



UNIVERSIDADE FEDERAL FLUMINENSE
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E DE
TELECOMUNICAÇÕES

MARCOS VINICIUS MORO SIQUEIRA

Aplicação de Modelos de Aprendizado de Máquina para Previsão de Energia Eólica

NITERÓI

2024

UNIVERSIDADE FEDERAL FLUMINENSE
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E DE
TELECOMUNICAÇÕES

MARCOS VINICIUS MORO SIQUEIRA

Aplicação de Modelos de Aprendizado de Máquina para Previsão de Energia Eólica

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Telecomunicações da Universidade Federal Fluminense, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica e de Telecomunicações. Área de concentração: Modelagem e Análise de Sistemas de Energia Elétrica.

Orientador:
Vitor Hugo Ferreira

NITERÓI
2024

Ficha catalográfica automática - SDC/BEE
Gerada com informações fornecidas pelo autor

S618a Siqueira, Marcos Vinicius Moro
Aplicação de Modelos de Aprendizado de Máquina para
Previsão de Energia Eólica / Marcos Vinicius Moro Siqueira. -
2024.
157 f.: il.

Orientador: Vitor Hugo Ferreira.
Dissertação (mestrado)-Universidade Federal Fluminense,
Escola de Engenharia, Niterói, 2024.

1. Aprendizado de máquina. 2. Modelos computacionais. 3.
Séries temporais. 4. Geração eólica. 5. Produção
intelectual. I. Ferreira, Vitor Hugo, orientador. II.
Universidade Federal Fluminense. Escola de Engenharia. III.
Título.

CDD - XXX

MARCOS VINICIUS MORO SIQUEIRA

APLICAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA PARA PREVISÃO DE
ENERGIA EÓLICA

Dissertação de Mestrado apresentada ao Programa de Pós Graduação em Engenharia Elétrica e de Telecomunicações da Universidade Federal Fluminense como requisito parcial para a Obtenção do Grau de Mestre em Engenharia Elétrica e de Telecomunicações.

Área de concentração: Sistemas de Energia Elétrica.

BANCA EXAMINADORA

Prof. Dr. Vitor Hugo Ferreira - Orientador
Universidade Federal Fluminense - UFF

Prof. Dr. Angelo Cesar Colombini
Universidade Federal Fluminense - UFF

Prof. Dr. Alexandre Rasi Aoki
Universidade Federal do Paraná - UFPR

Prof. Dr. Hugo Valadares Siqueira
Universidade Tecnológica Federal do Paraná - UTFPR

Niterói
(Agosto/2024)

ESCOLA DE ENGENHARIA

COORDENAÇÃO DO PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA E DE TELECOMUNICAÇÕES

Ata da sessão de julgamento de Dissertação de Mestrado na área de concentração Sistemas de Energia Elétrica, do aluno Marcos Vinicius Moro Siqueira.

Aos vinte dias do mês de agosto de dois mil e vinte e quatro, no prédio da Escola de Engenharia, reuniu-se a Comissão Examinadora, designada na forma regimental pela Coordenação do Curso, para a defesa da Dissertação de Mestrado apresentada pelo aluno Marcos Vinicius Moro Siqueira, sob o título “***Aplicação de Modelos de Aprendizado de Máquina para Previsão de Energia Eólica***”, como requisito para obtenção de grau de Mestre em Engenharia Elétrica e de Telecomunicações – área de concentração Sistemas de Energia Elétrica. A Comissão foi presidida pelo professor doutor Vitor Hugo Ferreira, tendo como membros da banca os professores doutores, Angelo Cesar Colombini, Alexandre Rasi Aoki e Hugo Valadares Siqueira. Aberta a sessão pública, foi concedido ao candidato o tempo de 45 (quarenta e cinco) minutos para a exposição do trabalho, sendo a defesa seguida pela arguição de cada examinador. A seguir, a Comissão reuniu-se para deliberar sobre o texto da dissertação e sua defesa oral, concluindo pela aprovação da mesma *sem restrições*, e nos termos do Regulamento Geral dos Cursos de Pós-Graduação desta Universidade, foi lavrada a presente ata, lida e julgada, conforme vai assinada pelos membros da Comissão.

Niterói, 20 de agosto de 2024.

Prof. Dr. Vitor Hugo Ferreira- Orientador
Universidade Federal Fluminense - UFF

Prof. Dr. Angelo Cesar Colombini
Universidade Federal Fluminense - UFF

Prof. Dr. Alexandre Rasi Aoki
Universidade Federal do Paraná - UFPR

Prof. Dr. Hugo Valadares Siqueira
Universidade Tecnológica Federal do Paraná - UTFPR

“A capacidade de acessar o futuro é o maior diferencial competitivo que o ser humano pode atingir”

José Siqueira

Agradecimentos

Agradeço à minha família e minha namorada, pelo incentivo e pelos valores que me foram passados.

Agradeço ao professor Vitor Hugo Ferreira pelo papel de professor e orientador no mestrado, pelo apoio, competência e conhecimento compartilhado.

Agradeço aos demais membros da banca. Professores Aoki, Ângelo e Hugo, pela dedicação e orientação.

Agradeço a Universidade Federal Fluminense pela oportunidade de realização deste trabalho.

Agradeço aos meus colegas de trabalho da Tradener pelas lições, paciência e conhecimentos teóricos.

Agradeço a todos que direta ou indiretamente fizeram parte de minha formação.

Resumo

A viabilização da energia eólica como uma fonte confiável de energia demanda a superação dos desafios impostos pela sua volatilidade, inerente à natureza estocástica dos ventos. A previsão de longo prazo oferece vantagens estratégicas para a gestão de empreendimentos de geração de energia, como a capacidade de desenvolver uma estratégia de gestão de portfólio de energia, permitindo a interação entre os contratos de compra e venda negociados e a curva de produção do empreendimento de geração. Isso possibilita maximizar os lucros ao alinhar a produção de energia com as condições de mercado favoráveis. Além disso, a otimização anual da operação, assegura que a usina opere de maneira eficiente, ajustando a produção para atender à demanda esperada e às variações sazonais. O planejamento estratégico e a gestão eficaz de riscos são facilitados, pois a previsibilidade a longo prazo permite que as empresas identifiquem e mitiguem potenciais riscos, garantindo a sustentabilidade e o crescimento contínuo no setor. O objetivo desse trabalho é a modelagem e aplicação de previsores de geração de energia eólica a partir de técnicas de aprendizado de máquina, tais como *eXtreme Gradient Boosting*, Perceptron de Múltiplas Camadas, Regressão por Vetores Suporte, Regressão Linear Rígida e Florestas Aleatórias. No treinamento dos modelos são consideradas séries temporais de geração histórica mensal do Complexo Eólico Pindaí, com potência instalada de 79,9 MW e situado na região nordeste do Brasil, juntamente com dados meteorológicos provenientes de modelos de reanálise, tais como temperatura do ar, umidade relativa, precipitação, pressão da superfície, velocidade do vento a 10 metros, velocidade do vento a 100m e rajadas de vento. As metodologias propostas são aplicadas à previsão de geração eólica mensal do ano de 2023 e as saídas comparadas sistematicamente através de métricas de avaliação, para determinar a modelagem mais adequada para o problema de previsão. Os resultados evidenciam a superioridade dos modelos Perceptron de Múltiplas Camadas, Regressão por Vetores Suporte e *eXtreme Gradient Boosting*, que alcançaram um Coeficiente de Determinação (R^2) de 0,86, 0,85 e 0,95, um Erro Percentual Absoluto Médio (MAPE) de 6,29%, 6,57% e 3,90%, uma Raiz do Erro Quadrático Médio (RMSE) de 0,56, 0,59 e 0,35 e um Erro Absoluto Médio (MAE) de 0,48, 0,52 e 0,29, respectivamente.

Palavras-chaves: Aprendizado de Máquina, *eXtreme Gradient Boosting*, Rede Neural Perceptron de Múltiplas Camadas, Máquinas de Vetor Suporte, Regressão Linear Rígida, Florestas Aleatórias, Complexo Eólico Pindaí.

Abstract

Ensuring the reliability of wind energy as a dependable source requires overcoming challenges posed by the inherent volatility and stochastic nature of wind patterns. Long-term forecasting provides strategic advantages in managing energy generation projects, enabling the development of effective portfolio management strategies. This allows for the synchronization of power production with favorable market conditions, optimizing profits through the strategic alignment of energy contracts with the generation profile of the wind farm. Moreover, annual operational optimization ensures that the facility operates efficiently by adjusting production to meet expected demand and seasonal variations. Strategic planning and risk management are also enhanced, as long-term predictability allows companies to identify and mitigate potential risks, ensuring sustained growth and stability in the sector. The aim of this study is to model and apply wind energy generation predictors using machine learning techniques, including eXtreme Gradient Boosting, Multilayer Perceptron, Support Vector Regression, Ridge Regression and Random Forests. The models are trained using historical monthly generation data from the Pindaí Wind Complex, which has an installed capacity of 79.9 MW and located in the northeastern region of Brazil, along with meteorological data from reanalysis models, such as air temperature, relative humidity, precipitation, surface pressure, wind speed at 10 meters, wind speed at 100 meters, and wind gusts. These methodologies are applied to forecast monthly wind generation for the year 2023, and the outputs are systematically compared using evaluation metrics to determine the most suitable modeling approach. The results highlight the superiority of the Multilayer Perceptron, Support Vector Regression, and eXtreme Gradient Boosting models, which achieved coefficients of determination (R^2) of 0.86, 0.85, and 0.95, mean absolute percentage errors (MAPE) of 6.29%, 6.57%, and 3.90%, root mean square errors (RMSE) of 0.56, 0.59, and 0.35, and mean absolute errors (MAE) of 0.48, 0.52, and 0.29, respectively.

Keywords: Machine Learning, eXtreme Gradient Boosting, Multilayer Perceptron, Support Vector Machines, Ridge Linear Regression, Random Forests, Pindaí Wind Complex.

Lista de Figuras

FIGURA 1 - AEROGERADOR WOBLEN WINDPOWER NO COMPLEXO PINDAÍ. FONTE: [95].....	49
FIGURA 2 - FORÇA DE SUSTENTAÇÃO DA LÂMINA. FONTE: ADAPTADO DE [98].....	50
FIGURA 3 - CONCEITO DE VELOCIDADE RELATIVA.....	51
FIGURA 4 - VELOCIDADE PERPENDICULAR AO CILINDRO DE AR UNIFORME. FONTE: ADAPTADO DE [101].....	52
FIGURA 5 - COMPONENTES PRINCIPAIS DO AEROGERADOR HAWT	57
FIGURA 6 - CURVA DE POTÊNCIA DOS AEROGERADORES E-92. FONTE: [123].....	64
FIGURA 7 - TRANSITÓRIO DE VELOCIDADE DOS AEROGERADORES E-92. FONTE: [124].....	64
FIGURA 8 - GERAÇÃO HORÁRIA HISTÓRICA DO COMPLEXO PINDAÍ.....	66
FIGURA 9 - GERAÇÃO MENSAL HISTÓRICA DO COMPLEXO PINDAÍ [MWh]	71
FIGURA 10 - LOCALIZAÇÃO DO COMPLEXO PINDAÍ. FONTE: [130].....	77
FIGURA 11 - VARIÁVEIS METEOROLÓGICAS HISTÓRICAS EXTRAÍDAS DO MODELO DE REANÁLISE <i>BEST MATCH</i> . FONTE: [30]	79
FIGURA 12 - VARIÁVEIS METEOROLÓGICAS HISTÓRICAS EXTRAÍDAS POR API JSON	79
FIGURA 13 - VARIÁVEIS METEOROLÓGICAS PREVISTAS EXTRAÍDAS POR API JSON. FONTE: [30].....	80
FIGURA 14 - VARIÁVEIS METEOROLÓGICAS PREVISTAS EXTRAÍDAS POR API JSON	81
FIGURA 15 - FLUXOGRAMA DO ALGORITMO DE PREVISÃO DE LONGO PRAZO.....	83
FIGURA 16 - GERAÇÃO MENSAL HISTÓRICA DO COMPLEXO PINDAÍ [MWh]	98
FIGURA 17 - <i>BOXPLOTS</i> DA DISTRIBUIÇÃO HISTÓRICA DE GERAÇÃO NO PERÍODO DE 2014 A 2023 EM MWh. PINDAÍ I (SUPERIOR ESQUERDO), PINDAÍ II (SUPERIOR DIREITO), PINDAÍ III (INFERIOR ESQUERDO) E PINDAÍ IV (INFERIOR DIREITO).....	100
FIGURA 18 - CORRELAÇÃO DE VARIÁVEIS EXÓGENAS COM A GERAÇÃO DO COMPLEXO DE PINDAÍ.....	102
FIGURA 19 - PESO ATRIBUÍDO ÀS VARIÁVEIS EXÓGENAS.....	103
FIGURA 20 - DISPERSÃO ENTRE A GERAÇÃO HISTÓRICA E A VELOCIDADE DO VENTO A 100M. PINDAÍ I (SUPERIOR ESQUERDO), PINDAÍ II (SUPERIOR DIREITO), PINDAÍ III (INFERIOR ESQUERDO) E PINDAÍ IV (INFERIOR DIREITO)	104

FIGURA 21 - SEPARAÇÃO DO CONJUNTO DE DADOS EM TREINAMENTO, VALIDAÇÃO E TESTE	107
FIGURA 22 - CURVAS DE APRENDIZADO REFERENTE AOS MODELOS – XGBOOST, REGRESSÃO RÍGIDA E FLORESTAS ALEATÓRIAS.....	109
FIGURA 23 - CURVAS DE VALIDAÇÃO REFERENTE AOS MODELOS – XGBOOST, REGRESSÃO RÍGIDA E FLORESTAS ALEATÓRIAS.....	110
FIGURA 24 - EVOLUÇÃO DO COEFICIENTE DE DETERMINAÇÃO EM FUNÇÃO DO HIPERPARÂMETRO C - REGRESSÃO POR VETORES SUPORTE.....	111
FIGURA 25 - EVOLUÇÃO E VARIAÇÕES DO COEFICIENTE DE DETERMINAÇÃO AO LONGO DAS ÉPOCAS - MLP	113
FIGURA 26 – PREVISÃO DE GERAÇÃO DE PINDAÍ I - 2023	114
FIGURA 27 - PREVISÃO DE GERAÇÃO DE PINDAÍ II - 2023	115
FIGURA 28 - PREVISÃO DE GERAÇÃO DE PINDAÍ III - 2023	115
FIGURA 29 - PREVISÃO DE GERAÇÃO DE PINDAÍ IV – 2023	116

Lista de Tabelas

TABELA 1 - HIPERPARÂMETROS E DEFINIÇÕES - REGRESSÃO LINEAR RÍGIDA.....	20
TABELA 2 - HIPERPARÂMETROS E DEFINIÇÕES - FLORESTAS ALEATÓRIAS	23
TABELA 3 - HIPERPARÂMETROS E DEFINIÇÕES – REGRESSÃO POR VETORES SUPORTE	26
TABELA 4 - HIPERPARÂMETROS E DEFINIÇÕES - PERCEPTRON DE MÚLTIPLAS CAMADAS....	30
TABELA 5 - HIPERPARÂMETROS E DEFINIÇÕES - <i>EXTREME GRADIENT BOOSTING</i>	39
TABELA 6 - ANOMALIAS SAZONAIS DE TEMPERATURA DESDE 2000. FONTE: [94]	48
TABELA 7 - GERAÇÃO MENSAL HISTÓRICA DO COMPLEXO PINDAÍ [MWh].....	71
TABELA 8 - ESTATÍSTICAS DESCRITIVAS DOS DADOS DE GERAÇÃO	73
TABELA 9 - MODELOS METEOROLÓGICAS E PROVEDORES INTERNACIONAIS. FONTE: [30] ..	76
TABELA 10 - PARÂMETROS E DESCRIÇÃO DAS VARIÁVEIS HISTÓRICAS. FONTE: ADAPTADO DE [30].....	78
TABELA 11 - MODELOS METEOROLÓGICOS UTILIZADOS PARA PREVISÃO. FONTE: [30].....	79
TABELA 12 - PARÂMETROS E DESCRIÇÃO DAS VARIÁVEIS PREVISTAS	80
TABELA 13 – VALIDAÇÃO CRUZADA DOS HIPERPARÂMETROS - REGRESSÃO LINEAR RÍGIDA	84
TABELA 14 – VALIDAÇÃO CRUZADA DOS HIPERPARÂMETROS – FLORESTAS ALEATÓRIAS..	86
TABELA 15 - VALIDAÇÃO CRUZADA DOS HIPERPARÂMETROS – REGRESSÃO POR VETORES SUPORTE.....	87
TABELA 16 - VALIDAÇÃO CRUZADA DOS HIPERPARÂMETROS - PERCEPTRON DE MÚLTIPLAS CAMADAS.....	89
TABELA 17 - CONFIGURAÇÃO DOS HIPERPARÂMETROS – <i>EXTREME GRADIENT BOOSTING</i>	91
TABELA 18 - ESTATÍSTICA DA GERAÇÃO MENSAL [MWh]	98
TABELA 19 - TESTES AUGMENTED DICKEY-FULLER PARA O COMPLEXO EÓLICO PINDAÍ ..	105
TABELA 20 - GERAÇÃO PREVISTA E REALIZADA PARA O COMPLEXO EÓLICO PINDAÍ - 2023 [MWmed]	116
TABELA 21 – MÉTRICAS DE AVALIAÇÃO PARA A PREVISÃO DO COMPLEXO EÓLICO PINDAÍ	118
TABELA 22 - RESULTADO MÉDIO DE TODOS OS PARQUES – MÉTRICAS DE AVALIAÇÃO.....	120

Lista de Abreviaturas e Siglas

ADAM	<i>Adaptive Moment Estimation</i>
ADF	<i>Augmented Dickey-Fuller</i>
AEMO	<i>Australian Energy Market Operator</i>
AMO	Oscilação Multidecadal do Atlântico
API	<i>Application Programming Interface</i>
ARIMA	<i>Autoregressive Integrated Moving Average</i>
ASAS	Anticiclone Subtropical do Atlântico Sul
CFD	Dinâmica dos Fluidos Computacional
CERRA	<i>Copernicus Regional Reanalysis for Europe</i>
CNN	<i>Convolutional Neural Networks</i>
CPU	<i>Central Processing Unit</i>
DBSCAN	<i>Density-based Spatial Clustering of Applications with Noise</i>
DCD	<i>Dual Coordinate Descent</i>

DFIG	<i>Doubly-Fed Induction Generator</i>
DQN	<i>Deep Q-Network</i>
DWD	<i>Deutscher Wetterdienst</i>
ECMWF	<i>European Centre for Medium-Range Weather Forecast</i>
ENSO	<i>El Niño-Southern Oscillation</i>
ERA5	<i>ECMWF Reanalysis v5</i>
FFNN	<i>Feedforward Neural Networks</i>
GAN	<i>Generative Adversarial Network</i>
GBM	<i>Gradient Boosting Machine</i>
GPU	<i>Graphics Processing Unit</i>
HAWTs	<i>Horizontal Axis Wind Turbine</i>
JSON	<i>JavaScript Object Notation</i>
k-NN	<i>k-Nearest Neighbors</i>
LASSO	<i>Least Absolute Shrinkage and Selection Operator</i>
LOCF	<i>Last Observation Carried Forward</i>

LS-SVM	<i>Least Squares Support Vector Machines</i>
LSTM	<i>Long Short-Term Memory</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MAR	<i>Missing at Random</i>
MCAR	<i>Missing Completely at Random</i>
MDP	<i>Markov Decision Process</i>
ML	<i>Machine Learning</i>
MLP	<i>Multilayer Perceptron</i>
MNAR	<i>Missing Not at Random</i>
MQO	Mínimos Quadrados Ordinários
MSL	Pressão do Ar Atmosférico a Nível Médio do Mar
NOAA	<i>National Oceanic and Atmospheric Administration</i>
NOCB	<i>Next Observation Carried Backward</i>
NWP	<i>Numerical Weather Prediction</i>
ONS	Operador Nacional do Sistema

ODP	Oscilação Decadal do Pacífico
PCA	<i>Principal Component Analysis</i>
PMSG	<i>Permanent Magnet Synchronous Generator</i>
RBF	<i>Radial Basis Function</i>
ReLU	<i>Rectified Linear Unit</i>
RF	<i>Random Forest</i>
RNN	<i>Recurrent Neural Network</i>
RPM	Rotação por Minuto
R ²	Coefficiente de Determinação
SCADA	<i>Supervisory Control and Data Acquisition</i>
SGD	<i>Stochastic Gradient Descent</i>
SHAP	<i>SHapley Additive exPlanations</i>
SMO	<i>Sequential Minimal Optimization</i>
SSR	<i>Sum of Squares of Residuals</i>
SST	<i>Total Sum of Squares</i>

SVM	<i>Support Vector Machine</i>
SVR	<i>Support Vector Regression</i>
TCL	<i>Central Limit Theorem</i>
TSM	Temperaturas da Superfície do Mar
VAWTs	<i>Vertical-Axis Wind Turbine</i>
XGBoost	<i>eXtreme Gradient Boosting</i>
ZCIT	Zona de Convergência Intertropical

Sumário

Capítulo 1 -	Introdução	1
1.1	Motivações	4
1.2	Objetivos	5
1.3	Estrutura do Trabalho	5
Capítulo 2 -	Referencial Teórico.....	7
2.1	Revisão da Literatura.....	7
2.1.1	Regressão Linear	16
2.1.1.1	Regressão Linear Rígida	18
2.1.1.2	Regressão Linear Lasso.....	20
2.1.2	Florestas Aleatórias	22
2.1.3	Máquina de Vetor Suporte	24
2.1.4	Perceptron de Múltiplas Camadas.....	28
2.1.5	eXtreme Gradient Boosting	32
2.2	Contexto Climático e Operação de Turbinas Eólicas.....	41
2.2.1	Natureza do Vento	42
2.2.1.1	Volatilidade e Previsibilidade	43
2.2.1.2	Regimes Eólicos Sinóticos e de Mesoescala na Bahia.....	44
2.2.1.3	Variabilidade Interanual e Fenômenos ENSO	46
2.2.2	Turbinas Eólicas	48
2.2.2.1	Energia e Eficiência das Turbinas Eólicas	52
2.2.2.2	Projetos e Características das Turbinas Eólicas.....	54
Capítulo 3 -	Materiais e Métodos	62
3.1	Descrição do Complexo Eólico e Conjunto de Dados.....	63

3.2	Modelos de Reanálise de Previsão Meteorológica com Alta Resolução	76
3.3	Machine Learning.....	82
3.3.1	Regressão Linear Rígida	84
3.3.2	Florestas Aleatórias	85
3.3.3	Regressão por Vetores Suporte	87
3.3.4	Perceptron de Múltiplas Camadas.....	88
3.3.5	eXtreme Gradient Boosting	89
3.4	CrITÉrios de AvaliaÇão dos Resultados.....	91
3.5	Considerações Gerais	94
Capítulo 4 -	Resultados.....	97
4.1	Resultados da Análise da Série Temporal.....	97
4.2	Resultados da Etapa de Treinamento e Teste dos Modelos.....	107
4.3	Resultado da Previsão do Ano de 2023	114
4.4	Resultados e Discussão	121
Capítulo 5 -	Conclusões e Trabalhos Futuros	126
Bibliografia.....		129

Capítulo 1 - Introdução

A construção de usinas eólicas se justifica pela necessidade de atender à crescente demanda por energia no Brasil, reduzir as emissões de gases do efeito estufa e mitigar os riscos associados ao atual modelo energético brasileiro por meio da diversificação de suas fontes de energia. Segundo dados do Operador Nacional do Sistema Elétrico (ONS), a capacidade da energia eólica instalada alcançou 20.655 MW médios em 2022, ampliando para 25.491 MW médios ao término de 2023, representando um incremento anual de 23% [1]. Tal crescimento, particularmente notável no subsistema Nordeste, deve-se a investimentos substanciais, consolidando o Brasil como líder global em geração eólica.

