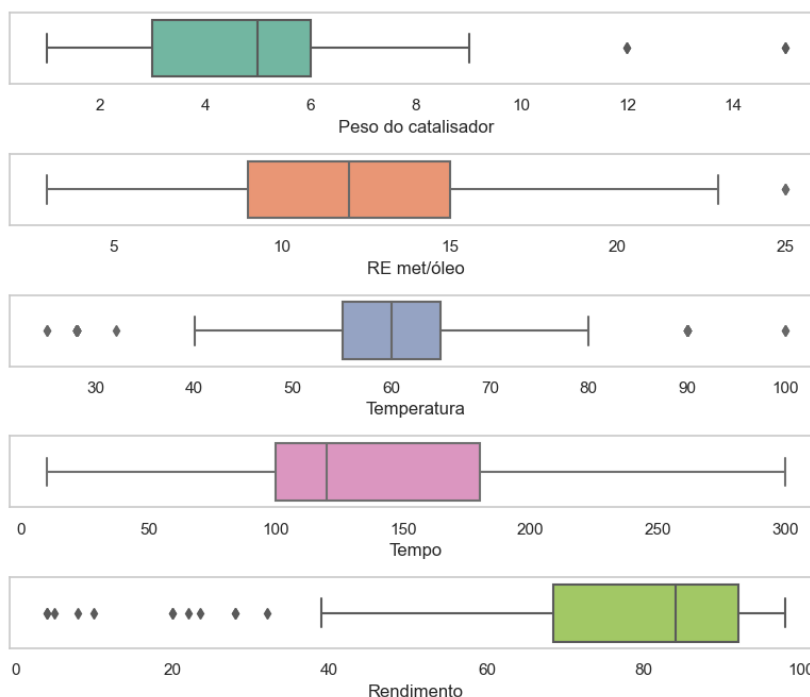


Fonte: elaborado pelo Autor, 2024

Os boxplots da Figura 19, fornecem uma representação visual da distribuição de dados de cada variável numérica, que servem para verificar as tendências, detectar dados discrepantes (*outliers*) e comparar as distribuições para entender como os dados se comparam entre si. As extremidades das caixas no gráfico horizontal representam os quartis. O quartil 1 está localizado à esquerda da caixa e o quartil 3 à direita. Esses quartis oferecem uma visão geral da concentração dos dados e a linha no interior da caixa corresponde à mediana ou valor central do conjunto de dados. Os limites superior e inferior são as linhas que se estendem além das caixas do boxplot. Eles representam os extremos do conjunto de dados e são calculados a partir dos quartis e utilizando o intervalo interquartil, valores fora desses

extremos são considerados como discrepantes. O limite superior corresponde ao maior valor do conjunto de dados, enquanto o limite inferior corresponde ao menor valor (SUN; GENTON, 2011).

Figura 19 – Boxplots das variáveis numéricas



Fonte: elaborado pelo Autor, 2024

Foi identificada uma assimetria negativa na distribuição dos dados do peso do catalisador, através da mediana. Além disso, observou-se uma maior dispersão em valores abaixo de 5% m/m, com uma variação entre 3 e 6% m/m. Os valores extremos do peso do catalisador estão compreendidos entre 1 e 9% m/m, sendo que valores acima de 9% m/m são considerados discrepantes (*outliers*). Neste caso, foram identificados como *outliers* os valores de 12 e 15% m/m.

Da mesma forma, foi constatado que a distribuição do conjunto de dados da razão estequiométrica entre o metanol e óleo apresenta uma distribuição mais simétrica. Além disso, a maioria dos dados está distribuído no intervalo de 9 e 15 com mediana igual a 12, sendo que as extremidades inferior e superior são 3 e 23, respectivamente. Neste conjunto de dados há apenas um indício de valores discrepantes. Na temperatura, também há uma distribuição semelhante, ou seja, simétrica, além disso os dados são mais concentrados. A maior concentração está

na faixa dos 55 e 65 °C, com mediana em 60 °C. Os valores extremos são 40 e 80 °C, no entanto esta variável possui valores discrepantes, tanto abaixo do limite inferior entre 20 e 40 °C, como também acima do limite superior, entre 80 a 100 °C.

Para o tempo, observa-se que o conjunto de dados possui uma assimetria positiva, com maior dispersão em valores acima de 120 minutos, variando entre 100 e 180 minutos. Além disso, os valores extremos do tempo estão compreendidos entre 10 e 300 minutos, sem indícios de valores discrepantes. A análise da distribuição dos dados revela uma amplitude maior na distribuição de dados do tempo, resultando em uma variação de escala dos dados em comparação com as demais variáveis. Essa diferença pode prejudicar o desempenho de alguns algoritmos de aprendizagem de máquina que são sensíveis à escala dos dados.

Por último, para o rendimento foi constatado que a distribuição tem assimetria negativa e por meio da mediana, observou-se uma maior dispersão em valores abaixo 85% com uma variação entre 38 e 92%. Os valores extremos do rendimento estão compreendidos entre 38 e 98%, sendo que valores abaixo de 38% são considerados discrepantes (*outliers*).

Assim, decidiu-se que os dados do peso do catalisador, da razão estequiométrica do óleo e álcool e o tempo de reação deveriam passar por transformações, visando melhorar o desempenho dos modelos a serem utilizados. Para corrigir a distribuição dos dados, no caso do peso do catalisador e da razão estequiométrica do óleo e álcool, foi aplicado o logaritmo nas respectivas colunas, enquanto a técnica MinMax foi utilizada para normalizar a coluna do tempo de reação.

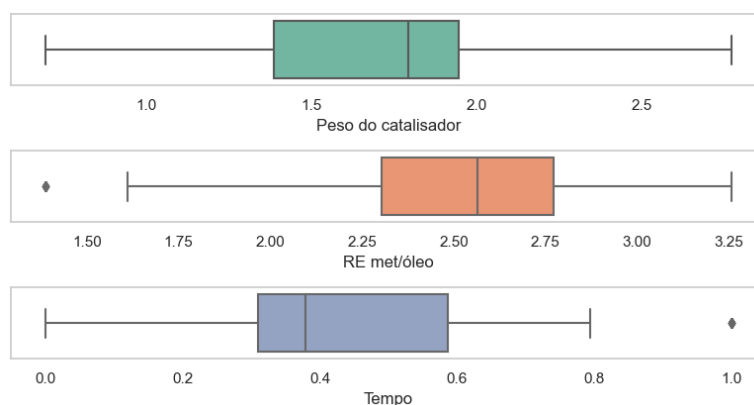
Na Tabela 9, observa-se que a aplicação do logaritmo corrigiu a assimetria e a curtose das colunas transformadas. Os valores próximos de zero após a transformação indicam uma tendência para uma distribuição mais simétrica. Este resultado sugere que a transformação logarítmica foi eficaz em corrigir as distorções presentes na distribuição dos dados. Além disso, conforme observado no boxplot da Figura 20, a normalização do tempo foi eficaz, resultando em uma escala entre 0 e 1. O processo de transformação dos dados também foi capaz de eliminar a maioria dos valores discrepantes.

Tabela 9 – Assimetria e curtose dos dados após as transformações

Variáveis	Antes da Correção		Após a Correção	
	Assimetria	Curtose	Assimetria	Curtose
Peso do Cat. (% m/m)	1,16	2,41	-0.06	-0.065
RE metanol/óleo	0,2	-1,06	-0.47	-0,58

Fonte: elaborado pelo Autor, 2024

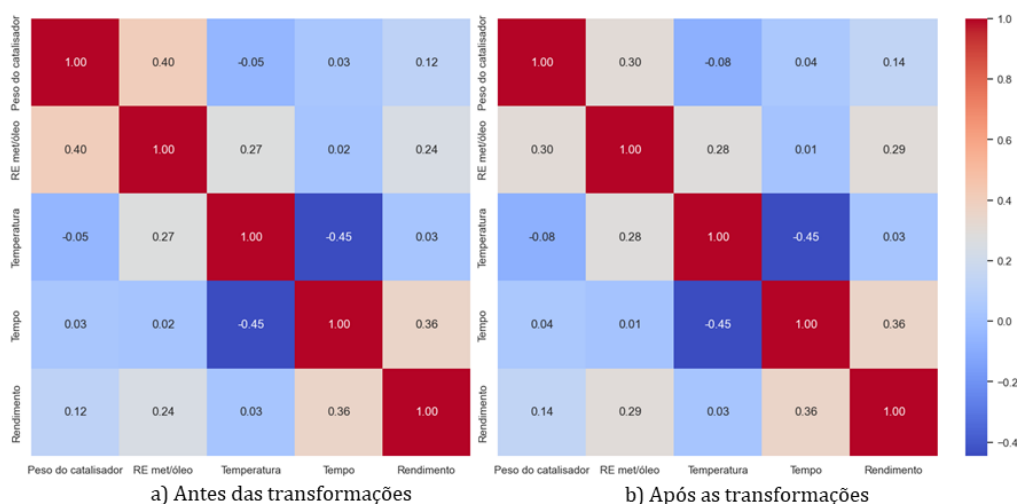
Figura 20 – Boxplots das variáveis numéricas após as transformações



Fonte: elaborado pelo Autor, 2024

A correlação de Pearson foi utilizada para examinar a relação que há entre as variáveis preditoras, presumindo uma relação linear (Figura 21). Esta medida serve para diagnosticar a presença de multicolinearidade entre as variáveis e para analisar a importância dos *features*.

