



**UNIVERSIDADE DA INTEGRAÇÃO INTERNACIONAL DA LUSOFONIA AFRO-
BRASILEIRA**
INSTITUTO DE ENGENHARIAS E DESENVOLVIMENTO SUSTENTÁVEL
CURSO DE ENGENHARIA DE ENERGIAS

LIZANDRA RÉGIA MIRANDA DA SILVA

**CLASSIFICAÇÃO DO POTENCIAL ENERGÉTICO DE COMBUSTÍVEIS
ORIUNDOS DA BIOENERGIA COM O AUXÍLIO DE APRENDIZAGEM DE
MÁQUINA**

REDENÇÃO

2023

LIZANDRA RÉGIA MIRANDA DA SILVA

CLASSIFICAÇÃO DO POTENCIAL ENERGÉTICO DE COMBUSTÍVEIS ORIUNDOS
DA BIOENERGIA COM O AUXÍLIO DE APRENDIZAGEM DE MÁQUINA

Monografia apresentada ao curso de Engenharia de Energias do Instituto de Engenharias e Desenvolvimento Sustentável da Universidade da Integração Internacional da Lusofonia Afro-Brasileira, como requisito para obtenção do Título de Bacharel em Engenharia de Energias.

Orientadora: Profa. Dra. Ada Amélia Sanders Lopes.

Coorientador: Prof. Dr. Alisson Pessoa Guimarães.

REDENÇÃO

2023

Universidade da Integração Internacional da Lusofonia Afro-Brasileira
Sistema de Bibliotecas da UNILAB
Catalogação de Publicação na Fonte.

Silva, Lizandra Regia Miranda da.

S586c

Classificação do potencial energético de combustíveis oriundos da bioenergia com o auxílio de aprendizagem de máquina / Lizandra Regia Miranda da Silva. - Redenção, 2023.

66f: il.

Monografia - Curso de Engenharia de Energias, Instituto de Engenharias e Desenvolvimento Sustentável, Universidade da Integração Internacional da Lusofonia Afro-Brasileira, Redenção, 2023.

Orientadora: Prof.^a Dr.^a. Ada Amelia Sanders Lopes.

1. Biomassa. 2. Aprendizagem - Máquina. 3. Regressão Logística. I. Lopes, Ada Amelia Sanders. II. Título.

CE/UF/BSCA

CDD 634.9

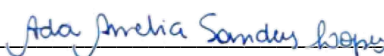
LIZANDRA RÉGIA MIRANDA DA SILVA

CLASSIFICAÇÃO DO POTENCIAL ENERGÉTICO DE COMBUSTÍVEIS
ORIUNDOS DA BIOENERGIA COM O AUXÍLIO DE APRENDIZAGEM DE
MÁQUINA

Trabalho de Conclusão de Curso
apresentado ao curso de Engenharia de
Energias do Instituto de Engenharias e
Desenvolvimento Sustentável da
Universidade da Integração
Internacional da Lusofonia Afro-
Brasileira, como requisito para obtenção
do Título de Bacharel em Engenharia
de Energias.

Aprovada em: 25/01/2023

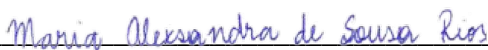
BANCA EXAMINADORA



Prof. Dra. Ada Amélia Sanders Lopes (Orientadora)
Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB)



Prof. Dr. Antonio Alisson Pessoa Guimarães (Coorientador)
Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB)



Prof. Dra. Maria Alexandra de Sousa Rios
Universidade Federal do Ceará (UFC)

A Deus.
Aos minha mãe, meus avós, meu padrinho e
minhas tias.

AGRADECIMENTOS

A minha mãe, Danielle Miranda, ao meu padrinho, Wilquer Pontes; por ter me proporcionado qualidade de ensino, pelo apoio e pela dedicação.

A minhas tias, Virlane Oliveira e Viviane Pontes e aos demais familiares; por constituírem o núcleo que torcem pelo meu desempenho e dedicação ao longo da vida.

Aos meus avós, Maria das Graças e Raimundo Fernandes; por serem meus exemplos e que, infelizmente, não podem estar presentes, mas tenho certeza que estão orgulhosos.

Aos amigos e companheiros de graduação, professores e técnicos; por contribuírem com minha trajetória acadêmica. Dedico especialmente este momento a: Fabiana Silva.

A minhas amigas do coração que foram minha base aos longo desses 4 anos, Vitória Moreno, Marisabel França e Bruna Coelho; por tantas noites em claro estudando e nos divertindo, estas amizades que levarei para a vida após a graduação.

Aos meus professores; agradeço por todo o carinho que recebi e acolhimento, por facilitarem a aprendizagem e por serem ímpares em sua ética profissional e por serem exemplos a serem seguidos. Dedico este momento aos professores: Silvia Helena, Cristiane Souza, Marcia Farias, Rejane Felix, Cleiton Santos, George Mamede, Olímpio Carneiro, Allberson Dantas e Vandilberto Pereira.

A minha querida professora orientadora, Ada Amelia Sanders Lopes; que por vezes foi um ombro amigo, me auxiliou diversas vezes e é uma das principais motivadoras pela realização deste trabalho. Um exemplo íntegro de profissional, de mulher e de mãe, de tamanha ética, responsabilidade e afetividade na qual me espelho e me serve de exemplo. Agradeço por ter me acompanhado durante esse trabalho e, principalmente, por me acompanhar durante o período de graduação e agora neste trabalho. Agradeço a todos os ensinamentos, a paciência e o compromisso. *Merci.*

Ao meu coorientador, Antonio Alisson Pessoa Guimarães; por ter me convidado para ser meu orientador durante dois anos de iniciação científica. Agradeço ao professor por toda dedicação e pela oportunidade de conhecer sua metodologia de ensino que me estimulou a permanência no curso. Ademais agradeço por ter aceitado o convite para

participar da orientação deste trabalho, por toda disponibilidade e por não medir esforços em torcer para o meu sucesso, muito obrigada.

Ao Instituto de Engenharia e Desenvolvimento Sustentável (IEDS) em parceria com o Centro Acadêmico de Engenharia de Energias, neste onde fui vice-presidente e presidente, que por meio de sua colaboração e gestão, auxiliando da melhor forma possível os estudantes do curso de Engenharia de Energias para obter suas respectivas carreiras.

Ao PIBIC-UNILAB pela oportunidade de participar como bolsista e na iniciação científica, pois a base do conhecimento adquirido durante o processo possibilitou a realização do presente trabalho. Ao meu orientador de pesquisa Prof. Dr. Alisson Guimarães pela excelente condução no projeto, por todo suporte e apoio.

Ao Grupo de Inovações Tecnológicas e Especialidades Químicas (GRINTEQUI/UFC), por ter sido o grupo responsável a adquirir a maioria dos dados coletados para a realização deste trabalho. Dedico este momento a Profa. Dra. Maria Alexandra Rios.

Agradeço à banca por ter aceitado o convite para participação da minha defesa para a obtenção do título de Engenheira de Energias.

*“São as nossas escolhas, mais do que as
nossas capacidades, que mostram quem
realmente somos.”*

- J.K. Rowling (Alvo Dumbledore, HP7)

RESUMO

A biomassa é uma fonte de energia renovável que pode ser precursora na produção de combustível e pode ser obtida a partir de matéria orgânica, como plantas e resíduos animais. Ela pode ser utilizada para gerar energia elétrica em usinas termoelétricas ou para aquecimento em caldeiras ou fornos. A vantagem da biomassa como combustível é que pode ser produzida localmente, reduzindo a dependência de combustíveis fósseis importados e ajudando a diminuir as emissões de gases de efeito estufa. Neste contexto, faz-se a análise do potencial energético de biomassas oriundas da bioenergia para obtenção de biocombustíveis, para isso utiliza-se a técnica de Análise Imediata, na qual verifica os seguintes parâmetros: teor de umidade, teor de materiais voláteis, teor de cinzas e teor de carbono fixo. Outro parâmetro utilizado para análise de potencial energético é a obtenção do poder calorífico. Entretanto, verifica-se que, para a situação, demanda-se tempo para analisar e obter dados. Assim, o presente trabalho teve como objetivo desenvolver um método para avaliar e indicar resultados que classificam a biomassa com potencial ou não à combustível. Especificamente, coletaram-se os resultados obtidos em trabalhos disponibilizados na bibliografia e construiu-se um acervo de dados, nos quais foram submetidos a um algoritmo de Aprendizagem de Máquina, ferramenta capaz de compreender a situação proposta e obter saídas desejáveis para o problema. Diante disto, escolheu-se o modelo de algoritmo de Regressão Logística e construiu-se três cenários distintos e avaliou-se a sua aplicabilidade quanto a sua fase de treinamento e sua fase de teste, por fim, comparou-se com a teoria adquirida. Desta forma, o cenário A considera três parâmetros para análise: teor de umidade, teor de materiais voláteis e teor de cinzas. Adiante, o cenário B considera os parâmetros de A com a inclusão de mais um parâmetro, o teor de carbono fixo. Por fim, a construção do cenário C considera os mesmos parâmetros do cenário B com a inclusão de mais um parâmetro, o poder calorífico. Em termos de resultado, obteve-se que o cenário C desempenhou vantagem em relação aos cenários A e B, respectivamente. Além disso, nota-se que os resultados obtidos possuem coerência ao compará-los com a bibliografia. Portanto, verifica-se que o algoritmo consegue atender a demanda proposta por ter conseguido classificar corretamente todas as amostras nas fases de treinamento e de teste e poderá ser utilizado para classificar o potencial energético a partir das características informadas sobre os dados.

Palavras-chave: Biomassa; Aprendizagem de Máquina; Regressão Logística.

ABSTRACT

Biomass is a renewable energy source that can be a precursor in the production of fuel and can be obtained from organic matter such as plants and animal waste. It can be used to generate electricity in thermoelectric power plants or for heating boilers or furnaces. The advantage of biomass as a fuel is that it can be produced locally, reducing dependence on imported fossil fuels and helping to reduce greenhouse gas emissions. In this context, an analysis of the energy potential of biomass derived from bioenergy to obtain biofuels is carried out, for which the Immediate Analysis technique is used, in which the following parameters are verified: moisture content, volatile material content, ash and fixed carbon content. Another parameter used for energy potential analysis is obtaining the calorific value. However, it appears that, for the situation, time is required to analyze and obtain data. Thus, the present work aimed to develop a method to evaluate and indicate results that classify biomass with potential or not for fuel. Specifically, the results obtained from works available in the bibliography were collected and a collection of data was built, in which they were submitted to a Machine Learning algorithm, a tool capable of understanding the proposed situation and obtaining desirable outputs for the problem. In view of this, the Logistic Regression algorithm model was chosen and three different scenarios were built and its applicability was evaluated regarding its training phase and its test phase, finally, it was compared with the acquired theory. Thus, scenario A considers three parameters for analysis: moisture content, volatile material content and ash content. Further on, scenario B considers the parameters of A with the inclusion of one more parameter, the fixed carbon content. Finally, the construction of scenario C considers the same parameters as scenario B with the inclusion of one more parameter, the calorific value. In terms of results, it was found that scenario C had an advantage over scenarios A and B, respectively. In addition, it is noted that the results obtained are coherent when compared with the bibliography. Therefore, it appears that the algorithm is able to meet the proposed demand for having managed to correctly classify all samples in the training and test phases and can be used to classify the energy potential from the characteristics informed about the data.

Keywords: Biomass; Machine Learning; Logistic Regression.

LISTA DE FIGURAS

Figura 1	– Diagrama de blocos com as etapas do algoritmo do exemplo.....	26
Figura 2	– Modelo genérica de uma matriz de confusão.....	34
Figura 3	– Etapas presentes na metodologia com as respectivas especificações.....	35
Figura 4	– Fluxograma com as etapas estabelecidas ao algoritmo.....	39
Figura 5	– Matriz de confusão sobre o resultado na fase de treinamento para 3 parâmetros.....	43
Figura 6	– Matriz de confusão sobre o resultado na fase de teste para 3 parâmetros.....	43
Figura 7	– Matriz de confusão sobre o resultado na fase de treinamento para 4 parâmetros	45
Figura 8	– Matriz de confusão sobre o resultado na fase de teste para 4 parâmetros.....	45
Figura 9	– Matriz de confusão sobre o resultado na fase de treinamento para 5 parâmetros	46
Figura 10	– Matriz de confusão sobre o resultado na fase de teste para 4 parâmetros.....	47

LISTA DE GRÁFICOS

Gráfico 1 – Convergência do Algoritmo RL para 3 parâmetros.....	42
Gráfico 2 – Convergência do Algoritmo RL para 4 parâmetros.....	44
Gráfico 3 – Convergência do Algoritmo RL para 5 parâmetros.....	46
Gráfico 4 – Comparativo entre os resultados para a fase de treinamento [em %].....	47
Gráfico 5 – Comparativo entre os resultados para a fase de teste [em %].....	48

LISTA DE TABELAS

Tabela 1 – Biomassas utilizadas para composição do acervo.....	36
Tabela 2 – Faixas de atuação inferidas para cada parâmetro em porcentagem	38

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
ASTM	<i>American Society for Testing and Materials</i>
NBR	Norma Brasileira Regulamentar
TU	Teor de Umidade
TC	Teor de Cinzas
TMV	Teor de Materiais Voláteis
TCF	Teor de Carbono Fixo
PC	Poder Calorífico
ANEEL	Agência Nacional de Energia Elétrica
GEE	Gases Efeito Estufa
IA	Inteligência Artificial
PCS	Poder Calorífico Superior
PCI	Poder Calorífico Inferior
AD	Árvore de Decisão
KNN	<i>K-Nearest Neighbors</i>
CO ₂	Dióxido de Carbono
GRINTEQUI	Grupo de Inovações Tecnológicas e Especialidades Químicas

SUMÁRIO

1.	INTRODUÇÃO	14
2.	OBJETIVOS	16
2.1.	Objetivo Geral	16
2.2.	Objetivo Específico	16
3.	REFERENCIAL TEÓRICO	17
3.1.	Energia Renovável	17
3.2.	Biomassa	19
3.3.	Análise imediata para o estudo do Potencial Energético	20
3.3.1.	<i>Teor de Umidade</i>	20
3.3.2.	<i>Teor de Materiais Voláteis</i>	21
3.3.3.	<i>Teor de Cinzas</i>	22
3.3.4.	<i>Teor de Carbono Fixo</i>	23
3.3.1.	<i>Poder Calorífico</i>	23
3.4.	Introdução à Aprendizagem de Máquina (AM)	24
3.4.1.	Modelos de Aprendizagem de Máquina (AM)	24
3.4.2.	Funcionamento do algoritmo com Aprendizagem de Máquina (AM)	25
3.4.3.	Ferramentas que auxiliam na resolução em problemas com AM	26
3.4.3.1.	<i>K-Nearest Neighbors (KNN)</i>	26
3.4.3.2.	<i>Naive Bayes</i>	27
3.4.3.3.	<i>Clustering</i>	28
3.4.3.4.	<i>Árvore de Decisão</i>	28
3.4.4.	Desafios dos algoritmos que utilizam a Aprendizagem de Máquina	30
3.4.5.	Aplicações que auxiliam a análise de resultados obtidos pela AM	32
4.	METODOLOGIA	35
4.1.	Coleta de dados	35
4.2.	Processamento de dados	37
4.3.	Treinamento do algoritmo	38
5	RESULTADOS E DISCUSSÃO	41
5.1.	CENÁRIO A	42
5.2.	CENÁRIO B	42
5.3.	CENÁRIO C	42

5.4.	SÍNTESE COMPARATIVA ENTRE OS CENÁRIOS A,B E C.....	48
6	CONSIDERAÇÕES FINAIS.....	53
	REFERÊNCIAS	54
	ANEXO A – DADOS COLETADOS.....	60

1. INTRODUÇÃO

O histórico da humanidade está fortemente associado a energia, desde aos modelos primitivos de transportes até a necessidade residencial, no qual foi possível estabelecer um cenário favorável ao investimento de tecnologias voltadas ao desenvolvimento energético (CARVALHO, 2014). Dessa forma, é notório que a qualidade de vida da população pode estar associada ao consumo energético, onde percebe-se a colocação dos países desenvolvidos em relação a sua produção de energia, seja ela de origem não renovável ou renovável. Mediante a isso, surge uma demanda ascendente por fontes alternativas de energia, para contribuir positivamente a necessidade da população.