A geração de energia eólica está intrinsecamente ligada a processos estocásticos de grandezas meteorológicas, refletindo a natureza aleatória e imprevisível dessas variáveis. A estocasticidade no contexto da energia eólica diz respeito à dependência direta de variáveis climáticas que são, por sua própria essência, voláteis e dinâmicas. A velocidade do vento, principal motor das turbinas eólicas, pode variar significativamente, desde brisas suaves até rajadas intensas que ultrapassam 25 m/s em curtos intervalos de tempo. Essas flutuações na intensidade do vento tornam o processo de geração de energia altamente sensível às condições meteorológicas momentâneas. Além da velocidade do vento, outros fatores climáticos contribuem para a eficiência da geração de energia. Variações de temperatura, por exemplo, alteram a densidade do ar e impactam diretamente a capacidade das turbinas de converter a energia cinética do vento em energia elétrica. A umidade relativa, por sua vez, influencia a densidade atmosférica e, conseqüentemente, a eficácia aerodinâmica das pás das turbinas. Essas variáveis não são constantes e apresentam variações sazonais significativas, como as mudanças de temperatura no Oceano Pacífico, que podem alterar os padrões de precipitação e, assim, impactar a produção de energia. Adicionalmente, as flutuações na pressão atmosférica podem alterar tanto a velocidade quanto a direção do vento, introduzindo mais uma camada de incerteza no processo de geração [2].

Nesse cenário de alta variabilidade e incerteza, a previsão de longo prazo surge como uma ferramenta estratégica que visa alcançar objetivos tanto para a operação eficiente dos parques eólicos quanto para o planejamento de longo prazo no setor de energia renovável. A capacidade de antecipar a curva de geração futura ao longo de meses ou anos permite que

empresas desenvolvam estratégias de gestão de portfólio de energia mais eficientes e financeiramente mais lucrativas. Por meio dessa previsão, é possível alinhar a produção de energia com as condições de mercado mais favoráveis, otimizando a negociação de contratos de compra e venda de energia. Ao ajustar a curva de produção às oportunidades de mercado, é possível maximizar os lucros, garantindo que a energia gerada seja comercializada em momentos de maior demanda e preços mais elevados.

Em termos de planejamento estratégico, essa antecipação da geração futura e dos comportamentos climáticos possibilita a otimização do cronograma de manutenção preventiva e corretiva, assegurando que as turbinas estejam em condições ideais de operação nos períodos de maior geração esperada. A manutenção planejada, respaldada por previsões de longo prazo, minimiza a ocorrência de inatividade não programada, que poderia resultar em perdas de produção e, conseqüentemente, de receita. Além disso, ao evitar falhas inesperadas, que poderiam exigir intervenções corretivas emergenciais, os operadores reduzem o risco de comprometimento da eficiência operacional e prolongam a vida útil dos equipamentos, considerando que a maioria dos processos decisórios e administrativos em parques eólicos, tais como a formulação de orçamentos, alocação de recursos e programação de manutenções, é realizada em ciclos mensais ou trimestrais [3].

A integração de modelos climáticos com técnicas de aprendizado de máquina pode oferecer soluções para enfrentar esses desafios, que incluem as flutuações nas condições climáticas e seu impacto na eficiência dos aerogeradores. Ao combinar dados históricos de padrões climáticos com dados em tempo real, algoritmos de aprendizado de máquina podem identificar padrões ocultos e prever com maior precisão as condições futuras.

Nesse contexto atual, marcado pela elevada disponibilidade de dados meteorológicos e operacionais, a ciência de dados, especialmente o aprendizado de máquina (*machine learning*), emerge como uma ferramenta para a análise preditiva de séries temporais. Estes conceitos aplicam algoritmos e técnicas de mineração de dados, como redes neurais artificiais, regressão por vetores suporte e árvores de decisão, para extrair padrões e *insights* a partir de grandes volumes de dados. Além disso, a ciência de dados possibilita a implementação de sistemas de monitoramento em tempo real, utilizando técnicas de processamento de fluxo contínuo para detectar anomalias operacionais e otimizar a manutenção preditiva dos aerogeradores, uma abordagem que possibilita o aumento da

eficiência e da vida útil dos equipamentos, maximizando a produção de energia eólica e contribuindo para a confiabilidade do fornecimento de energia renovável [4].

Dentro do espectro das técnicas de aprendizado de máquina, existem técnicas como a Regressão Linear Rígida (*Ridge*) e Florestas Aleatórias (*Random Forest*). A regressão linear rígida é uma variante da regressão linear clássica, introduzindo um termo de regularização que penaliza coeficientes elevados, especialmente em conjuntos de dados onde as variáveis são altamente correlacionadas. Por outro lado, as florestas aleatórias, agregam múltiplas árvores de decisão independentes, cada uma construída a partir de amostras aleatórias do conjunto de dados e subconjuntos de características, para capturar complexas interações não lineares e variáveis importantes que podem ser difíceis de identificar em modelos lineares.

Além disso, as técnicas de Máquinas de Vetor Suporte (*Support Vector Machine* - SVM) e Redes Neurais Perceptron de Múltiplas Camadas (*Multilayer Perceptron* - MLP) são amplamente valorizadas por suas capacidades em problemas de classificação e regressão [5]. Regressão por Vetores Suporte (*Support Vector Regression* - SVR) são reconhecidos por sua robustez em espaços de alta dimensionalidade e pela eficácia em otimizar hiperplanos de separação, aproveitando os recursos dos *kernels* para tratar de problemas não linearmente separáveis [6]. A MLP é composta por várias camadas de neurônios interconectados, capaz de modelar relações não-lineares complexas graças às suas camadas ocultas, que diferenciam essa rede de um perceptron simples. O treinamento do utiliza o algoritmo de retropropagação para ajustar os pesos sinápticos e minimizar erros, permitindo à rede generalizar a partir dos dados, resultando em desempenhos significativos em tarefas como previsão de séries temporais e reconhecimento de padrões [7].

O *Gradient Boosting*, por sua vez, destaca-se por sua eficácia em uma ampla gama de aplicações, posicionando-o como estado da arte dos métodos de classificação e regressão. O *eXtreme Gradient Boosting* (XGBoost) é uma implementação avançada e escalável desta técnica, consolidando-se como uma ferramenta considerada moderna nos estudos referentes à ciência de dados. Um fator importante do XGBoost é atribuído à sua escalabilidade, suportando operações mais rápidas que métodos convencionais em máquinas individuais, sendo dimensionado para processar bilhões de exemplos em ambientes distribuídos ou com memória restrita [8].

Sendo assim, nesse capítulo tem-se por objetivo apresentar as motivações para se estudar a modelagem de técnicas de aprendizado de máquina no contexto de previsão de geração eólica e enumerar os objetivos principais deste trabalho.

1.1 Motivações

A motivação para a realização deste trabalho baseia-se em avanços no estado da arte em técnicas de aprendizado de máquina e sua aplicação na previsão de geração de energia renovável. Para isso, são empregados modelos de aprendizado de máquina, contribuindo para a otimização da produção de energia e visando maximizar a eficiência e a rentabilidade dos parques eólicos. O Complexo Eólico Pindaí, especificamente, serve como um caso de estudo real para a aplicação destas técnicas de previsão.

Antecipar a geração de energia na escala de longo prazo é fundamental para lidar com a variação intrínseca das fontes renováveis, dado que a produção energética está sujeita às variações de velocidade do vento, dificultando o planejamento e a gestão da rede elétrica. A previsão permite aos operadores e gestores ajustarem as operações, integrando de forma eficiente a energia eólica no portfólio energético da carteira de energia da comercializadora e garantindo uma distribuição otimizada da energia gerada. Do ponto de vista financeiro, as previsões de longo prazo desempenham um papel na mitigação de riscos e na maximização do retorno sobre o investimento. Para os investidores e operadores de parques eólicos, o conhecimento da geração futura possibilita firmar contratos de venda de energia personalizados e mais vantajosos, ajustando estratégias de manutenção para reduzir custos operacionais e otimizar a alocação de recursos. Isso resulta em maior estabilidade financeira do empreendimento, tornando-o atraente para investimentos futuros, financiamentos e eventualmente *retrofits* de equipamentos obsoletos. No campo da engenharia, a previsão de longo prazo representa um desafio no que diz respeito à modelagem de sistemas dinâmicos, análise de grandes volumes de dados e à aplicação de técnicas computacionais através de algoritmos de aprendizado de máquina.

1.2 Objetivos

O objetivo geral deste trabalho consiste no desenvolvimento de métodos de previsão que auxiliem na tomada de decisões estratégicas no âmbito do empreendimento, especificamente para o complexo eólico e sua operação comercial. Para tal, leva-se em consideração a integração de técnicas de análise de *Big Data*, engenharia de dados e uso de computação avançada incluindo técnicas de aprendizado de máquina, sendo possível capturar, armazenar e processar grandes volumes de dados, a fim de realizar prospecções futuras a respeito da geração de energia do complexo eólico.

Os objetivos específicos consistem em desenvolver, implementar e validar modelos de previsão de geração de energia eólica em um período de 12 meses à frente. Para isso, propõe-se o desenvolvimento de cinco modelos de aprendizagem supervisionada, em linguagem de programação Python, tais como *eXtreme Gradient Boosting*, Perceptron de Múltiplas Camadas, Regressão por Vetores Suporte, Regressão Linear Rígida e Florestas Aleatórias. Como insumo, serão utilizados dados de geração histórica do Complexo Eólico Pindaí e variáveis meteorológicas derivadas de modelos de reanálise de alta resolução, combinando conjuntos de dados e técnicas de previsão numérica do tempo. Para avaliar a eficácia das cinco metodologias propostas, os resultados são comparados de maneira sistemática utilizando métricas de avaliação, com o objetivo de determinar a modelagem mais adequada para o problema de previsão. A análise de desempenho focaliza na minimização de funções de erro e na avaliação de métricas tradicionais, incluindo o Coeficiente de Determinação (R^2), o Erro Percentual Absoluto Médio (MAPE), Raiz do Erro Quadrático Médio (RMSE) e o Erro Absoluto Médio (MAE).

1.3 Estrutura do Trabalho

O trabalho foi definido em capítulos da seguinte forma. No primeiro e presente capítulo, é introduzido o problema de pesquisa abordado, assim como as motivações e objetivos. O capítulo 2 concerne a revisão da literatura referente aos conceitos fundamentais do aprendizado de máquina, suas categorias de aprendizado, horizonte de previsão, testes de

estacionariedade e comparação de trabalhos semelhantes como *benchmark* para os resultados que foram obtidos nesse estudo. Além disso, abordagens do referencial teórico das metodologias de aprendizado de máquina propostas no trabalho, tais como Regressão Linear Rígida, Florestas Aleatórias, Regressão por Vetores Suporte, Perceptron de Múltiplas Camadas e *eXtreme Gradient Boosting*. Em sequência pela apresentação da conceituação teórica referente a natureza do vento, descrevendo conceitos como variação, previsibilidade, funcionamento das turbinas eólicas relativas a características, projetos e eficiência das turbinas. Seguindo pela parte conceitual sobre a climatologia dos regimes e variabilidade de vento. O capítulo 3 descreve as metodologias utilizadas neste estudo, tais como o levantamento do conjunto de dados referente ao Complexo Eólico Pindaí, utilização dos modelos meteorológicos de reanálise de alta resolução, modelagem das estruturas dos algoritmos de aprendizado de máquina aplicados na previsão e descrição da avaliação das métricas de erro. No capítulo 4 são avaliados os resultados obtidos da análise das séries temporais, da etapa de treinamento dos modelos e das previsões futuras para o ano de 2023. Por fim, o capítulo 5 pontua as conclusões do trabalho, com sugestões para melhorias e projetos futuros.

Capítulo 2 - Referencial Teórico

O referencial teórico é fundamental como a base sobre a qual todos os conceitos, métodos e interpretações do estudo são construídos. No contexto da previsão de geração de energia eólica, o uso de técnicas de *Machine Learning* torna-se eficaz devido à capacidade de lidar com grandes volumes de dados, identificar padrões complexos e adaptar-se a mudanças nas condições ambientais e operacionais. O perfil estocástico dos padrões de vento e de variáveis meteorológicas sugere a aplicação de modelos capazes de analisar dados no escopo da energia eólica.

2.1 Revisão da Literatura

O termo *Machine Learning* (ML) é definido por Marsland [9] e Bishop [10] como um conjunto de técnicas matemáticas que operam através de um processo iterativo de aprendizagem, empregando poder computacional para melhor prever padrões e comportamentos ou realizar classificações e criação de grupos. Conforme o avanço científico, Mitchell [11] forneceu uma definição mais formal: "diz-se que um programa de computador aprende com a experiência E em relação a alguma classe de tarefas T e mede o desempenho P , se seu desempenho em tarefas em T , medido por P , melhora com a experiência E ". Essa definição envolve os elementos do aprendizado de máquinas, como tarefas, desempenho e experiência, destacando a necessidade da melhoria ao longo do tempo. A evolução fez com que fossem cruzadas as linhas entre vários conceitos, no qual Alpaydin [12] enfatiza que o ML se baseia no reconhecimento de padrões e na teoria de aprendizagem computacional em inteligência artificial, bem como na mineração de dados e estatísticas, sendo essas influências convergindo para produzir sistemas capazes de realizar predições (modelagem preditiva) ou tomada de decisão (inferência).

Um dos componentes desse processo de aprendizagem é a noção de generalização, sendo a capacidade de um modelo ter um bom desempenho em dados novos depois de ter estimado o relacionamento entre as variáveis a partir da sua interação com um conjunto de dados de treinamento. O campo do ML se esforça por modelos que se ajustam bem aos dados

de treinamento e por modelos que possam extrapolar esses dados para fazer previsões ou decisões precisas em ambientes gerais, uma abordagem elaborada nos livros clássicos de Bishop [10] e Haykin [7]. À medida que o poder computacional cresceu e os conjuntos de dados se tornaram maiores e mais complexos, o campo do ML se expandiu a partir da criação do *deep learning*. Esse aprendizado é caracterizado como um subconjunto dentro do ML, ganhando destaque por seu sucesso em tarefas que envolvem redes neurais em larga escala, como explorado por Goodfellow, Bengio e Courville em [13], fornecendo uma visão geral das técnicas de aprendizagem profunda que alcançaram o sucesso em campos como visão computacional e processamento de linguagem natural.

O aprendizado de máquina possui três categorias, sendo elencadas como aprendizagem supervisionada, não supervisionada e por reforço. O aprendizado supervisionado pode ser classificado como um subconjunto do aprendizado de máquina, dependendo do conceito de utilizar conjuntos de dados rotulados para treinar algoritmos para classificar dados ou prever resultados com regressão. Para isso, o algoritmo de aprendizado itera sobre um determinado conjunto de dados com resultados conhecidos para aprender os pesos que podem ser aplicados a novos dados para fazer previsões, desde aplicações como o reconhecimento de fala até a previsões no mercado financeiro. O referencial teórico para a aprendizagem supervisionada baseia-se no estabelecimento de um modelo matemático que mapeia entradas para saídas desejadas, a partir da construção de uma função de custo que o algoritmo de aprendizagem visa minimizar. Para isso, são ajustados os pesos através de métodos como a descida de gradiente, com o objetivo de descobrir a função que mais se aproxima da relação entre as variáveis de entrada e as variáveis de saída. Isso permite que o modelo faça previsões sobre novos dados [14].

Um dos trabalhos que esclarece os fundamentos estatísticos de métodos de aprendizagem supervisionada aborda as metodologias de regressão linear, classificação e técnicas de regularização por Hastie, Tibshirani e Friedman [14]. A regressão linear é um dos métodos existentes para predição de variáveis contínuas, explorada no trabalho de Rencher e Schaalje em [15]. Este estudo fornece uma análise de modelos de regressão linear, discutindo os pressupostos da linearidade, a importância dos mínimos quadrados e a interpretação dos coeficientes de regressão. No âmbito da classificação, pode ser citado o livro de Bishop [10], aprofundando nos algoritmos projetados para categorizar dados em

classes predefinidas, discutindo abordagens probabilísticas como o classificador *Naive Bayes* e máquinas de vetor de suporte. Um aspecto da aprendizagem supervisionada é a possibilidade de ocorrência do fenômeno do *overfitting*. Nesta situação, um modelo torna-se muito ajustado com os dados de treinamento e perde sua capacidade de generalizar para novos conjuntos de dados. Para evitar esse problema, se faz necessário o equilíbrio entre o viés e a variância de um modelo, conceito que é articulado por Singh, Asari e Rajasekaran [16].

A interação entre viés e variância é um componente na compreensão do desempenho e da capacidade de generalização desses modelos, conceitos que estão relacionados à habilidade de um modelo se ajustar aos dados de treinamento e generalizar para dados não observados. Um viés elevado é indicativo de um modelo demasiadamente simples, incapaz de capturar a complexidade intrínseca dos dados, levando a uma subestimação dos valores verdadeiros. Por outro lado, uma alta variância sinaliza um modelo que, embora seja preciso no ajuste aos dados de treinamento, falha em sua aplicação a conjuntos de dados novos e não observados anteriormente, devido a uma complexidade exagerada que abrange até mesmo o ruído nos dados de treinamento [17]. A relação entre viés e variância é tradicionalmente articulada através do *trade-off* entre ambos: uma redução no viés geralmente implica um aumento na variância, e vice-versa. O objetivo é alcançar o equilíbrio ótimo, configurando um modelo que se ajuste de maneira adequada aos dados de treinamento enquanto mantém uma capacidade robusta de generalização para novos conjuntos de dados, de maneira que o modelo não possua um baixo viés e alta variância (resultando em *overfitting*) e nem um alto viés e baixa variância (resultando em *underfitting*).

As técnicas de regularização são uma resposta ao dilema do *overfitting*, introduzindo restrições no algoritmo de aprendizagem para impedir que os coeficientes se ajustem ao ruído nos dados de treinamento. O artigo de referência de Tibshirani [18] introduz o método de *Lasso* (operador de seleção e encolhimento mínimo absoluto), sendo uma técnica de regularização que aumenta a precisão e a interpretabilidade de modelos de regressão alterando a complexidade do modelo.

O aprendizado não supervisionado, por sua natureza, foca na identificação de estruturas e padrões em conjuntos de dados que não possuem rótulos ou variáveis-alvo previamente definidas. Sem a direção explícita de uma saída conhecida, os algoritmos de

aprendizado não supervisionado necessitam explorar e discernir a organização interna dos dados, revelando *insights* ocultos ou correlações inesperadas. Esse tipo de abordagem é amplamente aplicado em contextos diversos, como a segmentação de produtos em um caixa automático de uma loja de roupas, na qual os algoritmos agrupam itens com base em suas características visuais ou funcionais, por exemplo.