Figura 21 – Mapa de calor e correlações de Pearson antes e depois das transformações



Fonte: elaborado pelo Autor, 2024

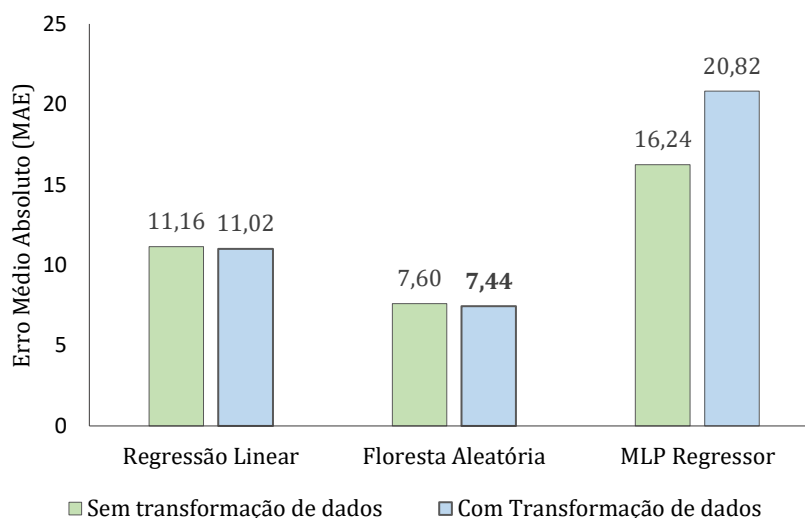
Analisando os dados da correlação linear de Pearson, observa-se um leve aumento nas correlações após as transformações, tanto as correlações entre o rendimento e o peso do catalisador, quanto ao rendimento e à razão estequiométrica entre metanol e óleo. No entanto, a força dessas correlações não foi afetada após as transformações da natureza dos dados.

Além disso, é possível identificar correlações fracas e moderadas entre as variáveis preditoras. Entre o peso do catalisador e a razão estequiométrica, encontramos uma correlação moderada ($r = 0,40$); para a razão estequiométrica e a temperatura ($r = 0,27$) há uma correlação fraca. A análise da temperatura e do tempo revela uma correlação linear moderada, porém negativa ($r = -0,45$). Em outras palavras, à medida que a temperatura aumenta, o tempo de reação diminui, indicando que há uma relação inversa entre essas variáveis. Esse fenômeno ocorre devido ao aumento na temperatura que acelera a taxa de reação, uma vez que este processo é endotérmico. Portanto, elevar a temperatura permite que a reação atinja a mesma conversão em um período de tempo mais curto (BARROS; JARDINE, 2021).

As correlações lineares entre o rendimento e as variáveis numéricas foram todas positivas: tempo ($r = 0,36$), temperatura ($r = 0,03$), RE metanol e óleo ($r = 0,24$) e peso dos catalisadores ($r = 0,12$). No entanto, apenas a correlação com o tempo indicou uma relação moderada, enquanto as demais apresentaram correlações baixas.

6.3 TREINAMENTO DE MODELOS DE AM

Após o treinamento dos modelos sem ajuste de hiperparâmetros (modelo padrão), os resultados de MAE indicam que o modelo de Floresta Aleatória apresentou maior precisão, seguido pela Regressão Linear e MLP Regressor, respectivamente conforme ilustrado na Figura 22. Quando um modelo tem um MAE baixo, significa que ele está fazendo previsões que estão mais próximas dos valores reais.

Figura 22 – Comparação do desempenho dos modelos antes e depois das transformações

Fonte: elaborado pelo Autor, 2024

Na Figura 22, podem ser observados os modelos treinados, eles sugerem que a transformação dos dados resultou em melhorias no desempenho da Regressão Linear e da Floresta Aleatória. No entanto, observou-se que o desempenho do MLP Regressor foi inferior após a transformação dos dados. Este modelo, por sua complexidade pode apresentar várias razões para um desempenho inferior após a transformação dos dados. Podem incluir desde sensibilidade à escala, como possíveis casos de *overfitting*. No entanto, a explicação mais provável é que os hiperparâmetros pré-definidos do modelo se ajustaram melhor aos dados sem transformação do que aos dados transformados. Isso ocorre porque esses valores são definidos antes do treinamento e podem não ser adequados para os novos dados após a transformação.

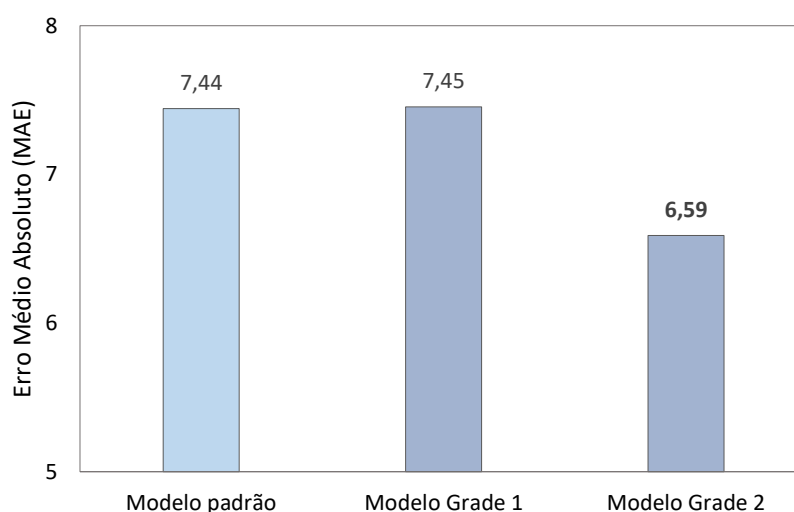
6.4 AJUSTE DE HIPERPARÂMETROS E AVALIAÇÃO DOS MODELOS

Para o ajuste dos hiperparâmetros dos modelos da Floresta Aleatória e MLP Regressor, foram utilizadas duas grades de busca no espaço de hiperparâmetros para cada modelo, de acordo com a variação do número de iterações foi obtido o melhor conjunto de hiperparâmetros, para cada grade.

Na Figura 23, é possível observar os valores de MAE para as grades 1 e 2 utilizadas na Floresta Aleatória, em comparação com o modelo padrão (sem ajuste de hiperparâmetros). Pode-se notar uma melhora significativa no desempenho do

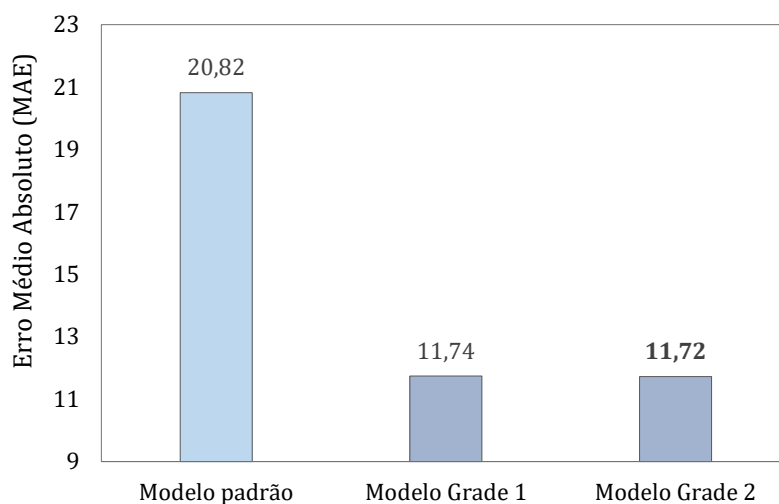
modelo ao utilizar a grade 2 para a busca do melhor conjunto de hiperparâmetros. Isso é evidenciado pelo menor valor de MAE, que foi de 6,59, em comparação com o MAE de 7,44 do modelo padrão e de 7,45 do modelo utilizando a grade 1. Esse resultado foi alcançado ao utilizar 100 iterações, resultando nos seguintes hiperparâmetros obtidos: `max_depth = 20`, `max_features = sqrt`, `n_estimators = 500`, `random_state = 42`.

Figura 23 – Comparação do desempenho da Floresta Aleatória utilizando hiperparâmetros padrão, com os modelos ajustados



Fonte: elaborado pelo Autor, 2024

Já na Figura 24, é possível observar os valores de MAE para as grades 1 e 2 utilizadas no MLP Regressor, em comparação com o modelo padrão (sem ajuste de hiperparâmetros). Se observa uma melhora significativa no desempenho do modelo nas grades 1 e 2 com valores bem próximos para a busca do melhor conjunto de hiperparâmetros. Isso é evidenciado pelo menor valor de MAE, que foi de 11,72 na grade 2 e 11,74 na grade 1, em comparação com o MAE de 20,82 do modelo padrão. Esse resultado foi alcançado ao utilizar 500 iterações para a grade 1 e 1000 para a grade 2, resultando nos seguintes hiperparâmetros obtidos: `activation = tanh`, `alpha = 0.01`, `beta_2 = 0.8`, `learning_rate_init = 0.01`, `max_iter = 500`, `momentum = 0.6`, `nesterovs_momentum = False`, `power_t = 0.7`, `random_state = 42`, `shuffle = False`, `tol = 0.001`.

Figura 24 – Comparação do desempenho do MLP Regressor utilizando hiperparâmetros padrão, com os modelos ajustados

Fonte: elaborado pelo Autor, 2024

O desempenho dos modelos foi avaliado utilizando a Regressão Linear como linha base, e comparando com os modelos de Floresta Aleatória e MLP Regressor já ajustados com os melhores hiperparâmetros, conforme demonstrado na Tabela 10. Os resultados revelam que a Floresta Aleatória teve um melhor desempenho em relação aos outros modelos.