Neste âmbito, os biocombustíveis aderem ao competitivo mercado energético, ganhando notoriedade, por serem uma opção alternativa à utilização de combustíveis fósseis (MORELLO *et al*, 2020). É neste cenário em que as biomassas, por serem de origem animal e vegetal, desempenham função de destaque, principalmente devido à facilidade e abundância no mundo e, não menos importante, a inserção das biomassas no contexto energético global reduz a forte dependência dos combustíveis fósseis, impulsionando a diversidade na matriz energética (ANEEL, 2022). Logo, nota-se uma vasta literatura na qual há variadas pesquisas voltadas a análise de diferentes tipos de biomassa como (SIMÃO NETO, 2021), (BIZERRA, 2017) e (VASCONCELOS, 2017).

Entretanto, o processo para obtenção de dados que elegem a biomassa como biocombustível é lento, visto que cada análise para caracterizar as propriedades físico-químicas onde demanda-se muitas horas para obtenção dos resultados, além de que o pesquisador, após a conclusão da análise, necessita comparar os seus dados com trabalhos semelhantes na literatura. Basicamente, define-se as análises a partir de um método chamado de Análise Imediata (BIZERRA, 2017), onde são considerados valores para o teor de umidade, o teor de carbono fixo, o teor de materiais voláteis e o teor de cinzas. Além disto, outro parâmetro que auxilia na caracterização do biocombustível é a obtenção do poder calorífico.

Nesta perspectiva, este trabalho destina-se a classificar o potencial à combustível, a partir de uma específica ferramenta computacional de classificação oriunda da área de Aprendizagem de Máquinas denominada de Regressão Logística (SANTOS, 2017). O algoritmo é baseado no paradigma de aprendizagem com supervisão, em que faz-se necessário ter um conjunto de objetos ou atributos (dados de entrada) para os quais as saídas desejadas são conhecidas e comparadas com as saídas produzidas pelo sistema (DINIZ, C.A. & LOUZADA NETO, 2000).

Especificamente, a ideia é classificar o potencial à combustível, segundo um determinado limiar, no qual determinará duas possibilidades de classe mutuamente exclusivas: ter ou não potencial à combustível a partir da Análise Imediata e do Poder Calorífico de diferentes biomassas, levando-se em consideração todos os parâmetros adquiridos para essa classificação. Noutros termos, a pesquisa objetiva fornecer uma ferramenta de fácil manuseio, funcionando, inclusive, como um pré-resultado da análise feita em bancada, servindo como uma ferramenta eficiente na otimização de tempo. Desta forma, a base de dados irá atuar como treinamento do algoritmo, e, posteriormente, ela terá capacidade de classificar uma biomassa se ela tem ou não potencial a combustível, servindo como parâmetro de comparação na realização dos experimentos de bancada.

Nestas circunstâncias, frisa-se que a proposta possui originalidade, pois nota-se a escassez de trabalhos acadêmicos que contemplem a área de Aprendizagem de Máquina voltada a biocombustíveis e correlacionados. Mediante a isto, este trabalho pretende servir de precursor para trabalhos voltados a esta linha de pesquisa.

Adiante, será indispensavelmente evidenciada a literatura que fomenta os combustíveis oriundos da bioenergia, dando ênfase a metodologia aplicada em termos de Análise Imediata e do Poder Calorífico. Outrossim, será esclarecido como a inclusão da máxima quantidade de parâmetros se torna essencial para a obtenção de resultados mais precisos, visto que para a literatura e para o algoritmo quanto maior a base de informações, melhor caracterização do potencial energético da biomassa oriunda da bioenergia.

Dando continuidade, no trabalho será explicado os termos básicos que contemplam a disciplina de Aprendizagem de Máquina. Para esta seção será destacado o princípio de seu funcionamento, seus modelos existentes e suas diferentes ferramentas facilmente encontradas na literatura que são para solucionar problemas propostos.

Por fim, após ser possível obter as informações necessárias sobre o problema (o de classificação) e sobre a ferramenta para solução (ao utilizar a Aprendizagem de Máquina), será definido qual algoritmo melhor atende a proposta requerida, partindo da análise de quantas classes, possíveis valores que deseja-se prever, serão necessárias para resposta desejada. Desta forma, será definido um cenário onde melhor atenderá a proposta em função da qualidade do ambiente inserido.

Resumidamente, no Capítulo 1 deste trabalho será apresentada a teoria na qual engloba a problemática, no Capítulo 2 será abordada a metodologia adotada para a solução proposta e, por fim, no Capítulo 3 será discutido os resultados obtidos.

2. OBJETIVOS

2.1. Objetivo Geral

Desenvolver uma ferramenta computacional para, a partir dos parâmetros desejados, indicar o potencial à combustível de uma determinada biomassa.

2.2. Objetivos Específicos

- a) Estabelecer, a partir da literatura, uma base de dados que contemple a ferramenta computacional;
- b) Compreender melhor a bibliográfica acerca da Análise Imediata e Poder Calorífico;
- c) Entender o funcionamento do algoritmo Regressão Logística;
- d) Estabelecer os parâmetros de entrada para o algoritmo;
- e) Definir as variáveis de entrada do algoritmo, em termos dos critérios da Análise Imediata e Poder Calorífico;
- f) Definir a quantidade de entradas (atributos ou propriedades);
- g) Definir o limiar de classificação;
- h) Avaliar o comportamento da convergência do algoritmo;
- i) Apresentar as biomassas utilizadas;
- j) Avaliar a precisão do poder de classificação do método, em termos de matrizes de confusão para 3 parâmetros, para 4 parâmetros e, por fim, para 5 parâmetros;
- k) Discutir os resultados;
- l) Analisar o potencial energético a partir dos parâmetros estudados em Análise Imediata e Poder Calorífico.

3. REFERENCIAL TEÓRICO

3.1. Energia Renovável

A energia renovável consiste na obtenção, a partir de recursos naturais, do melhor aproveitamento energético mediante fontes alternativas e com possibilidades de fácil reposição da natureza. Scholten e Bosman (2013 *apud* BARBOSA, 2015) reiteram que energia renovável é uma energia derivada de processos naturais que são ou podem ser constantemente reabastecidos. Estas incluem energia solar, eólica, biomassa, geotérmica, hidrelétrica, energia maremotriz e biocombustíveis.

Vê-se, portanto, uma discussão acerca da geopolítica dos recursos naturais renováveis como matriz energética (SANTOS, 2016). Por exemplo, tem-se a obtenção de energia a partir de quedas d'água, na qual se obtém o aproveitamento a partir do potencial hidráulico; a energia a partir da força dos ventos, ou seja, energia eólica; a energia a partir da radiação solar, resultando na energia solar e, finalmente, a biomassa na qual é uma fonte renovável de energia que contempla fontes mais tradicionais, como a lenha e o subprodutos de animais e vegetais, bem como as fontes consideradas mais modernas, na oportunidade de utilização do etanol como combustível automotivo.

Vale considerar que a energia é um recurso indispensável ao funcionamento das economias, sendo necessária para a produção de bens e para o provimento de todos os serviços (ALMEIDA, 2013, p. 1).

Tolmasquim, Guerreiro e Gorini (2007), em uma prospectiva do setor energético brasileiro para o período 2005 à 2030, apontam que a competitividade econômica dos países e a qualidade de vida de seus cidadãos serão, ainda mais intensamente, influenciadas pela energia, pois, em um mercado global, tendo em face as crescentes preocupações com o meio ambiente, essa influência se mostra cada vez mais decisiva. Nesse contexto, as economias que melhor se posicionam quanto ao acesso a recursos energéticos de baixo custo e de baixo impacto ambiental obtêm importantes vantagens comparativas.

Quando um recurso natural passa a ser escasso, torna-se um recurso natural estratégico e potencialmente vital para o desenvolvimento de atividades econômicas uma vez que o componente de conflito da geopolítica dos recursos naturais acontece quando em alguns territórios há abundância e, em outros, escassez.

Nessa perspectiva, com a característica principal de renovação em curto prazo, a energia renovável ganha destaque no cenário mundial devido, principalmente, a possibilidade

de independência de cada país para produzir sua própria energia e, com isso, reduzir as taxas de importação, além de contribuir para a redução de gases nocivos ao meio ambiente, como os gases associados ao efeito estufa.

No caso dos recursos naturais renováveis para a produção de energia, sua importância geopolítica está, principalmente, na possibilidade de minimizar a dependência da importação de energia, além de que com a diversificação da matriz energética, os países que investem em produção de energia renovável tendem a diminuir sua vulnerabilidade com as altas dos preços internacionais dos combustíveis fósseis. (BARBOSA, 2015, p. 202)

No contexto da geopolítica, a disponibilidade de energia nas condições de quantidade e qualidade adequadas, a custos competitivos, tem-se constituído em um dos mais importantes pré-requisitos para o desenvolvimento econômico das nações. Tendo em vista, a energia tem sido tratada como um bem de natureza estratégica, e não por acaso a segurança energética tornou-se um dos temas mais relevantes da agenda mundial.

A dependência da importação de recursos energéticos por parte de algumas nações e a necessidade de redução dos Gases de Efeito Estufa (GEE) na atmosfera gerou uma crescente mobilização em torno da alteração da matriz energética, pois a geopolítica da energia começa a adquirir novas dimensões, onde a eficiência energética e o maior domínio de tecnologias das energias renováveis e políticas de criação de novos mercados para as mesmas, tornam-se cada vez mais urgentes.

Barbosa (2015) destaca que, nos debates internacionais contemporâneos, a questão energética vem sendo tratada a partir de múltiplas visões: geopolítica, militar, econômica, ou meramente comercial. Contudo, Rifkin (2011, p. 107) revela que a importância de analisar a questão energética sob uma ótica política é mais urgente, pois os regimes de energia moldam a natureza das civilizações (como são organizadas, como os benefícios do comércio são distribuídos, como o poder político é exercido e como as relações sociais são conduzidas). A demanda por energia tem aumentado consideravelmente por parte de países emergentes, de forma a viabilizar o desenvolvimento de suas economias.

Por essas razões, torna-se imprescindível que a comunidade internacional aprimore e expanda o uso de fontes renováveis de energia nas suas mais diversas aplicações (BARBOSA, 2015). A mudança do atual paradigma energético é de fundamental importância para o funcionamento de todo o domínio produtivo, mais especialmente para o setor de transportes, cuja dependência em relação ao petróleo responde por cerca de 98% da demanda

mundial de combustíveis, que é o foco deste trabalho, no que diz respeito ao potencial energético de diferentes biomassas, destinadas à produção desses combustíveis.

3.2. Biomassa

Desde a crise do petróleo, em 1970, percebe-se que ascendentemente há uma procura por fontes de energia na utilização de matérias-primas alternativas às convencionais, onde o cenário, até então, era dominado pelos combustíveis fósseis (BONASSA *et al.*, 2015).

No tocante ao fator ambiental, desde a referida década, Mascarenhas (2010) ressalta a preocupação com o efeito estufa, a escassez dos combustíveis fósseis e a importância de se tornar as fontes renováveis como matriz energética, tem levantado discussões sobre desenvolvimento sustentável e o futuro do planeta. Com isso, de forma a ser utilizada também como uma ferramenta para busca da mitigação das mudanças climáticas, a biomassa se destaca por ser uma das fontes de energia que auxilia na redução dos impactos ambientais, já que se origina totalmente de material orgânico.

A biomassa, que pode ser caracterizada como qualquer matéria orgânica que possa ser transformada em energia térmica, mecânica ou elétrica é utilizada desde os primórdios da civilização humana. Além disso, considera-se como uma das fontes para produção de energia com maior potencial de crescimento nos próximos anos, pois, recentemente, passou a ser considerada uma alternativa para a diversificação da matriz energética mundial e consequente redução da dependência de combustíveis fósseis (GOLDEMBERG, 2009).

Diante do contexto, a biomassa é historicamente utilizada como um biocombustível, competindo à madeira a posição em destaque da matéria-prima de fácil disponibilidade e abundância no mundo. Já que possui origem animal ou vegetal, inclusive de resíduos urbanos e de resíduos agrícolas, Santos, Nascimento e Alves (2017) escrevem que a biomassa pode ser designada para a geração de calor, pois possui potencial à combustível, podendo ser empregada na geração de eletricidade além de ter aplicabilidade como um combustível líquido ou como um combustível gasoso.

Com foco nesta investigação, reitera-se que o recente interesse na energia da biomassa tem dado ênfase em aplicações que produzem combustíveis líquidos para o setor de transportes (biocombustíveis). O uso eficaz destas novas matérias-primas de biomassa para a coprodução local de aquecimento, eletricidade e combustível de transporte, também teria um impacto profundo na capacidade das populações rurais de acessar formas de energia modernas e mais limpas (SANTOS, 2016).

O progresso no desenvolvimento de alternativas para a energia de biomassa, além de aliviar a pressão em recursos finitos de combustíveis fósseis, pode reduzir os custos de mitigação de emissões de carbono (MORELLO, *et al*, 2020). Os combustíveis como etanol e biodiesel são um dos principais produtos provenientes da biomassa. Esses biocombustíveis, além de mais econômicos, são sustentáveis, pois sua produção não emite gases poluentes na atmosfera.

Em relação ao aspecto econômico, quando adotada a Biomassa como fonte alternativa de energia em substituição à energia proveniente de combustíveis fósseis, diminuí-se os gases tóxicos e poluentes ao meio ambiente, diminuindo consequentemente os gastos públicos com a saúde da população em geral.

No aspecto social, Morello *et al.* (2020), comentam que o uso da biomassa como fator energético, tem-se notado como grande gerador de empregos desenvolvendo as regiões menos favorecidas, impactando na economia regional através do aumento da receita local, reduzindo então o êxodo rural. O efeito não surtiu somente no campo, mas nas cidades que iriam prover dos bens e serviços da indústria urbana.

O conceito de desenvolvimento sustentável vem se ampliando a cada dia e o uso de biocombustíveis provenientes da Biomassa contribuem positivamente para a redução da emissão dos gases de efeito estufa.

Acredita-se que é preocupação de todos visa caminhos para amenizar a emissão de gases de efeito estufa, de modo a diminuir o impacto ambiental causado na produção de energia. Como alternativa, consideramos que a biomassa possa fazer um diferencial considerável, pois é uma fonte renovável significativa.

3.3. Análise Imediata e Poder Calorífico para o estudo do Potencial Energético

A Análise Imediata consiste na determinação, em conteúdo percentual, da massa do material com potencial para combustível, se ramificando em: Teor de Umidade (TU), Teor de Materiais Voláteis (TMV), Teor de Carbono Fixo (TCF) e Teor de Cinzas (TC). Em relação ao Poder Calorífico (PC), este funciona como outro parâmetro que auxilia a caracterização da biomassa. Para esse tipo de análise, comumente são analisados os parâmetros nos quais serão devidamente esclarecidos na sequência do texto.

3.3.1. Teor de Umidade (TU)

O Teor de Umidade (TU) está relacionado a quantidade de água presente na biomassa. Esta classificação pode estar atrelada a quantidade úmida ou seca, podendo ser analisada experimentalmente pesando as amostras antes e após submetê-las à secagem.

Suscintamente, após a biomassa em estudo passar pelo tratamento adequado a suas especificidades, a metodologia adotada para estabelecer esse parâmetro permite alocar determinada quantidade da amostra em cadinhos, feitos de porcelana, separando-os em duplicatas para, posteriormente, se submeterem à secagem em fornos com altas temperaturas, geralmente em torno de 750 °C (PESSOTTI, 2017).

Ademais, consiste em deixar os cadinhos dentro do forno por um determinado tempo, esperar eles resfriarem no dessecador e realizar as pesagens, utilizando uma balança analítica para efeitos de comparação e análise; verificar a quantidade de massa que possui cada cadinho antes e depois de ir ao forno. Ao concluir essa primeira parte do processo, destina-se uma determinada quantidade da amostra do objeto de estudo em cada cadinho, repetindo-se o processo de secagem e pesagem em condições de temperatura diferentes, em torno de 105 °C (PESSOTTI, 2017).

Repete-se o processo para cada duplicata e anota-se os valores, para que seja possível utilizar o Equacionamento (1) que permite calcular o Teor de Umidade, baseando-se pela norma NBR 16550, dada a seguir:

$$TU [\%] = \frac{MAu - MAs}{MAu} \times 100 \quad (1)$$

De acordo com a Equação (1) Simão Neto (2021) define que a partir do equacionamento é possível obter os valores acerca do TU para cada amostra e suas duplicatas, onde MAu representa a massa da amostra úmida e MAs representa a massa da amostra seca. Por fim, como resultado é dado em porcentagem. Para esse parâmetro, tem-se que, resultados próximos a 100% não são desejáveis para o problema, pois umidade elevada atrapalha o poder calorífico da amostra, dificultando a possibilidade de queima do combustível.

