



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de Rosana

CHRISTIAN CALANDRIN
PEDRO MARK BIANCHI

**IMPLEMENTAÇÃO DE UM MODELO DE APRENDIZADO DE
MÁQUINA APLICADO PARA PREVISÃO DO CONSUMO DE
ENERGIA ELÉTRICA NA REGIÃO NORDESTE DO BRASIL**

Rosana

2023

CHRISTIAN CALANDRIN
PEDRO MARK BIANCHI

**IMPLEMENTAÇÃO DE UM MODELO DE APRENDIZADO DE
MÁQUINA APLICADO PARA PREVISÃO DO CONSUMO DE
ENERGIA ELÉTRICA NA REGIÃO NORDESTE DO BRASIL**

Trabalho de Conclusão de Curso (TCC)
apresentado ao Conselho de Curso de
Engenharia de Energia, da Faculdade de Ciências
e Engenharia, da Universidade Estadual Paulista
– UNESP - Câmpus de Rosana, para obtenção do
título de Bacharel em Engenharia de Energia.

Orientador: Prof. Dr. Kleber Rocha de Oliveira

Rosana
2023

C142i Calandrin, Christian
Implementação de um modelo de aprendizado de máquina aplicado para previsão do consumo de energia elétrica na região Nordeste do Brasil / Christian Calandrin; Pedro Mark Bianchi. -- Rosana, 2023

106 p.

Trabalho de conclusão de curso (Bacharelado - Engenharia de Energia) - Universidade Estadual Paulista (Unesp), Faculdade de Engenharia e Ciências, Rosana
Orientador: Kleber Rocha de Oliveira

1. Energia. 2. Machine Learning. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Engenharia e Ciências, Rosana. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

CHRISTIAN CALANDRIN
PEDRO MARK BIANCHI

**IMPLEMENTAÇÃO DE UM MODELO DE APRENDIZADO DE
MÁQUINA APLICADO PARA PREVISÃO DO CONSUMO DE
ENERGIA ELÉTRICA NA REGIÃO NORDESTE DO BRASIL**

Trabalho de Conclusão de Curso (TCC) apresentado ao Conselho de Curso de Engenharia de Energia, da Faculdade de Engenharia de Ciências, da Universidade Estadual Paulista – UNESP - Câmpus de Rosana, para obtenção do título de Bacharel em Engenharia de Energia.

Banca Examinadora

Prof. Dr. Kleber Rocha de Oliveira
UNESP – Câmpus de Rosana
Orientador

Prof. Dr. Leonardo Henrique Faria Macedo Possagnolo
UNESP – Câmpus de Rosana

Prof. Dr. Guilherme Pina Cardim
UNESP – Câmpus de Rosana

Rosana, 13 de novembro de 2023

Àqueles que, apesar de todas as adversidades, sempre acreditaram que eu seria capaz de atingir meus objetivos.

AGRADECIMENTOS

Ao concluir esta etapa fundamental de minha jornada acadêmica, gostaria de expressar minha profunda gratidão a todos que desenvolveram para o sucesso deste Trabalho de Conclusão de Curso. Esta conquista não seria possível sem o apoio constante e importante de pessoas especiais em minha vida.

À minha Família agradeço imensamente aos meus pais, Liliane e Sérgio, cujo amor incondicional e encorajamento foram os alicerces desta jornada. Sua dedicação em mim proporciona as melhores oportunidades e em me inspira a buscar o conhecimento que é fundamental. Agradeço também aos meus irmãos, Nicolas e Caio e demais familiares pelo apoio constante e compreensão nos momentos difíceis.

À UNESP Rosana minha sincera gratidão à Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP, Campus de Rosana, pela excelência acadêmica e pelos recursos oferecidos. Os professores e funcionários desempenharam papéis cruciais na minha formação, proporcionando um ambiente educacional estimulante e enriquecedor. Também obtivemos recursos e infraestrutura que tornaram possível a realização deste trabalho.

Ao meu Orientador, Prof. Dr. Kleber Rocha de Oliveira, um agradecimento especial ao meu orientador, cuja orientação sábia e apoio incansável foram fundamentais para o desenvolvimento deste trabalho. Suas contribuições e *insights* foram inestimáveis, guiando-me através dos desafios acadêmicos e incentivando meu crescimento profissional.

Aos Amigos e Colegas às amizades cultivadas ao longo dessa jornada, meu agradecimento sincero. As trocas de conhecimento, o apoio mútuo e as experiências compartilhadas enriqueceram significativamente minha trajetória acadêmica. Este é um momento de celebração e reconhecimento da importância de cada pessoa que fez parte desta jornada. A todos que, de alguma forma, desenvolvi para este capítulo da minha vida, meus mais profundos agradecimentos. Que esta conquista seja compartilhada por todos nós.

Muito obrigado.

Christian Calandrin

AGRADECIMENTOS

Com grande gratidão, encerro esta etapa crucial da minha jornada acadêmica, reconhecendo o papel fundamental de todos aqueles que contribuíram para o êxito deste Trabalho de Conclusão de Curso. Primeiramente agradeço a minha mãe, Erica Fernanda da Silva, que lutou para que eu pudesse me formar. Quero agradecer também ao meu pai, Wanderlei Marques Bianchi, que nunca deixou de me apoiar em meus estudos.

Gostaria de Agradecer também a Universidade Estadual Paulista “Júlio de Mesquita Filho”, que fiz parte por cinco anos, e aos professores e colegas, agradeço pela oportunidade de aprendizado e crescimento acadêmico. As experiências compartilhadas e os desafios superados foram fundamentais para o desenvolvimento deste trabalho. Também agradeço a FAPESP e CNPq pelos recursos disponibilizados durante minhas pesquisas.

Reservo um agradecimento especial aos meus orientadores de Iniciação Científica Prof. Dr. Wallace Casaca e Prof. Dr. Marilaine Colnago eeo orientador do presente trabalho, Prof. Dr Kleber Rocha de Oliveira, pela orientação sábia, paciência e pela dedicação em orientar durante toda minha jornada acadêmica. Suas contribuições foram inestimáveis.

Também gostaria de estender meus agradecimentos aos amigos e colegas que estiveram ao meu lado, oferecendo apoio e compartilhando conhecimento.

Por fim, agradeço a todos que, de alguma forma, contribuíram para o sucesso deste trabalho. Que este trabalho seja apenas o início de uma grande jornada.

Muito obrigado por tudo.

Pedro Mark Bianchi

“Uma mente que se expande para uma nova ideia, jamais retornará para o seu tamanho inicial.”

- Albert Einstein

RESUMO

Por ser um grande consumidor de energia elétrica, o Brasil sempre visa a maior diversificação da matriz energética nacional, a fim de não ser dependente apenas de uma só fonte. Diante disso, desenvolver métodos cada vez mais eficazes para prever o consumo elétrico do país ou região é fundamental para ajudar as agências governamentais, geradoras e concessionárias do setor elétrico a criar um planejamento mais eficiente dos custos de geração de eletricidade para atender às demandas imediatas, além de ser de grande importância para estabelecer um horizonte de consumo mais confiável para seus consumidores. Diante deste panorama, o projeto visa a implementação de um modelo de *machine learning* (ML) que possa prever o consumo de energia elétrica futuro de uma determinada região, neste caso a região Nordeste. Baseando-se em bases de dados acessíveis de anos anteriores da Operador Nacional do Sistema Elétrico (ONS) e do Instituto Nacional de Meteorologia (INMET), pode-se desenvolver este modelo de aprendizagem de máquina preditiva, visando ter uma estimativa do consumo futuro na região. De acordo com os resultados, foi possível construir um modelo de *gradient boosting* que resultou em previsões precisas, com MAPE de 2,23% no ano de 2022.

Palavras-chave: Análise Preditiva; Consumo de Energia Elétrica; *Machine Learning*.

ABSTRACT

As a major consumer of electrical energy, Brazil always aims for greater diversification of the national energy matrix, in order not to be dependent on a single source. Therefore, developing increasingly effective methods to predict the country's or region's electrical consumption is crucial to assist governmental agencies, power generators, and electric sector concessionaires in creating a more efficient plan for electricity generation costs to meet immediate demands. Additionally, it is of great importance to establish a more reliable consumption forecast for consumers. In this context, the project aims to implement a machine learning model capable of predicting the future electricity consumption of a specific region, in this case, the Northeast region. By relying on accessible databases from previous years provided by the Operador Nacional do Sistema Elétrico (ONS) and the Instituto Nacional de Meteorologia (INMET), this predictive machine learning model can be developed to estimate future consumption in the region. According to the results, it was possible to build a gradient boosting model that yielded accurate predictions, with a MAPE of 2.23% in the year 2022.

Keywords: Electricity Consumption Forecast; Machine Learning; Predictive Analysis.

LISTA DE FIGURAS

Figura 1 - Emissões de CO2 Mundial.....	24
Figura 2 - Fornecimento de Energia Mundial por Diferentes Fontes.....	25
Figura 3 - Comparativo entre Consumo de Energia	26
Figura 4 - Consumo de Energia por Habitante em Cada País.....	27
Figura 5 - Comparativo entre IDHs.....	27
Figura 6 - Comparativo entre PIBs	30
Figura 7 - Histórico e Expectativa do Consumo de Energia Mundial	31
Figura 8 - Capacidade Instalada Global de Energia Solar Fotovoltaica	32
Figura 9 - Capacidade Instalada Global de Energia Eólica	34
Figura 10 - Oferta Interna de Energia Por Fontes	35
Figura 11 - Emissões de CO2 no Brasil	36
Figura 12 - Comparativo de Emissões de CO2 entre o Mundo e o Brasil.....	37
Figura 13 - Geração de Energia no Brasil	38
Figura 14 - Histórico de Consumo de Energia no Brasil	39
Figura 15 - Histórico Perdas Técnicas no SDEE	39
Figura 16 - Histórico de Perdas Não Técnicas no Brasil	40
Figura 17 - Comparativo entre Produção, Consumo, Perdas Técnicas e Não Técnicas	40
Figura 18 - Participação Setorial no Consumo de Energia	42
Figura 19 - Crescimento do Consumo de Energia no setor Industrial.....	42
Figura 20 - Evolução do PIB no Brasil.....	44
Figura 21 - Comparativo entre Produção e Consumo de Energia no Brasil.....	45
Figura 22 - Crescimento da Energia Solar no Brasil.....	46
Figura 23 - Produção de Energia Eólica no Brasil	48
Figura 24 - Produção Anual de Energia Térmica por Área de Coletor Solar.....	51
Figura 25 - Mapa do Potencial de Geração de Energia Solar Fotovoltaica.....	52
Figura 26 - Densidade de Potência para 10 (Dez) metros.....	53
Figura 27 - Densidade de Potência para 50 (Cinquenta) metros	54
Figura 28 - Densidade de Potência para 100 (Cem) metros.....	54
Figura 29 - Densidade de Potência para 150 (Cento e Cinquenta) metros.....	55
Figura 30 - Densidade de Potência para 200 (Duzentos) metros	55
Figura 31 - Frequência de Ocorrência de Diferentes Intervalo	60

Figura 32 - Exemplo de Boxplot	62
Figura 33 - Exemplo de Gráfico Quantil-Quantil	63
Figura 34 - Fluxograma da Metodologia Implementado no Projeto	76
Figura 35 - Mapa do Sistema Interligado Nacional	79
Figura 36 - Correlação de Pearson para Variável Target	81
Figura 37 - Histograma da Variável Target.....	82
Figura 38 - BoxPlot da Variável Target.....	83
Figura 39 - Q-Q Plot da Variável Target.....	83
Figura 40 - Distribuição da Variável Carga Durante as Semanas (2011 a 2022)	84
Figura 41 - Distribuição da Variável Carga Durante os meses (2011 a 2022)	85
Figura 42 - Variáveis mais importantes para o GB	89
Figura 43 - Gráfico das Predições para o Ano de 2022	90
Figura 44 - Gráfico das Predições feitas pelo GB no Ano de 2022.....	91
Figura 45 - Resíduo (GWh) das Predições do Modelo GB no Ano de 2022	93

LISTA DE TABELAS

Tabela 1 - Aplicação Dummy Coding	68
Tabela 2 - Variáveis Dummy	68
Tabela 3 - Análise Estatística Descritiva da Variável Target.....	80
Tabela 4 - MAPE (%) Média Mensal da Carga Prevista para o Ano de 2022	87
Tabela 5 - MAE (GWh), MSE (GWh ²) e MAPE (%) ao longo do Ano de (2022).	88

LISTA DE ABREVIATURAS E SIGLAS

IOT	<i>Internet of Things</i>
IA	Inteligência Artificial
GEEs	Gases de Efeito Estufa
IEA	<i>International Energy Agency.</i>
IDH	Índice de Desenvolvimento Humano
ONU	Organização das Nações Unidas
PIB	Produto interno Bruto
P&D	Pesquisa e Desenvolvimento
BEN	Balanco Energético Nacional
PTs	Perdas Técnicas
PNTs	Perdas Não Técnicas
SDEE	Sistema de Distribuição de Energia Elétrica
EPE	Empresa de Pesquisa Energéticas
IBGE	Instituto Brasileiro de Geografia e Estatística
ABSOLAR	Associação Brasileira de Energia Solar
ANEEL	Agência Nacional de Energia Elétrica
ABEEÓLICA	Associação Brasileira de Energia Eólica
INPE	Instituto de Pesquisas Espaciais
GWA	<i>Global Wind Atlas</i>
KDD	<i>Knowledge Data Discovery</i>
ML	<i>Machine Learning</i>
MMS	Média Móvel Simples
MME	Média Móvel Exponencial
SVM	Máquina de Vetores de Suporte
RNAs	Redes Neurais Artificiais

MAE	Erro Médio Absoluto
MSE	Erro Quadrático Médio
MAPE	Erro Percentual Absoluto Médio
ONS	Operador Nacional do Sistema
INMET	Instituto Nacional de Meteorologia
SIN	Sistema Interligado Nacional
AED	Análise Exploratória de Dados

LISTA DE UNIDADES

TWh	Tera watt hora
TW	Tera watt
W/mês	Watt por mês
kW	Kilo watt
GW	Giga watt
GWh	Giga watt hora
mm	Milímetros
mB	Mili Bar
°C	Celsius
m/s	Metros por segundo

SUMÁRIO

1	INTRODUÇÃO	19
2	REVISÃO BIBLIOGRÁFICA	22
2.1	PANORAMA A NÍVEL INTERNACIONAL	22
2.1.1	O Problema das Mudanças Climáticas	22
2.1.2	O Cenário Energético Internacional	24
2.1.3	O Consumo de Energia e sua Relação com os Habitantes	26
2.1.4	Correlação do PIB e Fatores Econômicos com a Energia	29
2.1.5	A Matriz Energética Mundial	30
2.1.6	Cenário da Energia Solar Mundial	32
2.1.7	Cenário da Energia Eólica Mundial	33
2.2	PANORAMA A NÍVEL NACIONAL	34
2.2.1	Comparativo do Cenário Brasileiro com o Cenário Mundial	35
2.2.2	Sistema de Distribuição de Energia Elétrica Brasileira (SDEE)	37
2.2.3	A Matriz Energética Brasileira	41
2.2.4	Relação do PIB com a Energia	43
2.2.5	Cenário da Energia Solar no Brasil	45
2.2.6	Cenário da Energia Eólica no Brasil	47
2.3	PANORAMA A NÍVEL DA REGIÃO NORDESTE	49
2.3.1	A Região Nordeste	49
2.3.2	Classificação da Matriz Energética Nordestina	49
2.3.3	Setor Energético Solar no Nordeste	50
2.3.4	Setor Energético Eólico no Nordeste	53
2.4	TRABALHOS RELACIONADOS	56
3.0	KNOWLEDGE DATA DISCOVERY	57
3.1	Estatística Descritiva	58
3.2	Coeficiente de Variação	59
3.3	Histograma	59
3.4	Coeficiente de Correlação	61
3.5	BoxPlot	61
3.6	Q-Q PLOT	63
3.7	ENGENHARIA DE CARACTERÍSTICAS	64
3.7.1	Normalização de Dados e Análise de Variação em <i>BASE LOG</i>	65
3.7.2	Médias Móveis	66

3.7.3	Transformação de Variáveis Categóricas	67
3.8	REGRESSÃO LINEAR	69
3.9	K-NEIGHBORS	70
3.10	MÁQUINAS DE VETORES DE SUPORTE.....	71
3.11	REDES NEURAIS ARTIFICIAIS.....	72
3.12	AUMENTO DE GRADIENTE	73
3.13	FLORESTAS RANDÔMICAS.....	74
3.14	MÉTRICAS DE VALIDAÇÃO	75
4	METODOLOGIA E MATERIAIS DE USO	76
4.1	FERRAMENTAL UTILIZADO.....	76
4.2	BASE DE DADOS	77
5	RESULTADOS E DISCUSSÃO	80
5.1	DESCOBERTA DE CONHECIMENTO	80
5.2	ENGENHARIA DE CARACTERÍSTICAS.....	85
5.3	SEPARAÇÃO DOS DADOS E NORMALIZAÇÃO.....	86
5.4	AVALIAÇÃO DA PERFORMANCE DOS MODELOS.....	86
6	CONCLUSÃO	95
7	REFERÊNCIAS.....	96

1 INTRODUÇÃO

Nos últimos anos, é observado um grande avanço voltado ao setor energético, tendo um notável crescimento no consumo de energia elétrica em escalada global. A priori, esta importância se deve ao fato de ser atribuída a uma diversa combinação de fatores, sendo o desenvolvimento tecnológico um dos principais impulsionadores. Com a proteção de dispositivos eletrônicos, a expansão da infraestrutura de tecnologia da informação e a crescente dependência de soluções digitais em diversos setores da sociedade, acabou-se desenvolvendo um aumento exponencial na demanda de energia elétrica (GALICKI, 2022). Outrora, novas indústrias voltadas ao setor da mobilidade elétrica e as energias renováveis, trouxeram uma demanda adicional por eletricidade, impulsionando ainda mais o setor de energia elétrica e se adaptar para atender essas necessidades emergentes.

A mesma medida que nota-se o crescimento da demanda do setor elétrico, surge uma grande necessidade de adotar novas práticas sustentáveis e cada vez mais eficientes no consumo de energia. Algumas iniciativas visam promover cada vez mais a eficiência energética, a transição por fontes de energias renováveis e a implementação de tecnologias cada vez mais limpas. Além disso, a busca por novas soluções cada vez mais inovadoras e a conscientização ambiental vem moldando como a sociedade vem enfrentando sobre o consumo de energia elétrica, incentivando cada vez mais a transição para um futuro cada vez mais sustentável e resiliente (LOPES e NETO, 2022).

A transformação cada vez mais repentina digital e a revolução industrial em curso, estão entre os principais tópicos do crescimento incessante do consumo de energia elétrica. A *Internet of Things* (Internet das Coisas - IOT), Inteligências Artificiais (IAs) e a automação de tarefas, estão ganhando cada vez mais proeminência em vários setores, surgindo assim uma dependência adicional por uma infraestrutura energética mais robusta que esteja preparada para suportar a demanda necessária para suprir toda a cadeia da sociedade (KAGAN, OLIVEIRA e ROBBA, 2021).

Paralelamente a estes fatores, órgãos governamentais, e diversas organizações estão buscando intensificar seus esforços para descarbonizar o setor elétrico, para proporcionar cada vez mais fontes de energia renováveis para implementar medidas cada vez mais eficientes e longevas. Esta busca por soluções

mais sustentáveis e a conscientização sobre os efeitos das mudanças climáticas, traz uma remodelação de novas políticas energéticas em muitas partes do mundo, uma vez que este tema vem se tornando cada vez mais promissor, importante e significativo.

Considerando a magnitude desses desafios e a necessidade urgente de prever e gerenciar efetivamente o consumo de eletricidade, este trabalho visa enfrentar os desafios da previsão de carga no contexto específico do Nordeste do Brasil. O objetivo é desenvolver e avaliar modelos de aprendizagem que possam prever com precisão o consumo regional de eletricidade, contribuindo assim para a eficiência e sustentabilidade do setor energético da região.

A previsão de carga desempenha um papel importante na otimização do fornecimento de energia. Ao prever os padrões de consumo, as empresas de serviços públicos podem planejar eficientemente a produção e distribuição de energia, evitar a falta ou o excesso de oferta e alcançar operações mais econômicas e ambientalmente sustentáveis. Adicionalmente, a capacidade de prever picos de procura permite uma gestão mais eficaz dos recursos energéticos, reduzindo o risco de sobrecargas e interrupções de fornecimento (JÚNIOR, 2022).

Nesse contexto, o objetivo deste trabalho é desenvolver um modelo de previsão de carga altamente preciso e confiável, considerando a complexidade dos fatores envolvidos. Para isso, será utilizado variáveis relacionadas a energia elétrica nordestina e dados meteorológicos na construção do modelo. A integração destes fatores proporciona uma compreensão mais abrangente dos padrões de consumo e permite uma adaptação mais eficaz às flutuações sazonais e climáticas.

Além disso, este estudo também visa comparar a eficácia de diferentes algoritmos de aprendizado de máquina na previsão de carga elétrica. Ao explorar modelos como o *Support Vector Regressor*, *Multi Layer Perceptron*, *Random Forest* e *Gradient Boosting*, pretende-se descobrir quais modelos oferecem o melhor desempenho para o contexto específico da região Nordeste. Esta análise comparativa é importante para a tomada de decisões na implementação de um sistema de previsão de carga mais preciso e eficiente.

Em última análise, este estudo visa contribuir não apenas para o maior desenvolvimento da previsão de carga, mas também para o fortalecimento do setor energético em todo o Nordeste do Brasil, podendo ser replicado para as demais regiões do país. A implementação de sistemas de previsão mais precisos e

eficientes não só otimizará as operações dos serviços públicos, mas também beneficiará os consumidores, garantindo um fornecimento de energia mais estável e fiável.

2 REVISÃO BIBLIOGRÁFICA

Esta seção do trabalho proporciona uma análise do sistema energético mundial e nacional, destacando as principais peculiaridades do setor energético. Também são discutidos alguns avanços científicos recentes, que servirão de base para a construção deste trabalho. Logo em seguida, serão revisadas algumas técnicas estatísticas aplicadas à descoberta de conhecimento. São discutidos também os procedimentos envolvidos na identificação e análise de padrões em conjuntos de dados, visando a extração de informações relevantes. Além disso, explora-se a influência dos vieses na etapa de engenharia de recursos, destacando a importância da geração de novas variáveis como um componente essencial desse processo, e finalmente, também são exploradas as aplicações dos modelos de ML.

2.1 PANORAMA A NÍVEL INTERNACIONAL

A seção é dedicada a demonstrar os fatores e parâmetros relacionados ao cenário internacional, a fim de explicar todo o cenário energético e climático mundial atualmente. Para que seja possível ter uma visão parcial do cenário macroeconômico energético, econômico e climático.

2.1.1 O Problema das Mudanças Climáticas

Com o avanço da tecnologia, a sociedade progrediu em inúmeros aspectos. Dentre todas as evoluções que a sociedade passou, torna-se imprescindível citar a Revolução Industrial, uma vez que ela ocasionou numa disruptiva dos modelos de operação em vigência nas indústrias e fábricas do século XVIII, na Inglaterra. Devido ao uso do carvão, como sendo a principal fonte de energia da época, a troca dos

produtos manufaturados por produtos oriundos das manufaturas ocasionou um grande conflito entre a classe operária e os empresários da época.

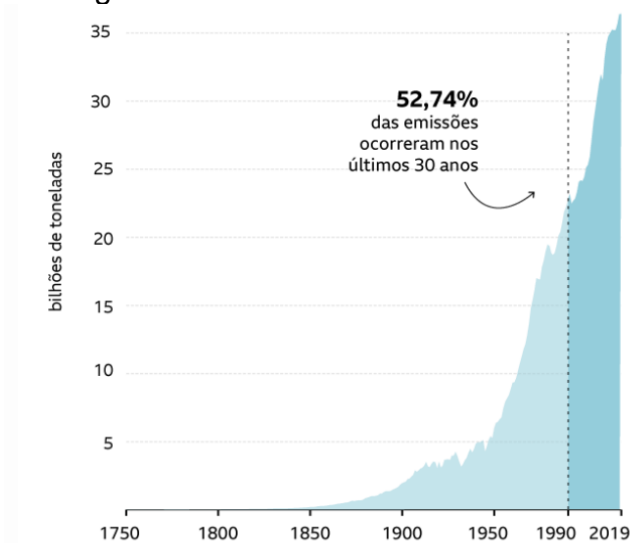
Embora a Revolução Industrial tenha trazido inúmeros benefícios para o avanço da sociedade, houve um custo e um impacto significativo no meio ambiente. O fato das indústrias e fábricas da época utilizavam carvão mineral e outros combustíveis fósseis resultou em emissões crescentes de dióxido de carbono e entre outros gases que, de certa forma, contribuem para o aquecimento global e aumento das emissões de gases do efeito estufa (GEEs), causando um aumento da temperatura global (BALOGUN, MARKS, *et al.*, 2022).

A priori, o problema das mudanças climáticas é uma das questões mais urgentes e que vem assolando a sociedade nestes últimos tempos. Apesar de terem sido observadas desde a Revolução Industrial, só foram notados os problemas que as mudanças climáticas carregam, ao final do século XX. Sendo assim, houve uma enorme negligência em relação ao problema, e como muitos governos e empresas vieram ignorando os riscos envolvidos, esperou-se chegar a níveis alarmantes para se poder tomar medidas contra essas emissões (SALOMÃO e ALMEIDA, 2022).

A partir da década de 1990, houve uma comoção da comunidade científica a fim de mobilizar um alerta a nível mundial para se tomar medidas para a solução dos problemas dessas emissões. Precisamente no ano de 1997, surgiu-se no Japão em uma convenção o Protocolo de Kyoto, que teve grande importância e foi considerado um marco em estabelecer metas e práticas para a redução de GEEs nos países industrializados.

Após os desdobramentos, as mudanças climáticas se tornaram uma das principais pautas do mundo, tanto por parte dos governos, empresas e indivíduos, mostrando cada vez mais exigências para conscientizar a urgência de planos de ação. A Figura 1, apresenta as emissões de CO₂ ao longo das décadas até os dias atuais, utilizando a base de dados disponíveis pela *Global Carbon Project*.

Figura 1 - Emissões de CO2 Mundial



Fonte: (*Global Carbon Project, 2022*)

A transição das matrizes energéticas para um cenário de fontes renováveis, como a solar e a eólica, tem se acelerado cada vez mais, no qual, muitos países têm estabelecido metas deveras ambiciosas para mitigar essas emissões.

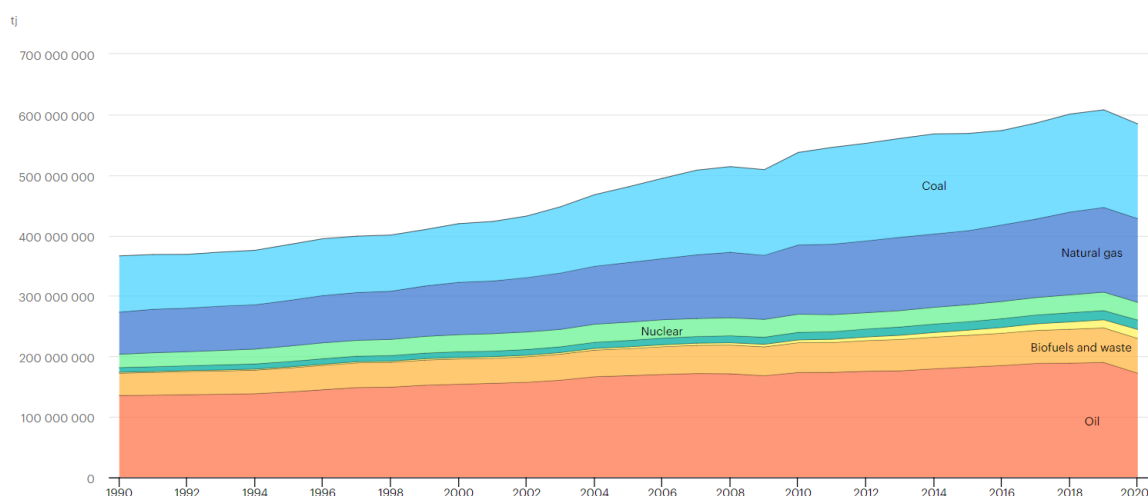
2.1.2 O Cenário Energético Internacional

O cenário energético mundial, vem passando por mudanças significativas ao longo das décadas. Devido ao aumento da população mundial, a demanda por energia aumentou de maneira exponencial, junto a isso há a necessidade de reduzir as emissões de GEEs. A partir do começo do século XXI, houve um aumento de na geração de energia a partir de fontes renováveis, como os setores de energia solar e energia eólica, sendo grandes promissores para geração de energia para diminuir progressivamente a produção de energia oriundas de combustíveis fósseis, tais como: carvão mineral, petróleo e gás natural.

Muitos países industrializados, têm investido cada vez mais em políticas e investimentos em tecnologias para promover a transição energética para fontes mais limpas e sustentáveis, como: energia nuclear, hidrelétricas, solar e eólica. A sociedade vem caminhando progressivamente para se chegar em uma matriz energética global cada vez mais sustentável.

