

UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUIMICA

AGAONE DONIZETE SILVA

**PREVISÃO DAS PROPRIEDADES DO BIODIESEL E SEU PERFIL IDEAL DE  
ÁCIDOS GRAXOS POR MEIO DE APRENDIZADO DE MÁQUINA**

Uberlândia

2024

AGAONE DONIZETE SILVA

**PREVISÃO DAS PROPRIEDADES DO BIODIESEL E SEU PERFIL IDEAL DE  
ÁCIDOS GRAXOS POR MEIO DE APRENDIZADO DE MÁQUINA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Química da Universidade Federal de Uberlândia como requisito parcial para obtenção do título de mestre em Engenharia Química

Linha de pesquisa: Modelagem, Controle e Otimização de Processo

Orientadora: Prof. Dra. Sarah Arvelos Altino

Uberlândia

2024

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU  
com dados informados pelo(a) próprio(a) autor(a).

S586 2024	<p>Silva, Agaone Donizete, 1955- Previsão das propriedades do biodiesel e seu perfil ideal de ácidos graxos por meio de aprendizado de máquina [recurso eletrônico] / Agaone Donizete Silva. - 2024.</p> <p>Orientadora: Sarah Arvelos Altino. Dissertação (Mestrado) - Universidade Federal de Uberlândia, Pós-graduação em Engenharia Química. Modo de acesso: Internet. Disponível em: <a href="http://doi.org/10.14393/ufu.di.2024.686">http://doi.org/10.14393/ufu.di.2024.686</a> Inclui bibliografia. Inclui ilustrações.</p> <p>1. Engenharia química. I. Altino, Sarah Arvelos ,1986- , (Orient.). II. Universidade Federal de Uberlândia. Pós-graduação em Engenharia Química. III. Título.</p> <p style="text-align: right;">CDU: 66.0</p>
--------------	---

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091

Nelson Marcos Ferreira - CRB6/3074

AGAONE DONIZETE SILVA

**PREVISÃO DAS PROPRIEDADES DO BIODIESEL E SEU PERFIL IDEAL DE  
ÁCIDOS GRAXOS POR MEIO DE APRENDIZADO DE MÁQUINA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Química da Universidade Federal de Uberlândia como requisito parcial para obtenção do título de mestre em Engenharia Química

Linha de pesquisa: Modelagem, Controle e Otimização de Processo

Uberlândia, 26/09/2024

Banca Examinadora:

---

Prof. Dra. Sarah Arvelos Altino (orientadora) - FEQUI/UFU

---

Prof. Dr. Diego Andrade Lemos – DEQ/UFSCar

---

Prof. Dr. Rubens Gedraite – FEQUI/UFU



**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**  
Coordenação do Programa de Pós-Graduação em Engenharia  
Química  
Av. João Naves de Ávila, 2121, Bloco 1K, Sala 206 - Bairro Santa Mônica, Uberlândia-MG,  
CEP 38400-902  
Telefone: (34)3239-4249 - www.ppgeq.feq.ufu.br - secppgeq@feq.ufu.br



## ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-graduação em:	Engenharia Química				
Defesa de:	Mestrado Acadêmico, 8/2024, PPGEQ				
Data:	26 de setembro de 2024	Hora de início:	10:00	Hora de encerramento:	11:23
Matrícula do Discente:	12222EQU001				
Nome do Discente:	Agaone Donizete Silva				
Título do Trabalho:	Previsão das propriedades do biodiesel e seu perfil ideal de ácidos graxos por meio de aprendizado de máquina				
Área de concentração:	Desenvolvimento de Processos Químicos				
Linha de pesquisa:	Modelagem, controle e otimização de processos químicos				
Projeto de Pesquisa de vinculação:	Projeto auxiliado por computador de processos industriais				
ODS-ONU:	ODS 7 - Energia Limpa e Acessível, ODS 9 - Indústria, Inovação e Infraestrutura				

Reuniu-se por meio de webconferência, a Banca Examinadora designada pelo Colegiado do Programa de Pós-graduação em Engenharia Química, assim composta: Professores Doutores: Diego Andrade Lemos - DEQ/UFSCar, Rubens Gedraite - PPGEQ/UFU e Sarah Arvelos Altino - PPGEQ/UFU, orientadora do candidato.

Iniciando os trabalhos a presidente da mesa, Prof<sup>a</sup> Dr<sup>a</sup> Sarah Arvelos Altino, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao discente a palavra para a exposição do seu trabalho. A duração da apresentação do discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir, a presidente concedeu a palavra, pela ordem sucessivamente aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

Aprovado

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente

ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Sarah Arvelos Altino, Professor(a) do Magistério Superior**, em 26/09/2024, às 11:24, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Rubens Gedraite, Professor(a) do Magistério Superior**, em 26/09/2024, às 11:25, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Diego Andrade Lemos, Usuário Externo**, em 26/09/2024, às 11:25, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://www.sei.ufu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **5716824** e o código CRC **DE999F42**.

**Referência:** Processo nº 23117.062653/2024-56

SEI nº 5716824

*Dedico este trabalho à minha esposa Jane que sempre me apoiou nas minhas decisões e ao meu filho Bryan que incentivou o meu trabalho.*

## **AGRADECIMENTOS**

*Agradeço à professora Dra Sarah Arvelos Altino pelo incentivo, motivação e sua amizade, no decorrer da orientação nesta caminhada acadêmica.*

*Aos colegas, que trilharam comigo nessa jornada, sempre em busca de conhecimentos.*

*Agradeço também às coordenadoras do curso de Pós-Graduação em Engenharia Química da Universidade Federal de Uberlândia por permitirem o meu trancamento.*



*"Qualquer caminho é apenas um caminho e não constitui insulto algum – para si mesmo ou para os outros – abandoná-lo quando assim ordena o seu coração. Olhe cada caminho com cuidado e atenção. Tente-o tantas vezes quantas julgar necessárias... Então, faça a si mesmo e apenas a si mesmo uma pergunta: possui esse caminho um coração? Em caso afirmativo, o caminho é bom. Caso contrário, esse caminho não possui importância alguma."*

Carlos Castenada

## RESUMO

O biodiesel é um combustível produzido a partir de recursos renováveis, como óleos vegetais e gorduras animais. Ele é considerado uma alternativa mais limpa e sustentável ao diesel convencional, que é produzido a partir de petróleo. O biodiesel pode ser definido em termos das misturas de ésteres como um combustível composto por ésteres monoalquil de ácidos graxos de cadeia longa, os quais são produzidos, em geral por uma reação de transesterificação. Esta dissertação aborda a previsão das propriedades do biodiesel e seu perfil ideal de ácidos graxos por meio de aprendizado de máquina. As propriedades de interesse são a densidade, a viscosidade cinemática e o calor de combustão. O estudo buscou criar modelos que prevejam as propriedades do biodiesel com base em sua composição de ácidos graxos, permitindo identificar as melhores composições para o biocombustível em questão. O modelo empregado foi o de k-vizinhos mais próximos (KNN), um regressor clássico dentro do aprendizado de máquina. A metodologia detalha a base de dados utilizada para criação do modelo, a qual foi consultada na literatura e expandida via interpolação, computando 136 amostras e sua composição em termos de ésteres derivados de ácidos de 8 até 18 carbonos na cadeia, contendo insaturações ou não. Foi realizada a análise básica dos dados, visando-se comparar o banco de dados original e o banco de dados expandido. O modelo KNN foi preciso em realizar a regressão das 3 propriedades de interesse. Métodos de interpretabilidade foram utilizados para checar sua aderência ao fenômeno físico, a saber: escalas de importância e gráficos de dependência parcial. Observou-se que o modelo é capaz de representar de forma razoável os efeitos esperados em termos de composição de ésteres: misturas ricas em ácidos graxos saturados de cadeia curta são menos viscosas, mais densas e tem menor calor de combustão. O método de evolução diferencial permitiu buscar por composições de mistura que rendem os maiores valores possível de calor de combustão, estando em torno de 40 MJ/kg.

**Palavras-chave:** biodiesel; ácidos graxos; aprendizado de máquina; propriedades físico-químicas.

## ABSTRACT

Biodiesel is a fuel produced from renewable resources, such as vegetable oils and animal fats. It is considered a cleaner and more sustainable alternative to conventional diesel, which is produced from petroleum. Biodiesel can be defined in terms of ester mixtures as a fuel composed of monoalkyl esters of long-chain fatty acids, which are generally produced by a transesterification reaction. This dissertation addresses the prediction of biodiesel properties and its ideal fatty acid profile through machine learning. The properties of interest are density, kinematic viscosity, and heat of combustion. The study aimed to create models that predict the properties of biodiesel based on its fatty acid composition, allowing the identification of the best compositions for the biofuel in question. The model employed was the k-nearest neighbors (KNN), a classic regressor within machine learning. The methodology details the database used for model creation, which was consulted in the literature and expanded via interpolation, computing 136 samples and their composition in terms of esters derived from acids with 8 to 18 carbons in the chain, containing unsaturations or not. Basic data analysis was performed to compare the original database and the expanded database. The KNN model was accurate in performing the regression of the 3 properties of interest. Interpretability methods were used to check its adherence to the physical phenomenon, namely: importance scales and partial dependence plots. It was observed that the model is capable of reasonably representing the expected effects in terms of ester composition: mixtures rich in short-chain saturated fatty acids are less viscous, denser, and have lower heat of combustion. The differential evolution method allowed the search for mixture compositions that yield the highest possible values of heat of combustion, around 40 MJ/kg.

Keywords: biodiesel; fatty acids; machine learning; physicochemical properties.

## LISTA DE QUADROS

Quadro 1. Propriedades complementares atribuídas ao biodiesel em comparação ao óleo diesel.....	23
Quadro 2. Principais matérias-primas utilizadas para produção de biodiesel em diferentes países .....	25
Quadro 3. Esquema da reação global de transesterificação .....	27
Quadro 4. Principais ácidos graxos presentes em óleos vegetais e gorduras animais.....	27
Quadro 5. Principais ésteres metílicos presentes no biodiesel .....	28
Quadro 6. Padrões para algumas propriedades de biodieseis segundo a EN 14214, ASTM D6751 e ANP 45/2014. ....	29
Quadro 7. Características gerais dos principais trabalhos publicados com o objeto de previsão de viscosidade cinemática de biodieseis pelo uso de modelos de aprendizado de máquina.....	35
Quadro 8. Características gerais dos principais trabalhos publicados com o objeto de previsão do calor de combustão de biodieseis pelo uso de modelos de aprendizado de máquina.....	37
Quadro 9. Características gerais dos principais trabalhos publicados com o objeto de previsão da densidade de biodieseis pelo uso de modelos de aprendizado de máquina. ....	38

## LISTA DE ILUSTRAÇÕES

Figura 1. Histogramas representando a propriedade viscosidade cinemática do banco de dados original (a) e aumentado (b); porcentagem de C8 no banco de dados original (c) e aumentado (d). .....	49
Figura 2. Gráfico de pares para viscosidade cinemática, calor de combustão e densidade, respectivamente - banco de dados original. ....	50
Figura 3. Gráfico de pares para viscosidade cinemática, calor de combustão e densidade, respectivamente - banco de dados interpolado. ....	51
Figura 4. Mapa de calor para o coeficiente de Pearson para frações de ésteres, viscosidade cinemática, calor de combustão e densidade, respectivamente – (a) Banco de dados original; (b) Banco de dados interpolado. ....	52
Figura 5. Evolução do valor da função objetivo para os 500 passos de busca para uma simulação do modelo KNN para a estimativa do valor da densidade do conjunto de treino. ....	54
Figura 6. Valores reais <i>versus</i> valores preditos para (a) densidade, (b) calor de combustão e (c) viscosidade cinemática do conjunto de teste utilizando o modelo KNN. ....	56
Figura 7. Importâncias das características para previsão da (a) densidade, (b) calor de combustão e (c) viscosidade cinemática do conjunto de teste utilizando o modelo KNN. ...	59
Figura 8. Gráficos de dependência parcial das características avaliadas no modelo KNN para o ajuste de densidade do conjunto de teste. ....	60
Figura 9. Gráficos de dependência parcial das características avaliados no modelo KNN para o ajuste do calor de combustão do conjunto de teste. ....	61
Figura 10. Gráficos de dependência parcial das características avaliados no modelo KNN para o ajuste da viscosidade do conjunto de teste. ....	62

## LISTA DE TABELAS

Tabela 1. Banco de dados: Composição em ésteres metílicos de 34 amostras de biodiesel	40
Tabela 2. Grandes estatísticas básicas do banco de dados original (34 dados) e do banco aumentado (136 dados).....	53
Tabela 3. Parâmetros ótimos para o modelo KNN e qualidade do ajuste obtidos nas etapas de treinamento e teste. ....	55
Tabela 4. Resultados obtidos para a composição de ésteres visando maximizar o calor de combustão da mistura. Valores calculados de densidade e viscosidade para cada uma das composições obtidas pelo algoritmo de evolução diferencial. ....	63

## LISTA DE ABREVIATURAS E SIGLAS

ABE	Análise Básica Exploratória
ABNT	Associação Brasileira de Normas Técnicas
AC	Ácidos Graxos
AD	Árvores de Decisão
AE	Análise Exploratória
AM	Aprendizado de Máquina
ANN	Redes Neurais Artificiais
ANP	Agência Nacional do Petróleo
AR	Árvores de Regressão
ASTM	<i>American Society for Testing and Material</i>
CCP	Coefficiente de Correlação de Pearson
CFPP	Ponto de Entupimento
CMS	Carbon Molecular Sieves
CN	Número de Cetano
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CNTP	Condições Normais de Temperatura e Pressão
CP	Coefficiente de Pearson
DE	Evolução Diferencial (do inglês <i>Differential Evolution</i> )
EN	European Standards
FA	Florestas Aleatórias
FAEEs	Ésteres Etilícos de Ácidos Graxos
FAMEs	Ésteres Metílicos de Ácidos Graxos
GP	Programação Genética
HM	Heatmap (Mapa de calor)
IA	Inteligência Artificial
IP	Importâncias de Permutação
KNN	KNowledge Now (Conhecimento agora)
MAE	Erro Absoluto Médio (do inglês <i>Mean Absolute Error</i> )
ML	<i>Machine Learning</i>
MSE	Erro Quadrático Médio (do inglês <i>Mean Squared Error</i> )
PGR	Processo Gaussiano de Regressão
$R^2$	Coefficiente de Determinação
RF	Random Forest
RLM	Regressão Linear Múltipla

RNAs	Redes Neurais Artificiais
RVS	Regressão de Vetores de Suporte
SGB	Stochastic Gradient Boosting
SVM	Máquina de Vetores de Suporte ( <i>Support Vector Machine</i> )
TAE	Teoria do Aprendizado Estatístico
UE	União Europeia
UFU	Universidade Federal de Uberlândia
USA	<i>Unit States of American</i>
VC	Validação Cruzada



## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>18</b>
<b>2. OBJETIVOS.....</b>	<b>21</b>
2.1. Objetivo Geral.....	21
2.2. Objetivos Específicos .....	21
<b>3. REVISÃO DE LITERATURA.....</b>	<b>22</b>
3.1. Introdução ao Biodiesel.....	22
3.1.1. Definições .....	22
3.1.2. Importância do biodiesel como fonte de energia renovável .....	23
3.2. Produção de biodiesel.....	25
3.3. Padrão de qualidade dos biodieseis.....	28
3.4. Necessidade de Modelagem.....	30
3.5. Aprendizado de Máquina em Modelagem de Propriedades de Biodieseis.....	32
3.6. Previsão da viscosidade cinemática de biodieseis .....	33
3.7. Previsão do calor de combustão de biodieseis.....	34
3.8. Previsão da densidade de biodieseis.....	36
<b>4. METODOLOGIA.....</b>	<b>39</b>
4.1. Base de dados e características.....	39
4.2. Ambiente computacional e expansão do banco de dados.....	41
4.3. Análise básica dos dados .....	42
4.4. Algoritmos de Aprendizado de Máquina .....	43
4.5. Avaliação do Modelo .....	44
4.6. Avaliação da composição ótima.....	46
<b>5. RESULTADOS E DISCUSSÃO.....</b>	<b>48</b>
5.1. Análise básica dos dados .....	48
5.2. Resultados para o modelo de aprendizado de máquina.....	54
<b>6. CONCLUSÃO E CONSIDERAÇÕES FINAIS.....</b>	<b>65</b>
<b>7. REFERENCIAS.....</b>	<b>66</b>

## 1. INTRODUÇÃO

Biodiesel é um combustível produzido a partir de recursos renováveis, como óleos vegetais e gorduras animais, que pode ser usado em motores a diesel sem a necessidade de modificações significativas. Ele é considerado uma alternativa mais limpa e sustentável ao diesel convencional produzido a partir de petróleo. O biodiesel é comumente produzido por meio de um processo chamado transesterificação, no qual o óleo ou gordura é combinado com um álcool, geralmente metanol, para produzir ésteres de ácidos graxos, que são então separados da glicerina. O biodiesel pode ser usado puro (B100) ou misturado com diesel convencional em diferentes proporções, como B20 (20% de biodiesel e 80% de diesel) (ISSARIYAKUL; DALAI, 2014).

O uso de biodiesel tem como benefícios a redução das emissões de gases de efeito estufa, a diminuição da dependência do petróleo e a promoção da agricultura local. Essas vantagens do biodiesel em relação ao diesel fóssil têm despertado a atenção mundial em relação à aplicação do biodiesel no setor de transporte e como um potencial substituto do diesel fóssil. Isso é evidenciado pelo fato de que, em 2020, a produção global de biodiesel atingiu 143,9 bilhões de litros, aproximadamente o dobro do que foi produzido em 2010 (IEA, 2020; SAJJADI; RAMAN; ARANDIYAN, 2016).

Algumas das propriedades importantes do biodiesel que são comumente estudadas e frequentemente relatadas na literatura incluem viscosidade, densidade, número de cetano, poder calorífico, ponto de turvação, ponto de fluidez, ponto de entupimento a frio, ponto de fulgor e estabilidade à oxidação (BUKKARAPU; KRISHNASAMY, 2022; KRISHNASAMY; BUKKARAPU, 2021). Devido à variabilidade das propriedades do biodiesel produzido em todo o mundo, a conformidade regulatória e os padrões foram estabelecidos para garantir a qualidade do produto e a confiança do usuário. Os dois padrões amplamente referidos são as Especificações Padrão Americanas para Combustível Biodiesel (B100) para Combustíveis Destilados, ASTM D6751, e o Padrão Europeu para Biodiesel, EN14214 (AFDC, [s.d.]; EUROPEAN STANDARDS, [s.d.]). A qualidade do biodiesel no Brasil é regulamentada pela Resolução ANP (Agência Nacional do Petróleo, Gás Natural e Biocombustíveis) nº 920/2023. Esta resolução estabelece as especificações e as diretrizes

para o controle de qualidade do biodiesel no país (AGÊNCIA NACIONAL DO PETRÓLEO, 2023).