Tabela 10 – Comparação dos modelos ajustados utilizando diferentes métricas

Métricas	Regressão Linear	Floresta Aleatória	MLP Regressor
R ²	0,264	0,63	0,048
MAE	11,129	6,59	11,72
MAPE	0,149	0,094	0,188
RMSE	13,25	9,33	15,816

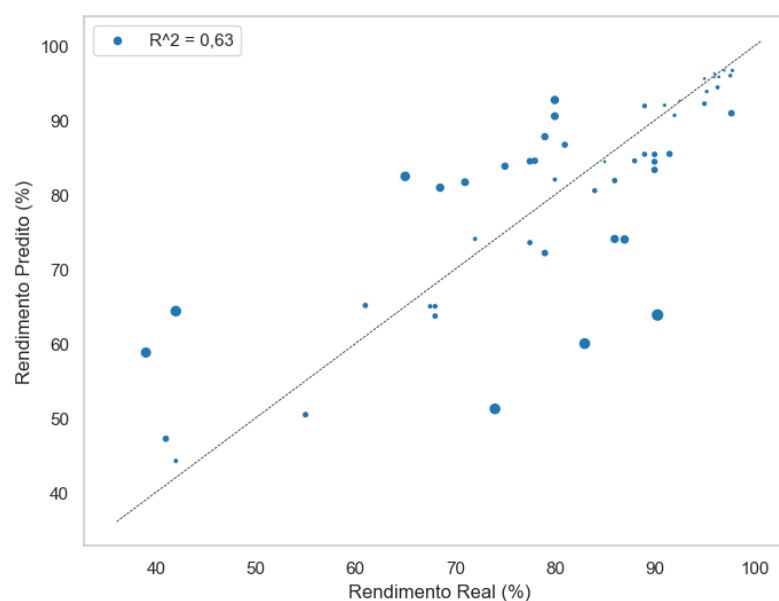
Fonte: elaborado pelo Autor, 2024

Analisando o coeficiente de determinação (R²), é possível verificar que a Floresta Aleatória obteve um valor superior, indicando sua capacidade de explicar uma porcentagem significativa da variabilidade dos dados, como pode ser observada na Figura 25. Além disso, este modelo apresentou valores menores de MAE, MAPE e RMSE, demonstrando maior precisão nas predições do rendimento de biodiesel em comparação com a Regressão Linear e MLP Regressor. Esses resultados, sugerem que a Floresta Aleatória é a abordagem mais eficaz para esse

problema específico, destacando sua utilidade na modelagem e previsão de processos relacionados à produção de biodiesel.

O desempenho entre os modelos de Regressão Linear e MLP Regressor foram comparáveis em termos de precisão, uma vez que o MAE e RMSE apresentaram valores próximos. No entanto, a Regressão Linear ainda se destaca em relação ao MLP Regressor, pois ele foi capaz de explicar uma porcentagem maior da variabilidade dos dados, alcançando um R^2 de 0,264.

Figura 25 – Gráfico de dispersão dos valores de rendimentos reais vs. preditos



Fonte: elaborado pelo Autor, 2024

A Figura 25 oferece uma representação visual das previsões do modelo de Floresta Aleatória treinado. Cada ponto no gráfico corresponde a uma observação no conjunto de dados. Além disso, é possível observar que o modelo é mais preciso para valores entre 90 e 100% de rendimento. Esse padrão pode ser atribuído à distribuição dos dados utilizados no treinamento, que possuem uma maior concentração em faixas de rendimento mais elevadas, essa distribuição de dados ocorreu porque os dados coletados nos artigos, focavam principalmente na otimização do rendimento, resultando em uma quantidade maior de observações nesses intervalos em comparação com os rendimentos mais baixos.

O desempenho superior da Floresta Aleatória em comparação com outros modelos, pode estar relacionado a várias características como na capacidade de lidar eficientemente com relações não-lineares, especialmente considerando que as reações de transesterificação do biodiesel não seguem um comportamento linear. Além disso, o modelo demonstra robustez na presença de *outliers*, o que é especialmente relevante neste trabalho. Outro ponto é a flexibilidade na escolha de hiperparâmetros oferecida por este modelo. Por outro lado, a Regressão Linear possivelmente apresentou um desempenho inferior, uma vez que este modelo presume uma relação linear entre as variáveis de entrada e saída. Além disso, a presença de *outliers* e multicolinearidade entre os dados podem ter dificultado as previsões do modelo.

No MLP Regressor, por ser um modelo complexo, isso pode ter dificultado a obtenção de um ajuste adequado aos dados, uma vez que requer uma quantidade significativa para o seu treinamento. No entanto, no estudo conduzido por Lima et al. (2022), utilizando 254 dados com os mesmos parâmetros deste estudo e adicionando outras variáveis, como viscosidade, densidade do óleo e volume da reação, obtiveram bons resultados. Logo, a inclusão desses parâmetros possivelmente contribuiu significativamente para a capacidade de predição do modelo.

6.5 VALIDAÇÃO E AVALIAÇÃO DOS PARÂMETROS DO BIODIESEL

A validação do modelo que apresentou o melhor desempenho (Floresta Aleatória) foi realizada através da escolha de parâmetros de experimentos de dois artigos acadêmicos diferentes. Um deles utilizou um catalisador de CaO puro (VAZ, 2015), enquanto o outro experimento utilizou um catalisador de CaO não puro (TANG et al., 2013). Na Tabela 11, podem ser observados os dados da comparação entre o rendimento previsto e o rendimento experimental. Os resultados revelaram que os valores experimentais obtidos são próximos do rendimento de biodiesel previsto pelo modelo de Floresta Aleatória, confirmando assim a adequação deste modelo escolhido para a previsão do rendimento de biodiesel.

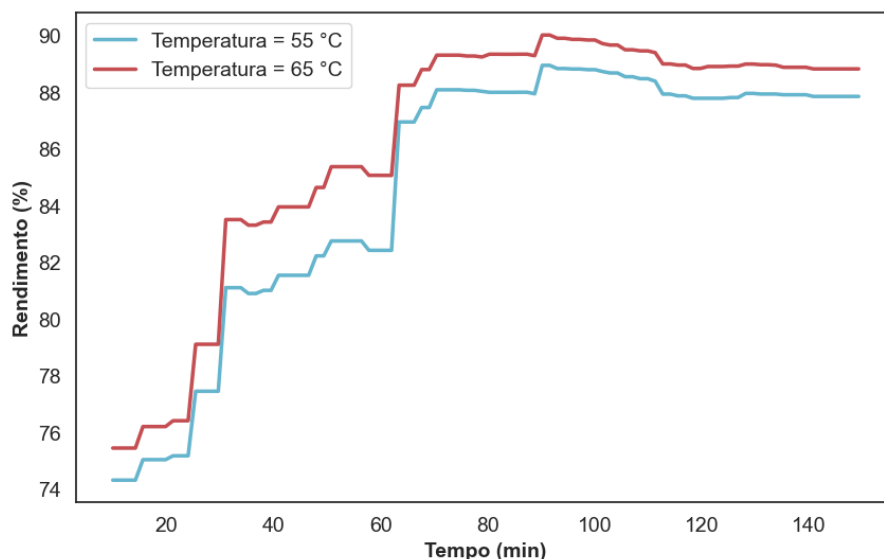
Tabela 11 – Validação do modelo com valores experimentais da literatura

Autor	Matéria Prima	Catalisador	Condições da Reação	Rendimento Predito (%)	Rendimento Experimental (%)	Erro Percentual
VAZ, 2015	Óleo de soja	CaO puro	Temperatura: 65°C, Peso do catalisador: 4% m/m, RE: 6:1, Tempo: 120 min	85,8	81	5,93%
TANG et al., 2013	Óleo de soja	CaO não puro	Temperatura: 65°C, Peso do catalisador: 5% m/m, RE: 15:1, Tempo: 180 min	89,6	99,5	9,95%

Fonte: elaborado pelo Autor, 2024

Os valores da Tabela 11 revelam que o modelo treinado realizou predições com erros percentuais inferiores a 10%. Esse desempenho era esperado, considerando que o erro médio absoluto (MAE) foi de 6,59 para a predição do rendimento. Esse valor indica que, em média, as previsões estão a uma distância de 6,59 unidades do valor real. No entanto, como esse erro é uma média, os valores podem variar, como demonstrado nos casos apresentados na tabela, onde alguns erros são menores e outros são maiores que esse valor médio. Por exemplo, no primeiro e no segundo experimento, os erros foram de 4,8 e 9,9 unidades, respectivamente, entre os valores previstos e reais.

Para avaliar os parâmetros do biodiesel, foram realizadas simulações de reações usando o modelo de Floresta Aleatória, foram empregados óleo de cozinha como matéria-prima e CaO não puro para fins de comparação. Na Figura 26, observa-se que o aumento da temperatura contribui para a conversão dos triglicerídeos em ésteres metílicos. Isso se deve ao fato de que a reação de transesterificação é endotérmica e requer temperaturas de reação mais elevadas para promover uma taxa de reação eficaz. Portanto, temperaturas mais altas podem reduzir a viscosidade e favorecer a velocidade da reação, resultando em rendimentos de biodiesel mais elevados. Além disso, em baixas temperaturas, a reação entre catalisadores, metanol e óleo pode ser dificultada, o que impede a transferência eficiente de massa do catalisador heterogêneo (OYEKUNLE et al., 2023).