Segundo a IEA (*International Energy Agency* - Agência Internacional de Energia), houve no ano de 2022 uma geração de energia elétrica próxima de 25.721 TWh (terawatt por hora), na qual deste montante do consumo mundial beirou 21.371 TWh, fazendo a desconsideração das perdas técnicas que naturalmente ocorrem na transmissão de energia ao redor do mundo. Contudo, pode-se analisar o Figura 2, que representa o fornecimento de energia mundial para diferentes fontes de energia.

Figura 2 - Fornecimento de Energia Mundial por Diferentes Fontes



Fonte: (International Energy Agency, 2022).

De acordo com a base de dados da IEA os maiores produtores de energia do mundo são: China, Estados Unidos, Índia, Rússia e Japão. No entanto, quando se faz o *ranking* em termos de consumo de energia, há uma pequena mudança, sendo ela: China, Estados Unidos, Índia, Japão e Rússia. Estes países, são os responsáveis por grande parte do consumo e da produção global de energia. Porém há uma demanda crescente com o passar dos anos por energia em países em desenvolvimento, tais como: Brasil, Indonésia e África do Sul. Estes países vêm aumentando suas populações e expandindo suas matrizes para suprir a capacidade de produção de energia por meios de fontes renováveis (IEA, 2023).

A visão de um cenário de gestão de energia mais sustentável, mostra-se como uma pauta cada vez mais relevante, devido a promoção de políticas e incentivos para a transição para fontes de energia mais limpas e reduções das emissões de GEEs. Investimentos públicos e privados, transformam o cenário científico destes países, uma vez que promovem o avanço da tecnologia mais eficiente para tecnologias já

existentes e para elaboração de projetos de novas formas de obtenção de energias renováveis, armazenamento de energia, aparelhos eletroeletrônicos mais eficientes no consumo de energia e afins (NETO e LOPES, 2022).

2.1.3 O Consumo de Energia e sua Relação com os Habitantes

A princípio, para que haja longevidade na vida do ser humano, é necessário que haja qualidade de vida. A qualidade de vida trata-se de um método utilizado para mensurar as condições de vida de um ser humano, sendo um conjunto de condições que contribuem para o bem-estar físico e espiritual dos indivíduos em sociedades.

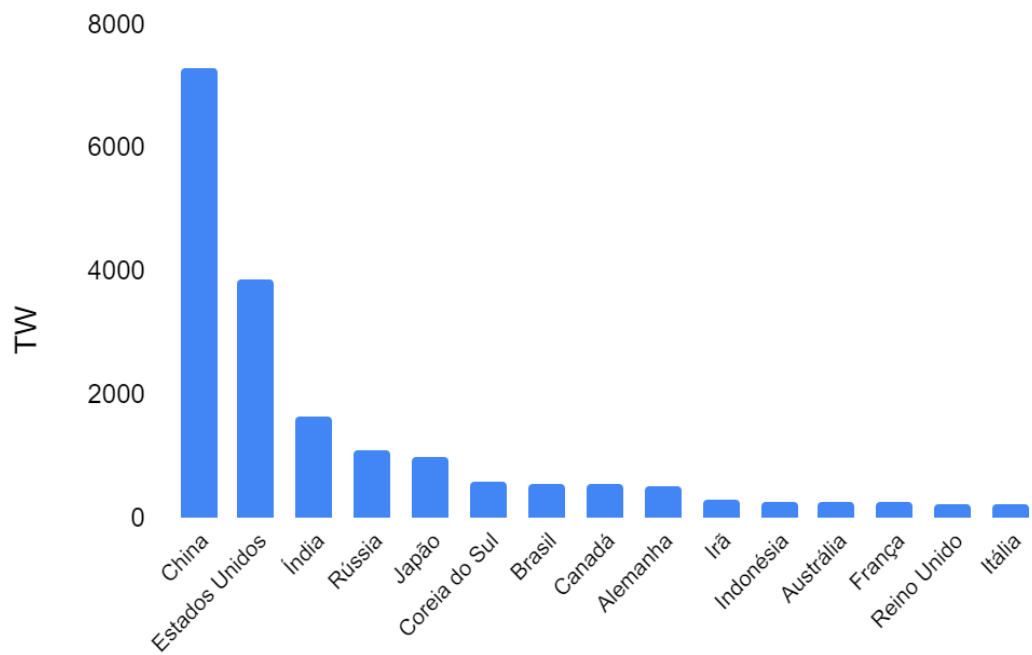
Alguns fatores são considerados de extrema importância para que se alcance essa qualidade de vida, dentre os quais pode-se citar: Família, Tratamento de Recursos Básicos (água e esgoto), Saúde Mental, Espiritualidade, Trabalho, Vida Sexual, Vida Social, Saúde Física e Saúde Mental.

Quanto maior a qualidade de vida do indivíduo, ou de uma sociedade, maior será o Índice de Desenvolvimento Humano (IDH). Este índice trata-se de uma medida comparativa usada para classificar os países pelo seu grau de desenvolvimento humano, e para classificar os países em desenvolvidos, emergentes e subdesenvolvidos.

O IDH varia de 0 (zero) a 1 (um), quanto mais próximo de 1 (um) mais desenvolvido é o país, e conseqüentemente maior será a qualidade de vida daquela população, e quanto mais próximo de 0 (zero) mais subdesenvolvido é o país, conseqüentemente menor será a qualidade de vida. Com isso, pode-se expor pelo Figura 3 os 15 (quinze) maiores consumidores de energia no mundo (base de dados IEA), a

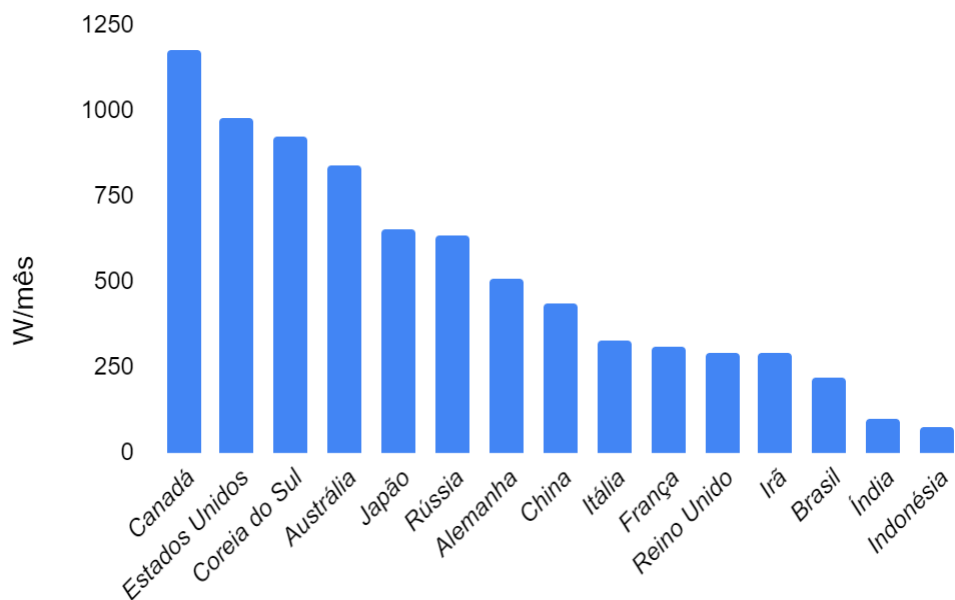
Figura 4 expõem o consumo de energia por habitante em cada país (base de dados IEA), e assimilar ambos os Figuras com o Figura 5 (base de dados da Organização das Nações Unidas - ONU), a compreensão de um comparativo entre IDHs para estes países que mais consomem energia.

Figura 3 - Comparativo entre Consumo de Energia



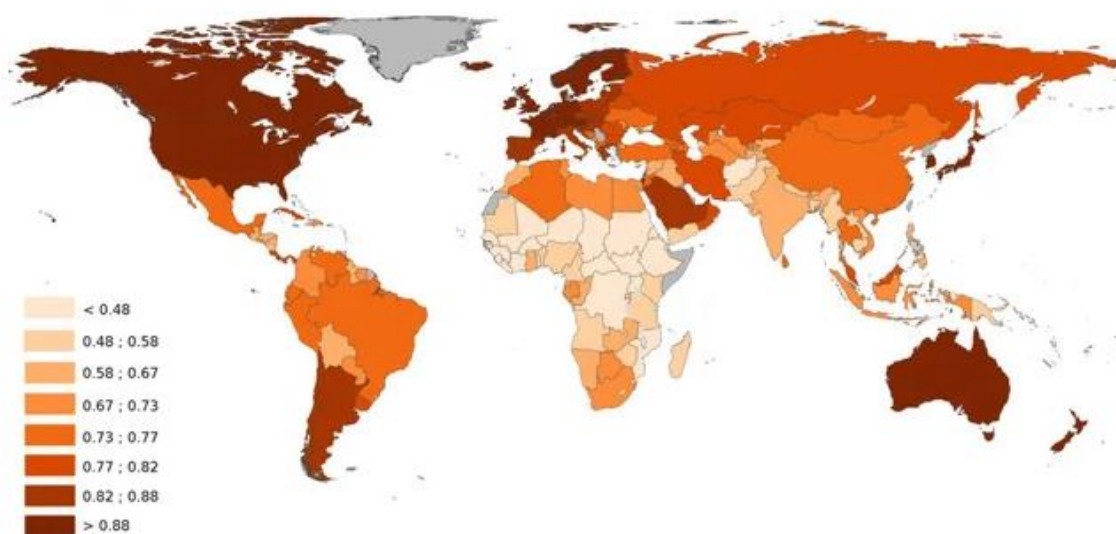
Fonte: Elaborado pelos Autores.

Figura 4 - Consumo de Energia por Habitante em Cada País



Fonte: Elaborado pelos Autores.

Figura 5 - Comparativo entre IDHs



Fonte: (ONU, 2019)

Desta maneira, pode-se relacionar a demanda de energia elétrica com o IDH da seguinte maneira: quando se analisa os países desenvolvidos, nota-se que a demanda energética desses países é alta em relação aos em desenvolvimento, logo o IDH desses países tende a ser superiores dos subdesenvolvidos.

A fim de elucidar o raciocínio, pode-se usar os Estados Unidos e o Brasil, como exemplos. Os EUA é uma país altamente industrializado e desenvolvido, com um PIB referente ao ano de 2022 de US\$ 26,13 trilhões, com uma renda per capita de US\$ 70.248,63, e uma demanda energética per capita de 1,560 kW.

Por outro lado, o Brasil, é um país considerado emergente, com um PIB de 2022 de US\$ 1,93 trilhões, com uma renda per capita de R\$ 46.154,60 e uma demanda energética per capita de 0,226 kW. Outra informação relevante para análise é o próprio IDH de cada país, que no caso dos EUA é de 0,926 e no caso do Brasil é de 0,765.

Ao analisar os dados e as informações disponíveis, torna-se evidente que os Estados Unidos desfrutam de um desenvolvimento superior e de uma qualidade de vida mais elevada em comparação com o Brasil. Essa disparidade é destacada pela constatação de que a demanda energética de carga nos EUA é aproximadamente 6,5 vezes maior do que no Brasil. Uma demanda energética mais elevada permite que os indivíduos utilizem uma variedade maior de equipamentos elétricos, tais como ar-condicionado, aquecedores, chuveiros elétricos, lava-louças, máquinas de lavar e

geladeiras. Isso, por sua vez, contribui para uma melhoria significativa na qualidade de vida da população, uma vez que o conforto individual é benéfico maior.

O consumo de energia por habitante é um indicador importante da qualidade de vida em um país. Um alto consumo de energia pode indicar um alto nível econômico e acesso a tecnologias mais modernas. Porém, este consumo excessivo de energia também pode trazer impactos negativos na qualidade de vida e no meio ambiente.

2.1.4 Correlação do PIB e Fatores Econômicos com a Energia

Em geral, países que possuem maior PIB, tendem a ser países que consomem mais energia, uma vez que têm as atividades econômicas e maior consumo por parte da população. Diante disso, existem exceções para este cenário, como no caso de países que possuem uma alta taxa de eficiência energética e baixa de combustíveis fósseis.

Além disso, outros fatores que influenciam o cenário energético com o cenário econômico, se tem relação com os índices de exportação de produtos, pois estão correlacionados com o consumo de energia, como é no caso de produtos siderúrgicos e químicos no geral. Portanto, essa relação pode ser influenciada e afetada por diversos fatores e variáveis, como o nível de tecnologia do país, a competitividade no mercado internacional e afins (WAHEED, SARWAR e CHEN, 2019).

Vale ressaltar que os investimentos feitos no setor de energia afetam tanto o consumo quanto a produção de energia em um país, na qual países que investem mais em fontes renováveis de energia podem ter uma relação mais eficiente entre o consumo de energia e o PIB, por exemplo. Outrora, investimentos em infraestrutura energética podem aumentar capacidade de produção e distribuição de energia, o que pode afetar tanto o consumo quanto a economia com um todo.

Com este cenário, para elucidar por meio do Figura 6, um comparativo entre o PIB dos maiores países do mundo é apresentado.

Figura 6 - Comparativo entre PIBs



Fonte: (IBGE, 2022).

2.1.5 A Matriz Energética Mundial

Durante os anos, há constante preocupação sobre as questões ambientais e climáticas que permeiam a sociedade, como a necessidade de praticar a redução das emissões de gases de efeito estufa. Com isso, tem se levado a uma maior adoção por fontes de energia renovável no mundo.

Segundo dados da IEA, as participações das fontes de energia renováveis no mundo vêm se tornando cada vez mais relevantes, caminhando aos poucos para garantir o protagonismo no cenário energético mundial. Estas fontes de energia têm crescido gradualmente nos últimos anos, passando de 18% de representatividade no ano de 2009, para 26% em 2018.

Essa tendência acaba sendo impulsionada por diversos fatores, na qual se inclui o avanço da tecnologia, o aumento da competitividade dos preços e liberdade econômica das fontes de energia renováveis em relação às fontes de energia não renováveis, e as políticas públicas que vêm sendo implementadas para incentivar o uso dessas fontes.

Com isso, alguns países vêm estudando planos para traçar estratégias para incentivar a produção e consumo de energia renovável. Dentre eles, alguns têm estabelecido metas ambiciosas de energia renovável para as próximas. Um exemplo disso é a União Europeia, que pretende chegar a 32% de participação de energias renováveis em sua matriz energética até 2030 (IEA, 2023).

Outros países têm investido pesadamente em projetos de Pesquisa e Desenvolvimento (P&D) para aumento da eficiência energética das fontes de energia

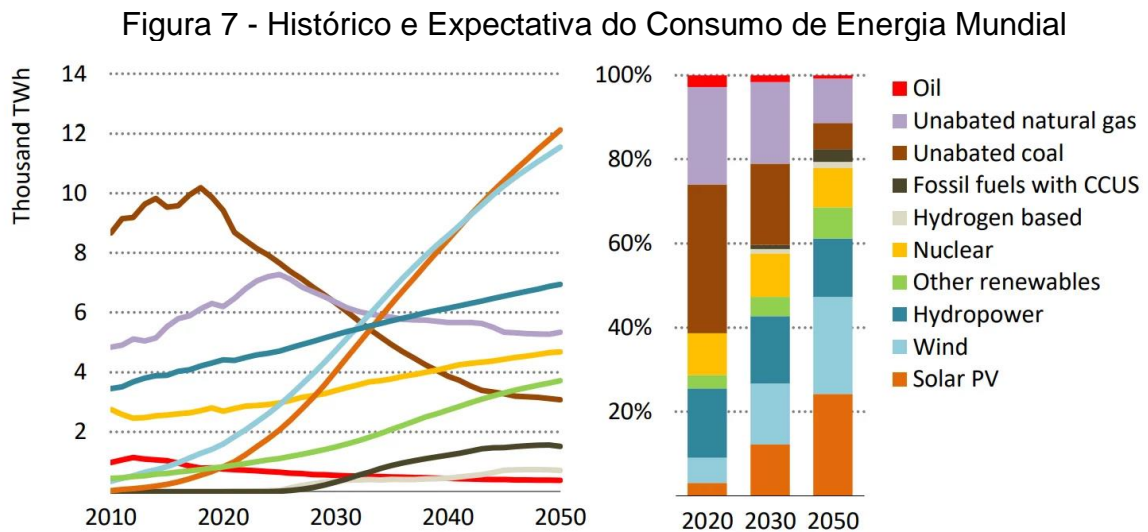
já existentes e da criação de novas fontes de energias renováveis, como é o caso da China, que atualmente lidera o mercado mundial de energia em termos de capacidade instalada.

Contudo, existem diversas convenções internacionais que visam incentivar a produção e consumo de energia renovável. O exemplo mais famoso, é o Acordo de Paris, que possui como sua principal pauta o limite do aumento da temperatura global a menos de 2°C acima dos níveis pré-industriais, como esforços para limitar o aumento a 1,5°C.

O acordo prevê uma redução das emissões de GEEs, e um aumento de estímulos para a eficiência energética, com a implementação de novas tecnologias para beneficiar o setor de energia e aumentar o uso de energias renováveis.

Segundo dados da *Bloomberg New Energy Finance*, durante o ano de 2020 foram investidos em todo o mundo, em energias renováveis, um total recorde de US\$ 303,5 bilhões. Este valor representa um aumento de 2%, em relação ao ano de 2019, o que reflete o crescente interesse de investidores em fontes de energia limpa.

Diante deste panorama a Figura 7 explana uma estimativa futura do consumo mundial de energia comercializada para os próximos anos.



Fonte: (IEA, 2023).

Em *suma*, pode-se considerar que a matriz energética mundial vem se comportando de uma forma gradual mais renovável durante os últimos anos, sendo impulsionada por uma série de fatores, na qual se inclui a adoção de políticas públicas, o avanço da tecnologia e o aumento da competitividade dos preços das energias

renováveis. Além disso, as convenções e agendas climáticas mundiais, vem estabelecendo metas cada mais ambiciosas para gerar um estímulo significativo e um crescimento gradual do setor.

2.1.6 Cenário da Energia Solar Mundial

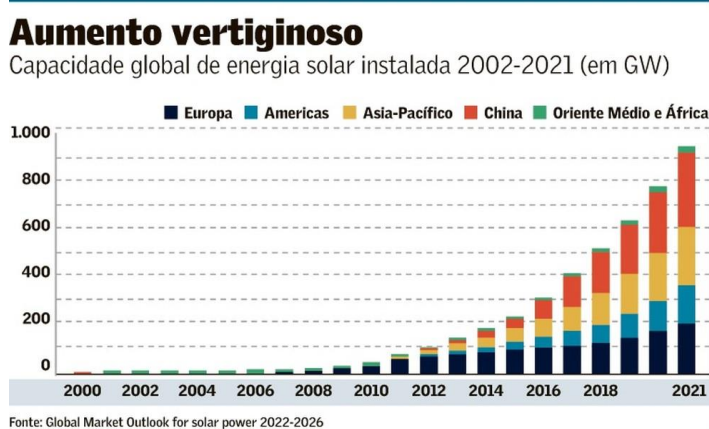
A energia fotovoltaica é considerada uma das grandes tecnologias que causaram uma disrupção no setor energético, além de ser uma das mais promissoras em todo o mundo, com um grande potencial para redução de emissões de GEEs, e contribuir para a transição para uma economia mais sustentável.

Durante a virada do milênio, este tipo de energia vem causando um impacto significativo na capacidade instalada em todo o mundo, fazendo com que impulse principalmente pela queda dos custos de produção de energia solar e pelo aumento da eficiência dos painéis.

De acordo com o a IEA, a capacidade global de energia solar fotovoltaica cresceu em média cerca de 40% ao ano entre 2010 e 2020, fazendo com que fosse alcançado um total de 760 GW no final do ano de 2020.

Diante disso, segundo a base de dados do *Global Market Outlook for Solar Power*, foi desenvolvida a Figura 8, enfatizando o crescimento vertiginoso da capacidade instalada deste tipo de energia, em relação aos anos de 2000 a 2021, em GW.

Figura 8 - Capacidade Instalada Global de Energia Solar Fotovoltaica



Fonte: (*Global Market Outlook, 2022*)

O que faz a energia solar ser o que ela é hoje, é o fato de ser altamente vantajosa com sua modularidade e escalabilidade, na qual permite que seja implantada em quaisquer tipos de instalações, sejam industriais, comerciais e residenciais, contando até mesmo com grandes usinas. Além de diversificar a matriz energética, ela reduz a dependência de combustíveis fósseis (SILVA, 2023).

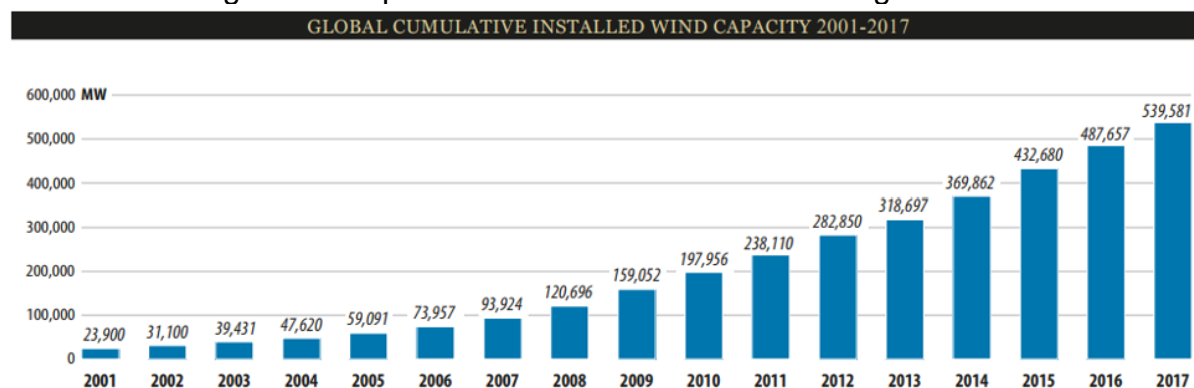
Com o estímulo a esta tecnologia faz com que haja um desenvolvimento de tecnologias que sejam voltadas para a produção, armazenagem e distribuição deste tipo de energia. Nesse contexto, a fabricação desses painéis tem estimulado crescentemente o avanço de tecnologias inovadoras para a produção de materiais semicondutores, os quais apresentam potenciais aplicações em diversos setores, incluindo a eletrônica.

Com o estímulo a essa produção, seus custos de produção têm tornado a energia cada vez mais competitiva em relação aos combustíveis fósseis, se tem levado ao aumento de uma demanda por essa tecnologia em todo o mundo. A energia fotovoltaica pode trazer benefícios como a total isolamento de uma microgeração da rede, sendo denominado este sistema de *OffGrid*, que trata-se de um sistema totalmente isolado da rede (SILVA, 2023).

2.1.7 Cenário da Energia Eólica Mundial

Por ser uma fonte de energia renovável bastante versátil, a energia eólica pode ser instalada em dois tipos de sistemas: *Onshore* e *Offshore*. Seus impactos no setor energético são bastantes expressivos, tendo crescido rapidamente em todo mundo. Segundo a ABEEólica, a capacidade de geração de energia eólica em todo o mundo aumentou cerca de 17 GW em 2000 para mais de 758 GWh em 2021. Diante desse panorama, pode-se analisar o Figura 9 abaixo, para compreender o comportamento da energia eólica no mundo.

Figura 9 - Capacidade Instalada Global de Energia Eólica



Fonte: (ABEEólica, 2022).

Com o passar dos anos, a energia eólica vem se tornando cada vez mais competitiva em relação às fontes de energia convencionais, tais como carvão, gás natural e petróleo. Isso se deve ao fato da grande redução de custo de instalação e de operação em relação às demais. Além disso, por sempre estar evoluindo, a energia eólica tem tornado as turbinas cada vez mais eficientes e mais confiáveis. Uma vez que, em regiões em que a velocidade do vento é intermitente, a produção pode ser feita em qualquer hora do dia.

Por ser uma fonte renovável, que não emite quaisquer tipos de resíduos ou emissões de GEEs na atmosfera, a construção de parques eólicos pode ter impactos ambientais locais, como a interferência na migração de aves e perturbação de habitats terrestres e marinhos.

Apesar de todos estes fatores, a energia eólica tem gerado milhares de empregos ao redor do mundo, principalmente para as áreas rurais. Além dos parques, podem impulsionar o desenvolvimento econômico de regiões remotas e a melhorar a infraestrutura energética em países em desenvolvimento.

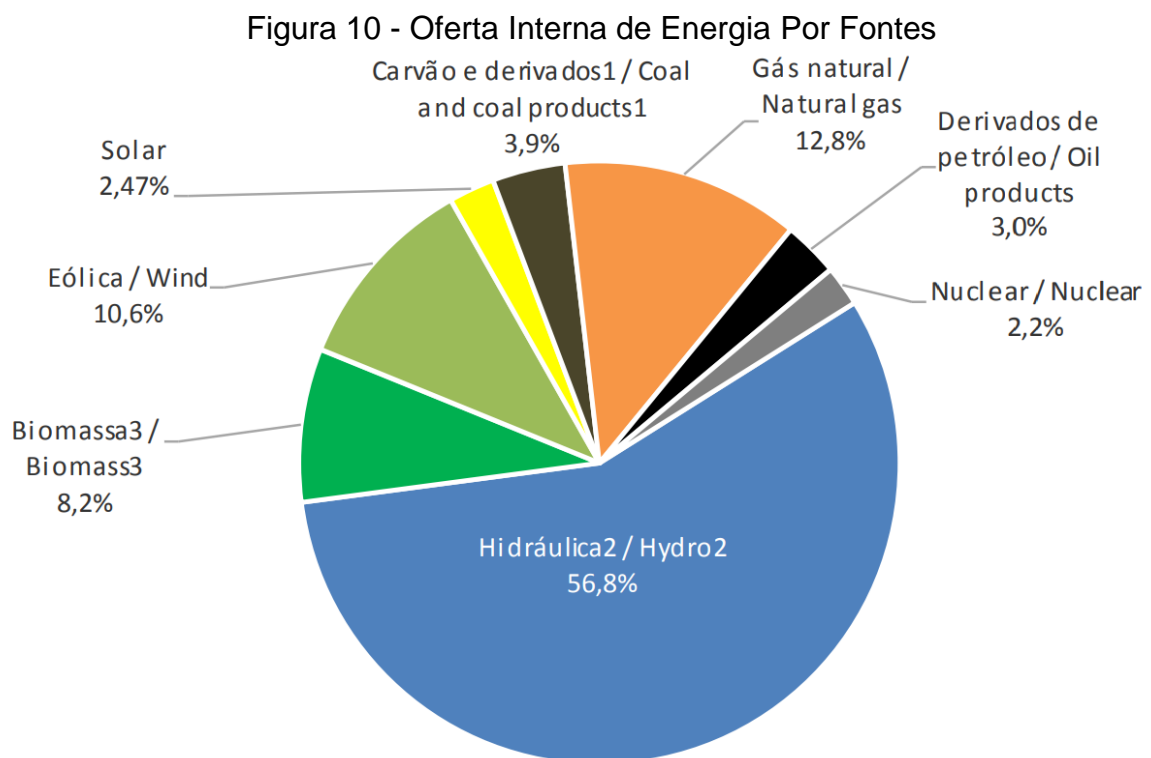
2.2 PANORAMA A NÍVEL NACIONAL

Esta seção, visa esclarecer de maneira um pouco mais profunda os cenários na qual o Brasil está inserido, e sua posição em relação aos demais países do mundo. Estes cenários visam explicar de maneira complexa e específica os fatores

econômicos, energéticos e climáticos que possuem no país. Pode-se compreender ainda mais a fundo os desafios e alternativas que o país enfrenta.

2.2.1 Comparativo do Cenário Brasileiro com o Cenário Mundial

De acordo com o Balanço Energético Nacional (BEN), relatório desenvolvido pela Empresa de Pesquisa Energética (EPE), o Brasil representa cerca de 3,17% de todo consumo de energia que foi consumida no mundo para o ano de 2022. Dentre isso, as fontes de energia renováveis representam cerca de 78,1% da oferta interna de energia no Brasil, na qual é a resultante do somatório dos montantes referentes à produção de energia nacional, podendo ser analisado pela Figura 10, que explana a oferta interna de energia.