Na literatura, é indicado que as propriedades físico-químicas do biodiesel variam principalmente devido a diferenças na composição de seus ésteres de ácidos graxos (FAEs: *fatty acid esters*), que por sua vez são influenciados pela composição dos ácidos graxos presentes na matéria-prima utilizada. O biodiesel pode ser produzido a partir de mais de centenas de culturas oleaginosas (comestíveis e não comestíveis), gorduras animais e lipídios de algas, contribuindo ainda mais para sua variabilidade (MAHESHWARI et al., 2022).

Dada a importância da composição de FAEs na determinação das propriedades do combustível, a comunidade de pesquisa em biodiesel se concentrou em duas questões-chave. Em primeiro lugar, é possível prever as propriedades do biodiesel diretamente com base em sua composição de FAEs ou ácidos graxos? Em segundo lugar, se tal relação existir, é possível determinar a composição ótima de ácidos que atenda aos requisitos regulatórios das propriedades do combustível por meio de engenharia reversa? (AKRAM et al., 2022; PYDIMALLA et al., 2023; SUVARNA et al., 2022).

Embora ambas as questões sejam proeminentes, elas são frequentemente tratadas como tópicos de pesquisa individuais e raramente houve um esforço para trabalhar neles coletivamente. No contexto para otimizar a composição de ácidos graxos do biodiesel; vários pesquisadores têm misturado os biodieseis transesterificados de várias matérias-primas, a fim de melhorar a estabilidade oxidativa, a viscosidade e as propriedades de fluxo de baixa temperatura do biodiesel (BRAHMA et al., 2022; SAJJADI; RAMAN; ARANDIYAN, 2016).

Na era da análise centrada em dados, conceitos e ferramentas de aprendizado de máquina (*Machine Learning*, ML) podem ser usados para tratar desses dois objetivos de pesquisa a partir de um único ponto de vista. A aplicação de ML na pesquisa de biodiesel não é nova. Uma revisão recente descreveu os principais trabalhos sobre a aplicação de ML na pesquisa de biodiesel ao longo das últimas três décadas e concluiu que as redes neurais artificiais (ANN) estão entre os algoritmos mais investigados nesse domínio. No entanto, apesar de uma pesquisa substancial nesse campo, lacunas definitivas ainda existem. Em primeiro lugar, grande parte dos trabalhos anteriores se concentrou exclusivamente no

desenvolvimento de modelos preditivos com alta precisão, sem muita ênfase na generalização do modelo. Além disso, exceto por alguns estudos, houve poucos relatos sobre a aplicação e adequação de algoritmos de ML básicos neste campo de pesquisa (AGHBASHLO et al., 2021).

Em termos de previsão e modelos de aprendizado de máquina, o único trabalho disponível na literatura (até onde se sabe) que trata da previsão de propriedades e ao mesmo tempo de atendimento aos órgãos regulatórios, foi publicado recentemente (SUVARNA et al., 2022). Neste trabalho, os autores desenvolveram modelos preditivos para o número de cetano (CN) e o ponto de entupimento a frio (CFPP), onde os algoritmos *extreme gradient boost* (XGB) e *random forest* (RF) tiveram o melhor desempenho. Um modelo classificador para a estabilidade oxidativa (OS) foi elaborado para prever se ela atenderia ou não aos limites da ASTM e da UE.

## 2. OBJETIVOS

### *2.1. Objetivo Geral*

Desenvolver modelos preditivos para mapear as propriedades do biodiesel como uma função de suas composições de ácidos graxos que possam ser usados para engenharia reversa e determinar a(s) composição(ões) ótima(s) do biocombustível de interesse.

### *2.2. Objetivos Específicos*

- Analisar um banco de dados diverso contendo perfis de ácidos graxos e as propriedades chave delimitadas nas normas ASTM D6751, EN 14214 e ANP 45/2014;
- Aplicar o algoritmo KNN, para previsão de propriedades de biodieseis, a saber: densidade, calor de combustão e viscosidade cinemática;
- Aplicar técnicas de interpretabilidade de modelos para avaliar a aderência dos modelos aos dados: análise de importâncias de permutação e gráficos de dependência parcial.

### 3. REVISÃO DE LITERATURA

#### 3.1. *Introdução ao Biodiesel*

##### 3.1.1. Definições

A ideia de usar biocombustíveis em motores a diesel surgiu com Rudolf Diesel em 1900, na Feira Mundial de Paris. Óleo de amendoim foi usado como combustível, mostrando a possibilidade de usar óleos locais. O diesel fóssil é amplamente usado devido à sua adaptabilidade e alta eficiência em diversos setores. O motor a diesel queima o combustível pelo calor da compressão do ar no cilindro, aproveitando suas propriedades favoráveis. O esgotamento dos recursos é uma preocupação, levando agricultores a buscar alternativas para a produção agrícola (OGUNKUNLE; AHMED, 2019).

O biodiesel tornou-se um candidato atraente para consideração na busca por combustíveis alternativos. É um combustível biodegradável, renovável, não tóxico e ecologicamente correto que libera quantidades negligenciáveis de hidrocarbonetos aromáticos e sulfatos na combustão. As matérias-primas necessárias para sua síntese estão disponíveis durante todo o ano e geralmente não são muito caras. A Sociedade Americana para Testes e Materiais, International (ASTM International) definiu o biodiesel como "um combustível composto por ésteres monoalquil de ácidos graxos de cadeia longa, derivados de óleos vegetais renováveis ou gorduras animais, que atendem aos requisitos da ASTM D6751" (AKRAM et al., 2022). O nome "Biodiesel" foi popularizado nos Estados Unidos em 1992 pela *National Soy Diesel Development Board*, o que abriu caminho para a comercialização do biodiesel (KARMAKAR; KARMAKAR; MUKHERJEE, 2010).

Muitos pesquisadores testaram a combustão de biodiesel em motores a diesel de automóveis usando diferentes proporções de misturas de biodiesel com diesel. De acordo com alguns relatórios, misturas de combustível acima de 20% geraram problemas de manutenção e, ocasionalmente, danificaram o motor. Biodieseis em misturas mais baixas, B2 e B5, no combustível diesel têm sido usados com sucesso em motores a diesel existentes,

bem como em novos projetos saindo da linha de montagem (OGUNKUNLE; AHMED, 2019). Além de ter seu uso plausível em motores, há benefícios ambientais adicionais pelo uso do mesmo nos motores. Como, por exemplo, quando comparado ao diesel de petróleo, as características do biodiesel reduzem as emissões de gases de escape (VENKATESAN et al., 2023). No Quadro 1 são apresentadas estas vantagens adicionais.

Quadro 1. Propriedades complementares atribuídas ao biodiesel em comparação ao óleo diesel.

<b>Características</b>	<b>Propriedades complementares</b>
Características químicas apropriadas	Livre de enxofre e compostos aromáticos, alto número de cetanos, ponto de combustão apropriado, excelente lubricidade, não tóxico e biodegradável
Ambientalmente benéfico	Nível de toxicidade compatível ao sal ordinário, com diluição tão rápida quanto a do açúcar (Departamento de Agricultura dos Estados Unidos). Reduz sensivelmente as emissões de: (a) partículas de carbono (fumaça), (b) monóxido de carbono, (c) óxidos sulfúricos e (d) hidrocarbonetos policíclicos aromáticos.
Economicamente competitivo	Complementa todas as novas tecnologias do diesel com desempenho similar e sem a exigência da instalação de uma infraestrutura ou política de treinamento.
Reduz aquecimento global	O gás carbônico liberado é absorvido pelas oleaginosas durante o crescimento, o que equilibra o balanço negativo gerado pela emissão na atmosfera.
Economicamente atraente	Permite a valorização de subprodutos de atividades agroindustriais, aumento na arrecadação regional de impostos, aumento da fixação do homem ao campo e de investimentos complementares em atividades rurais.
Regionalização	Pequenas e médias plantas para produção de biodiesel podem ser construídas.

Fonte: Ramos (1999)

### 3.1.2. Importância do biodiesel como fonte de energia renovável

O desenvolvimento da economia requer enormes quantidades de insumos e de recursos naturais, especialmente energia, que é considerada o principal impulsionador do crescimento econômico. Somente em 2014, o consumo mundial de energia cresceu a, aproximadamente, 13 bilhões de toneladas equivalentes de petróleo, aumentando em 22% e 54%, respectivamente, em comparação com 2004 e 1994. A demanda rapidamente crescente acelerou a exploração de recursos energéticos do meio ambiente natural, trazendo desafios graves de escassez de energia e mudança climática. Esperava-se que as reservas provadas de petróleo, gás natural e carvão do mundo, até o final de 2014, atendessem a 53, 54 e 110 anos de produção global, respectivamente. A conservação dos recursos energéticos tornou-se, portanto, uma preocupação urgente no mundo atual. Enquanto isso, como as emissões de

gases de efeito estufa relacionadas à energia representam mais de 80% das emissões antropogênicas globais, a economia de energia também é essencial para a mitigação do aquecimento global (CHEN; WU, 2017).

O biodiesel oferece uma alternativa mais ambientalmente amigável ao combustível diesel tradicional, potencialmente reduzindo as emissões de gases de efeito estufa em até 60% e gerando menos poluentes atmosféricos, como material particulado, monóxido de carbono e hidrocarbonetos. Derivado de óleos vegetais ou gorduras animais renováveis, o biodiesel promove práticas sustentáveis, contrastando com os combustíveis fósseis finitos que contribuem para as mudanças climáticas. Sua biodegradabilidade e menor toxicidade aumentam a segurança ambiental, minimizando os riscos de contaminação em caso de vazamentos. Apesar desses benefícios ecológicos, é necessária atenção para abordar preocupações de sustentabilidade associadas ao biodiesel. A expansão do cultivo de oleaginosas, especialmente em regiões tropicais, aumenta o risco de desmatamento, impactando a biodiversidade e contribuindo para as mudanças climáticas. A natureza intensiva em água da produção de biodiesel, que requer irrigação, pode levar ao estresse hídrico e conflitos em regiões escassas (SEAF ELNASR et al., 2024).

Uma variedade de fontes lipídicas pode ser usada para produzir biodiesel. Óleos vegetais, gorduras animais, algas marinhas e óleos de cozinha usados podem ser a fonte dos lipídios. Aproximadamente 80% a 90% do custo de produção de biodiesel está relacionado aos materiais brutos. As matérias-primas disponíveis devem ser fornecidas a um preço razoável para garantir a produção contínua de biodiesel. Óleo comestível também tem sido comumente usados como fonte primária de biodiesel. No mundo, o óleo comestível representa 99% do biodiesel. Os óleos comestíveis típicos usados para biodiesel são: soja (20%), óleo de girassol (15%), óleo de palma (5%) e óleo de colza (60%). No Quadro 2 são apresentadas as principais fontes de ácidos graxos para a produção de biodiesel em alguns países do mundo. A utilização de óleo vegetal e gordura animal como matéria-prima essencial na fabricação de biodiesel enfrenta obstáculos legais e ambientais notáveis. Contudo, os progressos recentes em tecnologia têm sido substanciais, especialmente no que diz respeito ao aproveitamento de óleos residuais, gorduras orgânicas e óleos não



comestíveis, abrindo novas perspectivas para a produção sustentável de biodiesel (VIGNESH et al., 2021).

Quadro 2. Principais matérias-primas utilizadas para produção de biodiesel em diferentes países

<b>País</b>	<b>Matérias-primas</b>
Argentina e Brasil	Soja
Canadá	Colza/mostarda/Gordura animal/Gordura Vegetal/Soja/Sebo
China	Óleo de cozinha usado
France	Girassol/Colza
Alemanha	Colza
Grécia	Algodão
Índia	Laranja/Pinhão manso/Colza/Soja/Amendoim/Girassol
Indonésia	Palma/coco/pinhão manso
Islândia	Gorduras animal /óleo de frituras
Itália	Girassol e Colza
Japão	Óleo usado
Malásia	Palma
México	Óleo usado/Gordura animal
Nova Zelândia	Óleo usado
Filipinas	Pinhão manso / Coco
Espanha	Linhaça/Girassol
Suécia	Colza
Tailândia	Palma/Coco/Colza
Reino Unido	Óleo usado/Colza
Estados Unidos	Soja/Amendoim/Óleo usado

Fonte: Atabani et al. (2012)

### *3.2. Produção de biodiesel*

Os principais métodos de produção de biodiesel são pirólise, microemulsificação, diluição e transesterificação. De todos esses métodos, a transesterificação permanece o

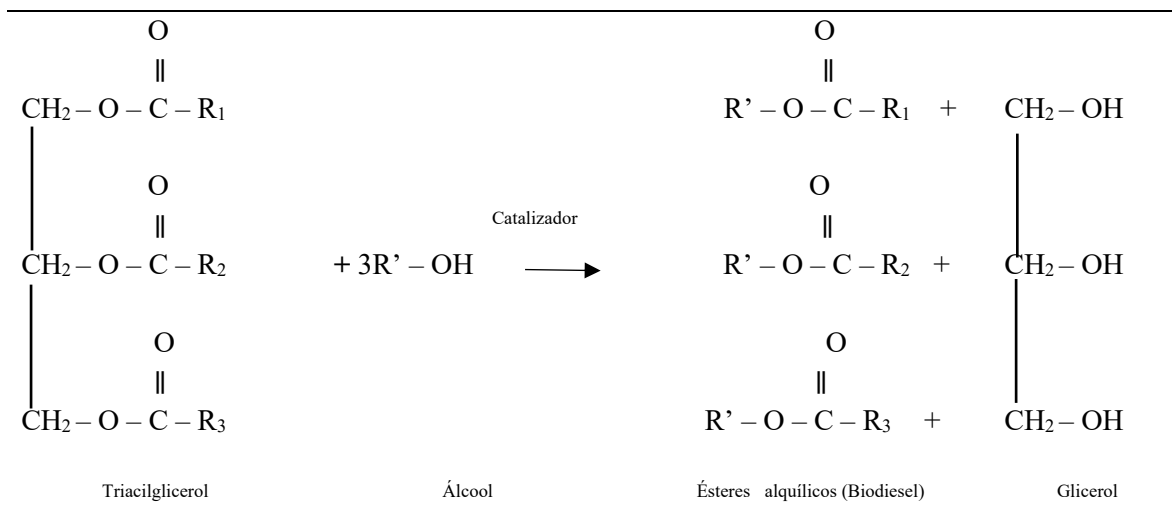
processo viável que foi adotado até agora e é o método mais comum usado na produção de biodiesel (OGUNKUNLE; AHMED, 2019).

Em uma usina de biodiesel, o processo de transesterificação compreende várias etapas. Primeiro, o óleo vegetal (ou outra fonte de ácidos graxos) é combinado com o álcool de cadeia curta, geralmente etanol (obtido da cana-de-açúcar) ou metanol (produzido a partir de biomassa de madeira) e é então tratado com um catalisador. Este catalisador facilita a reação química entre o óleo e o álcool. Posteriormente, o óleo é separado da glicerina (que pode ser usada na fabricação de sabonetes) e filtrado. Óleos e gorduras configuram-se como misturas de diversos componentes. Contudo, majoritariamente são compostos por moléculas de triglicerídeos, os quais consistem em longas cadeias de ácidos graxos ligadas a uma molécula de glicerina. Quando reagem com o álcool, as três cadeias de ácidos graxos se separam da glicerina, originando assim uma molécula de éster (LISBOA et al., 2014).

A qualidade do produto obtido na transesterificação é influenciada pelo tratamento adequado do processo para cada tipo específico de óleo ou gordura utilizados. Embora a transesterificação seja uma reação reversível em princípio, durante a produção de ésteres alquílicos de óleos ou gorduras, ou seja, biodiesel, a reversão da reação, pode ser desconsiderada devido à formação de glicerol, que não é miscível no produto final, resultando em um sistema de duas fases (KNOTHE, 2006).

Os grupos representados por  $R_1$ ,  $R_2$  e  $R_3$  no Quadro 3, que podem variar entre si, são cadeias de ácidos graxos ligados ao glicerol, sendo os ácidos mais comumente encontrados listados no Quadro 4. Na molécula do biodiesel, o grupo  $R'$  representa a cadeia de hidrocarboneto do álcool usado como reagente. A proporção dos ácidos graxos modificada as características e propriedades físico-químicas dos biodieseis. No Quadro 4 são apresentados os principais ácidos graxos presentes em óleos vegetais e gorduras animais, destacando os ácidos com 16-18 carbonos e seus nomes usuais. Nota-se que os ácidos oleico, linoleico e linolênico possuem 18 carbonos em sua cadeia. Contudo, o ácido oleico é insaturado e os ácidos linoleico e linolênico não o são. No Quadro 5 são apresentados os nomes dos principais ésteres metílicos de ácidos graxos, produzidos, então, a partir de metanol.

Quadro 3. Esquema da reação global de transesterificação



Fonte: Adaptado de Lisboa et al. (2014)

Quadro 4. Principais ácidos graxos presentes em óleos vegetais e gorduras animais

Ácido	Fórmula Química
Palmítico	R = $-(\text{CH}_2)_{14} - \text{CH}_3$
Esteárico	R = $-(\text{CH}_2)_{16} - \text{CH}_3$
Oléico	R = $-(\text{CH}_2)_7 - \text{CH} = \text{CH} - (\text{CH}_2)_7 - \text{CH}_3$
Linoléico	R = $-(\text{CH}_2)_7 - \text{CH} = \text{CH} - \text{CH}_2 - \text{CH} = \text{CH} - (\text{CH}_2)_4 - \text{CH}_3$
Linolênico	R = $-(\text{CH}_2)_7 - \text{CH} = \text{CH} - \text{CH}_2 - \text{CH} = \text{CH} - \text{CH}_2 - \text{CH} = \text{CH} - \text{CH}_2 - \text{CH}_3$

Fonte: Atabani et al. (2012)

Quadro 5. Principais ésteres metílicos presentes no biodiesel

<b>Ésteres Metílicos de Ácidos Graxos Nomenclatura (FAME)</b>	
<b>Ésteres saturados de cadeia curta</b>	
Hexanoato de Metila	C6:0
Heptanoato de Metila	C7:0
Octonoato de Metila	C8:0
Decanoato de Metila	C10:0
Laurato de Metila	C12:0
Miristato de Metila	C14:0
<b>Ésteres saturados de cadeia longa</b>	
Palmitato de Metila	C16:0
Margarato de Metila	C17:0
Estearato de Metila	C18:0
Araquidato de Metila	C20:0
Behenato de Metila	C22:0
Lignocerato de Metila	C24:0
<b>Ésteres insaturados</b>	
Sorbato de Metila	C6:2
Palmitoleato de Metila	C16:1
Oleato de Metila	C18:1
Linoleato de Metila	C18:2
Linolenato de Metila	C18:3
Eicosenoato de Metila	C20:1
Eicosadienoato	C20:2
Erucado Metílico	C22:1

Fonte: Atabani et al. (2012)

### 3.3. Padrão de qualidade dos biodieseis

O biodiesel pode ter padrões de especificação diferentes em várias partes do mundo. Esses padrões variam dependendo das legislações e regulamentações locais, as características dos veículos, as condições climáticas, e as matérias-primas disponíveis para a produção de biodiesel.