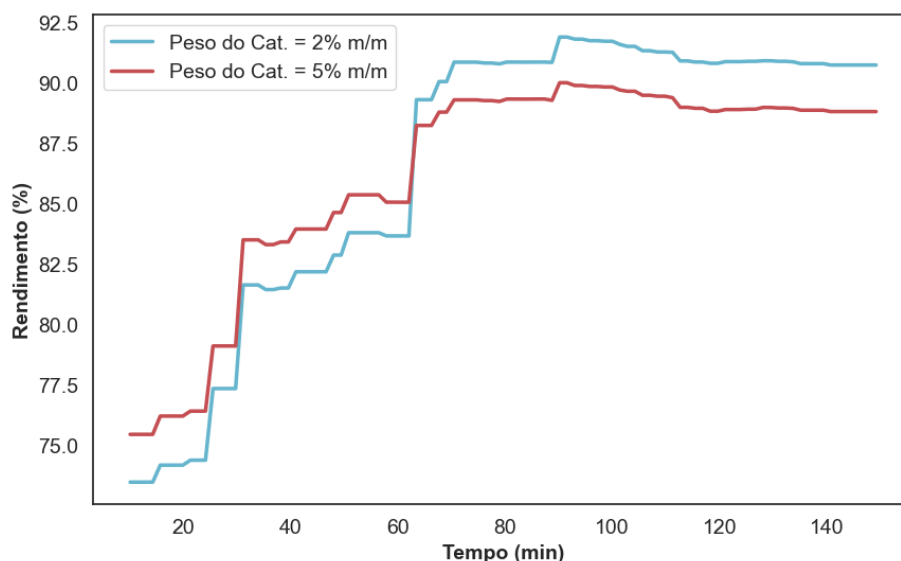
Figura 26 – Curvas do rendimento de biodiesel a diferentes temperaturas

Fonte: elaborado pelo Autor, 2024

Nota: As condições de reação foram: RE: 9:1, Peso cat.: 5% m/m, Tipo de catalisador: CaO não puro, Fonte: Óleo de cozinha

Conforme explicado por Moraes et al. (2013), o aumento da temperatura favorece a conversão da reação até certo ponto, porém essa conversão começa a diminuir ao longo do tempo devido ao consumo dos reagentes, o que promove uma reação inversa. Dessa forma, tempos de reação mais longos podem resultar no consumo contínuo de reagentes, além da formação de sabões. Esse fenômeno é observado na figura por volta de 90 minutos para ambas as temperaturas, logo após atingir a conversão máxima.

Com relação à dosagem do catalisador, foram realizadas duas simulações com o modelo, utilizando pesos distintos nas mesmas condições reacionais, como pode ser observado na Figura 27. Neste experimento, foi constatado que o aumento da quantidade de catalisador de 2 para 5% m/m afetou negativamente o rendimento do biodiesel, um comportamento inesperado, uma vez que com aumento da quantidade de catalisador há também um aumento na quantidade de sítios ativos disponíveis, que facilita a conversão do óleo (triglicerídeos) em ésteres de ácidos graxos (biodiesel) (SAHAR et al., 2022).

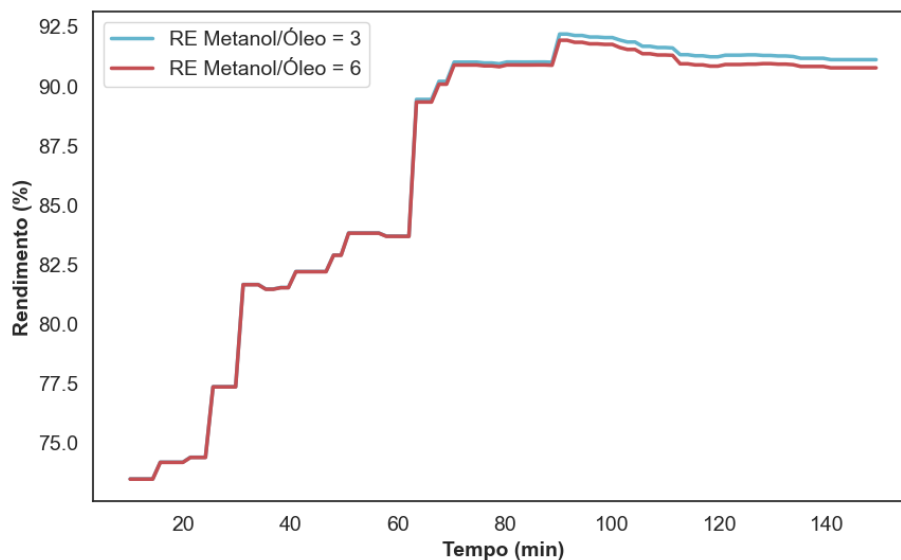
Figura 27 – Curvas do rendimento de biodiesel a diferentes pesos de catalisador

Fonte: elaborado pelo Autor, 2024

Nota: As condições de reação foram: RE: 9:1, Temperatura: 65°C, Tipo de catalisador: CaO não puro, Fonte: Óleo de cozinha

Entretanto, na figura, é evidente que o catalisador com maior concentração, teve inicialmente um aumento na taxa de reação, isto pode ser observado entre 30 e 60 minutos. De acordo com Li et al. (2022), ao aumentar a quantidade de sítios ativos disponíveis, há uma aceleração na taxa de reação para atingir o equilíbrio. Entretanto, a previsão do rendimento pelo modelo utilizando diferentes massas de catalisador não foi satisfatória.

Para comparar o efeito da razão estequiométrica entre o metanol e o óleo, foram utilizadas duas proporções, 3:1 e 6:1, respectivamente. Na Figura 28, é possível verificar que houve uma pequena melhora para a razão 3:1, isto pode ser causado pelo fato de que para cada razão estequiométrica há uma temperatura ótima associada (MORAIS et al., 2013).

Figura 28 – Curvas do rendimento de biodiesel a diferentes razões estequiométricas de metanol e óleo

Fonte: elaborado pelo Autor, 2022

Nota: As condições de reação foram: Peso cat.: 2% m/m, Temperatura: 65°C, Tipo de catalisador: CaO não puro, Fonte: Óleo de cozinha

Os resultados do treinamento dos modelos indicam que a Floresta Aleatória demonstra um desempenho superior na predição do rendimento de biodiesel em comparação com os modelos de Regressão Linear e MLP Regressor neste trabalho. No entanto, em um estudo recente conduzido por Agrawal e Dhawane (2024), que treinaram 11 modelos de aprendizagem de máquina com dados coletados de várias fontes da literatura de forma semelhante a este trabalho, concluíram que o modelo CatBoost apresentou o melhor desempenho, seguido pela Floresta Aleatória, com valores de RMSE de 4,29 e 4,55, respectivamente.

A obtenção desses modelos com melhores desempenhos pode estar relacionada a várias razões. Isso pode incluir a metodologia empregada na revisão sistemática ou até mesmo a quantidade de dados coletados. Também para esses 11 modelos, foram considerados outros parâmetros que influenciam no rendimento além daqueles utilizados neste trabalho, tais como densidade, viscosidade e diferentes tipos de catalisador, juntamente com sua categoria, assim como o tipo de óleo e sua categoria associada.

Cabe ressaltar que cada matéria-prima possui o seu próprio teor de ácidos graxos livres (FFAs). Por exemplo, o estudo intitulado "*Estudo à Seção de Transesterificação na Produção de Biodiesel*", conduzido por Pereira (2022), utilizou um modelo de redes neurais e levou em consideração o tipo de óleo utilizado e seu teor de FFAs, esse estudo obteve um alto desempenho no modelo em teste, usando dados provenientes da indústria Prio SGPS, S.A. A categorização dos óleos, considerando esse fator, permite fornecer ao modelo informações relevantes sobre a reação. Isso ocorre porque o teor de FFAs pode afetar diretamente a eficiência da reação, seja formando subprodutos ou até mesmo inibindo a atuação do catalisador.

No mesmo trabalho, o autor também levou em consideração que a escolha entre um catalisador ácido ou básico é primariamente influenciada pela quantidade de ácidos graxos livres (FFAs) presentes no óleo de alimentação. Para a catálise heterogênea ácida, é possível utilizar matéria-prima com alto teor de FFAs. Por outro lado, na catálise heterogênea básica, óleos com teor elevado de FFAs não são tolerados para a produção de biodiesel. O autor do trabalho mencionado também classificou o catalisador considerando este fator.

Para analisar o desempenho dos algoritmos de aprendizagem de máquina na predição do rendimento de biodiesel, este trabalho não considerou esses fatores devido à limitação de dados disponíveis na literatura. Em vez disso, foram utilizados apenas os parâmetros mais comuns em experimentos, para poder padronizar o banco de dados. Apesar de que isso possa limitar o modelo, isto permitiu uma previsão mais eficaz. Portanto, a obtenção de dados é uma etapa crucial que pode ser reestruturada para fornecer informações mais precisas e, conseqüentemente, melhorar o desempenho dos modelos.

Considerando todos esses fatores, conclui-se que o modelo da Floresta Aleatória se mostrou mais eficaz, sendo construída e treinada com o banco de dados coletado e utilizando os hiperparâmetros mais otimizados encontrados, demonstrou ser capaz de fornecer boas aproximações do rendimento para a reação de síntese de biodiesel a partir dos dados de entrada que foram óleo utilizado, tipo e quantidade de catalisador, razão estequiométrica metanol/óleo, temperatura e tempo, embora ainda existam possíveis pontos de melhoria a serem considerados.