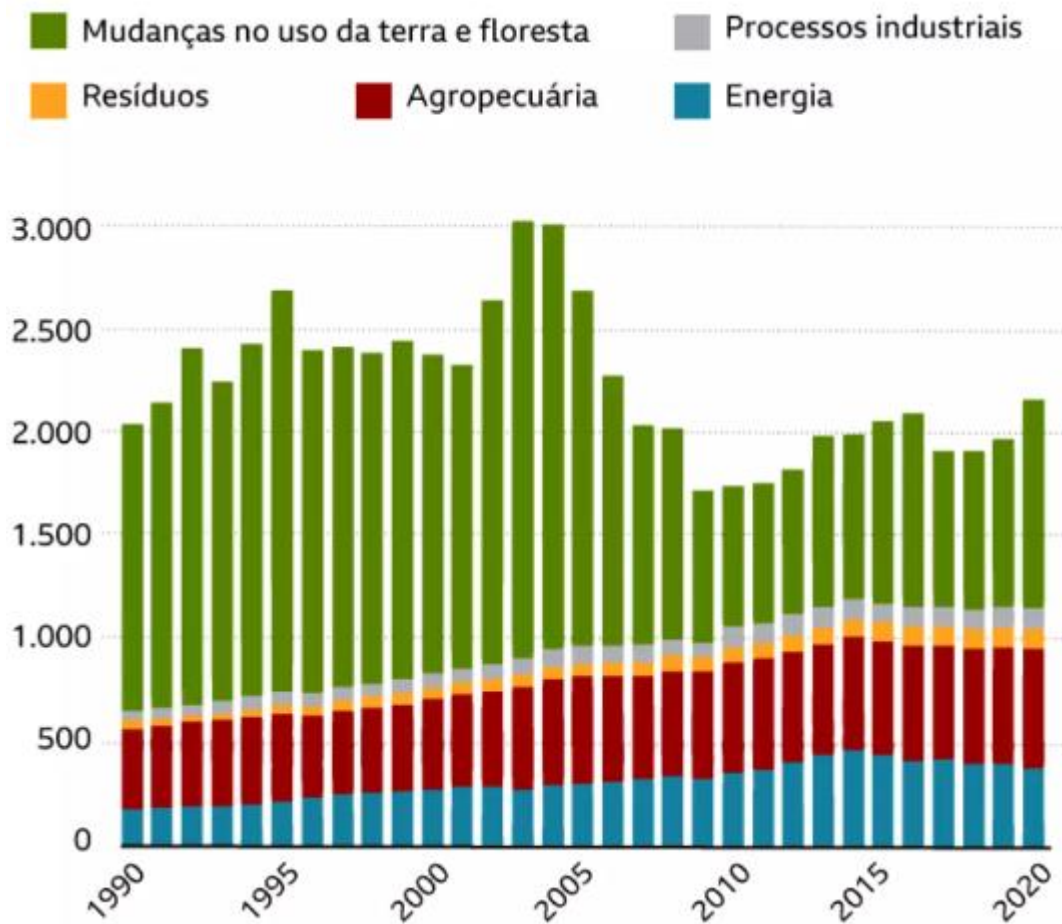


Fonte: (EPE, 2023).

Já no cenário mundial, a participação dessas fontes de energia renováveis representa cerca de 12,5% de toda produção de energia. Contudo, nota-se o poder expressivo das fontes renováveis no setor energético brasileiro, e como ele é

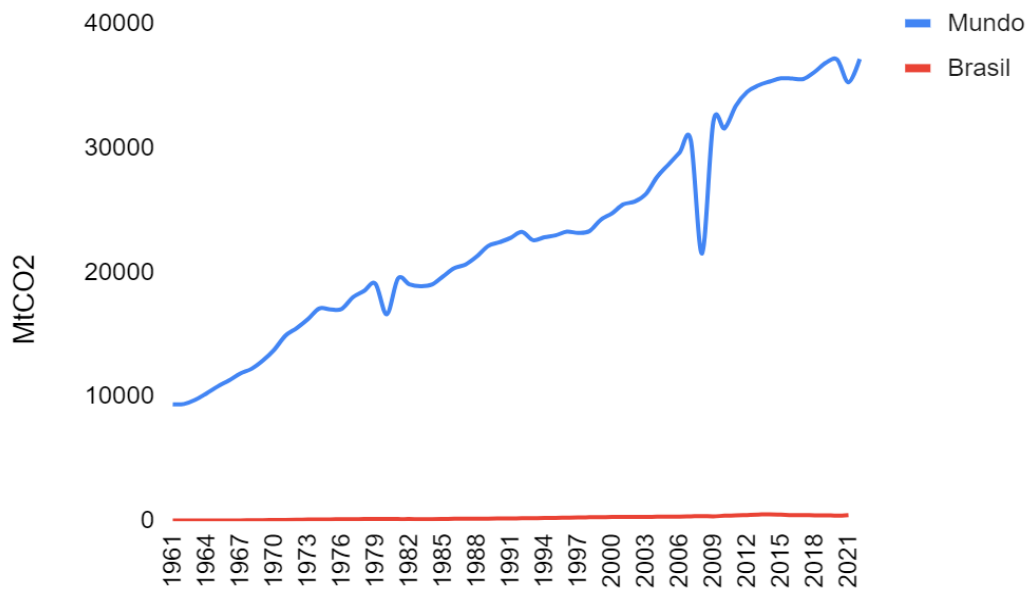
representativo em relação ao cenário mundial. Porém, por ter essa expressiva representatividade destas fontes, não mitiga o país em relação às emissões de GEEs. Diante disso, a Figura 11 expressa essas emissões no Brasil, e a Figura 12 faz um comparativo entre as emissões do mundo e as emissões do Brasil, ambos usando as bases de dados disponíveis na *Global Carbon Project*.

Figura 11 - Emissões de CO2 no Brasil



Fonte: (*Global Carbon Project*, 2022).

Figura 12 - Comparativo de Emissões de CO2 entre o Mundo e o Brasil



Fonte: Elaborado pelos Autores.

2.2.2 Sistema de Distribuição de Energia Elétrica Brasileira (SDEE)

Inicialmente, o sistema elétrico brasileiro é composto por três partes, sendo elas (KAGAN, OLIVEIRA e ROBBA, 2021):

- Geração: Parte responsável pela produção de energia por meio de usinas, podendo ser produzida de diferentes fontes renováveis ou não renováveis;
- Transmissão: Parte responsável pela transferência de energia gerada pela usina, para centros de distribuição, este processo é feito em alta tensão;
- Distribuição: Parte final responsável por entregar a energia vinda da Geração e Transmissão para o consumidor final, este processo é feito em baixa tensão.

Da mesma forma como qualquer outro sistema, o SDEE não está isento de perdas de energia, na qual são classificadas em dois tipos, sendo elas:

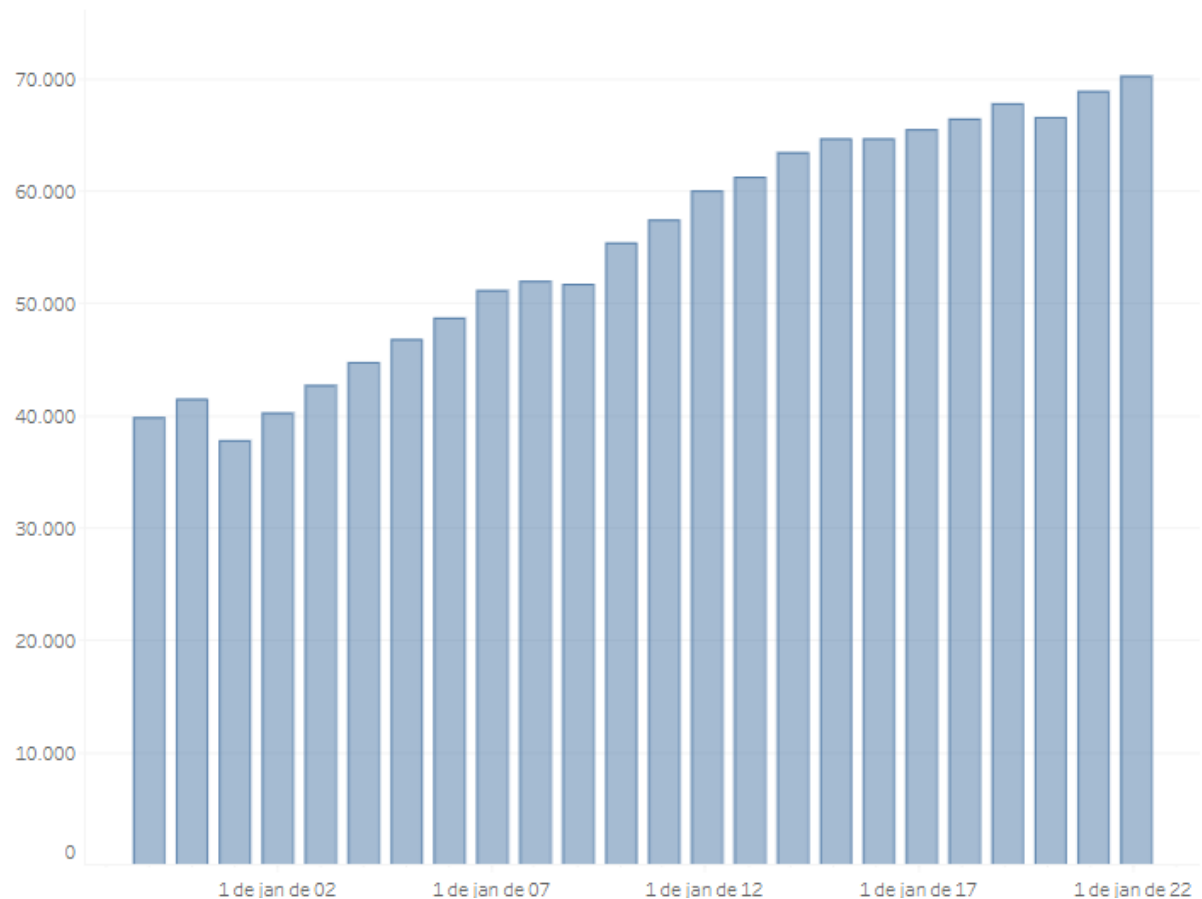
- Perdas Técnicas (PTs): São perdas esperadas e calculadas no SDEE, ou seja, inevitáveis, na qual são derivadas de fatores técnicos como o Efeito Joule, que converte a energia elétrica em calor nos condutores, por exemplo;

- Perdas Não Técnicas (PNTs): São perdas inesperadas no SDEE, também conhecida também como perdas comerciais, ou seja, evitáveis, na qual este tipo de perdas é o resultado obtido por meio da diferença entre o total de perdas na rede pelas perdas técnicas. Estão presentes neste tipo de perdas: furtos de energia, ligações clandestinas, erro de leitura por parte dos leituristas, adulteração de medidores de energia e afins.

Diante disso, pode-se utilizar a base de dados disponível pela ONS, referentes a produção de energia (Figura 13), consumo de energia (Figura 14), PTs (Figura 15), PNTs (Figura 16) e o comparativo entre as Figuras listadas acima (Figura 17).

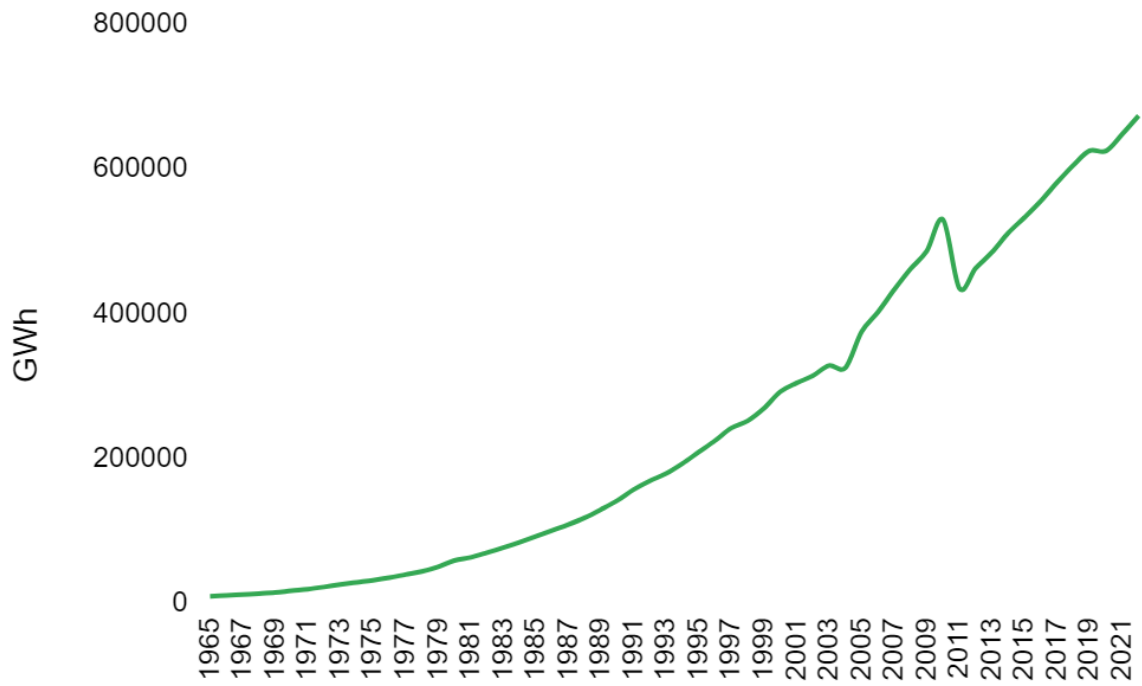
Figura 13 – Geração de Energia no Brasil

Geração de Energia (MWmed)



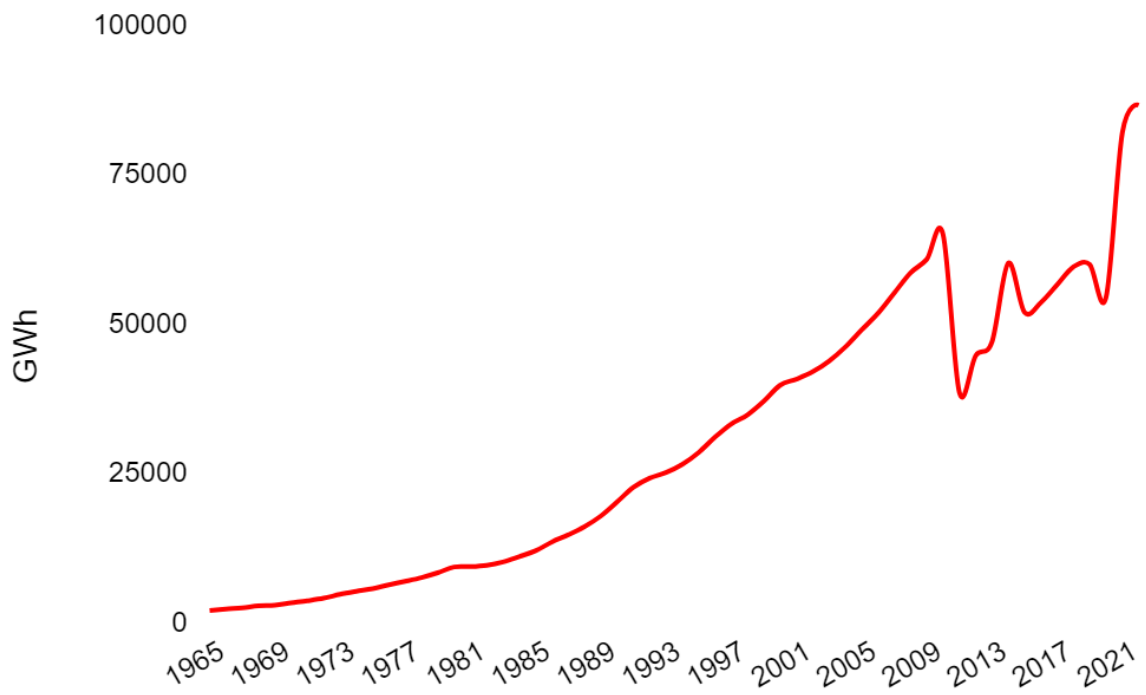
Fonte: (ONS, 2023).

Figura 14 - Histórico de Consumo de Energia no Brasil



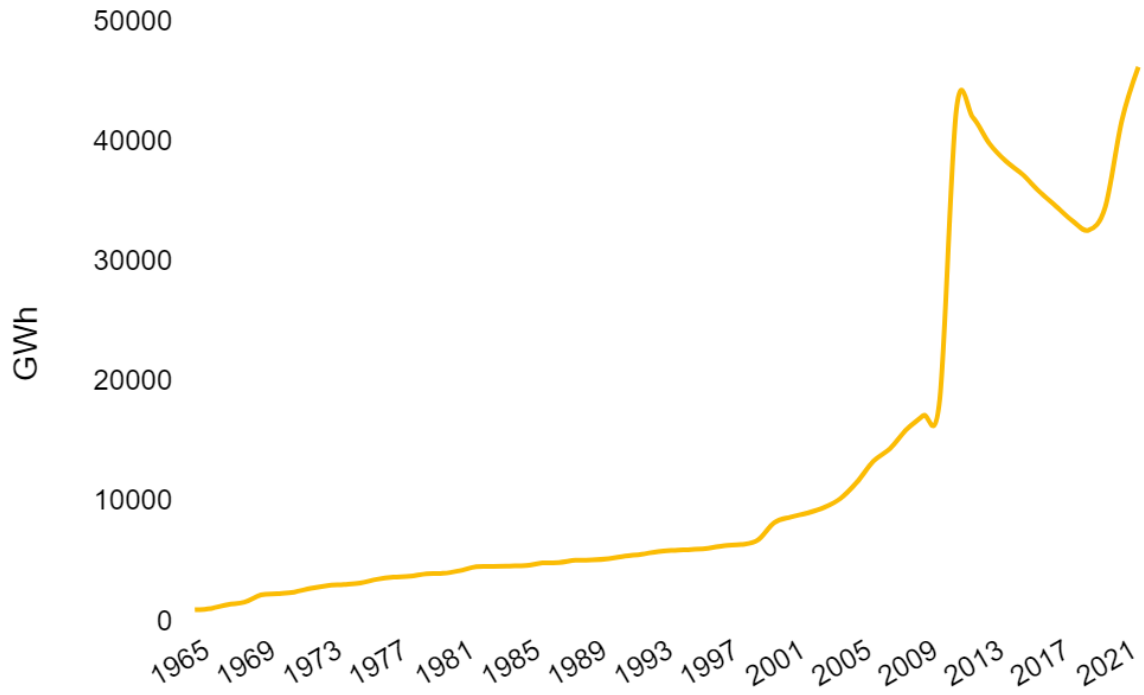
Fonte: Elaborado pelos Autores.

Figura 15 - Histórico Perdas Técnicas no SDEE



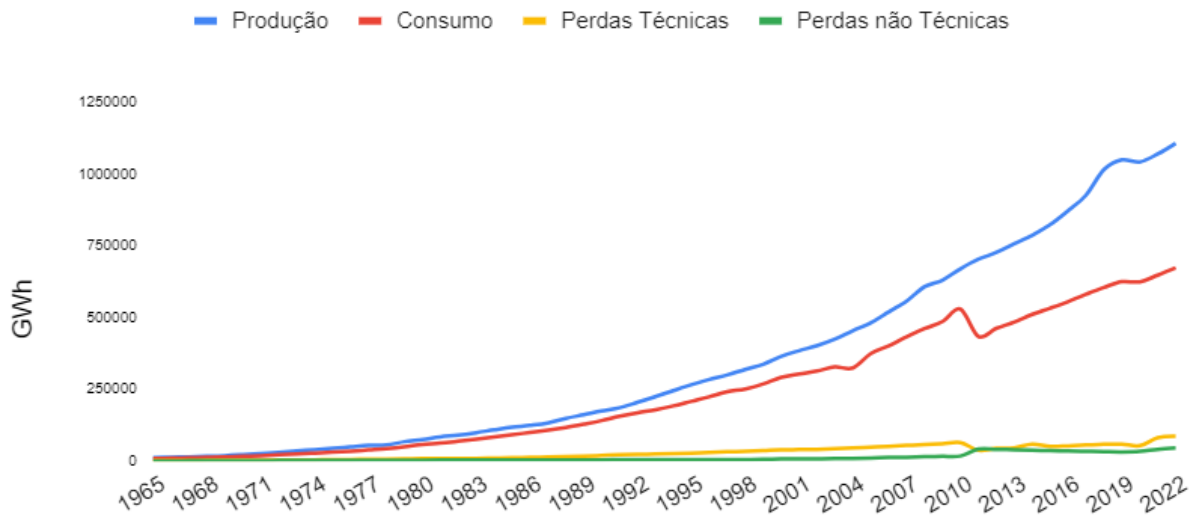
Fonte: Elaborado pelos Autores.

Figura 16 - Histórico de Perdas Não Técnicas no Brasil



Fonte: Elaborado pelos Autores.

Figura 17 - Comparativo entre Produção, Consumo, Perdas Técnicas e Não Técnicas



Fonte: Elaborados pelos Autores.

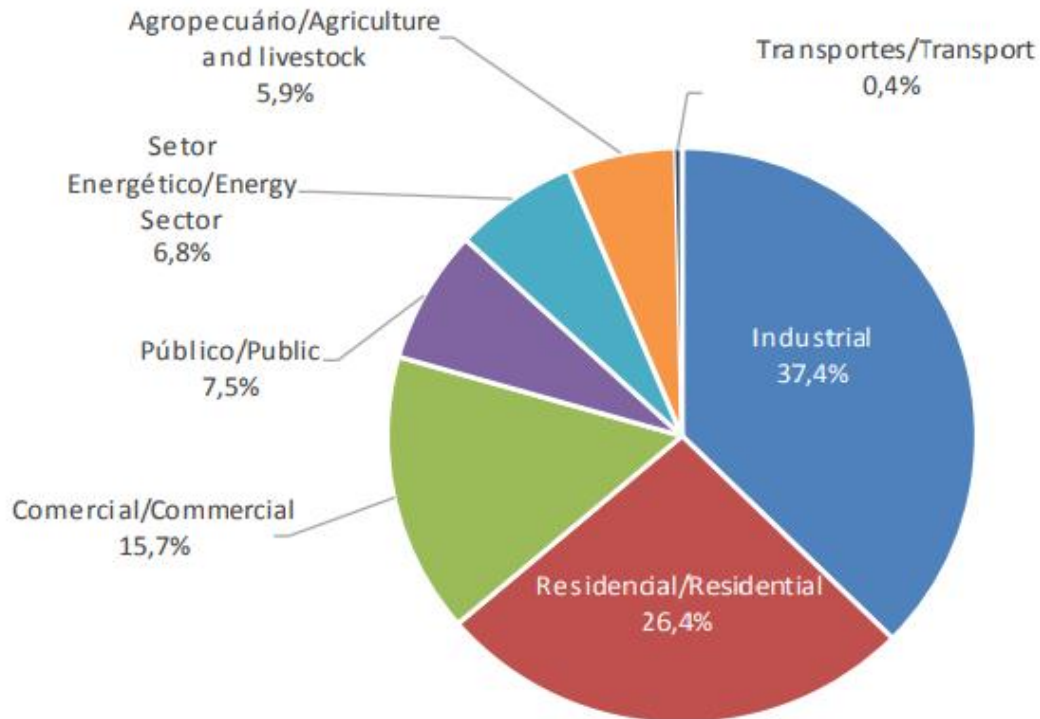
2.2.3 A Matriz Energética Brasileira

Pelo Brasil possuir uma matriz energética bastante diversificada, e proeminentemente renovável faz o país uma das referências em geração de energia. Como descrito pela Figura 10, a matriz energética brasileira possui uma forte presença da energia hidrelétrica, na qual o país aproveita das bacias hidrográficas para produção de energia.

Porém, as fontes de energia não renováveis também possuem sua parcela, na qual a produção de petróleo e gás natural possui a função da exploração de reservas *offshore* na costa brasileira. Sendo assim, o país vem fazendo investimentos para alternativas mais renováveis e sustentáveis, como: solar, eólica e biomassa (ANEEL, 2023).

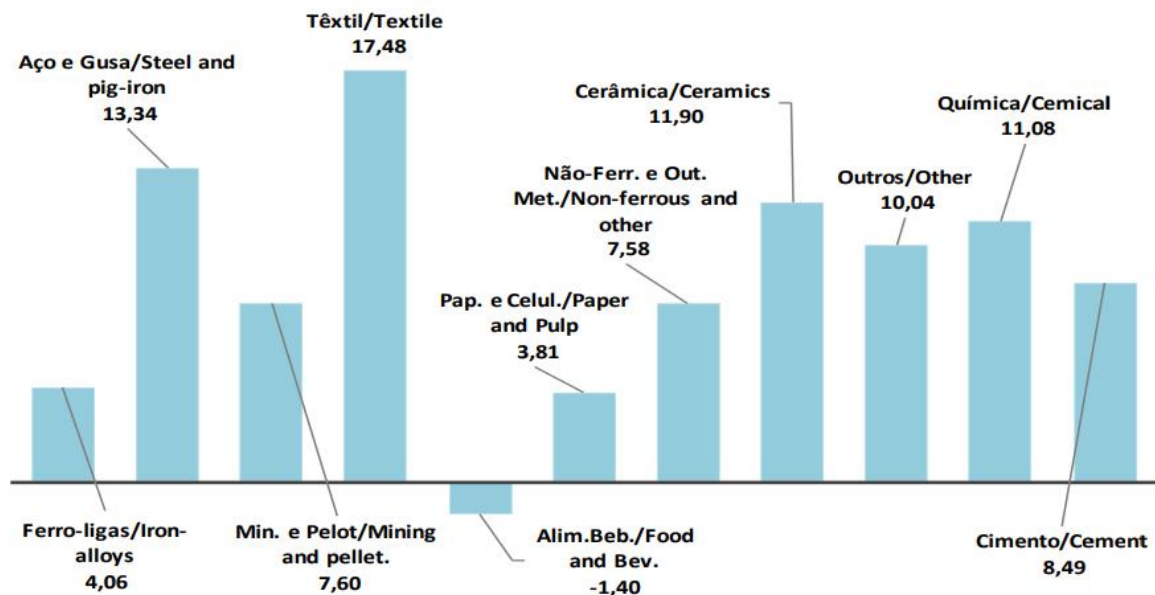
Contudo, o consumo e participação setorial no consumo de energia no Brasil, pode ser expressado na Figura 18. Onde, se pode compreender que o setor industrial tem bastante relevância no consumo, seguido do setor residencial. Além disso, pela Figura 19, pode-se ver o crescimento do consumo de energia no setor industrial e os principais setores das indústrias que consomem essa energia em relação ao ano anterior de 2021.

Figura 18 - Participação Setorial no Consumo de Energia



Fonte: (EPE, 2022).

Figura 19 - Crescimento do Consumo de Energia no setor Industrial



Fonte: (EPE, 2022).

Como analisado nas Figuras acima (18 e 19), pode perceber a força e o impacto que o setor da indústria tem no consumo de energia, na qual se destaca a indústria têxtil, aço e gusa e cerâmica. Apenas estes três setores, representam um somatório de aproximadamente 43% de todo consumo da indústria (EPE, 2022).

2.2.4 Relação do PIB com a Energia

O Produto Interno Bruto (PIB) de um país, trata-se de um indicador de bens e serviços finais em um determinado período, no caso 1 (um) ano. O órgão que é responsável por medir esse indicador é o Instituto Brasileiro de Geografia e Estatística (IBGE).

Para análise do PIB, é necessário que haja a composição de três fatores principais para se levar em conta no cálculo, sendo eles:

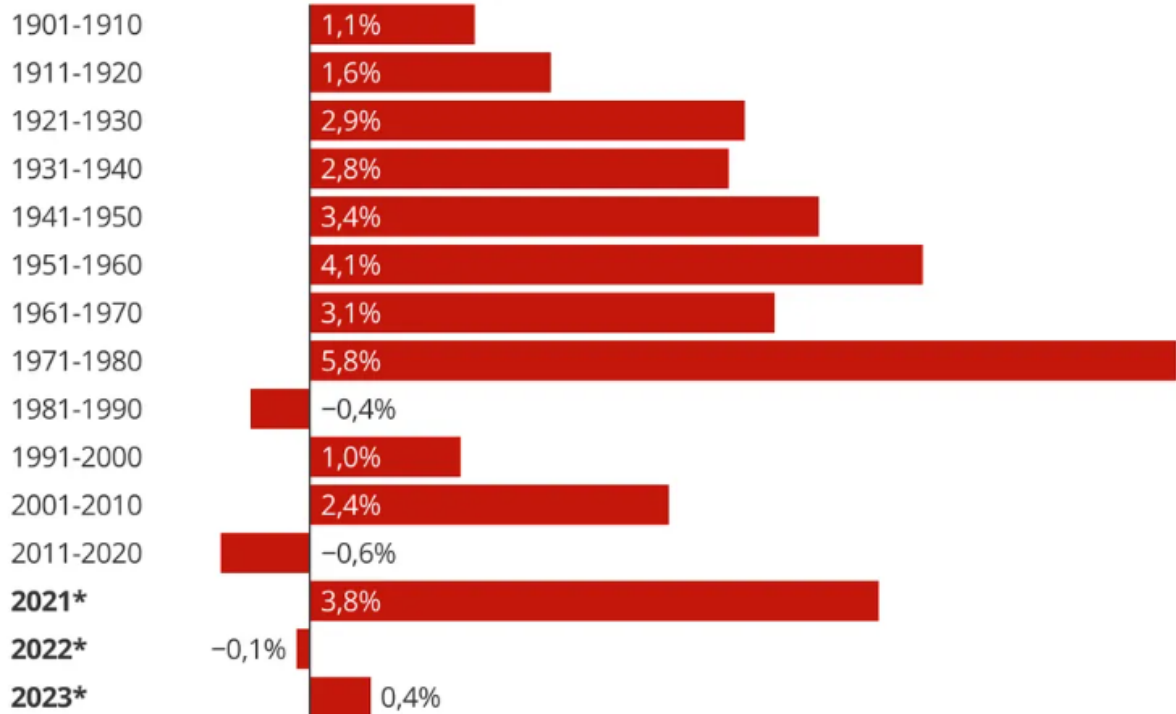
- Setor Primário: agricultura, pecuária, pesca, extração mineral;
- Setor Secundário: engloba todos os setores da indústria;
- Setor Terciário: compreende os produtos e serviços comercializados, sendo este o maior e que atinge o maior impacto no país, representando cerca de $\frac{2}{3}$ do PIB brasileiro (IBGE, 2023).