Um importante conceito aplicado à aprendizagem não supervisionada consiste em descobrir a distribuição ou geometria subjacente dos dados. Um exemplo disso é o *clustering*, na qual os algoritmos tentam agrupar pontos de dados de tal forma que aqueles dentro do mesmo cluster sejam mais semelhantes entre si do que aqueles em outros clusters. Métodos como *K-means* (técnica de *clustering* baseada em centro) e *Density-based spatial clustering of applications with noise* (DBSCAN), um algoritmo de *clustering* baseado em densidade, estão entre as abordagens utilizadas para produzir esses agrupamentos, com conceitos estabelecidos em trabalhos do Aggarwal e Reddy em [19].

A redução de dimensionalidade é um aspecto presente tanto na aprendizagem supervisionada quanto na não supervisionada, visando simplificar o conjunto de dados e reduzindo o número de variáveis consideradas, melhorando a eficiência das operações subsequentes e permitindo uma melhor visualização. No trabalho de Jolliffe em [20], são apresentadas técnicas como a Análise de Componentes Principais (*Principal Component Analysis* - PCA), que transforma os dados em um novo sistema de coordenadas para destacar as direções da variância máxima. Este método é robusto em cenários onde as variáveis originais são altamente correlacionadas ou quando a quantidade de variáveis é tão grande (dimensionalidade alta) que dificulta a identificação das relações entre as observações.

Outra abordagem dentro da aprendizagem não supervisionada é o uso de redes neurais para aprender codificações eficientes dos dados de entrada. São redes projetadas para reconstruir suas entradas a partir de representações compactadas e codificadas e, ao fazer isso, podem aprender representações de recursos. O trabalho de Hastie, Tibshirani e Friedman [14] descreve como essas redes neurais aprendem a capturar as características mais significativas dos dados para tarefas como *denoising* (eliminação de ruídos) ou detecção de anomalias.

A evolução da aprendizagem não supervisionada possibilitou a introdução de técnicas de aprendizagem profunda para *clustering*, que alavancam o poder representacional

das redes neurais, conforme demonstrado nos trabalhos de Xie, Girshick e Farhadi em [21]. Além disso, surgiram estruturas em que duas redes neurais são treinadas simultaneamente em uma configuração competitiva para gerar novas instâncias de dados que são indistinguíveis dos dados reais. Essas estruturas são conhecidas como Redes Adversariais Gerativas (*Generative Adversarial Networks* - GAN), conforme apresentado por Goodfellow *et al.* em [22].

Em relação a aprendizagem por reforço, por sua vez, emergiu no campo do aprendizado de máquina com foco em processos sequenciais de tomada de decisão, na qual um agente interage com o ambiente para maximizar a recompensa cumulativa. Sua formalização como uma abordagem computacional é vista no livro *Reinforcement Learning: An Introduction*, trabalho dos autores Sutton e Barto [23] e considerado como um dos textos fundamentais na área. O trabalho aborda a interação entre um agente e seu ambiente, através do processo chamado de Decisão de Markov (*Markov Decision Process* - MDP), que fornece uma maneira de modelar problemas de tomada de decisão onde os resultados são parcialmente aleatórios e parcialmente controlados pelo tomador de decisão. Um exemplo é o algoritmo *Q-learning*, introduzido por Watkins [24], sendo um método que o agente aprende o valor de tomar uma determinada ação em um determinado estado através da interação direta com o ambiente.

Outro avanço refere-se ao desenvolvimento da *Deep Q-Network* (DQN) por Mnih *et al.* [25], que combinou *deep learning* com *Q-learning*. Essa abordagem permitiu o treinamento de agentes de aprendizagem por reforço para jogar jogos de computador a partir de entradas de pixels, demonstrando o potencial dos métodos de aprendizagem por reforço profundo para lidar com entradas sensoriais de alta dimensão e tarefas complexas de tomada de decisão.

No âmbito da previsão de séries temporais, a aprendizagem por reforço pode ser utilizada para otimizar tarefas sequenciais de tomada de decisão, onde as recompensas futuras dependem de ações tomadas em um contexto temporal. Um exemplo é a abordagem de Moody e Saffell em [26], ao aplicarem a técnica à negociação financeira, demonstrando como os agentes de aprendizagem por reforço podem aprender estratégias de negociação que maximizam o lucro com base em dados históricos de preços. Outra aplicação é na gestão de redes inteligentes por Liu *et al.* [27], em que os agentes aprendem a tomar decisões sobre

distribuição e consumo de energia com base em padrões de demanda e oferta que variam no tempo. Além disso, avanços na aprendizagem por reforço têm explorado a integração do aprendizado hierárquico e do meta-aprendizado para melhorar a escalabilidade e adaptabilidade dos agentes. A aprendizagem por reforço hierárquica, como discutido por Barto e Mahadevan [28], envolve a decomposição de tarefas complexas em subtarefas mais simples, permitindo uma aprendizagem mais eficiente e a transferência de conhecimento entre tarefas.

Dentro do campo do aprendizado de máquina, uma definição importante é a escolha do horizonte de previsão. Esse horizonte refere-se ao período futuro para o qual um modelo de previsão está sendo treinado para fornecer estimativas. Os horizontes de previsão de séries temporais são categorizados em três períodos: curto, médio e longo prazo, cada um servindo a diferentes propósitos e influenciados por fatores específicos [29].

No contexto da geração eólica, as previsões de curto prazo referem-se a períodos que se estendem de algumas horas até alguns dias, geralmente 72 horas. O objetivo do horizonte de curtíssimo prazo pode ser definido através da otimização da operação diária dos parques eólicos, planejamento de manutenção de curto prazo, ajuste de previsões para o mercado de energia de curto prazo e gerenciamento de cargas.

Em relação ao médio prazo, as previsões abrangem de vários dias até algumas semanas, geralmente de 3 dias a 1 mês. Nesse horizonte, o objetivo consiste no planejamento operacional semanal, otimização de manutenção preventiva, planejamento da integração de energia eólica com outras fontes de energia e gestão de recursos energéticos.

A previsão de longo prazo compreende de vários meses até anos à frente. Neste contexto, o objetivo consiste no planejamento estratégico de longo prazo, avaliação de investimentos em infraestrutura, desenvolvimento de novos projetos de energia eólica e estudos de viabilidade econômica. A maioria dos processos de planejamento estratégico e operacional, como orçamento, alocação de recursos e programações para manutenção, é conduzida mensalmente ou trimestralmente, de maneira que a agregação a esses intervalos garante que as informações derivadas dos dados sejam diretamente aplicáveis a esses processos. Um desafio na previsão de longo prazo é por conta do aumento da incerteza e o impacto potencial de variáveis exógenas, como mudanças em padrões de chuvas em determinada região devido a efeitos climáticos como El Niño e La Niña, por exemplo [29].

Levando em consideração esses efeitos climáticos, são utilizadas variáveis meteorológicas derivadas de modelos de reanálise, que englobam dados históricos e previstos, adaptados ao horizonte específico relacionado ao Complexo Eólico Pindaí. Modelos de reanálise de previsão meteorológica de alta resolução são desenvolvidos por serviços meteorológicos internacionais, combinando conjuntos de dados e técnicas de previsão numérica do tempo (*Numerical Weather Prediction* - NWP), para fornecer previsões atmosféricas. Esses modelos integram dados de satélites, sistemas de radar, estações meteorológicas, aviões e boias, utilizando métodos de assimilação dinâmica que aumentam a confiabilidade e a precisão da previsão. Para isso, existem provedores que oferecem previsões que combinam a amplitude dos modelos globais com os modelos de mesoescala, como a Open-Meteo. A previsão numérica do tempo consiste em cálculos que simulam o comportamento da atmosfera com base em princípios fundamentais da dinâmica dos fluidos, exigindo grandes recursos computacionais e são normalmente realizados em supercomputadores, envolvendo a resolução de equações que descrevem o movimento e a interação de partículas atmosféricas [30].

Modelos meteorológicos de alta resolução, utilizam uma combinação de modelos globais (normalmente com resoluções em torno de 11 km) e modelos de mesoescala (com resoluções finas de 1 km). Os modelos são projetados para capturar a dinâmica atmosférica em diferentes escalas, desde sistemas meteorológicos amplos até fenômenos localizados como tempestades ou microclimas, sendo desenvolvidos por serviços meteorológicos internacionais, como o *Deutscher Wetterdienst* (DWD) na Alemanha, NOAA nos Estados Unidos e *Météo-France* na França.

Outro ponto no campo do aprendizado de máquina e séries temporais é relativo à estacionariedade. Seu conceito é aplicado para diversos métodos de previsão estatística que exigem que os dados tenham propriedades estatísticas constantes ao longo do tempo, sendo nos testes tradicionais a verificação de uma estrutura constante de média, variância e autocorrelação. Para métodos como a regressão linear rígida, por exemplo, garantir a estacionariedade é uma etapa recomendável, pois esses modelos assumem que a relação entre as variáveis independentes e a variável dependente é linear e constante ao longo do tempo. Se os dados não forem estacionários, as propriedades estatísticas, como a média e a variância podem mudar ao longo do tempo, o que pode levar a previsões inconsistentes. Em

contraste, as metodologias como florestas aleatórias, regressão por vetores suporte, perceptron de múltiplas camadas e XGBoost, não impõem tal restrição [31].

O teste de *Augmented Dickey-Fuller* (ADF) aborda se uma série temporal é predominantemente estacionária de tendência, ou tem uma raiz unitária, o que significa que exhibe tendências estocásticas. A equação de regressão usada no teste de ADF é descrita na expressão (1):

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-1} + \epsilon_t \quad (1)$$

onde y_t representa a série temporal, Δ é o operador da primeira diferença, t é a tendência temporal e ϵ_t é o termo de erro. O parâmetro δ_i se refere ao coeficiente sobre as diferenças defasadas da série, destinado a capturar a autocorrelação. Os testes de hipótese primária no procedimento ADF concentram-se no parâmetro γ . A hipótese nula H_0 considera que $\gamma = 0$, indicando a presença de uma raiz unitária e, portanto, não estacionária. A hipótese alternativa H_1 sugere que $\gamma < 0$, sendo estacionariedade. A estatística de teste derivada da regressão é então comparada com valores críticos da distribuição de *Dickey-Fuller*, de maneira que se a estatística de teste for mais negativa que o valor crítico, a hipótese nula é rejeitada, sugerindo que a série está estacionária [31].

A prática de reversão, ou seja, a aplicação do processo inverso para retornar os dados à sua escala e contexto originais, é solidamente fundamentada na literatura de séries temporais e modelagem estatística. Box e Jenkins [32] discutem a necessidade de reverter as transformações para interpretar os resultados no contexto original ao usar modelos em séries temporais financeiras e econômicas. Além disso, Hamilton [31] reforça a importância de retornar os dados à sua forma original para a análise final e interpretação dos resultados, destacando que a transformação é um passo intermediário necessário, mas a compreensão completa do fenômeno estudado requer a reversão da série transformada.

A eficácia de qualquer modelo de previsão precisa ser rigorosamente avaliada para validar suas previsões e determinar sua aplicabilidade em cenários reais. A análise das métricas de erro encontradas em estudos semelhantes oferece uma base de referência para validar os resultados obtidos, fornecendo um ponto de comparação essencial para determinar a eficácia dos modelos que serão desenvolvidos neste trabalho. No trabalho de Camelo *et al.*, [33] os autores buscaram realizar previsões de séries temporais da velocidade do vento

em termos de médias mensais no nordeste brasileiro, para quatro locais distintos (Fortaleza, Parnaíba, São Luís e Natal). Foram testados diversos modelos, incluindo o Autorregressivo Integrado de Médias Móveis (ARIMA), o modelo de Holt-Winters (HW) e a utilizando Redes Perceptron de Múltiplas Camadas. Os menores resultados referentes ao método da MLP apresentaram um MAPE de 6,52%, RMSE de 0,37 e MAE de 0,28.

No trabalho de Paula *et al.*, [34] foram realizadas previsões de longo prazo de geração de energia eólica em parques eólicos no nordeste do Brasil, utilizando modelos de aprendizado de máquina como Florestas Aleatórias, Redes Neurais Recorrentes (*Recurrent Neural Network* - RNN) e *eXtreme Gradient Boosting*. De acordo com as métricas de erro utilizadas no trabalho, como MAPE e MAE, o modelo de Florestas Aleatórias resultarem em um MAPE de 12,47% e MAE de 3,93, enquanto o XGBoost em 10,39% e 3,35, respectivamente. A rede recorrente apresentou os menores índices, com MAPE de 4,55% e MAE de 1,34.

Por fim, outro trabalho com aplicação de modelos de aprendizado de máquina para previsão de geração de energia eólica de longo prazo foi apresentado pelos autores Magesh *et al.* [35]. Foram aplicadas as técnicas de Regressão Linear, Florestas Aleatórias e Árvores de decisão, sendo comparadas pelas métricas de R^2 , RMSE, MAE e MSE. Os melhores resultados foram alcançados pela metodologia do *Random Forest*, resultando em um R^2 de 0,9264, RMSE de 0,4447, MAE de 0,3839 e MSE de 0,19776.

A revisão de trabalhos semelhantes na área de previsão de geração eólica permite observar, por exemplo, que valores de RMSE abaixo de 0,5, R^2 acima de 0,90, MAPE abaixo de 6% e MAE abaixo de 0,40 são indicativos de um desempenho preditivo robusto.

Com base nessas observações, o foco deste trabalho será nos modelos de aprendizado supervisionado, escolha que se justifica pela eficácia e pela relevância desses modelos para o estudo em questão. Entre os modelos considerados estão a Regressão Linear Rígida, Florestas Aleatórias, Regressão por Vetores Suporte, Perceptron de Múltiplas Camadas e *eXtreme Gradient Boosting*. Estas técnicas constituem a base do aprendizado de máquina no contexto da análise de séries temporais e são reconhecidas por sua capacidade de lidar com dados complexos e fornecer previsões precisas. Além disso, a inclusão de abordagens avançadas como o XGBoost representa um avanço significativo no estado da arte, destacando sua aplicação na previsão de geração de energia renovável.

No contexto da previsão de geração de energia eólica, o uso extensivo de dados meteorológicos derivados de modelos de reanálise de alta resolução ainda representa uma área insuficientemente explorada, especialmente quando integrado a técnicas avançadas de aprendizado de máquina. Além disso, são feitas análises exaustivas dos hiperparâmetros dos modelos de aprendizado de máquina, enfoque que permite capturar as dinâmicas atmosféricas que influenciam a geração de energia e otimizar os modelos para maximizar sua precisão e generalização.

2.1.1 Regressão Linear

As origens da regressão linear remontam ao século 19, com o trabalho do matemático e estatístico inglês chamado Francis Galton sobre correlação e regressão em direção à média, lançando as bases para a modelagem estatística das relações entre variáveis [36]. A regressão linear simples é a forma mais antiga de análise de regressão, na qual procura-se modelar a relação linear entre uma variável dependente e uma única variável independente. Sua representação matemática é descrita pela fórmula (2):

$$y = \beta_0 + \beta_1 x + \epsilon \quad (2)$$

Nesta formulação, y representa a variável dependente, x a variável independente, β_0 o intercepto em y , β_1 a inclinação da linha e ϵ o termo de erro que captura o desvio dos valores observados da linha de melhor ajuste. O objetivo é estimar os parâmetros β_0 e β_1 que minimizam a soma dos resíduos quadrados, fornecendo uma linha de melhor ajuste que prediz a variável dependente com base na variável independente. De uma perspectiva de aprendizado estatístico, a regressão linear simples é o exemplo inicial de aprendizado supervisionado, contendo uma estrutura em que o modelo é treinado em um conjunto de dados rotulados, consistindo em pares de entrada e saída. O trabalho de James *et al.* [37] ressalta a importância da compreensão da regressão linear simples para suas aplicações em modelagem preditiva e seu papel inicial para entendimento e possíveis avanços para modelos mais complexos.

Apesar de sua simplicidade, diversos estudos foram realizados em áreas de aplicação como economia, finanças e engenharia por sua utilidade na modelagem de variáveis contínuas, na avaliação da contribuição de cada variável dependente e nas análises de tendências. Porém, sua aplicabilidade é limitada por pressupostos que incluem linearidade, independência de erros, homocedasticidade e normalidade dos termos de erro [38]. Violações dessas suposições podem levar a estimativas tendenciosas ou imprecisas, limitando a eficácia do modelo na captura de relações complexas.

A suposição de linearidade assume que existe uma relação linear entre a variável independente (x) e a variável dependente (y), implicando que as mudanças em x resultem em mudanças proporcionais no valor esperado de y . Em termos práticos, essa relação pode ser visualizada como uma linha reta quando y é projetado juntamente com x . A violação dessa suposição, podendo ser curva ou exponencial, prejudica a precisão do modelo uma vez que a regressão linear simples é projetada para ajustar uma linha reta aos dados.

A independência de erros requer que os resíduos do modelo (definidos como as diferenças entre os valores observados e os valores previstos pela regressão linear) sejam independentes entre si. Em outras palavras, o valor de um resíduo para uma observação não deve depender do valor de outro resíduo. Isso é importante em dados de séries temporais em que pontos de dados sequenciais podem estar correlacionados, caso os resíduos não sejam independentes (condição conhecida como autocorrelação), isso pode levar a uma subestimação do erro padrão e, conseqüentemente, a inferências errôneas sobre a significância das variáveis [39].

A homocedasticidade refere-se à suposição de que a variância em torno da linha de regressão é a mesma para todos os valores da variável preditora (x), uniformidade que garante que o modelo permaneça confiável em todos os níveis de x . A heterocedasticidade, que é considerada a violação dessa suposição, indica que o modelo não representa igualmente todas as observações levando a estimativas ineficientes [40].

A normalidade dos termos de erro estipula que os resíduos do modelo são normalmente distribuídos, sendo importante para a validação de estatísticas inferenciais associadas à regressão, como o cálculo de intervalos de confiança e o teste de hipóteses sobre os coeficientes de regressão. O Teorema do Limite Central (*Central Limit Theorem* - TCL) assegura que a distribuição amostral da média tende à normalidade com um tamanho

amostral suficientemente grande, em que desvios significativos da normalidade nos resíduos podem afetar os intervalos de confiança e os valores de p (medida de significância), tornando-os não confiáveis [41].

O reconhecimento e o teste desses pressupostos são partes do processo de apuração na análise de regressão linear. Para isso, são utilizadas ferramentas como gráficos residuais, o teste de autocorrelação de Durbin-Watson e o teste de normalidade de Shapiro-Wilk, auxiliando na avaliação dessas condições. Quando violações são detectadas, são aplicadas estratégias alternativas de modelagem, como transformar variáveis para alcançar linearidade ou empregar métodos de regressão não-linear para mitigar o impacto da heterocedasticidade ou resíduos não normais, para garantir a validade e a confiabilidade das inferências do modelo [42] e [43].

2.1.1.1 Regressão Linear Rígida

Uma das evoluções das técnicas de regularização no aprendizado estatístico foi o desenvolvimento da Regressão Linear Rígida ou *Ridge Regression* em inglês. O objetivo dessa modelagem consistia no tratamento dos desafios da multicolinearidade e eventualmente do *overfitting* em modelos lineares. A regressão linear rígida possui relação com a regularização de Tikhonov, como uma solução para a instabilidade e falta de confiabilidade das estimativas em modelos de regressão linear causada por preditores altamente correlacionados. Essa técnica modifica o objetivo dos mínimos quadrados adicionando uma penalidade proporcional ao quadrado da magnitude dos coeficientes. Esta restrição está relacionada com a premissa da existência de alguma regularidade nos dados, visando balancear o efeito do dilema viés-variância e mitigando o risco de *overfitting*. A introdução da regressão rígida é abordada no trabalho de Hoerl e Kennard [44], na qual a incorporação de um parâmetro de regularização α , que controla a magnitude da penalização aplicada aos coeficientes, proporciona uma alternativa aos mínimos quadrados ordinários (MQO). Esta abordagem melhora o desempenho preditivo e a interpretabilidade do modelo, especialmente em situações em que os preditores são altamente correlacionados. Na regressão rígida, os coeficientes são estimados minimizando uma função de custo que inclui

uma penalização para os coeficientes. A equação da estimativa dos coeficientes considerando o critério estatístico utilizado é descrita pela fórmula da pseudoinversa de Moore-Penrose (3):

$$\beta^{ridge} = (X^T X + \alpha I)^{-1} X^T y \quad (3)$$

onde I é a matriz identidade, x a matriz de *design* (ou matriz de características), y o vetor de respostas observadas e α é o parâmetro de regularização que controla a magnitude da penalização aplicada aos coeficientes [45].

A aplicação desta técnica de regressão abrange diversos domínios, desde modelagem preditiva em finanças até processamento de linguagem natural e reconhecimento de imagem em aprendizado de máquina, por exemplo. A técnica de regularização permite a obtenção de soluções estáveis mesmo quando a matriz inversa não existe, o que ocorre quando o número de entradas é maior que o número de observações. A integração desta regressão em estruturas de aprendizado de máquina foi facilitada por avanços em técnicas computacionais para otimização de hiperparâmetros, como validação cruzada e algoritmos de busca em grade, permitindo a seleção eficiente da variável α e otimizando o desempenho do modelo em diversos conjuntos de dados para tarefas de previsão [14].

Por outro lado, a regressão rígida depende em demasia da seleção do parâmetro de regularização λ , exigindo um ajuste calibrado para equilibrar a relação entre o viés e variância. Embora a regressão possa reduzir a magnitude dos coeficientes, ela não produz modelos esparsos, pois não define os coeficientes exatamente como zero. Essa limitação foi um dos problemas que estimulou o desenvolvimento de técnicas alternativas de regularização, como a Regressão *Lasso*, introduzida no trabalho de Aggarwal e Reddy [19].