7 CONSIDERAÇÕES FINAIS

Este estudo utilizou técnicas de aprendizagem de máquina para prever o rendimento do biodiesel usando dados de reação de transesterificação. Um conjunto de dados foi coletado de diferentes artigos acadêmicos selecionados, abordando certos parâmetros envolvendo diferentes matérias-primas. Em seguida, o conjunto de dados foi particionado em conjuntos de treinamento e teste para facilitar o treinamento do modelo, otimização e avaliação de precisão, respectivamente. Os algoritmos de Regressão Linear, Floresta Aleatória e o MLP Regressor foram testados para determinar sua eficácia na predição do rendimento do biodiesel.

Além disso, realizou-se uma otimização de hiperparâmetros para ajustar os modelos e melhorar seu desempenho preditivo. Através desta pesquisa, foram obtidas percepções importantes sobre o impacto de vários aspectos na predição do rendimento do biodiesel durante o processo de transesterificação de matérias-primas de biomassa. Os hiperparâmetros da grade 2, com 100 iterações, do modelo de Floresta Aleatória ajustado, foi o que mostrou o melhor desempenho com um valor de MAE mais baixo. Além disso, a validação mostrou uma previsão próxima do rendimento real experimental e a análise dos parâmetros do biodiesel foi possível através do modelo treinado, demonstrando que as curvas do rendimento seguem um comportamento esperado de acordo com a literatura.

Apesar destes fatores, o desempenho do modelo ainda precisa ser aprimorado em vários aspectos como a obtenção de uma base de dados maior, além de adicionar alguns outros parâmetros da reação, como o tipo de reator utilizado, a rotação, além de categorizar os óleos por teor de FFAs e também a categoria do catalisador utilizado. Além disso, é recomendado para trabalhos futuros o teste de novos modelos e a utilização de dados provenientes de indústrias que produzem biodiesel. Este modelo treinado possui aplicações práticas para predições de sistemas reacionais. Ele fornece aos pesquisadores informações para o planejamento e ajuste dos parâmetros operacionais na produção de biodiesel por catálise heterogênea. Essas predições podem ser utilizadas para otimizar o processo de produção, controlar custos e até mesmo orientar a tomada de decisões estratégicas.

REFERÊNCIAS BIBLIOGRÁFICAS

ABDUL MUTALIB, A. A. et al. SiO₂-Rich Sugar Cane Bagasse Ash Catalyst for Transesterification of Palm Oil. **BioEnergy Research**, v. 13, n. 3, p. 986–997, set. 2020.

ADAMES, L. V. et al. Produção de hidrogênio em reator anaeróbico de fluxo contínuo utilizando glicerol bruto oriundo da produção de biodiesel. **Matéria (Rio de Janeiro)**, v. 26, n. 2, 2021.

AGHBASHLO, M. et al. Machine learning technology in biodiesel research: A review. **Progress in Energy and Combustion Science**, v. 85, p. 100904, jul. 2021.

AGRAWAL, P.; R., G.; DHAWANE, S. H. Prediction of Biodiesel Yield Employing Machine Learning: Interpretability Analysis via Shapley Additive Explanations. **Fuel**, v. 359, p. 130516, mar. 2024.

AL-MUHTASEB, A. H. et al. Biodiesel production over a catalyst prepared from biomass-derived waste date pits. **Biotechnology Reports**, v. 20, p. e00284, 1 dez. 2018.

ALPAYDIN, E. **Introduction to Machine Learning**. 4. ed. Cambridge Massachusetts: The MIT Press, 2020.

AMENAGHAWON, A. N.; EVBARUNEGBE, N. I.; OBAHIAGBON, K. Optimum biodiesel production from waste vegetable oil using functionalized cow horn catalyst: A comparative evaluation of some expert systems. **Cleaner Engineering and Technology**, v. 4, 1 out. 2021.

ANWAR, M. et al. The potential of utilising papaya seed oil and stone fruit kernel oil as non-edible feedstock for biodiesel production in Australia - A review. **Energy Reports**, v. 5, p. 280–297, 1 nov. 2019.

ARRAIS, K. F. **Construção de um pipeline de dados utilizando serviços da nuvem**. 2022. Trabalho de Conclusão de Curso (Engenharia de Computação) - Universidade Federal de São Carlos, São Carlos.

ATHAR, M.; ZAIDI, S. A review of the feedstocks, catalysts, and intensification techniques for sustainable biodiesel production. **Journal of Environmental Chemical Engineering**, v. 8, n. 6, p. 104523, dez. 2020.

BARROS, T. D.; JARDINE, J. G. **Transesterificação**. Disponível em: <https://www.embrapa.br/agencia-de-informacao->

tecnologica/tematicas/agroenergia/biodiesel/tecnologia/transesterificacao. Acesso em: 31 mar. 2024.

BASUMATARY, S. et al. Utilization of renewable and sustainable basic heterogeneous catalyst from *Heteropanax fragrans* (Kesseru) for effective synthesis of biodiesel from *Jatropha curcas* oil. **Fuel**, v. 286, p. 119357, fev. 2021.

BEDIR, Ö.; DOĞAN, T. H. Use of sugar industry waste catalyst for biodiesel production. **Fuel**, v. 286, p. 119476, fev. 2021.

BHATTACHARYYA, S. **Ridge and Lasso Regression: A Complete Guide with Python Scikit-Learn**. Towards Data Science, mar. 2018. Disponível em: <https://towardsdatascience.com/ridge-and-lasso-regression-a-complete-guide-with-python-scikit-learn-e20e34bcbf0b>. Acesso em: 31 mar. 2024.

BIGOTO, M. A. R.; ELENO, L. T. F. **Avaliação de modelos de machine learning para predição da temperatura crítica de supercondutores**. Trabalho de Conclusão de Curso (Engenharia Física) - Escola de Engenharia de Lorena, Universidade de São Paulo, Lorena, 2020.

BISHOP, C. M. **Pattern Recognition and Machine Learning**. 1. ed. NY: Springer New York, 2006.

BRINK, Henrik; RICHARDS, Joseph; FETHEROLF, Mark. **Real-world machine learning**. Simon and Schuster, 2016.

CABELLO, C.; S., R.; A., Z. Catalizadores heterogêneos utilizados para la obtención de biodiesel. Afinidad. **Journal of Chemical Engineering Theoretical and Applied Chemistry**, v. 74, n. 577, mar. 2017.

CAI, J. et al. Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest. **Applied Energy**, v. 262, p. 114566, 15 mar. 2020.

CÂMARA DOS DEPUTADOS DO BRASIL. **Câmara aprova projeto dos “combustíveis do futuro”**. Agência Câmara de Notícias, mar. 2024. Disponível em: <https://www.camara.leg.br/noticias/1043361-camara-aprova-projeto-dos-combustiveis-do-futuro/#:~:text=A%20C%C3%A2mara%20dos%20Deputados%20aprovou,gasolina%20e%20ao%20diesel%2C%20respectivamente>. Acesso em: 10 fev. 2024.

CARDOSO, T. S. et al. Uma revisão da utilização de catalisadores heterogêneos para a produção de biodiesel. **Brazilian Applied Science Review**, v. 4, n. 1, p. 240–276, 2020.

CARVALHO, A. et al. **Inteligência Artificial—uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC, 2011.

CHOUHAN, A. P. S.; SARMA, A. K. Modern heterogeneous catalysts for biodiesel production: A comprehensive review. **Renewable and Sustainable Energy Reviews**, v. 15, n. 9, p. 4378–4399, 1 dez. 2011.

CORDEIRO, C. S. et al. Catalisadores heterogêneos para a produção de monoésteres graxos (biodiesel). **Química Nova**, v. 34, n. 3, p. 477–486, 2011.

COSTA, Patrícia Pinto Kalil Gonçalves. Catalisadores químicos utilizados na síntese de biodiesel. Brasília, **Embrapa Agroenergia**, 2011.

DABDOUB, M. J.; BRONZEL, J. L.; RAMPIN, M. A. Biodiesel: visão crítica do status atual e perspectivas na academia e na indústria. **Química Nova**, v. 32, n. 3, p. 776–792, 2009.

DEMIRBAS, A. Political, economic and environmental impacts of biofuels: A review. **Applied Energy**, v. 86, n. SUPPL. 1, p. S108–S117, 1 nov. 2009.

DEMIRBAS, A.; FATIH DEMIRBAS, M. Importance of algae oil as a source of biodiesel. **Energy Conversion and Management**, v. 52, n. 1, p. 163–170, 1 jan. 2011.

FABIEN, M. **A Guide to Hyperparameter Optimization (HPO)**. Disponível em: <https://maelfabien.github.io/machinelearning/Explorium4/#>. Acesso em: 1 mar. 2024.

FRANÇA, M. G. **Compressão de imagens usando redes neurais artificiais**. Trabalho de conclusão de curso (Bacharelado em Engenharia Elétrica) - Brasília: Universidade de Brasília - UnB, 2019.

GAUR, R. Kanwar; GOYAL, R. A review: Effect on performance and emission characteristics of waste cooking oil Biodiesel-diesel blends on IC engine. **Materials Today: Proceedings**, 28 abr. 2022.

GÉRON, A. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow**. Second Edition. Sebastopol, CA: O'Reilly Media, Inc., 2019.

GNANAPRAKASAM, A. et al. Recent Strategy of Biodiesel Production from Waste Cooking Oil and Process Influencing Parameters: A Review. **Journal of Energy**, v. 2013, p. 1–10, 2013.

GODOY, D. **Hiperparâmetros em Ação Parte I - Funções de Ativação**. Disponível em: <https://medium.com/ensina-ai/hiper-par%C3%A2metros-em->

a%C3%A7%C3%A3o-parte-i-fun%C3%A7%C3%B5es-de-ativa%C3%A7%C3%A3o-cd37600e46c6. Acesso em: 29 fev. 2024.