3.3.2. Teor de Materiais Voláteis (TMV)

O Teor de Materiais Voláteis (TMV) é a classificação para o potencial à combustão da biomassa em estudo. Os MV representam a facilidade ou a dificuldade da amostra de se submeter a queima, pois à medida em que se obtém um maior valor para esse parâmetro,

maior será a reatividade e o potencial à combustível da amostra (BIZERRA, 2017). Para este parâmetro, analisa-se o resultado da amostra obtida por meio do Teor de Umidade.

Tomando-se como referência a norma ASTM D3175, a qual adota base seca para a determinação do TMV (em termos de porcentagem) tem-se que tal parâmetro é descrito de acordo com a Equação (2).

$$TMV [\%] = \left[\frac{Mi - Mf}{Mi - Mc} \times 100 \right] - TU \quad (2)$$

Em que Mc representa a massa do cadinho com tampa; Mf a massa final do cadinho de porcelana com tampa e com a amostra; Mi a massa inicial do cadinho de porcelana com tampa e com a amostra; por fim, TU refere-se ao Teor de Umidade. Os dados obtidos para massa são geralmente estabelecidos em gramas (SIMÃO NETO, 2021).

Diante do exposto, os resultados desejáveis para TMV ocorrem quando estes se apresentam elevados, visto que esse parâmetro tem característica relacionada à volatilidade da biomassa a ser acionada, no qual um alto TMV implica em uma baixa energia para a queima, o que é uma característica favorável ao poder calorífico, já que, a biomassa, devido ao seu elevado teor de voláteis, é facilmente inflamável e queima rapidamente, mas é difícil de ser controlada, o que pode afetar o processo de combustão.

3.3.3. Teor de Cinzas (TC)

O Teor de Cinzas (TC), refere-se a porção remanescente de fração inorgânica após a calcinação, no qual verifica-se que um elevado resultado não é desejado para uma biomassa com potencial à combustão, visto que representa um menor poder calorífico do biocombustível, chegando a inviabilizar a utilização da biomassa para este fim.

Ponte *et al* (2019) definem que a característica que contempla o maior resultado em teores de cinzas evidencia que ela não é interessante para um biocombustível, pois as cinzas formadas no processo de queima podem se estabelecer dentro do combustor, uma vez que durante o processo de queima é notória elevadas temperaturas, acarretando no amolecimento das cinzas e assim, podem causar aglomeração destas em locais que prejudicam a convecção de energia.

Baseado pela norma NBR 16550, o TC é dado pela Equação (3).

$$TC [\%] = \frac{MC}{MAS} \times 100 \quad (3)$$

Onde MC representa a massa das cinzas, geralmente representada em gramas, MAS representa a massa da amostra na base seca (SIMÃO NETO, 2021).

3.3.4. Teor de Carbono Fixo (TCF)

O Teor de Carbono Fixo (TCF) classifica o material resultante após decomposição dos voláteis, após retirar as Cinzas e a Umidade, sendo a maior parte do aproveitamento energético da biomassa quando ela é utilizada como combustível. Considera-se, para esse parâmetro, que quanto maior for o valor obtido, mais lento será o processo de queima do combustível produzido (GIRÃO, 2016).

Segundo a norma ASTM D3172, dado também em termos de porcentagem, o TCF é obtido conforme a Equação (4).

$$TCF [\%] = 100 - TC - TMV \quad (4)$$

Cuja expressão depende do TC, Teor de Cinzas e TMV, Teor de Materiais Voláteis. Importante comentar que todas as variáveis envolvidas na Equação (4) são também dadas em porcentagem (SIMÃO NETO, 2021).

3.3.5. Poder Calorífico (PC)

O Poder Calorífico (PC) está relacionado a quantidade de calor que está disponível para liberação a partir da massa analisada com potencial a combustível. Com base nessa perspectiva, entende-se que na composição química da amostra durante o processo de queima, uma determinada quantidade de água é liberada, enquanto a quantidade restante fica contida na amostra (GADELHA, 2016).

Portanto, de acordo com o contexto, obtém-se: o Poder Calorífico Superior (PCS) e o Poder Calorífico Inferior (PCI), no qual o primeiro adquire a água condensada a partir do processo de queima, para readequar o calor latente de vaporização, em condições de combustão completa. No caso do PCI, este se refere ao calor gerado nas mesmas condições que o PCS, entretanto, para esta situação, a água está em estado gasoso, não incluindo calor latente de vaporização (GIRÃO, 2016).

3.4. Introdução à Aprendizagem de Máquina (AM)

A Aprendizagem de Máquina (AM) lida com reconhecimento de padrões, e, para isso, utiliza algoritmos computacionais que possuem diversas finalidades. Peres *et al* (2012) definem que modelos de aprendizagem podem ser determinados a partir de um conjunto de dados ou até por representações de experiência. Com isso, essa área de estudo permite desenvolver uma Inteligência Artificial (IA) capaz de prever uma solução a um determinado problema a partir de um aspecto proposto, partindo do pressuposto da inclusão de dados e, posteriormente, análise destes, verificando, inclusive, a relação entre acertos e entre erros. Dessa forma, com a AM, o algoritmo consegue desempenhar sua função eficientemente, aprendendo com os erros à medida em que ocorre iterações.

3.4.1. Modelos de Aprendizagem de Máquina (AM)

Existem modelos diferentes para AM, sendo elas: modelo supervisionado, modelo não-supervisionado e, por fim, modelo por reforço. Basicamente, a diferença entre os três consiste em que no primeiro, o algoritmo ajustará os parâmetros inseridos a partir do erro e mediante a resposta obtida já esperada (MAHESH, 2020), ou seja, os dados de entrada são rotulados e são designados para aprender determinada função que irá obter uma saída desejada.

É importante salientar que, durante o processamento do algoritmo, onde 100% é a margem total, destina-se uma porcentagem de dados para treinamento e a porcentagem restante de dados é utilizada para teste.

Portanto, para o primeiro modelo, a partir do erro obtido na fase de teste, o algoritmo já conhece o resultado esperado e o compara com o resultado obtido (MAHESH, 2020). O segundo modelo caracteriza o ajuste sem o conhecimento do resultado esperado, ele apenas indica o resultado obtido com base na maximização de medidas de qualidade na resposta obtida. O terceiro modelo representa uma modelagem de iteração que, a partir de uma situação dinâmica com um objetivo específico, o algoritmo converge ao melhor resultado possível otimizado, com base na tentativa e erro, repetidas vezes. Nesta perspectiva, em situações onde há quantidade menores de dados, opta-se pelo modelo supervisionado, entretanto, Mahesh (2020) afirma que a aprendizagem pelo modelo não-supervisionado

oferece melhor desempenho e resultados otimizados para a situação onde há grande quantidade do acervo de dados.

3.4.2. Funcionamento do algoritmo com Aprendizagem de Máquina (AM)

A Aprendizagem de Máquina permite solucionar problemas de classificação e de regressão. Dentro desta área de conhecimento, utiliza-se o reconhecimento de padrões para analisar cada situação e, para isso, existem propriedades que são imprescindíveis durante processo. Dentre tais propriedades, tem-se: padrões, classes, limiar de ativação e saída.

Castro e Prado (2022) definem padrões como uma propriedade que consiste em um agrupamento de objetos identificados devido a sua verossimilhança, assim, um problema de reconhecimento de padrão pode ser utilizado em tarefas de classificação ou de categorização.

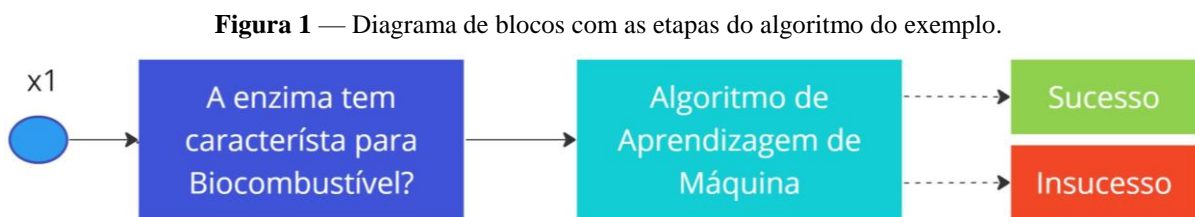
Adiante, entende-se por classe um conjunto de atributos que possuem objetos em comum (Haykin, 2001). Dessa forma, pode-se inferir que, para existir uma classe, deve-se, primeiramente, avaliar qual padrão pertence o objeto, pois, posteriormente, este objeto será adequado a uma classe que o contemple recorrentemente.

Durante o estudo de Aprendizagem de Máquina é familiar a propriedade referente ao limiar de ativação. Este termo refere-se aquele que maximiza o ganho de informação resultante da segmentação (Haykin, 2001). O limiar de ativação é recursivamente aplicado a cada partição obtida até que o critério de parada seja atingido. Logo, em outras palavras, o limiar representa os parâmetros que fomentam a base de análise para o algoritmo, com caráter comparativo.

Existem métricas específicas que delimitam a estrutura do algoritmo com relação na definição da resposta de saída. Normalmente, encontra-se algoritmos de AM com saída binária ou multiclasse. Para a saída binária, 0 ou 1, tem-se que a ferramenta irá adotar apenas duas alternativas como resposta obtida, geralmente associa-se a 0 respostas tidas como classe “não” e associa-se a 1 respostas tidas como classe “sim”. Em contrapartida, para a classificação multiclasse, considera-se situações onde a resposta desejada pode ser mais de duas classes. Por exemplo, predizer se uma enzima possui características para biocombustível ou não, entende-se que é um problema com saída binária, entretanto, na classificação multiclasse o método irá inferir, qual a enzima será analisada.

Um algoritmo de classificação, de modo geral, consiste em um conjunto de valores de entrada, que se submete a regras específicas para cada modelo e, por fim, irá classificar o algoritmo em termos de classes, com saída binária ou multiclasse (CHEIN, 2019). Para

ilustrar, imagina-se o exemplo dito anteriormente em relação às enzimas. Neste âmbito, há uma saída referente ao sucesso e outro ao insucesso do problema. Sugere-se, para a melhor compreensão da modelagem, que os dados escolhidos para a fase de treinamento já sejam previamente rotulados, de modo que seja fácil a análise do algoritmo na fase de teste. A Figura 1 representa, em diagrama de blocos, as etapas que envolvem a estrutura de um algoritmo:



Fonte — Autora, 2023.

A Figura 1 apresenta x_1 como representação do conjunto de valores submetidos a regras do algoritmo proposto. Dessa forma, um algoritmo que resolve problemas de classificação se diferenciará de outro de acordo com o classificador utilizado e a quantidade de classes desejadas, logo, cada algoritmo terá sua particularidade. Pode-se incluir, para cada objeto de estudo, várias ferramentas capazes de solucionar o problema, como será visto em diante.

Já para problema de regressão possui valores reais, Chein (2019) define que geralmente em forma decimal, como saída. Tem-se uma variável independente ou conjunto de variáveis independentes e uma variável dependente o que é a resposta desejada, dadas as variáveis independentes.

3.4.3. Ferramentas que auxiliam na resolução em problemas com AM

3.4.3.1. *K-Nearest Neighbors* (KNN)

O *K-Nearest Neighbors* (KNN), no português, K vizinhos mais próximos, é um algoritmo de aprendizado de máquina supervisionado simples que pode ser usado para resolver problemas de classificação, onde K é um valor determinável. Kramer (2013) mostra que o método é fácil de implementar e entender, mas tem uma grande desvantagem de se tornar significativamente mais lento à medida em que cresce.

O KNN funciona com base em quanto o dado fornecido pode ser similar ao outro dado, estabelecendo assim, a verossimilhança em dados ditos como vizinhos. Este método funciona basicamente em como os dados se comportam estabelecendo uma menor distância entre si sobre os dados disponíveis para análise e, conseqüentemente, mostra o resultado a partir da similaridade (Beckmann *et al.* 2015). Em outros termos, o algoritmo busca classificar cada objeto a partir de um conjunto de dados, levando em consideração a sua distância em relação aos seus vizinhos.

Pode-se dizer que o método KNN é um dos mais simples e mais utilizados algoritmos de Aprendizagem de Máquina. Em poucas palavras, consiste em classificar um novo exemplo, calculando a distância entre os K e exemplos mais próximos a ele.

O método KNN armazena todas as amostras de treinamento, incluindo suas características, em um espaço de acordo com suas métricas sem processamento ou cálculo (LIU, 2018 *apud*, SOUSA, et al, 2017). Quando o modelo recebe um objeto para ser previsto, coloca o novo objeto nesse espaço. O modelo, então, faz previsões olhando para os vizinhos mais próximos do novo objeto, ou seja, o método KNN é um algoritmo que classifica todas as estatísticas disponíveis com base em uma medida de similaridade.

O algoritmo de classificação é realizado de acordo com as seguintes etapas (Beckmann *et al.* 2015): 1) Identificar um objeto não classificado; 2) Calcular a distância, geralmente euclidiana; 3) Selecionar os K vizinhos cujo possui menor distância; 3) Verificar as classes nas quais os objetos possuíram menor distância e contabilizar a quantidade de vezes em que tais classes foram identificadas. 4) Repete-se o processo, para validação e, a partir disso, classifica-se o objeto não identificado a classe que mais teve ocorrência.

3.4.3.2. Naive Bayes

Considerado um classificador multinominal, o modelo computacional Naïve Bayes está inserido no contexto de Aprendizagem de Máquinas (*Machine Learning*), no qual estabelece como premissa a suposição de independência entre as variáveis do problema, pré-definindo a probabilidade dentre as observações realizadas a partir da teoria dos grafos.

Fernandes e Filho (2019) apontam que o classificador Naive Bayes é um dos principais algoritmos da área, sendo para muitas pessoas o primeiro contato com esse universo. A fórmula do classificador determina apenas os passos a serem seguidos para o aprendizado e, por ser muito simples e rápido, possui um desempenho relativamente maior do

que outros classificadores. Além disso, o Naive Bayes só precisa de um pequeno número de dados de teste para concluir classificações com uma boa precisão (LIRA, et al, 2019).

Ainda segundo Lira (et al, 2019), o Teorema de Bayes fornece uma forma de calcular a probabilidade posterior $P(C | X)$ a partir de $P(C)$, $P(x)$ e $P(X | c)$.

Por ser mais simples e envolver uma base estatística fácil de entender, ele é o preferido em muitos casos. Um dos destaques do uso é o fato de que é fácil ilustrar também. Se houver uma base de exemplo pequena, é possível realizar os cálculos e seguir o procedimento do algoritmo.

3.4.3.3. Clustering

Conhecido por sua característica de Análise de Agrupamento de Dados, o método *Clustering* segundo Lior Rokach e Oded Maimon (2005) consiste em introduzir dados na fase de treinamento, na tentativa de relacioná-los em grupos por grau de semelhança com um novo dado inserido o algoritmo busca padrões existentes nele e cataloga-o no respectivo grupo. É uma ferramenta não-supervisionada, geralmente aplicada em situações onde deseja-se encontrar padrões inesperados nos dados a partir de x parâmetros de entrada.

3.4.3.4. Árvore de Decisão (AD)

Decision Trees ou Árvore de Decisão (AD) é uma das ferramentas de Aprendizagem de Máquina capaz de classificar, a partir do aprendizado de máquina supervisionado para prever classes, onde encaixa-se classes binárias, ou pode prever valores numéricos (DINIZ, 2000). O diferencial desse modelo consiste em estabelecer uma relação de hierarquia, a partir dos *decision nodes*, definido por nós, onde há uma ramificação até obter, de fato, o resultado final que se dá por meio dos *leaf-nodes* (nós-folhas).

Em relação ao seu funcionamento, este algoritmo estabelece vários nós, e, para cada nó, haverá um resultado desejado que seguirá caminhos variados, por meio dos nós-folha. É válido destacar que é possível que estes caminhos possam estabelecer um mesmo resultado, nos quais irão se diferenciar pela estrutura, já que busca-se rotas mais otimizadas.