A otimização de hiperparâmetros envolve a busca sistemática pela melhor combinação de hiperparâmetros que minimize o erro em um conjunto de validação, garantindo que o modelo generalize bem para dados não vistos. Os hiperparâmetros abordados para a regressão rígida são detalhados na Tabela 1, tais como *alpha*, *fit_intercept*, *normalize*, *max_inter*, *tol*, *solver* e *random_state*.

Tabela 1 - Hiperparâmetros e Definições - Regressão Linear Rígida

Hiperparâmetros	Definição
<i>alpha</i>	Controla a força do termo de regularização L2.
<i>fit_intercept</i>	Especifica se a interceptação do modelo deve ser calculada. Se definido como falso, nenhuma interceptação será usada nos cálculos, pressupondo que os dados estejam centralizados. Isso pode afetar a capacidade do modelo de ajustar dados com média diferente de zero.
<i>normalize</i>	Determina se os regressores devem ser normalizados antes de ajustar o modelo. A normalização pode melhorar a estabilidade numérica do modelo e garantir que o termo de regularização seja aplicado uniformemente entre as variáveis.
<i>max_iter</i>	Define o número máximo de iterações de convergência.
<i>tol</i>	Define o limite para a convergência. Valores menores levam a soluções mais precisas, mas podem aumentar o tempo computacional necessário para a convergência.
<i>solver</i>	Especifica o algoritmo a ser usado para ajustar o modelo.
<i>random_state</i>	Garante a reprodutibilidade definindo uma semente para o gerador de números aleatórios.

2.1.1.2 Regressão Linear Lasso

O desenvolvimento da Regressão *Lasso* (*Least Absolute Shrinkage and Selection Operator*) foi introduzida por Aggarwal e Reddy em [19] e influenciou significativamente o aprendizado estatístico, oferecendo uma metodologia para a criação de modelos na presença de multicolinearidade e alta dimensionalidade. Essa técnica amplia o conceito de regularização ao introduzir uma penalidade *L1* sobre o valor absoluto dos coeficientes, ao

contrário da penalidade $L2$ da regressão rígida em seus coeficientes. Essa distinção permite que os coeficientes sejam reduzidos e defina alguns como zero, realizando a seleção de variáveis. A função objetivo *Lasso* é definida pela expressão (4):

$$\min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (4)$$

onde λ é o parâmetro de regularização que controla o grau de retração aplicado aos coeficientes.

O trabalho aborda o duplo papel na regularização e seleção de variáveis, destacando sua utilidade em modelos carregados pela maldição da dimensionalidade [19]. Posteriormente, trabalhos teóricos exploraram as propriedades da Regressão *Lasso*, incluindo sua consistência, esparsidade e desempenho preditivo. Um destes trabalhos foi a introdução da Regressão *Elastic Net* por Zou e Hastie em [46], ampliando a estrutura de *Lasso* e combinando as penalidades $L1$ e $L2$ para superar limitações de preditores altamente correlacionados ou quando o número de preditores excede o número de observações. Em bioinformática, por exemplo, a técnica de *Lasso* tem sido importante na identificação de preditores em conjuntos de dados de alta dimensão, como dados de expressão gênica, ajudando a identificar genes relevantes associados a certas doenças, visto em Hastie, Tibshirani e Friedman em [14]. No mercado financeiro, um exemplo é o trabalho de Brodie *et al.* [47], no qual ilustra como técnicas de regularização semelhantes a *Lasso* podem contribuir para a seleção estável de carteiras, destacando a relevância do método na gestão do risco de carteiras financeiras.

Um dos problemas da técnica está na determinação do valor ótimo para o parâmetro de regularização λ , impactando o desempenho do modelo e a seleção de preditores. Para isso, técnicas de validação cruzada são utilizadas como manobra de desvio, mesmo sendo considerado computacionalmente intensivo para grandes conjuntos de dados. A evolução contínua é caracterizada por esforços para aumentar sua aplicabilidade e eficácia, incluindo avanços recentes como o *Lasso* adaptativo, que atribui pesos diferentes à penalidade $L1$ para cada coeficiente, melhorando a precisão de seleção de variáveis do método [48].

2.1.2 Florestas Aleatórias

As Florestas Aleatórias, ou *Random Forest* (RF) em inglês, originaram-se através do conceito das árvores de decisão e do *bagging* (ensacamento). O artigo de Breiman [49] introduziu as bases ao propor um conjunto de árvores de decisão que são construídas individualmente, usando amostras *bootstrap* (técnica de reamostragem estatística que envolve amostragem aleatória de um conjunto de dados com substituição) dos dados e subconjuntos aleatórios de características em cada divisão. Essa abordagem retrata a alta variância e as tendências de *underfitting* das árvores de decisão individuais, melhorando a precisão da previsão por meio da agregação de múltiplas árvores. A base teórica das Florestas Aleatórias está enraizada na lei dos grandes números, em que o erro de previsão do conjunto diminui à medida que o número de árvores aumenta, desde que as árvores não estejam correlacionadas.

O centro desta técnica consiste em construir uma grande quantidade de árvores de decisão durante o tempo de treinamento e emitir o modo das classes (no caso de classificação) ou a previsão média (em regressão) das árvores individuais. A aleatoriedade é inserida de duas maneiras: através do *bootstrapping* dos dados e através da seleção de um subconjunto aleatório de entradas para divisão em cada nó, garantindo a diversidade entre as árvores do conjunto. Essa metodologia é detalhada no trabalho de Hastie, Tibshirani e Friedman [14] e enfatiza o equilíbrio que as Florestas Aleatórias estabelecem entre a redução do viés e da variância, destacando a eficácia em uma ampla gama de conjuntos de dados e configurações de problemas.

A capacidade das Florestas Aleatórias de lidar com dados de alta dimensão, pontuação de importância de variáveis e capacidade de interpretação de modelos a torna muito usual, como no contexto da previsão de séries temporais empregadas para capturar relações não lineares complexas e interações entre variáveis defasadas no tempo, como estudos que ilustram a utilidade das Florestas Aleatórias na previsão de séries temporais financeiras através da adaptabilidade às complexidades dos dados temporais, no trabalho de Sezer e Ozbayoglu em [50].

Os hiperparâmetros abordados para as florestas aleatórias são detalhados na Tabela 2, tais como *n_estimators*, *criterion*, *max_depth*, *min_samples_split*, *min_samples_leaf*, *min_weight_fraction_leaf*, *max_features*, *max_leaf_nodes*, *min_impurity_decrease*, *bootstrap*, *n_jobs*, *random_state*, *verbose* e *max_samples*.

Tabela 2 - Hiperparâmetros e Definições - Florestas Aleatórias

Hiperparâmetros	Definição
<i>n_estimators</i>	Define o número de árvores na floresta.
<i>criterion</i>	Determina a função usada para medir a qualidade de uma divisão.
<i>max_depth</i>	Define a profundidade máxima das árvores individuais.
<i>min_samples_split</i>	Especifica o número mínimo de amostras necessárias para dividir um nó interno.
<i>min_samples_leaf</i>	Define o número mínimo de amostras necessárias para estar em um nó folha.
<i>min_weight_fraction_leaf</i>	Define a fração mínima ponderada da soma total dos pesos necessários para estar em um nó folha.
<i>max_features</i>	Determina o número de recursos a serem considerados ao procurar a melhor divisão.
<i>max_leaf_nodes</i>	Limita o número de nós folha em cada árvore.
<i>min_impurity_decrease</i>	Define um limite para a divisão de um nó.
<i>bootstrap</i>	Determina se amostras de <i>bootstrap</i> são usadas ao construir árvores.
<i>n_jobs</i>	Especifica o número de trabalhos a serem executados em paralelo.

<i>random_state</i>	Garante a reprodutibilidade definindo uma semente para o gerador de números aleatórios.
<i>verbose</i>	Controla o detalhamento da saída do modelo durante o treinamento.
<i>max_samples</i>	Especifica o número de amostras a serem extraídas do conjunto de dados total para treinar cada estimador base.

2.1.3 Máquina de Vetor Suporte

O conceito das Máquinas de Vetor de Suporte ou *Support Vector Machines* em inglês, tem suas raízes na década de 1960. A ideia básica do SVM, que envolve a separação linear de dados em um espaço de características por um hiperplano, foi introduzida por Vladimir Vapnik e Alexey Chervonenkis na União Soviética no começo dos anos 60 [51]. No entanto, foi na década de 1990 que a técnica ganhou notoriedade e foi amplamente difundida, especialmente após a introdução dos *kernels* não-lineares. Esses *kernels* permitiram que os SVMs fossem aplicados a problemas não-lineares, mapeando os dados para um espaço de alta dimensão onde uma separação linear poderia ser encontrada [52].

O artigo de Cortes e Vapnik [53] em 1995, lançou as bases para a aplicação prática do SVM, detalhando o uso de hiperplanos para separar pontos de dados em um espaço de alta dimensão. Isso foi seguido por uma extensa pesquisa sobre os algoritmos de otimização usados para resolver o problema SVM, como o algoritmo Otimização Sequencial Mínima (*Sequential Minimal Optimization* - SMO) proposto por Platt [54] em 1998, que melhorou a eficiência computacional do SVM.

No centro da abordagem das máquinas de vetor suporte está o conceito de encontrar um hiperplano que melhor separe os pontos de dados de diferentes classes. O hiperplano é definido por um subconjunto dos dados de treinamento, conhecidos como vetores de suporte, que são os elementos críticos que influenciam a posição e a orientação do hiperplano. O algoritmo pode ser formulado como um problema de otimização, onde o objetivo é encontrar um hiperplano ou uma função de regressão que se ajusta aos dados de modo que a maioria

dos pontos esteja dentro de uma faixa especificada pela margem de tolerância [53]. O problema de otimização pode ser expresso pela fórmula (5):

$$\text{minimizar } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (5)$$

onde w é o vetor de pesos, o hiperparâmetro C controla a penalização dos desvios, ε é a margem de tolerância e $\xi_i + \xi_i^*$ são as variáveis de folga que permitem que algumas previsões estejam fora da margem de tolerância.

Para lidar com dados não linearmente separáveis, o SVM emprega o *kernel*. Os *kernels* são funções que transformam os dados de entrada em um espaço de maior dimensão, onde se espera que os dados se tornem linearmente separáveis. Os *kernels* mais comuns utilizados são: polinomial, linear, função de base radial (*Radial Basis Function* - RBF) e sigmoide. Cada *kernel* tem suas próprias características e aplicações. Por exemplo, o *kernel* linear é usado quando os dados são aproximadamente linearmente separáveis no espaço original, enquanto o *kernel* RBF é usado para capturar relações complexas e não lineares nos dados [53].

No contexto da previsão de séries temporais, a natureza não-linear e não-estacionária de muitos conjuntos de dados torna o SVM uma escolha adequada devido à sua flexibilidade. Essa técnica tem sido usada, por exemplo, para prever ações nos mercados financeiros. Isso foi abordado na pesquisa de Tay e Cao em [55], destacando a eficácia na previsão de séries temporais financeiras em relação a métodos tradicionais como os modelos autorregressivos integrados de média móvel.

Outra abordagem é relativa a problemas que exigem previsões em tempo real. Suykens e Vandewalle [56] introduziram o *Least Squares Support Vector Machines* (LS-SVM), reformulando o problema SVM em um conjunto de equações lineares, tornando-o mais eficiente computacionalmente. A introdução de novas técnicas de otimização aprimorou ainda mais os recursos do SVM, como o método de descida de coordenadas duplas (*Dual Coordinate Descent* - DCD) e a descida de gradiente estocástico (*Stochastic Gradient Descent* - SGD), permitindo que o SVM seja dimensionado para grandes conjuntos de dados e resolver problemas com milhões de exemplos de treinamento de forma eficiente [57].

A Regressão por Vetores Suporte (*Support Vector Regression - SVR*) foi concebida para lidar com problemas de regressão, estendendo a capacidade do SVM para prever valores contínuos em vez de rótulos discretos. A abordagem básica da SVR envolve a criação de um hiperplano ou função de regressão que minimize o erro preditivo enquanto mantém a simplicidade do modelo através de uma penalização baseada na norma dos coeficientes [52].

O funcionamento baseia-se em uma função de perda específica, chamada função de perda *epsilon*-insensível, que ignora erros menores que um determinado valor *epsilon*. Essa característica confere à SVR a capacidade de se concentrar em grandes desvios, enquanto pequenas variações são desconsideradas, promovendo um modelo que é robusto a ruídos nos dados [58]. Além disso, a SVR utiliza vetores de suporte, que são os pontos de dados que se encontram fora da faixa de *epsilon*, para definir a função de regressão, o que garante que o modelo seja definido apenas pelos dados mais críticos, reduzindo o risco de *overfitting* [59].

A principal vantagem da SVR sobre outras técnicas de regressão reside em sua flexibilidade e robustez ao lidar com dados de alta dimensionalidade e problemas não-lineares. Essa flexibilidade é obtida através da utilização de funções de *kernel*, que mapeiam os dados de entrada para um espaço de alta dimensão onde uma função de regressão linear pode ser aplicada [60] e [61].

A configuração e execução da pesquisa em grade (*Grid Search*) permite a otimização dos hiperparâmetros da SVR, testando diversas combinações para encontrar a melhor configuração que maximiza o desempenho do modelo. Os hiperparâmetros testados são detalhados na Tabela 3, tais como *C*, *gamma*, *kernel*, *degree*, *coef0*, *epsilon* e *shrinking*.

Tabela 3 - Hiperparâmetros e Definições – Regressão por Vetores Suporte

Hiperparâmetros	Definição
<i>C</i>	Controla a penalização dos erros de predição. Valores menores permitem uma margem maior com mais erros, incentivando um modelo mais generalizado. Valores maiores tentam minimizar os erros de predição, o que pode levar ao <i>overfitting</i>
<i>gamma</i>	Define a influência de um único exemplo de treinamento. Valores altos resultam em uma influência mais localizada, enquanto valores

	<p>baixos fazem com que a influência se estenda mais longe. No contexto do <i>kernel</i> RBF, <i>gamma</i> controla o alcance de cada ponto de suporte.</p>
<i>kernel</i>	<p>O <i>kernel</i> é uma função que transforma os dados de entrada em um espaço de maior dimensão, onde se espera que os dados se tornem linearmente separáveis. Cada <i>kernel</i> tem suas características específicas: o <i>kernel</i> linear é usado quando os dados são aproximadamente linearmente separáveis; o <i>kernel</i> RBF para capturar relações não lineares; o <i>kernel poly</i> (polinomial) é adequado para dados que seguem uma tendência polinomial; e o <i>kernel</i> sigmoide é usado em casos específicos de classificação binária.</p>
<i>degree</i>	<p>Parâmetro relevante apenas para o <i>kernel</i> polinomial e define o grau do polinômio utilizado na transformação dos dados.</p>
<i>coef0</i>	<p>Usado nos <i>kernels</i> polinomial e sigmoide, ajustando a complexidade do modelo (termo independente adicionado na função de <i>kernel</i>).</p>
<i>epsilon</i>	<p>Define uma faixa de valores em que não há penalização associada à função de perda na regressão SVR. Esse parâmetro controla a largura da margem de tolerância, onde não há penalidade para erros de predição.</p>
<i>shrinking</i>	<p>Heurística que pode ser ativada ou desativada para acelerar o treinamento do modelo. Quando ativada, a heurística de <i>shrinking</i> ignora vetores de suporte que não são mais relevantes, acelerando o processo de otimização.</p>

2.1.4 Perceptron de Múltiplas Camadas

O Perceptron de Múltiplas Camadas (*Multilayer Perceptron* - MLP) é uma classe de redes neurais artificiais *feedforward* (*Feedforward Neural Networks* - FFNN). A MLP consiste em pelo menos três camadas de nós, sendo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. A MLP distingue-se do perceptron de camada única em sua capacidade de resolver problemas não-lineares. As origens das MLP remontam à década de 1940 com o trabalho de McCulloch e Pitts [62], que propuseram um modelo computacional para redes neurais baseado em algoritmos e limiares. A técnica de retropropagação (*backpropagation*) em redes neurais de múltiplas camadas foi proposta por Werbos, em 1974 [63]. Este trabalho é considerado um marco na história do aprendizado de máquina e redes neurais, pois a retropropagação se tornou o algoritmo fundamental para o treinamento de redes neurais profundas, amplamente utilizadas nas aplicações. A retropropagação é um algoritmo de otimização que utiliza o método do gradiente descendente para minimizar a função de custo de uma rede neural. A ideia central do algoritmo é calcular o gradiente do erro em relação aos pesos da rede, propagando esse erro a partir da camada de saída para as camadas anteriores. Isso permite que os pesos sejam ajustados de forma iterativa para minimizar o erro global da rede [63].

Os conceitos fundamentais de redes neurais foram extensivamente explorados no livro *Neural Networks and Learning Machines* de Haykin [7]. Este livro é reconhecido como referência no campo das redes neurais artificiais e aprendizado de máquina, fornecendo uma base teórica e compreensão profunda dos princípios que governam o funcionamento e a aplicação dessas redes.

A MLP consiste em várias camadas de nós, cada uma conectada à próxima camada. Cada nó, ou neurônio na rede, aplica uma soma ponderada de suas entradas e transmite o resultado através de uma função de ativação não-linear. Haykin discute a importância das funções de ativação não lineares nas camadas ocultas da MLP, como a função sigmoide e a tangente hiperbólica. Essas funções introduzem a não linearidade necessária para que a MLP possa modelar e aprender relações complexas nos dados, superando a limitação do perceptron de camada única, que resolve problemas linearmente separáveis. A escolha da

função de ativação pode afetar a capacidade da rede de modelar relacionamentos complexos [64]

A arquitetura da MLP é caracterizada pelo número de camadas e pelo número de neurônios em cada camada. A camada de entrada recebe os dados, as camadas ocultas transformam a entrada em um conjunto de recursos e a camada de saída produz as previsões finais. A complexidade do modelo aumenta com o número de camadas e neurônios, permitindo capturar padrões intrincados nos dados. O livro *Pattern Recognition and Machine Learning* de Bishop [10], fornece uma explicação detalhada das arquiteturas de redes neurais e seus fundamentos teóricos, tornando-se também uma referência fundamental no campo do aprendizado de máquina e reconhecimento de padrões.

O treinamento da MLP envolve um processo de ajuste dos pesos das conexões sinápticas para minimizar o erro entre as previsões da rede e os valores reais dos dados de treinamento. Este processo é conduzido principalmente pelo algoritmo de *backpropagation*, que calcula o gradiente da função de perda em relação a cada peso da rede [65]. O algoritmo, juntamente com um método de otimização como a descida de gradiente, ajusta iterativamente os pesos na direção que reduz o erro, permitindo que a rede aprenda a partir dos dados [64].

A descida de gradiente é um método de otimização que atualiza os pesos da rede com base na magnitude e na direção dos gradientes calculados. A taxa de aprendizagem η é o hiperparâmetro que controla o tamanho dos passos dados de cada iteração em direção ao mínimo da função de perda. Uma taxa de aprendizagem muito alta pode fazer com que o modelo oscile em torno do mínimo ou até mesmo divirja, enquanto uma taxa muito baixa pode resultar em um processo de treinamento muito lento, podendo permanecer preso em mínimos locais [66].

As técnicas de regularização são utilizadas em redes neurais de múltiplas camadas para prevenir o overfitting, melhorar a generalização do modelo e assegurar que ele generalize bem em novos. A regularização $L2$, também conhecida como penalização de queda de peso, adiciona uma penalidade proporcional ao quadrado da magnitude dos coeficientes dos pesos na função de perda [67]. Isso incentiva a rede a manter os pesos pequenos, o que pode melhorar a generalização ao reduzir a complexidade do modelo.

Outra técnica recentemente introduzida é o *dropout*, que envolve a omissão aleatória de neurônios durante o treinamento [68]. Isso força a rede a não depender excessivamente de nenhum neurônio específico, promovendo uma rede mais robusta e menos propensa ao *overfitting*. Seu funcionamento ocorre a partir do desligamento aleatório de uma fração dos neurônios durante o treinamento. Isso significa que, em cada iteração, diferentes subconjuntos de neurônios são usados para atualizar os pesos. O efeito é que o modelo não se torna excessivamente dependente de qualquer conjunto específico de neurônios ou características, forçando a rede a aprender representações mais robustas e distribuídas.

A configuração e execução do modelo perceptron de múltiplas camadas leva em consideração a escolha e definição de parâmetros importantes. Com isso, os parâmetros testados são detalhados na Tabela 4, tais como camada de entrada, camada oculta, camada de saída, quantidade de neurônios, função de ativação, tamanho do lote, coeficiente L2 de regularização, otimizador, função de perda, taxa de aprendizado, taxa de *dropout*, *early stopping* e bibliotecas do Python.

Tabela 4 - Hiperparâmetros e Definições - Perceptron de Múltiplas Camadas

Parâmetros	Definição
Camada de entrada	Primeira camada da rede que recebe os dados de entrada. Cada neurônio corresponde a uma característica dos dados.
Camada oculta	Camadas intermediárias que aprendem representações internas dos dados. Contêm neurônios que processam e transformam as informações recebidas.
Camada de saída	Última camada da rede que produz a predição final. Em problemas de regressão, geralmente possui um neurônio com uma função de ativação linear.
Quantidade de neurônios	Número de neurônios em cada camada, ajustado para capturar a complexidade do problema. Determinado frequentemente através de experimentação e validação.