GOHAIN, M. et al. Carica papaya stem: A source of versatile heterogeneous catalyst for biodiesel production and C–C bond formation. **Renewable Energy**, v. 147, p. 541–555, mar. 2020.

GUPTA, K. K. et al. Machine Learning-Based Predictive Modelling of Biodiesel Production - A Comparative Perspective. **Energies**, v. 14, n. 4, p. 1122, 20 fev. 2021.

GURUNATHAN, B.; RAVI, A. Biodiesel production from waste cooking oil using copper doped zinc oxide nanocomposite as heterogeneous catalyst. **Bioresource Technology**, v. 188, p. 124–127, 1 jul. 2015.

HASTIE, T. et al. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. Springer, 2009. v. 2.

HUANG, J. et al. Fabrication of hollow cage-like CaO catalyst for the enhanced biodiesel production via transesterification of soybean oil and methanol. **Fuel**, v. 290, p. 119799, abr. 2021.

JAIN, K. **Tree Based Algorithms: A Complete Tutorial from Scratch (in R & Python)**. Disponível em: <https://www.analyticsvidhya.com/blog/2016/04/tree-based-algorithms-complete-tutorial-scratch-in-python/>. Acesso em: 10 fev. 2024.

JAMES, Gareth et al. **An Introduction to Statistical Learning**. New York: Springer, 2013.

JUNQUEIRA, P. G. et al. Economic and environmental analysis of the cumene production process using computational simulation. **Chemical Engineering and Processing - Process Intensification**, v. 130, p. 309–325, ago. 2018.

KATARIA, J.; MOHAPATRA, S. K.; KUNDU, K. Biodiesel production from waste cooking oil using heterogeneous catalysts and its operational characteristics on variable compression ratio CI engine. **Journal of the Energy Institute**, v. 92, n. 2, p. 275–287, abr. 2019.

KHATIBI, M.; KHORASHEH, F.; LARIMI, A. Biodiesel production via transesterification of canola oil in the presence of Na–K doped CaO derived from calcined eggshell. **Renewable Energy**, v. 163, p. 1626–1636, jan. 2021.

KOTSIANTIS, S. B. et al. Supervised machine learning: A review of classification techniques. **Emerging Artificial Intelligence Applications in Computer Engineering**, v. 160, n. 1, p. 3–24, 2007.

KUYUCU, A. K. **Artificial Neural Networks: Basics and Principles**. Disponível em: <https://medium.com/ai-in-plain-english/dl-tutorial-2-artificial-neural-networks-basics-and-principles-40355d1331c1>. Acesso em: 7 mar. 2024.

LANI, N. S.; NGADI, N.; INUWA, I. M. New route for the synthesis of silica-supported calcium oxide catalyst in biodiesel production. **Renewable Energy**, v. 156, p. 1266–1277, ago. 2020.

LASKAR, I. B. et al. Taming waste: Waste *Mangifera indica* peel as a sustainable catalyst for biodiesel production at room temperature. **Renewable Energy**, v. 161, p. 207–220, dez. 2020.

LEE, E. **An Intro to Hyper-parameter Optimization using Grid Search and Random Search**. Disponível em: <https://medium.com/@cjl2fv/an-intro-to-hyper-parameter-optimization-using-grid-search-and-random-search-d73b9834ca0a>. Acesso em: 7 mar. 2024.

LEE, T.-H.; ULLAH, A.; WANG, R. Bootstrap Aggregating and Random Forest. **Macroeconomic Forecasting in the Era of Big Data**, v. 52, p. 389–429, mar. 2019.

LEUNG, D. Y. C.; GUO, Y. Transesterification of neat and used frying oil: Optimization for biodiesel production. **Fuel Processing Technology**, v. 87, n. 10, p. 883–890, out. 2006.

LEUNG, D. Y. C.; WU, X.; LEUNG, M. K. H. A review on biodiesel production using catalyzed transesterification. **Applied Energy**, v. 87, n. 4, p. 1083–1095, 1 abr. 2010.

LI, Y. et al. Role of oxygen vacancy on activity of Fe-doped SrTiO₃ perovskite bifunctional catalysts for biodiesel production. **Renewable Energy**, v. 199, p. 1258–1271, nov. 2022.

LIMA, J. L. M. et al. Intensificação Da Produção De Biodiesel Por Meio De Redes Neurais Artificiais. XXX Congresso de Iniciação Científica. **Anais**. UNICAMP: Pró-reitoria de pesquisa Unicamp, 2022.

LÍRIO COSTA, L. H. **AutoML e sua contribuição para o desenvolvimento de modelos para a Aprendizagem Profunda**. Trabalho de Conclusão de Curso - Universidade Federal de Ouro Preto, Instituto de Ciências Exatas e Aplicadas, Departamento de Computação e Sistemas, João Monlevade, 2024.

LIU, X. et al. Highly active and durable Ca-based solid base catalyst for biodiesel production. **Fuel**, v. 302, p. 121094, out. 2021.

LIU, Y. et al. Biofuels for a sustainable future. **Cell**, v. 184, n. 6, p. 1636–1647, 18 mar. 2021.

LORA, E. E. S.; VENTURINI, O. J. **Biocombustíveis**. Second Edition ed. Rio de Janeiro: Interciência, 2012. v. Vol.1 e Vol.2.

MAHESH, B. Machine Learning Algorithms - A Review. **ResearchGate**, v. 1, mar. 2019.

MAHMOUD, H. R.; EL-MOLLA, S. A.; IBRAHIM, M. M. Biodiesel production via stearic acid esterification over mesoporous ZrO₂/SiO₂ catalysts synthesized by surfactant-assisted sol-gel auto-combustion route. **Renewable Energy**, v. 160, p. 42–51, 1 nov. 2020.

MCKINNEY, W. **Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython**. "O'Reilly Media, Inc.", 2012.

MEHTA, A. **Demystifying Hyperparameter Tuning: GridSearchCV and RandomizedSearchCV**. Disponível em: <https://medium.com/@dancerworld60/demystifying-hyperparameter-tuning-gridsearchcv-and-randomizedsearchcv-2123bf3fb6c8>. Acesso em: 31 jan. 2024.

MIRABOUTALEBI, S. M.; KAZEMI, P.; BAHRAMI, P. Fatty Acid Methyl Ester (FAME) composition used for estimation of biodiesel cetane number employing random forest and artificial neural networks: A new approach. **Fuel**, v. 166, p. 143–151, fev. 2016.

MOAYEDI, H. et al. Feature validity during machine learning paradigms for predicting biodiesel purity. **Fuel**, v. 262, p. 116498, fev. 2020.

MORAIS, F. R. et al. Influência da Temperatura e Razão Molar na Produção Contínua de Biodiesel. **Scientia Plena**, v. 9, abr. 2013.

MÜLLER, A. C.; GUIDO, S. **Introduction to Machine Learning with Python: A Guide for Data Scientists**. "O'Reilly Media, Inc.", 2016.

NIGAM, P. S.; SINGH, A. Production of liquid biofuels from renewable resources. **Progress in Energy and Combustion Science**, v. 37, n. 1, p. 52–68, 1 fev. 2011.

OYEKUNLE, D. T. et al. Heterogeneous catalytic transesterification for biodiesel production: Feedstock properties, catalysts and process parameters. **Process Safety and Environmental Protection**, v. 177, p. 844–867, set. 2023.

PARIDA, S.; SINGH, M.; PRADHAN, S. Biomass wastes: A potential catalyst source for biodiesel production. **Bioresource Technology Reports**, p. 101081, maio 2022.

PENG, J. et al. Machine Learning Techniques for Personalised Medicine Approaches in Immune-Mediated Chronic Inflammatory Diseases: Applications and Challenges. **Frontiers in Pharmacology**, v. 12, p. 720694, 30 set. 2021.

PEREIRA, B. M. F. **Estudo à Secção de Transesterificação na produção de biodiesel**. Dissertação de Mestrado em Engenharia Química e Biologia - Politécnica de Lisboa, Instituto Superior de Engenharia de Lisboa, Departamento de Engenharia Química, Lisboa, 2022.

PEREIRA, F. DE A. **Processo de produção e purificação do biodiesel de amendoim**. Dissertação de Mestrado - Universidade Federal de Alagoas, Centro de Tecnologia, Maceió, 2015.

POLIKAR, R. **Ensemble learning. Ensemble machine learning: Methods and applications**, p. 1-34, 2012.

PRAKASH, P. **Understanding Baseline Models in Machine Learning**. Disponível em: https://medium.com/@preethi_prakash/understanding-baseline-models-in-machine-learning-3ed94f03d645. Acesso em: 6 fev. 2024.

PRASAD, S. D. **The Importance of a Machine Learning Pipeline**. Disponível em: <https://embeddedcomputing.com/technology/ai-machine-learning/the-importance-of-a-machine-learning-pipeline>. Acesso em: 21 mar. 2024.

PREMANAND S, P. **Artificial Neural Networks - Better Understanding**. Disponível em: <https://www.analyticsvidhya.com/blog/2021/06/artificial-neural-networks-better-understanding/>. Acesso em: 14 mar. 2024.

PROVOST, F.; FAWCETT, T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. **Big Data**, v. 1, n. 1, p. 51–59, mar. 2013.

RAMOS, L. P. et al. Biodiesel Production Technologies. **Revista Virtual de Química**, v. 3, n. 5, 2011.