Pode-se analisar o crescimento do PIB brasileiro, por meio do Figura 20 que representa o histórico do PIB desde o ano de 1965, na qual a base de dados foi retirada do IBGE.

Figura 20 - Evolução do PIB no Brasil

Evolução do PIB per capita no Brasil

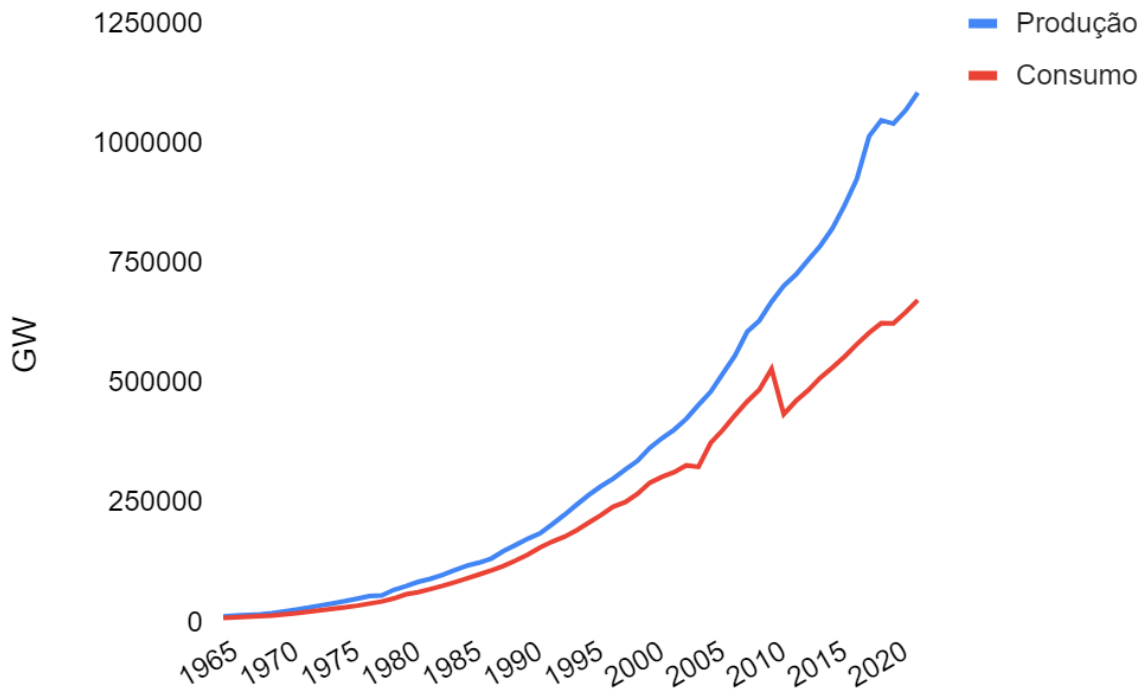
Taxa de crescimento anual médio, por década, e projeções para até 2023, em %



Fonte: (IBGE, 2023).

Nota-se que durante os anos de 1965 a 1990, o PIB do Brasil teve um leve crescimento, até passar da década de 90 (noventa) inicia um crescimento de maneira significativa até os dias atuais. Contudo, pode-se relacionar o setor de energia, com o crescimento, a Figura 21 explana a produção e o consumo de energia desde a década de 1965, na qual pode-se notar o crescimento bastante significativo a partir da década de 90 (noventa). Com isso pode-se relacionar o crescimento acelerado do Brasil a partir desta década até os dias atuais, com os incentivos e investimentos nos setores de energia (IEA, 2023).

Figura 21 - Comparativo entre Produção e Consumo de Energia no Brasil

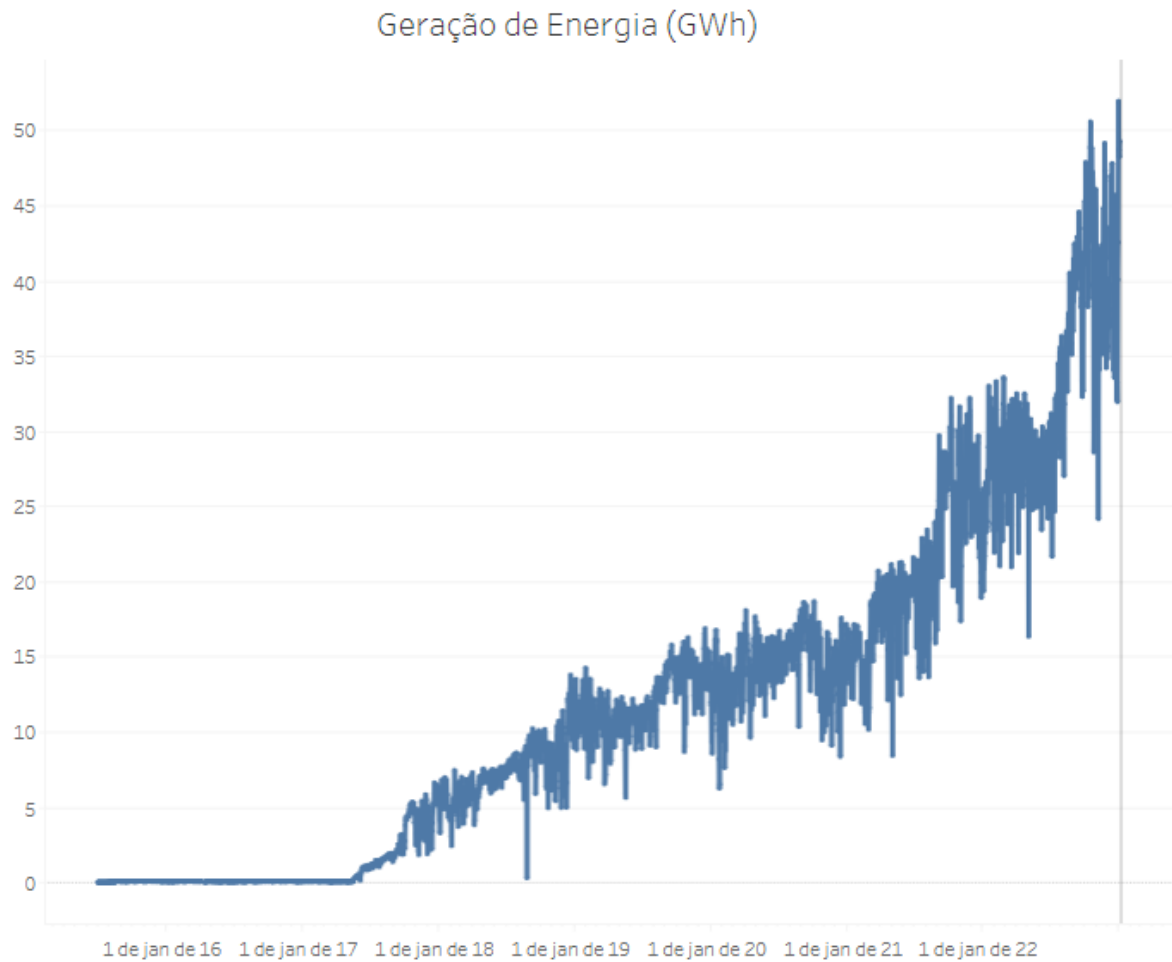


Fonte: Elaborado pelos Autores.

2.2.5 Cenário da Energia Solar no Brasil

Durante os anos 2000, houve um crescimento significativo na produção e consumo de energia no Brasil. Diante deste cenário, foram feitos diversos estudos e investimentos em uma maior variação das fontes de energia no país, a fim de não depender somente da energia hidrelétrica. Contudo, a energia solar surgiu como uma solução prática para estimular a produção de energia e contribuir para o suprimento de demanda energética no país. Pode-se analisar o crescimento da energia solar no Brasil através da Figura 22, na qual foi utilizado a base de dados do ONS, abaixo.

Figura 22 - Crescimento da Energia Solar no Brasil



Fonte: (ONS, 2023).

Este crescimento significativo da energia solar, pode ser explicado pelo aumento da mini e microgeração de energia. Essas práticas, vêm se tornando cada vez mais comuns no Brasil, devido ao aumento da conscientização sobre a importância de se investir em fontes de energia renováveis, a fim de estimular o crescimento da sustentabilidade. Estas modalidades de geração de energia, fazem com que pequenos consumidores, preferencialmente os consumidores residenciais, onde podem produzir a energia que consomem, por meio de painéis solares, podendo ser instalados nas próprias residências, ou até mesmo em empresas ou propriedades rurais (MAURICIO, 2023).

A Agência Nacional de Energia Elétrica (ANEEL) em 2012, criou a regulamentação para mini e microgeração de energia solar, e desde então tem se expandido gradualmente. Além disso, a ANEEL estabeleceu algumas regras para fazer a conexão deste sistema para os sistemas da rede elétrica, fazendo com que a

energia excedente injetada que o consumidor produza pelos seus painéis, faz gerar créditos que possam ser abatidos na conta de energia, ou seja, parte dessa energia injeta gera créditos para o consumidor utilizá-las durante os períodos de ausência de luz solar (SILVA, 2023).

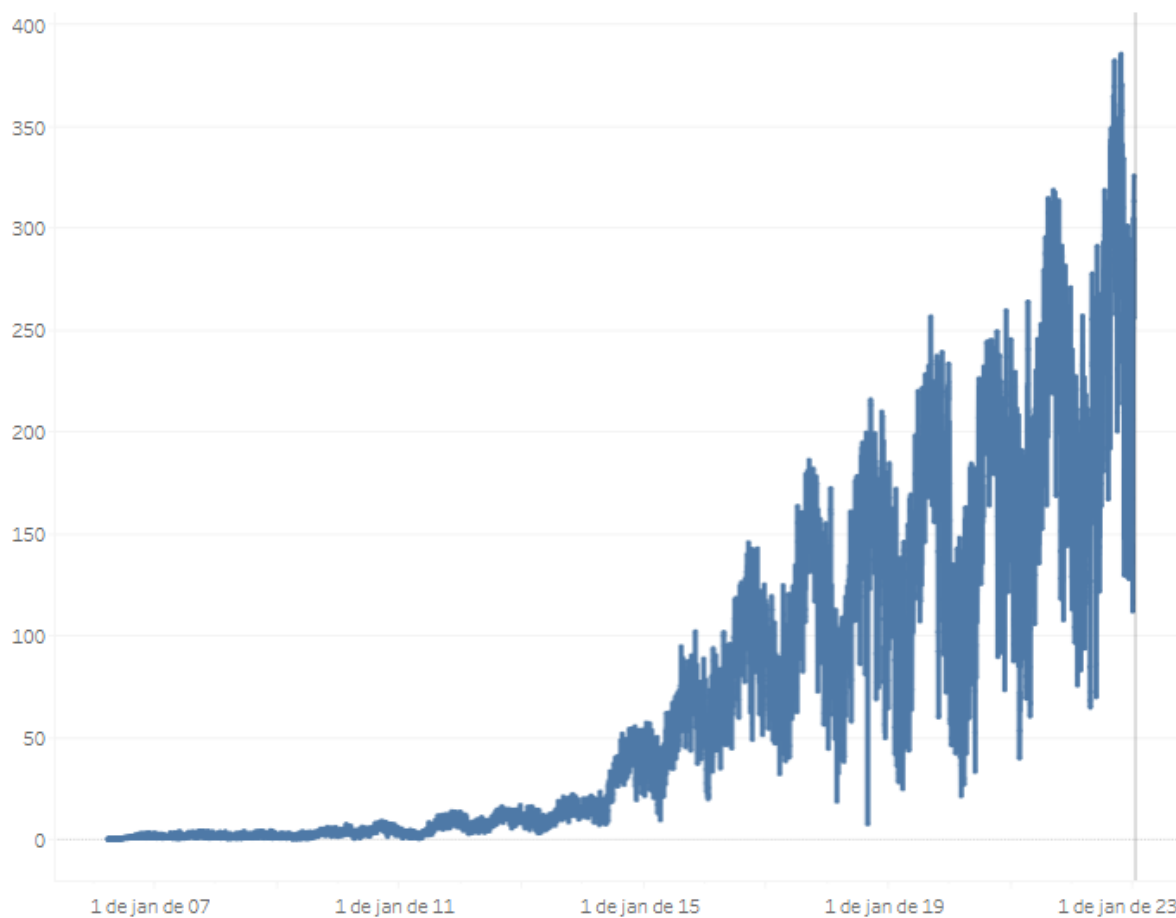
A mini e microgeração de energia solar é uma das fontes de energia promissoras para o crescimento da produção de energia futura no Brasil. Atualmente, diversas instituições financeiras oferecem formas de financiamento a juros mais baixos que os cobrados no mercado, para estimular o crescimento desta prática de geração. Além disso, o país possui um enorme potencial para essa produção, especialmente em regiões próximas à Linha do Equador, como é o caso da região Nordeste, pelos altos índices de irradiação solar (ALMEIDA e ALMEIDA, 2022).

2.2.6 Cenário da Energia Eólica no Brasil

Durante os últimos anos, pelo mesmo motivo da energia solar, a energia eólica tem ganhado bastante notoriedade na matriz energética brasileira por conta do baixo custo da implantação deste tipo de energia, em relação às demais outras fontes, e por ser uma fonte de energia, que em regiões propícias, pode gerar energia de maneira intermitente. Além disso, este tipo de fonte de energia, pode ser gerado tanto *Onshore* (implantação das turbinas eólicas em terra), quanto *offshore* (implantação das turbinas eólicas em alto mar).

De acordo com a Associação Brasileira de Energia Eólica (ABEEólica), o Brasil é o 6º maior produtor de energia eólica do mundo, com uma capacidade instalada atual de 24 GW, e com uma previsão até o ano de 2028 o país terá 44,78 GW. Partindo de 13,2% de participação na matriz energética brasileira para 20%. Diante disso, pode-se analisar pela Figura 23 o crescimento da produção de energia eólica no Brasil, por meio da base de dados do ONS.

Figura 23 - Produção de Energia Eólica no Brasil
Geração de Energia (GWh)



Fonte: (ONS, 2023).

Dentre a extensão territorial do país, o Brasil possui algumas regiões propícias para o desenvolvimento e incentivo de uma maior produção deste tipo de energia. Dentre elas, se destaca especialmente a região Nordeste, apresenta o maior potencial para produção de energia, com destaque específico para o estado do Rio Grande do Norte. Porém vale ressaltar que a região Sul, também tem seu potencial energético, com destaque para os estados de Santa Catarina e Rio Grande do Sul (ABEEÓLICA, 2023).

O Governo Federal, por sua vez, vem adotando medidas para incentivar a expansão da energia eólica no país, fazendo ações como a realização de leilões de energia elétrica provindas dessa fonte energética, na qual possibilita a contratação e incentivos de novos empreendimentos e parques eólicos. Além disso, com a implantação de políticas públicas e parcerias público-privada de incentivos de redução fiscal e benefícios financeiros (ANEEL, 2023).

2.3 PANORAMA A NÍVEL DA REGIÃO NORDESTE

Nos tópicos abaixo, será discutido o motivo pelo qual o objetivo de estudo é a região Nordeste do país. A região, torna-se uma grande promissora para ser líder no segmento brasileiro em produção e geração de energia elétrica por meio de energias renováveis, uma vez que o clima contribui fortemente para a produção de energia elétrica solar e eólica.

2.3.1 A Região Nordeste

A região Nordeste do Brasil, é composta por 9 (nove) estados, sendo eles: Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte e Sergipe. Além disso, esta região é classificada em terceira colocação de maior região em termos territoriais, e considerada a segunda região mais populosa do país com mais 58 milhões de habitantes (IBGE, 2023).

Desde sempre, a região sofre e enfrenta desafios socioeconômicos, como fatores fenótipos, a seca e fatores de infraestrutura, no qual impacta diretamente no desenvolvimento da região. Contudo, nos últimos anos, a região tem apresentado crescimentos significativos tanto no setor econômico quanto no setor energético, destacando o setor de serviços, indústrias e agronegócio (MATSUBARA, 2020).

2.3.2 Classificação da Matriz Energética Nordestina

A região Nordeste, como destacado nos tópicos acima, mostra como a região pode ser referencial não só nacional, mas também referência mundial na produção de energia renovável, principalmente para as fontes solar e eólica. No entanto, no cenário

energético da região, tem representatividade de cerca de 30% de toda a capacidade instalada de geração de energia elétrica no país (MARCOS e JÚNIOR, 2021).

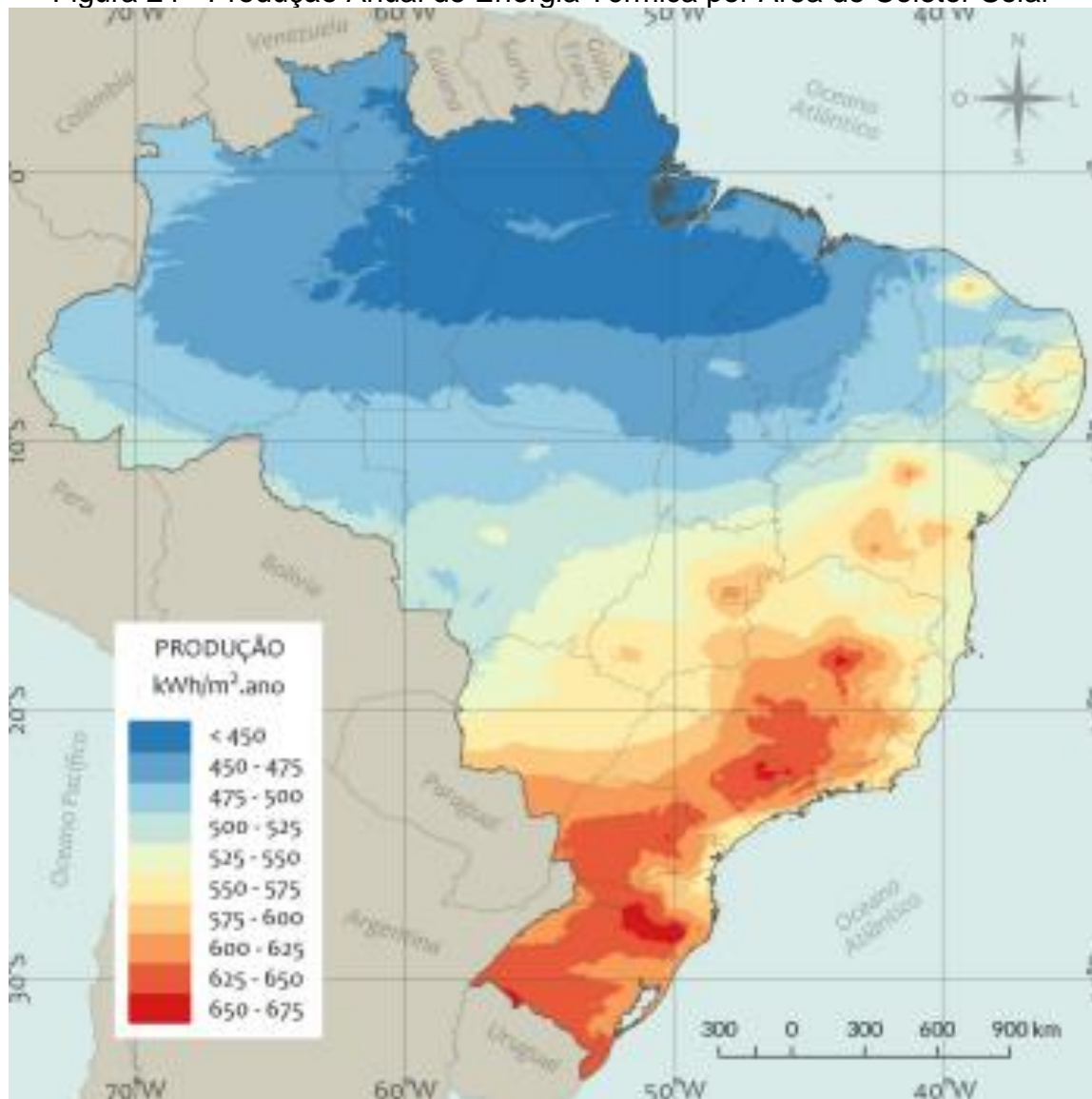
Além disso, o Nordeste também possui importantes hidrelétricas na matriz energética, como a Usina Hidrelétrica de Paulo Afonso, responsável por grande parte da energia produzida na região. Contudo, a região também conta com usinas térmicas movidas a óleo combustível, gás natural e carvão natural e carvão mineral, porém essas fontes vêm diminuindo ao longo dos anos, em vertente do crescimento de incentivos de fontes renováveis de energia sustentável (KAGAN, OLIVEIRA e ROBBA, 2021).

Entretanto, a região Nordeste vem investido em projetos de energias renováveis e sustentáveis, como criação de parques e usinas eólicas e solares, bem como o incentivo às gerações distribuídas de energia elétrica. A transição energética na região vem em iniciativas para o desenvolvimento de modelos mais sustentáveis e ecológicos (COSTA, 2022).

2.3.3 Setor Energético Solar no Nordeste

Por ser uma região promissora, o Nordeste brasileiro vem se destacando no cenário energético solar, por possuir uma das maiores irradiações solares no país, segundo o Instituto Nacional de Pesquisas Espaciais (INPE), na qual a região tem atraído investimentos em grandes projetos de geração de energia solar. Além disso, pode-se analisar pela Figura 24 a produção anual de energia térmica por área de coletores solares.

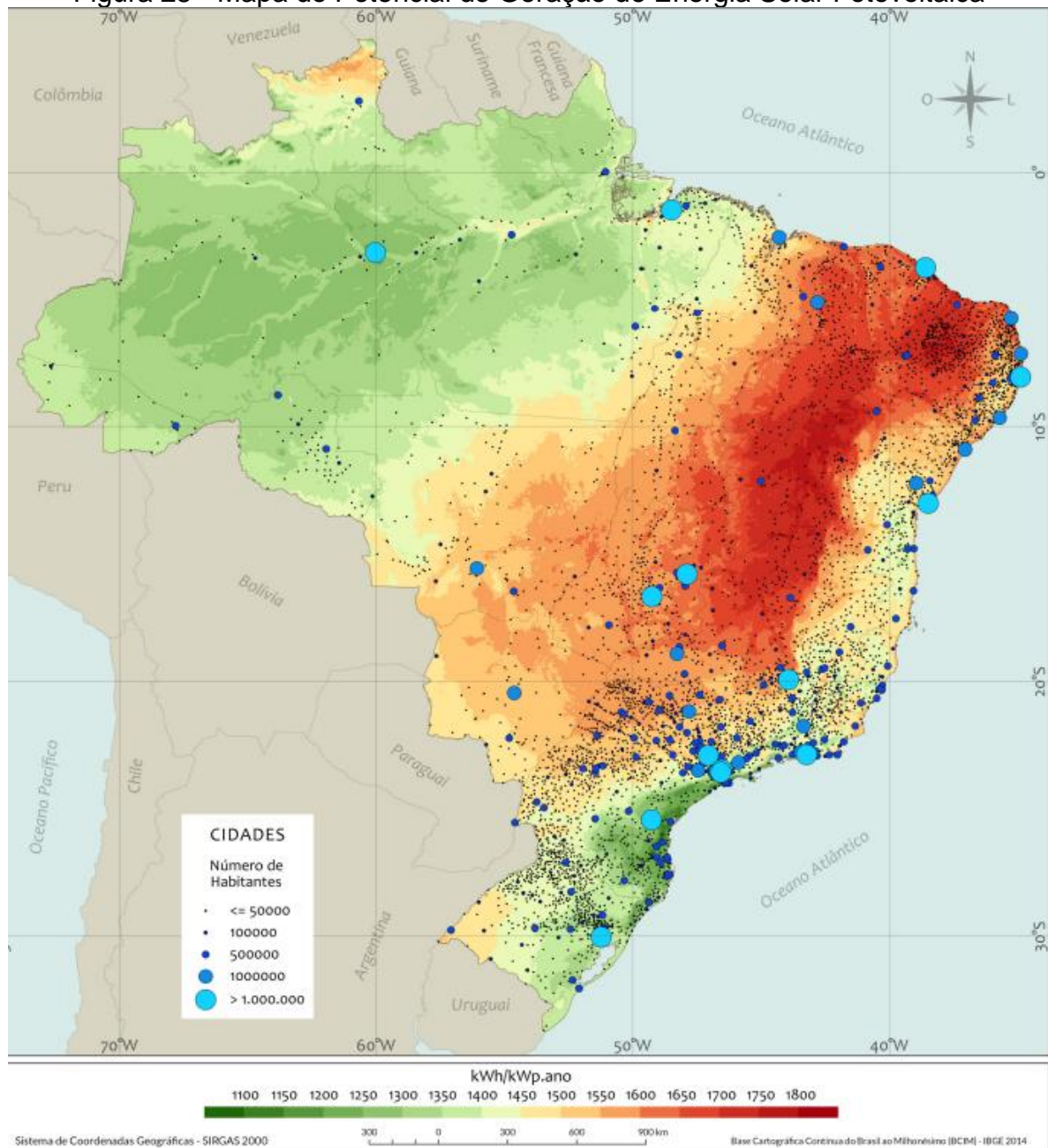
Figura 24 - Produção Anual de Energia Térmica por Área de Coletor Solar



Fonte: (INPE, 2023).

Analisando a região Nordeste do país, por meio da Figura 1, pode-se ver que a região possui uma pequena proeminência de produção de energia solar por meio de coletores solares. Porém, ao analisar a Figura 25, pode-se ver a irradiação solar para produção de energia fotovoltaica, e com isso, dá para se reparar o grande destaque da região para produção de energia, fazendo com que haja uma grande possibilidade de investimentos em mini e microgeração de energia.

Figura 25 - Mapa do Potencial de Geração de Energia Solar Fotovoltaica



Fonte: (INPE, 2023).

Portanto, a energia solar na região vem trazendo a criação de novos postos de trabalho nos setores de manutenção, instalação e operação de usinas e sistemas fotovoltaicos. Além disso, destaca projetos de grande porte nas áreas de energia distribuída, que vem contribuindo para a mitigação dos impactos ambientais e tirar as dependências de fontes não renováveis (SILVA, 2023).

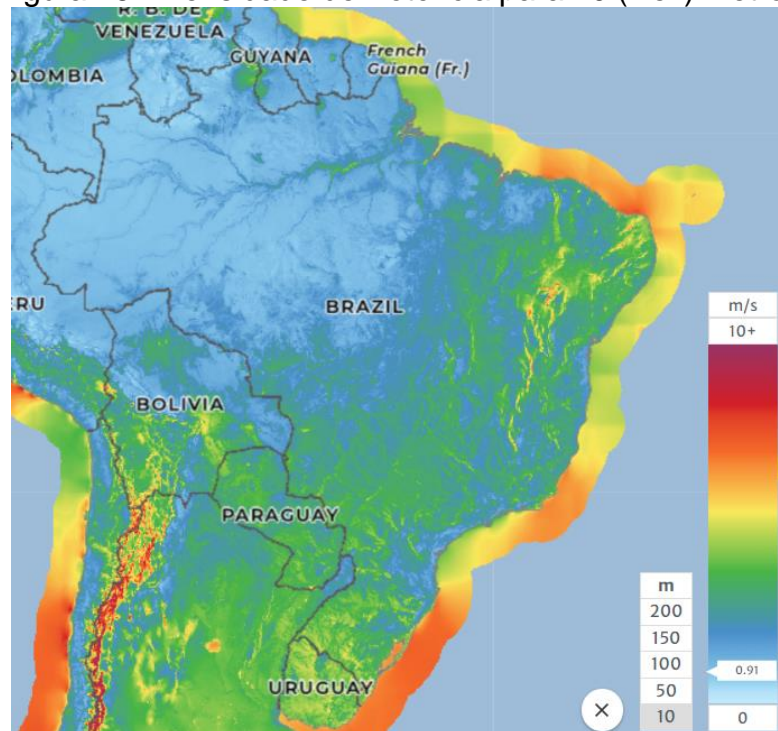
2.3.4 Setor Energético Eólico no Nordeste

A região Nordeste do Brasil, é uma das maiores promissoras em relação a produção de energia eólica no país, quiçá umas das maiores internacionalmente. Segundo a ABEEólica, a região acaba concentrando cerca de 80% de toda a capacidade instalada de energia eólica no país. Isso se deve ao fato de possuir maiores capacidade de áreas ventosas.