Função de ativação	Função que transforma a entrada de um neurônio antes de passá-la para a próxima camada. ReLU é comum para camadas ocultas e linear para a camada de saída.
Tamanho do lote	Número de amostras de treinamento usadas em uma única atualização dos pesos do modelo. Tamanhos comuns variam de 32 a 256.
Coefficiente L2 de regularização	Penaliza grandes pesos no modelo para prevenir <i>overfitting</i> , mantendo os pesos pequenos.
Otimizador	Algoritmo que ajusta os pesos do modelo durante o treinamento. Adam é popular devido à sua eficiência e baixos requisitos de memória.
Função de perda	Métrica usada para calcular o erro entre as previsões e os valores reais. MSE é comum em regressão.
Taxa de aprendizado	Controla o tamanho dos passos dados na direção dos gradientes durante a atualização dos pesos. Pequenas taxas de aprendizado conduzem a uma convergência mais estável.
Taxa de <i>dropout</i>	Fração de neurônios desativados aleatoriamente durante o treinamento para prevenir <i>overfitting</i> .
<i>Early stopping</i>	Técnica para interromper o treinamento se o desempenho no conjunto de validação não melhorar após um número definido de épocas consecutivas.
Bibliotecas	<i>Keras</i> e <i>TensorFlow</i> são bibliotecas usadas para construir e treinar redes neurais. <i>Keras</i> fornece uma interface de alto nível, e <i>TensorFlow</i> executa operações de aprendizado profundo.

2.1.5 eXtreme Gradient Boosting

O *eXtreme Gradient Boosting* (XGBoost) estabeleceu-se como um método de aprendizado de máquina eficiente para uma gama de tarefas preditivas, em razão de seu desempenho em dados estruturados e sua eficácia na previsão de séries temporais. Originado da necessidade de implementações de alto desempenho de aumento de gradiente, seu início pode ser rastreado até a introdução das *Gradient Boosting Machine* (GBM) por Friedman em [69], no qual o foco era minimizar uma função de perda diferenciável para um problema de aprendizagem preditiva. A jornada do GBM para o XGBoost envolveu uma série de refinamentos destinados a melhorar a velocidade e o desempenho. O aumento do gradiente estocástico de Friedman [69], que introduziu a randomização no processo de ajuste para melhorar as taxas de convergência e a precisão preditiva, lançou as bases para estas melhorias futuras. Dessa forma, o XGBoost se baseou nesses avanços, introduzindo uma formalização de um modelo mais regularizado para controlar o *overfitting*, proporcionando melhor desempenho. Essa introdução foi feita por Chen e Guestrin [8], definido como um sistema escalável para aumento de gradiente, capaz de resolver milhões de instâncias além do alcance de outros algoritmos na época.

A força do XGBoost está em sua escalabilidade e seu manuseio de estruturas de dados esparsas, assim como sua capacidade de gerenciar vários tipos de problemas de modelagem preditiva de forma eficiente. Sua estrutura de aumento de gradiente é otimizada através do uso de derivadas de segunda ordem, que fornecem uma aproximação diversificada da curvatura da função de perda do que os métodos tradicionais de aumento de gradiente. Essa abordagem tem sido detalhada na literatura que explora a profundidade e a amplitude dos métodos de aprendizagem de conjunto dentro do aprendizado de máquina, no trabalho desenvolvido por Rencher e Schaalje [15].

Uma das características dessa técnica são seus recursos de otimização de sistema, incluindo computação paralela e distribuída, que o torna mais rápido e escalável do que outras implementações de aumento de gradiente. Ele emprega um algoritmo sensível à esparsidade para lidar com diferentes formatos de dados, o que é benéfico ao lidar com conjuntos de dados extensos com valores ausentes. Além disso, o *design* do sistema leva em

consideração a consciência de cache, que envolve a otimização dos algoritmos e das estruturas de dados para maximizar o uso eficiente da hierarquia de memória cache do hardware, melhorando assim a eficiência e a velocidade de processamento. Este *design* organiza as estruturas de dados e os algoritmos de maneira a aproveitar ao máximo a arquitetura do *hardware*, conforme abordado em detalhes por Chen e Guestrin [8].

No domínio da predição de séries temporais, o XGBoost tem abordado os desafios de autocorrelação presentes nos dados temporais, capturando dependências não-lineares e sem a necessidade da especificação manual de variáveis defasadas, sendo considerada como estado da arte dentro das predições em aprendizado de máquina. Nos últimos anos, estudos experimentais têm aplicado o XGBoost a uma série de problemas como previsão de séries temporais financeiras, demonstrando assertividade na previsão de preços de ações, no trabalho de Krauss, Do e Huck [70], superando modelos econométricos clássicos. No âmbito da previsão de demanda de energia, a técnica tem sido empregada para prever o consumo de eletricidade com um grau significativo de precisão, beneficiando-se da capacidade do algoritmo de modelar relações de carga não lineares, abordado por Wang *et al.* [71] e resultados interessantes em relação a demanda de varejo, analisado por Bojer em [72].

Em contrapartida, o modelo possui uma alta complexidade e a interpretabilidade de seus resultados têm sido objetos de estudo, pela sua natureza de conjunto que combina as previsões de muitas árvores de decisão, tornando a interpretação das decisões do modelo complexas. Para isso, avanços no campo da ciência de dados como os valores SHAP (*SHapley Additive exPlanations*), introduzidos por Lundberg e Lee em [73], começaram a fornecer *insights* sobre a importância dos recursos e os processos de tomada de decisão dentro dos modelos XGBoost, tornando-os mais transparentes, de maneira que é atribuído a cada recurso um valor de importância para uma previsão específica. A versatilidade do método em lidar com sazonalidades e tendências sem a especificação explícita de componentes temporais o diferencia dos modelos tradicionais de séries temporais, abordado nos estudos de Hyndman e Athanasopoulos [74]. Além disso, estudos de Taylor e Letham [75] demonstram que esse método funciona bem quando combinado com técnicas de decomposição, aprimorando suas capacidades preditivas ao permitir que ele se concentre na modelagem de resíduos que contêm padrões complexos e não lineares.

A análise de séries temporais exige a inclusão de dados históricos e outras características temporais conforme definido pelo modelo. A relevância desse processo de preparação de dados é enfatizada na literatura, de maneira que a importância de converter as séries em um problema de aprendizado supervisionado é mencionada em Bontempi *et al.* [76]. A exploração de Abassi [77] sobre a utilização do *eXtreme Gradient Boosting* para previsão de carga de curto prazo apresenta uma abordagem no domínio das operações de redes inteligentes, aproveitando os dados semanais de séries temporais do Operador do mercado de energia australiano (*Australian Energy Market Operator - AEMO*), de forma que a inovação metodológica está na dupla aplicação do XGBoost para extração e previsão de recursos.

Portanto, o algoritmo *eXtreme Gradient Boosting* [8] é baseado em árvore de decisão, usando a estrutura do aumento do gradiente. Essa técnica está situada no espectro dos classificadores de conjunto chamado *Ensemble*, combinando múltiplos modelos preditivos para aprimorar a acurácia em relação a modelos individuais, com a premissa de que modelos fracos podem ser integrados para formar um modelo forte.

Dentre os algoritmos para aprendizado em conjunto (*ensemble*), os dois principais métodos são chamados de *bagging* e *boosting*. O *bagging*, abreviação de *Bootstrap Aggregating*, introduzido por Breiman em [78], baseia-se na ideia de melhorar a estabilidade e a precisão do modelo treinando vários submodelos em diferentes subconjuntos do conjunto de dados original. Esses subconjuntos são gerados por meio de amostragem *bootstrap*, processo que permite a substituição, garantindo diversidade entre os conjuntos de treinamento. A previsão final é obtida agregando as previsões de todos os submodelos, normalmente através de votação simples para tarefas de classificação ou média para regressão. O apelo teórico do *bagging* reside em sua capacidade de reduzir a variância sem aumentar substancialmente o viés, agregando vários modelos treinados em subconjuntos variados de dados para mitigar o risco de *underfitting* em modelos com alta variância.

O *boosting*, em contrapartida, opera com a premissa de melhorar sequencialmente o desempenho do conjunto, concentrando-se em casos que os modelos anteriores classificaram erroneamente. Essa técnica envolve o treinamento de uma sequência de submodelos simples, com cada submodelo subsequente ajustando seu foco com base nos erros de seus antecessores. A agregação desses submodelos resulta em um modelo de conjunto forte que

exibe viés e variância reduzidos. A matemática do *boosting* está inserida em sua função objetivo, que busca minimizar a perda exponencial, oferecendo um caminho para a compreensão do mecanismo de correção de erros do algoritmo [78].

Dessa forma, as árvores de decisão funcionam como blocos de construção (submodelos), sendo cada árvore composta por nós e ramificações. Sua construção inicia-se com um único nó raiz e, a partir disso, ocorre o particionamento em subconjuntos com base em variáveis de entrada selecionadas. O particionamento continua até que um critério de interrupção seja atendido, resultando em previsões finais chamadas de folhas [78]. Matematicamente [8], considerando um conjunto de dados D com n observações e entradas de dimensão m , o objetivo é dividir D em subconjuntos D_L e D_R que minimizam uma função de perda L . Para uma dada divisão, a estrutura da árvore pode ser representada na fórmula (6):

$$\begin{aligned} D_L &= \{(x_i, y_i) \mid x_{ij} \leq \theta\} \\ D_R &= \{(x_i, y_i) \mid x_{ij} > \theta\} \end{aligned} \quad (6)$$

onde x_{ij} é j -ésimo variável da i -ésima observação e θ é o limite para a divisão. A construção da árvore continua recursivamente para cada nó filho até que os critérios de parada sejam atendidos.

Os critérios de divisão determinam como são realizados os particionamentos dos dados em cada nó, para encontrar a melhor divisão com base na otimização da função de perda. Para problemas de regressão, a função de perda é normalmente o erro quadrático médio (MSE), sendo definido pela expressão (7):

$$L = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

onde y_i são os valores reais e \hat{y}_i são os valores previstos. O ganho G de uma divisão no nó j determina a melhor variável e o melhor limite de divisão, podendo ser calculado na fórmula (8):

$$G = \frac{1}{2} \left(\frac{(\sum_{i \in D_L} g_i)^2}{|D_L| + \lambda} + \frac{(\sum_{i \in D_R} g_i)^2}{|D_R| + \lambda} - \frac{(\sum_{i \in D} g_i)^2}{|D| + \lambda} \right) - \gamma \quad (8)$$

onde $g_i = \frac{\partial L}{\partial \hat{y}_i}$ é o gradiente da função de perda em relação a previsão \hat{y}_i , λ é o parâmetro de regularização controlando a complexidade e γ é a redução mínima de perdas necessárias

para fazer uma divisão. Cada somatório representa, respectivamente, a soma dos gradientes para as divisões das observações D_L , D_R e D .

O ganho quantifica o impacto de dividir os dados em subconjuntos na redução das perdas em comparação com a perda original. Um ganho mais alto indica uma divisão mais eficaz, o que resulta em melhor desempenho preditivo. Ao avaliar sistematicamente o ganho para todas as possíveis divisões, o XGBoost constrói árvores de decisão que otimizam a função de perda, ao mesmo tempo que controlam a complexidade do modelo por meio da regularização [8].

As técnicas de regularização introduzem penalidades à função de perda, punindo modelos excessivamente complexos e promovendo a generalização. O XGBoost utiliza duas formas principais de regularização, a regularização $L1$ (*Lasso*) e regularização $L2$ (*Ridge*). Essas técnicas são aplicadas aos pesos das folhas das árvores de decisão, que são os valores previstos pelas folhas das árvores. O objetivo é limitar esses pesos a valores baixos, reduzindo a sensibilidade do modelo às variações nos dados de treinamento.

A regularização $L1$ adiciona uma penalidade proporcional ao valor absoluto dos pesos das folhas. Esta forma de regularização incentiva a dispersão no modelo, o que significa que leva alguns pesos a serem exatamente zero. Isso resulta em um modelo mais simples, onde apenas as variáveis mais importantes são retidas, realizando efetivamente a seleção de variáveis.

Por sua vez, a regularização $L2$ adiciona uma penalidade proporcional ao quadrado dos pesos das folhas. Ao contrário da regularização $L1$, $L2$ não promove a dispersão, mas reduz os pesos para zero. Essa forma de regularização suaviza o modelo, distribuindo o impacto das variáveis individuais de maneira mais uniforme pelas árvores de decisão. A regularização $L2$ é particularmente eficaz em cenários onde todas as características têm alguma relevância, pois não elimina nenhuma, mas reduz a sua influência, evitando assim que qualquer característica única tenha dominância no modelo [8].

Em aplicações práticas, os parâmetros de regularização no XGBoost, denotados como α para regularização $L1$ e λ para regularização $L2$, são ajustados para alcançar o melhor desempenho. Esses parâmetros normalmente são definidos por meio de validação cruzada, onde diferentes valores são testados e selecionados aqueles que minimizam o erro

de validação. Esse processo de ajuste garante que a força da regularização seja ideal para o conjunto de dados e a tarefa específicos.

O desempenho do XGBoost está condicionado ao ajuste de seus hiperparâmetros, que controlam diversos aspectos do comportamento do modelo. A otimização de hiperparâmetros envolve a seleção do melhor conjunto de parâmetros que maximizam a precisão preditiva do modelo, minimizando o *overfitting* e o custo computacional.

Os hiperparâmetros podem ser categorizados em três tipos principais: parâmetros específicos da árvore, de reforço e de regularização. Os hiperparâmetros específicos da árvore determinam a estrutura e a complexidade das árvores de decisão individuais. Isso inclui parâmetros como a profundidade máxima das árvores, que controla a profundidade de crescimento de cada árvore, e o peso mínimo do filho, que especifica a soma mínima dos pesos de instância necessários em um filho. Os parâmetros de reforço influenciam o processo geral de reforço, incluindo a taxa de aprendizagem, que determina o tamanho do passo em cada iteração, e o número de estimadores, que define o número total de árvores no modelo. Os parâmetros de regularização, são os termos de regularização $L1$ e $L2$, abordados anteriormente.

O objetivo da otimização de hiperparâmetros é minimizar a função objetivo, que compreende dois componentes: a perda de treinamento e o termo de regularização [8]. Matematicamente, isso é apresentado na expressão (9):

$$Obj(\theta) = L(\theta) + \Omega(\theta) \quad (9)$$

Na qual θ representa o espaço dos hiperparâmetros, $L(\theta)$ é perda de treinamento (quantificando o erro de previsão do modelo nos dados de treinamento) e $\Omega(\theta)$ denota o termo de regularização (penalizando a complexidade do modelo para mitigar o *overfitting*).

Uma escolha comum de L é o erro quadrático médio, dado pela fórmula (10):

$$L(\theta) = \sum_i (y_i - \hat{y}_i)^2 \quad (10)$$

Outra função de perda normalmente usada é a perda logística, expressa por (11):

$$L(\theta) = \sum_i (y_i \ln(1 + e^{\hat{y}_i}) + (1 + y_i) \ln(1 + e^{\hat{y}_t})) \quad (11)$$

O processo de otimização de hiperparâmetros pode ser abordado através de diversas estratégias, cada uma com suas metodologias e implicações computacionais, como: *grid search*, *random search*, otimização *bayesiana* e otimização baseada em gradiente [79] e [80].

- *Grid Search*: a pesquisa em grade é a abordagem mais direta para a otimização de hiperparâmetros, envolvendo uma busca exaustiva através de um subconjunto específico do espaço de hiperparâmetros. O usuário define uma grade de valores de hiperparâmetros e avalia cada combinação possível desses parâmetros. Para um conjunto de n hiperparâmetros, a pesquisa em grade avalia o desempenho do modelo para cada ponto em uma grade n -dimensional no espaço de hiperparâmetros, selecionando a combinação que produz o melhor desempenho de acordo com uma métrica predeterminada. A formulação explora sistematicamente o espaço $\Theta = \Theta_1 \times \Theta_2 \times \dots \times \Theta_i$, na qual Θ_i é o conjunto de valor para o i -ésimo hiperparâmetro, sendo o número de combinações definido por $Y = \prod_{n=1}^N \Theta_n$.
- *Random Search*: aborda algumas das ineficiências da pesquisa em grade por amostragem aleatória do espaço de hiperparâmetros, reconhecendo que nem todos os hiperparâmetros são igualmente importantes para o desempenho de um modelo. Em vez de pesquisar através de uma grade pré-definida, a busca aleatória coleta combinações de hiperparâmetros a partir de uma distribuição de probabilidade (uniforme ou log-uniforme) sobre o espaço do hiperparâmetro. Se $f(\Theta)$ representa o desempenho do modelo com hiperparâmetro Θ , a busca aleatória visa maximizar $f(\Theta)$ sobre $\Theta \in \Theta_1 \times \Theta_2 \times \dots \times \Theta_i$, onde os valores para Θ são sorteados aleatoriamente de acordo com uma distribuição específica.
- Otimização *Bayesiana*: a otimização *bayesiana* é uma abordagem baseada em modelos probabilísticos para otimização global, construindo um modelo probabilístico da função objetivo para selecionar os hiperparâmetros mais promissores a serem avaliados na função objetivo verdadeira. Ocorre a introdução de um modelo substituto (na maioria das vezes um processo

gaussiano) para aproximar a função objetivo e uma função de aquisição para decidir onde amostrar em seguida. Essa abordagem equilibra a exploração do espaço hiper paramétrico com a exploração de regiões boas conhecidas. Dado um modelo substituto $p(f|\Theta, D)$, na qual D são os dados observados, a otimização bayesiana atualiza iterativamente o modelo com base no $f(\Theta)$ observado e seleciona novos Θ para avaliar otimizando uma função de aquisição $a(\Theta|D)$.

- Otimização Baseada em Gradiente: estratégias de otimização baseadas em gradiente ajustam os hiperparâmetros seguindo o gradiente da função objetivo em relação aos hiperparâmetros, particularmente adequado para hiperparâmetros contínuos. São utilizadas as derivadas da função objetivo em relação aos hiperparâmetros para atualizá-las iterativamente na direção esperada para diminuir o objetivo. Dada a função objetivo $f(\Theta)$, a regra de atualização para um hiperparâmetro $\Theta_i \in \Theta$ na etapa t pode ser expressa como $\Theta_i^{(t+1)} = \Theta_i^{(t)} - \alpha \frac{\partial f}{\partial e_i^{(t)}}$, onde α é a taxa de aprendizado.

Os hiperparâmetros que serão considerados na modelagem do algoritmo são definidos na Tabela 5, tais como *n_estimators*, *max_depth*, *learning_rate*, *colsample_bytree*, *subsample*, *min_child_weight*, *gamma*, *reg_lambda*, *reg_alpha*, *n_jobs*, *lags_grid*, *grid_search_forecaster*, *steps*, *refit*, *random_state* e *lags*.

Tabela 5 - Hiperparâmetros e Definições - *eXtreme Gradient Boosting*

Hiperparâmetros	Definição
<i>n_estimators</i>	Especifica o número de árvores a serem usadas no modelo, impactando sua complexidade e potencial de <i>overfitting</i> .
<i>max_depth</i>	Controla a profundidade máxima permitida para cada árvore, servindo como restrição para regular a capacidade de aprendizado do modelo e evitar <i>overfitting</i> .
<i>learning_rate</i>	Reduz o impacto de cada árvore individual e controla a taxa de aprendizado, equilibrando entre velocidade e precisão.

<i>colsample_bytree</i>	Determina a fração de funcionalidades a serem utilizadas para cada árvore, influenciando a diversidade dos modelos e auxiliando na prevenção do <i>overfitting</i> do modelo.
<i>subsample</i>	Define a proporção do conjunto de dados de treinamento a ser amostrados aleatoriamente para cada árvore, promovendo velocidade de treinamento, qualidade do modelo e redução da variância.
<i>min_child_weight</i>	Define a soma mínima dos pesos de todas as observações necessárias em um nó filho para continuar o particionamento, usada para controlar o <i>overfitting</i> aumentando o viés.
<i>gamma</i>	Especifica a redução mínima de perdas necessária para fazer uma divisão adicional em um nó folha, controlando o crescimento da árvore e melhorando a simplicidade.
<i>reg_lambda</i>	Aplica a regularização L2 nos pesos das árvores, auxiliando no tratamento da multicolinearidade e melhorando o desempenho, reduzindo o <i>overfitting</i> .
<i>reg_alpha</i>	Implementa regularização L1 para adicionar dispersão, podendo ser benéfico em modelos com alta dimensionalidade para reduzir a complexidade.
<i>n_jobs</i>	Determina a quantidade de <i>threads</i> paralelas utilizadas pelo algoritmo, otimizando recursos computacionais para agilizar os processos de treinamento.
<i>lags_grid</i>	Configuração de matriz que especifica as defasagens de pontos de tempo passados a serem usados como preditores, a fim de aprimorar a capacidade do modelo de capturar a dinâmica temporal na previsão de séries temporais.

<i>grid_search_forecaster</i>	Ferramenta projetada para pesquisar automaticamente diversas combinações de hiperparâmetros para encontrar a configuração de modelo ideal, maximizando o desempenho.
<i>steps</i>	Quantidade de passos à frente que o modelo deve prever, afetando o escopo e a granularidade da previsão.
<i>refit</i>	Permite que o modelo seja reajustado com novos dados, atualizando o aprendizado à medida que novos dados se tornam disponíveis.
<i>random_state</i>	Semente do gerador aleatório.
<i>lags</i>	Etapas de tempo anteriores de uma variável como características de entrada para modelar dependências temporais.