RASCHKA, Sebastian; MIRJALILI, Vahid. **Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2**. Packt publishing ltd, 2019.

RITVO, A. R. et al. **Regression Analysis**. In: **Encyclopedia of Autism Spectrum Disorders**. New York, NY: Springer, 2013. p. 2538–2538.

ROSCHAT, W. et al. The synthesis of a high-quality biodiesel product derived from Krabok (Irvingia Malayana) seed oil as a new raw material of Thailand. **Fuel**, v. 308, p. 122009, 15 jan. 2022.

SABOYA, R. M. A. **Produção de biodiesel empregando catalisadores nanoestruturados do tipo SBA-15 modificada com lantânio**. Dissertação de Mestrado - Universidade Federal do Ceará, Departamento de Engenharia Química, Fortaleza, 2012.

SAHAR, J. et al. Highly efficient heteropoly acid decorated SnO₂@Co-ZIF nanocatalyst for sustainable biodiesel production from Nannorrhops ritchiana seeds oil. **Renewable Energy**, v. 198, p. 306–318, out. 2022.

SAINI, A. **Decision Tree Algorithm - A Complete Guide**. Analytics Vidhya, 2024. Disponível em: <https://www.analyticsvidhya.com/blog/2021/08/decision-tree-algorithm/>. Acesso em: 14 mar. 2024.

SALLET, Cíntia Letícia; ALVIM, Augusto Mussi. Biocombustíveis: uma análise da evolução do biodiesel no Brasil. **Economia & Tecnologia (UFPR)**, 2011.

SANDOVAL, G. et al. Enzyme-Catalyzed Production of Biodiesel as Alternative to Chemical- Catalyzed Processes: Advantages and Constraints. **Current Biochemical Engineering**, v. 4, n. 2, 31 jul. 2017.

SCHWEIDTMANN, A. M. et al. Machine Learning in Chemical Engineering: A Perspective. **Chemie Ingenieur Technik**, v. 93, n. 12, p. 2029–2039, 22 dez. 2021.

SIMEONE, Osvaldo et al. A brief introduction to machine learning for engineers. **Foundations and Trends® in Signal Processing**, v. 12, n. 3-4, p. 200-431, 2018.

SINGH, D. et al. A comprehensive review of physicochemical properties, production process, performance and emissions characteristics of 2nd generation biodiesel feedstock: *Jatropha curcas*. **Fuel**, v. 285, p. 119110, fev. 2021.

SINGH, D. et al. A review on feedstocks, production processes, and yield for different generations of biodiesel. **Fuel**, fev. 2020.

SMANIOTTO, G. C. **Super-resolução de áudio utilizando redes neurais artificiais**. Trabalho de Conclusão de Curso - Universidade Federal do Rio Grande do Sul, Escola de Engenharia, Curso de Graduação em Engenharia Elétrica, Porto Alegre, janeiro de 2023.

SOARES, T. M. **Estudo da adição de lítio a nanocatalisadores de óxido de cálcio para a produção de biodiesel**. Dissertação de Mestrado em Engenharia Química - Universidade Federal do Espírito Santo, Centro de Ciências Agrárias e Engenharias, Alegre, ES, 2018.

SUN, Y.; GENTON, M. G. Functional Boxplots. **Journal of Computational and Graphical Statistics**, v. 20, n. 2, p. 316–334, jan. 2011.

SUTTON, R. S.; BARTO, A. G. **Reinforcement learning: An introduction**. MIT press, 2018.

SUVARNA, G. Linear Regression in machine learning. Disponível em: <https://www.analyticsvidhya.com/blog/2021/06/linear-regression-in-machine-learning/>. Acesso em: 23 jul. 2022.

TANG, Y. et al. Biodiesel production from vegetable oil by using modified CaO as solid basic catalysts. **Journal of Cleaner Production**, v. 42, p. 198–203, mar. 2013.

UMENWEKE, G. C. et al. Machine learning methods for modeling conventional and hydrothermal gasification of waste biomass: A review. **Bioresource Technology Reports**, v. 17, p. 100976, 1 fev. 2022.

VAZ, L. **Produção de biodiesel utilizando óxido de cálcio e zirconato de sódio livre e suportado em materiais poliméricos**. Dissertação de Mestrado em Engenharia Química - Universidade Federal de Uberlândia, Faculdade de Engenharia Química, Uberlândia, MG, 26 de março de 2015.

VICENTE, G.; MARTÍNEZ, M.; ARACIL, J. Integrated biodiesel production: a comparison of different homogeneous catalysts systems. **Bioresource Technology**, v. 92, n. 3, p. 297–305, 1 maio 2004.

VIEIRA, S.; PINAYA, W. H. L.; MECHELLI, A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. **Neuroscience & Biobehavioral Reviews**, v. 74, p. 58–75, mar. 2017.

WANG, Z. et al. The role of machine learning to boost the bioenergy and biofuels conversion. **Bioresource Technology**, v. 343, p. 126099, 1 jan. 2022.

XING, Y. et al. A review on machine learning application in biodiesel production studies. **International Journal of Chemical Engineering**, v. 2021, p. 1-12, 2021.

YAASHIKAA, P. R.; KUMAR, P. S.; KARISHMA, S. Bio-derived catalysts for production of biodiesel: A review on feedstock, oil extraction methodologies, reactors and lifecycle assessment of biodiesel. **Fuel**, v. 316, p. 123379, 15 maio 2022.

YEHOSHUA, R. **Random Forests**. 2023. Disponível em: <https://medium.com/@roiyeho/random-forests-98892261dc49>. Acesso em: 14 mar. 2024.

ZORNOZA, J. **Decision Trees Explained**. 2020. Disponível em: <https://towardsdatascience.com/decision-trees-explained-3ec41632ceb6>. Acesso em: 14 mar. 2024.

APÊNDICE A – TABELA COM OS DADOS UTILIZADOS PARA O TREINAMENTO DOS MODELOS DE AM

Nº de dados	Óleo utilizado	Tipo de catalisador	Peso de Catalisador (% m/m)	RE de Met./Óleo	Temperatura (°C)	Tempo (min)	Rendimento (%)
1	Girassol	CaO puro	1	15	60	120	4
2	Girassol	CaO puro	2	15	60	120	50
3	Girassol	CaO puro	3	15	60	120	72
4	Girassol	CaO puro	4	15	60	120	82
5	Girassol	CaO puro	5	15	60	120	87.64
6	Girassol	CaO puro	6	15	60	120	83
7	Girassol	CaO puro	7	15	60	120	80
8	Girassol	CaO puro	5	6	60	120	8
9	Girassol	CaO puro	5	9	60	120	69
10	Girassol	CaO puro	5	12	60	120	81
11	Girassol	CaO puro	5	15	60	120	90
12	Girassol	CaO puro	5	18	60	120	80
13	Girassol	CaO puro	5	15	60	60	5
14	Girassol	CaO puro	5	15	60	120	87.64
15	Girassol	CaO puro	5	15	60	180	82
16	Girassol	CaO puro	5	15	60	240	78
17	Girassol	CaO puro	5	15	60	300	77.5
18	Girassol	CaO puro	5	9	90	10	49
19	Girassol	CaO puro	5	9	90	20	74
20	Girassol	CaO puro	5	9	90	30	83
21	Girassol	CaO puro	5	9	90	40	85
22	Girassol	CaO puro	5	9	90	60	86
23	Girassol	CaO puro	5	3	90	60	70
24	Girassol	CaO puro	5	5	90	60	83
25	Girassol	CaO puro	5	7	90	60	87
26	Girassol	CaO puro	5	9	60	60	42
27	Girassol	CaO puro	5	9	65	60	60
28	Girassol	CaO puro	5	9	70	60	84
29	Girassol	CaO puro	1	9	90	60	20
30	Girassol	CaO puro	2	9	90	60	39
31	Girassol	CaO puro	3	9	90	60	86
32	Óleo de cozinha	CaO puro	2	3	60	180	75
33	Óleo de cozinha	CaO puro	2	6	60	180	86.06
34	Óleo de cozinha	CaO puro	2	9	60	180	95.23
35	Óleo de cozinha	CaO puro	2	12	60	180	95.23
36	Óleo de cozinha	CaO puro	2	15	60	180	90
37	Óleo de cozinha	CaO puro	1	9	60	180	85
38	Óleo de cozinha	CaO puro	2	9	60	180	95.23
39	Óleo de cozinha	CaO puro	3	9	60	180	95.23
40	Óleo de cozinha	CaO puro	4	9	60	180	95.23
41	Óleo de cozinha	CaO puro	2	9	60	60	80