Na Europa, os padrões de qualidade para biodiesel são principalmente definidos pela norma EN 14214. Nos Estados Unidos, o padrão de especificação para biodiesel é o ASTM

D6751<sup>1</sup> (SAKTHIVEL et al., 2018). No Brasil, o biodiesel deve atender à norma ANP 45/2014<sup>2</sup>.

Sobre estas normas, destacam-se os limites mínimos e máximos apresentados no Quadro 6. É importante salientar que várias outras propriedades são listadas nas normas, as quais, por conveniência, foram omitidas.

Quadro 6. Padrões para algumas propriedades de biodieseis segundo a EN 14214, ASTM D6751 e ANP 45/2014.

Propriedade	Unidade	EN 14214	ASTM D6751	ANP 45/2014
Ponto de fulgor	°C	Mínimo de 101	Mínimo de 130	Mínimo de 100
Densidade à 15 °C	kg/m <sup>3</sup>	860-900	Max 880	850 - 900
Viscosidade à 40 °C	mm <sup>2</sup> /s	3,5 – 5,0	1,9 - 6,0	3,0 - 6,0
Número de cetano	-	Mínimo de 51	Mínimo de 47	-
Estabilidade oxidativa	-	Mínimo de 3h	-	Mínimo de 8 h
Ponto de bolha	°C	-	100 - 615	-
Teor de ácido linoleico	g/g	Máximo de 12%	-	-

Fonte: Adaptado de Sakthivel et al. (2018)

No Quadro 6 é destacado que algumas propriedades podem ter limite monitorado apenas por algumas agências. Além disso, os limites permitidos podem variar de uma para a outra. Dentre as propriedades interesse desta dissertação, estão a densidade, a viscosidade e o poder calorífico.

A densidade é uma propriedade essencial do combustível, importante para calcular a quantidade de combustível injetada pelos sistemas de injeção, garantindo a combustão adequada. A densidade do biodiesel varia conforme a matéria-prima, o método de produção e o perfil dos ésteres metílicos. Ela é vital no dimensionamento dos bicos injetores, afetando

<sup>1</sup> <https://www.astm.org/d6751-20a.html>

<sup>2</sup> <https://atosoficiais.com.br/anp/resolucao-n-45-2014?origin=instituicao&q=45/2014>

diretamente a operação do motor e a atomização do combustível, impactando a eficiência térmica. A densidade do biodiesel é um dos maiores obstáculos para sua ampla utilização (GÜLÜM; BILGIN, 2015; SINGH et al., 2024).

A viscosidade é uma das características fundamentais de um combustível, que indica sua capacidade de fluir. Como representa a resistência ao fluxo, a viscosidade desempenha um papel importante na atomização e penetração do biodiesel vaporizado. Devido à sua estrutura química maior e massa molecular elevada, a viscosidade do biodiesel é de 10 a 15 vezes maior do que a dos combustíveis fósseis convencionais. Combustíveis mais viscosos causam atomização insuficiente, resultando em menor eficiência térmica e depósitos de fuligem. Por outro lado, uma viscosidade reduzida leva à formação de gotículas de combustível mais finas, facilitando a injeção do combustível na câmara de combustão. O processo de transesterificação é amplamente utilizado para reduzir a viscosidade dos ésteres metílicos (FERREIRA et al., 2021; GÜLÜM; BILGIN, 2015; SINGH et al., 2024).

Calor de combustão (CC) é precisamente definido como a quantidade de energia liberada pela combustão de uma unidade de valor de combustíveis. Portanto, um CC mais alto é um fator desejável para um motor de combustão interna. O CC biodiesel é inferior ao do diesel convencional. Embora o Calor de combustão não seja especificado nas normas ASTM D6751 e EN 14214, ele é prescrito na norma EN 14213 (biodiesel para fins de aquecimento) com um valor mínimo de 35 MJ/kg (ATABANI et al., 2013).

### *3.4.Necessidade de Modelagem*

A variabilidade na composição do biodiesel resulta em variações nas propriedades físicas e químicas. Com base nas pesquisas realizadas até agora em diferentes tipos de motores, conclui-se que a substituição do diesel por biodiesel resulta em menores emissões de monóxido de carbono (CO) e hidrocarbonetos não queimados (HC), maior consumo específico de combustível e uma pequena penalidade nas emissões de óxidos de nitrogênio (NOx). As variações no tipo de ésteres metílicos e suas proporções relativas no biodiesel são responsáveis para influenciar as propriedades fundamentais do combustível, como

viscosidade, densidade e tensão superficial, e, portanto, as características de pulverização do combustível. Consequentemente, motores a diesel operados com biodiesel produzido a partir de diferentes matérias-primas resultam em características de desempenho e emissão com variações significativas (KRISHNASAMY; BUKKARAPU, 2021).

A predição das propriedades físico-químicas do biodiesel é fundamental por diversas razões. Em primeiro lugar, é essencial para garantir a qualidade do produto final, assegurando que o biodiesel produzido atenda aos padrões de qualidade e especificações regulatórias, o que é crucial para seu uso como combustível. Além disso, a capacidade de prever as propriedades físico-químicas permite otimizar os processos de produção, tornando-os mais eficientes e eficazes na obtenção das características desejadas do biodiesel. Isso também pode contribuir para a redução de custos associados aos testes laboratoriais extensivos (SAXENA; JAWALE; JOSHIPURA, 2013; YU; ZHAO, 2021a).

Outro aspecto importante é o desenvolvimento de novas fontes de matéria-prima para a produção de biodiesel. Ao prever as propriedades físico-químicas a partir de diferentes matérias-primas, é possível explorar novas fontes de matéria-prima, ampliando as opções disponíveis e promovendo a sustentabilidade. Além disso, a predição das propriedades físico-químicas ajuda a garantir a segurança e a confiabilidade do processo de produção, minimizando o risco de falhas ou produtos inadequados (ELGHARBAWY et al., 2021; YAŞAR, 2020; YU; ZHAO, 2021b).

A medição direta das propriedades físico-químicas do biodiesel enfrenta vários desafios. Um dos principais desafios é a variedade de propriedades que precisam ser medidas, como densidade, viscosidade, ponto de fulgor e teor de água, entre outras. Essas propriedades podem ser influenciadas pela complexidade das amostras de biodiesel, que podem conter uma variedade de compostos e impurezas, dificultando a medição precisa. Além disso, algumas propriedades exigem equipamentos específicos e técnicas laboratoriais especializadas, o que pode tornar o processo demorado e caro. Garantir a padronização e a precisão das medições também é um desafio, devido à variedade de métodos disponíveis e à necessidade de calibração adequada dos equipamentos (BHARTI et al., 2024; SINGH et al., 2024).

### *3.5. Aprendizado de Máquina em Modelagem de Propriedades de Biodiesel*

A Inteligência Artificial (IA) tem sido cada vez mais aplicada para a resolução de problemas em diversos setores da economia e da pesquisa científica e tecnológica, também com várias aplicações bem-sucedidas na gestão pública e atividades que levem a benefícios sociais. Dentro da IA, uma área de estudo que merece destaque especial é a de Aprendizado de Máquina (AM). Aprendizado de Máquina é uma área de pesquisa da Inteligência Artificial que visa ao desenvolvimento de programas de computador com a capacidade de aprender a executar uma dada tarefa com sua própria experiência (FACELI et al., 2011).

Isso resulta na concepção de programas com capacidade de aprender autonomamente, utilizando um conjunto de dados que representa experiências anteriores. Essa é uma área de pesquisa interdisciplinar que abrange inteligência artificial, probabilidade e estatística, teoria da complexidade computacional, teoria da informação, filosofia, psicologia, neurobiologia, entre outros campos. Exemplos de tarefas de Aprendizado de Máquina incluem classificação e agrupamento de dados, e previsão de séries temporais. Para ilustrar o conceito básico de Aprendizado de Máquina, podemos considerar um programa de computador encarregado de realizar uma tarefa simples, como distinguir entre três variedades diferentes de flores de uma mesma espécie. Em vez de codificar um programa utilizando todo o conhecimento sobre as variedades da flor em questão, são apresentadas ao programa características botânicas das três flores. Esse programa implementa um algoritmo de Aprendizado de Máquina, que, por meio de um processo de treinamento, aprende a caracterizar uma flor com base nessas características. Assim como os seres humanos aprendem a diferenciar as variedades de flores observando suas características, o programa de Aprendizado de Máquina também aprende a tarefa por meio dessas características (GERON, 2019).



### *3.6. Previsão da viscosidade cinemática de biodieseis*

A viscosidade é um dos parâmetros físicos mais críticos do biodiesel, pois afeta diretamente a qualidade da atomização do combustível nos motores a diesel. Este parâmetro, altamente dependente da temperatura, impacta significativamente a qualidade da mistura ar-combustível, o comportamento da combustão, os parâmetros de desempenho e as características de emissão dos motores a diesel que operam com combustível biodiesel. Apesar de a viscosidade do biodiesel ser uma ordem de magnitude menor do que a de suas matérias-primas originais, sua fluidez ainda é inferior à do petrodiesel (GUI; LEE; BHATIA, 2008; GÜLÜM; BILGIN, 2015).

Diversos fatores influenciam a viscosidade do biodiesel, como o comprimento da cadeia dos ácidos graxos, o comprimento da cadeia do álcool, o nível de saturação, a configuração das ligações duplas, a ramificação da cadeia, além da presença de impurezas (glicerídeos, glicerol, álcool, água, catalisadores) e produtos de oxidação. Um aumento no comprimento das cadeias de ácidos graxos, das cadeias de álcool e no grau de saturação tende a elevar a viscosidade do biodiesel. A presença de ramificações nas cadeias e produtos de oxidação também aumenta a viscosidade. Além disso, o biodiesel com ligações duplas cis possui menor viscosidade em comparação com aquele que possui ligações trans, enquanto a posição da ligação dupla tem pouco impacto sobre essa propriedade (FERREIRA et al., 2021; REFAAT, 2009)

Além disso, um combustível com alta viscosidade leva à formação de gotas maiores durante a injeção, prejudica a vaporização das gotas injetadas, estreita o ângulo do spray de injeção e aumenta a penetração longitudinal do spray de combustível. A alta viscosidade do biodiesel resulta em mais problemas operacionais em climas frios devido a um aumento drástico dessa propriedade à medida que a temperatura diminui. Isso, por sua vez, promove reações de polimerização, aumenta a fumaça/exaustão de partículas, aumenta os depósitos de carbono dentro do cilindro e corrói partes metálicas/não metálicas do motor, além de acelerar a diluição do óleo. Por outro lado, um combustível com viscosidade muito baixa não consegue lubrificar adequadamente os elementos móveis do sistema de injeção de

combustível, levando a vazamentos ou aumento do desgaste. A viscosidade é frequentemente relatada em termos de viscosidade dinâmica (DV) e viscosidade cinemática (KV). A DV é uma medida da resistência ou atrito interno de um líquido ao fluxo ou cisalhamento quando submetido a deformação. A KV é simplesmente a DV de um líquido dividida pela sua densidade. Considerando a importância crucial da viscosidade do biodiesel para o desempenho do motor e as características de emissão, todas as normas bem conhecidas de biodiesel frequentemente especificam os limites superior e inferior para essa propriedade (DAS et al., 2018; SONG et al., 2024; YUAN; HANSEN; ZHANG, 2009).

A literatura indica que os primeiros trabalhos a respeito da modelagem de viscosidade de biodieseis utilizaram correlações para realizar a previsão, muitas vezes empíricas ou semi-empíricas (MEJÍA; SALGADO; ORREGO, 2013; YUAN; HANSEN; ZHANG, 2009). O Quadro 7 apresenta um resumo sobre os principais trabalhos publicados na área de previsão de viscosidade de biodieseis utilizando-se modelos de aprendizado de máquina.

### *3.7. Previsão do calor de combustão de biodieseis*

Calor de combustão superior é a quantidade de calor liberada por uma determinada quantidade de combustível uma vez que ele é completamente queimado, e os produtos da combustão são resfriados de volta a uma temperatura padrão de 25 °C. A combustão completa significa que, por exemplo, todos os átomos de carbono e hidrogênio são convertidos em dióxido de carbono e água, respectivamente. Calor de combustão inferior é calculado subtraindo o calor de vaporização da água gerada durante a combustão do valor de aquecimento superior. Geralmente, o calor de combustão do biodiesel é cerca de 10% menor do que o do petrodiesel, principalmente devido ao seu oxigênio estrutural (DEMIRBAS, 2009).

Quadro 7. Características gerais dos principais trabalhos publicados com o objeto de previsão de viscosidade cinemática de biodieseis pelo uso de modelos de aprendizado de máquina.

Modelo (Referência)	Banco de dados e Qualidade do ajuste	Características gerais
RNA (MENG; JIA; WANG, 2014)	105, $R^2$ de 0,9774	O estudo revelou que a composição de ésteres metílicos de ácidos graxos (FAMES) tem um impacto significativo nas propriedades do biodiesel, especialmente na viscosidade. Os ésteres mais comuns encontrados foram C16:0, C18:0, C18:1, C18:2 e C18:3, enquanto alguns FAMES menos comuns podem ser predominantes em certas amostras. O modelo ANN foi validado com base em dados experimentais, mostrando que ele pode ser uma ferramenta eficaz para prever a viscosidade de biodiesel. Os métodos de Knothe–Steidley e Ramírez-Verduzco, propostos anteriormente, tendem a subestimar a viscosidade cinemática em muitos casos, o que pode ser atribuído à contribuição de viscosidade de componentes menores que não foram considerados adequadamente por esses métodos.
RNA (GIWA et al., 2015)	55, $R^2$ de 0,958	Os cinco constituintes mais prevalentes (FA/FAME) dos biodieseis obtidos a partir dos dados coletados foram selecionados para uso. Estes são os ácidos palmítico (C16:0), esteárico (C18:0), oleico (C18:1), linoleico (C18:2) e linolênico (C18:3).
RNA (BARRADAS FILHO et al., 2015)	198, RMSE de 0,55 mm <sup>2</sup> /s	As variáveis de entrada foram as porcentagens dos 13 FAMES mais comuns em biodieseis. O uso de RNAs para prever a viscosidade foi otimizado com sucesso. O modelo baseado em RNAs apresentou desempenho significativamente melhor em comparação aos métodos lineares descritos na literatura. Essas previsões podem ser feitas antes ou após a transesterificação, ou durante o armazenamento, fornecendo uma ferramenta útil para avaliar a matéria-prima.
SVM, ANFIS e RNA (AMINIAN; ZARENEZHAD, 2018)	191, $R^2$ de 0,9818 para ANFIS	Neste artigo, os modelos foram treinados por diferentes algoritmos de otimização e foram desenvolvidos para prever com precisão a viscosidade de biodiesel em comparação com os modelos de viscosidade de biodiesel mais utilizados na literatura. As entradas do modelo foram a fração mássica de biodiesel, a temperatura, o número médio de átomos de carbono e ligações duplas nos FAME que constituem os biodieseis.

Notavelmente, o oxigênio estrutural do biodiesel compensa parcialmente seu conteúdo energético, uma vez que impulsiona o processo de combustão em direção à conclusão. O calor de combustão do biodiesel aumenta com o aumento do número de

carbonos nas moléculas do combustível, aumentando a razão de carbono e hidrogênio em relação ao oxigênio e diminuindo o número de ligações duplas (KNOTHE, 2005). Não há especificação para o calor de combustão do biodiesel nas normas ASTM D6751 e EN 14214, embora o valor mínimo seja de 35 MJ/kg de acordo com a EN 14213.

O calor de combustão superior de um determinado combustível de biodiesel pode ser determinado experimentalmente usando um simples calorímetro e, conseqüentemente, seu calor de combustão inferior pode ser calculado matematicamente. No entanto, determinar o poder calorífico de um determinado biodiesel a priori usando modelos precisos e abordagens computacionais pode fornecer informações úteis sobre sua adequação para motores a diesel e evitar medições experimentais laboriosas e errôneas (SADRAMELI; SEAMES; MANN, 2008). O Quadro 8 apresenta um resumo sobre os principais trabalhos publicados na área de previsão de calor de combustão de biodieseis utilizando-se modelos de aprendizado de máquina.

### *3.8. Previsão da densidade de biodieseis*

A densidade é uma propriedade essencial do combustível, importante para calcular a quantidade de combustível injetada pelos sistemas de injeção, garantindo a combustão adequada. Ela é vital no design dos bicos injetores, afetando diretamente a operação do motor e a atomização do combustível, impactando a eficiência térmica. A densidade do biodiesel é um dos maiores obstáculos para sua ampla utilização (GÜLÜM; BILGIN, 2015; SINGH et al., 2024).

O comprimento da cadeia e o grau de insaturação influenciam fortemente a densidade do biodiesel. Quanto maior o grau de insaturação e menor o comprimento da cadeia, maior será a densidade do biodiesel (RAMOS et al., 2009). Embora a densidade do biodiesel possa ser facilmente medida usando equipamentos simples como picnômetro e densímetro a uma dada temperatura, este é um processo demorado porque essas medições são realizadas em várias temperaturas (ALICKE et al., 2015).

Quadro 8. Características gerais dos principais trabalhos publicados com o objeto de previsão do calor de combustão de biodieseis pelo uso de modelos de aprendizado de máquina.

Modelo (Referência)	Banco de dados e Qualidade do ajuste	Características gerais
SVM, (CORRAL BOBADILLA et al., 2018)	56, RMSE de 10,58	Através dos dados obtidos nos experimentos, foram construídos modelos de regressão para prever as propriedades do biodiesel, como viscosidade, turbidez, densidade, poder calorífico e rendimento. Modelos de regressão linear e máquinas de vetor de suporte (SVM) com diferentes kernels (linear, polinomial e radial) foram usados para criar as previsões. Os modelos de regressão mais precisos foram utilizados como base para a aplicação de algoritmos genéticos, visando otimizar o processo de produção. O foco foi maximizar o poder calorífico e o rendimento do biodiesel, ao mesmo tempo em que se minimizava a viscosidade, densidade e turbidez. A combinação ótima de variáveis foi identificada, mostrando que as melhores condições de produção foram alcançadas com uma dosagem de catalisador em torno de 1% em peso, resultando em um biodiesel com maiores valores de rendimento e poder calorífico, além de propriedades físicas aprimoradas.
RNA, (BUKKARAPU; KRISHNASAMY, 2021)	40, não informado	Os resultados mostraram que o poder calorífico aumentou com o comprimento das cadeias de ésteres metílicos e com o maior número de ligações simples. Com base nesses dados experimentais, foi desenvolvida uma abordagem de regressão linear múltipla para prever as propriedades do biodiesel a partir de sua composição, e essa foi validada com amostras adicionais. Além disso, RNAs foram utilizadas para prever as propriedades dos biodieseis com base em suas composições. Este estudo sugere que a investigação pode ser expandida para estudar outros efeitos das variações de composição no biodiesel, visando encontrar uma composição ideal para aplicações em motores automotivos.