2.2 Contexto Climático e Operação de Turbinas Eólicas

A complementaridade dos modelos de aprendizado de máquina deve ser fundamentalmente atrelada às variáveis climáticas e ao contexto meteorológico, pois a eficácia desses modelos depende da correta integração dos fenômenos atmosféricos com os padrões de geração de energia eólica. A compreensão dos conceitos relacionados à dinâmica do vento é essencial para capturar a interação complexa entre as condições meteorológicas, como variações de velocidade, direção do vento e outros fatores climáticos, e a resposta das turbinas eólicas. A volatilidade e a previsibilidade dos ventos, elementos fundamentais para o desempenho das turbinas eólicas, são abordados teoricamente, explorando as características de projeto das turbinas e sua eficiência energética.

No caso do Complexo Eólico Pindaí, utilizado como estudo de caso para a aplicação das metodologias de aprendizado de máquina, esses aspectos se tornam ainda mais relevantes. Localizado no município de Pindaí, na Bahia, o complexo está sujeito a diversos

sistemas atmosféricos que afetam tanto os regimes eólicos sinóticos quanto os de mesoescala, impactando diretamente a eficiência operacional. A análise da variabilidade diurna, sazonal e interanual do vento, com destaque para fenômenos como El Niño e La Niña, é importante para entender as flutuações nos padrões de vento e sua influência na geração de energia. Compreender essa variabilidade atmosférica permite otimizar a operação e gestão do complexo eólico, e orientar a escolha das técnicas e parâmetros nos modelos de previsão de geração, garantindo um aproveitamento mais eficiente da energia eólica na região.

2.2.1 Natureza do Vento

O vento é um componente do sistema atmosférico da Terra influenciado pela dinâmica das variações de pressão atmosférica, diferenciais de temperatura e rotação do planeta. Esses fatores interagem de maneiras complexas produzindo inúmeros padrões de vento, desde brisas suaves até rajadas de alta intensidade como tornados ou furacões. Os fundamentos teóricos do vento estão diretamente ligados aos princípios da física e ciência atmosférica, baseando-se nas leis da termodinâmica, da dinâmica dos fluidos e do efeito de Coriolis para explicar o movimento do ar de áreas de alta pressão para áreas de baixa pressão [81]. Chama-se vento à componente horizontal ($\bar{V}_z = u\bar{i} + v\bar{j}$) do vetor velocidade ($\bar{V} = u\bar{i} + v\bar{j} + w\bar{k}$) do ar. A caracterização do vento em qualquer ponto (ϕ, λ, z) da atmosfera requer dois parâmetros: sua direção e sua velocidade (módulo). Ambas são grandezas instantâneas e pontuais, pois o escoamento do ar depende das condições atmosféricas (que variam no espaço e no tempo). Nas proximidades da superfície-atmosfera o vento é altamente influenciado pelas características geométricas e pelo estado de aquecimento da própria superfície subjacente [82].

No centro da geração eólica está o conceito de gradiente de pressão, princípio abordado nos trabalhos fundamentais da meteorologia, que detalham o movimento do ar em direção a regiões de menor pressão em resposta às variações espaciais da pressão atmosférica. Esses gradientes são originados através do aquecimento desigual da superfície da Terra, influenciado por fatores como o ângulo de insolação solar e o albedo da superfície

[83]. A temperatura possui um papel importante na determinação dos padrões de vento, atuando como a força motriz por trás da criação de gradientes de pressão. A disparidade no aquecimento entre o equador e os polos configura as células de circulação primária da Terra, levando ao estabelecimento de grandes cinturões de vento como os ventos alísios, ocidentais e polares. O trabalho de Lorenz sobre a circulação geral da atmosfera mostra como o aquecimento diferencial contribui para esses padrões de vento em grande escala, enquanto fenômenos locais como brisas marítimas e terrestres exemplificam a influência da temperatura em sistemas eólicos de menor escala [84]. As nuvens são indicadores e moduladores dos padrões de vento, com sua formação ligada à dinâmica dos sistemas de baixa pressão e do ar ascendente. A relação entre formação de nuvens, movimentos atmosféricos e sistemas de pressão é complexa, com nuvens afetando os padrões locais de vento através de mecanismos como o albedo e o efeito estufa, interações em profundidade que mostram o papel das nuvens no sistema climático terrestre e sua influência na dinâmica atmosférica [85].

A interação entre sistemas de alta e baixa pressão forma a base para grande parte da variabilidade nos padrões de vento observados, de maneira que os sistemas de alta pressão, associados ao ar descendente, normalmente trazem condições meteorológicas claras e estáveis, enquanto os sistemas de baixa pressão, caracterizados pela ascensão do ar, estão ligados à nebulosidade, precipitação e tempo chuvoso. Um exemplo é o efeito Coriolis, fenômeno que descreve como a rotação da Terra influencia o movimento dos objetos sobre sua superfície, como correntes de ar ou de água. Este efeito faz com que o caminho destes objetos pareça curvar-se para a direita no hemisfério norte e para a esquerda no hemisfério sul, sendo esta deflexão exercida sobre os objetos pela rotação da Terra em velocidades e latitudes diferentes [86].

2.2.1.1 Volatilidade e Previsibilidade

A volatilidade e a previsibilidade do vento são fatores que influenciam a geração de energia eólica devido à natureza da dinâmica atmosférica, caracterizada por interações não lineares em várias escalas, desde a circulação atmosférica global até as influências locais do

terreno. Os fundamentos teóricos da previsibilidade do vento consideram as contribuições sobre a teoria do caos [87], que aborda os limites da previsão do tempo de longo prazo devido à sensível dependência das condições iniciais da atmosfera. Este princípio de sistemas caóticos, apresenta implicações na previsão de padrões de vento, impactando a precisão e a confiabilidade da previsão da produção de energia eólica.

A variabilidade espaço-temporal do vento é influenciada por fenômenos atmosféricos de macroescala, a partir da interação entre sistemas de alta e baixa pressão e fatores de microescala, como topografia e rugosidade superficial do terreno. Os modelos de previsão numérica do tempo levam em consideração as equações de Navier-Stokes para simular fluxos atmosféricos, com capacidades avançadas de previsão de vento. Esses modelos, no entanto, enfrentam limitações relacionadas aos recursos computacionais e à resolução dos dados de entrada, levando a incertezas nas previsões de velocidade e direção do vento para aplicações no ramo da energia eólica [88]. A integração da energia nas redes elétricas destaca a importância da previsão da produção para que os operadores de rede possam assegurar a estabilidade do sistema energético e otimizar a diversificação das fontes de energia. Isso abrange a gestão do armazenamento de energia, estratégias de resposta à demanda e considerações de viabilidade econômica e operacional, decisões que estão ligadas às previsões de vento, influenciando a seleção do local e o *design* das turbinas [89].

2.2.1.2 Regimes Eólicos Sinóticos e de Mesoescala na Bahia

No centro dos regimes de ventos de escala sinótica da Bahia estão o Anticiclone Subtropical do Atlântico Sul (ASAS) e os ventos alísios. O Anticiclone Subtropical do Atlântico é um sistema de alta pressão localizado sobre o Oceano Atlântico que exerce uma influência significativa sobre o clima da Bahia, promovendo condições estáveis e secas em toda a região. Esse anticiclone facilita os ventos alísios de nordeste, fundamentais para definir as características climáticas do litoral baiano, de maneira que esses ventos, originados do anticiclone, são consistentes e persistentes, caracterizados por sua natureza carregada de umidade ao atravessarem o Oceano Atlântico antes de atingirem a costa brasileira. A

modulação dessas forças sinóticas pela dinâmica estocástica de massas de ar polares introduz variabilidade nos padrões de vento, particularmente no sul da Bahia, levando a perturbações periódicas na influência anticiclônica [90].

Os regimes eólicos de mesoescala da Bahia são afetados pela variada topografia do estado, incluindo suas regiões costeiras, planaltos e cadeias montanhosas [90]. Estas características geográficas locais dão origem a fenômenos eólicos distintos, a saber:

1. Brisas do Mar e da Terra: as variações diurnas de temperatura entre a terra e o oceano geram brisas marítimas e terrestres locais, presentes em áreas costeiras. Durante o dia, a terra aquece mais rapidamente do que o oceano, levando a uma diferença de pressão que impulsiona a brisa marítima para o interior. Por outro lado, à noite, a inversão dos gradientes de temperatura facilita a brisa terrestre, que flui da terra para o mar. Essas brisas moderam as temperaturas ao longo da costa e alteram os padrões climáticos locais, formando precipitações.
2. Brisas do Vale da Montanha: semelhante às brisas do mar e da terra, as brisas do vale da montanha surgem do aquecimento e resfriamento diferenciados de áreas montanhosas e vales adjacentes. Durante o dia, o ar ao longo das encostas das montanhas aquece mais rapidamente do que o ar na mesma altitude sobre o vale, levando a ventos ascendentes. À noite, o processo se inverte, com ar mais frio e denso descendo para o vale, criando uma brisa do vale, fenômeno representativo na região das chapadas na Bahia, influenciando os padrões de vento locais.
3. Jatos noturnos: outra característica do regime de ventos de mesoescala da Bahia é o desenvolvimento de jatos noturnos de baixo nível, sendo caracterizados por correntes de ar em movimento rápido que se formam perto da superfície à noite, em regiões onde o terreno induz uma aceleração do vento. A formação de jatos noturnos está diretamente ligada à estabilização da atmosfera durante a noite, com jatos contribuindo para o aumento da velocidade do vento.

2.2.1.3 Variabilidade Interanual e Fenômenos ENSO

Como visto anteriormente, a variabilidade diurna do vento na Bahia é em grande parte impulsionada pelo aquecimento diferencial da superfície da Terra pelo sol. Durante o dia, a radiação solar aquece a terra mais rápido do que os corpos d'água, criando um gradiente de temperatura entre a terra e o oceano. Isto resulta no desenvolvimento da brisa marítima, à medida que o ar mais frio e denso sobre a água se move em direção à terra, substituindo o ar mais quente que sobe e se move para completar a circulação. Durante a noite, a inversão desse gradiente de temperatura facilita a brisa terrestre, onde o ar flui da superfície terrestre mais fria em direção ao oceano mais quente [82].

A variabilidade sazonal dos ventos na Bahia é influenciada por padrões de circulação atmosférica de maior escala, incluindo o movimento da Zona de Convergência Intertropical (ZCIT) e a presença do sistema de Alta Pressão do Atlântico Sul. À medida que a ZCIT se desloca para o norte e para o sul com as estações, ela altera os padrões de precipitação e as direções dos ventos em toda a região. Durante os meses de verão da Bahia, a ZCIT está posicionada mais ao sul, influenciando o aumento dos ventos alísios de nordeste, o que pode levar a condições mais secas em algumas partes do estado [91]. Por outro lado, nos meses de inverno, a ZCIT se move para o norte e a influência do sistema de Alta Pressão do Atlântico Sul torna-se mais pronunciada, potencialmente levando ao aumento da precipitação e alteração dos padrões de vento ao longo da costa. Essas mudanças sazonais são importantes para a agricultura, ditando os ciclos de plantio e colheita, e para a energia renovável, influenciando a disponibilidade de energia eólica.

No centro da variabilidade climática interanual estão fenômenos como o *El Niño-Southern Oscillation* (ENSO), a Oscilação Multidecadal do Atlântico (AMO) e a Oscilação Decadal do Pacífico (ODP). O fenômeno ENSO, englobando a fase quente (El Niño) e a fase fria (La Niña), destaca-se como a dinâmica climática que afeta as variações anuais da temperatura média global. A essência do ENSO não está nos valores absolutos de temperatura, mas nas anomalias de temperatura relativa no Oceano Pacífico tropical central, que desencadeiam alterações atmosféricas com efeitos de longo alcance em todo o globo. As

implicações dessas anomalias de temperatura são profundas, impactando os climas sazonais e contribuindo para perturbações ambientais e socioeconômicas em todo o mundo [92].

A definição operacional empregada pela Administração Atmosférica e Oceânica dos Estados Unidos (*National Oceanic and Atmospheric Administration* - NOAA) caracteriza as condições de El Niño como Temperaturas da Superfície do Mar (TSM) no Pacífico tropical central $0,5^{\circ}\text{C}$ mais quentes do que a média, e as condições de La Niña como temperaturas $0,5^{\circ}\text{C}$ mais frias do que a média. Essas condições devem persistir ao longo de cinco temporadas consecutivas para constituir um episódio ENSO completo.

Portanto, durante os eventos de El Niño, o aquecimento das águas do Oceano Pacífico leva a uma reconfiguração dos padrões de circulação atmosférica, muitas vezes resultando em chuvas reduzidas e temperaturas mais altas no nordeste do Brasil, incluindo a Bahia. O fenômeno tende a enaltecendo as condições de seca, afetando os recursos hídricos, a agricultura, aumentando a produção de energia eólica e reduzindo a geração hidrelétrica, devido à alteração dos regimes eólicos. Por outro lado, os eventos de La Niña, caracterizados por temperaturas mais baixas do que a média da superfície do mar no Pacífico, estão geralmente associados ao aumento das chuvas e temperaturas mais baixas no nordeste do Brasil, condições que reduzem o potencial de produção de energia eólica por influenciar os padrões de vento e tendendo a levar a condições agrícolas mais favoráveis pelo aumento da disponibilidade hídrica [93].

A Tabela 6 representa o período histórico de TSM desde o ano de 2000. A cor vermelha é caracterizada por períodos quentes ($TSM > 0,5^{\circ}\text{C}$) e a cor azul representa os períodos frios ($TSM < 0,5^{\circ}\text{C}$).

Tabela 6 - Anomalias Sazonais de Temperatura desde 2000. Fonte: [94]

Year	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
2000	-1.7	-1.4	-1.1	-0.8	-0.7	-0.6	-0.6	-0.5	-0.5	-0.6	-0.7	-0.7
2001	-0.7	-0.5	-0.4	-0.3	-0.3	-0.1	-0.1	-0.1	-0.2	-0.3	-0.3	-0.3
2002	-0.1	0.0	0.1	0.2	0.4	0.7	0.8	0.9	1.0	1.2	1.3	1.1
2003	0.9	0.6	0.4	0.0	-0.3	-0.2	0.1	0.2	0.3	0.3	0.4	0.4
2004	0.4	0.3	0.2	0.2	0.2	0.3	0.5	0.6	0.7	0.7	0.7	0.7
2005	0.6	0.6	0.4	0.4	0.3	0.1	-0.1	-0.1	-0.1	-0.3	-0.6	-0.8
2006	-0.9	-0.8	-0.6	-0.4	-0.1	0.0	0.1	0.3	0.5	0.8	0.9	0.9
2007	0.7	0.2	-0.1	-0.3	-0.4	-0.5	-0.6	-0.8	-1.1	-1.3	-1.5	-1.6
2008	-1.6	-1.5	-1.3	-1.0	-0.8	-0.6	-0.4	-0.2	-0.2	-0.4	-0.6	-0.7
2009	-0.8	-0.8	-0.6	-0.3	0.0	0.3	0.5	0.6	0.7	1.0	1.4	1.6
Year	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
2010	1.5	1.2	0.8	0.4	-0.2	-0.7	-1.0	-1.3	-1.6	-1.6	-1.6	-1.6
2011	-1.4	-1.2	-0.9	-0.7	-0.6	-0.4	-0.5	-0.6	-0.8	-1.0	-1.1	-1.0
2012	-0.9	-0.7	-0.6	-0.5	-0.3	0.0	0.2	0.4	0.4	0.3	0.1	-0.2
2013	-0.4	-0.4	-0.3	-0.3	-0.4	-0.4	-0.4	-0.3	-0.3	-0.2	-0.2	-0.3
2014	-0.4	-0.5	-0.3	0.0	0.2	0.2	0.0	0.1	0.2	0.5	0.6	0.7
2015	0.5	0.5	0.5	0.7	0.9	1.2	1.5	1.9	2.2	2.4	2.6	2.6
2016	2.5	2.1	1.6	0.9	0.4	-0.1	-0.4	-0.5	-0.6	-0.7	-0.7	-0.6
2017	-0.3	-0.2	0.1	0.2	0.3	0.3	0.1	-0.1	-0.4	-0.7	-0.8	-1.0
2018	-0.9	-0.9	-0.7	-0.5	-0.2	0.0	0.1	0.2	0.5	0.8	0.9	0.8
2019	0.7	0.7	0.7	0.7	0.5	0.5	0.3	0.1	0.2	0.3	0.5	0.5
Year	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
2020	0.5	0.5	0.4	0.2	-0.1	-0.3	-0.4	-0.6	-0.9	-1.2	-1.3	-1.2
2021	-1.0	-0.9	-0.8	-0.7	-0.5	-0.4	-0.4	-0.5	-0.7	-0.8	-1.0	-1.0
2022	-1.0	-0.9	-1.0	-1.1	-1.0	-0.9	-0.8	-0.9	-1.0	-1.0	-0.9	-0.8
2023	-0.7	-0.4	-0.1	0.2	0.5	0.8	1.1	1.3	1.6	1.8	1.9	2.0

2.2.2 Turbinas Eólicas

A transição das estruturas de moinhos de vento, que historicamente convertiam a energia do vento em força mecânica como bombeamento de água e moagem de grãos, para as turbinas eólicas modernas, transformando energia em eletricidade, marca a evolução na capacidade humana de aproveitar o vento. Esta evolução permitiu que a energia gerada por turbinas eólicas alimente uma ampla gama de sistemas elétricos, desde instalações residenciais isoladas até grandes redes de serviços públicos e residências.

O Complexo Eólico de Pindaí é composto por quatro parques eólicos que totalizam 34 aerogeradores Wobben Windpower, cada um com 98 metros de altura e 92 metros de diâmetro [95], ilustrado na Figura 1.



Figura 1 - Aerogerador Wobben Windpower no Complexo Pindaí. Fonte: [95]

O funcionamento das turbinas eólicas requer um entendimento dos princípios aerodinâmicos que influenciam sua operação. As turbinas utilizam a força de elevação aerodinâmica para gerar torque rotacional, convertendo energia mecânica em elétrica através de um gerador. A conversão direta, contudo, enfrenta a variabilidade inerente do recurso eólico: o vento não pode ser armazenado e, portanto, a energia gerada por uma turbina eólica flutua de acordo com a disponibilidade imediata do vento. Sistemas elétricos conectados a turbinas eólicas devem adaptar-se a essa flutuabilidade, uma tarefa que varia em complexidade de acordo com o tamanho e a natureza da rede à qual a turbina está conectada [96].

O *design* das lâminas das turbinas utiliza análises aerodinâmicas para otimizar sua eficiência. As lâminas possuem seções transversais de aerofólio que consistem em diferentes tamanhos e formas desde a raiz até a ponta, ou seja, a tecnologia de aerofólio simples faz com que a lâmina da turbina eólica gire, produzindo uma força de sustentação quando um

fluido se move sob o aerofólio [97], e desta forma, a turbina eólica atinge a rotação básica que permite o seu funcionamento, conforme ilustrado na Figura 2.

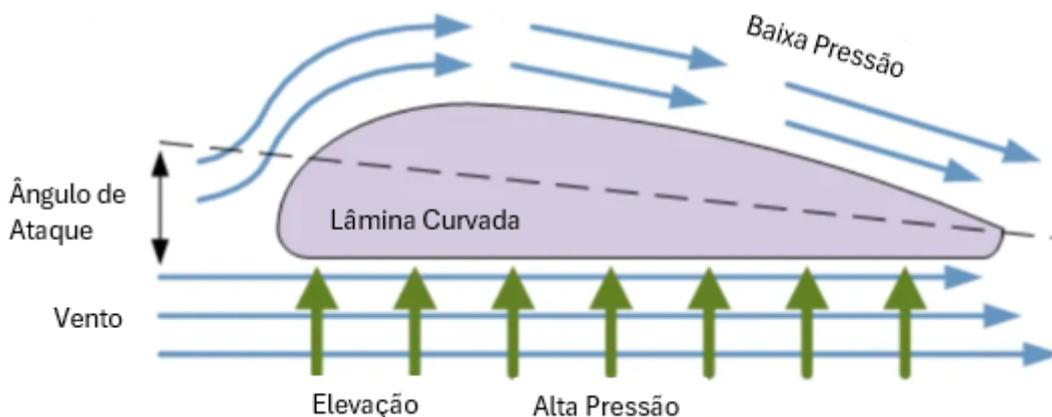


Figura 2 - Força de Sustentação da Lâmina. Fonte: Adaptado de [98]

A lâmina da turbina eólica interage com o vento de maneira relativa, como mostrado na Figura 3. Para capturar a velocidade relativa do vento, a lâmina móvel é posicionada em um ângulo de alinhamento com essa velocidade e, à medida que a lâmina se move em direção à ponta, a velocidade relativa do vento se inclina ainda mais em direção a essa ponta, resultando em uma torção contínua ao longo da extensão da lâmina, desde a raiz até a ponta. No entanto, essa rotação não pode ser diretamente transmitida a um gerador devido à baixa velocidade de rotação por minuto (RPM) das pás da turbina eólica, que é mantida para evitar problemas de ruído e resistência mecânica. Dado esse limite de velocidade, não é possível gerar eletricidade em uma frequência significativa diretamente do gerador. Portanto, antes de conectar ao gerador, a velocidade precisa ser aumentada usando uma caixa multiplicadora, empregando um sistema de engrenagens planetárias para elevar a velocidade a um nível adequado [99].