42	Óleo de cozinha	CaO puro	2	9	60	120	90.6
43	Óleo de cozinha	CaO puro	2	9	60	180	95.23
44	Óleo de cozinha	CaO puro	2	9	60	240	95.23
45	Óleo de cozinha	CaO puro	2	9	60	300	95.23
46	Soja	CaO puro	1	15	65	120	90.3
47	Soja	CaO puro	2	15	65	120	92
48	Soja	CaO puro	3	15	65	120	97
49	Soja	CaO puro	4	15	65	120	94
50	Soja	CaO puro	5	15	65	120	91
51	Soja	CaO puro	6	15	65	120	90
52	Soja	CaO puro	7	15	65	120	89
53	Soja	CaO puro	3	6	65	120	4
54	Soja	CaO puro	3	9	65	120	80
55	Soja	CaO puro	3	12	65	120	92
56	Soja	CaO puro	3	15	65	120	97
57	Soja	CaO puro	3	18	65	120	92.1
58	Soja	CaO puro	3	21	65	120	91
59	Soja	CaO puro	3	15	65	20	22
60	Soja	CaO puro	3	15	65	40	46
61	Soja	CaO puro	3	15	65	60	68
62	Soja	CaO puro	3	15	65	80	88
63	Soja	CaO puro	3	15	65	100	91
64	Soja	CaO puro	3	15	65	120	98
65	Soja	CaO puro	3	15	65	200	96
66	Soja	CaO puro	3	15	50	120	89
67	Soja	CaO puro	3	15	55	120	91
68	Soja	CaO puro	3	15	60	120	92.5
69	Soja	CaO puro	3	15	65	120	97.8
70	Soja	CaO puro	3	15	70	120	96
71	Soja	CaO não puro	3	6	28	60	23.5
72	Soja	CaO não puro	3	6	28	120	40
73	Soja	CaO não puro	3	6	28	180	52.5
74	Soja	CaO não puro	3	6	28	240	61
75	Soja	CaO não puro	3	6	28	300	65
76	Soja	CaO não puro	4	6	28	60	28
77	Soja	CaO não puro	4	6	28	120	47
78	Soja	CaO não puro	4	6	28	180	63
79	Soja	CaO não puro	4	6	28	240	75
80	Soja	CaO não puro	4	6	28	300	79
81	Soja	CaO não puro	5	6	28	60	32
82	Soja	CaO não puro	5	6	28	120	57
83	Soja	CaO não puro	5	6	28	180	72
84	Soja	CaO não puro	5	6	28	240	81.5
85	Soja	CaO não puro	5	6	28	300	86
86	Soja	CaO não puro	6	6	28	60	40
87	Soja	CaO não puro	6	6	28	120	67.5
88	Soja	CaO não puro	6	6	28	180	84
89	Soja	CaO não puro	6	6	28	240	96

90	Soja	CaO não puro	6	6	28	300	98
91	Soja	CaO não puro	6	4	28	240	71
92	Soja	CaO não puro	6	5	28	240	89
93	Soja	CaO não puro	6	6	28	240	98
94	Soja	CaO não puro	6	7	28	240	96.5
95	Soja	CaO não puro	6	8	28	240	94
96	Soja	CaO não puro	6	6	28	60	41
97	Soja	CaO não puro	6	6	28	120	68
98	Soja	CaO não puro	6	6	28	180	88
99	Soja	CaO não puro	6	6	28	300	96
100	Óleo de cozinha	CaO não puro	3	6	65	20	20
101	Óleo de cozinha	CaO não puro	3	6	65	40	42
102	Óleo de cozinha	CaO não puro	3	6	65	60	55
103	Óleo de cozinha	CaO não puro	3	6	65	80	60
104	Óleo de cozinha	CaO não puro	3	6	65	105	64
105	Óleo de cozinha	CaO não puro	3	3	65	20	10
106	Óleo de cozinha	CaO não puro	3	3	65	40	28
107	Óleo de cozinha	CaO não puro	3	3	65	60	41
108	Óleo de cozinha	CaO não puro	3	3	65	80	54
109	Óleo de cozinha	CaO não puro	3	3	65	105	62
110	Óleo de cozinha	CaO não puro	3	9	65	20	42
111	Óleo de cozinha	CaO não puro	3	9	65	40	86
112	Óleo de cozinha	CaO não puro	3	9	65	60	90
113	Óleo de cozinha	CaO não puro	3	9	65	80	95
114	Óleo de cozinha	CaO não puro	3	9	65	105	98
115	Óleo de cozinha	CaO não puro	3	12	65	20	42
116	Óleo de cozinha	CaO não puro	3	12	65	40	86
117	Óleo de cozinha	CaO não puro	3	12	65	60	90
118	Óleo de cozinha	CaO não puro	3	12	65	80	95
119	Óleo de cozinha	CaO não puro	3	12	65	105	98
120	Canola	CaO não puro	3	9	40	180	72.5
121	Canola	CaO não puro	3	9	50	180	97.6
122	Canola	CaO não puro	3	9	60	180	93.5
123	Canola	CaO não puro	3	9	70	180	90
124	Canola	CaO não puro	3	6	50	180	96
125	Canola	CaO não puro	3	9	50	180	96.7
126	Canola	CaO não puro	3	12	50	180	91
127	Canola	CaO não puro	3	15	50	180	80
128	Canola	CaO não puro	1	9	50	180	73.3
129	Canola	CaO não puro	3	9	50	180	97.6
130	Canola	CaO não puro	5	9	50	180	90
131	Canola	CaO não puro	7	9	50	180	86
132	Canola	CaO não puro	3	9	50	120	71
133	Canola	CaO não puro	3	9	50	240	95
134	Canola	CaO não puro	3	9	50	300	90
135	Jatrofa	CaO não puro	7	12	65	65	97.75
136	Jatrofa	CaO não puro	9	12	65	55	97.4
137	Jatrofa	CaO não puro	3	12	65	185	91.96

138	Jatrofa	CaO não puro	7	3	65	185	89.45
139	Jatrofa	CaO não puro	7	6	65	115	93.12
140	Jatrofa	CaO não puro	7	9	65	85	96.45
141	Jatrofa	CaO não puro	7	15	65	65	97.21
142	Jatrofa	CaO não puro	7	18	65	75	97.43
143	Jatrofa	CaO não puro	7	12	32	205	97.63
144	Jatrofa	CaO não puro	7	12	45	135	96.32
145	Jatrofa	CaO não puro	7	12	55	95	97.52
146	Jatrofa	CaO não puro	7	12	75	60	97.71
147	Óleo de cozinha	CaO não puro	3	20	60	120	75
148	Óleo de cozinha	CaO não puro	6	20	60	120	80
149	Óleo de cozinha	CaO não puro	9	20	60	120	85
150	Óleo de cozinha	CaO não puro	12	20	60	120	78
151	Óleo de cozinha	CaO não puro	15	20	60	120	68
152	Óleo de cozinha	CaO não puro	9	5	60	120	64
153	Óleo de cozinha	CaO não puro	9	10	60	120	71
154	Óleo de cozinha	CaO não puro	9	15	60	120	79
155	Óleo de cozinha	CaO não puro	9	20	60	120	85
156	Óleo de cozinha	CaO não puro	9	25	60	120	77
157	Óleo de cozinha	CaO não puro	9	20	60	60	71.5
158	Óleo de cozinha	CaO não puro	9	20	60	120	85
159	Óleo de cozinha	CaO não puro	9	20	60	180	85
160	Óleo de cozinha	CaO não puro	9	20	60	240	80
161	Óleo de cozinha	CaO não puro	9	20	60	300	74.5
162	Óleo de cozinha	CaO não puro	9	20	25	120	63.5
163	Óleo de cozinha	CaO não puro	9	20	40	120	77.5
164	Óleo de cozinha	CaO não puro	9	20	60	120	85
165	Óleo de cozinha	CaO não puro	9	20	80	120	68.5
166	Óleo de cozinha	CaO não puro	9	20	100	120	53
167	Óleo de cozinha	CaO não puro	3	20	60	120	83
168	Óleo de cozinha	CaO não puro	6	20	60	120	93.5
169	Óleo de cozinha	CaO não puro	9	20	60	120	90
170	Óleo de cozinha	CaO não puro	12	20	60	120	82.5
171	Óleo de cozinha	CaO não puro	15	20	60	120	78.5
172	Óleo de cozinha	CaO não puro	6	5	60	120	67
173	Óleo de cozinha	CaO não puro	6	10	60	120	82
174	Óleo de cozinha	CaO não puro	6	15	60	120	94
175	Óleo de cozinha	CaO não puro	6	20	60	120	93
176	Óleo de cozinha	CaO não puro	6	25	60	120	80
177	Óleo de cozinha	CaO não puro	6	15	60	60	82.5
178	Óleo de cozinha	CaO não puro	6	15	60	120	94
179	Óleo de cozinha	CaO não puro	6	15	60	180	91.5
180	Óleo de cozinha	CaO não puro	6	15	60	240	80.5
181	Óleo de cozinha	CaO não puro	6	15	60	300	79.5
182	Óleo de cozinha	CaO não puro	6	15	25	120	70.5
183	Óleo de cozinha	CaO não puro	6	15	40	120	84
184	Óleo de cozinha	CaO não puro	6	15	60	120	94.5
185	Óleo de cozinha	CaO não puro	6	15	80	120	73

186	Óleo de cozinha	CaO não puro	6	15	100	120	59.8
187	Palma	CaO não puro	5	9	65	180	53
188	Palma	CaO não puro	5	12	65	180	57
189	Palma	CaO não puro	5	15	65	180	73
190	Palma	CaO não puro	5	18	65	180	86.5
191	Palma	CaO não puro	5	20	65	180	91.2
192	Palma	CaO não puro	5	23	65	180	89
193	Palma	CaO não puro	3	20	65	180	54
194	Palma	CaO não puro	4	20	65	180	55.5
195	Palma	CaO não puro	5	20	65	180	90
196	Palma	CaO não puro	6	20	65	180	92
197	Palma	CaO não puro	7	20	65	180	89
198	Palma	CaO não puro	6	20	65	60	58
199	Palma	CaO não puro	6	20	65	120	74
200	Palma	CaO não puro	6	20	65	180	93.6
201	Palma	CaO não puro	6	20	65	240	87.5
202	Palma	CaO não puro	6	20	65	300	86
203	Palma	CaO não puro	6	20	50	180	65
204	Palma	CaO não puro	6	20	55	180	78.5
205	Palma	CaO não puro	6	20	60	180	86.5