A *Global Wind Atlas*, mostra por meio de um atlas as potências por regiões do país para produção de energia eólica, e para as alturas de 10 (dez), 50 (cinquenta), 100 (cem), 150 (cento e cinquenta) e 200 (duzentos) metros, a região leva o maior potencial, tanto *onshore* quanto *offshore*. Essas densidades de potência podem ser analisadas e visualizadas, nas Figuras: Figura 26, Figura 27,

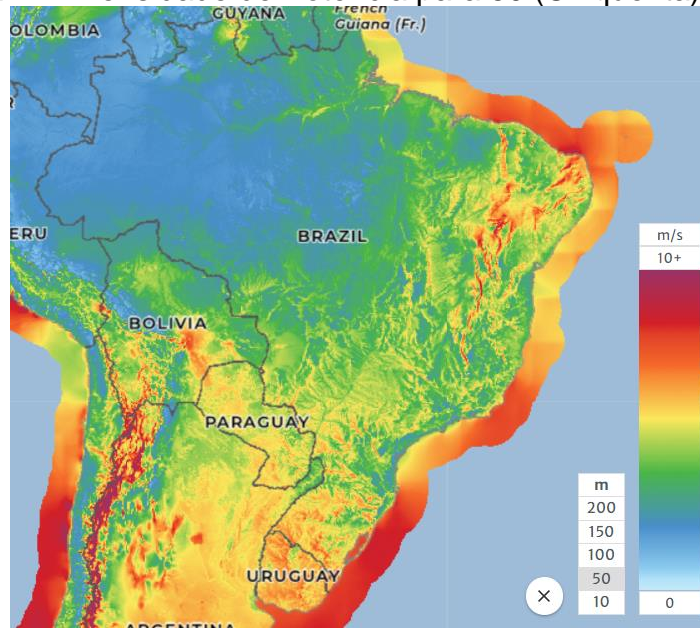
Figura 28, Figura 29, e Figura 30.

Figura 26 - Densidade de Potência para 10 (Dez) metros



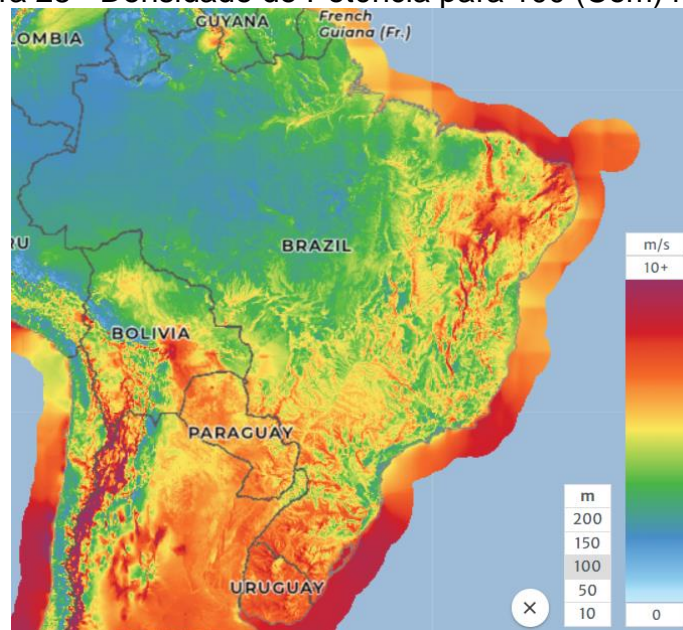
Fonte: (*Global Wind Atlas*, 2023).

Figura 27 - Densidade de Potência para 50 (Cinquenta) metros



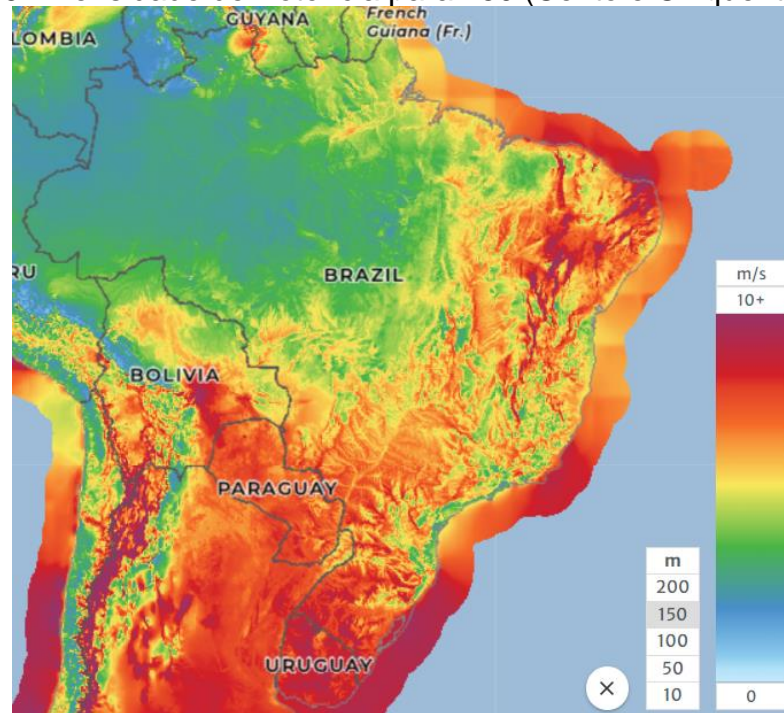
Fonte: (Global Wind Atlas, 2023).

Figura 28 - Densidade de Potência para 100 (Cem) metros



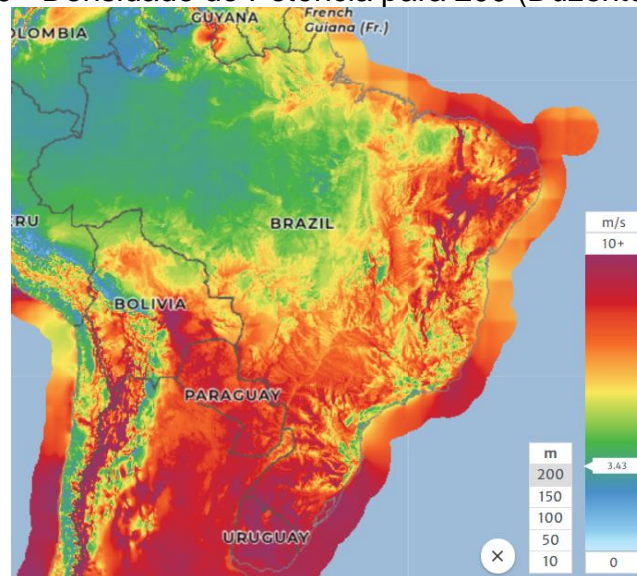
Fonte: (Global Wind Atlas, 2023).

Figura 29 - Densidade de Potência para 150 (Cento e Cinquenta) metros



Fonte: (*Global Wind Atlas, 2023*).

Figura 30 - Densidade de Potência para 200 (Duzentos) metros



Fonte: (*Global Wind Atlas, 2023*).

Como pode ser analisado pelas Figuras acima, os estados mais proeminentes para geração deste tipo de energia são: Ceará, Bahia, Rio Grande do Norte e Pernambuco, possuindo um enorme potencial e destaque para o estado do Rio Grande do Norte possuindo a maior capacidade instalada.

Além disso, essa energia demonstra grande propensão para sua produção, e geração de empregos para a região. A energia eólica vem contribuindo para renda de diversas comunidades do interior da região, garantindo o acesso e a oportunidade de elas poderem usufruir dos benefícios desta geração (COSTA, 2022).

A instalação de parques eólicos tanto em áreas rurais e em cenários *offshore*, vem trazendo olhares de diversas empresas do setor de todo o mundo. No Brasil, diversas empresas nacionais investem neste tipo de geração, sendo algumas delas: Casa dos Ventos, CPFL Renováveis e Omega Energia. Estes parques, podem possuir um sistema de armazenamento de energia, para poder garantir uma forma mais eficiente e de estabilidade no SDEE (FREIRE e FONTGALLAND, 2022).

2.4 TRABALHOS RELACIONADOS

Com o grande aumento de interesse no desenvolvimento de técnicas de machine learning para as mais diversas áreas, o setor elétrico não ficaria de fora. Alguns trabalhos atuais estudam de formas diferentes de prever o consumo elétrico de um edifício único, cidades inteiras, regiões ou até mesmo países inteiros. O trabalho de (CHEN, WANG, *et al.*, 2023) estudou a biblioteca de uma universidade pública (*YunTech*), usando o consumo elétrico total e os dados mensais de carga da biblioteca ao longo de cinco anos para construir um modelo preditivo utilizando a teoria de Fourier e regressão simples para comparações.

Em (GONZÁLEZ-BRIONES, HERNÁNDEZ e CORCHADO, 2019) é analisada a previsão de consumo de energia em uma situação mais específica (uma residência) por meio de modelos de aprendizado de máquina. Foram avaliados diversos modelos, como K-Nearest Neighbors, Regressão Linear, Árvore de Decisão, entre outros. Utilizando dados diários de consumo de uma residência com dois moradores, o estudo revelou que a Regressão Linear e a Regressão de Vetores de Suporte alcançaram uma precisão de 85,7%. A inclusão do dia anterior no processo de treinamento foi um fator contribuinte. O trabalho também propõe a expansão das variáveis no conjunto de dados, como temperatura externa e radiação solar, para análises futuras.

De forma mais abrangente, (BEDI e TOSHNIWAL, 2019) apresenta um *framework* de aprendizado profundo para previsão de demanda de eletricidade de

Chandigarh, na Índia. O modelo proposto, baseado em janelas de tempo, e múltiplas entradas e saídas, aprende ativamente as dependências nos dados históricos sem fazer quaisquer pressupostos prévios, superando outros modelos de previsão de ponta. O trabalho também oferece uma revisão da literatura sobre análise de energia e métodos de previsão. O *framework* apresenta uma solução promissora para lidar com padrões de demanda de eletricidade de longo prazo. No entanto, também são discutidos os desafios e limitações potenciais do uso de métodos baseados em IA para a previsão de demanda de eletricidade.

O trabalho realizado em (PESANTEZ, LI, *et al.*, 2023) comparou os métodos de previsão de demanda de eletricidade da área de Chicago, Estados Unidos. O estudo empregou diversos modelos, incluindo redes neurais e *Random Forest*. Os dados de entrada incluíram variáveis meteorológicas horárias de estações próximas, sendo comprovada a eficiência desta prática posteriormente. Os resultados indicaram que os modelos têm aplicabilidade prática para gerenciamento de demanda e operação eficiente do sistema elétrico. No entanto, há algumas limitações, como a falta de dados socioeconômicos detalhados.

Finalmente, em (LEME, CASACA, *et al.*, 2019), muito semelhante ao que o presente trabalho propõe, é realizado um estudo de previsão de carga elétrica na Rede Interligada Brasileira como um todo, com base em dados meteorológicos e outros dados de energia. Foram aplicados três modelos de ML e, em seguida, otimizados baseando-se em conjunto com hiper parâmetros ideais para fornecer previsões precisas diárias/mensais. O melhor modelo que os autores conseguiram foi o *Gradient Boosting* conseguindo ótimos resultados.

3.0 KNOWLEDGE DATA DISCOVERY

A descoberta de conhecimento, conhecida como KDD, representa um estudo mais aprofundado do domínio dos dados estruturados e semiestruturados. Este processo vai muito além da mera análise estatística e envolve a aplicação de técnicas avançadas para descobrir padrões ocultos, estabelecer conexões e fazer previsões informadas. O KDD amplia os limites tradicionais do aprendizado de máquina usando uma abordagem multidisciplinar. Isto inclui técnicas de visualização e análise

exploratória de dados para uma compreensão abrangente dos acontecimentos (SHU e YE, 2023). O KDD permite prever o comportamento futuro com base em registros passados e transformar dados em insights valiosos que informam a tomada de decisões estratégicas. Além disso, esta abordagem revela padrões que podem desempenhar um papel importante na compreensão e resolução de problemas complexos em vários campos.

3.1 ESTATÍSTICA DESCRITIVA

A estatística descritiva desempenha um papel fundamental no campo do aprendizado de máquina, fornecendo *insights* importantes sobre as características e distribuição dos conjuntos de dados. Inclui um conjunto de técnicas quantitativas utilizadas para resumir e compreender as variáveis que compõem um determinado conjunto de dados. Medidas de tendência central, como média, mediana e moda, e medidas de dispersão, como mínimo e máximo, desvio padrão e variância, fornecem uma visão geral abrangente da tendência central e da variabilidade dos seus dados. Além disso, percentis, quartis e métricas de distribuição fornecem informações sobre a distribuição e o formato dos seus dados e ajudam a identificar possíveis discrepâncias e anomalias (FISHER e MARSHALL, 2009).

Esses métodos estatísticos servem como uma base importante para as etapas subsequentes do processo de aprendizado de máquina. Elas ajudam na preparação e pré-processamento dos dados, permitindo decisões informadas sobre limpeza de dados, normalização e engenharia de características. Além disso, a Estatística Descritiva fornece uma compreensão preliminar das relações entre variáveis, abrindo caminho para análises e técnicas de modelagem mais avançadas.

Além disso, vale ressaltar que as suposições subjacentes a muitos métodos estatísticos podem ter um impacto significativo no desempenho e na confiabilidade dos modelos de Aprendizado de Máquina. Por exemplo, a suposição de normalidade na distribuição das variáveis é crucial em várias análises estatísticas. Compreender e validar essas suposições é uma etapa essencial para garantir a robustez e a precisão do modelo. Fundamentalmente, as estatísticas descritivas servem como um alicerce no caminho para aplicações de aprendizado de máquina, fornecendo a base

necessária para esforços de modelagem subsequentes (RAMACHANDRAN e TSOKOS, 2021).

3.2 COEFICIENTE DE VARIAÇÃO

O coeficiente de variação (CV), também conhecido como desvio padrão relativo (RSD), é um elemento fundamental da teoria das probabilidades e da estatística. Vai além da mera análise numérica, fornecendo uma medida padronizada de variação que é independente da magnitude da variável que está sendo analisada. Observe que a interpretação do desvio padrão pode ser subjetiva e depende do tamanho da variável que está sendo analisada. Nesse contexto, o CV atua como um balanceador, nivelando o campo de análise e permitindo uma compreensão mais precisa da dispersão dos dados (ABDI, 2010).

O coeficiente de variação é dado pela seguinte equação:

$$CV = \frac{s}{x} \quad (1)$$

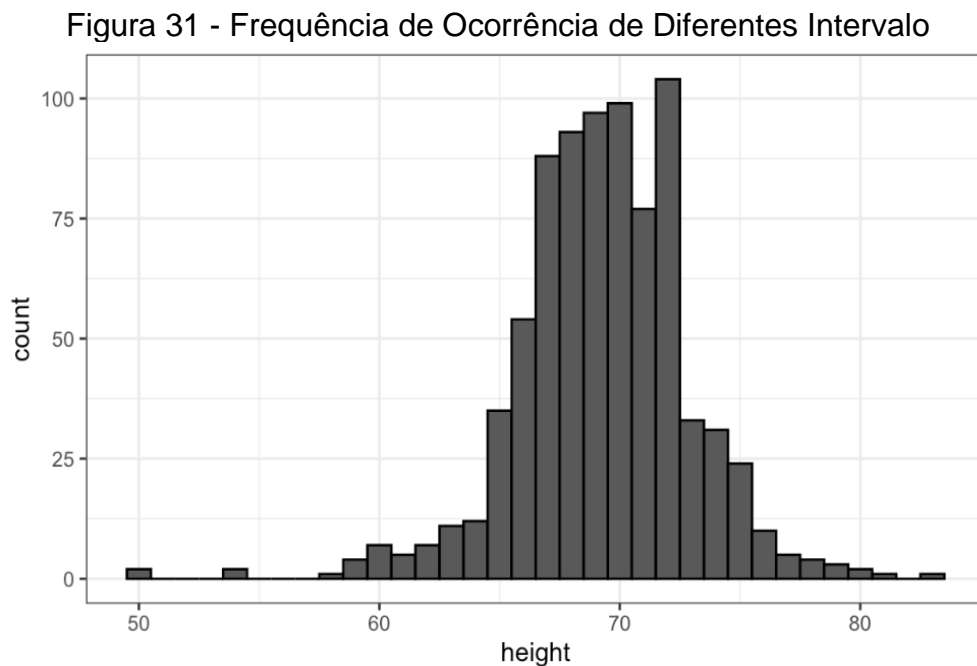
onde s é o desvio padrão e x , a média dos valores.

Além disso, o CV utiliza linguagem percentual, permitindo quantificar o desvio padrão em relação à média. Este aspecto pode ser importante em diversas aplicações, desde a avaliação de riscos financeiros até a interpretação de resultados experimentais. Embora um CV inferior a 25% indique frequentemente um conjunto de dados relativamente homogêneo, é importante notar que este limiar pode variar dependendo do contexto e do tipo de dados que estão a ser analisados. Portanto, o coeficiente de variação é uma ferramenta importante para obter *insights* valiosos que facilitam a tomada de decisões dentro da problemática.

3.3 Histograma

Um histograma é uma representação gráfica que desempenha um papel importante na análise de dados. Este tipo de gráfico fornece informações valiosas para

pesquisadores e analistas, examinando a distribuição de frequência das variáveis. Consiste em barras verticais, cada uma representando a frequência de ocorrência de um determinado intervalo de valores. Essa visualização intuitiva permite a identificação de padrões, tendências e características notáveis nos dados, tornando-se essencial para a exploração de dados e tomada de decisões fundamentadas (LAUTENCHLEGER, FLECK e STAMM, 2015).



Fonte: (IRIZARRY, 2019).

Por exemplo, ao analisar um conjunto de dados de altura populacional, um histograma como o mostrado na Figura 31 destaca a frequência de ocorrência de diferentes intervalos de altura. As barras mais altas indicam as áreas mais comuns e fornecem uma visão clara da distribuição de altura. Desta forma, os histogramas representam de forma visual e intuitiva a dispersão dos dados, permitindo ver se a distribuição é simétrica, assimétrica, unimodal ou bimodal, e identificar valores discrepantes que podem não ser óbvios à primeira vista. Esta ferramenta desempenha um papel importante em diversas áreas e contribui para uma base sólida e confiável para as conclusões tiradas da investigação e análise de dados.

3.4 Coeficiente de Correlação

O coeficiente de Pearson, também chamado de coeficiente de correlação, é uma medida importante para determinar o grau de associação linear entre duas variáveis quantitativas representado por ρ , esse coeficiente assume valores de -1 a 1, indicando correlações negativas e positivas, respectivamente. Valores próximos de zero indicam ausência de dependência linear e valores próximos de -1 ou 1 indicam relação forte. Por exemplo, se ao analisar o tempo de estudo e o desempenho de um grupo de alunos, ρ está próximo de 1, o que significa que o desempenho tende a melhorar à medida que o tempo de estudo aumenta e vice-versa. Se ρ estiver próximo de -1 indicaria que mais tempo de estudo está associado a notas mais baixas (FILHO e JÚNIOR, 2009). O coeficiente de Pearson pode ser exemplificado pela equação a seguir:

$$\rho = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}} \quad (2)$$

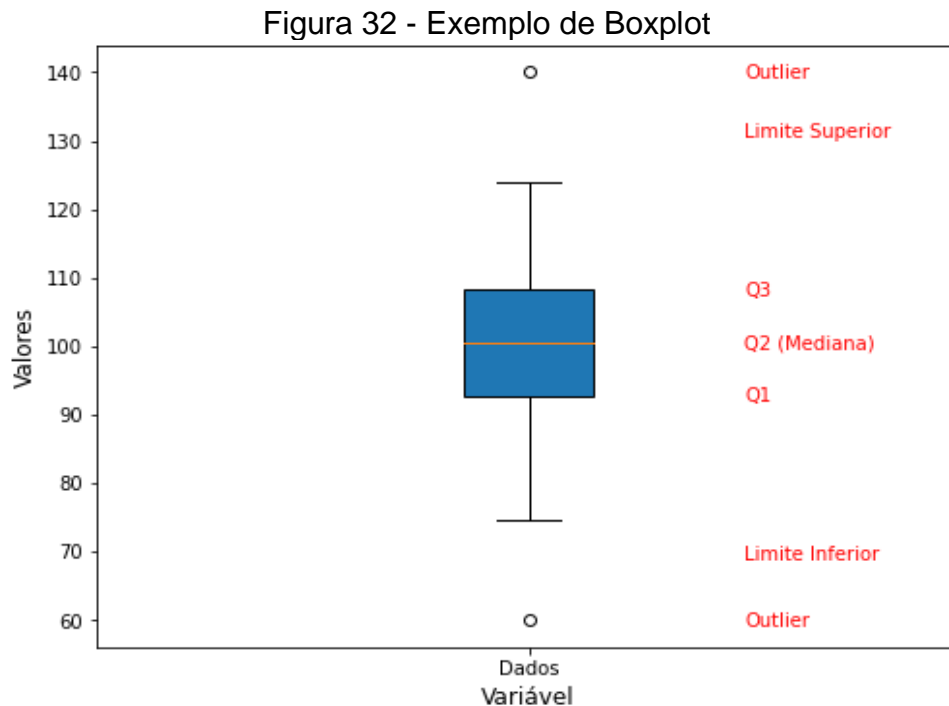
onde ρ representa o valor da correlação; X e Y são as variáveis dadas como amostras.

Para simplificar a interpretação, o grau de correlação pode ser dividido em regiões específicas. Por exemplo, uma correlação próxima de +/- 0,9 é considerada muito forte, e uma correlação entre +/- 0,7 e 0,9 é considerada forte. Uma correlação entre +/- 0,5 e 0,7 é considerada moderada e uma correlação entre +/- 0,3 e 0,5 é considerada fraca. Finalmente, uma correlação próxima de zero indica uma relação praticamente insignificante. Essa medição matemática é a base de pesquisas que visam compreender e quantificar as relações entre variáveis, fornecendo *insights* valiosos para análises estatísticas e diversos tipos de pesquisas.

3.5 BoxPlot

O gráfico *boxplot*, é uma poderosa ferramenta de análise estatística que representa visualmente a distribuição de um conjunto de dados. Consiste em cinco estatísticas principais: mínimo, primeiro quartil (Q1), mediana (Q2), terceiro quartil (Q3) e máximo. Essas estatísticas são utilizadas para determinar o intervalo da caixa

(representando os interquartis) e os "*whiskers*" (linhas que se estendem da caixa até os valores mínimo e máximo dentro do intervalo permitido (BRUNESE, MARTINELLI, *et al.*, 2020). Os outliers (dado que se desvia do resto da distribuição) são representados como pontos individuais fora dos limites dos *whiskers*. A Figura 32 exemplifica um *boxplot*:



Fonte: Elaborado pelos Autores.

Neste exemplo pode-se observar os dois limites:

$$\text{Limite Inferior: } LI = QL - k_1 C_1 \quad (3)$$

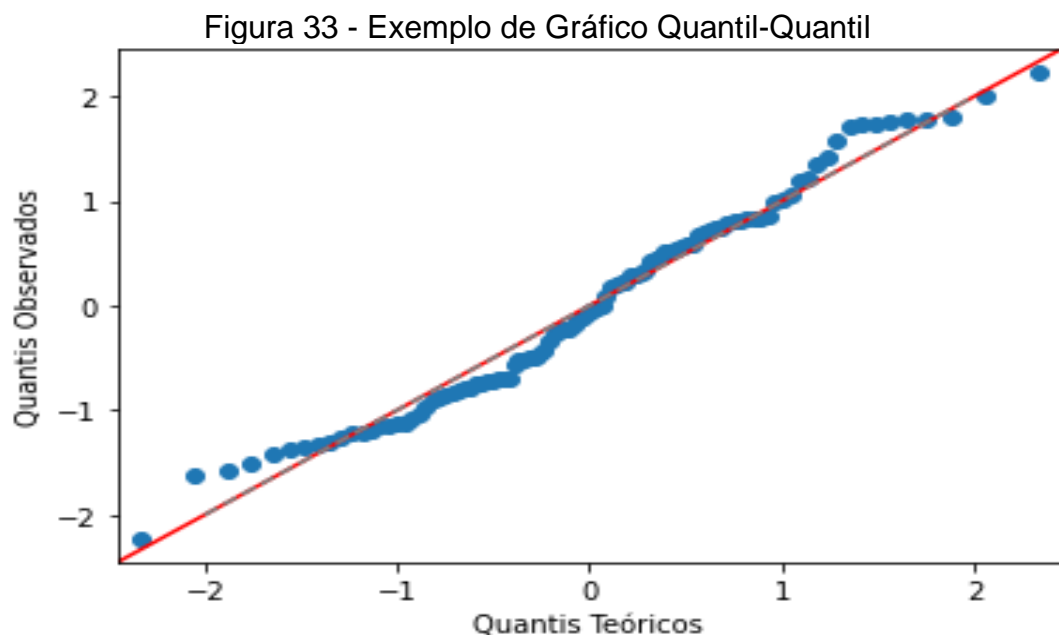
$$\text{Limite Superior: } LS = QU - k_2 C_2 \quad (4)$$

Onde $C_1 = C_2 = IQR$, e os coeficientes $k_1 = k_2 = 1,5$ (BISONG, 2019).

Os *boxplots* são uma ferramenta valiosa para identificar a propagação e a presença de outliers em um conjunto de dados, fornecendo uma visão clara das características da distribuição. A criação de um *boxplot* envolve o cálculo dos quartis (Q1, Q2, Q3) e a determinação dos limites superior e inferior do *whisker*. A fórmula de cálculo dos quartis é a base desta construção. Q1 é o valor que separa os primeiros 25% dos dados, Q2 é a mediana e Q3 é o limite para os primeiros 75% dos dados. Os *whiskers* são determinados usando uma fórmula que usa quartis e intervalos interquartis.

3.6 Q-Q PLOT

Um gráfico *quantil-quantil* (gráfico QQ) é uma ferramenta essencial na análise estatística que permite comparar os valores ordenados de uma variável com os *quantis* de uma determinada distribuição teórica, como a distribuição normal. Ao traçar esses pontos, você pode determinar se as duas distribuições correspondem. Se as distribuições coincidirem, os pontos formarão uma linha reta com inclinação 1 até a origem. Esta representação visual ajuda a avaliar quão bem a distribuição teórica se ajusta aos dados empíricos e fornece insights sobre a tendência central, dispersão e medidas de assimetria das duas distribuições (AUGUSTIN, SAULEAU e WOOD, 2012). A Figura 33 mostra um exemplo desse tipo de gráfico.



Fonte: Elaborado pelos Autores.

Além do diagrama QQ, outra métrica importante é o coeficiente de determinação, muitas vezes referido como R^2 . Este coeficiente fornece uma medida da proporção de variação na variável dependente que pode ser prevista a partir das variáveis independentes no modelo de regressão linear (CHIODE, 2021). Em outras palavras, R^2 quantifica o quão próximo os pontos de dados se ajustam à linha de regressão. Os valores variam de 0 a 1, onde 1 indica uma correspondência exata. Um R^2 maior indica que o modelo de regressão é mais capaz de explicar a variação da variável dependente com base nas variáveis independentes. Portanto, o coeficiente

de determinação desempenha um papel importante na avaliação da qualidade do ajuste de um modelo estatístico.

3.7 ENGENHARIA DE CARACTERÍSTICAS

A engenharia de características representa algumas etapas importantes no processo de preparação e modelagem de dados em pesquisas estatísticas e de aprendizado de máquina (AM). Para compreender esta abordagem, é importante primeiro delinear o conceito de variáveis, que são as características que despertam o interesse pelos dados. Em estatística, essas variáveis podem ser classificadas em dois grupos principais: variáveis contínuas e variáveis categóricas (KALIYADAN e KULKARNI, 2019).

- **Variáveis Contínuas:** Variáveis contínuas permeiam o universo da estatística e do AM, sendo capazes de assumir uma infinidade de valores em um intervalo contínuo da reta real. Dessa forma, sua natureza permite medições precisas, embora seja impraticável enumerar todos os valores possíveis dentro desse intervalo. Por meio dessas variáveis, é possível obter informações concretas e fundamentais para diversos estudos;
- **Variáveis Categóricas:** Variáveis categóricas são um tipo de variável estatística que representa diferentes categorias ou grupos. Ao contrário das variáveis numéricas, que assumem valores quantitativos, as variáveis categóricas são qualitativas e representam características que podem ser divididas em grupos distintos.

A engenharia de características desempenha um papel fundamental na formulação e seleção dos dados, modelo e elementos mais relevantes para a tarefa em questão. Esta etapa é muito importante porque a falta de variáveis que forneçam informações suficientes pode impactar negativamente o desempenho do modelo de ML. Por outro lado, ter muitas variáveis não relacionadas pode tornar o modelo computacionalmente intensivo e difícil de treinar. Portanto, a engenharia e seleção de recursos é uma etapa estratégica e crítica na construção de modelos estatísticos e de ML eficazes.

3.7.1 Normalização de Dados e Análise de Variação em *BASE LOG*

A conversão da variação em base logarítmica, conhecida como *Log Return*, desempenha um papel importante na análise estatística, pois permite a normalização das variáveis. Esta técnica permite que as variáveis sejam expressas como medidas comparáveis, geralmente percentagens. Esta é a base para avaliar relações analíticas entre diferentes variáveis, mesmo que tenham origem em conjuntos de valores contraditórios. Essa normalização é essencial para a realização de diversas análises estatísticas multidimensionais e técnicas de aprendizado de máquina, e cria uma base sólida para interpretação e comparação de dados.

A variação, denotada por r , num determinado momento i é calculada utilizando uma fórmula simplificada que leva em consideração o valor da variável naquele momento e o valor anterior da variável, como podemos ver na equação (5):

$$r_i = \frac{v_i - v_{i-1}}{v_i} \quad (5)$$

Esta equação representa uma medida da taxa de mudança ao longo do tempo e fornece informações valiosas para avaliação quantitativa e análise de tendências. Além disso, a aplicação de uma função logarítmica complementa esta abordagem, comprimindo o intervalo de números maiores e expandindo o intervalo de números menores (ZHAN, ZHAO, *et al.*, 2021). Essa transformação é importante para identificar oscilações mais significativas ao longo do período de interesse. A função logarítmica, que é o inverso da função exponencial, desempenha um papel fundamental neste processo. Isso permite uma exibição de dados mais eficaz e interpretável, tornando-os um recurso valioso na análise estatística aplicada.