Magee (1964) afirma que a principal característica desse modelo está atrelada a entropia e inconsistência nos dados, estabelecendo uma alta ou baixa desorganização que é justamente referente aos cálculos necessários para normalização de dados e, por fim, busca-se homogeneizar os dados para que eles se tornem potenciais ganhos em informação, caso contrário, se a quantidade for pequena, infere-se que os dados estão altamente desorganizados,

portanto, mais distantes do resultado esperado. Logo, para uma relação entre variáveis independentes, chamadas e preditoras e entre variáveis dependentes, chamadas de *target*, entende-se que o domínio de aprendizagem do algoritmo está relacionada a qualidade do relacionamento entre os dados da variável preditora e entre a variável *target*.

3.4.3.6. Regressão Logística (RL)

A Regressão Logística (RL) baseia-se em um modelo estatístico de classe multivariada ou de classe binária, que se desenvolve a partir de uma função linear com base em uma combinação linear de parâmetros de entrada (SANTOS *et al*, 2005). Este algoritmo é considerado um tipo de modelo usado para prever a probabilidade de uma observação pertencer a uma determinada classe. Diante do exposto, a RL é chamada de "regressão" porque tenta prever uma probabilidade, mas é chamado de "logística" porque essa previsão é feita com base na função logística. A função logística produz uma "curva S" que é usada para modelar a relação entre as variáveis independentes, preditoras, e a variável dependente, a classe a ser prevista.

A função logística é utilizada no modelo para modelar a relação entre as variáveis independentes e a variável dependente, com base nisso a classe é definida, geralmente modelando a probabilidade em um certo limiar (SANTOS *et al*, 2005). A partir dos coeficientes estimados no modelo, é possível calcular a probabilidade de uma observação pertencer a uma determinada categoria, dado um conjunto de valores para as variáveis independentes.

Diante disso, Mendes e Vega (2011) ressalta que a RL considera um conjunto de análises de dados e estima o resultado com base em uma saída categórica. A partir dele, é possível obter resultados prévios, funcionando como uma segunda opinião, para situações de variáveis complexas a partir de uma análise aleatória. A RL vem com a proposta de tentar resolver problemas matemáticos de grande complexidade, constituindo um método de otimização que busca os máximos e mínimos de uma equação.

Por meio da RL, os dados formarão matrizes de atributos que se direcionarão ao processamento de dados, e, posteriormente, obtém-se uma saída de desejada. O algoritmo escolhido para o presente trabalho consiste em um modelo supervisionado com curvas capaz de oferecer melhor desempenho para o resultado (SANTOS *et al*, 2005). Pode-se justificar a eficiência do modelo com a característica dele de realizar o treinamento dos dados já

utilizando os dados de saída, ou seja, o algoritmo já conhece o resultado final para cada objeto realizado e, a partir disso, consegue ter melhor precisão na resposta de saída.

Este algoritmo necessita de um parâmetro chamado de norma do vetor gradiente da função objetivo, cuja definição consiste em ser é uma medida da taxa de variação da função objetivo em relação aos parâmetros do modelo. Nesta perspectiva, o vetor gradiente é um vetor que aponta na direção de maior aumento da função objetivo e sua norma é a magnitude desse vetor (Mendes e Vega, 2011). Em outras palavras, é uma medida da rapidez com que a função objetivo está mudando em relação aos parâmetros do modelo. Esta norma é considerada o vetor das derivadas parciais da função objetivo em relação a cada parâmetro do modelo. A norma do vetor gradiente é utilizada para determinar a direção e a magnitude do passo de otimização.

3.4.4. Desafios dos algoritmos que utilizam a Aprendizagem de Máquina

O algoritmo baseado na Aprendizagem de Máquina, por vezes, é submetido à diversas situações nas quais obtém-se resultados não satisfatórios que, posteriormente, podem ser decisivos para a qualidade da resposta obtida (COSTA e PAIXÃO, 2018). Diante disto, vê-se a necessidade de explicar, de forma resumida, alguns desafios enfrentados pelos algoritmos de AM que serão explicados ao longo desta seção. Conforme o que foi dito, lista-se alguns casos mencionados por Ludermir (2021).

- Insuficiência nos dados

Dados insuficientes: podem afetar negativamente o aprendizado de máquina causando a baixa precisão, ou seja, afetam modelos de aprendizado de máquina necessitam de uma quantidade suficiente de dados para treinar e ajustar os parâmetros. Com dados insuficientes, os modelos não têm a capacidade de aprender as tendências e relações subjacentes nos dados, o que pode levar a resultados incoerentes.

Sobreajuste: modelos de aprendizado de máquina podem sobreajustar aos dados de treinamento quando há poucos dados disponíveis. Isto pode fazer com que o modelo se ajuste ao ruído ao invés de aprender as tendências gerais e relações subjacentes nos dados.

Incapacidade de lidar com variações: modelos de aprendizado de máquina precisam de dados suficientes para lidar com variações nos dados e ambientes. Com dados insuficientes, os modelos podem ser sensíveis a variações e não ser capazes de generalizar para novos dados.

Falta de capacidade de lidar com complexidade: modelos complexos precisam de mais

dados para serem capazes de aprender e generalizar corretamente. Com dados insuficientes, os modelos podem não ser capazes de aprender a complexidade subjacente aos dados.

É importante ter uma quantidade suficiente de dados para garantir que o modelo de aprendizado de máquina seja preciso e adeque bem para novos dados. Isso pode incluir coletar mais dados, gerar dados sintéticos ou selecionar uma técnica de aprendizado de máquina que requer menos dados.

- Dados de baixa qualidade

Dados de baixa qualidade podem afetar negativamente o aprendizado de máquina de várias maneiras, tais como:

Baixa precisão: dados de baixa qualidade podem conter erros, incompletos ou informações irrelevantes, o que pode levar a uma precisão insuficiente do modelo.

Sobrepreenchimento: dados de baixa qualidade podem incluir ruído ou variações aleatórias, o que pode levar ao sobrepreenchimento do modelo. Isto pode fazer com que o modelo se ajuste ao ruído ao invés de aprender as tendências gerais e relações subjacentes nos dados.

Desbalanceamento de classes: se os dados de treinamento estão desbalanceados, ou seja, uma classe tem muito mais exemplos do que outra, isso pode afetar negativamente a capacidade do modelo de prever corretamente para as classes menos representadas.

Falta de representatividade: Se os dados de treinamento não são representativos da população geral, o modelo pode não generalizar bem para novos dados.

Inconsistência: dados de baixa qualidade podem ser inconsistentes, como variações na forma como as informações são coletadas, medidas ou registradas. Isso pode causar problemas para a modelagem e a interpretação dos resultados.

É importante ter uma boa qualidade dos dados para garantir que o modelo de aprendizado de máquina seja preciso e se adeque bem para novos dados. Neste momento, inclui-se verificar a precisão, a consistência e a representatividade dos dados, bem como lidar com desbalanceamento de classes.

- Overfitting

Overfitting ocorre quando um modelo de aprendizado de máquina muito se ajusta aos dados de treinamento. Isto significa que o modelo é capaz de reproduzir muito bem os dados de treinamento, mas não se adequa bem para novos dados. Em outras palavras, o modelo se

torna muito complexo e se ajusta a variações aleatórias nos dados de treinamento, ao invés de aprender as tendências gerais e relações subjacentes. Este fato resulta em uma alta precisão no conjunto de treinamento mas baixa precisão no conjunto de teste.

Existem várias maneiras de lidar com o *overfitting*, inclui-se: reduzir a complexidade do modelo, adicionar mais dados de treinamento, utilizar técnicas de regularização, que restringem a capacidade de ajuste do modelo ou validação cruzada, para selecionar o modelo que tem melhor desempenho geral.

- Underfitting

Underfitting ocorre quando um modelo de aprendizado de máquina pouco se ajusta aos dados de treinamento. Isto significa que o modelo não é capaz de reproduzir os dados de treinamento de maneira adequada, e também não se adequa bem para novos dados. Em outras palavras, o modelo é muito simples e não consegue capturar as tendências e relações subjacentes nos dados. Deste modo, a resultante é uma baixa precisão tanto no conjunto de treinamento quanto no conjunto de teste.

Existem várias maneiras de lidar com o *underfitting*, dentre elas: aumentar a complexidade do modelo, adicionando mais características relevantes ou utilizando modelos mais flexíveis.

3.4.5. Aplicações que auxiliam a análise dos resultados a partir da AM

Na otimização de Aprendizado de Máquina, o objetivo é encontrar os valores ótimos dos parâmetros do modelo que minimizam a função objetivo. Para isso, verifica-se a necessidade de apresentar mecanismos de aplicação que auxiliam a análise dos resultados a partir da AM. Nesta seção serão brevemente explicados os termos (LUDEMIR, 2021): precisão da taxa de convergência, gráficos de convergência, análise de épocas e matrizes de confusão.

- Análise das épocas

Em um algoritmo de AM, uma época representa uma iteração completa através do conjunto de dados de treinamento. Durante uma época, o algoritmo de otimização atualiza os parâmetros do modelo com base nos gradientes calculados a partir do conjunto de dados de treinamento.

Analisa-se, portanto, o cenário onde as épocas são uma forma de avaliar o desempenho do modelo durante o processo de treinamento, logo é possível identificar se o modelo está se ajustando corretamente aos dados e se está convergindo para uma solução ótima. Uma forma de analisar o comportamento das épocas é plotar a função desejada ao longo das épocas. Se a função desejada estiver caindo a cada época, isso indica que o modelo está se ajustando corretamente aos dados e está se movendo em direção a uma solução ótima.

- Gráficos de convergência

Analisar o gráfico de convergência em regressão logística é uma forma de avaliar se o modelo está convergindo para uma solução ótima. O gráfico de convergência mostra como a função poderá mudar ao longo das épocas de treinamento a partir de uma norma de vetor gradiente.

Esta norma de vetor gradiente é frequentemente usada como uma medida da precisão da solução encontrada, e uma norma de gradiente menor indica que a solução é mais precisa. Alguns algoritmos de otimização, como o algoritmo RL, usam a norma do vetor gradiente para controlar a taxa de aprendizado, ou seja, a velocidade com que os parâmetros são atualizados.

Se a função dada estiver caindo a cada época, em função da norma do vetor gradiente, isso indica que o modelo está se ajustando corretamente aos dados e está se movendo em direção a uma solução ótima. Em contrapartida, se o gráfico apresentar uma tendência de estabilidade ou até apresentar um aumento gradativo, isto pode ser um indicativo de que o modelo está tendo dificuldade em convergir para uma solução ótima, o que pode ser causado por problemas como a inicialização dos parâmetros ou falta de dados suficientes.

Adiante, a partir da análise do gráfico de convergência, é possível identificar se o modelo está convergindo de forma adequada e se é necessário fazer ajustes no modelo, como aumentar o número de épocas, mudar a taxa de aprendizado, ou utilizar outras técnicas de otimização. Além disso, é possível identificar se o modelo está sofrendo de *overfitting* ou *underfitting*, caso a função de custo esteja muito baixa no conjunto de treinamento e muito alta no conjunto de validação.

- Matriz de confusão

A matriz de confusão é uma tabela que é utilizada para descrever a performance de um

modelo de classificação. Ela mostra a frequência com que cada classe foi prevista corretamente ou incorretamente. Este método é uma ferramenta útil para entender como o modelo está se comportando e identificar quais classes estão sendo previstas incorretamente com mais frequência. Ele também é útil para calcular medidas de performance como precisão e taxa de acerto.

A matriz de confusão é uma ferramenta comumente utilizada para avaliar o desempenho de um modelo de classificação, incluindo a RL. Ela mostra a quantidade de previsões corretas, VERDADEIROS POSITIVOS (VP), e VERDADEIROS NEGATIVOS (VN), e as previsões incorretas, FALSOS POSITIVOS (FP) e FALSOS NEGATIVOS (FN), feitas pelo modelo. Este tipo de matriz é uma forma útil de avaliar o desempenho do modelo pois ele permite visualizar de forma clara os erros cometidos pelo modelo, o que permite identificar onde o algoritmo precisa ser aprimorado. A Figura 2 indica a representação deste modelo de aplicação.

Figura 2 — Modelo genérica de uma matriz de confusão.

		PREVISTA	
		SIM	NÃO
REAL	SIM	VP	FP
	NÃO	FN	VN

Fonte — Autora, 2023.

- Precisão na taxa de convergência

A precisão do algoritmo na fase de treinamento e na fase teste de regressão logística pode ser medida a partir na quantidade de verdadeiros positivos e o verdadeiros negativos obtidos no final do processo. O verdadeiro positivo é o número de vezes em que o modelo prevê corretamente uma classe positiva e o verdadeiro negativo é o número de vezes em que o modelo prevê corretamente uma classe negativa.

Desta forma, a precisão pode ser calculada como a razão entre a soma entre os verdadeiros, positivos e negativos, pelo número total de amostras, logo, a Equação (5)

representa esta relação.

$$\frac{VP + VN}{X} \times 100 \quad (5)$$

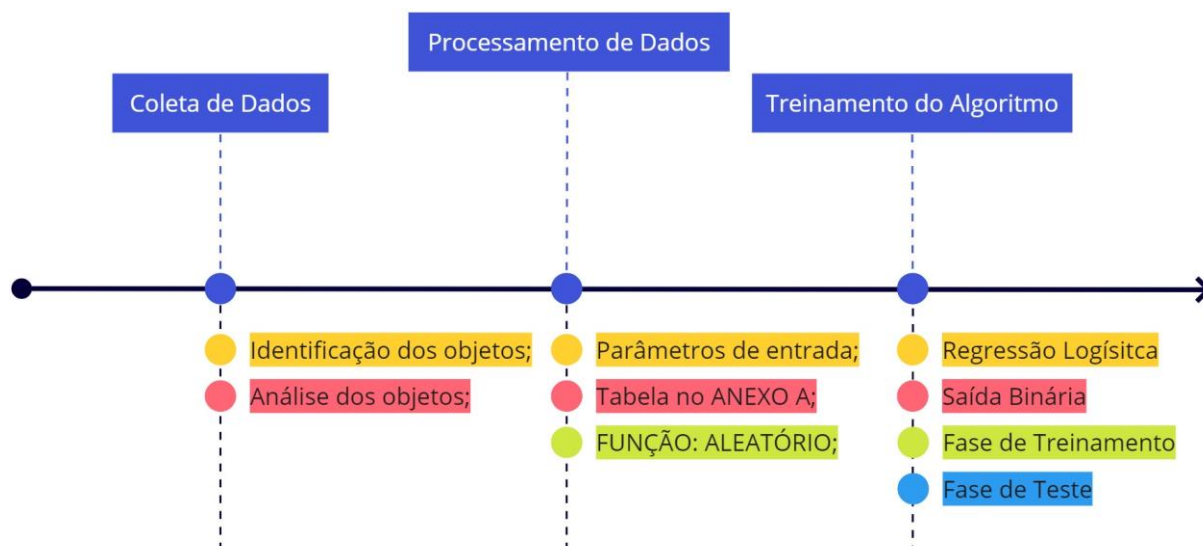
A partir da Equação (5) define-se que VP representa o VERDADEIRO POSITIVO, VN representa o VERDADEIRO NEGATIVO, X representa o número total de amostras destinadas para a precisão.

4. METODOLOGIA

O presente trabalho propõe um classificador do potencial à combustível de uma biomassa qualquer, a partir dos parâmetros: teor de umidade, materiais voláteis, teor de cinzas, carbono fixo e poder calorífico superior. Desse modo, desenvolveu-se uma ferramenta com base no algoritmo Regressão Logística (RL) capaz de interpretar o problema de classificação que contribua de forma eficaz ao estudo.

A metodologia consiste em três etapas fundamentais: Coleta de Dados, Processamento de Dados e Treinamento do Algoritmo. A Figura 3 ilustra a ordem que foram realizados os estágios bem como descreve as principais características de cada seção.

Figura 3 — Etapas presentes na metodologia com as respectivas especificações.



Fonte — Autora, 2023.

Adiante, será possível compreender melhor cada etapa, nas quais serão abordadas a seguir. É válido informar que foram adotados três cenários distintos para verificar qual das situações apresentadas melhor se encaixa como solução para a problemática deste trabalho.