Assim, estimar a densidade do biodiesel a partir de suas propriedades físico-químicas usando técnicas avançadas de aprendizado de máquina é um assunto de pesquisa interessante. Entre as pesquisas, destacam-se diversos trabalhos nos quais o modelo mais empregado foi o de Redes Neurais Artificiais (RNAs) conforme indica o Quadro 9.

Quadro 9. Características gerais dos principais trabalhos publicados com o objeto de previsão da densidade de biodieseis pelo uso de modelos de aprendizado de máquina.

Modelo (Referência)	Banco de dados e		Características gerais
	Qualidade do	ajuste	
SVM, (CORRAL BOBADILLA et al., 2018)	56, RMSE de	10,58	Idem Quadro 7.
RNA (GIWA et al., 2015)	55, R <sup>2</sup> de 0,994		Idem Quadro 7.
(ROCABRUNO-VALDÉS; RAMÍREZ-VERDUZCO; HERNÁNDEZ, 2015)	780, R <sup>2</sup> de 0,9921		O perfil de FAMEs, temperatura e número de átomos de carbono e número de átomos de hidrogênio foram entradas para o modelo.
RNA, (BUKKARAPU; KRISHNASAMY, 2021)	40, não informado		Idem Quadro 7.
RNA e regressão linear, (GANESHMOORTHY; MUTHUKANNAN; THIRUGNANASAMBANDAM, 2019)	90, MAPE de	5,27% para RNA	Neste estudo, a modelagem de redes neurais artificiais (RNAs) e de regressão linear foi usada para prever a densidade e a viscosidade de misturas de biodiesel de óleo de peixe com diesel em várias temperaturas. A mistura foi avaliada em diferentes frações volumétricas de biodiesel (10% a 100%), e os resultados experimentais mostraram que as propriedades de densidade e viscosidade do biodiesel de óleo de peixe se sincronizam com as do diesel em misturas B40 a temperaturas entre 40-50°C. Os resultados da modelagem mostraram que a rede neural artificial apresentou uma performance superior em relação à regressão linear.
RNA, (BALABIN; LOMAKINA; SAFIEVA, 2011)	103, R <sup>2</sup> de 0,9583		Variáveis de entrada: Dados de espectroscopia no infravermelho próximo pré-processados.
Stochastic Gradient Boosting (SGB) e Programação Genética (GP), (ABOOALI; SOLEIMANI; GHOLAMREZA-RAVI, 2020)	2117, R <sup>2</sup> de 0,9998		Neste estudo, foram desenvolvidos métodos inovadores para estimar quatro propriedades importantes de ésteres de ácidos graxos metílicos e etílicos (FAMEs e FAEEs) de cadeia longa. Variáveis de entrada: Massa molar, ponto de fusão normal de ésteres de ácidos graxos, temperatura, pressão. Utilizando abordagens matemáticas avançadas, como Stochastic Gradient Boosting (SGB) e Programação Genética (GP), foram aplicados modelos em grandes conjuntos de dados. Entre os métodos, os modelos baseados em SGB apresentaram previsões mais precisas e confiáveis em comparação com a programação genética.

## 4. METODOLOGIA

### *4.1. Base de dados e características*

Os dados a serem modelados foram consultados no trabalho de Bukkarapu e Krishasamy (2021). A seguir, descreve-se o modo de obtenção das amostras pelos autores citados. O biodiesel de palma foi adquirido de um fornecedor local de biodiesel. Ésteres metílicos puros, nomeadamente octanoato de metilo (C8:0), metilo decanoato (C10:0), laurato de metila (C12:0), miristato de metila (C14:0), palmitato de metila (C16:0), estearato de metila (C18:0), oleato de metila (C18:1) e linoleato de metila (C18:2) de 99% pureza foram adquiridos. Estes ésteres metílicos são os constituintes da maioria dos tipos de biodiesel, conforme foi citado na revisão de literatura desta dissertação.

Cada éster metílico puro foi misturado com biodiesel de palma em 10%, 15% e 20% (por volume) e agitou-se vigorosamente para garantir uma mistura homogênea. Trinta e quatro amostras de biodiesel com composição variada foram preparadas e designados conforme mostrado na Tabela 1. Se observado, a porcentagem em massa de linolenato de metila (C18:3) em biodiesel de palma puro é apenas 0,07%, como pode ser visto na Tabela - 1. Seu percentual de massa nas amostras restantes permanece zero, o que transmite quão menos significativo é. Portanto, o linolenato de metila puro não foi misturado com biodiesel de palma.

Das amostras em questão, os autores determinaram a viscosidade cinemática [ $\text{mm}^2/\text{s}$ ], o calor de combustão [ $\text{MJ}/\text{kg}$ ], a densidade [ $\text{g}/\text{cm}^3$ ] e a tensão superficial [ $\text{mN}/\text{m}$ ] usando procedimentos de teste padrão. Na presente dissertação, focou-se na modelagem de três das propriedades citadas no início deste parágrafo, exceto a tensão superficial tendo em vista que esta propriedade não é regulamentada pelas normas técnicas usadas nos Estados Unidos, no continente Europeu e no Brasil.

Tabela 1. Banco de dados: Composição em ésteres metílicos de 34 amostras de biodiesel

		Éster metílico (massa %)								
Amostra		C8:0	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2	C18:3
1.	Palma	0	0	0,74	0,93	37,20	3,20	47,42	10,37	0,07
2.	MO10	8,78	0	0	0,59	37,23	2,11	42,24	9,01	0
3.	MO15	14,9	0	0	0,35	36,66	0,81	38,66	8,60	0
4.	MO20	23,5	0	0	0,28	30,40	1,89	35,97	7,87	0
5.	MD10	0	10,32	1,27	1,37	37,72	2,64	38,18	8,40	0,05
6.	MD15	0	17,03	1,18	1,19	33,77	0,99	37,80	7,95	0,05
7.	MD20	0	22,64	1,12	1,07	30,66	2,65	34,31	7,50	0
8.	ML10	0	0,42	9,83	0,74	34,56	2,89	41,14	10,39	0,02
9.	ML15	0	0,15	16,19	0,81	33,91	2,52	36,98	9,34	0,06
10.	ML20	0	0,17	19,67	0,69	31,71	2,78	35,78	9,18	0
11.	MM10	0	1,21	0,74	8,98	32,97	3,59	43,14	9,34	0
12.	MM15	0	0,42	0,78	13,2	32,08	3,30	41,25	8,92	0
13.	MM20	0	0,25	0,75	19,9	31,47	3,07	36,75	7,72	0
14.	MP10	0	0,06	0,61	1,02	50,20	3,12	37,77	7,13	0
15.	MP15	0	0	0,43	0,92	52,85	2,89	35,76	7,12	0
16.	MP20	0	0,10	1,00	1,01	54,01	2,51	34,30	7,05	0
17.	MS10	0	0	0,53	1,10	34,15	12,93	42,86	8,39	0
18.	MS15	0	0	0,48	0,93	32,15	19,12	39,68	7,87	0
19.	MS20	0	0	0,43	0,76	29,97	23,50	37,42	7,87	0
20.	MOL10	0	0	0,32	0,74	33,37	2,77	53,79	8,97	0
21.	MOL15	0	0	0,43	0,83	32,06	2,36	56,23	8,06	0
22.	MOL20	0	0	0,32	0,75	30,40	0,99	59,87	7,65	0
23.	MLI10	0	0	0,35	0,84	31,98	3,43	41,91	21,45	0
24.	MLI15	0	0	0,35	0,71	31,47	3,02	39,34	25,13	0
25.	MLI20	0	0	0,33	0,67	29,72	3,07	37,78	28,40	0
26.	MO100	100	0	0	0	0	0	0	0	0
27.	MD100	0	100	0	0	0	0	0	0	0
28.	ML100	0	0	100	0	0	0	0	0	0
29.	MM100	0	0	0	100	0	0	0	0	0
30.	MP100	0	0	0	0	100	0	0	0	0
31.	MS100	0	0	0	0	0	100	0	0	0
32.	MOL100	0	0	0	0	0	0	100	0	0
33.	MLI100	0	0	0	0	0	0	0	100	0
34.	MLNI100	0	0	0	0	0	0	0	0	100

Fonte. Adaptado pelo autor.



#### *4.2. Ambiente computacional e expansão do banco de dados*

O banco de dados foi tratado computacionalmente, analisado e modelado no ambiente online do Google Colaboratory, utilizando a linguagem de programação Python (versão 3.7.13). Essa plataforma, oferecida pela Google, tem manipulação facilitada em aplicações de Aprendizado de Máquina, fornecendo um ambiente de desenvolvimento integrado na web. Para realizar essas tarefas, foram empregadas diversas bibliotecas computacionais, incluindo scikit-learn, pandas, numPy, seaborn e matplotlib, para tratamento dos dados, execução da modelagem e visualização dos resultados.

Os dados coletados da literatura e explicitados na Tabela 1 foram lidos pelo programa através da biblioteca pandas e um objeto do tipo DataFrame foi criado. Para além das informações apresentadas na Tabela 1, também foram computados os resultados referentes à viscosidade cinemática [ $\text{mm}^2/\text{s}$ ], o calor de combustão [ $\text{MJ}/\text{kg}$ ] e a densidade [ $\text{g}/\text{cm}^3$ ].

Para garantir um bom desempenho, os modelos modernos de aprendizado de máquina geralmente requerem grandes quantidades de dados de qualidade. Enquanto isso, os processos de coleta e anotação de dados geralmente são realizados manualmente e/ou consomem muito tempo e recursos (o que é o caso da produção dos diferentes tipos de biodiesel). A qualidade e representatividade dos dados utilizado para uma determinada tarefa são geralmente ditadas pela disponibilidade natural de dados limpos no domínio específico, bem como pelo nível de expertise dos desenvolvedores envolvidos. Atualmente, a ampliação de dados é a maneira mais eficaz de amenizar esse problema. O principal objetivo da ampliação de dados é aumentar o volume, qualidade e diversidade dos dados de treinamento (MUMUNI; MUMUNI, 2022).

Métodos de interpolação têm sido amplamente utilizados para auxiliar na geração de características adicionais para aumentar os dados no espaço de características (LI et al., 2020; MUMUNI; MUMUNI, 2022; OH; HAN; JEONG, 2020; SZLOBODNYIK; FARKAS, 2021). Os métodos geram características adicionais com base nas relações entre as características vizinhas existentes. Isso, portanto, resulta em informações extras que preenchem possíveis lacunas nas representações. Nesse sentido, neste trabalho aplicou-se o método de interpolação no banco de dados coletado da literatura. Através de testes

preliminares, foi delimitado que o fator de aumento de 4 vezes seria suficiente para o treinamento dos modelos de aprendizado de máquina de interesse. A biblioteca utilizada foi a numpy, função `interp`<sup>3</sup>. Sendo assim, após esta etapa do pré-processamento dos dados, o banco de dados figurou com 136 amostras (eram 34 inicialmente).

#### 4.3. Análise básica dos dados

Inicialmente, procedeu-se com a visualização dos dados para traçar uma análise comparativa entre os dados originais e o conjunto aumentado. Para isto, foram construídos histogramas, gráficos de pares e também foram avaliados os coeficientes de Pearson para ambos os conjuntos. Estas tarefas foram executadas através das seguintes funções: `matplotlib.hist`<sup>4</sup>, `seaborn.pairplot`<sup>5</sup> e `pandas.corr()`<sup>6</sup>, respectivamente.

O coeficiente de correlação de Pearson, representado pela letra ( $r$ ), é uma medida estatística que expressa o grau de correlação linear entre duas variáveis quantitativas. Seu valor varia entre -1 e 1, onde 1 indica uma correlação positiva perfeita, -1 indica uma correlação negativa perfeita, e 0 indica que não há correlação linear. Tecnicamente, o coeficiente é calculado como a covariância das duas variáveis dividida pelo produto dos seus desvios padrão. É importante notar que o coeficiente de Pearson só mede relações lineares e pode ser afetado por *outliers*. Além disso, uma correlação alta não implica causalidade (MYATT; JOHNSON, 2014). Matematicamente, é expresso pela Equação (1).

---

<sup>3</sup> <https://numpy.org/doc/stable/reference/generated/numpy.interp.html>

<sup>4</sup> [https://matplotlib.org/stable/api/\\_as\\_gen/matplotlib.pyplot.hist.html](https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.hist.html)

<sup>5</sup> <https://seaborn.pydata.org/generated/seaborn.pairplot.html>

<sup>6</sup> <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.html>

$$r = \frac{\sum(xi - \bar{x})(yi - \bar{y})}{\sqrt{\sum(xi - \bar{x})^2 \sum(yi - \bar{y})^2}} \quad (1)$$

Na qual:

$x_i$  e  $y_i$  são valores das variáveis;

$\bar{x}$  e  $\bar{y}$  são as médias dos valores ( $x$ ) e ( $y$ )

Por fim, para comparar o banco de dados original e o aumentado, foi utilizada a função `pandas.describe`<sup>7</sup> para cálculo das métricas estatísticas básicas de ambos os bancos de dados.

#### 4.4. Algoritmos de Aprendizado de Máquina

Para esta dissertação, o algoritmo selecionado foi o dos k-vizinhos mais próximos (KNN, *K-Nearest Neighbors*). Este algoritmo foi escolhido por motivos diversos, dentre os quais, destacam-se (BELYADI; HAGHIGHAT, 2021; ÇAKIR et al., 2023; MOBARAK et al., 2023; ZOHURI; MOSSAVAR-RAHMANI; BEHGOUNIA, 2022):

- Há facilidade de interpretação → Simples de entender e implementar, já que faz previsões baseadas na média dos k vizinhos mais próximos.
- É um modelo flexível → Funciona bem para dados não lineares e não assume nenhuma distribuição subjacente dos dados.
- É um algoritmo robustos a dados com ruídos → O impacto do ruído pode ser minimizado ajustando o valor de  $k$ .

---

<sup>7</sup> <https://pandas.pydata.org/pandas/docs/stable/reference/api/pandas.DataFrame.describe.html>

- Adaptabilidade a Diferentes Tipos de Dados → Pode ser aplicado a dados categóricos e contínuos sem necessidade de muitas transformações.
- Aplicabilidade a Pequenos e Médios Conjuntos de Dados → É mais adequado para conjuntos de dados de pequeno a médio porte devido à sua simplicidade computacional.
- Capacidade de Capturar Interações Complexas → Embora simples, pode capturar interações complexas ao considerar a proximidade no espaço multidimensional.

O KNN para regressões é implementado pela biblioteca scikit-learn<sup>8</sup>. Nesta biblioteca os parâmetros são:

- *n\_neighbors*: Número de vizinhos a serem usados por padrão para consultas de k-vizinhos; deve ser um número inteiro.
- *weights*: Função de peso usada na previsão. Valores possíveis: ‘uniform’: pesos uniformes. Todos os pontos em cada vizinhança são ponderados igualmente.
- ‘distance’: pondera os pontos pelo inverso de sua distância. Neste caso, os vizinhos mais próximos de um ponto de consulta terão uma influência maior do que os vizinhos que estão mais distantes.
- *metric*: Métrica utilizada para computação da distância dos vizinhos. Neste trabalho, avaliamos a possibilidade de métricas tipo cosseno (‘cosine’), euclidiana (‘euclidean’) e ‘manhattan’.

#### 4.5. Avaliação do Modelo

As métricas de avaliação são essenciais para medir o desempenho dos modelos de aprendizado de máquina e podem variar dependendo do tipo de problema (classificação, regressão etc.). Algumas métricas comuns para regressão são o coeficiente de determinação ( $R^2$ ), o Erro Quadrático Médio (MSE - *Mean Squared Error*) e o Erro Absoluto Médio (MAE - *Mean Absolute Error*), os quais foram avaliados nesta dissertação para mensuração do sucesso dos modelos de aprendizado de máquina e são calculados com base nas Equações

---

<sup>8</sup> <https://scikit-learn.org/stable/modules/neighbors.html>

(2-5). Nestas equações,  $n_{sample}$  indica o número de dados (amostras) avaliadas,  $\hat{y}_i$  os valores preditos e  $y_i$  os valores reais da variável alvo.  $\bar{y}_i$  indica o valor médio de  $y_i$ .

$$MAE(y, \hat{y}) = \frac{1}{n_{sample}} \sum_{i=0}^{n_{samples}-1} |y_i - \hat{y}_i| \quad (2)$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (3)$$

Na qual:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (4)$$

$$MSE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}} (y_i - \hat{y}_i)^2 \quad (5)$$

*Overfitting*, também conhecido como sobreajuste, é quando um modelo de aprendizado de máquina se adapta demais aos dados de treinamento, capturando o ruído ou flutuações aleatórias presentes nesses dados em vez de aprender as tendências gerais. Como resultado, o modelo pode ter um desempenho excelente nos dados de treinamento, mas falha em generalizar para novos dados não vistos, levando a previsões imprecisas. Isso geralmente ocorre quando o modelo é muito complexo em relação à quantidade e variedade dos dados de treinamento. Para evitar o *overfitting*, é crucial utilizar técnicas como validação e garantir que o modelo seja treinado com um conjunto de dados representativo e suficientemente grande. *Underfitting*, ou subajuste, por outro lado, é quando um modelo de aprendizado de

máquina é muito simples para capturar a complexidade dos dados de treinamento (FACELI et al., 2011). Para obter ajustes suficientes, foi selecionado 80% do banco de dados para treinamento dos modelos e 20% para teste. Durante o treinamento, dividiu-se o banco de dados em 4 grupos para trabalhar-se a validação do modelo. A função objetivo de busca dos melhores parâmetros maximizou a média dos coeficientes de determinação dos 4 grupos.

A busca do ponto de máximo foi realizada utilizando-se o pacote computacional Optuna (AKIBA et al., 2019). O modelo foi treinado 100 vezes a partir de diferentes sementes (*seeds*) para escolha aleatória dos conjuntos de treino e teste. Os resultados apresentados referem-se à média das simulações.

#### *4.6. Avaliação da composição ótima*

Após a delimitação dos parâmetros ótimos, avaliou-se a partir do modelo otimizado quais são as composições requeríveis de misturas de ésteres para melhor atender às normas técnicas vigentes da ASTM, EN e ANP. A função ‘predict’ permite fazer o cálculo das propriedades de interesse tendo-se valores de N de misturas fictícios. Ou seja, de posse do modelo, o problema foi invertido para encontrar o(s) perfil(s) de ácidos graxos final e ideal.

Modelos de caixa preta apresentam relações indistintas e complexas entre as entradas e saídas, que são frequentemente difíceis de interpretar e, para dizer o mínimo, são computacionalmente intensivas para resolver. Por essa mesma razão, métodos de otimização meta-heurísticos são preferidos em relação à otimização matemática clássica para resolver tais modelos de caixa preta, pois eles geram uma série de ótimos locais próximos ao ótimo global (desde que o ótimo global não seja encontrado); os quais podem então ser considerados como a solução ótima (SUVARNA et al., 2022).