Figura 3 - Conceito de Velocidade Relativa

Na nacele, existe um freio destinado a interromper a rotação da pá durante condições de rajadas de ventos, sendo essencial para garantir a segurança e a estabilidade da turbina. A eletricidade gerada é então conduzida através dos cabos em direção à base, onde um transformador *step-up* está localizado para elevar a tensão. Para otimizar a extração de energia, as lâminas devem ser direcionadas para enfrentar o vento da maneira mais eficiente possível. Considerando a natureza variável do vento, um sistema de controle é instalado no topo da nacele, monitorando constantemente a velocidade e a direção do vento e enviando essas informações para um controlador eletrônico. O sinal é emitido para o mecanismo de orientação (guinada), que ajusta a posição da nacele para alinhá-la corretamente e corrigindo o erro conforme a direção atual do vento. Além disso, o ângulo de velocidade relativa do vento também sofre alteração, acionando o mecanismo de inclinação da lâmina, garantindo o alinhamento adequado com referência à velocidade relativa do vento, estando sempre no ângulo ideal de ataque com fluxo relativo do vento [100].

2.2.2.1 Energia e Eficiência das Turbinas Eólicas

A conversão da energia eólica em energia elétrica baseia-se na energia cinética característica à movimentação de massas de ar. A equação fundamental que governa esta conversão é descrita na expressão (12):

$$E_c = \frac{1}{2}mv^2 \quad (12)$$

onde E_c representa a energia cinética do vento, m denota a massa do ar em movimento e v é a velocidade do vento. Esta equação retrata a proporcionalidade direta da energia eólica tanto da massa do ar em movimento quanto do quadrado de sua velocidade, destacando o impacto quadrático da velocidade do vento sobre o potencial energético [101].

A análise do fluxo de massa de ar através do rotor de uma turbina eólica horizontal fornece informações sobre a quantificação do potencial de energia eólica. Quando o vento se move perpendicularmente ao rotor da turbina, ele pode ser conceituado como um cilindro de ar uniforme com uma área transversal A e uma velocidade v , como na Figura 4. O fluxo de massa m , através do rotor é, portanto, um produto da densidade do ar (ρ), da área transversal do cilindro (A) e da velocidade do vento (v), expressa na equação (13):

$$m = \rho Av \quad [kg/s] \quad (13)$$

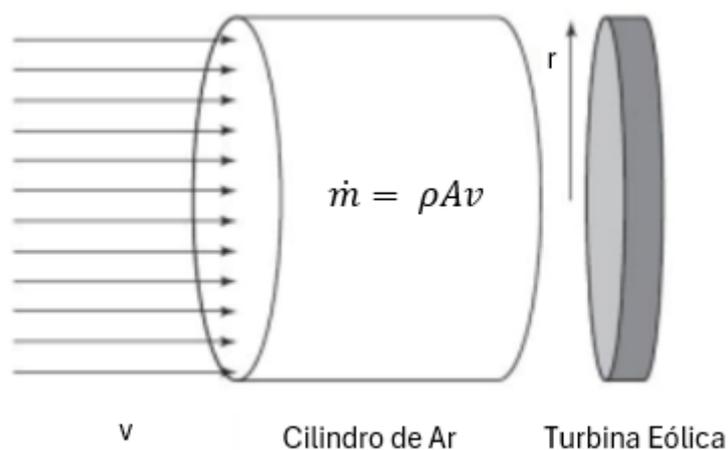


Figura 4 - Velocidade Perpendicular ao Cilindro de Ar Uniforme. Fonte: Adaptado de [101]

A área transversal A corresponde diretamente à área varrida do rotor da turbina eólica, que determina a capacidade de captura de energia da turbina. A área varrida é definida pelo círculo descrito pelas pás rotativas e é proporcional ao quadrado do raio do rotor, o que enfatiza a importância do comprimento da lâmina na eficiência da extração de energia eólica [101].

Substituindo a equação (2) na equação (1) é obtida a expressão (14):

$$E = \rho_v = \frac{1}{2}mv^2 = \frac{1}{2}\rho Av^3 \quad [W] \quad (14)$$

A relação entre o raio do rotor da turbina eólica e a potência que pode ser extraída do vento é definida na equação (15), onde ρ_v denota a potência disponível no vento, ρ é a densidade do ar, r representa o raio do rotor e v é a velocidade do vento. Esta equação é derivada dos princípios da dinâmica dos fluidos e da conservação de energia, abordando a dependência cúbica da potência na velocidade do vento e a dependência quadrática do raio do rotor, ilustrando o impacto do tamanho do rotor na extração de energia eólica [101].

$$\rho_v = \frac{1}{2}\rho\pi r^2 v^3 \quad [W] \quad (15)$$

A conservação da massa, dada a hipótese de estado estacionário e considerando o fluido (no caso, o ar) como incompressível, é representada pela equação (16).

$$m = \rho A_2 v_2 \quad (16)$$

Esta equação indica que a vazão mássica (m) através da turbina é constante em diferentes seções do volume de controle, refletindo o princípio de que a massa é conservada em um sistema fechado. À medida que o ar passa pela turbina, sua velocidade diminui ($v_0 > v_r > v_2$), enquanto a área de secção transversal pela qual flui aumenta ($A_0 < A_r < A_2$), ou seja, a dinâmica é crítica para as turbinas eólicas, pois está subjacente ao mecanismo pelo qual a energia cinética é extraída do vento [102].

A potência capaz de ser extraída pelo rotor (P) é a diferença entre a potência do vento a montante (P_0) e a jusante (P_2), apresentada na equação (17):

$$P = P_0 - P_2 = \frac{1}{2}m(v_0^2 - v_2^2) = \frac{1}{2}\rho(A_0 v_0^2 - A_2 v_2^2) \quad (17)$$

A conservação do momento, aplicando a segunda lei de Newton, determina que a força exercida pelo vento sobre a turbina é $F = mv_0 - mv_2 = m(v_0 - v_2)$. Essa força,

multiplicada pela velocidade relativa no rotor (v_r), resulta em outra expressão (18) para potência [102]:

$$P = FV_r = m(v_0 - v_2)v_r \quad (18)$$

O limite de Betz, introduzido pelo físico alemão Albert Betz em 1919, é um princípio na engenharia de turbinas eólicas que estabelece a eficiência máxima teórica na conversão da energia cinética do vento em energia mecânica. Este conceito é derivado da análise das equações de potência do vento e da otimização das condições que maximizam a extração de energia, considerando a conservação da massa e da energia no fluxo de ar através da turbina. Betz demonstrou que, para maximizar a potência extraída do vento, a velocidade do ar ao passar pelo rotor da turbina (v_r) deve ser aproximadamente dois terços da velocidade do vento antes de atingir a turbina (v_0). Essa relação otimiza a extração de energia, balanceando a desaceleração do vento com a quantidade de energia capturada.

O resultado dessa análise é que a eficiência máxima teórica, ou coeficiente de potência máxima ($C_{pmáx}$), que uma turbina eólica pode alcançar é de 59,3% [102] e [103], conforme a expressão (19).

$$\frac{\text{Máxima Potência Extraída}}{\text{Potência Disponível}} = \frac{\rho}{\frac{1}{2}\rho A_r v_0^3} = \frac{\frac{8}{27}}{\frac{1}{2}} = \frac{16}{27} = 0,593 = C_{pmáx} \quad (19)$$

2.2.2.2 Projetos e Características das Turbinas Eólicas

As primeiras turbinas exploraram alternativas de projeto diferentes, incluindo turbinas de eixo vertical (*Vertical-Axis Wind Turbine* - VAWTs). À medida que a indústria evoluiu, a maioria dos conceitos de projeto padronizou turbinas eólicas de eixo horizontal (*Horizontal-Axis Wind Turbine* - HAWTs) com uma, duas ou três pás. Nas turbinas de eixo vertical, as forças aerodinâmicas de sustentação e arrasto são os dois princípios que classificam esse projeto, de maneira que o eixo de rotação está configurado perpendicularmente ao plano do solo, também denominado como turbinas de fluxo cruzado,

podendo operar mesmo quando rotacionado em 90 graus desde que o fluxo de vento seja perpendicular ao eixo do rotor [104] e [105].

O tipo de VAWT baseado em arrasto é chamado de tipo Savonius, em homenagem ao inventor finlandês Sigurd Johannes Savonius, que aprimorou as turbinas de arrasto no início do século 19 [106]. Este tipo de turbina é uma das mais simples, sendo um dispositivo de rotor movido por arrasto composto por duas ou três pás. Outro tipo de VAWT é a turbina Darrieus, concebida e patenteada pelo engenheiro aeronáutico francês Georges Jean Marie Darrieus em outubro de 1926 [107]. Esta turbina consiste em duas ou três pás curvas com perfil de aerofólio na seção transversal, montadas em um eixo rotativo vertical, sendo esse *design* das pás curvas com objetivo de mitigar as tensões de flexão devido às forças centrífugas. A turbina Darrieus é considerada mais eficiente que a Savonius e requer menos material de construção, no entanto, suas principais desvantagens incluem forças oscilantes que causam fadiga, necessidade de proteção contra condições extremas de vento e dificuldades de partida automática [104].

O *design* predominante das turbinas eólicas é o de eixo horizontal, onde o eixo de rotação está alinhado paralelamente ao solo, como é observado nos aerogeradores do complexo de Pindaí, compreendendo três componentes primordiais: uma torre, uma nacela e um rotor. A torre desempenha a função de suporte estrutural, elevando o rotor à altura necessária, enquanto a nacela abriga o gerador, a caixa de engrenagens e outros elementos mecânicos essenciais para sustentar o rotor. O cubo da turbina estabelece a ligação mecânica entre as pás e o eixo de transmissão, conectando o cubo à nacela, e conjuntamente com as pás, formando o rotor da turbina. Diversas abordagens de projeto foram investigadas para a torre da turbina, sendo as alternativas principais construídas por uma treliça ou uma torre de tubo. Embora as primeiras turbinas tenham adotado torres de treliça, as torres tubulares de aço são atualmente mais comuns, construídas pela laminação de chapas de aço planas para alcançar o diâmetro necessário e unindo-as por soldagem. Alguns projetos empregam uma torre relativamente fina, reforçada com cabos para conferir rigidez adicional e resistência lateral, porém isso é raro em turbinas de grande porte [108].

As turbinas HAWT requerem um mecanismo para alinhar adequadamente o rotor em relação ao vento, a partir de dois métodos principais de orientação (passivo e ativo), visando maximizar a produção de energia. Uma orientação passiva depende da força gerada pelo

vento para girar a nacelle, comumente alcançada por meio de uma cauda localizada atrás do rotor, que captura a força do vento. Em contrapartida, uma orientação ativa depende de um mecanismo de controle para girar a nacelle mecanicamente. A maioria das turbinas modernas emprega a orientação ativa para simplificar o projeto, reduzindo os graus de liberdade. As turbinas que orientam o rotor na frente da nacelle são chamadas de *upwind* e são mais comuns em turbinas de grande porte [108].

Grande parte das turbinas é projetada para gerar sustentação, utilizando pás com seções transversais em formato de aerofólio, diferentemente das primeiras turbinas, que dependiam do arrasto para acionar o rotor [109]. As turbinas de sustentação podem extrair uma porção maior da energia cinética do vento, e como medida de segurança, essas turbinas geralmente são equipadas com freios aerodinâmicos nas pontas das pás. Além disso, todas as turbinas de grande porte contam com um sistema de freio mecânico para conferir resistência à rotação do rotor. As características fundamentais de desempenho dos aerogeradores HAWTs são [110]:

1. Coeficiente de Potência e Eficiência: o coeficiente de potência representa a eficácia com que a energia eólica é convertida em energia elétrica. Este coeficiente é influenciado por diversos fatores, como a velocidade do vento, o *design* das pás e os sistemas de controle da turbina. Além disso, a eficiência global da turbina é um parâmetro crítico, refletindo a capacidade de aproveitamento da energia eólica.
2. Desempenho Aerodinâmico: o desempenho aerodinâmico dos HAWTs é determinado pelo *design* das pás e das características do rotor. A forma e torção das pás influenciam as forças de sustentação e arrasto, afetando diretamente a eficiência da turbina, utilizando como métricas a curva de potência e as velocidades de corte, para avaliar o desempenho aerodinâmico.
3. Desempenho Estrutural: a integridade estrutural das HAWTs garante uma operação segura e confiável ao longo de sua vida útil. A análise estrutural envolve a avaliação das cargas, tensões e deformações experimentadas pelos componentes da turbina, utilizando métodos como a análise por elementos finitos para garantir que os componentes sejam projetados para suportar as cargas esperadas.

4. Desempenho Operacional: o desempenho operacional engloba a disponibilidade da turbina para gerar energia, com métricas de confiabilidade que avaliam a probabilidade de falha e o tempo de inatividade. Estratégias de manutenção, como o monitoramento de condições, minimizam o tempo de inatividade e maximizam a produção de energia.
5. Avaliação de Impacto Ambiental: a avaliação do impacto ambiental considera fatores como emissões sonoras, impacto visual, colisões com aves e morcegos, e efeitos ecológicos, aplicando medidas de mitigação como modificações aerodinâmicas das pás e estratégias para mitigar colisões com pássaros.

Cada componente de um aerogerador desempenha um papel na conversão de energia eólica em energia elétrica, garantindo a eficiência, confiabilidade e vida útil da turbina eólica [111]. Os componentes principais são: cubo do rotor, hub, pá do rotor, caixa de engrenagens, freio, gerador, nacele, controle de direção, torre e fundação, como mostra a Figura 5.

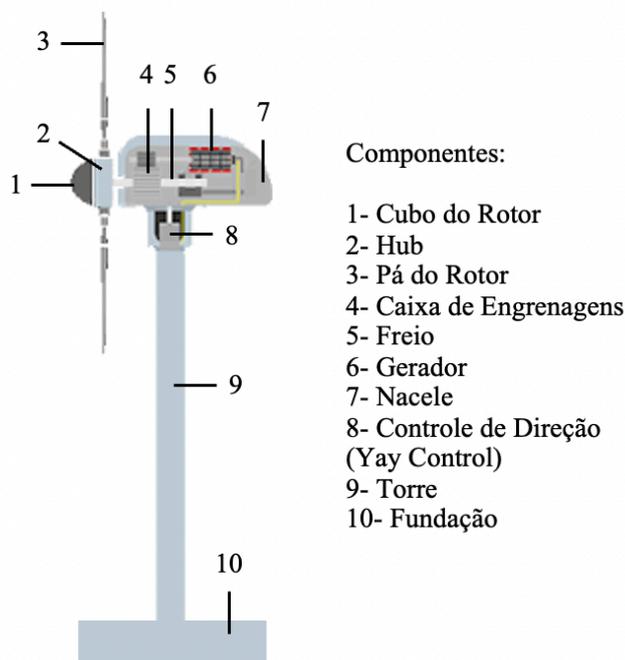


Figura 5 - Componentes Principais do Aerogerador HAWT

O rotor é um componente composto pelas pás e pelo cubo com o objetivo de equilibrar a eficiência aerodinâmica e a integridade estrutural do aerogerador, para maximizar a captura de energia. Na maioria dos projetos, as pás são construídas a partir de materiais compósitos, como poliéster reforçado com fibra de vidro ou fibra de carbono,

oferecendo um equilíbrio entre resistência e flexibilidade [111]. Seu projeto utiliza de princípios aerodinâmicos semelhantes às asas de avião, apresentando uma forma de aerofólio que facilita a geração de sustentação à medida que o vento passa sobre elas, sendo sua base matemática derivada do princípio de Bernoulli e da equação de elevação (20):

$$L = \frac{1}{2} \rho v^2 A C_L \quad (20)$$

onde L é a força de elevação, ρ é a densidade do ar, v é a velocidade do vento, A é a área da lâmina e C_L é o coeficiente de sustentação dependente da forma da lâmina e do ângulo de ataque.

O comprimento da lâmina afeta a capacidade de captura de energia da turbina, pois a energia gerada é proporcional ao quadrado do raio do rotor ($P \propto r^2$) e ao cubo da velocidade do vento ($P \propto v^3$). Portanto, o aumento do diâmetro do rotor aumenta a área varrida ($A = \pi r^2$), capturando assim, mais energia eólica. Por exemplo, uma turbina com diâmetro de rotor de 92 metros, como a Enercon E-92 (turbina do Complexo Pindaí, abordada na seção 4.1), tem uma área varrida de aproximadamente 6.648 m², o que influencia muito sua produção de energia [111].

O cubo, que conecta as pás ao eixo principal, deve ser robusto o suficiente para lidar com as cargas dinâmicas transmitidas pelas pás, mantendo um alinhamento adequado. Sua estrutura é construída de ferro fundido ou aço forjado e incorporado mecanismos de controle de passo que permitem que os ângulos da lâmina sejam ajustados, otimizando o ângulo de ataque. O *design* do cubo também integra o sistema de controle de passo, ajustando o ângulo das pás para otimizar a captura de energia e proteger a turbina durante rajadas de vento. Esse sistema inclui atuadores hidráulicos ou elétricos que mudam o ângulo de passo da lâmina com base nas entradas do sistema de controle da turbina, na qual a precisão do mecanismo de arremesso é fundamental para manter o ângulo de ataque ideal. Além disso, o cubo deve ser projetado para facilitar sua manutenção, pois sua falha pode levar a danos na turbina e perdas econômicas substanciais, sendo dessa forma, frequentemente submetido a testes de fadiga em grande escala, para validar sua durabilidade sob condições de carregamento cíclicas prolongadas [111].

A forma do cubo e sua integração com a nacele e as pás do rotor são otimizadas para reduzir as perdas aerodinâmicas e garantir um fluxo de ar suave sobre a estrutura da turbina.

Para isso, são feitas simulações de dinâmica dos fluidos computacional (*Computational Fluid Dynamics* - CFD) para analisar seu desempenho aerodinâmico, identificando possíveis efeitos aerodinâmicos adversos, como separação de fluxo e derramamento de vórtices, que poderiam reduzir a eficiência da turbina. O sistema de controle, o trem de força e o gerador são componentes presentes dentro da estrutura da nacele, que serve como unidade de alojamento, sendo montada no topo da torre e projetada para girar em torno de seu eixo vertical para enfrentar o vento, movimento chamado de guiada. Dentro da nacele, são integrados sistemas de amortecimento de vibração e gerenciamento térmico para proteger o equipamento sensível de cargas mecânicas e térmicas. O sistema de controle é um conjunto de *hardware* e *software* projetados para otimizar o desempenho da turbina e garantir sua segurança, estando presentes neste sistema sensores, atuadores e algoritmos de controle, que monitoram e ajustam as funções da turbina em tempo real. Matematicamente, os algoritmos de controle dependem de malhas de *feedback* e modelos preditivos para ajustar dinamicamente o ângulo de *pitch* a partir de dados de anemômetros [111].

O trem de força é responsável por transmitir a energia mecânica das pás do rotor para o gerador, incluindo um eixo principal, a caixa de engrenagens e os mecanismos de acoplamento. O eixo principal conecta o cubo do rotor à caixa de engrenagens (caixa multiplicadora), que aumenta a velocidade de rotação para um nível adequado para a geração de eletricidade. As caixas de engrenagens geralmente empregam sistemas de engrenagens planetárias, que fornecem alta eficiência de transmissão de torque em uma forma compacta, envolvendo o cálculo do torque T transmitido através das engrenagens, utilizando a relação $T = \frac{P}{\omega}$, onde P é a potência de entrada e ω é a velocidade angular [111].

A maioria das turbinas eólicas usa geradores de indução duplamente alimentados (*Doubly-Fed Induction Generator* - DFIG) ou geradores síncronos de ímã permanente (*Permanent Magnet Synchronous Generator* - PMSG). Os DFIG são utilizados por sua capacidade de operar em uma ampla faixa de velocidades e sua alta eficiência na conversão de energia eólica. Seu funcionamento é baseado na interação entre os enrolamentos do estator e do rotor, utilizando os princípios da indução eletromagnética [112]. A potência elétrica P é dada pela expressão (21):

$$P = \frac{1}{2} \rho A v^3 C_p \eta \quad (21)$$

onde ρ é a densidade do ar, A é a área varrida do rotor, v é a velocidade do vento, C_p é o coeficiente de potência e η é a eficiência do gerador e do trem de força. Os PMSGs, por outro lado, são utilizados em certos projetos pela sua alta eficiência e confiabilidade, especialmente em configurações de acionamento direto que eliminam a necessidade de uma caixa de câmbio.

O sistema de controle também incorpora mecanismos de segurança, como sistemas de frenagem para parar o rotor em emergências ou durante a manutenção. Esses freios podem ser aerodinâmicos, usando ajustes de passo da lâmina para travar o rotor, ou mecânicos, usando freios a disco.

A estrutura principal serve como suporte estrutural primário dentro da nacela, sendo projetado para suportar as cargas mecânicas geradas pela operação da turbina, como o peso dos componentes da nacela, as forças aerodinâmicas no rotor e as cargas dinâmicas resultantes das rajadas de vento e rotação da turbina. O sistema de guinada é responsável por girar a nacela para alinhar o rotor com a direção do vento, sendo composto por rolamentos de guinada, mecanismo de acionamento de guinada e os freios de guinada. Seu mecanismo de acionamento é formado por motores elétricos ou hidráulicos e controlado pelo sistema de controle da turbina, aplicando torque aos rolamentos de guinada para girar a nacela com base em dados de direção do vento em tempo real, fornecidos por anemômetros montados no topo da nacela. A modelagem matemática do acionamento de guinada envolve o cálculo do torque necessário T para superar as forças aerodinâmicas e o atrito, usando a relação $T = F_r r + F_a h$, onde F_r é a força radial, r é o raio do rolamento de guinada, F_a é a força axial e h é a altura do ponto de aplicação de força a partir do plano do rolamento. Os freios de guinada são implementados para travar a nacela no lugar uma vez que ela esteja corretamente alinhada com a direção do vento, evitando a rotação em caso de turbulência ou falha mecânica [111].