4.1. Coleta de Dados

Na literatura é possível identificar diversos trabalhos que envolvem a Análise Imediata de diferentes biomassas, nas quais é possível extrair resultados favoráveis à biomassas com potencial à combustível. Nestes termos, iniciou-se uma vasta pesquisa bibliográfica para coleta de dados, catalogando todas as amostras que possuíam todos os parâmetros referentes à Análise Imediata, são eles: Teor de Umidade (TU), Teor de Materiais Voláteis (TMV), Teor de

Cinzas (TC) e Teor de Carbono Fixo (TCF) além do Poder Calorífico (PC). Foram coletadas 72 amostras. Salienta-se que, das 72 amostras coletadas, 64 experimentos foram originalmente produzidos pelo Grupo de Inovações Tecnológicas e Especialidades Químicas (GRINTEQUI/UFC) em seu respectivo laboratório, na cidade de Fortaleza-CE, o que viabilizou significativamente para a construção do acervo de dados.

Diante do contexto, no que compete ao acervo de dados, considerou-se apenas as biomassas que haviam todos os parâmetros citados, ou seja, não há dados incompletos. Importante comentar que dados ausentes influem negativamente no comportamento do algoritmo, e na resposta dele ao problema de classificação.

A Tabela 1 indica quais biomassas foram utilizadas para a coleta de dados, ou seja, inicia-se o processo de identificação dos objetos.

Tabela 1 — Biomassas utilizadas para composição do acervo.

BIOMASSAS	FONTE	LOCAL
CASCA DE CÔCO VERDE	(MALVEIRA, <i>et al</i> , 2017)	UFC - FORTALEZA
BAGAÇO DE CANA	(MALVEIRA, <i>et al</i> , 2017)	UFC - FORTALEZA
BAGAÇO DE CANA-DE-AÇÚCAR	(GIRÃO, 2016)	UFC - FORTALEZA
PALHA DE CARNAÚBA <i>IN NATURA</i>	(BIZERRA, 2017)	UFC - FORTALEZA
PALHA DE CARNAÚBA CARBONIZADA	(BIZERRA, 2017)	UFC - FORTALEZA
EUCALIPTO (<i>Eucalyptus globulus</i>)	(VASCONCELOS, 2017)	UFC - FORTALEZA
MARMELEIRO (<i>Croton sonderianus</i>)	(VASCONCELOS, 2017)	UFC - FORTALEZA
IPÊ ROXO (<i>Handroanthus impetiginosus</i>)	(VASCONCELOS, 2017)	UFC - FORTALEZA
PALHA DA CARNAÚBA	(LIMA, 2018)	UFC - FORTALEZA
TALO DA CARNAÚBA	(LIMA, 2018)	UFC - FORTALEZA
<i>Mimosa tenuiflora (Willd) Poir</i>	(GONDIM, 2018)	UFC - FORTALEZA
<i>Prosopis juliflora</i>	(GONDIM, 2018)	UFC - FORTALEZA
MOFUMBO	(OLIVEIRA, 2018)	UFC - FORTALEZA
ANGICO	(OLIVEIRA, 2018)	UFC - FORTALEZA
CASCA DE CASTANHA DE CAJU	(VIDAL, 2018)	UFC - FORTALEZA
BAGAÇO DA CANA-DE-AÇÚCAR	(VIDAL, 2018)	UFC - FORTALEZA
CASCA DE CASTANHA DE CAJU	(GADELHA, 2016)	UNILAB - CE
CASCA DE CÔCO BABAÇU	(GADELHA, 2016)	UNILAB - CE
CAVACOS	(GUERI, <i>et al</i> 2014)	UNICENTRO - PR
CASCAS DAS SEMENTES DA PLANTA NIM	(SIMÃO NETO, 2021)	UNILAB - CE
<i>Acacia auriculiformis</i>	(COSTA, <i>et al</i> 2009)	UFAM/INPA - AM

Fonte — Autora, 2023.

Nota-se que a Tabela 1 indica devidamente os autores que obtiveram os resultados para cada tipo de biomassa, além de identificar o local no qual foi realizada a pesquisa. Posteriormente, com os dados já armazenados, inicia-se a análise dos objetos de estudo. Verificou-se que houve uniformidade nos valores obtidos para cada parâmetro da Análise Imediata e Poder Calorífico, ou seja, se os dados estão normalizados. Este procedimento é necessário para avaliar se há algum objeto discrepante em relação aos demais objetos, já que

se um conjunto de objetos que, posteriormente, formar um padrão de normalização inadequada, poderá gerar resultados que não possuirão validação, o que prejudicará a classificação sugerida pelo algoritmo RL ao questionar a sua aplicabilidade.

4.2. Processamento de Dados

Na literatura, a obtenção de dados cujo o valor é baixo para potencial a combustível são escassos, sendo necessário inferir valores que sirvam de referência para esse cenário. Nesta perspectiva, na Seção 3.3. do presente trabalho é explicado os conceitos sobre os parâmetros indicativos à combustão e, com isso, traz a noção em termos de valores considerados como ótimo para cada parâmetro da Análise Imediata e Poder Calorífico, caracterizando um bom potencial à combustão. Resumidamente, para os teores: TU e TC, prefere-se resultados com teores baixos, referindo-se a valores mais próximos possíveis a 0%; para os teores: TMV e TCF além do PC, adotou-se resultados com teores elevados, referindo-se a valores mais próximos possíveis a 100%. Visto a ausência na literatura de objetos que não continham dados de Análise Imediata para biomassas com baixo potencial à combustível, foi necessário gerar dados nos quais, embasados no ponto ótimo para cada parâmetro, definem-se em uma situação onde se encaixam como ponto péssimo para combustão.

Por conseguinte, gerou-se valores aleatórios definidos como dados de baixo potencial à combustível e construiu-se uma tabela necessária à aprendizagem do algoritmo. Esta tabela está apresentada no ANEXO A. Nota-se que todos os autores que constam na Tabela 1 somam-se ao ANEXO A.

O critério de análise foi considerar valores extremos para cada parâmetro. Por exemplo, para TU, espera-se um resultado de porcentagem baixa, próximas a 0%, pois a umidade não é desejável para caracterizar o potencial energético de determinada Biomassa. Logo, para definir um baixo potencial para esse parâmetro, considerou-se resultados de porcentagem elevada, próximas a 100%, o mesmo critério foi adotado para os parâmetros: TMV, TC, TCF e PC.

Adiante, adequou-se o TU como parâmetro 1, o TMV como parâmetro 2, o TC, como parâmetro 3, o TCF como parâmetro 4 e, por fim, o PC como parâmetro 5. Desta forma, adotou-se faixas de atuação para cada parâmetro, conforme a Tabela 2:

Tabela 2 — Faixas de atuação inferidas para cada parâmetro.

PARÂMETROS	FAIXAS	FUNÇÃO
TU	(70-99)%	= ALEATÓRIO () * (70-99) +99
TMV	(10-50)%	= ALEATÓRIO () * (10-50) +50
TC	(30-99)%	= ALEATÓRIO () * (30-99) +99
CF	(60-100)%	= ALEATÓRIO () * (60-100) +100
CPS	(1-30) MJ/kg	= ALEATÓRIO () * (1-30) + 30

Fonte: Autora, 2023.

A Tabela 2 mostra as faixas inferidas, com possíveis respostas para biomassas que não possuem potencial a combustível, simultaneamente apresenta a FUNÇÃO ALEATÓRIO, na qual pode ser aplicada em uma planilha comum do Pacote Office, com a ferramenta Excel. A partir da faixa de atuação indicada, pode-se gerar números aleatórios que contemplem-na. Logo, a partir da aplicação, foi possível gerar mais 72 dados para serem considerados como baixo potencial à combustão, após isso, eles foram adicionados ao acervo de dados coletados, totalizando, enfim, 144 amostras.

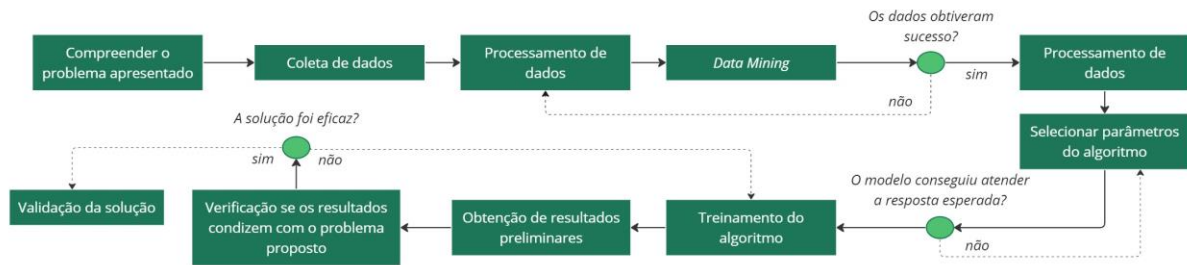
A saída do algoritmo para o problema é uma resposta de classe binária supervisionada, pois, a resposta desejada se resume em duas classes relacionadas a biomassa ter ou não potencial à combustível. Para melhor compreensão do algoritmo, define-se como valores de saída 1: para resultados que obtiveram resposta positiva para o problema de classificação, sendo ela a variável de sucesso; e de saída 0: para resultados que obtém resposta negativa para o problema de classificação, correspondendo a variável de insucesso.

4.3. Treinamento do Algoritmo

O treinamento do algoritmo inicia a partir dos dados tabelados, conforme o ANEXO A. Os objetos foram agrupados em uma matriz para que a rede regressiva, posteriormente, consiga identificá-los. Após isso, realiza-se o estudo sobre os as características do modelo de classificação ou de regressão que melhor compreenda o resultado para este trabalho.

Diante do contexto, a Figura 4 representa o fluxograma que contempla todas as etapas que constituem a funcionalidade do algoritmo, e, a partir dela, é possível analisar e compreender a ferramenta.

Figura 4 — Fluxograma com as etapas estabelecidas ao algoritmo.



Fonte — Autora, adaptado (Costa e Paixão, 2018).

A funcionalidade do algoritmo se dá a partir da abstração do problema em relação a seu contexto, isso possibilitará estabelecer qual será a resposta desejada. O presente trabalho deseja classificar o potencial energético de combustíveis que possuem como característica ser oriundos da bioenergia, isto implica em coletar os dados necessários e, posteriormente, verificar os parâmetros estabelecidos que proporcionaram esse rótulo.

Após tal verificação já realizada na Seção 4.2., inicia-se a exploração dos dados em busca de obter conhecimento acerca deles, pois trazem consigo a determinada informação valiosa e trivial sobre suas características, o que agrega o significado deles para o contexto.

Em continuidade, o algoritmo é sujeito a várias iterações, onde repete-se o processo x vezes em busca de convergir para a resposta esperada, por isso frisa-se que a escolha do modelo de Aprendizagem de Máquina por Regressão Logística (RL) do tipo supervisionado é uma excelente escolha para este trabalho, visto a quantidade baixa de dados coletados e a confiável excelência de sua respectiva obtenção. Caso a resposta seja afirmativa para o sucesso dos dados, o algoritmo prossegue para a próxima etapa, caso a resposta seja negativa, o algoritmo repete as etapas anteriores até buscar a convergência. Salienta-se que o ciclo de repetição está contido na fases de treinamento e é validada na fase de teste.

Acerca do algoritmo RL, o ponto ótimo dele se dá a partir do vetor gradiente, com o propósito de alcançar o valor mínimo da função, referente ao erro. Métodos de otimização são baseados no vetor gradiente, com o propósito de maximizar ou minimizar funções objetivo de várias variáveis. Noutros termos, para problemas de maximização por exemplo, tais métodos são processos iterativos que se baseiam na ideia de seguir a direção do gradiente da função, mas também levando em conta as direções já percorridas, ou seja, ele utiliza a informação sobre as direções anteriores para construir uma base orto-normal de direções, que é usada para

calcular a direção de descida mais rápida. Isso permite que o algoritmo evite cair em mínimos locais e chegue mais rapidamente ao mínimo global.

A partir disso, também utiliza-se a função Hessiana permitindo novas iterações, convergindo para o resultado esperado. Diante do contexto, a Hessiana é uma matriz de segunda derivada de uma função de várias variáveis. Ela é usada para descrever as propriedades de uma função no ponto de mínimo ou máximo. Esta função é definida como o Jacobiano da gradiente da função.

O algoritmo de ferramenta de Regressão Logística, foi separado com duas matrizes principais: a matriz de entrada, onde serão armazenados os valores referentes aos parâmetros das biomassas; e a que se refere somente a saída dessas biomassas, indicando 1 para biomassas que tem potencial à combustível e 0 para biomassas que não tem potencial à combustível. Além disso, definiu-se os limiares de classificação do algoritmo em termos das características conforme anteriormente explanado na Tabela 2. Por fim, a regularização do algoritmo consiste na possibilidade da ferramenta de criar uma região de decisão que pode ser uma curva bem próxima do esperado a partir dos dados obtidos nas fases de treinamento e de teste dada pela Equação (5).

$$\frac{VP + VN}{X} \times 100 \quad (5)$$

Por fim, definiu-se que a para melhor interpretar os resultados fornecidos pelo algoritmo, será utilizada matrizes de confusão, pois são adequadas há poucos recursos e dados de treinamento disponíveis, logo elas podem ser usadas para comparar as duas fases apresentadas e serem imprescindíveis para a escolha do cenário ideal como resposta da problemática.

5. RESULTADOS E DISCUSSÃO

Para analisar os resultados obtidos de forma eficaz, foram considerados três cenários distintos para a aplicabilidade do problema, estes cenários serão explicados a seguir.

5.1. CENÁRIO A

O primeiro cenário considera apenas **três variáveis** como parâmetros: TU, TMV e TC. A escolha dos parâmetros para este momento se deu a partir da análise das Equações (1), (2) e (3), verifica-se que para obter cada teor, as condições dependem diretamente dos valores de massa e, para o TMV, necessita-se do TU, por isso este parâmetro foi necessário. Diante do que foi dito, o primeiro cenário poderá ser identificado, ao longo do texto, como **cenário A**.

5.2. CENÁRIO B

Já ao segundo cenário adotou-se **quatro variáveis**, TU, TMV, TC e TCF. Neste caso, acrescenta-se mais uma variável para verificar o desempenho do algoritmo quando se introduz mais uma informação acerca dos dados. Diante do que foi dito, o segundo cenário poderá ser identificado, ao longo do texto, como **cenário B**.

5.3. CENÁRIO C

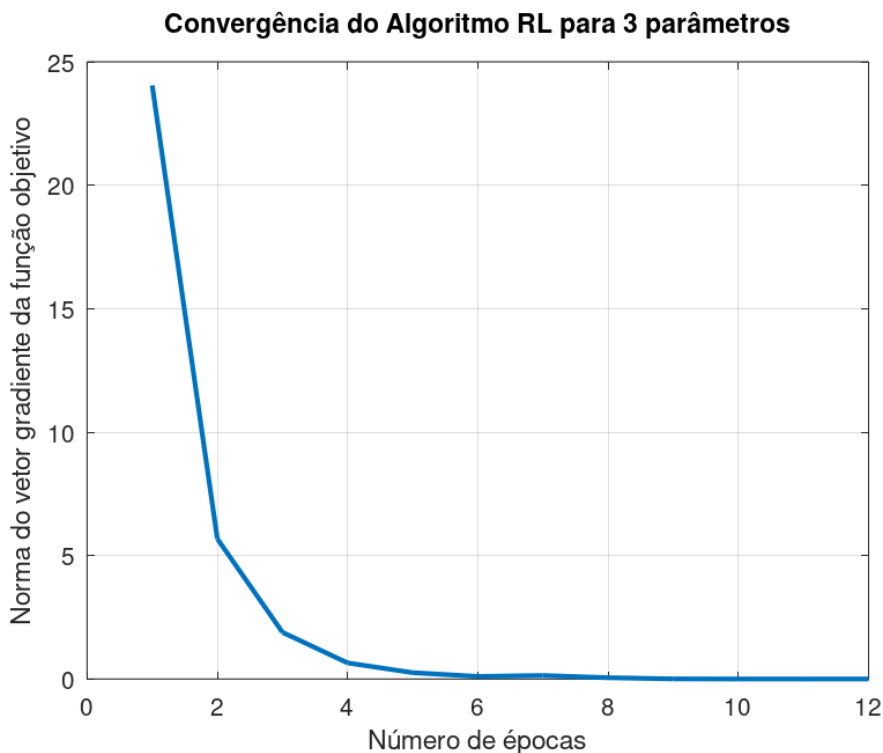
Por fim, o terceiro cenário contempla **cinco variáveis**, ou seja, estão incluídos TU, TMV, TC, TCF e PC. Para todos os três cenários foram considerados as 144 amostras coletadas, nas quais podem ser visualizadas no ANEXO A. Diante do que foi dito, o terceiro cenário poderá ser identificado, ao longo do texto, como **cenário C**.