O método de evolução diferencial (DE) consiste em uma técnica de otimização baseada em população. Ele funciona iterativamente, mantendo uma população de soluções candidatas que evoluem ao longo do tempo. Em cada iteração, novas soluções são geradas combinando soluções existentes (por meio de operações como mutação e cruzamento) e, em seguida, uma seleção é realizada para determinar quais soluções sobreviverão para a próxima geração (ALI; KAJEE-BAGDADI, 2009). O objetivo é explorar o espaço de busca de forma eficiente

para encontrar o ótimo global ou uma aproximação próxima dele. Como das características em estudo neste trabalho, duas têm faixa de trabalho possíveis (densidade e viscosidade) e a outra pode ser a máxima possível (que é o calor de combustão com um mínimo de 35 MJ/kg), procedeu-se da seguinte maneira: buscou-se a combinação de mistura ésteres de ácidos graxos que rendesse o maior valor de calor de combustão possível e, na sequência, procedeu-se com a conferência do cálculo de densidade e de viscosidade do fluido.

Para procurar o valor máximo a ser predito de calor de combustão, utilizou-se o pacote computacional `scipy.optimize.differential_evolution`<sup>9</sup>. Os parâmetros empregados para a busca do ótimo foram os padrões da biblioteca, não sendo necessária nenhuma alteração. Somente foi informado ao algoritmo que a soma das frações de ácidos somaria em 100. Foram realizadas 10 corridas para busca do ponto de ótimo, uma vez que se trata de um método de busca estocástico.

---

9

[https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.differential\\_evolution.html](https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.differential_evolution.html)

## 5. RESULTADOS E DISCUSSÃO

### 5.1. Análise básica dos dados

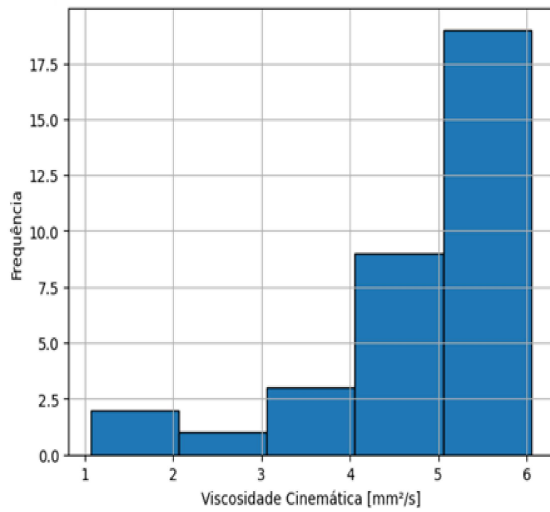
Na Figura 1 são apresentados os histogramas referentes à propriedade viscosidade cinemática para o banco de dados original e o aumentado, bem como também à porcentagem do ácido graxo C8. A partir da Figura 1 se pode verificar que o formato da distribuição não se alterou durante a expansão do banco de dados, o que é um bom indicativo, pois mostra que as características intrínsecas se mantiveram. Demais histogramas dos bancos de dados foram omitidas desta dissertação, pois apresentam conclusões semelhantes.

Nas Figuras 2 e 3 são apresentados os gráficos de pares para todas as variáveis estudadas para o banco original e aumentado. As conclusões que podem ser obtidas a partir da visualização deste gráfico são similares às obtidas para a Figura 1. Através destes gráficos, notamos que não há padrão simples e claro sobre a correspondência entre as variáveis estudadas, por exemplo, linearidade. Isto mostra que este banco de dados é um candidato interessante para a modelagem a partir de aprendizado de máquina.

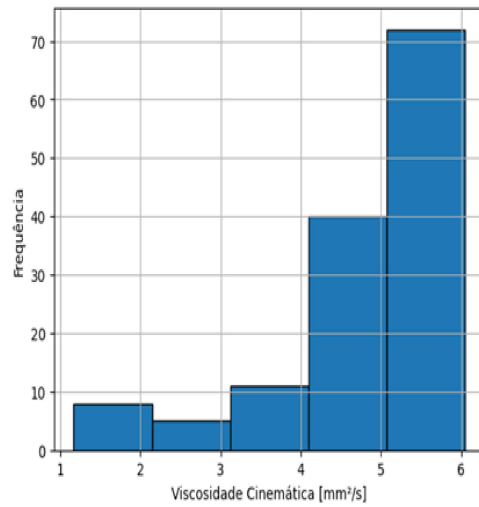
Visando avaliar o grau de correspondência linear entre as variáveis, foram calculados os coeficientes de Pearson entre as variáveis do banco de dados. Estes resultados são apresentados através de mapas de calor. Um mapa de calor, em inglês *heatmap*, é uma representação visual de dados que utiliza cores para destacar padrões e variações em um conjunto de dados. Na Figura 4(a) são mostrados os resultados para o banco de dados original e na 4(b) para o aumentado. É possível verificar que o coeficiente não sofre variação aparente com o aumento do número de dados, o que é um bom indicativo para validação da técnica. Além disso, também se pode notar que não há correlação linear forte para as variáveis em estudo.



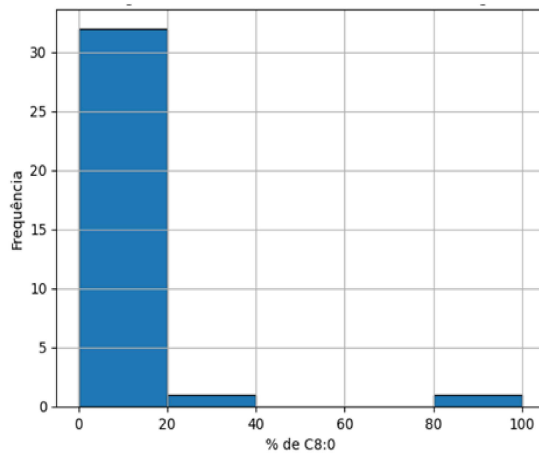
Figura 1. Histogramas representando a propriedade viscosidade cinemática do banco de dados original (a) e aumentado (b); porcentagem de C8 no banco de dados original (c) e aumentado (d).



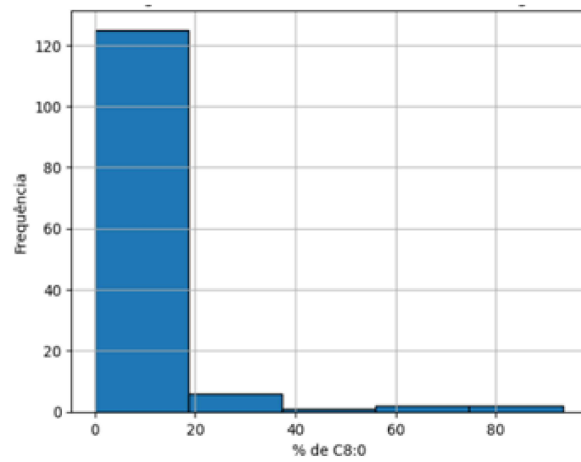
(a)



(b)



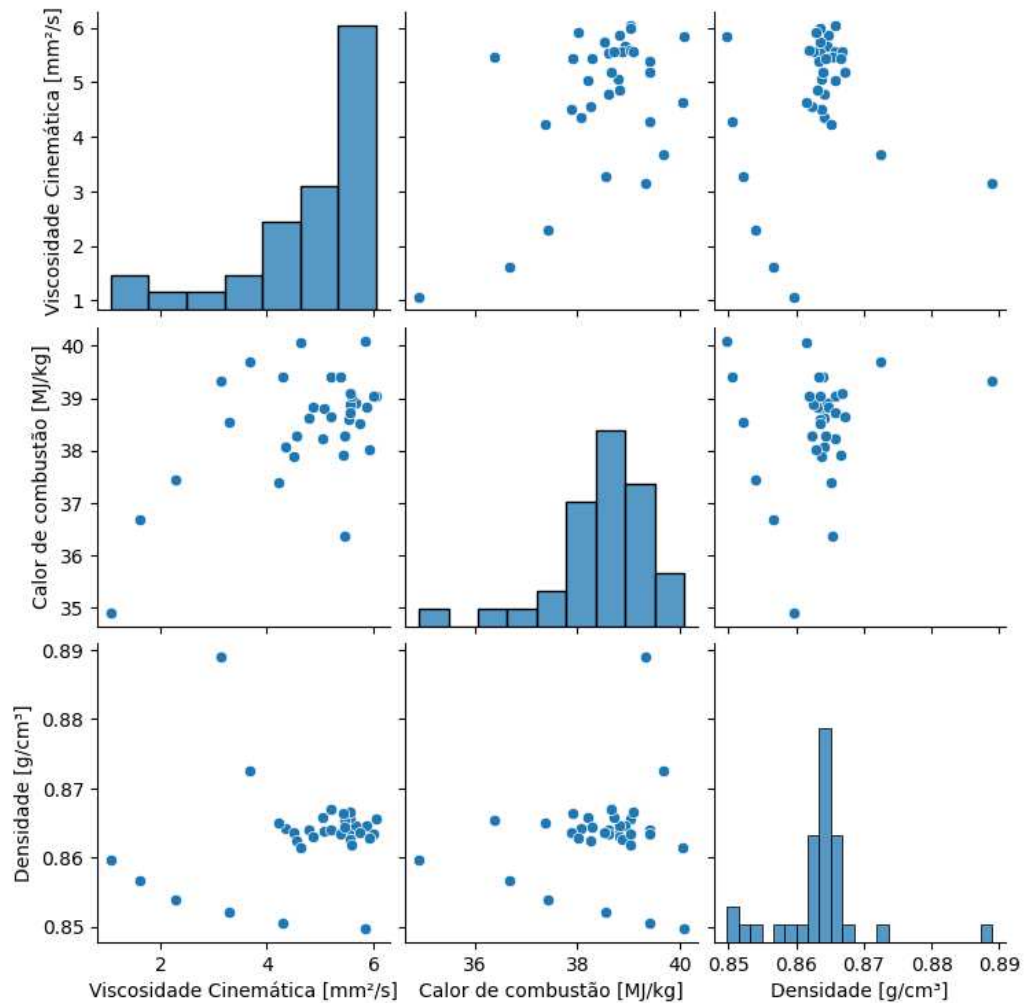
(c)



(d)

Fonte: Elaborada pelo autor.

Figura 2. Gráfico de pares para viscosidade cinemática, calor de combustão e densidade, respectivamente - banco de dados original.

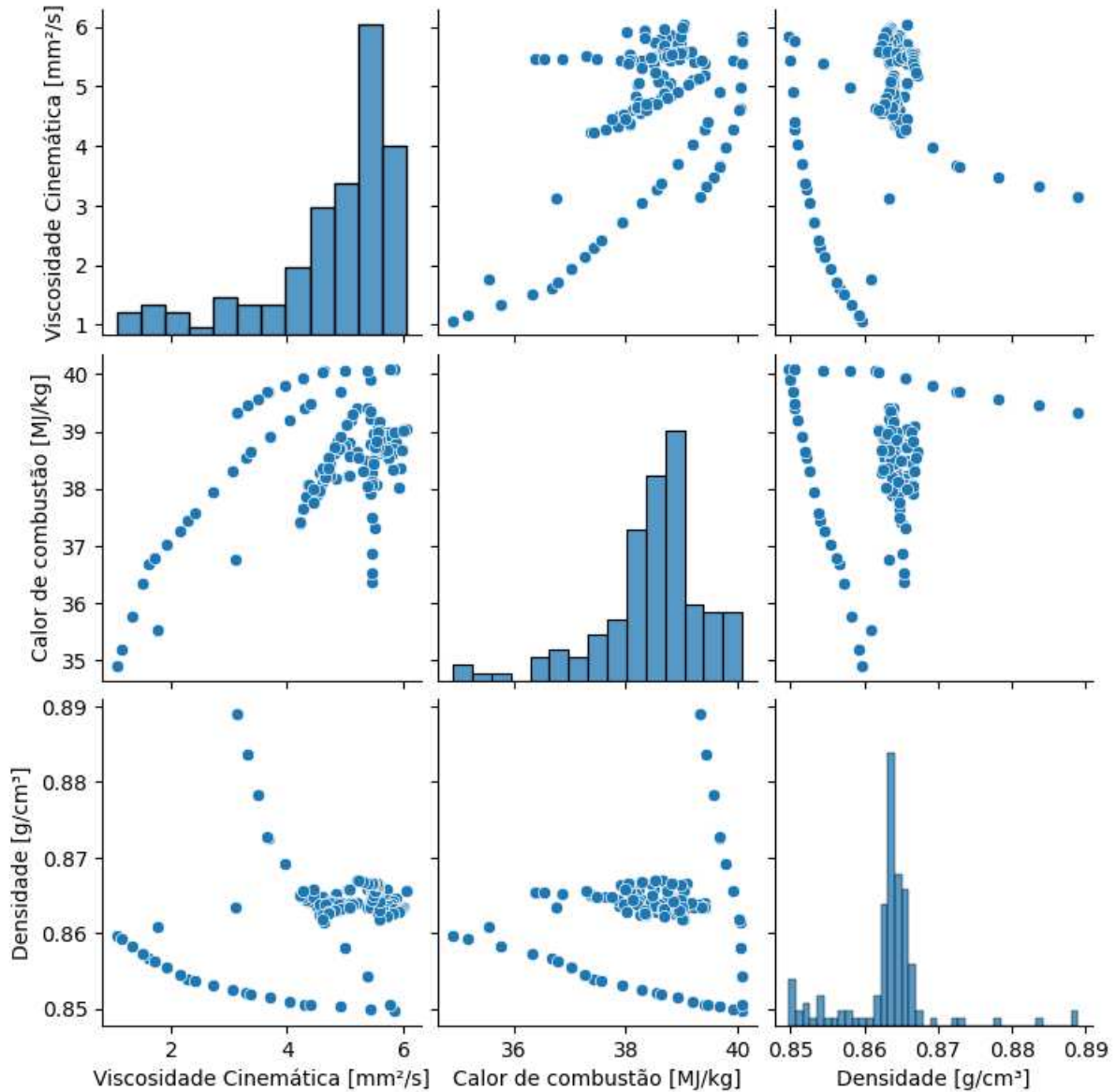


Fonte: Elaborada pelo autor.

Na Tabela 2 são apresentadas as métricas estatísticas básicas dos bancos de dados original e aumentado, os quais foram obtidos através da biblioteca pandas. describe. Se pode observar que as métricas são bastante parecidas. Se dois bancos de dados apresentam a mesma média e o mesmo desvio padrão, é possível afirmar que, estatisticamente falando,

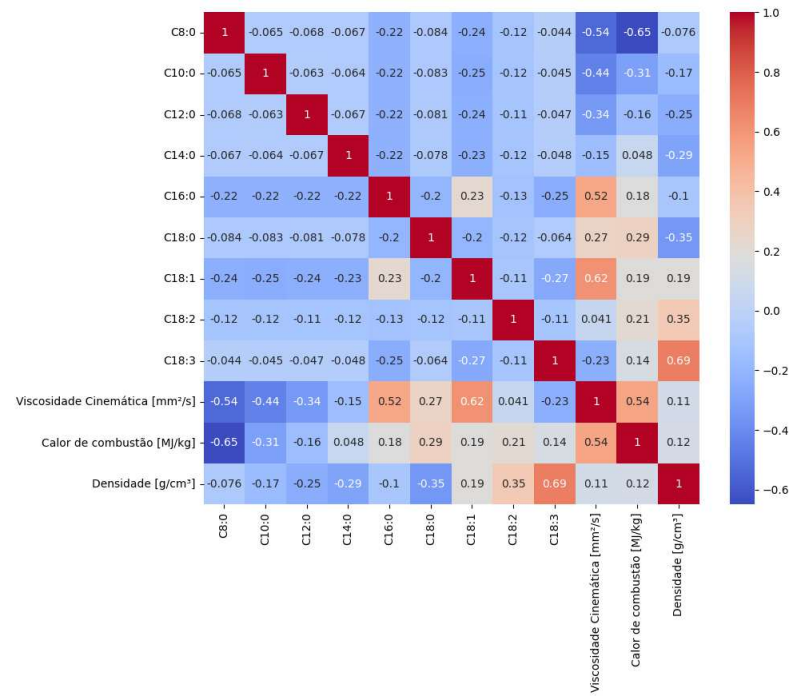
eles têm distribuições muito semelhantes em termos de dispersão dos dados em torno da média. É importante ressaltar que isso não significa necessariamente que os dados em si sejam idênticos (WALPOLE et al., 2006).

Figura 3. Gráfico de pares para viscosidade cinemática, calor de combustão e densidade, respectivamente - banco de dados interpolado.

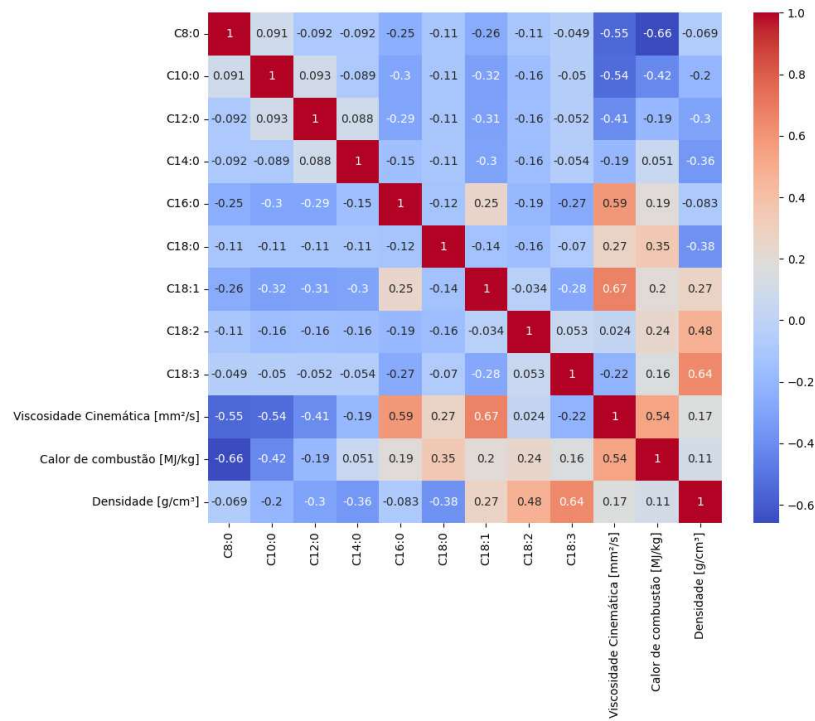


Fonte: Elaborada pelo autor.

Figura 4. Mapa de calor para o coeficiente de Pearson para frações de ésteres, viscosidade cinemática, calor de combustão e densidade, respectivamente – (a) Banco de dados original; (b) Banco de dados interpolado.



(a)



(b)

Fonte: Elaborada pelo autor.

Tabela 2. Grandes estatísticas básicas do banco de dados original (34 dados) e do banco aumentado (136 dados).