Já a normalização de dados desempenha um papel importante na preparação e pré-processamento de conjuntos de dados para modelos de aprendizado de máquina. Este é um passo importante para garantir que as diversas variáveis sejam de escala comparável, evitando assim que tenham um impacto desproporcional no modelo devido ao tamanho da variável. Sem a normalização adequada, características com escalas muito diferentes podem distorcer a interpretação dos resultados e afetar a precisão das previsões.

Entre os métodos de normalização disponíveis, o *min-max scaling* é um dos mais utilizados. Esta técnica fornece uma maneira equilibrada de transformar dados

em um intervalo específico (geralmente de 0 a 1), preservando ao mesmo tempo a interoperabilidade dos dados originais (SINSOMBOONTHONG, 2022). Para calcular o valor normalizado X_{norm} de um dado X em uma variável, aplica-se a seguinte equação (8):

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (8)$$

Onde:

- X_{norm} é o valor normalizado de X ;
- X_{min} é o menor valor observado na variável;
- X_{max} é o maior valor observado na variável.

A escala Min-Max é especialmente útil quando todos os atributos têm um impacto consistente no modelo, preservando ao mesmo tempo a interoperabilidade dos dados originais. Além disso, a estrutura relativa da distribuição de dados é preservada. Esta abordagem minimiza o impacto negativo que variáveis de diferentes escalas podem ter no desempenho do modelo. Ao usar o *min-max scaling* como parte integrante da engenharia de recursos em projetos de aprendizado de máquina, os modelos ganham uma base sólida para análises e previsões precisas e são mais robustos e sensíveis às sutilezas dos dados.

3.7.2 Médias Móveis

O uso de médias móveis é uma técnica bem estabelecida em diversos campos da estatística e da análise de dados, e desempenha um papel importante na suavização de séries temporais e na identificação de tendências. Esta técnica também pode ser aplicada em um contexto de aprendizado de máquina e pode ser usada para melhorar a precisão e estabilidade do modelo. As médias móveis são divididas em dois tipos principais: média móvel simples e exponencial.

As médias móveis simples (MMS) são uma abordagem fundamental para analisar padrões e tendências em séries temporais. Esta técnica calcula a média de uma série de observações dentro de uma janela móvel ao longo do tempo (MURRAY, GABBETT, *et al.*, 2017). A MMS é particularmente eficaz na redução de ruídos e flutuações nos dados temporais e em tornar as tendências mais visíveis. No contexto

do aprendizado de máquina, as aplicações de MMS ajudam a suavizar os dados de treinamento e facilitam o reconhecimento de padrões subjacentes pelos modelos.

A equação (6) para calcular a média móvel simples no tempo t é:

$$MMS_t = \frac{X_t + X_{t-1} + X_{t-2} + \dots + X_n}{n} \quad (6)$$

As médias móveis exponenciais (MME) melhoram a abordagem MMS ao atribuir pesos decrescentes a observações mais antigas. Isto significa que as observações mais recentes têm maior influência no valor médio e a MME torna-se mais sensível às mudanças na tendência. Este comportamento é particularmente útil em situações em que as alterações recentes são mais indicativas do comportamento futuro do que os dados históricos antigos (WANG e ZHANG, 2022).

A equação 7 para calcular a média móvel exponencial no tempo t é:

$$MME_t = (X_t - MME_{t-1}) * \left(\frac{2}{n+1}\right) + MME_{t-1} \quad (7)$$

3.7.3 Transformação de Variáveis Categóricas

A transformação de variáveis categóricas é uma etapa importante no pré-processamento de dados para modelos de aprendizado de máquina. Muitos algoritmos exigem entradas numéricas, mas os dados do mundo real geralmente incluem características categóricas, como cor, tipo de produto ou dia da semana. Para incorporar essas variáveis ao modelo, é essencial realizar uma transformação eficaz que preserve a informação categórica original.

Existem vários tipos de codificação para transformação de variáveis categóricas, uma delas é a codificação de rótulos (*Label Encoding*), que atribui a cada categoria um valor único sequencial. Embora seja uma abordagem simples, é adequada para variáveis ordinais, onde existe uma relação de ordem entre as categorias. Outra técnica é a Codificação *One-Hot* (*One-Hot Encoding*), que cria uma variável binária para cada categoria, resultando em uma matriz esparsa de zeros e uns. Isso é particularmente útil quando não existe uma ordem intrínseca nas categorias. Além disso, a codificação *Helmert* e a codificação *Backward Difference* são alternativas úteis, cada uma com suas aplicações específicas (POTDAR, PARDAWALA e PAI, 2017). Cada uma dessas técnicas desempenha um papel crucial

na preparação de dados categóricos para análises de *Machine Learning*, permitindo que os modelos compreendam e utilizem eficazmente essas informações para tomadas de decisões mais precisas e confiáveis.

O *Dummy Coding* se destaca entre as técnicas disponíveis por ser um método amplamente utilizado para representar variáveis categóricas em forma numérica. Esta abordagem consiste em criar uma variável binária (também chamada de “variável *dummy*”) para cada categoria presente na variável original (DALY, DEKKER e HESS, 2016). Por exemplo, se você tiver uma variável “cores” com três categorias: vermelho, verde e azul, a codificação fictícia criará três variáveis binárias (0 ou 1) para representar cada uma dessas cores, como mostra a Tabela 1 e Tabela 2 abaixo:

Tabela 1 - Aplicação Dummy Coding

Dados	Variável Categórica
1	Verde
2	Vermelho
3	Verde
4	Azul
5	Azul

Fonte: Elaborado pelos Autores.

Tabela 2 - Variáveis Dummy

Dados	Variável <i>Dummy</i>		
	Verde	Vermelho	Azul
1	1	0	0
2	0	1	0
3	1	0	0
4	0	0	1
5	0	0	1

Fonte: Elaborado pelos Autores.

Essas variáveis binárias indicam a presença ou ausência de cada categoria, permitindo que sejam integradas de forma significativa em modelos de aprendizado de máquina. A codificação fictícia preserva as informações da categoria original para

que sejam devidamente contabilizadas na análise e na previsão. Esta técnica é importante para garantir que os recursos categóricos contribuam efetivamente para o desempenho do modelo sem comprometer a interoperabilidade dos resultados. A aplicação estratégica de codificação fictícia permite que seu modelo lide com dados categóricos de maneira eficaz, resultando em análises mais precisas e previsões robustas.

3.8 REGRESSÃO LINEAR

A regressão linear é um dos métodos mais básicos e amplamente utilizados para análise e previsão de dados. Este modelo estabelece uma relação linear entre uma variável dependente e uma ou mais variáveis independentes. O objetivo é encontrar a linha reta que melhor se ajusta aos dados (MAULUD e ABDULAZEEZ, 2020). Esta reta é representada por uma equação linear e os coeficientes são estimados por métodos estatísticos.

A regressão linear de forma mais técnica pode ser representada pela equação 9:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (9)$$

Onde:

- Y é a variável dependente que queremos prever;
- $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ são os coeficientes do modelo que representam a interceptação e os pesos associados às variáveis independentes;
- X_1, X_2, \dots, X_n são as variáveis independentes que influenciam a variável dependente;
- ϵ é o termo de erro, representando a variação não explicada pelo modelo.

A estimativa dos coeficientes pode ser obtida através do método dos mínimos quadrados, que encontra os valores de que minimizam a função de perda. Isso é feito calculando as derivadas parciais da função de perda em relação a cada coeficiente e igualando-as a zero.

Uma vez que os coeficientes são estimados, o modelo pode ser usado para fazer previsões sobre a variável dependente com base nos valores das variáveis independentes. Além disso, testes estatísticos e métricas de avaliação, como o coeficiente de determinação R^2 , podem ser empregados para avaliar a qualidade do modelo e a relevância estatística das variáveis independentes.

3.9 K-NEIGHBORS

O regressor *K-Neighbors* é um modelo de regressão baseado em instâncias que opera com base na proximidade entre os pontos de dados. Este algoritmo adota uma abordagem simples, mas poderosa, para fazer previsões. Quando um novo ponto de dados de teste é recebido, o modelo identifica os k vizinhos mais próximos no espaço de características. A proximidade é calculada usando métricas como distância euclidiana, e k vizinhos são selecionados com base nesses cálculos.

Uma vez identificados os k vizinhos mais próximos, o regressor *K-Neighbors* estima o valor do ponto de dados de teste calculando a média dos valores alvo dos vizinhos selecionados (SHABANI, SAMADIANFARD, *et al.*, 2020). Isso pode ser expresso pela equação (10):

$$\hat{y} = \frac{1}{k} \sum_{i=1}^k y_i \quad (10)$$

Onde:

- \hat{y} é a estimativa do valor alvo para o ponto de dados de teste;
- k é o número de vizinhos mais próximos;
- y_i é o valor alvo de cada vizinho selecionado.

A equação (10) mostra como *K Neighbors Regressor* calcula a média dos k vizinhos mais próximos do alvo para chegar à estimativa final do ponto de dados de teste. Essa abordagem permite que o modelo capture relacionamentos complexos nos dados, tornando-o uma opção valiosa para problemas de regressão. Escolher um valor apropriado para k é importante para otimizar o desempenho do seu modelo para um determinado conjunto de dados.

3.10 MÁQUINAS DE VETORES DE SUPORTE

O modelo baseado em máquinas de vetores de suporte (SVM) é um poderoso algoritmo de aprendizado de máquina usado para classificação e regressão. O objetivo fundamental do SVM é encontrar um hiperplano no espaço de recursos que maximize a margem entre as classes de dados. Este hiperplano é escolhido para otimizar a separação entre classes e maximizar a distância entre os pontos de dados mais próximos (chamados vetores de suporte) para cada classe. Para problemas de classificação binária, o SVM procura o hiperplano que separa de forma mais eficaz as duas classes. Por outro lado, para problemas de classificação multiclasse, são aplicadas estratégias de um contra um ou um contra todos para estender a abordagem (LOU e XIE, 2023).

A relação que rege a construção do hiperplano em um SVM linear pode ser expressa pela equação (11):

$$w^T x + b = 0 \quad (11)$$

Onde:

- w representa o vetor de pesos que define a direção do hiperplano;
- x é o vetor de características do dado de entrada;
- b é o termo de polarização (bias) que ajusta a posição do hiperplano.

Para classificar um novo ponto de dados, pode-se usar a função de decisão 12:

$$f(x) = w^T x + b \quad (12)$$

Se $f(x)$ é positivo, o ponto é atribuído à classe 1; se é negativo, o ponto é atribuído à classe -1 (MAMMONE, TURCHI e CRISTIANINI, 2009).

Um dos grandes recursos do SVM é a capacidade de lidar com conjuntos de dados complexos que não são linearmente separáveis no espaço de recursos original. Isto é feito aplicando técnicas de mapeamento a um espaço de características de alta dimensão onde a separação de classes pode ocorrer linearmente. Esses mapeamentos são executados por uma função conhecida como "kernel" que calcula um produto escalar no espaço de recursos estendido. Alguns kernels comuns incluem funções de base linear, polinomial e radial (RBF).

Além disso, o SVM é conhecido por sua capacidade de lidar com dados de alta dimensão de maneira eficaz, tornando-o uma escolha popular para aplicativos ricos em recursos. A flexibilidade e versatilidade dos SVMs os tornam ferramentas valiosas em diversos campos, desde classificação de texto até reconhecimento de padrões de imagens.

3.11 REDES NEURAIS ARTIFICIAIS

Redes neurais artificiais (RNA) representam uma abordagem computacional poderosa inspirada em estruturas neurais biológicas. Semelhante ao cérebro humano, as RNAs usam neurônios artificiais para simplificar processos neurais complexos (JIN, KIM, et al., 2022). Durante o treinamento intensivo, o modelo adquire conhecimento por meio da experiência contínua, tornando-o uma ferramenta valiosa, principalmente quando a modelagem simplificada de problemas não é possível. Estes são poderosos quando se trata de previsões em cenários complexos ou quando se analisam padrões difíceis de detectar na prática (FENG, MA, et al., 2020).

As RNAs são responsáveis por armazenar esse conhecimento adquirido nos pesos da rede e estabelecer conexões entre os neurônios. Este processo customiza o comportamento da rede, mas em alguns casos pode ser difícil compreender completamente a complexidade das relações entre esses pesos (LORENCIN, ANDELIĆ, et al., 2019). Para projetar adequadamente uma RNA, é importante seguir etapas como definição de neurônios, escolha de padrões para a rede e aplicação de algoritmos de aprendizagem baseados em parâmetros estabelecidos.

Como parte do funcionamento de uma RNA, a equação (13) para calcular o estado de um neurônio no tempo t é dada por:

$$net_i(t) = \sum_{j=1}^n x_j w_{ij}(t) \quad (13)$$

Esta equação é importante para determinar a ativação do neurônio e é comparada a um limiar (θ) para determinar a saída do neurônio. As funções de ativação, como a linear, logística e hiperbólica, moldam o estado do neurônio com

base no somatório das entradas multiplicadas pelos seus pesos, proporcionando flexibilidade ao modelo (PARVEZ, SARWAT, *et al.*, 2020).

Essas redes podem ser classificadas em três tipos. Uma delas são as redes neurais convolucionais, que são frequentemente usadas para processamento de imagens. Redes neurais recorrentes são aplicadas a problemas de linguagem e comportamentais. As redes neurais multicamadas são mais adequadas para fazer previsões em dados ordenados. Para o escopo deste TCC, optou-se pelo uso de redes neurais multicamadas *perceptron*, que dividem os neurônios em camadas de entrada, intermediária e de saída, possibilitando a resolução de problemas complexos mediante uma estrutura bem equilibrada de variáveis e uma estratégia eficiente de seleção de dados (SERIN, SENER, *et al.*, 2020).

3.12 AUMENTO DE GRADIENTE

O aumento de gradiente é uma técnica avançada de aprendizado de máquina caracterizada pela capacidade de criar preditores altamente precisos a partir de vários modelos simples, os chamados "weak learners". Isso funciona de forma iterativa, com cada modelo subsequente tentando corrigir quaisquer erros remanescentes no modelo anterior. Este processo é realizado minimizando a função de custo usando técnicas de otimização, como algoritmos de descida de gradiente. Esta abordagem fornece níveis excepcionais de precisão e generalização, tornando o Gradient Boosting uma escolha popular para muitas aplicações de previsão e regressão (BENTÉJAC, CSÖRGÖ e MARTÍNEZ-MUÑOZ, 2021).

No centro do *Gradient Boosting* está a capacidade de compreender e explorar padrões de erro residuais de modelos anteriores. A cada iteração, o modelo tenta ajustar os pesos das observações para aumentar a importância das observações anteriormente mal estimadas. Isto levou a uma série de modelos focados em corrigir certas deficiências dos modelos anteriores. A combinação desses modelos, cada um focado em uma parte diferente do espaço de recursos, resulta em um preditor robusto e preciso (SILVA, SANTOS, *et al.*, 2023).

Uma das grandes vantagens do aumento de gradiente é que, além de ser capaz de lidar naturalmente com recursos e valores categóricos ausentes, ele pode lidar com

dados de vários tipos e tamanhos. Além disso, este modelo é altamente flexível e pode ser adaptado a diferentes métricas de avaliação, tornando-o versátil para diversas aplicações, desde classificação até regressão (UPADHYAY, MANERO, *et al.*, 2021).

3.13 FLORESTAS RANDÔMICAS

Random Forest é um poderoso algoritmo de aprendizado de máquina que pode lidar com uma ampla gama de tarefas, como classificação e regressão. Diferentes conjuntos de árvores de decisão são criados, cada um treinado em uma subamostra aleatória dos dados de treinamento e um subconjunto aleatório de recursos usados para divisão. Essa abordagem de conjunto permite que florestas aleatórias capturem uma gama mais ampla de padrões e relacionamentos nos dados, resultando em um modelo mais robusto e menos sujeito a *overfitting* (LI, HAN, *et al.*, 2019).

A força de uma floresta aleatória está na diversidade das árvores que a compõem. Cada árvore é construída de forma independente, o que significa que diferentes subamostras e subconjuntos de características são utilizados em cada uma delas. Isso cria modelos que exploram diferentes aspectos dos seus dados, permitindo que os conjuntos capturem nuances e complexidades que podem ser perdidas em um único modelo. Além disso, o “processo de ajuste” entre árvores ajuda a reduzir o impacto de outliers e erros individuais, contribuindo para a robustez do modelo final (PUNIA, NIKOLOPOULOS, *et al.*, 2020).

Outra vantagem importante das florestas aleatórias é que elas podem fornecer uma avaliação da importância das características no processo de tomada de decisão. Ao avaliar o ganho de impurezas ou a redução de erros, o algoritmo atribui a cada variável um valor de importância que indica sua contribuição para a precisão do modelo. Isto não só fornece informações valiosas sobre quais recursos são mais úteis, mas também facilita a interpretação e explicação do modelo para as partes interessadas e especialistas no assunto (KURDI, AMAKHCHAN e GHARINEIAT, 2021). Random Forest também é conhecida por sua eficiência computacional e escalabilidade. Ao distribuir o treinamento em árvore em paralelo, pode-se processar grandes conjuntos de dados e otimizar o uso dos recursos computacionais disponíveis. Essa capacidade de processar grandes quantidades de dados tornou as

florestas aleatórias uma escolha popular para aplicações onde o dimensionamento de dados é um desafio, como análise de big data e problemas de aprendizagem em nível empresarial.

3.14 MÉTRICAS DE VALIDAÇÃO

Ao avaliar a performance de modelos de Machine Learning, é crucial empregar métricas de validação que proporcionem uma medida objetiva e significativa do quão bem o modelo está se saindo. Três métricas amplamente utilizadas para essa finalidade são o Erro Absoluto Médio (MAE), o Erro Quadrático Médio (MSE) e o Erro Percentual Absoluto Médio (MAPE).

O MAE é uma métrica que calcula a média das diferenças absolutas entre as previsões do modelo e os valores reais (KHULLAR e SINGH, 2021). Sua equação (14) é dada por:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (14)$$

Onde y_i representa os valores reais, \hat{y}_i as previsões do modelo e n o número de observações.

Por sua vez, o MSE mede a média dos quadrados das diferenças entre as previsões e os valores reais (PRAYUDANI, HIZRIADI, *et al.*, 2019). Sua equação (15) é expressa por:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (15)$$

O MAPE, por sua vez, é particularmente útil em contextos de previsão, pois representa a percentagem média em que as previsões do modelo estão distantes dos valores reais (KHULLAR e SINGH, 2021). Sua equação (16) é definida como:

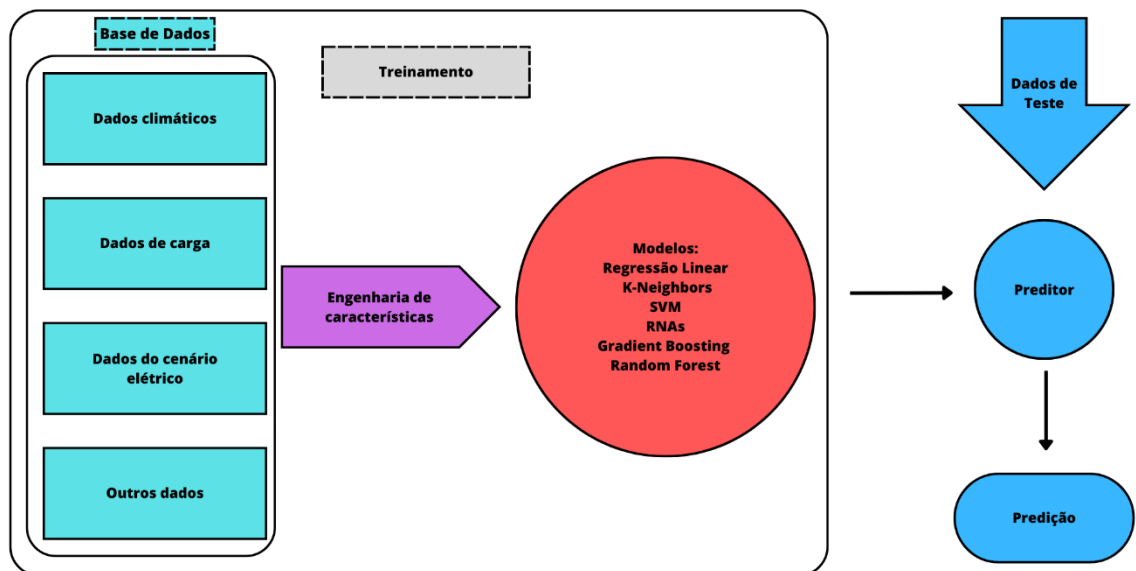
$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (16)$$

Assim, foi possível avaliar o efeito das técnicas de previsão e da infraestrutura computacional desenvolvida, especialmente no que se refere à capacidade de previsão, conformidade com os dados, e detecção de padrões e comportamentos.

4 METODOLOGIA E MATERIAIS DE USO

Esta seção detalha as abordagens metodológicas e os recursos utilizados para implementar o processo de aprendizado de máquina. Para facilitar a visualização, segue o fluxo de trabalho, apresentado na Figura 34, que ilustra esquematicamente o processo de aprendizado de máquina desde a preparação dos dados até a obtenção dos resultados da previsão. Este fluxograma serve como um guia básico para entender a implementação real das técnicas aplicadas no presente estudo.

Figura 34 - Fluxograma da Metodologia Implementado no Projeto



Fonte: Elaborado pelos Autores.

4.1 FERRAMENTAL UTILIZADO

No contexto do projeto, a aplicação de técnicas de Machine Learning para a predição de carga representa um avanço significativo. Utilizando o ambiente *Python*, com destaque para a integração do *Anaconda/Spyder*, foi possível desenvolver e avaliar modelos preditivos de forma eficiente. A escolha da biblioteca *Scikit-learn* se

mostrou crucial, dada sua versatilidade e robustez em lidar com problemas de séries temporais.

O *Scikit-learn* é uma ferramenta muito relevante para o projeto, oferecendo uma ampla gama de algoritmos supervisionados e não supervisionados. Uma interface consistente e facilidade de uso permitem a prototipagem ágil de modelos preditivos (KRAMER, 2016). Além disso, o suporte a diversas operações e manipulações de dados permitiu uma análise detalhada dos resultados obtidos.

A base técnica do projeto foi fortalecida através da integração do *Scikit-learn* com outras bibliotecas estabelecidas, como *NumPy*, *SciPy*, *Pandas* e *Matplotlib*. Essa variedade, proporcionou um ambiente robusto e confiável para executar todo o projeto, permitindo explorar os dados, construir modelos e avaliá-los de forma sistemática e eficaz (FUHRER, SOLEM e VERDIER, 2021). Portanto, esta escolha de infraestrutura acabou sendo a base para o sucesso da aplicação de técnicas de aprendizado de máquina na previsão de carga.

4.2 BASE DE DADOS

A fundamentação deste estudo foi construída sobre duas fontes de dados primordiais: o Operador Nacional do Sistema (ONS) e o Instituto Nacional de Meteorologia (INMET). Estas bases de dados apresentam informações cruciais para a previsão de carga no contexto do Nordeste brasileiro. O Operador Nacional do Sistema Elétrico (ONS) é uma entidade de grande relevância para o setor elétrico brasileiro (ONS, 2023). Trata-se de uma organização responsável por coordenar e supervisionar as atividades do Sistema Interligado Nacional (SIN), visando garantir a segurança e a confiabilidade no fornecimento de energia para todo o país. Sua atuação abrange desde a geração até a transmissão de energia elétrica, sendo fundamental para o funcionamento eficaz e estável do sistema.

O ONS desempenha diversas funções cruciais, incluindo o planejamento operacional, a programação da geração de energia, o controle da operação em tempo real e a fiscalização do mercado de energia elétrica. Além disso, é responsável por monitorar as condições operativas do sistema, prever a demanda de energia e coordenar a operação de usinas geradoras em todo o território nacional.

Dessa forma, o ONS desempenha um papel essencial na manutenção do equilíbrio e na eficiência do sistema elétrico brasileiro.

A partir da base de dados do ONS, foram coletados registros essenciais para a análise e previsão da carga na região Nordeste. Estes registros englobam informações vitais, tais como:

- Carga (GWh);
- Energia Armazenada (GWh);
- Geração Eólica (GWh);
- Geração de Energia (GWh);
- Intercâmbio de Energia (GWh);
- Demanda Máxima.

Esses dados detalhados fornecem um panorama abrangente das atividades energéticas na região, possibilitando análises mais refinadas e previsões mais precisas.

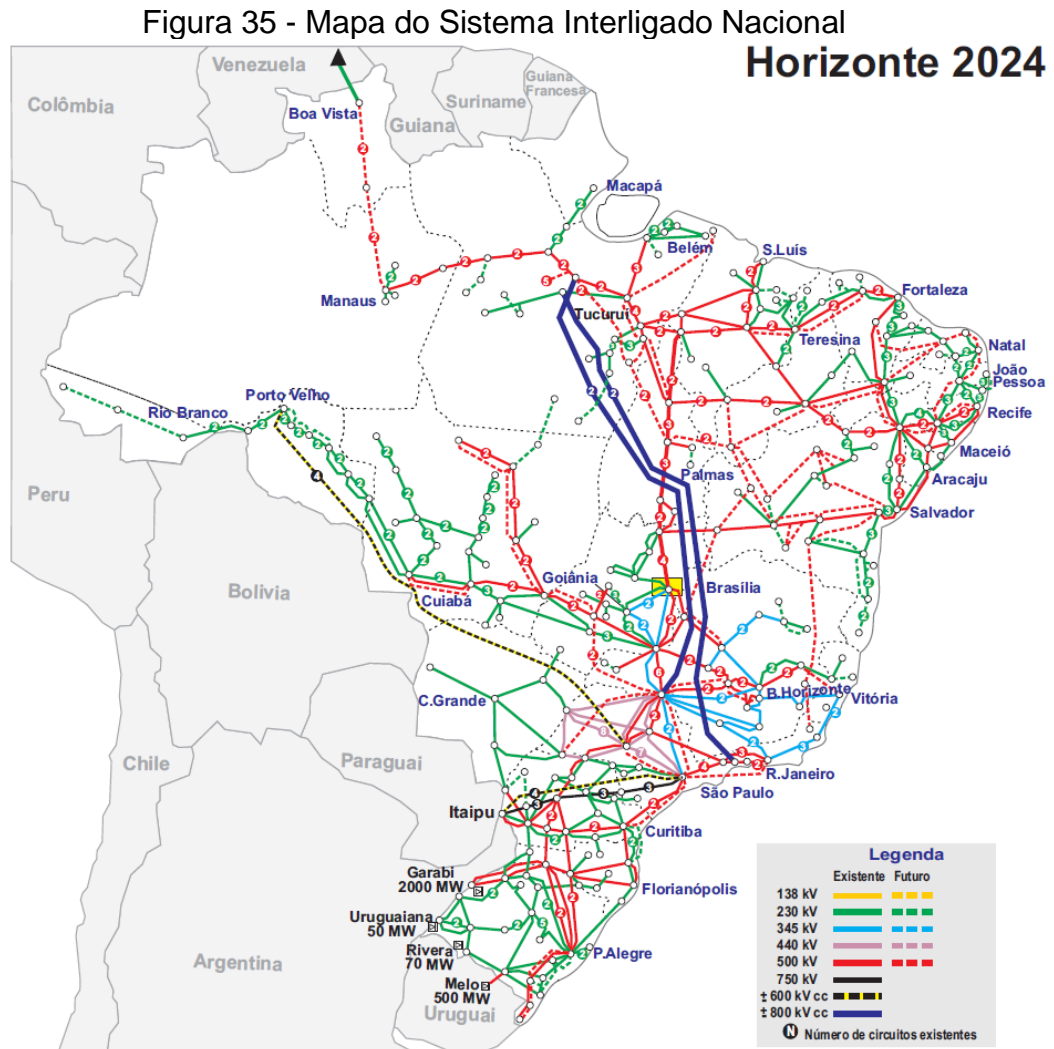
Além dos registros obtidos do ONS, os dados meteorológicos foram adquiridos junto ao Instituto Nacional de Meteorologia (INMET) (INMET, 2023). Nessa instância, foram coletadas informações provenientes das nove capitais nordestinas. Estas abrangem dados referentes à:

- Precipitação (mm);
- Pressão atmosférica (mB);
- Temperatura média (°C);
- Umidade relativa média (%).
- Velocidade média do vento (m/s).

Tais informações são cruciais para compreender o contexto climático no qual as operações energéticas ocorrem na região, proporcionando assim um embasamento sólido para a previsão de carga.

A utilização de variáveis relacionadas ao setor eólico, se deve ao fato de a matriz energética do Nordeste do Brasil ser influenciada por esse setor, que desempenha um papel preponderante na geração de eletricidade na região. A vasta extensão costeira e as condições climáticas favoráveis fazem do Nordeste uma das áreas mais propícias para a produção de energia eólica no país. Com muitos parques eólicos instalados ao longo do litoral nordestino, a região se destaca como líder nacional na geração de energia a partir dos ventos (GOUVÊA e SILVA, 2018).