Definiu-se, para três cenários que 70% das amostras destinam-se para a *fase de treinamento*, ou seja, 101 foram treinadas enquanto 30% restante foram destinadas para a *fase de teste*, ou seja, 43 amostras foram testadas, não havendo nenhum dado faltoso ou defeituoso.

Para cada cenário proposto foi plotado um gráfico, em termo do número de épocas que representa o comportamento do decrescimento da norma do vetor gradiente referente à função objetivo do problema, resultando no convergência do Algoritmo RL. Frisa-se que tais termos foram apresentados na Seção 3.4.3.6., na qual explica as características necessárias para a compreensão do algoritmo RL.

Após a plotagem do gráfico, insere-se matrizes de confusão que irão descrever a performance do modelo de classificação, indicando a frequência na qual a classe foi prevista correta ou incorretamente. Por fim, adiante foram descritos os resultados obtidos:

Gráfico 1 — Convergência do Algoritmo RL para 3 parâmetros.



Fonte — Autora, 2023.

De acordo com o Gráfico 1, percebe-se que ao longo do incremento das épocas, há um decrescimento da curva da Norma do vetor gradiente da função objetivo, até que ele se mantenha quase constante decaindo até zero. Subentende-se, a partir da análise do gráfico, que a RL iniciou a taxa de convergência entre 1 e 2 épocas, o que estabelece uma noção de tempo em que é necessário para a conversão total do algoritmo, concluindo a convergência e se tornando constante a partir de 12 épocas. Nota-se que necessitou-se de um número pequeno de épocas para que o algoritmo concluísse o procedimento, o que representa um fator positivo para o *software*, já que necessitou de pouco para conseguir se estabilizar e convergir.

Adiante, será analisado o comportamento do algoritmo durante suas fases de treinamento e de teste, para verificar a coerência do resultado de convergência em relação ao seu desenvolvimento. Para melhor explicar esse procedimento, estipula-se as matrizes de confusão que podem melhor representar as duas situações.

Figura 5 — Matriz de confusão sobre o resultado na fase de treinamento para 3 parâmetros.

		PREVISTA	
		SIM	NÃO
REAL	SIM	51	0
	NÃO	1	49

Fonte — Autora, 2023.

Nesta etapa, 101 objetos, escolhidos aleatoriamente pelo algoritmo, foram destinados para a fase de treinamento, A precisão do algoritmo foi de 99,01% na fase de treinamento, essa porcentagem foi obtida como relatório programável fornecido pelo algoritmo. Dessa forma, como resultado, a Figura 5 representa a obtenção de 51 objetos classificados como VERDADEIROS POSITIVOS, ou seja, ao conferir o resultado informado pelo algoritmo, verificou-se que a ferramenta conseguiu acertar 51 objetos que de fato possuíam o potencial energético de biocombustível. Dando prosseguimento, obteve-se 49 objetos classificados como VERDADEIROS NEGATIVOS, ou seja, ao analisar a saída obtida, verificou-se que a ferramenta acertou 49 objetos que não haviam o potencial energético a combustão. Ademais, o algoritmo identificou 1 objeto como FALSO POSITIVO, ou seja, na fase de treinamento, o algoritmo não obteve sucesso ao classificar esse objeto. Em outras palavras, a ferramenta classificou o dado como positivo, mas, ao conferir o gabarito, esta amostra na verdade não tinha potencial à combustão. Por fim, o algoritmo não identificou objetos classificados como FALSOS NEGATIVOS.

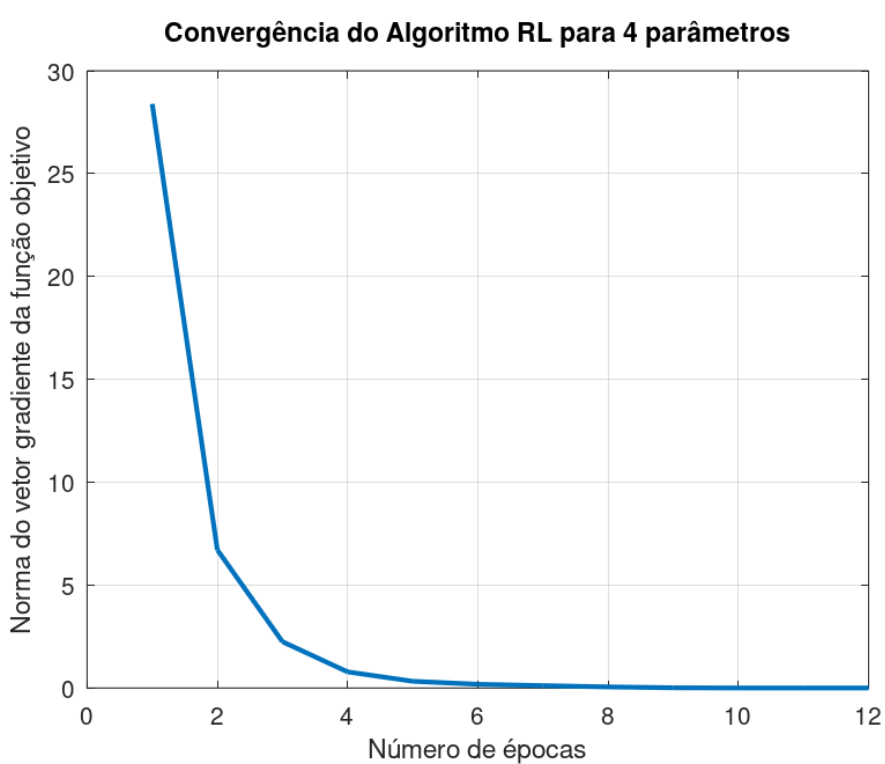
Figura 6 — Matriz de confusão sobre o resultado na fase de teste para 3 parâmetros.

		PREVISTA	
		SIM	NÃO
REAL	SIM	21	0
	NÃO	1	21

Fonte — Autora, 2023.

A compreensão acerca da matriz de confusão representada na Figura 4 é a mesma utilizada para explicar a Figura 6. Salienta-se que foram destinadas para essa etapa 43 objetos aleatórios a serem testados. Neste caso, a precisão do algoritmo na fase de teste foi de 97,67%. Nota-se que, em relação à fase de treinamento, a precisão na fase de teste para 3 parâmetros foi 1,34% inferior, no entanto esse resultado é plausível, visto que a quantidade de dados destinada a fase de treinamento supera o dobro de dados da fase de teste, logo, a precisão do algoritmo na fase atual se torna menor.

Gráfico 2 — Convergência do Algoritmo RL para 4 parâmetros.



Fonte — Autora, 2023.

A convergência representada no Gráfico 2 indica que o algoritmo levou 12 épocas para convergir, implicando no fato de que para 4 parâmetros, o algoritmo levou a mesma quantidade de tempo na convergência quando analisou-se o cenário A.

Figura 7 — Matriz de confusão sobre o resultado na fase de treinamento para 4 parâmetros.

		PREVISTA	
		SIM	NÃO
REAL	SIM	51	0
	NÃO	1	49

Fonte — Autora, 2023.

Ao realizar a análise exposta na Figura 7 para esse ciclo, nota-se que na fase de treinamento o algoritmo se comportou de forma similar em relação a situação com 3 parâmetros, pois nota-se que aqui também foi registrado 1 falso positivo, ou seja, apresentou um erro. Outrossim, a precisão do algoritmo foi de 99,01%, indicando uma taxa de convergência exatamente igual ao cenário A.

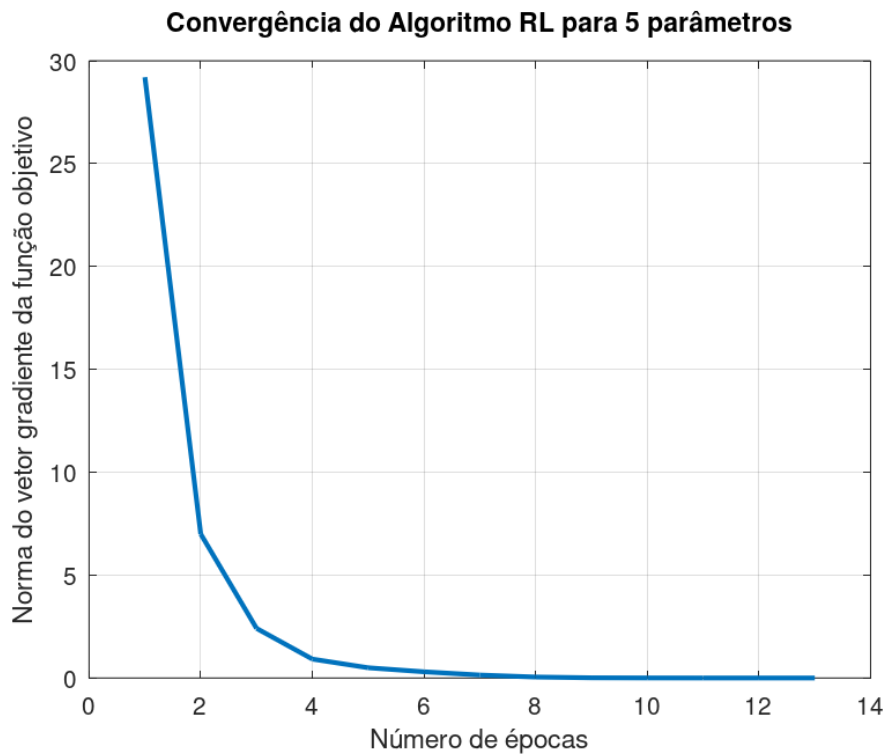
Figura 8 — Matriz de confusão sobre o resultado na fase de teste para 4 parâmetros.

		PREVISTA	
		SIM	NÃO
REAL	SIM	22	0
	NÃO	0	21

Fonte — Autora, 2023.

A Figura 8 indica a fase de teste para 4 parâmetros: TU, TMV, TC e TCF. Nota-se que o comportamento do algoritmo também foi similar ao cenário com 3 parâmetros, no que diz respeito ao FALSO NEGATIVO, entretanto não apresentou FALSO POSITIVO, o que indica uma taxa de convergência identificada de 100%, apontando que não houve erros para a fase de teste no cenário B.

Gráfico 3 — Convergência do Algoritmo RL para 5 parâmetros.



Fonte — Autora, 2023.

Por fim, a situação para 5 parâmetros é o cenário mais completo, no qual são consideradas todas as variáveis de entrada para o problema proposto. Percebe-se que a convergência do algoritmo inicia entre 1 a 2 épocas e se mantém constante a partir de 12 épocas. É importante lembrar que as épocas são o número de vezes que o algoritmo de aprendizado de máquina será executado sobre o conjunto de dados de treinamento.

Figura 9 — Matriz de confusão sobre o resultado na fase de treinamento para 5 parâmetros.

		PREVISTA	
		SIM	NÃO
REAL	SIM	52	0
	NÃO	0	49

Fonte — Autora, 2023.

Dando continuidade, o cenário C na fase de treinamento é representado pela Figura 9. Observa-se que o algoritmo, em relação aos cenários A e B, foi o único cenário que não apresentou erros, obtendo a precisão de convergência de 100%.

Figura 10 — Matriz de confusão sobre o resultado na fase de teste para 5 parâmetros.

		PREVISTA	
		SIM	NÃO
REAL	SIM	22	0
	NÃO	0	21

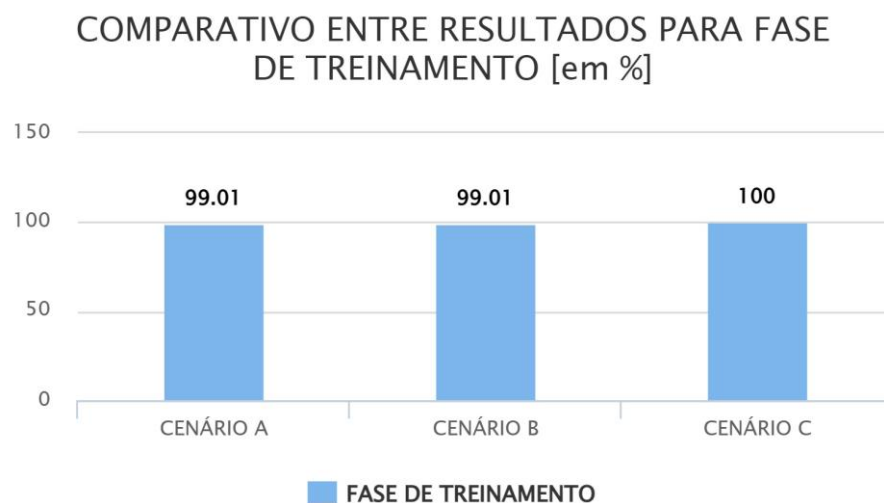
Fonte — Autora, 2023.

Finalmente, a Figura 10 define o cenário C na fase de teste, na qual obteve 100% de acerto já que não houve amostras classificadas como FALSO POSITIVO ou FALSO NEGATIVO, ou seja, a precisão do algoritmo foi de 100%.

5.3. SÍNTESE COMPARATIVA ENTRE OS CENÁRIOS A, B E C.

Ao realizar um levantamento geral é possível verificar, de forma mais sintética, a análise comparativa entre as precisões de convergência do algoritmo para os três cenários avaliados. O objetivo foi avaliar o desempenho do algoritmo ao inserir mais parâmetros para serem analisados, ou seja, verificar os resultados em um cenário com a menor quantidade de parâmetros, o cenário A, outro no qual representa uma melhora em relação ao cenário A no que acomete ao número de parâmetros, o cenário B, e por fim, o cenário mais completo e próximo do ideal para este estudo, nas condições estabelecidas, o cenário C.

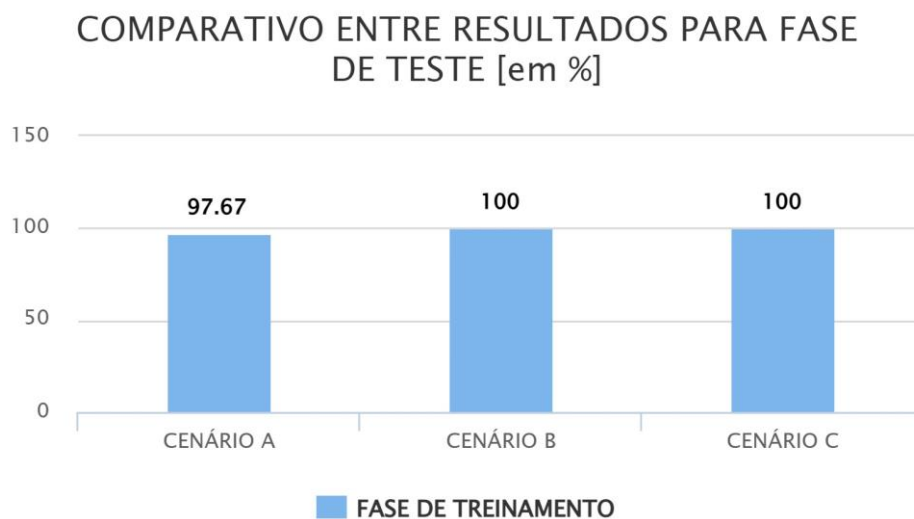
Gráfico 4 — Comparativo entre os resultados para a fase de treinamento [em %].



Fonte — Autora, 2023.

Dessa forma, percebe-se que o melhor cenário resultante é o cenário C, onde considera-se todas as cinco variáveis de entrada, única situação onde obteve-se 100% de acerto.

Gráfico 5 — Comparativo entre os resultados para a fase de teste [em %].



Fonte — Autora, 2023.

O Gráfico 5 representa resultados para os três cenários para a fase de teste, verifica-se que neste caso houve empate entre os cenários B e C, já o cenário A apresentou considerável desvantagem de 2,33%.

A principal interpretação da funcionalidade do algoritmo de se dá a partir dos resultados na fase de teste, pois este é o momento em que a ferramenta de Aprendizagem de Máquina valida o resultado esperado em uma situação onde ele desconhece a resposta obtida, para exemplificar, é similar a ação de realizar uma prova após ter adquirido conhecimento sobre o conteúdo. Entretanto, deve-se ressaltar a importância de ter coletado a precisão do algoritmo na fase de treinamento, pois é a etapa onde constata-se como o algoritmo se comportou durante a relação de acerto e erro, no processo de obter as informações acerca do problema de classificação e definir qual melhor solução, repetindo o processo, até obter a melhor saída possível.