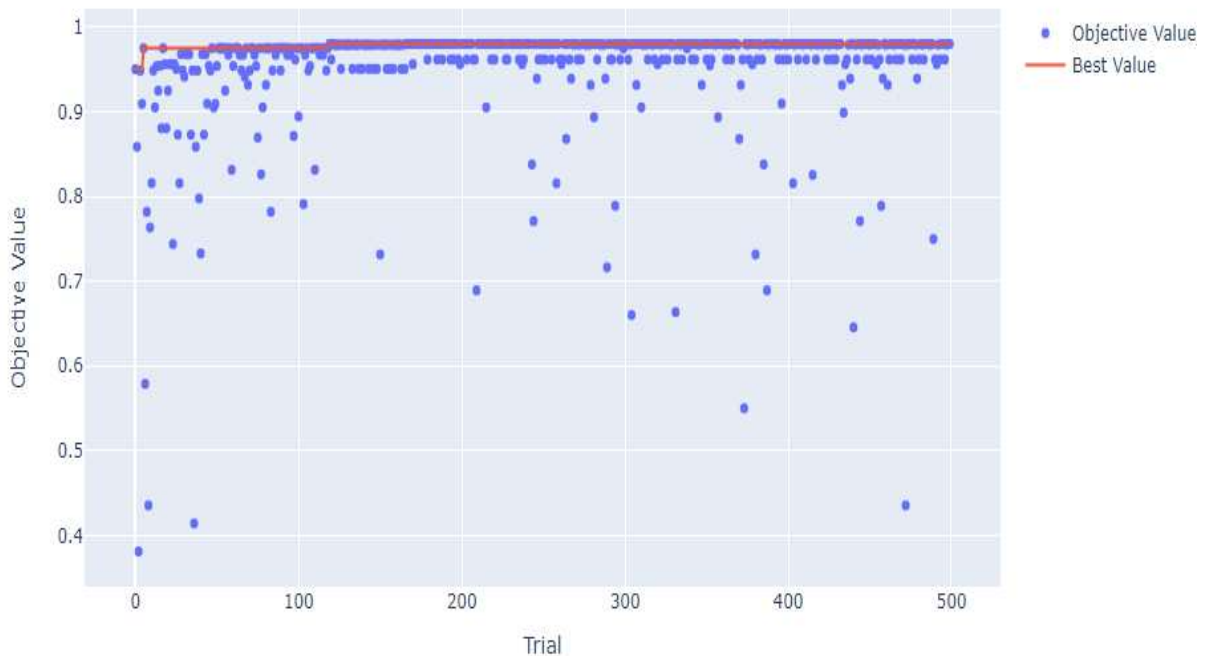
<b>Banco de dados original</b>							
	<b>Média</b>	<b>Desvio padrão</b>	<b>Valor Mínimo</b>	<b>1° Quartil (25%)</b>	<b>Mediana (50%)</b>	<b>3° Quartil (75%)</b>	<b>Valor Máximo</b>
<b>C8:0</b>	4,33	17,59	0,00	0,00	0,00	0,00	100,00
<b>C10:0</b>	4,49	17,60	0,00	0,00	0,00	0,17	100,00
<b>C12:0</b>	4,64	17,43	0,00	0,00	0,43	0,77	100,00
<b>C14:0</b>	4,72	17,33	0,00	0,30	0,76	1,02	100,00
<b>C16:0</b>	28,90	20,51	0,00	29,78	32,02	34,46	100,00
<b>C18:0</b>	6,24	17,34	0,00	0,86	2,65	3,11	100,00
<b>C18:1</b>	33,13	21,94	0,00	34,30	37,78	41,75	100,00
<b>C18:2</b>	10,58	17,11	0,00	7,07	7,91	9,14	100,00
<b>C18:3</b>	2,95	17,15	0,00	0,00	0,00	0,00	100,00
<b>Viscosidade Cinemática [mm<sup>2</sup>/s]</b>	4,78	1,24	1,07	4,40	5,20	5,57	6,05
<b>Calor de combustão [MJ/kg]</b>	38,50	1,04	34,91	38,10	38,69	39,05	40,09
<b>Densidade [g/cm<sup>3</sup>]</b>	0,86	0,01	0,85	0,86	0,86	0,87	0,89
<b>Banco de dados aumentado</b>							
<b>C8:0</b>	4,38	14,99	0,00	0,00	0,00	0,00	100,00
<b>C10:0</b>	4,55	15,10	0,00	0,00	0,00	0,30	100,00
<b>C12:0</b>	4,71	14,85	0,00	0,00	0,43	0,87	100,00
<b>C14:0</b>	4,78	14,77	0,00	0,36	0,79	1,07	100,00
<b>C16:0</b>	29,06	18,10	0,00	29,97	32,09	34,78	100,00
<b>C18:0</b>	6,31	14,86	0,00	1,15	2,65	3,20	100,00
<b>C18:1</b>	33,26	19,65	0,00	34,49	37,81	41,43	100,00
<b>C18:2</b>	10,66	13,67	0,00	7,12	8,11	9,27	100,00
<b>C18:3</b>	2,24	13,62	0,00	0,00	0,00	0,00	100,00
<b>Viscosidade Cinemática [mm<sup>2</sup>/s]</b>	4,78	1,17	1,07	4,44	5,17	5,57	6,05
<b>Calor de combustão [MJ/kg]</b>	38,49	0,95	34,91	38,11	38,67	39,00	40,09
<b>Densidade [g/cm<sup>3</sup>]</b>	0,86	0,06	0,85	0,86	0,86	0,87	0,89

Fonte: Elaborada pelo autor.

## 5.2. Resultados para o modelo de aprendizado de máquina

Primeiramente, será apresentado um gráfico referente ao processo de otimização dos parâmetros dos modelos. A Figura 5 é referente a uma das otimizações do modelo KNN para a previsão da densidade das misturas e mostra a evolução do valor da função objetivo com o decorrer das 500 iterações da busca pelo ponto de máximo. Nota-se pela Figura 5 que 500 é um número suficiente para encontrar este valor. Resultados similares foram obtidos para demais modelagens realizadas nessa dissertação e foram omitidos da seção de resultados.

Figura 5. Evolução do valor da função objetivo para os 500 passos de busca para uma simulação do modelo KNN para a estimativa do valor da densidade do conjunto de treino.



Fonte: Elaborada pelo autor.

Após a otimização dos parâmetros do modelo KNN, podemos observar os parâmetros ótimos apresentados na Tabela 3. Na Tabela 3 também são apresentadas as métricas do ajuste para os conjuntos de treino e teste. O número ótimo de vizinhos encontrado foi de 2 a 3. A média da distância destes dois vizinhos mais próximos precisa ser ponderada pelo inverso de sua distância, conforme o parâmetro *weights* igual a “distance”. Ou seja, os vizinhos mais

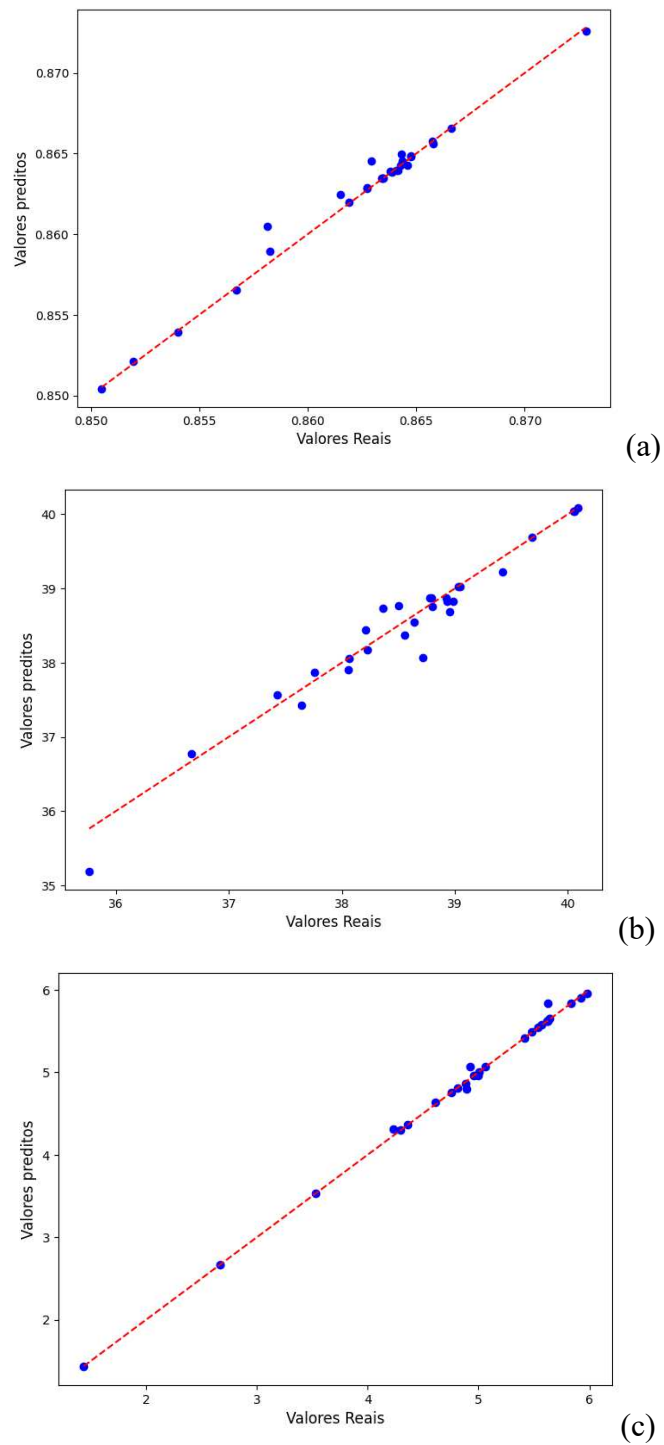
próximos de um ponto de consulta terão uma influência maior do que os vizinhos que estão mais distantes. Para todos os ajustes, foi observado um pequeno sobreajuste, uma vez que há diferença entre as métricas de treino e teste. Contudo, ambas as métricas de ajuste de treino e teste são satisfatoriamente altas. Na Figura 6 é mostrada uma representação dos valores reais e dos preditos pelos modelos. Observa-se boa conformidade, pois a proximidade dos pontos à reta bissetriz implica que os erros de previsão são pequenos e não tendem a um desvio sistemático. Além disso, há ausência de viés pois os pontos estão bem distribuídos em torno da reta. O modelo não tende a superestimar ou subestimar sistematicamente os valores reais. A boa distribuição em torno da linha bissetriz também sugere que o modelo é robusto e possui variância controlada. Isso significa que o modelo é capaz de generalizar bem para dados novos, mantendo a precisão das previsões.

Tabela 3. Parâmetros ótimos para o modelo KNN e qualidade do ajuste obtidos nas etapas de treinamento e teste.

Parâmetro		Alvo: Viscosidade Cinemática [mm <sup>2</sup> /s]	Alvo: Calor de combustão [MJ/kg]	Alvo: Densidade [g/cm <sup>3</sup> ]
<i>n_neighbors</i>		2	2	3
<i>weights</i>		<i>distance</i>	<i>distance</i>	<i>distance</i>
<i>metric</i>		<i>manhattan</i>	<i>manhattan</i>	<i>manhattan</i>
Métrica	Etapas			
R <sup>2</sup>	Treino	0,9748	0,8790	0,9799
	Teste	0,9970	0,9115	0,9822
MSE	Treino	0,1838	0,2674	0,0008
	Teste	0,0540	0,3569	0,0006
MAE	Treino	0,1118	0,1696	0,0004
	Teste	0,0241	0,1802	0,0003

Fonte: Elaborada pelo autor

Figura 6. Valores reais *versus* valores preditos para (a) densidade, (b) calor de combustão e (c) viscosidade cinemática do conjunto de teste utilizando o modelo KNN.



Fonte: Elaborada pelo autor.



A redução média na impureza é uma métrica usada para avaliar a importância de cada atributo (ou característica) em um modelo de aprendizado de máquina utilizando a biblioteca scikit-learn (sklearn). Esta métrica quantifica quanto a inclusão de um atributo específico reduz a impureza em média no modelo. Para o caso de regressões, a impureza avaliada é erro quadrático médio (*mean squared error* - MSE). Na Figura 7 é mostrada a lista de importâncias das características obtida para o problema de regressão da (a) densidade, (b) calor de combustão e (c) da viscosidade e representa a contribuição relativa de cada característica para a decisão tomada pelo modelo. Por padrão, a soma de todas as importâncias deve ser igual a 1 (ou 100%). No caso da densidade, a última característica, ou seja, o percentual de C18:3, é a mais importante, contribuindo com aproximadamente 45,99% da importância total. Isso significa que esta característica tem o maior impacto nas previsões do modelo. A sexta característica (percentual de C18:0) é a segunda mais importante, contribuindo com aproximadamente 17,58% da importância total. Salta aos olhos o fato de que C18:3 é o único éster derivado de um ácido poli-saturado (o mesmo é o que apresenta 3 saturações em sua cadeia). Fator também interessante é que os dois ésteres mais importantes são os de maior massa molecular.

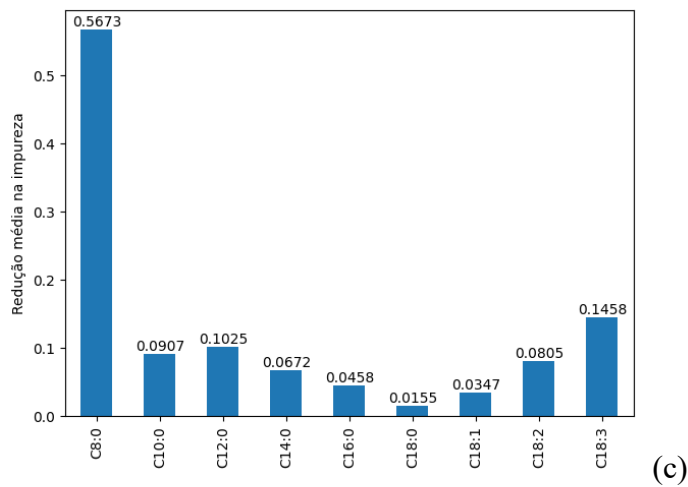
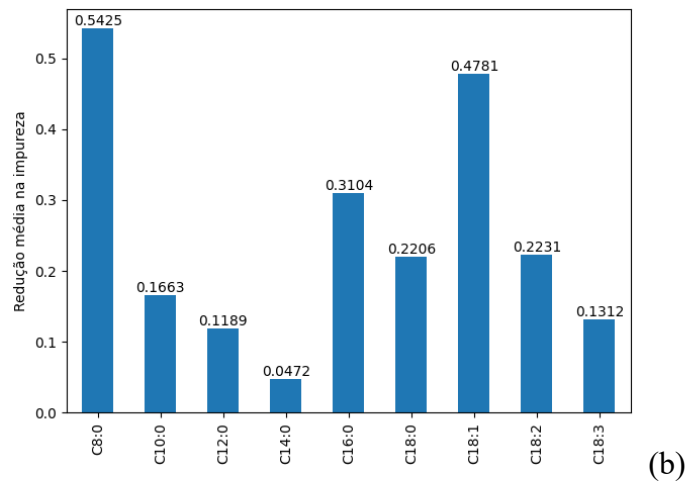
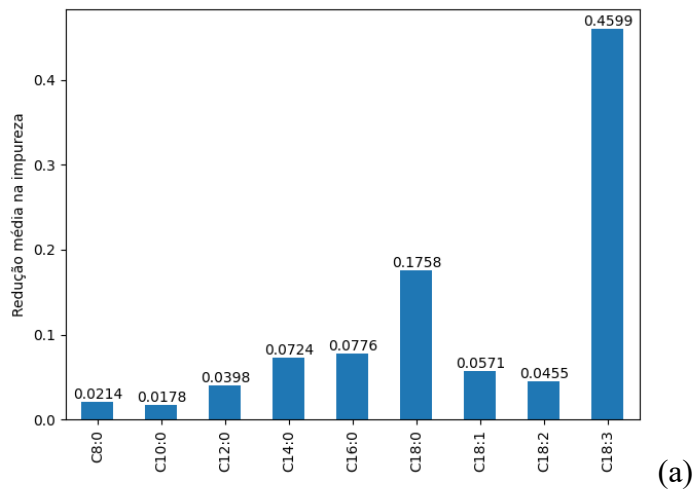
Na Figura 8 é apresentado o gráfico de dependência parcial das variáveis para o ajuste da densidade considerando o modelo KNN e o conjunto de teste. Com o aumento da fração de C18:2, notamos que a densidade aumenta. Quimicamente falando, ésteres saturados contêm apenas ligações simples entre os átomos de carbono. Isso permite que as moléculas se empilhem de maneira mais compacta, resultando em uma maior densidade. Por outro lado, ésteres saturados contêm uma ou mais ligações duplas entre os átomos de carbono, criando "dobras" ou "curvas" na cadeia. Essas ligações duplas impedem que as moléculas se alinhem tão compactamente, reduzindo a densidade (SAJJADI; RAMAN; ARANDIYAN, 2016; WEDLER; TRUSLER, 2023). Contudo, estes efeitos isolados não são visualizados para a maioria dos ésteres uma vez que nenhum deles compõe majoritariamente a mistura. C18:1 e C18:2 apresentam grande percentual de variação nos dados e a densidade aumenta com o aumento de sua proporção. Logo, os efeitos expressivos observados dizem respeito ao aumento da massa molar do sistema (quanto maior a massa, maior a densidade considerando

que as moléculas estão sujeitas à mesma temperatura e pressão) do que à presença de saturações na cadeia.

Na Figura 7(b) é apresentada a análise de sensibilidade do modelo (escala de importâncias) para a previsão do calor de combustão. A variável que mais afeta o modelo é a concentração de C8:0. O C8:0 possui uma cadeia de carbono relativamente curta, com apenas 8 átomos de carbono. Em geral, ésteres com cadeias mais curtas têm menores calores de combustão por unidade de massa em comparação com ésteres de cadeias mais longas. O C8:0 é um éster saturado, sem duplas ligações na cadeia de carbono. Os ésteres saturados geralmente possuem maior calor de combustão por unidade de massa em comparação com seus equivalentes insaturados devido à ausência de insaturações, que reduzem a energia total disponível. Tais observações nos levam a considerar que devido à sua cadeia curta, o C8:0 tem um efeito de redução do calor de combustão quando presente em altas concentrações. Isso porque, apesar de ser saturado, sua curta cadeia limita a quantidade total de energia que pode ser liberada durante a combustão. A Figura 9 mostra o gráfico de dependência parcial para o ajuste do calor de combustão considerando o modelo KNN e o conjunto de teste. A análise da Figura 9 nos confirma a expectativa de queda do calor de combustão com o aumento da fração de C8:0, efeito também observado para demais ésteres saturados.