Na Figura 35 é possível observar a abrangência do SIN, bem como as linhas de transmissão existentes e futuras aquisições:



Fonte: (ONS, 2023).

5 RESULTADOS E DISCUSSÃO

Esta análise centra-se numa compreensão mais profunda de cada questão através da aplicação de estatísticas descritivas. O objetivo principal é explorar padrões, tendências, anomalias e possíveis relações entre dados para obter *insights* valiosos. Para atingir esse objetivo, utilizou-se a análise exploratória de dados (AED), que permite uma visão mais clara das variáveis presentes na base de dados em questão.

5.1 DESCOBERTA DE CONHECIMENTO

A abordagem inicial centra-se em estudos descritivos quantitativos. Este foco permite uma análise detalhada das diversas variáveis existentes no banco de dados. A Tabela 1 mostra os principais indicadores estatísticos como média, desvio padrão, valores mínimo e máximo, quartis de 25%, 50% (mediana) e 75%. Estes indicadores desempenham um papel importante na análise subsequente e fornecem uma base sólida para a compreensão dos dados.

Uma análise detalhada do coeficiente de variação revelou que o percentual foi de 11,08%. Isso indica que as flutuações do consumo de energia elétrica no Nordeste são uniformes. Além disso, ao examinar os valores máximos e mínimos da variável alvo, foi possível analisar que a variação entre esses valores segue uma distribuição normal, girando em torno de 3 vezes o valor de desvio padrão acima e abaixo da média. Em particular, a proximidade da média e da mediana indica que a distribuição dos dados é suficientemente simétrica, tornando a análise realizada mais fiável.

Tabela 3 - Análise Estatística Descritiva da Variável Target

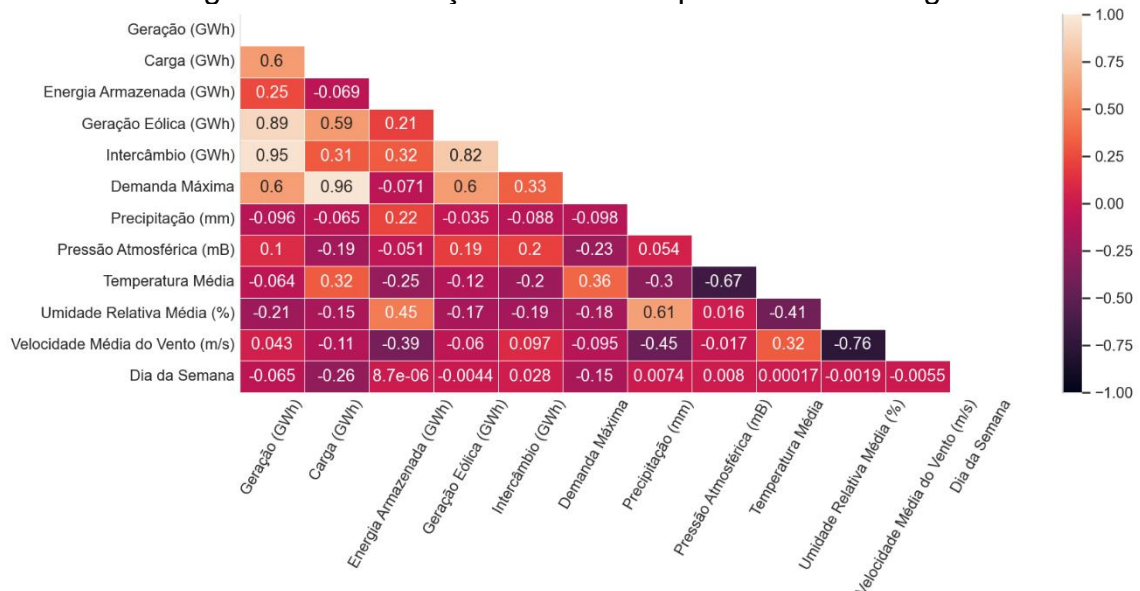
Carga (GWh)	
Média	239,04
Desvio Padrão	26,50
Valor Mínimo	160,85

Valor Máximo	307,87
25%	220,73
50%	239,52
75%	258,28

Fonte: Elaborado pelos Autores.

A análise de correlação de Pearson revelou relações significativas entre as variáveis do banco de dados. Variáveis relacionadas à demanda de energia elétrica, como a demanda máxima e a geração para atender essa demanda, apresentaram correlação positiva significativa com a variável de interesse, carga (GWh). Além disso, a geração elétrica eólica apresentou uma correlação positiva, refletindo o seu impacto no fornecimento de energia. Por outro lado, fatores meteorológicos, como velocidade do vento e umidade relativa, apresentaram correlação negativa com a carga, indicando uma interação inversamente proporcional. Estas correlações, que podem ser observadas na Figura 36, fornecem informações valiosas sobre a dinâmica dos sistemas energéticos e auxiliam a compreender as influências que afetam o consumo de energia.

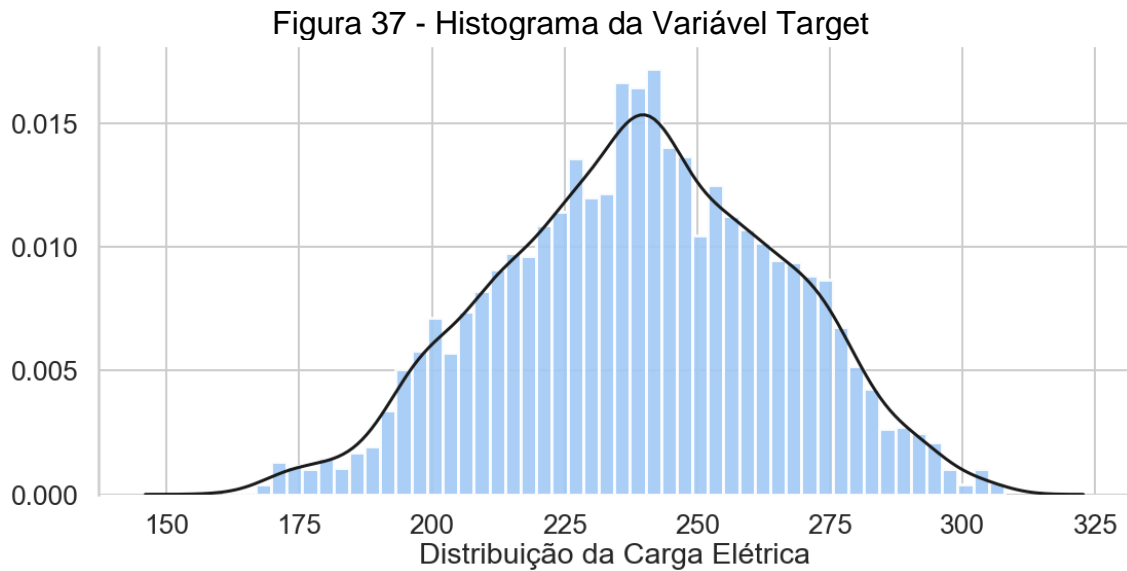
Figura 36 - Correlação de Pearson para Variável Target



Fonte: Elaborado pelos Autores.

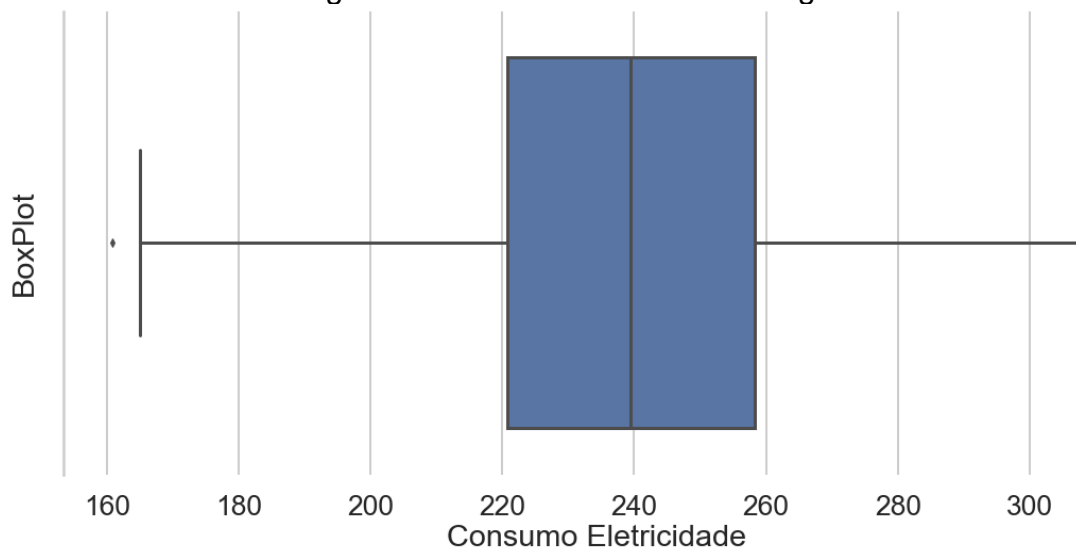
Analisando o histograma da carga nordestina abaixo na Figura 37, é possível observar a presença de um pico de frequência na faixa de 225 a 250 GWh, indicando

o consumo mais frequente na região. Também é possível analisar que não há outros picos, apontando para uma média de consumo consolidada, levando em consideração que a base contém dados de 2011 a 2022.



De acordo com o gráfico *BoxPlot* na Figura 38, observa-se um *outlier* abaixo do limite inferior. Investigando a base de dados, foi verificado que esse outlier se encontra no início da base (2011), onde a carga do nordeste era consideravelmente menor do que atualmente. Concluindo que o tal dado não representará danos aos futuros modelos que serão desenvolvidos em breve.

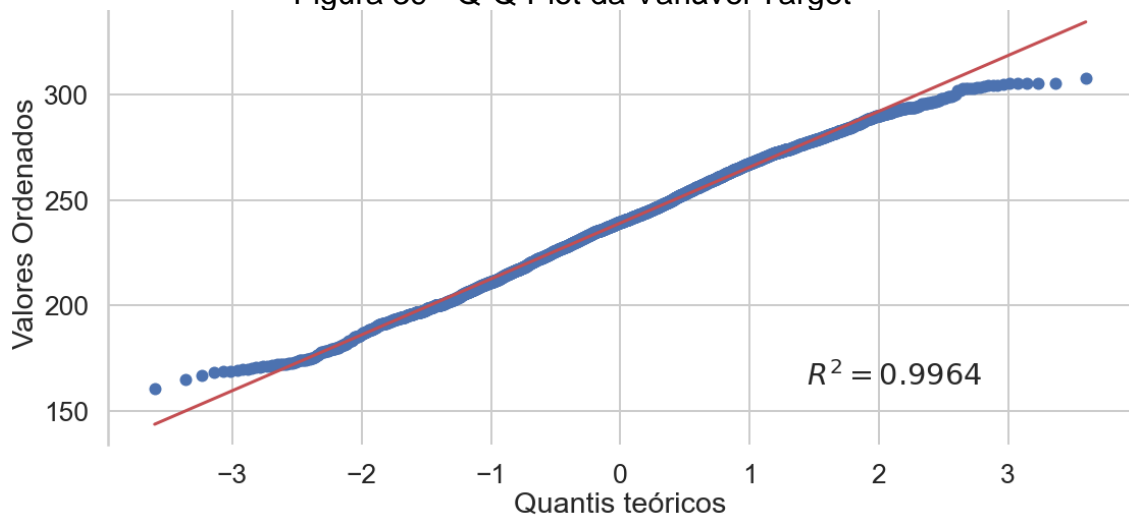
Figura 38 - BoxPlot da Variável Target



Fonte: Elaborado pelos Autores.

Ao examinar o gráfico Q-Q (gráfico *quantil-quantil*), na Figura 39, destaca-se a impressionante relação linear entre as distribuições quartis. Visualmente, é possível observar valores acima da linha que indicam uma relação positiva entre as distribuições quartis. A confirmação desta relação é feita pelo coeficiente de determinação calculado, que apresenta um valor de 0,9964. Este coeficiente reflete a elevada similaridade entre as distribuições avaliadas, e enfatiza a consistência e proximidade dos dados analisados.

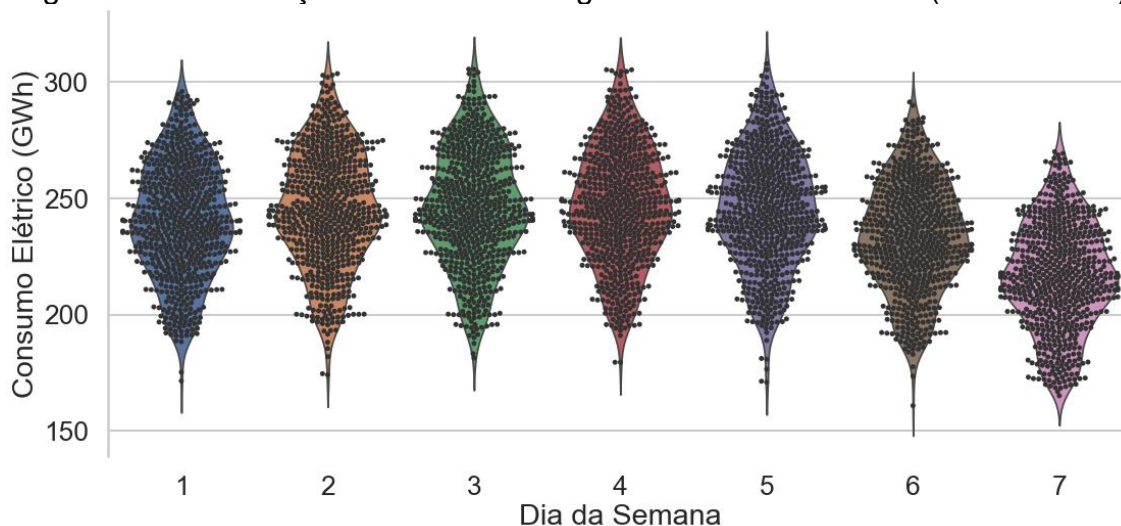
Figura 39 - Q-Q Plot da Variável Target



Fonte: Elaborado pelos Autores.

A seguir, é mostrado a distribuição da variável Carga (GWh) ao longo dos dias da semana (Figura 40) e dos meses (Figura 41). Onde 1 – Segunda-Feira; 2 – Terça-Feira; 3 – Quarta-Feira; 4 - Quinta-Feira; 5 – Sexta-Feira; 6 – Sábado; 7 – Domingo).

Figura 40 - Distribuição da Variável Carga Durante as Semanas (2011 a 2022)

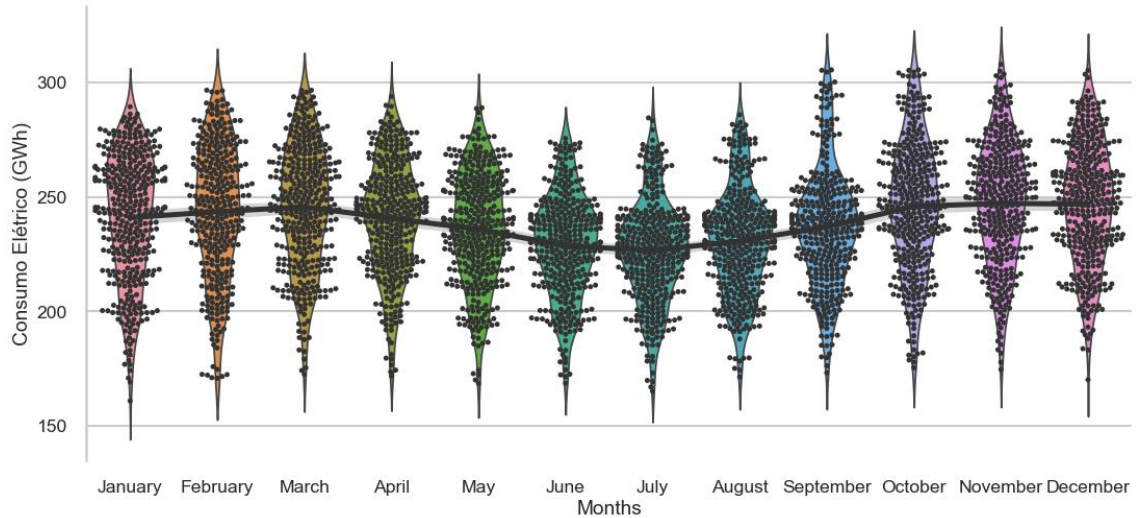


Fonte: Elaborado pelos Autores.

Ao analisar os dados sobre o consumo de eletricidade no Nordeste, algumas tendências interessantes emergem. De acordo com a média da carga variável alvo (GWh) nos dias de semana, observa-se que a demanda de energia elétrica costuma ser maior nos dias úteis do que nos finais de semana. Este padrão sugere uma ligação entre a atividade industrial e comercial e o consumo de energia, indicando que as indústrias e as empresas contribuem significativamente para o pico de consumo semanal.

Observando as médias anuais das variáveis, fevereiro, março e abril também se destacam como meses de alto consumo. A coincidência deste período com a transição da estação verão para o outono no Brasil sugere uma ligação entre as condições climáticas e o aumento do uso de sistemas de refrigeração, o que contribui para o aumento da demanda por eletricidade. Esta análise sazonal fornece informações valiosas sobre os fatores que influenciam a variação do consumo ao longo do ano.

Figura 41 - Distribuição da Variável Carga Durante os meses (2011 a 2022)



Fonte: Elaborado pelos Autores.

5.2 ENGENHARIA DE CARACTERÍSTICAS

Após a aplicação do KDD, duas técnicas foram utilizadas para gerar novas variáveis a partir das variáveis existentes. Essa abordagem é particularmente relevante para problemas de séries temporais, pois permite capturar informações e nuances adicionais nos dados que podem ser importantes para a precisão das previsões. Primeiramente, foi aplicado o retorno na escala logarítmica (*log return*) para períodos diários. Essa transformação fornece uma visão mais detalhada da variação dos dados. Adicionalmente, foi aplicada a técnica de médias móveis simples e exponencial considerando períodos diários e semanais. Essas novas variáveis fornecem informações mais sofisticadas e permitem que os modelos de aprendizado de máquina capturem padrões e tendências de maneira mais precisa e confiável.

A vetorização de variáveis categóricas é uma etapa importante na preparação de dados para modelagem. Neste contexto, as variáveis categóricas são o dia da semana e o mês. Em sua forma original, essas variáveis representam diferentes categorias com valores discretos, como os 7 dias da semana ou os 12 meses do ano. No entanto, para uma utilização eficaz em modelos de aprendizagem, é importante converter o modelo em uma representação numérica. Para isso, foram utilizadas técnicas de vetorização. Esse método consiste em criar um vetor binário para cada

categoria, onde cada posição do vetor representa uma categoria possível. A Seção 0 (*DUMMY CODING*) representa visualmente esse processo e mostra a conversão de variáveis categóricas em vetores binários.

5.3 SEPARAÇÃO DOS DADOS E NORMALIZAÇÃO

Como parte do projeto, foi importante dividir o banco de dados em dois conjuntos diferentes: um conjunto de treinamento e um conjunto de testes. O primeiro representa 90% do conjunto de dados disponível e foi utilizado para permitir a adaptação do modelo de aprendizado de máquina ao problema. Este conjunto abrangeu o período de 2011 a 2021 e serviu como base para o treinamento. Enquanto isso, o conjunto de testes, representando 10% dos dados restantes, cobriu todas as amostras de 2022. Esta análise cuidadosa é crítica para avaliar com precisão a eficácia do modelo.

Após a divisão dos dados, uma etapa importante foi aplicar o processo de normalização. Esta técnica desempenha um papel fundamental na preparação de dados para o treinamento de modelos. A padronização de todas as variáveis em uma escala de 0 a 1 visa reduzir os efeitos das diferentes unidades de medida entre as variáveis. Isto é especialmente importante para garantir que os modelos de aprendizagem automática sejam robustos e não favoreçam variáveis com escalas desequilibradas.

5.4 AVALIAÇÃO DA PERFORMANCE DOS MODELOS

Após a implementação do pipeline de pré-processamento e engenharia de recursos, os dados foram preparados para o treinamento do modelo de aprendizado de máquina (AM) proposto neste estudo. Utilizou-se a biblioteca *Scikit-learn* em *Python*, com configurações padrão para os parâmetros de cada modelo. Essa abordagem facilitou a prototipagem e os testes dos modelos. Métricas robustas, como

erro médio absoluto (MAE), erro quadrático médio (MSE) e erro médio absoluto percentual (MAPE), foram utilizadas para avaliar a eficácia do modelo. Essas métricas fornecem informações importantes sobre a precisão das previsões em relação aos valores reais da carga elétrica no contexto nordestino. Através de análises específicas para diferentes períodos de previsão, é possível entender o desempenho de cada modelo em diferentes cenários.

Além de avaliar o desempenho, foi fundamental analisar a importância das variáveis preditoras. Ao identificar os fatores que têm maior impacto na previsão da carga elétrica, compreende-se mais profundamente o problema. Essas informações podem não apenas validar o modelo, mas também orientar futuras estratégias de otimização e fornecer informações valiosas sobre o problema. Como mostra a Tabela 4, o MAPE foi calculado para cada mês, a fim de mostrar qual período os modelos tiveram mais dificuldades ou previram a carga elétrica de forma mais precisa.

Tabela 4 - MAPE (%) Média Mensal da Carga Prevista para o Ano de 2022

Mês	LR	KNR	RF	GB	SVR	MLP
Janeiro	2.386	2.900	2.245	2.277	2.711	2.485
Fevereiro	1.460	4.786	1.521	1.615	3.273	1.876
Março	2.462	4.079	2.481	2.380	3.101	2.337
Abril	1.450	3.874	1.283	1.547	3.644	1.584
Mai	2.067	3.502	2.095	2.262	3.403	2.262
Junho	3.322	4.004	3.749	2.761	2.950	2.446
Julho	1.966	2.010	1.934	1.823	2.058	2.698
Agosto	2.782	3.041	1.835	1.815	1.744	3.292
Setembro	3.078	4.851	2.302	2.724	3.071	3.074
Outubro	2.034	3.024	1.873	1.854	2.595	2.482
Novembro	2.296	1.740	2.875	2.814	2.395	2.755
Dezembro	2.697	2.975	2.864	2.903	3.294	3.148

Fonte: Elaborado pelos Autores.

De acordo com os dados obtidos, o melhor mês para o LR, RF, GB e MLP foi o mês de abril, com MAPE's indo de 1,283% a 1,584%. Já para o SVR o melhor mês foi agosto, com 1,744%. O KNR obteve 1,740% de MAPE em seu melhor mês (novembro). Os piores meses variaram muito de modelo para modelo, destacando-se o GB que obteve o MAPE abaixo dos 3% mesmo em seu pior mês (dezembro).

Com essas informações conclui-se que dos 6 modelos, 4 deles obtiveram seu melhor resultado no mês de abril, o que indica uma possível situação recorrente na base de dados que reflete isso. Ao analisar a Figura 14 é possível observar que abril é mês que a carga elétrica na região Nordeste começa a diminuir, podendo ser uma boa explicação para a ótima predição nesse período.

Para de fato selecionar o melhor modelo para futuras análises, foi construída a Tabela 5, que mostra todas as métricas calculadas para os modelos estudados.

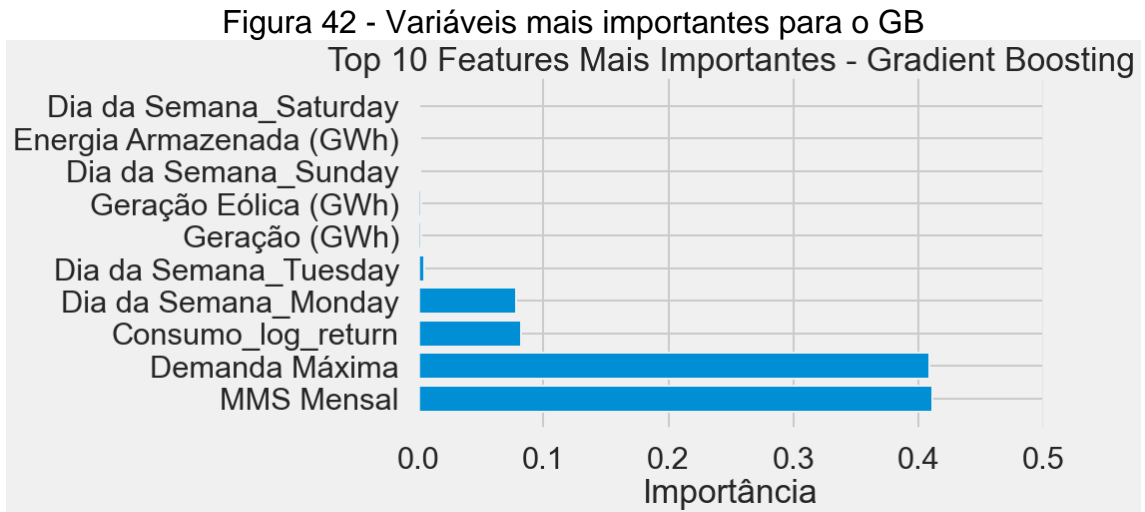
Tabela 5 - MAE (GWh), MSE (GWh²) e MAPE (%) ao longo do Ano de (2022).

Modelo	MAE (GWh)	MSE (GWh ²)	MAPE (%)
LR	6.142	61.726	2.338
KNR	9.124	129.466	3.385
RF	5.969	60.434	2.257
GB	5.932	56.026	2.234
SVR	7.703	93.139	2.848
MLP	6.732	71.300	2.543

Fonte: Elaborado pelos Autores.

A partir da análise da tabela acima, conclui-se que o melhor modelo, de acordo com todas as métricas adotadas, foi o GB, obtendo 5,932 GWh de MAE, 56,025 (GWh²) de MSE e 2,234% de MAPE.

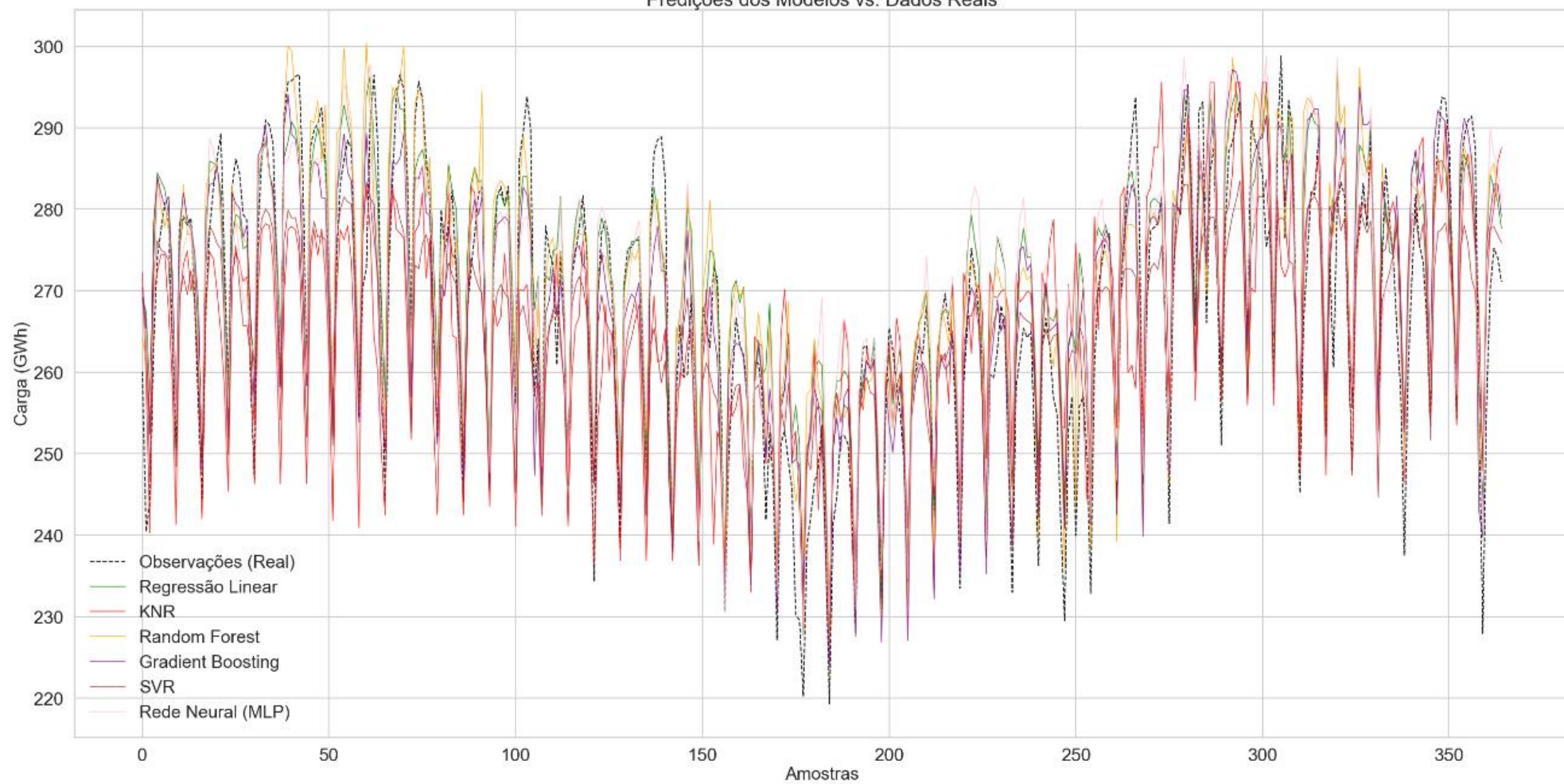
Com a Figura 42 é possível notar a importância das variáveis artificiais, pois a variáveis de mais importância para o melhor modelo testado foi a “MMS Mensal”, que indica a média móvel simples com janela de 30 dias. Logo em seguida a variável “Demanda Máxima”, e “Consumo_log_return”. Também há duas variáveis referentes a data da coleta dos dados de carga.