Diante da apresentação dos resultados, definir qual cenário é o ideal não depende somente dos melhores saídas desejadas, deve-se considerar também a referência teórica para validar a coerência do que for esperado. Para este momento, será necessário comparar os cenários com a teoria adquirida acerca do problema e, conseqüentemente, será definido o cenário ideal.

É válido destacar a importância da utilização das matrizes de confusão para analisar os resultados obtidos. Isto se justifica a partir da perspectiva onde a precisão do algoritmo é mais relevante que a velocidade de convergência, pois estas matrizes avaliam em qual classe o algoritmo teve mais dificuldade de prever. Percebe-se que para os cenários A e B o algoritmo apresentou falha no FALSO POSITIVO, ou seja ele não pode compreender com excelência a informação acerca dos valores considerados como negativo, pois ele considerou como combustível uma amostra que na realidade foi informada que não era. Esta característica é compreensível a partir da metodologia adotada, pois os valores para os objetos indicados ao algoritmo que não haviam potencial à combustível foram inferidos na metodologia deste trabalho, ou seja, é totalmente plausível que a ferramenta RL não compreenda de forma eficaz justamente essa classe.

As normas NBR 16550, utilizada para a caracterização química, e a ASTM D3172, utilizada para determinar o TCF, são as principais bases que constituem a origem dos dados coletados. A partir delas é possível obter a metodologia aplicável para estabelecer valores de Análise Imediata para os parâmetros: Teor de Umidade, Teor de Materiais Voláteis, Teor de Cinzas, Teor de Carbono Fixo e, por fim, Poder Calorífico.

O cenário A utilizou como acervo de dados com base em três parâmetros: TU, TMV, TC e obteve resultados excelentes na fase de treinamento, 99,01%, na fase de teste, 97,67%.

Contudo, classificar o potencial energético de combustíveis oriundos da bioenergia com foco apenas nessas três características é inconsistente.

A partir dos teores indicados em uma situação real, as informações adquiridas são de: dificuldade em obter uma camada protetora sobre o combustível que atrapalhe a ignição à combustão, isso porque um baixo TU implica na aptidão das chamas aumentarem a taxa à combustão; facilidade de ignição à combustão, pois um alto TMV indica a facilidade do combustível de inflamar rapidamente; por fim, liberação dos fluxos de ar proporcionando a combustão completa, devido a um baixo TC que isso diminui a o consumo do combustível elevando sua eficiência.

Neste enquadramento, é necessário relembrar a essência da combustão que é a liberação do dióxido de carbono (CO_2), por isso, para classificar se um combustível pode ou não ter potencial energético é primordial ter conhecimento sobre o valor do TCF, pois esse parâmetro é justamente o carbono presente no combustível que é liberado pelo CO_2 durante a queima. Nesta circunstância, apesar do cenário A apresentar excelentes resultados, ele não é o ideal para a solução do problema.

De acordo com o contexto, verifica-se, *a priori*, que o cenário B se enquadra em um cenário favorável como resposta desejada, pois considera quatro parâmetros: TU, TMV, TC e TCF, constata-se que esta situação corrige a lacuna apontada para o cenário A. O cenário B obteve resultado similar ao cenário A na fase de treinamento, entretanto obteve resultado melhor na fase de teste em relação ao cenário A, no qual foi 2,33% superior, ressaltando-se que o cenário B obteve 100% precisão na convergência do algoritmo. Portanto, o cenário B é elegível como solução à saída desejada.

Finalmente, o cenário C contempla todos os cinco parâmetros avaliados, para esta situação é seguro afirmar que a precisão do algoritmo é mais confiável em relação aos cenários A e B. Este cenário obteve na fase de treinamento e na fase de teste resultado de 100% na precisão de convergência. Além da precisão do algoritmo, é imprescindível mencionar que este caso faz-se uso do PC como parâmetro fundamental como indicativo de bom combustível. Portanto, diante das justificativas apresentadas, o cenário C é a melhor saída desejada para a solução da problemática. Destaca-se que mesmo em uma possibilidade onde os valores de precisão sejam inferiores em relação ao cenário B, o cenário C prossegue com a preferência de saída desejada, porque tem mais parâmetros que auxiliam o algoritmo a obter maior precisão sobre informações do problema e pode compreendê-lo melhor. Contudo,

em uma situação onde se desconhece o PC, o cenário B é uma alternativa viável para resposta ao trabalho, tanto ao que acomete ao algoritmo quanto ao que acomete a referência teórica.

Por conseguinte, pode-se inferir que o algoritmo de Aprendizagem de Máquina utilizando a ferramenta Regressão Logística pode desempenhar um papel eficaz para solucionar a problemática, já que o algoritmo conseguiu ser preciso ao verificar os objetos destinados ao teste. No entanto, frisa-se que, na possibilidade da obtenção de uma base maior de dados e, principalmente, na abstração de dados que indicam baixo potencial energético à combustão, estima-se que a precisão do algoritmo será mais aceitável e poderá obter resultados mais concisos, mesmo que ocorra a sua taxa de convergência seja inferior a obtida nesse trabalho.

6. CONSIDERAÇÕES FINAIS

Este trabalho buscou uma forma de propor uma solução eficaz para análise de biomassas com potencial a combustível. Para isto, escolheu-se utilizar um algoritmo de Aprendizagem de Máquina, a Regressão Logística, como ferramenta de classificação para a saída desejada. Diante disto, as variáveis de entrada tiveram como base a Análise Imediata e Poder Calorífico das amostras referenciadas à trabalhos acadêmicos. A partir da metodologia aplicada, a ferramenta conseguiu ser eficaz na otimização do processo, obteve excelentes resultados para todos os cenários propostos.

Contudo, para trabalhos futuros, recomenda-se a inserção de mais dados para compor o acervo, visto que a quantidade de objetos analisados é trivial para a melhora do desempenho na precisão do algoritmo. Com isto, espera-se que o comportamento da ferramenta obtenha melhor resultado em relação à fase de treinamento e à fase de teste.

Dando continuidade, para dar prosseguimento nessa perspectiva e nessa linha de pesquisa, recomenda-se a inclusão da Análise Elementar como parâmetro incluso. A Análise Elementar é o processo de analisar e quantificar a massa, em porcentagem, que representa o elemento químico para uma amostra. Na qual são obtidos valores para Carbono (C), Hidrogênio (H), Nitrogênio (N), Enxofre (S) e Oxigênio (O) presentes em uma amostra. Por ser usada para verificar a pureza de uma substância, além de avaliar a qualidade de um combustível, e até detectar a presença de impurezas e contaminantes, a Análise Elementar e o Poder Calorífico podem ser considerados parâmetros ímpares para possibilitar mais informações sobre os objetos ao algoritmo. Portanto, dessa forma, o trabalho poderá ser aperfeiçoado e otimizado.

REFERÊNCIAS

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. Disponível em: <http://www2.aneel.gov.br/arquivos/pdf/atlas_par2_cap4.pdf>. Acesso em: 08 dez. 2022.

ALMEIDA, Adriano Pires de. (2013), **Perspectivas teóricas da integração energética na América do Sul**. O Atlântico Sul como eixo de inserção internacional do Brasil. 21-23 de maio de 2013. Seminário Brasileiro de Estudos estratégicos Internacionais (SEBREEI), Porto Alegre.

BARBOSA, Gabriela Gonçalves. Recursos Naturais Renováveis e Produção de Energia. **Revista Política Hoje** - 1a Edição - Volume 23 - p. 193-215.

Beckmann, M., Ebecken, N., e Lima, B. P. (2015) A KNN Undersampling Approach for Data Balancing, **Journal of Intelligent Learning Systems and Applications**, vol.7, pp. 104-116.

BIZERRA, DEIBY UCHOA. **AVALIAÇÃO DA PALHA DE CARNAÚBA IN NATURA E CARBONIZADA PARA PRODUÇÃO DE COMBUSTÍVEIS SÓLIDOS: Análise Imediata, Poder Calorífico e Densificação da Biomassa**. Orientador: Profa. Dra. Maria Alessandra de Sousa Rios. 2017. TESE (Engenharia Ambiental) - Universidade Federal do Ceará, [S. l.], 2017.

BONASSA, Gabriela et al. **Subprodutos Gerados na Produção de Bioetanol: Bagaço, Torta de Filtro, água de Lavagem e Palhagem**. Revista Brasileira de Energias Renováveis, v. 4, n. 3, 2015.

CARVALHO, Joaquim Francisco de. Energia e sociedade. **Estudos Avançados**, v. 28, n. 82, p. 25–39, 2014. Disponível em: <<https://www.scielo.br/j/ea/a/GYqnmvDvtGtr8N5zdsYdj5j/?lang=pt>>. Acesso em: 10 dez. 2022.

CASTRO, Leandro Nunes de; FERRARI, Daniel Gomes. **Introdução à Mineração de Dados**. Rio de Janeiro: Saraiva, 2016. 682 p.

CHEIN, Flávia. **Introdução aos modelos de regressão linear Metodologias COLEÇÃO.** [s.l.: s.n., s.d.] Brasília, 2019. Disponível em: https://repositorio.enap.gov.br/bitstream/1/4788/1/Livro_Regress%C3%A3o%20Linear.pdf.

COSTA. L.L. Jônntas; PAIXÃO. C. Flávio. **Sistema para Auxiliar a Detecção da Doença de Parkinson Interpretando Sinais de Tremor de Repouso Parkinsoniano Utilizando Aprendizado de Máquinas.** Orientadora: Dra. Lourdes Mattos Brasil. Coorientador: Bel. Ithallo Junior Alves Guimarães. 2018. TESE. Universidade de Brasília (UnB).

DINIZ, C.A. & LOUZADA NETO, F. **Data Mining: uma introdução.** São Carlos: Associação Brasileira de Estatística, 2000.

FARINHAQUE, R. 1981. **Influência da umidade no poder calorífico da madeira de bracatinga (Mimosa scrabella, Benth) e aspectos gerais de combustão.** Curitiba: fupef, Série Técnica. 14 pp.

FERREIRA, Italo Tadeu, et al. **Estimativa Do Potencial Energético De Resíduos Celulósicos De Fabricação De Papel Através De Análise Imediata.** Revista Brasileira De Energias Renováveis, vol. 3, no. 4, 2014, <https://doi.org/10.5380/rber.v3i4.38618>.

FERNANDES, Luiz Gustavo Leão. **Utilização de Redes Neurais na Análise e Previsão de Séries Temporais.** 1995. 77 f. Dissertação (Mestrado) - Curso de Ciências de Computação, UFRGS, Rio Grande do Sul, 1995.

FERNANDES, Fernando Timoteo; FILHO, Alexandre Dias Porto Chiavegatto. **Perspectivas do uso de mineração de dados e aprendizado de máquina em saúde e segurança no trabalho Tema livre/Ensaio,** Rev. bras. saúde ocup, 44, 2019.

GADELHA, ANTONIA MABRYSA. **O MARKETING VERDE DOS COMBUSTÍVEIS SÓLIDOS: ESTUDO DE RESÍDUOS AGRÍCOLAS, PRODUÇÃO DE BRIQUETES E APLICABILIDADE NO MAÇICO DE BATURITÉ-CE.** Orientador: Profa. Dra. Ada Amelia Sanders Lopes. 2016. TESE (Engenharia de Energias) - Universidade da Integração Internacional da Lusofonia Afro-Brasileira, [S. l.], 2016.

GIRÃO, LUIZ EDUARDO. **BAGAÇO DE CANA-DE-AÇÚCAR: INVESTIGAÇÃO DO PODER CALORÍFICO E POTENCIAL ENERGÉTICO**. Orientador: Profa. Dra. Maria Alessandra de Sousa Rios. 2016. TESE (Engenharia de Energias Renováveis) - Universidade Federal do Ceará, [S. l.], 2016.

GOLDEMBERG, José. Biomassa e energia. **Quím. Nova** 32 (3), 2009.

GONDIM, FABRICIO OLIVEIRA. **POTENCIAL ENERGÉTICO DAS ESPÉCIES *Mimosa tenuiflora* (Willd) Poir E *Prosopis juliflora***. Orientador: Profa. Dra. Maria Alessandra de Sousa Rios. 2018. TESE (Engenharia Mecânica) - Universidade Federal do Ceará, [S. l.], 2018.

Haykin, Simon. **Redes Neurais – Princípios e Práticas**. 2. ed. Porto Alegre: Bookman, 2001.

LARSON, R, FARBER, B. **Estatística Aplicada**. 2ª edição. São Paulo: Pearson - Prentice Hall, 2004.

LIMA, RAFAEL NOGUEIRA. **AVALIAÇÃO DO POTENCIAL ENERGÉTICO DA PALHA E TALO DA CARNAÚBA**. Orientador: Profa. Dra. Maria Alessandra de Sousa Rios. 2018. TESE (Engenharia Mecânica) - Universidade Federal do Ceará, [S. l.], 2018.

LUDERMIR, Teresa Bernarda. **Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências**. **Estudos Avançados**, v. 35, n. 101, p. 85–94, 2021. Disponível em: <<https://www.scielo.br/j/ea/a/wXBdv8yHBV9xHz8qG5RCgZd/>>. Acesso em: 16 jan. 2023.

Rokach, L., Maimon, O. (2005). **Clustering Methods**. In: Maimon, O., Rokach, L. (eds) *Data Mining and Knowledge Discovery Handbook*. Springer, Boston, MA. https://doi.org/10.1007/0-387-25465-X_15

LIRA, Larissa Duarte de Britto, et al. **Análise do algoritmo naive bayes na classificação de amostras do banco de dados hepatite**. Congresso Nacional de Pesquisa e Ensino em Ciências, 2019, 12p.

MAGEE, John. **Decision Making And Problem Solving**. Havard Business Review.1964.

MAHESH, Batta. **Machine Learning Algorithms - A Review**. International Journal of Science and Research (IJSR). Ed. 1. Vol. 9. 2020.

MASCARENHAS, Andresa. A proteção ambiental nas atividades de distribuição e revenda de combustíveis. Doi: 10.5020/2317-2150.2004.v09n1p54. **Pensar - Revista de Ciências Jurídicas**, v. 9, n. 1, p. 54–70, 2020. Disponível em: <<https://ojs.unifor.br/rpen/article/view/748>>. Acesso em: 16 dez. 2022.

MORELLO, Marcelo. Et al. **Biomassa para produção de Energia Sustentável**. Revista Científica Multidisciplinar Núcleo do Conhecimento. Ano 05, Ed. 10, Vol. 16, pp. 81-102. Outubro de 2020.

MUDANÇAS CLIMÁTICAS E TRANSIÇÃO ENERGÉTICA. Epe.gov.br. Disponível em: <<https://www.epe.gov.br/pt/abcdenergia/clima-e-energia>>. Acesso em: 16 jan. 2023.

MENDES, Carlos; VEGA. **TÉCNICAS DE REGRESSÃO LOGÍSTICA APLICADA À ANÁLISE AMBIENTAL**. n. 1, p. 5–30, 2011. Disponível em: <<https://www.lume.ufrgs.br/bitstream/handle/10183/238205/000835911.pdf?sequence=1>>. Acesso em: 12 jan. 2023.

NETO, FRANCISCO SIMÃO. **ANÁLISE DO POTENCIAL ENERGÉTICO DAS CASCAS DAS SEMENTES DA PLANTA NIM (*Azadirachta indica*)**. Orientador: Profa. Dra. Ada Amelia Sanders Lopes. 2016. TESE (Engenharia de Energias) - Universidade da Integração Internacional da Lusofonia Afro-Brasileira, [S. l.], 2016.

NIED, A. **Treinamento de redes neurais artificiais baseado em sistemas de estrutura variável com taxa de aprendizado adaptativa**. 2007. Belo Horizonte, MG. Tese (Doutorado) – Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais.

OLIVEIRA, HENRIQUE BELCHIOR. **ANÁLISE IMEDIATA E AVALIAÇÃO DO PODER CALORÍFICO SUPERIOR DAS ESPÉCIES *Combretum leprosum* Mart. (MOFUMBO) e *Anadenanthera macrocarpa* (Benth.) Brenan (ANGICO)**. Orientador:

Profa. Dra. Maria Alexandra de Sousa Rios. 2018. TESE (Engenharia Mecânica) - Universidade Federal do Ceará, [S. l.], 2018.