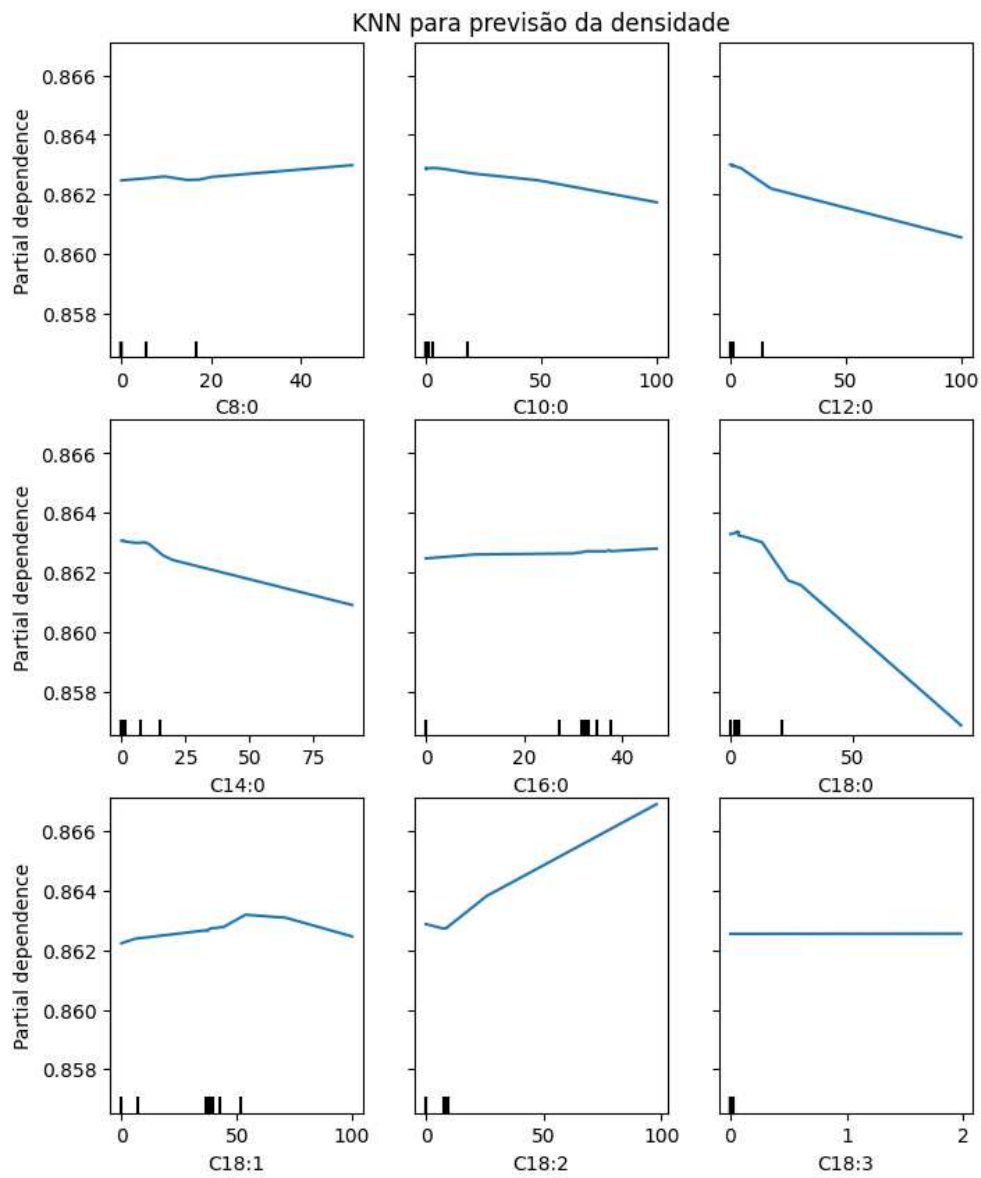
Na Figura 7(c) é mostrada a análise de sensibilidade do modelo (escala de importâncias) para a previsão da viscosidade do biodiesel. Assim como para o calor de combustão, a variável que mais afeta o modelo é a concentração de C8:0. A presença de cadeias curtas, como a do C8:0, reduz a viscosidade do biodiesel, uma vez que as interações intermoleculares (forças de van der Waals) são menores em cadeias curtas, resultando em menor resistência ao fluxo. Além disso, a ausência de insaturações no C8:0 contribui ainda mais para a redução da viscosidade, pois as moléculas lineares e saturadas se movem mais facilmente umas sobre as outras. Na Figura 10 está ilustrado o gráfico de dependência parcial para o ajuste da viscosidade considerando o modelo KNN e o conjunto de teste. Com base na análise da Figura 10 se pode observar a tendência de se diminuir a viscosidade da mistura com o aumento do teor de C8:0, bem como também pela presença de outros ésteres saturados de cadeia curta.

Figura 7. Importâncias das características para previsão da (a) densidade, (b) calor de combustão e (c) viscosidade cinemática do conjunto de teste utilizando o modelo KNN.



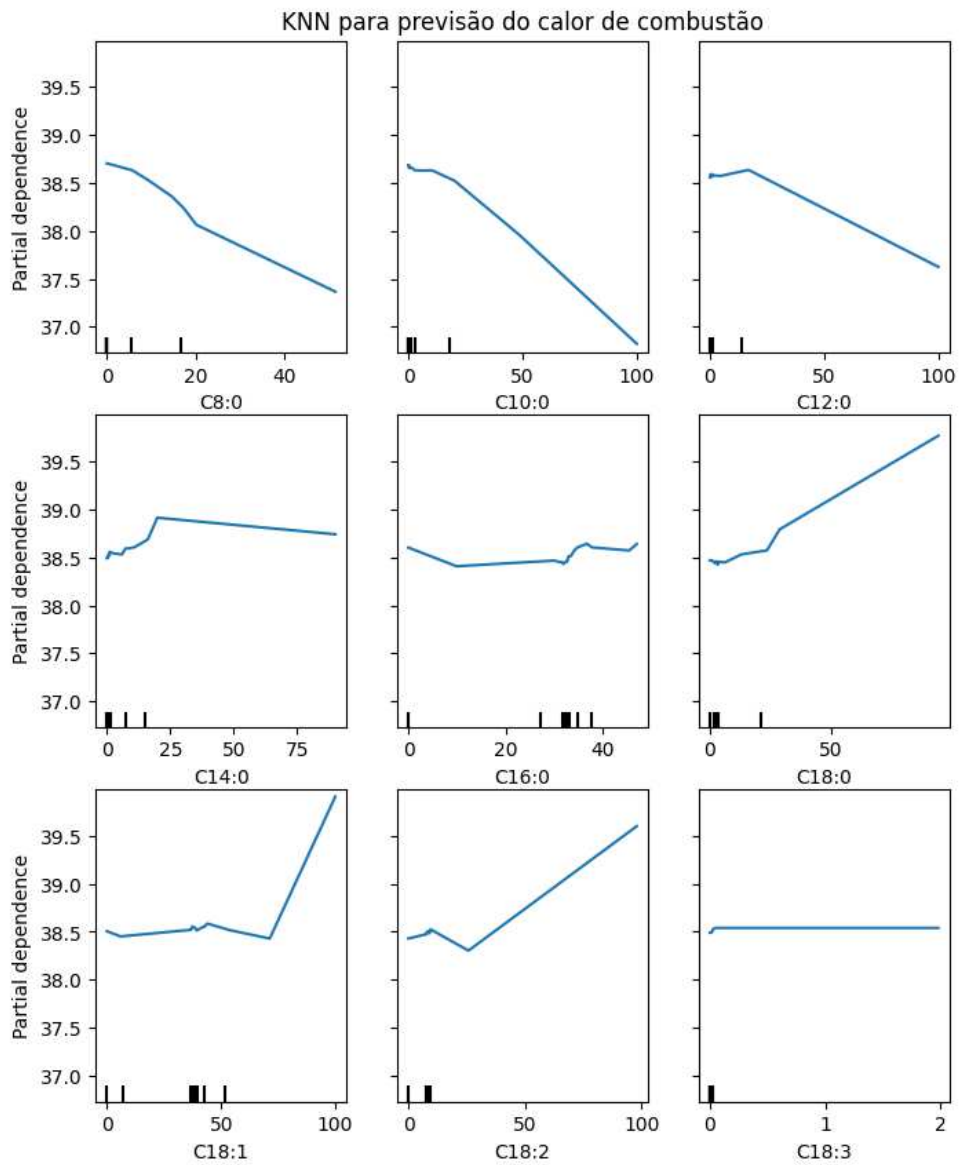
Fonte: Elaborada pelo autor

Figura 8. Gráficos de dependência parcial das características avaliadas no modelo KNN para o ajuste de densidade do conjunto de teste.



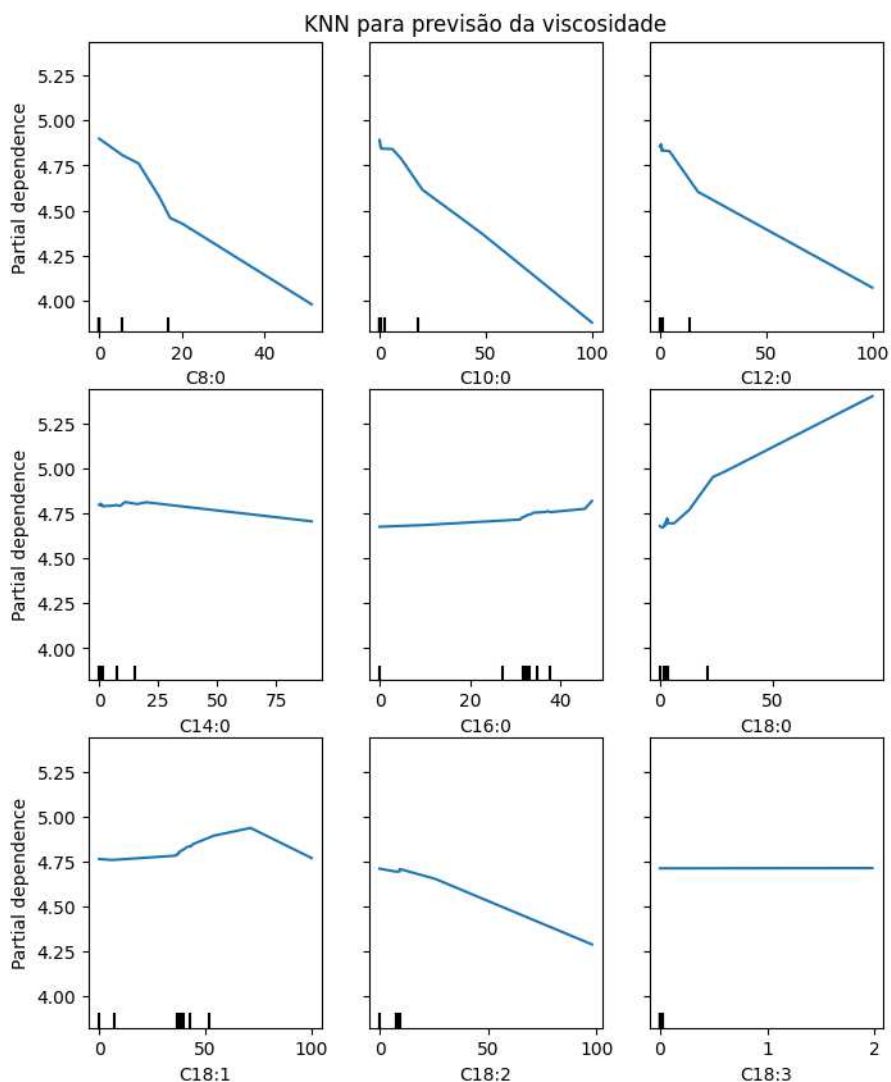
Fonte: Elaborada pelo autor

Figura 9. Gráficos de dependência parcial das características avaliados no modelo KNN para o ajuste do calor de combustão do conjunto de teste.



Fonte: Elaborada pelo autor

Figura 10. Gráficos de dependência parcial das características avaliados no modelo KNN para o ajuste da viscosidade do conjunto de teste.



Fonte: Elaborada pelo autor

No que tange à otimização da composição do biodiesel, observou-se que: as composições das misturas ricas em ésteres de cadeia longa insaturada e um éster de cadeia curta saturada e outro de cadeia média podem atender à busca dos maiores valores de calor de combustão possíveis. Os resultados das 10 corridas de otimização pelo algoritmo de evolução diferencial são apresentados na Tabela 4. Todas as composições estimadas têm valores de densidade e viscosidade que atendem a normas europeia e da ANP.

Tabela 4. Resultados obtidos para a composição de ésteres visando maximizar o calor de combustão da mistura. Valores calculados de densidade e viscosidade para cada uma das composições obtidas pelo algoritmo de evolução diferencial.

Corrida	C8:0	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2	C18:3	Calor de combustão otimizado [MJ/kg]	Densidade calculada [g/cm <sup>3</sup> ]	Viscosidade calculada [mm <sup>2</sup> /s]
1	12,9555	0,9458	0,9991	25,6413	0,7378	1,4300	0,0467	29,4214	27,8223	39,57	0,8770	3,3998
2	3,5574	7,0782	0,6333	31,5945	1,8008	11,9437	6,2713	3,2480	33,8727	39,51	0,8835	3,3897
3	16,8694	5,2240	22,1806	4,4984	0,5721	20,0163	3,7558	0,3929	26,4902	38,31	0,8601	4,5775
4	7,0889	2,2862	3,2684	1,2013	2,3379	77,9172	0,0028	5,2953	0,6017	40,11	0,8509	5,6519
5	2,0680	17,6679	4,3672	2,4550	1,3331	33,5512	3,0427	4,8051	30,7093	40,08	0,8719	4,4425
6	1,4592	0,8854	4,9687	36,4165	0,2992	8,4240	9,8336	33,9987	3,7143	39,86	0,8673	4,1262
7	6,8569	10,9981	3,2537	14,4218	2,9070	0,3801	15,7335	34,3755	11,0729	39,86	0,8667	4,1261
8	21,3037	16,2268	4,6371	6,3617	5,1618	1,4545	2,9166	8,4762	33,4612	39,51	0,8752	3,3997
9	5,1030	12,2081	4,3779	2,0482	0,2377	39,5917	5,2020	27,4045	3,8264	40,08	0,8516	5,4067
10	0,6760	29,8125	6,9642	12,7713	0,8197	36,3410	0,9910	0,6345	10,9893	40,08	0,8515	5,4078

Fonte: Elaborada pelo autor.

Sakthivel et al. (2018) realizaram ampla revisão de literatura a respeito das propriedades físico-químicas de biodieseis de terceira geração. Dentre os biodieseis listados, o calor de combustão pode variar de 20,8 (para biodiesel produzido a partir o óleo da pirólise da semente de Neem) até 45,63 MJ/kg (para biodiesel de alga *Spirulina Patensis*). Por outro lado, o biodiesel feito da alga tem viscosidade 12,4 mm<sup>2</sup>/s a 40 °C, o que está fora de especificação. Um biodiesel com características similares às apresentadas na Tabela 4 é o feito a partir de óleo de peixe (BEHÇET et al., 2015), o qual apresentou calor de combustão de 40,54 MJ/kg, 4,45 mm<sup>2</sup>/s de viscosidade e 0,881 g/cm<sup>3</sup> de densidade. Este biodiesel apresentou 31% de ésteres metílicos de ácidos graxos saturados e 54% de insaturados, proporção próxima à corrida 1 estimada neste trabalho.

Apesar de mostrar-se uma metodologia consistente para previsão de propriedades ótimas de um biodiesel, o uso do KNN tem uma limitação, conforme mostram os resultados. As previsões realizadas pelo modelo são baseadas nas médias dos vizinhos utilizados no treinamento. Ou seja, o algoritmo não consegue realizar extrapolações. Logo, os maiores valores de calor de combustão margeiam o máximo do próprio conjunto de treinamento que é o de 40 MJ/kg. Nesse sentido, recomenda-se para trabalhos futuros a avaliação de algoritmos que tenham estas capacidades, como, por exemplo, as árvores de decisão. Embora as árvores de decisão por si só geralmente não sejam boas em extrapolações, conjuntos de árvores como *Random Forests* ou *Gradient Boosting* podem, em alguns casos, capturar padrões complexos e fazer previsões razoáveis fora do intervalo de treinamento.



## **6. CONCLUSÃO E CONSIDERAÇÕES FINAIS**

A engenharia reversa aplicada ao biodiesel, por meio do desenvolvimento de modelos preditivos baseados na composição de ácidos graxos, demonstrou ser uma abordagem promissora para a otimização das propriedades do biocombustível. A capacidade de prever as características do biodiesel com base em suas composições permite identificar as misturas ideais que atendam aos requisitos de qualidade e desempenho, contribuindo para a produção de biocombustíveis mais eficientes e sustentáveis. Neste trabalho, foram identificadas 10 combinações de proporções de ácidos graxos que garantem, matematicamente, propriedades físico-químicas condizentes com a normativa atual no que tange à densidade, ao calor de combustão e à viscosidade cinemática.

## 7. REFERENCIAS

ABOOALI, D.; SOLEIMANI, R.; GHOLAMREZA-RAVI, S. **Characterization of physico-chemical properties of biodiesel components using smart data mining approaches.** *Fuel*, [s.l.], v. 266, p. 117075, 2020. ISSN: 00162361, DOI: 10.1016/j.fuel.2020.117075.

AFDC. **ASTM Biodiesel Specifications.** [s.d.]. Disponível em: <[https://afdc.energy.gov/fuels/biodiesel\\_specifications.html](https://afdc.energy.gov/fuels/biodiesel_specifications.html)>. Acesso em: 23/abr./23.

AGÊNCIA NACIONAL DO PETRÓLEO, G. N. E. B. **RESOLUÇÃO ANP Nº 920.** 2023. Disponível em: <<https://atosoficiais.com.br/anp/resolucao-n-920-2023-estabelece-a-especificacao-do-biodiesel-e-as-obrigacoes-quanto-ao-controle-da-qualidade-a-serem-atendidas-pelos-agentes-economicos-que-comercializem-o-produto-em-territorio-nacional?origin=instituicao>>. Acesso em: 03/abr./24.

AGHBASHLO, M. et al. **Machine learning technology in biodiesel research: A review.** *Progress in Energy and Combustion Science*, [s.l.], v. 85, p. 100904, 2021. ISSN: 03601285, DOI: 10.1016/j.pecs.2021.100904.

AKIBA, T. et al. **Optuna.** Em: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* New York, NY, USA: ACM, 2019. ISBN: 9781450362016, DOI: 10.1145/3292500.3330701.

AKRAM, F. et al. **Current trends in biodiesel production technologies and future progressions: A possible displacement of the petro-diesel.** *Journal of Cleaner Production*, [s.l.], v. 370, p. 133479, 2022. ISSN: 09596526, DOI: 10.1016/j.jclepro.2022.133479.

ALI, M. M.; KAJEE-BAGDADI, Z. **A local exploration-based differential evolution algorithm for constrained global optimization.** *Applied Mathematics and Computation*, [s.l.], v. 208, n° 1, p. 31–48, 2009. ISSN: 00963003, DOI: 10.1016/j.amc.2008.11.036.

ALICKE, A. A. et al. **Guidelines for the rheological characterization of biodiesel.** *Fuel*, [s.l.], v. 140, p. 446–452, 2015. ISSN: 00162361, DOI: 10.1016/j.fuel.2014.09.075.