Fonte: Elaborado pelos Autores.

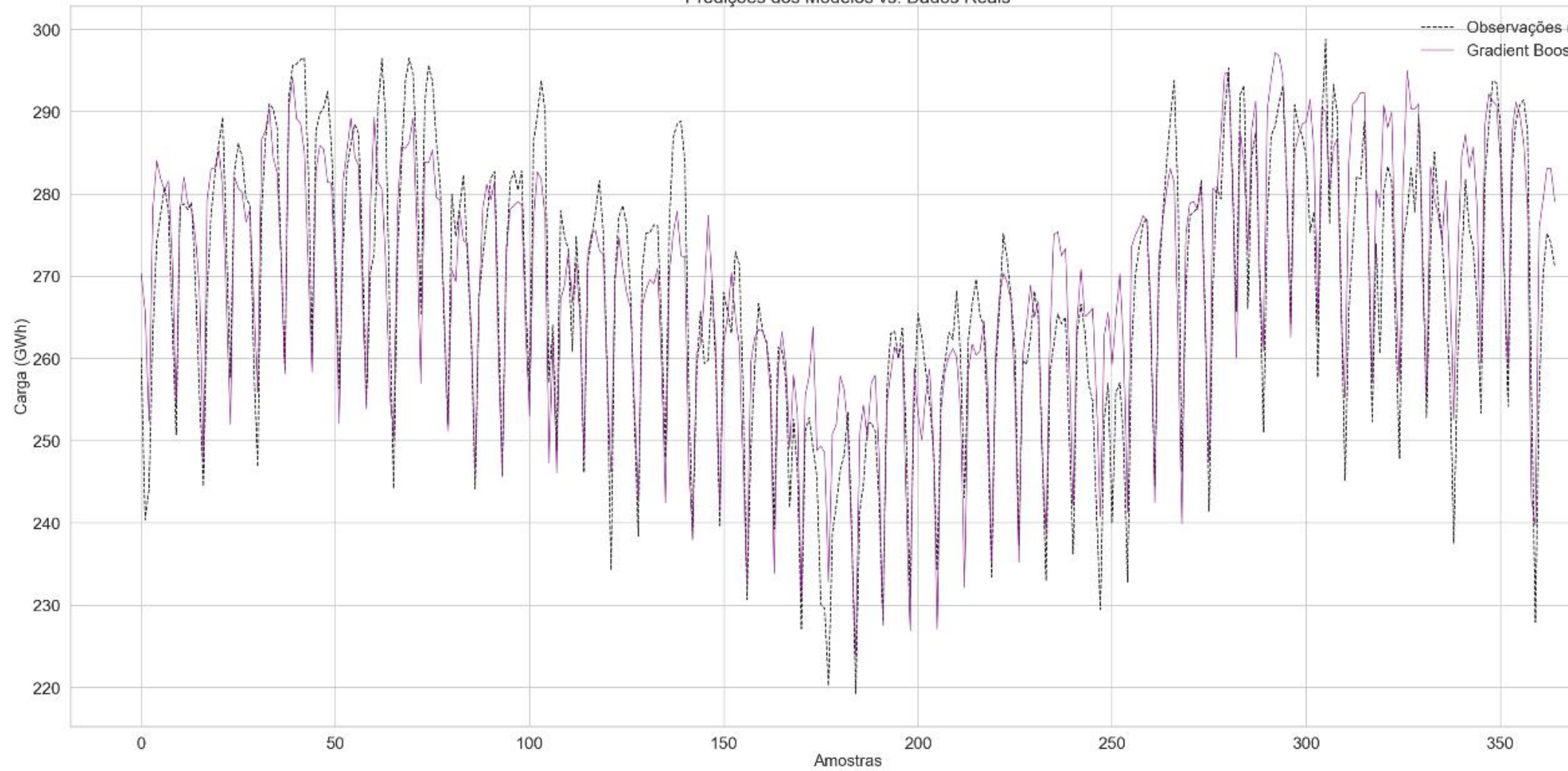
Na Figura 43, foram compiladas todas as previsões dos modelos analisados, juntamente com os valores reais do consumo de energia elétrica para todo o ano de 2022. Analisando o gráfico, há uma boa precisão de todos os modelos. Para fins de análise do melhor modelo, foi compilada a Figura 44, onde são mostradas apenas as previsões de carga do modelo GB em comparação com os dados reais.

Figura 43 - Gráfico das Predições para o Ano de 2022
Predições dos Modelos vs. Dados Reais



Fonte: Elaborado pelos Autores.

Figura 44 - Gráfico das Predições feitas pelo GB no Ano de 2022
Predições dos Modelos vs. Dados Reais



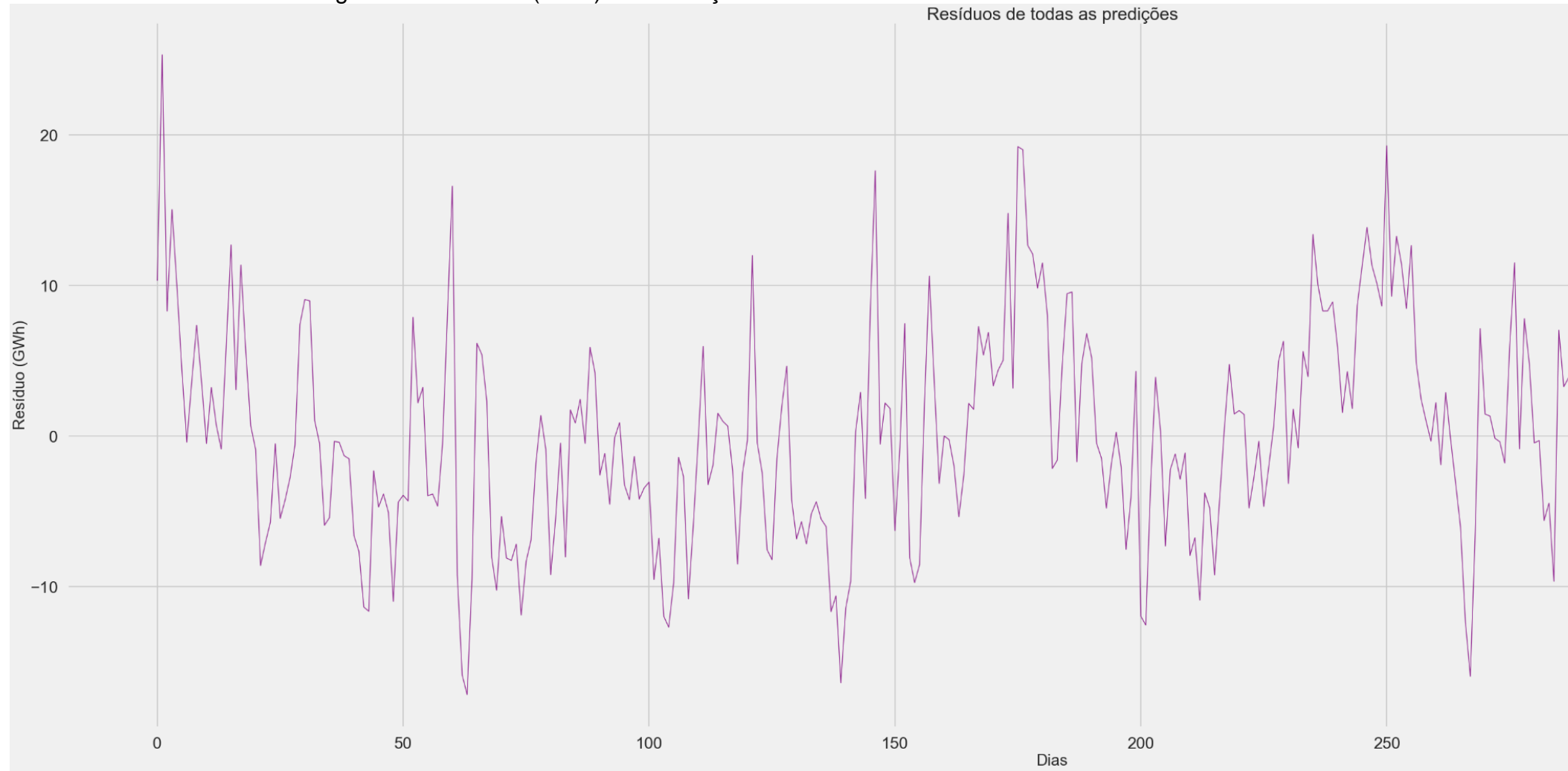
Fonte: Elaborado pelos Autores.

Com a Figura 44 fica mais claro de ver a melhor precisão adquirida entre os modelos. O GB desempenhou um bom papel nas previsões, conseguindo captar tanto a sazonalidade dos dados quanto a tendência durante o ano de 2022.

Isso segue com a Figura 45, que mostra os resíduos gerados pelo modelo GB com as previsões. Esse gráfico foi construído subtraindo os valores das previsões dos valores reais.

Figura 45 - Resíduo (GWh) das Predições do Modelo GB no Ano de 2022

Resíduos de todas as predições



Fonte: Elaborado pelos Autores.

Agora com esse gráfico (Figura 45), podemos ver que o erro do modelo em questão ficou entre 25 GWh e -15 GWh, o que demonstra uma boa assertividade do modelo, que por sua parte gerou valores condizentes com aqueles coletados (reais).

6 CONCLUSÃO

Este estudo adotou uma abordagem rigorosa para desenvolver e aplicar técnicas avançadas de aprendizado de máquina (AM) para previsão de cargas elétricas. Primeiro, foi fundamental criar um banco de dados que integrasse informações do Operador Nacional do Sistema Elétrico (ONS) e do Instituto Nacional de Meteorologia (INMET). Esta base de dados representou um marco importante no contexto da previsão de carga energética, proporcionando um conjunto de informações valiosas para o estudo.

A fase de descoberta de conhecimento (*Knowledge Data Discovery*) proporcionou uma compreensão quantitativa do comportamento das variáveis presentes na base de dados. A limpeza de dados removeu anomalias e garantiu a qualidade das informações usadas para modelar algoritmos de AM. Além disso, a aplicação de técnicas de engenharia de recursos foi essencial para criar variáveis a partir das existentes. Isto fortaleceu os modelos e contribuiu para a sua capacidade de regressão e previsibilidade.

Dentre as técnicas de AM existentes, foram selecionadas: *Linear Regression*, *K-Neighbors*, *Support Vector Regressor*, *Multi Layer Perceptron*, *Random Forest* e *Gradient Boosting*, realizando estudos teóricos e implementações computacionais dessas técnicas. A análise comparativa revela que o modelo *Gradient Boosting* se destacou, mostrando notável precisão, obtendo um MAPE de 2,23% nas previsões no ano de 2022. As variáveis artificiais geradas pela engenharia de recursos desempenharam um papel importante na ponderação do modelo ensemble, destacando a sua relevância na tomada de decisão de regressão.

Finalmente, a aplicação prática do resultado desta pesquisa é de grande importância para o setor energético. A previsão precisa das cargas elétricas permite que as concessionárias façam planos mais eficazes e informados. Isto significa minimizar custos e evitar possíveis racionamentos e penalidades para os concessionários. A utilização destas tecnologias AM representa um grande avanço na otimização dos processos de geração e transmissão de energia elétrica.

7 REFERÊNCIAS

- ABDI, H. Coefficient of variation. **Encyclopedia of research design**, 1, 2010.
- ABEEÓLICA. ABEEólica. **Associação Brasileira de Energia Eólica**, 2023. Disponível em: <<https://abeeolica.org.br/>>. Acesso em: 23 ago. 2023.
- ABSOLAR. ABSOLAR, 2023. Disponível em: <<https://www.absolar.org.br/>>. Acesso em: 15 ago. 2023.
- AGGARWAL, C. C. **Outlier Analysis**. 2. ed. [S.l.]: Springer, Cham, 2017.
- ALMEIDA, H. A. D.; ALMEIDA, E. D. C. V. D. Potencial da energia solar fotovoltaica no Semiárido nordestino. **Concilium**, 11 fev. 2022. 197–210. Disponível em: <<http://clium.org/index.php/edicoes/article/view/111>>. Acesso em: 15 maio 2023.
- ALMEIDA, H. A. D.; ALMEIDA, E. D. C. V. D. Potencial da Energia Solar Fotovoltaica no Semiárido Nordeste. **Concilium**, 22, 2022. 197-210.
- ALVARO VALDERRAMA, C. V. H. A. M. I. C. V. Machine learning applications for urban photovoltaic potential estimation: A survey. **Elsevier**, 14 Março 2023. Acesso em: 03 Julho 2023.
- ANEEL. Agência Nacional de Energia Elétrica. **Agência Nacional de Energia Elétrica**, 2023. Disponível em: <<https://www.gov.br/aneel/pt-br>>. Acesso em: 15 ago. 2023.
- AUGUSTIN, N. H.; SAULEAU, E.-A.; WOOD, S. N. On quantile quantile plots for generalized linear models. **Computational Statistics & Data Analysis**, 56, n. 8, 2012. 2404-2409.
- BALOGUN, A.-L. et al. Assessing the Potentials of Digitalization as a Tool for Climate Change Adaptation and Sustainable Development in Urban Centres. **Sustainable Cities and Society**, 53, 2022. 101888.

BATAINEH, A. A. et al. Advanced Misinformation Detection: A Bi-LSTM Model Optimized by Genetic Algorithms. **Electronics**, 2023. Acesso em: 27 jul. 2023.

BEDI, J.; TOSHNIWAL, D. Deep learning framework to forecast electricity demand. **Applied Energy**, 238, 2019. 1312-1326.

BENTÉJAC, C.; CSÖRGŐ, A.; MARTÍNEZ-MUÑOZ, G. A comparative analysis of gradient boosting algorithms. **Artif Intell Rev**, 54, 2021. 1937–1967.

BERMANN, C. Impasses e controvérsias da hidreletricidade. **ESTUDOS AVANÇADOS**, n. 21, 2007. 59.

BISONG, E. Building Machine Learning and Deep Learning Models on Google Cloud Platform. [S.l.]: Apress, Berkeley, CA, 2019. p. 151–165.

BRUNESE, L. et al. Machine learning for coronavirus covid-19 detection from chest x-rays. **Procedia Computer Science**, 176, 2020. 2212-2221.

CANTOR, G. A. R. INFLUÊNCIA DOS FATORES CLIMÁTICOS NO, 20 fev. 2017. Acesso em: 13 maio 2023.

CHEN, W. T. et al. Electricity consumption prediction - A case study of a university library. **Journal of Building Engineering**, v. 76, p. 106990, 2023.

CHIODE, A. D. S. Avaliação de propostas de coeficientes de determinação do tipo R^2 em modelos de regressão logística com resposta nominal, 2021.

CONNOR SCOTT, M. A. A. A. Machine learning for forecasting a photovoltaic (PV) generation system. **ScienceDirect**, 01 Setembro 2023. Acesso em: 03 Julho 2023.

COSTA, P. C. D. S. Tendências e ciclos médios da velocidade do vento e complementaridade energética em regiões próximas a empreendimentos eólicos no Nordeste do Brasil. **Dissertação de Mestrado - Universidade Federal do Rio Grande do Norte**, 2022.

COSTA, S. B. O. D. Avanços da implantação da energia solar no Brasil, uma análise sistemática, 2022.

DALY, A.; DEKKER, T.; HESS, S. Dummy coding vs effects coding for categorical variables: Clarifications and extensions. **Journal of Choice Modelling**, 21, 2016. 36-41.

DANTAS, S. G.; POMPERMAYER, F.. **VIABILIDADE ECONÔMICA DE SISTEMAS FOTOVOLTAICOS NO BRASIL E POSSÍVEIS EFEITOS NO SETOR ELÉTRICO**. [S.l.]: Ipea. 2018. p. 8-10.

EPE. **Balanco Energético Nacional (BEN)**. [S.l.]. 2022.

EPE. EPE. **Empresa de Pesquisa Energética**, 2023. Disponível em: <<https://www.epe.gov.br/>>. Acesso em: 15 ago. 2023.

FAGUNDES, J. C. F. Abordagem baseada em Árvores de Decisão para detecção e identificação de intrusões, 2022. Disponível em: <<https://repositorio.ufsc.br/handle/123456789/232654>>. Acesso em: 20 maio 2023.

FENG, X. et al. A multi-layer perceptron approach for accelerated wave forecasting in Lake Michigan. **Ocean Engineering**, 211, 2020.

FILHO, D. B. F.; JÚNIOR, J. A. D. S. Desvendando os Mistérios do Coeficiente de Correlação de Pearson. **Revista Política Hoje**, 18, 2009. 115-146.

FISHER, M. J.; MARSHALL, A. P. Understanding descriptive statistics. **Australian Critical Care**, 22, 2009. 93-97.

FREIRE, A. Í.; FONTGALLAND, I. L. Perspectivas e desafios econômicos da geração de energia eólica na região Nordeste do Brasil. **Research, Society and Development**, 11, 2022. 01-11.

FUHRER, C.; SOLEM, J. E.; VERDIER, O. **Scientific Computing with Python: High-performance scientific computing with NumPy, SciPy, and pandas**. 2. ed. [S.I.]: Packt Publishing, 2021.

GALICKI, G. D. PANORAMA SOBRE A MOBILIDADE ELÉTRICA E BENEFÍCIOS QUANTO A MITIGAÇÃO DOS EFEITOS DAS MUDANÇAS CLIMÁTICAS NO BRASIL. **Trabalho de conclusão de curso (Bacharel) - Universidade Tecnológica Federal do Paraná**, 2022.

GANAIÉ, M. A. et al. Ensemble deep learning: A review. **Engineering Applications of Artificial Intelligence**, 115, 2022.

GONZÁLEZ-BRIONES, A.; HERNÁNDEZ, G.; CORCHADO, J. M. Machine learning models for electricity consumption forecasting: A Review. **2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)**, 2019.

GOUVÊA, R. L. P. D.; SILVA, P. A. D. Desenvolvimento do setor eólico no Brasil. **Banco Nacional de Desenvolvimento Econômico e Social**, v. 25, n. 49, p. 81-118, jun. 2018.

GUNTURI, S. K.; SARKAR, D. Ensemble machine learning models for the detection of energy theft. **Electric Power Systems Research**, 192, 2021.

IBGE. Instituto Brasileiro de Geografia e Estatística, 2023. Disponível em: <<https://www.ibge.gov.br/acesso-informacao/institucional/trabalhe-conosco/37301-2023-03-agente-de-pesquisas-e-mapeamento.html>>. Acesso em: 15 ago. 2023.

IEA. INTERNATIONAL ENERGY AGENCY. **INTERNATIONAL ENERGY AGENCY**, 2023. Disponível em: <<https://www.iea.org/>>. Acesso em: 15 ago. 2023.

INMET, 2023. Disponível em: <<https://portal.inmet.gov.br/>>. Acesso em: 18 set. 2023.

IRIZARRY, R. A. Introduction to Data Science Data Analysis and Prediction Algorithms with R. [S.I.]: CRC Press, 2019.

JIN, H. et al. Optimization and analysis of bioenergy production using machine learning modeling: Multi-layer perceptron, Gaussian processes regression, K-nearest neighbors, and Artificial neural network models. **Energy Reports**, 8, 2022. 13979-13996.

JÚNIOR, J. R. M. PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA. **Dissertação de Mestrado - Universidade Estadual Paulista (UNESP)**, 2022.

KAGAN, N.; OLIVEIRA, C. C. B. D.; ROBBA, E. J. **Introdução aos sistemas de distribuição de energia elétrica**. [S.l.]: Blucher, 2021.

KAGAN, N.; OLIVEIRA, C. C. B. D.; ROBBA, E. J. **Introdução aos Sistemas de Distribuição de Energia Elétrica**. 2. ed. [S.l.]: Blucher, 2021.

KALIYADAN, F.; KULKARNI, V. Types of Variables, Descriptive Statistics, and Sample Size. **Indian Dermatol Online J**, 2019. 82-86.

KHULLAR, S.; SINGH, N. Water quality assessment of a river using deep learning Bi-LSTM methodology: forecasting and validation. **Environmental Science and Pollution Research**, 29, 2021. 12875–12889.

KHULLAR, S.; SINGH, N. Water quality assessment of a river using deep learning Bi-LSTM methodology: forecasting and validation. **Environmental Science and Pollution Research**, 29, 2022. 12875-12889.

KRAMER, O. Scikit-Learn. In: _____ **Machine Learning for Evolution Strategies**. [S.l.]: [s.n.], 2016. p. 45-53.

KURDI, F. T.; AMAKHCHAN, W.; GHARINEIAT, Z. Random Forest Machine Learning Technique for Automatic Vegetation Detection and Modelling in LiDAR Data. **International Journal of Environmental Sciences & Natural Resources**, 28, 2021.

LAUTENCHLEGER, E. P.; FLECK, D.; STAMM, P. R. FERRAMENTAS DA QUALIDADE: UMA ABORDAGEM CONCEITUAL. **SIEF - Semana Internacional de Engenharia e Economia FAHOR**, 2015.

LEME, J. V. et al. Towards Assessing the Electricity Demand in Brazil: Data-Driven Analysis and Ensemble Learning Models. **Energies**, 13, 2019. 1407.

LI, S. et al. Electricity Theft Detection in Power Grids with Deep Learning and Random Forests. **Hindawi**, 2019.

LOPES, A. X. F.; NETO, J. O. D. N. Governança Ambiental Como Paradigma Energético no Cenário Internacional. **Anais do EVINCI**, 8, 2022. 196-2010.

LORENCIN, I. et al. Genetic Algorithm Approach to Design of Multi-Layer Perceptron for Combined Cycle Power Plant Electrical Power Output Estimation. **Energies**, 12, 2019.

LOU, C.; XIE, X. Multi-view intuitionistic fuzzy support vector machines with insensitive pinball loss for classification of noisy data. **Neurocomputing**, 549, 2023.

MACIEL, J. N. et al. Analysis of Artificial Neural Networks for Forecasting Photovoltaic Energy Generation with Solar Irradiance. **Brazilian Archives of Biology and Technology**, 08 abril 2021. Acesso em: 03 julho 2023.

MAMMONE, A.; TURCHI, M.; CRISTIANINI, N. Support vector machines. **WIRES Comp Stat**, 2009. 283-289.

MARCOS, I. P.; JÚNIOR, A. P. P. Previsão do Consumo de Energia Elétrica na Região Nordeste do Brasil. **Revista De Engenharia E Pesquisa Aplicada**, 6, 2021. 21-30.

MATSUBARA, G. C. Impactos das mudanças climáticas futuras sobre a geração de energia renovável no Nordeste brasileiro. **Dissertação de Mestrado - Universidade Federal do Rio Grande do Norte**, 2020.

MAULUD, D.; ABDULAZEEZ, A. M. A Review on Linear Regression Comprehensive in Machine Learning. **JASTT**, 1, 2020. 140-147.

MAURICIO, F. R. C. Latifúndio eólico: energia renovável, green grabbing e modernização conservadora no Nordeste do Brasil. **SER Social**, 25, 2023.

MAZUMDAR, D.; NETO, M. P.; PAULOVICH, F. V. Random Forest Similarity Maps: A Scalable Visual. **Electronics**, 10, 2021. 2862.

MURRAY, N. B. et al. Calculating acute:chronic workload ratios using exponentially weighted moving averages provides a more sensitive indicator of injury likelihood than rolling averages. **British Journal of Sports Medicine**, 51, 2017. 749-754.

NETO, J. O. D. N.; LOPES, A. X. F. GOVERNANÇA AMBIENTAL COMO PARADIGMA ENERGÉTICO NO CENÁRIO INTERNACIONAL. **EVINCI**, 8, 2022. 196-210.

ONS, 2023. Disponível em: <<https://www.ons.org.br>>. Acesso em: 18 set. 2023.

PARVEZ, I. et al. Multi-layer Perceptron based Photovoltaic Forecasting for Rooftop PV Applications in Smart Grid. **SoutheastCon**, 2020.

PESANTEZ, J. E. et al. A comparison study of predictive models for electricity demand in a diverse urban environment. **Energy**, 2023. 129142.

POTDAR, K.; PARDAWALA, T. S.; PAI, C. D. A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. **International Journal of Computer Applications**, 175, 2017. 7-9.

PRAYUDANI, S. et al. Analysis Accuracy Of Forecasting Measurement Technique On Random K-Nearest Neighbor (RKNN) Using MAPE And MSE. **Journal of Physics: Conference Series**, 1361, 2019. 012089.

PUNIA, S. et al. Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. **International Journal of Production Research**, 2020. 1-16.

RAMACHANDRAN, K. M.; TSOKOS, C. P. Chapter 1 - Descriptive statistics. **Academic Press**, 2021. 1-40.

SALOMÃO, A.; ALMEIDA, R. G. Energia e Desenvolvimento Sustentável: marcos no regime internacional de mudanças climáticas. **RIGHT**, n. 8, 2022. 22-28.

SANTIAGO, J. V. A. O mercado emergente de energia solar fotovoltaica no Brasil entre 2012 e 2018: avanços, desafios e perspectivas, 24 jan. 2020. Acesso em: 13 maio 2023.

SERIN, G. et al. Deep Multi-Layer Perceptron based Prediction of Energy Efficiency and Surface Quality for Milling in The Era of Sustainability and Big Data. **Procedia Manufacturing**, 51, 2020. 1166-1177.

SHABANI, S. et al. Modeling Pan Evaporation Using Gaussian Process Regression K-Nearest Neighbors Random Forest and Support Vector Machines; Comparative Analysis. **Atmosphere**, 11, 2020.

SHU, X.; YE, Y. Knowledge Discovery: Methods from data mining and machine learning. **Social Science Research**, 110, 2023.

SILVA, A. P. D. et al. Risco de queda relacionado a medicamentos em hospitais. **Acta Paul Enferm**, 2023.

SILVA, B. A. O. D. Aproveitamento e potencial da energia solar no nordeste brasileiro. **Meio Ambiente (Brasil)**, 5, 2023. 055-066.

SILVA, L. R. D. J. R.; SHAYANI, R. A.; OLIVEIRA, M. A. G. D. ANÁLISE COMPARATIVA DAS FONTES DE ENERGIA SOLAR FOTOVOLTAICA, HIDRELÉTRICA E TERMELÉTRICA, COM LEVANTAMENTO DE CUSTOS

AMBIENTAIS. **CBENS**, 2018. Disponível em: <<https://anaiscbens.emnuvens.com.br/cbens/article/view/527/527>>. Acesso em: 15 maio 2023.

SINSOMBOONTHONG, S. Performance Comparison of New Adjusted Min-Max with Decimal Scaling and Statistical Column Normalization Methods for Artificial Neural Network Classification. **International Journal of Mathematics and Mathematical Sciences**, 2022.

SONKAVDE, G. Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications. **Financial Studies**, 13, 2023. 94. Acesso em: 15 maio 2023.

SOUZA, R. F. S. Detecção e classificação de falhas em rolamentos de motores elétricos baseado em árvores de decisão, 2022. Disponível em: <<https://repositorio.ufrn.br/handle/123456789/46252>>. Acesso em: 15 maio 2023.

TAMER KHATIB, D. H. M. Optimal Sizing of Standalone Photovoltaic System Using Improved Performance Model and Optimization Algorithm. **Sustainability**, 13 Março 2020. Acesso em: 03 Julho 2023.

TATIS, A. F. G. S. L.; CORRENTE, J. E.; FUMES-GHANTOUS, G. Análise exploratória gráfica para dados assimétricos com presença de pontos discrepantes. **Revista Brasileira de Iniciação Científica**, 9, 2022. Disponível em: <<https://periodicoscientificos.itp.ifsp.edu.br/index.php/rbic/article/view/540>>. Acesso em: 15 maio 2023.

UPADHYAY, D. et al. Gradient Boosting Feature Selection With Machine Learning Classifiers for Intrusion Detection on Power Grids. **IEEE Transactions on Network and Service Management**, 18, 2021. 1104-1116.

WAHEED, R.; SARWAR, S.; CHEN, W. The survey of economic growth, energy consumption and carbon emission. **Energy Reports**, 5, 2019. 1103-1115.

WANG, J.; ZHANG, S. An improved deep learning approach based on exponential moving average algorithm for atrial fibrillation signals identification. **Neurocomputing**, 513, 2022. 127-136.

ZHAN, Z. et al. Grabbing the Long Tail: A data normalization method for diverse and informative dialogue generation. **Neurocomputing**, 460, 2021. 374-384.

ZHANG, D.; KHALILI, A.; ASGHARIAN, M. Post-model-selection inference in linear regression models: An integrated review, 16, 2022. 86 - 136.

ZOUNEMAT-KERMANI, M. et al. Ensemble machine learning paradigms in hydrology: A review. **Journal of Hydrology**, 598, 2021.