PONTE, M.R.; GADELHA, A.M.T.; MACHADO, Y.L., et al. **Blendas de bagaço de cana-de-açúcar, podas de mangueira e cajueiro: caracterização das propriedades e investigação de seus potenciais energéticos.** Revista Matéria, v.24, n.2, 2019

RIFKIN, Jeremy. (2011), **The third revolution. How lateral power is transforming energy, the economy and the world.** New York: Palgrave Macmillan.

SANTOS, Alcione Miranda dos; SEIXAS, José Manoel de; PEREIRA, Basílio de Bragança; *et al.* **Usando redes neurais artificiais e regressão logística na predição da Hepatite A.** Revista Brasileira de Epidemiologia, v. 8, n. 2, p. 117–126, 2005. Disponível em: <<https://www.scielo.br/j/rbepid/a/wpHxNfpjJz4k9VHBRrzgVfw/?lang=pt&format=html>>. Acesso em: 11 jan. 2023.

SANTOS, José. Recursos naturais brasileiros: **Centro de Estudos Estratégicos do Exército: Artigos Estratégicos**, v. 1, n. 1, p. 89–111, 2016. Disponível em: <<http://www.ebrevistas.eb.mil.br/CEEEExArE/article/view/1197>>. Acesso em: 25 nov. 2022.

SANTOS, Guilherme Henrique Fávero; NASCIMENTO, Raphael Santos; ALVES, Geziele Mucio. Biomassa como energia renovável no Brasil. **Revista Uningá Review**, vol.29, n.2, pp.06-13 (Jan-Mar 2017).

SILVA, Ivan Nunes da; SPATTI, Danilo Hernane; FLAUZINO, Rogério Andrade. **Redes Neurais Artificiais para engenharia e ciências aplicadas.** 2. ed. Universidade de São Paulo: Artliber, 2016.

S. M. Peres, T. Rocha, M. R. C. B. Bís caro, H. H., and C. Boscaroli. **Tutorial sobre fuzzy-c-means efuzzy learning vector quantizations: Abordagens híbridas para tarefas de agrupamento e classificação.** 19(1):120–163, 2012.

SOUSA, Natalia Bastos, et al. **Agrupamento e Classificação de Consumidores de Energia Rural Utilizando Random Forest e K-Nearest Neighbors.** Universidade Federal de Santa Maria, v. 2, 1337.

TOALMASQUIM, Maurício; GUERREIRO, Amilcar; GORINI, Ricardo. **Matriz energética brasileira: uma prospectiva**. Novos estudos, CEBRAP (79) • Nov 2007.

VASCONCELOS, RAFAEL BARBOSA. **ANÁLISE IMEDIATA E AVALIAÇÃO DO PODER CALORÍFICO DAS ESPÉCIES: EUCALIPTO (*Eucalyptus globulus*), MARMELEIRO (*Croton sonderianus*) e IPÊ ROXO (*Handroanthus impetiginosus*)**. Orientador: Profa. Dra. Maria Aleksandra de Sousa Rios. 2017. TESE (Engenharia Mecânica) - Universidade Federal do Ceará, [S. l.], 2017.

VIDAL, MARIANA DE SOUZA. **ESTUDO DO POTENCIAL ENERGÉTICO DE BRIQUETES PRODUZIDOS A PARTIR DE RESÍDUOS AGRÍCOLAS DE CASCA DE CASTANHA DE CAJU E BAGAÇO DA CANA-DE-AÇÚCAR**. Orientador: Profa. Dra. Maria Aleksandra de Sousa Rios. 2018. TESE (Engenharia Ambiental) - Universidade Federal do Ceará, [S. l.], 2018.

ANEXO A – DADOS COLETADOS

FONTE	Parâmetro 1	Parâmetro 2	Parâmetro 3	Parâmetro 4	Parâmetro 5	Parâmetro 6
	RU	MV	TC	CF	PCF	SAÍDA
(MALVEIRA, <i>et al</i> , 2017)	10.98	77.84	3.2	8.07	35.84	1
(MALVEIRA, <i>et al</i> , 2017)	11.35	75.76	6.8	6.09	33.31	1
(MALVEIRA, <i>et al</i> , 2017)	11.12	79.56	4.8	4.52	33.42	1
(MALVEIRA, <i>et al</i> , 2017)	10.99	78.96	3.7	6.35	34.78	1
(MALVEIRA, <i>et al</i> , 2017)	11.23	79.09	6.6	3.08	31.98	1
(GIRÃO, 2016)	8.99	70.61	17.48	11.89	36.15	1
(GIRÃO, 2016)	10.74	63.45	20.31	16.23	37.14	1
(GIRÃO, 2016)	7.32	67.12	18.15	14.72	37.26	1
(BIZERRA, 2017)	2.38	49.27	17.27	31.1	44.63	1
(BIZERRA, 2017)	4.54	44.05	15.73	35.68	46.63	1
(BIZERRA, 2017)	7.75	37.69	19.85	34.72	43.36	1
(BIZERRA, 2017)	6.77	33	27.1	33.13	40.12	1
(BIZERRA, 2017)	7.13	68.01	8.61	16.26	39.08	1
(VASCONCELOS, 2017)	12.13	87.94	1.09	10.97	42.13	1
(VASCONCELOS, 2017)	11.95	85.75	1.13	13.12	43.15	1
(VASCONCELOS, 2017)	10.04	85.19	1.71	13.1	42.91	1
(VASCONCELOS, 2017)	9.66	86.98	1.73	11.29	42.03	1
(VASCONCELOS, 2017)	9.76	90.33	2.16	7.51	40.04	1
(VASCONCELOS, 2017)	8.52	88.07	0.99	10.94	42.15	1
(VASCONCELOS, 2017)	9.68	90.47	1.51	8.02	40.54	1
(VASCONCELOS, 2017)	7.14	89.15	1.02	9.83	41.6	1
(VASCONCELOS, 2017)	9.16	86.66	2.28	11.06	41.7	1
(VASCONCELOS, 2017)	10.58	86.74	1.68	11.58	42.19	1
(VASCONCELOS, 2017)	10.22	87.63	1.96	10.41	41.52	1
(VASCONCELOS, 2017)	9.31	87.54	2.23	10.24	41.33	1
(LIMA, 2018)	11.76	70.89	9.13	19.98	43.33	1
(LIMA, 2018)	11.6	81.12	2.8	16.08	43.92	1
(LIMA, 2018)	11.45	73.7	7.3	19.04	43.61	1
(GONDIM, 2018)	8.68	81.96	1.21	16.83	44.91	1
(GONDIM, 2018)	12.23	79.98	1.06	18.96	45.99	1
(GONDIM, 2018)	13.32	87.31	1.19	11.5	42.34	1
(GONDIM, 2018)	10.22	98.31	1.38	0.3	36.86	1
(OLIVEIRA, 2018)	8.7	75.36	2.12	22.51	47.29	1
(OLIVEIRA, 2018)	12.85	82.27	2.67	15.05	43.47	1
(OLIVEIRA, 2018)	8.3	77.22	1.55	22.51	47.99	1
(OLIVEIRA, 2018)	9.1	73.49	2.69	22.51	46.58	1
(OLIVEIRA, 2018)	12.4	78.3	3.11	15.05	41.99	1
(OLIVEIRA, 2018)	13.3	86.25	2.23	15.05	44.96	1
(VIDAL, 2018)	10.01	77.79	3.57	18.44	44.69	1
(VIDAL, 2018)	11.02	75.31	1.58	23.11	47.79	1
(GADELHA, 2016)	8.53	76.9	8.29	14.81	41.16	1
(GADELHA, 2016)	8.97	78.81	11.77	9.42	37.2	1
(GADELHA, 2016)	8.88	79.66	11.47	8.87	37.05	1

FONTE	Parâmetro 1	Parâmetro 2	Parâmetro 3	Parâmetro 4	Parâmetro 5	Parâmetro 6
	RU	MV	TC	CF	PCF	SAÍDA
(GADELHA, 2016)	11.3	82	10.22	7.77	37.01	1
(GADELHA, 2016)	11	81.2	14.24	4.55	33.88	1
(GADELHA, 2016)	10.63	81.7	10.08	8.21	37.27	1
(GADELHA, 2016)	11.18	81.8	11.21	6.98	36.24	1
(GADELHA, 2016)	9.55	78.5	9.55	11.94	39.28	1
(GADELHA, 2016)	10.28	80.57	10.57	8.85	37.39	1
(GADELHA, 2016)	10.51	78.7	10.51	10.78	38.35	1
(GADELHA, 2016)	10.48	81.2	10.48	8.31	37.17	1
(GADELHA, 2016)	9.89	77	11.38	11.61	38.41	1
(GADELHA, 2016)	9.65	77.4	14.32	8.27	35.64	1
(GADELHA, 2016)	9.29	78.7	12.33	8.96	36.76	1
(GADELHA, 2016)	9.49	77.3	10.97	11.72	38.62	1
(GADELHA, 2016)	10.82	79.1	9.53	11.36	39.01	1
(GADELHA, 2016)	10.78	75.6	9.41	14.99	40.81	1
(GADELHA, 2016)	10.81	77.9	8.97	13.13	40.09	1
(GADELHA, 2016)	10.74	77.2	8.85	13.94	40.52	1
(GADELHA, 2016)	55	84.2	1	14.8	44.01	1
(GUERI, et al 2014)	55	71.6	3.4	25	47.99	1
(GUERI, et al 2014)	55	68.1	10.2	21.7	43.74	1
(GUERI, et al 2014)	80	82.5	6.1	11.5	40.46	1
(GUERI, et al 2014)	55	68.3	11.5	20.2	42.51	1
(GUERI, et al 2014)	10.83	12.6	29	58.4	54.09	1
(GUERI, et al 2014)	12.68	81.76	3.8	14.44	42.74	1
(SIMÃO NETO, 2021)	10.4	92.59	7.41	24.82	55.58	1
(SIMÃO NETO, 2021)	7.33	82.29	0.46	17.25	45.4	1
(SIMÃO NETO, 2021)	10.89	80.73	0.55	8.07	36.97	1
(SIMÃO NETO, 2021)	11.36	81.36	2.66	6.09	35.47	1
(SIMÃO NETO, 2021)	11.34	30.42	0.48	69.01	70.34	1
(SIMÃO NETO, 2021)	7.59	84	0.59	15.41	44.46	1
(COSTA, et al 2009)	8.98	79.17	1.35	19.48	46.13	1
(COSTA, et al 2009)	76.74	17.89	81.67	63.79	22.2	1
AUTORA, 2023	76.04	30.75	78.83	61.67	9.49	0
AUTORA, 2023	89.41	17.27	76.95	82.84	11.32	0
AUTORA, 2023	80.66	39.38	53.97	68.08	29.16	0
AUTORA, 2023	73.68	16.69	78.22	60.83	8.78	0
AUTORA, 2023	89.61	32.27	94.24	89.41	15.68	0
AUTORA, 2023	91.38	15.47	37.96	83.45	6.56	0
AUTORA, 2023	77.37	23.18	36.75	99.57	12.11	0
AUTORA, 2023	74.7	46.4	44.15	73.4	13.02	0
AUTORA, 2023	90.22	36.62	46.04	78.24	14.72	0
AUTORA, 2023	82.78	29.81	40.6	87.67	24.15	0
AUTORA, 2023	96.91	10.65	61.52	77.88	16.7	0
AUTORA, 2023	85.8	27.82	67.45	64.08	11.4	0

FONTE	Parâmetro 1	Parâmetro 2	Parâmetro 3	Parâmetro 4	Parâmetro 5	Parâmetro 6
	RU	MV	TC	CF	PCF	SAÍDA
AUTORA, 2023	88.81	31.29	33.37	83.76	11.45	0
AUTORA, 2023	98.1	35.37	41.91	99.46	23.39	0
AUTORA, 2023	82.51	22.95	67.89	92.76	28.47	0
AUTORA, 2023	84.4	15.28	55.05	72.8	28.53	0
AUTORA, 2023	75.78	47.75	67.33	65.73	29.92	0
AUTORA, 2023	87.86	21.54	80.59	68.33	12.25	0
AUTORA, 2023	85.17	33.7	85.09	96.79	23.99	0
AUTORA, 2023	79.7	49.51	59.74	77.79	25.27	0
AUTORA, 2023	90.79	29.99	95.09	74.02	29.88	0
AUTORA, 2023	78.33	41.56	43.74	76.6	11.68	0
AUTORA, 2023	70.34	14.91	83.74	89.5	12.63	0
AUTORA, 2023	77.6	31.98	52.62	65.13	13.72	0
AUTORA, 2023	86.61	20.42	84.14	82.35	19.79	0
AUTORA, 2023	89.67	10.58	65.03	91.3	19.59	0
AUTORA, 2023	87.24	12.37	43.05	65.99	20.67	0
AUTORA, 2023	72.41	11.6	91.19	93.27	14.97	0
AUTORA, 2023	72.01	48.38	47.86	78.85	16.34	0
AUTORA, 2023	91.33	45.22	31.52	66.61	19.36	0
AUTORA, 2023	95.17	20.66	81	94.51	14.62	0
AUTORA, 2023	82.02	33.23	39.23	97.54	8.9	0
AUTORA, 2023	94.91	38.69	58.48	99.58	5.76	0
AUTORA, 2023	80.63	40.27	73.63	95.99	16.09	0
AUTORA, 2023	89.77	39.21	81.33	93.41	29.58	0
AUTORA, 2023	71.45	14.64	79.7	77.81	28.37	0
AUTORA, 2023	79.07	18.08	43.2	97.76	29.29	0
AUTORA, 2023	96.64	19.69	45.01	65.5	28.57	0
AUTORA, 2023	83.36	31.04	78.13	74.32	21.59	0
AUTORA, 2023	96.75	34.4	48.94	68.77	23.17	0
AUTORA, 2023	80.78	46.62	79.1	76.13	3.93	0
AUTORA, 2023	91.05	19.36	47.97	90.73	2.92	0
AUTORA, 2023	85.97	14.06	32.86	82.07	12.78	0
AUTORA, 2023	88.11	47.75	81.63	67.6	4.87	0
AUTORA, 2023	97.63	39.49	60.44	93.76	14.83	0
AUTORA, 2023	70.97	18.13	84.35	89.08	6.4	0
AUTORA, 2023	71.14	32.72	82.25	75.33	1.59	0
AUTORA, 2023	89.9	41	79.37	87.98	5.54	0
AUTORA, 2023	98.58	11.22	45.99	70.91	26.36	0
AUTORA, 2023	93.55	11.3	93.79	67.97	8.18	0
AUTORA, 2023	95.31	43.84	50.32	73.75	13.32	0
AUTORA, 2023	91.91	27.24	76.07	81.9	20.56	0
AUTORA, 2023	88.34	30.06	57.29	76.97	10.74	0
AUTORA, 2023	82.1	37.55	42.01	96.4	3.51	0
AUTORA, 2023	82.24	45.23	77.83	87.71	2.85	0

FONTE	Parâmetro 1	Parâmetro 2	Parâmetro 3	Parâmetro 4	Parâmetro 5	Parâmetro 6
	RU	MV	TC	CF	PCF	SAÍDA
AUTORA, 2023	84.52	36.78	75.89	62.42	10.47	0
AUTORA, 2023	75.43	34.56	45.57	76.25	20.63	0
AUTORA, 2023	90.69	24.99	35.57	73.33	24.68	0
AUTORA, 2023	97.61	39.73	70.34	86.54	19.77	0
AUTORA, 2023	82.83	45.66	45.97	90.92	19.55	0
AUTORA, 2023	86.48	37.43	67.19	85.22	13.23	0
AUTORA, 2023	73.97	36.85	43.82	70.85	29.13	0
AUTORA, 2023	91.2	19.64	32.53	68.75	3.7	0
AUTORA, 2023	70.93	13.72	39.05	61.05	9.45	0
AUTORA, 2023	75.53	43.74	96.71	82.36	21.24	0
AUTORA, 2023	78.19	29.31	49.56	93.7	21.04	0
AUTORA, 2023	91.59	37.57	96.9	94.89	28.22	0
AUTORA, 2023	73.57	31.2	73.53	99.7	16.42	0
AUTORA, 2023	80.57	40.86	87.72	87.67	8.26	0
AUTORA, 2023	96.17	26.02	49.4	78.75	4.42	0

Fonte — Autora, 2023.