AMINIAN, A.; ZARENEZHAD, B. **Accurate predicting the viscosity of biodiesels and blends using soft computing models.** *Renewable Energy*, [s.l.], v. 120, p. 488–500, 2018. ISSN: 09601481, DOI: 10.1016/j.renene.2017.12.038.

ATABANI, A. E. et al. **A comprehensive review on biodiesel as an alternative energy resource and its characteristics.** *Renewable and Sustainable Energy Reviews*, [s.l.], v. 16, n° 4, p. 2070–2093, 2012. ISSN: 13640321, DOI: 10.1016/j.rser.2012.01.003.

\_\_\_\_\_. **Non-edible vegetable oils: A critical evaluation of oil extraction, fatty acid compositions, biodiesel production, characteristics, engine performance and emissions production.** *Renewable and Sustainable Energy Reviews*, [s.l.], v. 18, p. 211–245, 2013. ISSN: 13640321, DOI: 10.1016/j.rser.2012.10.013.

BALABIN, R. M.; LOMAKINA, E. I.; SAFIEVA, R. Z. **Neural network (ANN) approach to biodiesel analysis: Analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy.** *Fuel*, [s.l.], v. 90, n° 5, p. 2007–2015, 2011. ISSN: 00162361, DOI: 10.1016/j.fuel.2010.11.038.

BARRADAS FILHO, A. O. et al. **Application of artificial neural networks to predict viscosity, iodine value and induction period of biodiesel focused on the study of oxidative stability.** *Fuel*, [s.l.], v. 145, p. 127–135, 2015. ISSN: 00162361, DOI: 10.1016/j.fuel.2014.12.016.

BEHÇET, R. et al. **Comparison of exhaust emissions of biodiesel–diesel fuel blends produced from animal fats.** *Renewable and Sustainable Energy Reviews*, [s.l.], v. 46, p. 157–165, 2015. ISSN: 13640321, DOI: 10.1016/j.rser.2015.02.015.

BELYADI, H.; HAGHIGHAT, A. **Supervised learning.** *Machine Learning Guide for Oil and Gas Using Python*. [s.l.]: Elsevier, 2021. p. 169–295. DOI: 10.1016/B978-0-12-821929-4.00004-4.

BHARTI, R. K. et al. **Evaluation of fuel properties for possible biodiesel output based on the fatty acid composition of oleaginous plants and microalgae.** *Science of The Total Environment*, [s.l.], v. 918, p. 170448, 2024. ISSN: 00489697, DOI: 10.1016/j.scitotenv.2024.170448.

BRAHMA, S. et al. **Biodiesel production from mixed oils: A sustainable approach towards industrial biofuel production.** *Chemical Engineering Journal Advances*, [s.l.], v. 10, p. 100284, 2022. ISSN: 26668211, DOI: 10.1016/j.ceja.2022.100284.

BUKKARAPU, K. R.; KRISHNASAMY, A. **A study on the effects of compositional variations of biodiesel fuel on its physiochemical properties.** *Biofuels*, [s.l.], v. 12, n° 5, p. 523–535, 2021. ISSN: 1759-7269, DOI: 10.1080/17597269.2018.1501638.

\_\_\_\_\_. **A critical review on available models to predict engine fuel properties of biodiesel.** *Renewable and Sustainable Energy Reviews*, [s.l.], v. 155, p. 111925, 2022. ISSN: 13640321, DOI: 10.1016/j.rser.2021.111925.

ÇAKIR, M. et al. **Accuracy assessment of RFerns, NB, SVM, and kNN machine learning classifiers in aquaculture.** *Journal of King Saud University - Science*, [s.l.], v. 35, n° 6, p. 102754, 2023. ISSN: 10183647, DOI: 10.1016/j.jksus.2023.102754.

CHEN, G. Q.; WU, X. F. **Energy overview for globalized world economy: Source, supply chain and sink.** *Renewable and Sustainable Energy Reviews*, [s.l.], v. 69, p. 735–749, 2017. ISSN: 13640321, DOI: 10.1016/j.rser.2016.11.151.

CORRAL BOBADILLA, M. et al. **Optimizing Biodiesel Production from Waste Cooking Oil Using Genetic Algorithm-Based Support Vector Machines.** *Energies*, [s.l.], v. 11, n° 11, p. 2995, 2018. ISSN: 1996-1073, DOI: 10.3390/en11112995.

DAS, M. et al. **Study on viscosity and surface tension properties of biodiesel-diesel blends and their effects on spray parameters for CI engines.** *Fuel*, [s.l.], v. 220, p. 769–779, 2018. ISSN: 00162361, DOI: 10.1016/j.fuel.2018.02.021.

DEMIRBAS, A. **Progress and recent trends in biodiesel fuels.** *Energy Conversion and Management*, [s.l.], v. 50, n° 1, p. 14–34, 2009. ISSN: 01968904, DOI: 10.1016/j.enconman.2008.09.001.

ELGHARBAWY, A. S. et al. **A REVIEW ON BIODIESEL FEEDSTOCKS AND PRODUCTION TECHNOLOGIES.** *Journal of the Chilean Chemical Society*, [s.l.], v. 66, nº 1, p. 5098–5109, 2021. ISSN: 0717-9707, DOI: 10.4067/S0717-97072021000105098.

EUROPEAN STANDARDS. **BS EN 14214:2012+A2:2019.** [s.d.]. Disponível em: <[https://www.en-standard.eu/bs-en-14214-2012-a2-2019-liquid-petroleum-products-fatty-acid-methyl-esters-fame-for-use-in-diesel-engines-and-heating-applications-requirements-and-test-methods/?gclid=CjwKCAjw0ZiiBhBKEiwA4PT9zxIHI5W4HFnbIRapkcj9-Z24RZ582ziTtxKgnubOCMmHPfEXv4W8mhoChQYQAvD\\_BwE](https://www.en-standard.eu/bs-en-14214-2012-a2-2019-liquid-petroleum-products-fatty-acid-methyl-esters-fame-for-use-in-diesel-engines-and-heating-applications-requirements-and-test-methods/?gclid=CjwKCAjw0ZiiBhBKEiwA4PT9zxIHI5W4HFnbIRapkcj9-Z24RZ582ziTtxKgnubOCMmHPfEXv4W8mhoChQYQAvD_BwE)>. Acesso em: 23/abr./23.

FACELI, K. et al. **Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina.** 1th ed. São Paulo: LTC, 2011. 1–394 p. ISBN: 8521618808.

FERREIRA, A. G. M. et al. **REVIEW: Models for predicting viscosities of biodiesel fuels over extended ranges of temperature and pressure.** *Fuel*, [s.l.], v. 287, p. 119544, 2021. ISSN: 00162361, DOI: 10.1016/j.fuel.2020.119544.

GANESHMOORTHY, V.; MUTHUKANNAN, M.; THIRUGNANASAMBANDAM, M. **Comparison of Linear Regression and ANN of Fish Oil Biodiesel Properties Prediction.** *Journal of Physics: Conference Series*, [s.l.], v. 1276, nº 1, p. 012074, 2019. ISSN: 1742-6588, DOI: 10.1088/1742-6596/1276/1/012074.

GERON, A. **Mãos à obra: aprendizado de máquina com Scikit-Learn & TensorFlow.** 1ª ed. São Paulo: Alta Books, 2019. 576 p.

GIWA, S. O. et al. **Prediction of selected biodiesel fuel properties using artificial neural network.** *Frontiers in Energy*, [s.l.], v. 9, n° 4, p. 433–445, 2015. ISSN: 2095-1701, DOI: 10.1007/s11708-015-0383-5.

GUI, M. M.; LEE, K. T.; BHATIA, S. **Feasibility of edible oil vs. non-edible oil vs. waste edible oil as biodiesel feedstock.** *Energy*, [s.l.], v. 33, n° 11, p. 1646–1653, 2008. ISSN: 03605442, DOI: 10.1016/j.energy.2008.06.002.

GÜLÜM, M.; BILGIN, A. **Density, flash point and heating value variations of corn oil biodiesel–diesel fuel blends.** *Fuel Processing Technology*, [s.l.], v. 134, p. 456–464, 2015. ISSN: 03783820, DOI: 10.1016/j.fuproc.2015.02.026.

IEA. **Transport biofuels.** <https://www.iea.org/reports/renewables-2020/transport-biofuels>. 2020. Disponível em: <<https://www.iea.org/reports/renewables-2020/transport-biofuels>>. Acesso em: 23/abr./23.

ISSARIYAKUL, T.; DALAI, A. K. **Biodiesel from vegetable oils.** *Renewable and Sustainable Energy Reviews*, [s.l.], v. 31, p. 446–471, 2014. ISSN: 13640321, DOI: 10.1016/j.rser.2013.11.001.

KARMAKAR, A.; KARMAKAR, S.; MUKHERJEE, S. **Properties of various plants and animals feedstocks for biodiesel production.** *Bioresource Technology*, [s.l.], v. 101, n° 19, p. 7201–7210, 2010. ISSN: 09608524, DOI: 10.1016/j.biortech.2010.04.079.

KNOTHE, G. **Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters.** *Fuel Processing Technology*, [s.l.], v. 86, n° 10, p. 1059–1070, 2005. ISBN: 0378-3820, ISSN: 03783820, DOI: 10.1016/j.fuproc.2004.11.002.

\_\_\_\_\_. **Analyzing biodiesel: standards and other methods.** *Journal of the American Oil Chemists' Society*, [s.l.], v. 83, n° 10, p. 823–833, 2006. ISSN: 0003-021X, DOI: 10.1007/s11746-006-5033-y.

KRISHNASAMY, A.; BUKKARAPU, K. R. **A comprehensive review of biodiesel property prediction models for combustion modeling studies.** *Fuel*, [s.l.], v. 302, p. 121085, 2021. ISSN: 00162361, DOI: 10.1016/j.fuel.2021.121085.

LI, Z. et al. **Random linear interpolation data augmentation for person re-identification.** *Multimedia Tools and Applications*, [s.l.], v. 79, n° 7–8, p. 4931–4947, 2020. ISSN: 1380-7501, DOI: 10.1007/s11042-018-7071-5.

LISBOA, P. et al. **Economic analysis of a plant for biodiesel production from waste cooking oil via enzymatic transesterification using supercritical carbon dioxide.** *The Journal of Supercritical Fluids*, [s.l.], v. 85, p. 31–40, 2014. ISSN: 08968446, DOI: 10.1016/j.supflu.2013.10.018.

MAHESHWARI, P. et al. **A review on latest trends in cleaner biodiesel production: Role of feedstock, production methods, and catalysts.** *Journal of Cleaner Production*, [s.l.], v. 355, p. 131588, 2022. ISSN: 09596526, DOI: 10.1016/j.jclepro.2022.131588.



MEJÍA, J. D.; SALGADO, N.; ORREGO, C. E. **Effect of blends of Diesel and Palm-Castor biodiesels on viscosity, cloud point and flash point.** *Industrial Crops and Products*, [s.l.], v. 43, n° 1, p. 791–797, 2013. ISSN: 09266690, DOI: 10.1016/j.indcrop.2012.08.026.

MENG, X.; JIA, M.; WANG, T. **Neural network prediction of biodiesel kinematic viscosity at 313K.** *Fuel*, [s.l.], v. 121, p. 133–140, 2014. ISSN: 00162361, DOI: 10.1016/j.fuel.2013.12.029.

MOBARAK, M. H. et al. **Scope of machine learning in materials research—A review.** *Applied Surface Science Advances*, [s.l.], v. 18, p. 100523, 2023. ISSN: 26665239, DOI: 10.1016/j.apsadv.2023.100523.

MUMUNI, A.; MUMUNI, F. **Data augmentation: A comprehensive survey of modern approaches.** *Array*, [s.l.], v. 16, p. 100258, 2022. ISSN: 25900056, DOI: 10.1016/j.array.2022.100258.

MYATT, G. J.; JOHNSON, W. P. **Making Sense of Data I: A Practical Guide to Exploratory Data Analysis and Data Mining.** 2nd ed. Hoboken: Wiley, 2014. 1–248 p.

OGUNKUNLE, O.; AHMED, N. A. **A review of global current scenario of biodiesel adoption and combustion in vehicular diesel engines.** *Energy Reports*, [s.l.], v. 5, p. 1560–1579, 2019. ISSN: 23524847, DOI: 10.1016/j.egy.2019.10.028.

OH, C.; HAN, S.; JEONG, J. **Time-Series Data Augmentation based on Interpolation.** *Procedia Computer Science*, [s.l.], v. 175, p. 64–71, 2020. ISSN: 18770509, DOI: 10.1016/j.procs.2020.07.012.

PYDIMALLA, M. et al. **Sustainable biodiesel: A comprehensive review on feedstock, production methods, applications, challenges and opportunities.** *Materials Today: Proceedings*, [s.l.], 2023. ISSN: 22147853, DOI: 10.1016/j.matpr.2023.03.593.

RAMOS, M. J. et al. **Influence of fatty acid composition of raw materials on biodiesel properties.** *Bioresource Technology*, [s.l.], v. 100, n° 1, p. 261–268, 2009. ISSN: 09608524, DOI: 10.1016/j.biortech.2008.06.039.

REFAAT, A. A. **Correlation between the chemical structure of biodiesel and its physical properties.** *International Journal of Environmental Science & Technology*, [s.l.], v. 6, n° 4, p. 677–694, 2009. ISSN: 1735-1472, DOI: 10.1007/BF03326109.

ROCABRUNO-VALDÉS, C. I.; RAMÍREZ-VERDUZCO, L. F.; HERNÁNDEZ, J. A. **Artificial neural network models to predict density, dynamic viscosity, and cetane number of biodiesel.** *Fuel*, [s.l.], v. 147, p. 9–17, 2015. ISSN: 00162361, DOI: 10.1016/j.fuel.2015.01.024.

SADRAMELI, S. M.; SEAMES, W.; MANN, M. **Prediction of higher heating values for saturated fatty acids from their physical properties.** *Fuel*, [s.l.], v. 87, n° 10–11, p. 1776–1780, 2008. ISSN: 00162361, DOI: 10.1016/j.fuel.2007.10.020.

SAJJADI, B.; RAMAN, A. A. A.; ARANDIYAN, H. **A comprehensive review on properties of edible and non-edible vegetable oil-based biodiesel: Composition, specifications and prediction models.** *Renewable and Sustainable Energy Reviews*, [s.l.], v. 63, p. 62–92, 2016. ISSN: 13640321, DOI: 10.1016/j.rser.2016.05.035.

SAKTHIVEL, R. et al. **A review on the properties, performance and emission aspects of the third generation biodiesels.** *Renewable and Sustainable Energy Reviews*, [s.l.], v. 82, p. 2970–2992, 2018. ISSN: 13640321, DOI: 10.1016/j.rser.2017.10.037.

SAXENA, P.; JAWALE, S.; JOSHIPURA, M. H. **A Review on Prediction of Properties of Biodiesel and Blends of Biodiesel.** *Procedia Engineering*, [s.l.], v. 51, p. 395–402, 2013. ISSN: 18777058, DOI: 10.1016/j.proeng.2013.01.055.

SEAF ELNASR, T. A. et al. **Sustainable biodiesel production from waste olive oil: Utilizing olive pulp-derived catalysts for environmental and economic benefits.** *Sustainable Chemistry and Pharmacy*, [s.l.], v. 37, p. 101426, 2024. ISSN: 23525541, DOI: 10.1016/j.scp.2024.101426.

SINGH, N. et al. **Progress and facts on biodiesel generations, production methods, influencing factors, and reactors: A comprehensive review from 2000 to 2023.** *Energy Conversion and Management*, [s.l.], v. 302, p. 118157, 2024. ISSN: 01968904, DOI: 10.1016/j.enconman.2024.118157.

SONG, Y. et al. **Study on spray characteristics of biodiesel alternative fuels for in-cylinder environment of diesel engine.** *Journal of the Energy Institute*, [s.l.], v. 113, p. 101507, 2024. ISSN: 17439671, DOI: 10.1016/j.joei.2023.101507.

SUVARNA, M. et al. **Predicting biodiesel properties and its optimal fatty acid profile via explainable machine learning.** *Renewable Energy*, [s.l.], v. 189, p. 245–258, 2022. ISSN: 09601481, DOI: 10.1016/j.renene.2022.02.124.

SZLOBODNYIK, G.; FARKAS, L. **Data Augmentation by Guided Deep Interpolation.** *Applied Soft Computing*, [s.l.], v. 111, p. 107680, 2021. ISSN: 15684946, DOI: 10.1016/j.asoc.2021.107680.

VENKATESAN, S. P. et al. **Study of emission characteristics of a diesel engine run by fuel blends of diesel, jatropha biodiesel and cetane improver.** *Materials Today: Proceedings*, [s.l.], 2023. ISSN: 22147853, DOI: 10.1016/j.matpr.2023.02.404.

VIGNESH, P. et al. **A review of conventional and renewable biodiesel production.** *Chinese Journal of Chemical Engineering*, [s.l.], v. 40, p. 1–17, 2021. ISSN: 10049541, DOI: 10.1016/j.cjche.2020.10.025.

WALPOLE, R. E. et al. **Probability and Statistics for Engineers and Scientists.** 8th ed. Upper Saddle River: Prentice Hall, 2006. 1–816 p.

WEDLER, C.; TRUSLER, J. P. M. **Review of density and viscosity data of pure fatty acid methyl ester, ethyl ester and butyl ester.** *Fuel*, [s.l.], v. 339, p. 127466, 2023. ISSN: 00162361, DOI: 10.1016/j.fuel.2023.127466.

YAŞAR, F. **Comparison of fuel properties of biodiesel fuels produced from different oils to determine the most suitable feedstock type.** *Fuel*, [s.l.], v. 264, p. 116817, 2020. ISSN: 00162361, DOI: 10.1016/j.fuel.2019.116817.

YU, W.; ZHAO, F. **Prediction of critical properties of biodiesel fuels from FAMES compositions using intelligent genetic algorithm-based back propagation neural**

**network.** *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, [s.l.], v. 43, n° 17, p. 2063–2076, 2021a. ISSN: 1556-7036, DOI: 10.1080/15567036.2019.1641575.

\_\_\_\_\_. **Prediction of critical properties of biodiesel fuels from FAMEs compositions using intelligent genetic algorithm-based back propagation neural network.** *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, [s.l.], v. 43, n° 17, p. 2063–2076, 2021b. ISSN: 1556-7036, DOI: 10.1080/15567036.2019.1641575.

YUAN, W.; HANSEN, A. C.; ZHANG, Q. **Predicting the temperature dependent viscosity of biodiesel fuels.** *Fuel*, [s.l.], v. 88, n° 6, p. 1120–1126, 2009. ISSN: 00162361, DOI: 10.1016/j.fuel.2008.11.011.

ZOHURI, B.; MOSSAVAR-RAHMANI, F.; BEHGOUNIA, F. **Artificial intelligence, machine learning, and deep learning use cases.** *Knowledge is Power in Four Dimensions: Models to Forecast Future Paradigm*. [s.l.]: Elsevier, 2022. p. 889–947. DOI: 10.1016/B978-0-323-95112-8.00028-3.