A torre e a fundação são componentes que garantem a estabilidade e funcionalidade de toda a estrutura, de maneira que a torre eleva a nacela e o rotor a alturas ideais para capturar a energia eólica com maior eficiência, enquanto a fundação ancora a turbina ao solo para proporcionar a estabilidade necessária e suportar cargas e condições ambientais. Matematicamente, a torre deve ser projetada para resistir a diversos tipos de cargas, tais como cargas axiais, de flexão e de torção. A carga axial P devido ao peso da nacela e do

rotor pode ser calculada usando a fórmula $P = mg$, onde m é a massa da nacele e do rotor, e g é a aceleração da gravidade. O momento de flexão M na base da torre devido às cargas de vento pode ser expresso como $M = Fh$, onde F é a força horizontal do vento atuando sobre o rotor e a nacele, e h é a altura da torre [111]. A força horizontal do vento F pode ser estimada através da expressão (22):

$$F = 0,5\rho AC_d V^2 \quad (22)$$

onde ρ é a densidade do ar, A é a área projetada do rotor, C_d é o coeficiente de arrasto e V é a velocidade do vento.

A resposta dinâmica da torre às vibrações induzidas pelo vento também é uma consideração a ser avaliada em projetos de turbinas eólicas, na qual a frequência natural f_n da torre deve ser suficientemente diferente da frequência de passagem da lâmina do rotor e outras frequências harmônicas para evitar ressonância, que pode causar vibrações intensas e fadiga estrutural. A frequência natural é aproximada usando a fórmula (23):

$$f_n = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad (23)$$

onde k é a rigidez da torre e m é a massa efetiva.

Este capítulo apresentou uma revisão da literatura relacionada aos principais conceitos do aprendizado de máquina, categorizando as técnicas em diferentes tipos de aprendizado e detalhando o horizonte de previsão aplicado a este estudo. Foram discutidos os testes de estacionariedade, essenciais para a análise de séries temporais, e a comparação de trabalhos semelhantes, utilizados como *benchmark* para avaliar os resultados obtidos. O capítulo trouxe um aprofundamento teórico sobre as metodologias de aprendizado de máquina aplicadas, como Regressão Linear Rígida, Florestas Aleatórias, Regressão por Vetores Suporte, Perceptron de Múltiplas Camadas e *eXtreme Gradient Boosting*. Também foi incluída a conceituação detalhada sobre a natureza do vento, abordando variação, previsibilidade, variabilidades interanuais, climatologia dos regimes e a variabilidade do vento, fatores fundamentais para a modelagem preditiva da geração eólica. Por fim, explorou-se o funcionamento das turbinas eólicas, com foco em suas características de projeto e eficiência.

Capítulo 3 - Materiais e Métodos

Nesse capítulo são apresentados os materiais e métodos utilizados no desenvolvimento deste estudo, detalhando cada uma das abordagens aplicadas. Inicialmente, é feita uma descrição do Complexo Eólico utilizado como objeto de estudo, com destaque para suas características operacionais e sua relevância no contexto da geração de energia. Em seguida, é abordado o conjunto de dados, com ênfase nas variáveis analisadas e na forma como foram processadas. A seguir, são explorados os modelos de reanálise de previsão meteorológica com alta resolução, que fornecem as bases climatológicas e meteorológicas para as previsões de geração de energia eólica. A seção dedicada ao machine learning introduz as técnicas aplicadas no estudo, sendo cada uma delas explicada em subseções específicas. São apresentados, os métodos de Regressão Linear Rígida, Florestas Aleatórias, Regressão por Vetores Suporte, Perceptron de Múltiplas Camadas e *eXtreme Gradient Boosting*, com uma descrição detalhada dos algoritmos, suas características e os motivos que justificam suas escolhas. Por fim, a seção de critérios de avaliação dos resultados discute as métricas utilizadas para avaliar o desempenho dos modelos preditivos, como o coeficiente de determinação (R^2), RMSE, MAE, e MAPE, fornecendo a base para a comparação dos resultados obtidos. Cada seção visa fornecer uma visão completa dos materiais e métodos aplicados, assegurando transparência e permitindo a replicação do estudo.

Portanto, este trabalho propõe o desenvolvimento de cinco previsores de geração eólica, desenvolvidos em linguagem de programação Python e base mensal. Os modelos utilizam uma combinação dos dados históricos de geração dos parques eólicos e de variáveis meteorológicas históricas e previstas. Estes dados meteorológicos são obtidos através de modelos de reanálise de alta resolução, acessados via a Open-Meteo API (*Application Programming Interface*) em formato JSON (*JavaScript Object Notation*) [30].

Para a regressão linear rígida, regressão por vetores de suporte e florestas aleatórias, utilizou-se a biblioteca *Scikit-Learn* [113], que oferece uma ampla gama de ferramentas para aprendizado de máquina e análise de dados, e fornece uma implementação otimizada para algoritmos *ensemble* no caso das florestas aleatórias. As redes neurais perceptron de múltiplas camadas foram desenvolvidas utilizando a biblioteca *TensorFlow* [114] em conjunto com o *Keras* [115], que é uma API de alto nível para construção e treinamento de redes neurais

profundas, facilitando a criação e o ajuste de modelos complexos. O *extreme Gradient Boosting* foi implementado utilizando a biblioteca *Scikit-Learn Skforecast* [116] e [117], especificamente projetada para este algoritmo e conhecida por sua performance e escalabilidade em tarefas de classificação e regressão. Além dessas bibliotecas específicas para cada modelo, foram utilizadas outras bibliotecas complementares para manipulação de dados, visualização e avaliação de desempenho, como *pandas*, *NumPy*, *matplotlib* e *seaborn* [118], [119], [120] e [121]. Adicionalmente, a biblioteca *Open-Meteo Requests* foi utilizada para obter previsões meteorológicas de alta resolução através da API Open-Meteo, permitindo a incorporação de variáveis meteorológicas previstas e históricas nos modelos de previsão [122].

O Complexo Eólico Pindaí, serve como um caso de estudo real para a aplicação destas técnicas de previsão, sendo apresentado na seção seguinte.

3.1 Descrição do Complexo Eólico e Conjunto de Dados

Localizado no município de Pindaí, no sul do estado da Bahia, o complexo eólico de Pindaí possui uma capacidade instalada de 79,9 MW, dividido em quatro parques distintos: Pindaí I com 16,45 MW, Pindaí II com 18,8 MW, Pindaí III com 21,15 MW e Pindaí IV com 25,5 MW.

Os aerogeradores do complexo foram construídos pela Wobben Windpower/ENERCON, especificamente do modelo Enercon E-92. As turbinas possuem uma potência nominal de 2.350,0 kW, sendo a curva de potência dos aerogeradores apresentada na Figura 6. Destaca-se a faixa entre 1 e 25 m/s em que o aerogerador tem sua partida e opera em transitório até a faixa de 50% da capacidade total de geração, faixa de velocidade onde o azimute (direção das pás/rotor) está sendo ajustado até que as pás fiquem perpendiculares ao vento incidente [123].

Cada turbina possui um diâmetro de 92,0 metros e uma altura de 98,0 metros, com uma área de varredura de 6.648,0 m² que maximiza a captação de vento, incluindo três pás, otimizando a conversão de energia cinética e a estabilidade mecânica das turbinas. O cubo

de cada turbina pode ser ajustado entre uma altura mínima de 85 metros e uma máxima de 138 metros, permitindo uma operação flexível que se adapta a diferentes condições de vento, enquanto a velocidade de ponta das pás atinge até 77 m/s, aumentando a eficiência da geração de energia, mantendo a integridade estrutural e o cumprimento das normas de segurança.

A segurança e a durabilidade são garantidas à medida que as operações são suspensas a uma velocidade de vento de 25,0 m/s, como mostra a Figura 7. Isto ocorre para a proteção do sistema contra danos potenciais causados por rajadas de vento, cujos *cut-in* e *cut-off* são de 2 e 25 m/s, respectivamente.

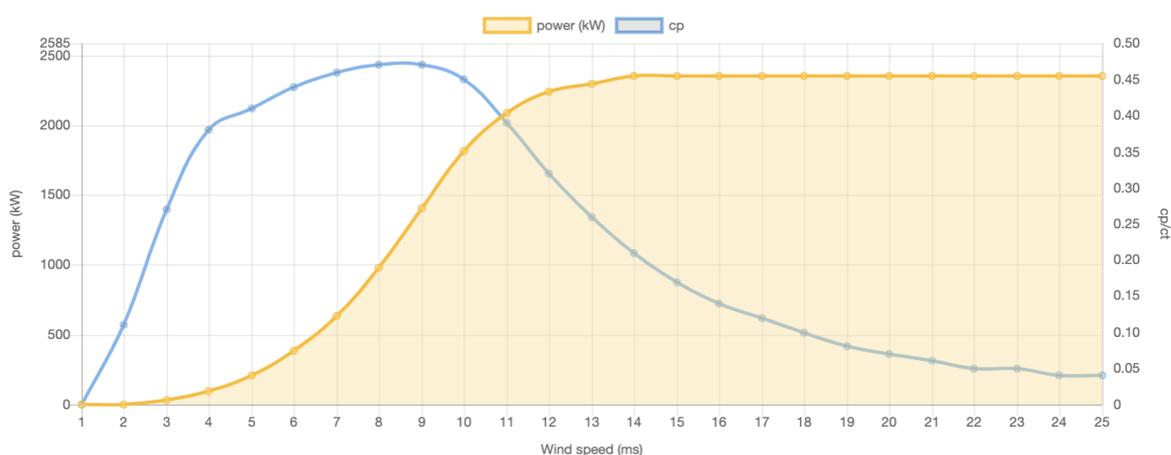


Figura 6 - Curva de Potência dos Aerogeradores E-92. Fonte: [123]

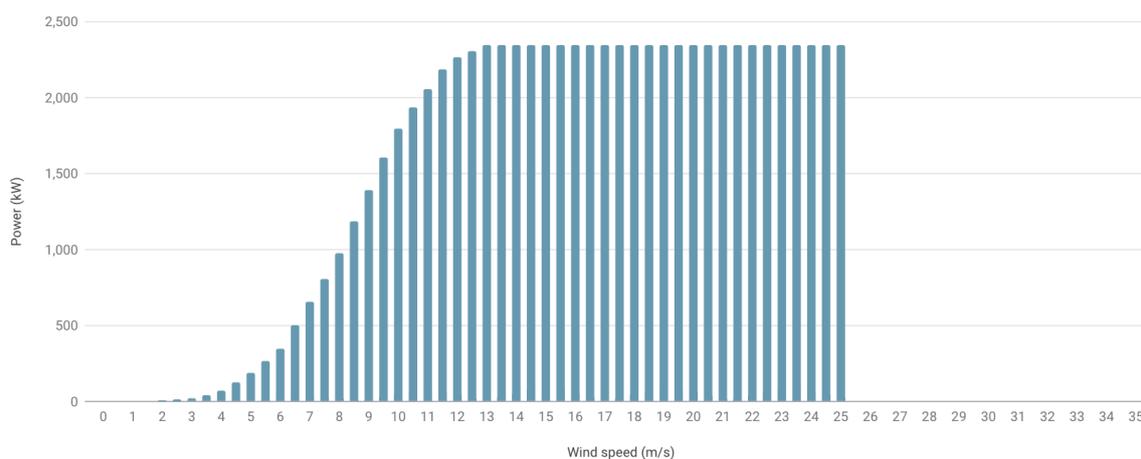


Figura 7 - Transitório de Velocidade dos Aerogeradores E-92. Fonte: [124]

Os dados de medição dos aerogeradores são armazenados em sistemas de controle chamados *Supervisory Control and Data Acquisition* (SCADA). Esses sistemas coletam e monitoram dados em tempo real, incluindo informações sobre a produção de energia, condições meteorológicas, desempenho dos aerogeradores, e status operacional [125]. Diante disso, o levantamento das séries de geração histórica foi feito a partir dos dados medidos da plataforma integrada de medição Way2, plataforma que realiza o monitoramento e gerenciamento dos dados de energia em tempo real. Os dados de geração são obtidos de maneira horária para cada parque eólico, no período de julho de 2014 a dezembro de 2023 em MWh, conforme apresentado na Figura 8. A curva em amarelo se refere ao parque eólico Pindaí I, em vermelho para Pindaí II, em azul Pindaí III e em verde o parque de Pindaí IV.

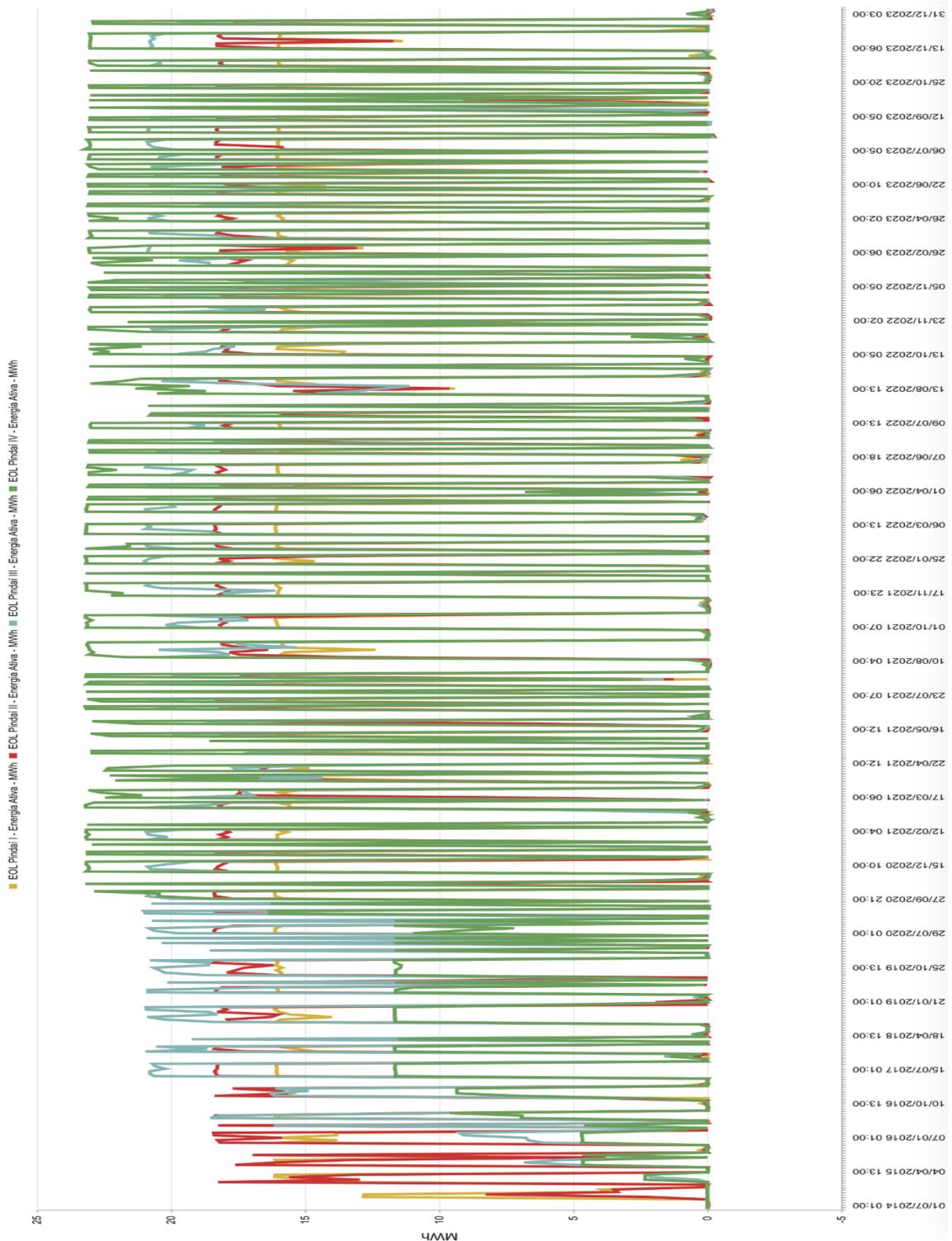


Figura 8 - Geração Horária Histórica do Complexo Pindaí

Dados de geração eólica horária contêm uma quantidade significativa de detalhes, capturando, por exemplo, flutuações causadas por mudanças meteorológicas em períodos muito curtos. Com mais de 80.000 horas no período de quase dez anos, essa discretização introduz uma grande parcela de volatilidade e ruído no conjunto de dados. Isso pode refletir em condições muito diferentes, dependendo da hora do dia, de padrões climáticos e influências geográficas locais, complicando o processo de reconhecimento de padrões. A alta resolução temporal significa que os métodos tradicionais de processamento de dados podem ser insuficientes e podem exigir técnicas mais robustas, como bancos de dados de séries temporais que são otimizados para lidar com grandes volumes de dados de sequência de forma eficiente. Isso exige recursos computacionais para armazenamento, processamento e análise para lidar com operações de dados em grande escala, incluindo processamento e armazenamento em tempo real.

A escolha de desenvolver modelos de previsão de geração eólica de longo prazo, especificamente com um horizonte temporal de 12 meses à frente, é uma decisão estratégica que visa alcançar objetivos tanto para a operação eficiente dos parques eólicos quanto para o planejamento de longo prazo no setor de energia renovável. A previsão com essa extensão temporal permite aos operadores e gestores obterem uma visão das tendências sazonais e interanuais que influenciam a disponibilidade de recursos eólicos. Visão importante para otimizar a operação dos parques eólicos ao longo do ano.

Um dos principais objetivos de se focar em previsões de longo prazo é o planejamento estratégico e a gestão de riscos. Conhecer antecipadamente os padrões de vento esperados para o próximo ano permite que os operadores de parques eólicos tomem decisões informadas sobre a manutenção preventiva e corretiva, evitando períodos de inatividade não planejada que poderiam comprometer a eficiência operacional. A maioria dos processos como orçamento, alocação de recursos e programações de manutenção, é conduzida em bases mensais ou trimestrais. A agregação mensal garante que as informações derivadas dos dados sejam diretamente aplicáveis a esses processos.

Além disso, essa informação é necessária para ajustar as estratégias de negociação de contratos de fornecimento de energia, assegurando que a produção eólica esteja alinhada com as demandas do mercado e as obrigações contratuais.

A previsão de longo prazo também desempenha um papel na avaliação e no planejamento de investimentos. Para investidores e desenvolvedores de projetos de energia eólica, a capacidade de prever a geração de energia com precisão ao longo de 12 meses permite uma melhor avaliação da viabilidade econômica dos projetos, influenciando decisões sobre a expansão da capacidade instalada e a alocação de recursos financeiros. Isso inclui a determinação de locais ideais para novos parques eólicos, onde os padrões de vento são mais favoráveis, e a previsão de retornos financeiros para atrair financiamentos.

Portanto, as metodologias de aprendizado de máquina desenvolvidas e aplicadas nesse trabalho consistem na previsão de geração eólica multi-passos direta, com um horizonte de 12 meses.

A previsão direta é uma abordagem na previsão de séries temporais multi-passos onde múltiplos modelos são treinados de forma independente, cada um destinado a prever um ponto futuro específico na série temporal. Diferentemente da abordagem recursiva, onde um único modelo é iterativamente aplicado para gerar previsões subsequentes, a previsão direta evita a propagação de erros entre os passos, uma vez que cada modelo prevê diretamente o horizonte desejado sem depender das previsões anteriores [126].

Os dados extraídos de sistemas de armazenamento, frequentemente apresentam valores errôneos devido a diversos problemas de medição. Sensores defeituosos ou mal calibrados podem registrar valores incorretos de produção de energia, velocidade do vento e outras métricas. Além disso, condições meteorológicas adversas como tempestades, podem causar leituras imprecisas ou falhas temporárias nos sensores. Problemas de comunicação, como interrupções na transmissão de dados entre os aerogeradores e o sistema central, também podem resultar em dados perdidos ou corrompidos. Diferenças nos relógios dos dispositivos de medição podem causar inconsistências nos dados temporais, como registros duplicados ou em horários incorretos. Durante manutenções e paradas programadas, os sensores podem não registrar dados ou registrar valores anômalos devido à ausência de geração de energia.

Diante desses problemas de medição, é importante realizar uma etapa de pré-processamento antes da agregação dos dados horários em base mensal na série temporal. Essa etapa consiste na verificação de dados faltantes ou nulos, identificando e tratando dados

ausentes que podem ocorrer devido a falhas nos sensores ou interrupções na transmissão de dados.

A natureza de dados faltantes ou nulos presentes em séries temporais podem ser classificada em três tipos [127] e [128]:

- *Missing Completely at Random* (MCAR): a probabilidade de um ponto de dados estar faltando é a mesma para todas as observações, de maneira completamente aleatória. Isso significa que a ausência de um dado não está relacionada a nenhuma outra informação ou valor dos dados. Por exemplo, um sensor pode falhar de vez em quando sem nenhum padrão específico.
- *Missing at Random* (MAR): a ausência dos dados está relacionada a outros dados que já foram observados, mas não ao valor específico que está faltando.
- *Missing Not at Random* (MNAR): os dados estão faltando por uma razão específica relacionada ao valor que está ausente. Isso significa que a ausência de dados depende diretamente do próprio valor que não foi registrado. Por exemplo, se um aerogerador para de funcionar quando a produção de energia é muito baixa, e por isso não há registro desse valor baixo, teremos uma situação de dados faltantes não aleatórios.

Identificar o tipo de ausência é importante, pois influencia a escolha dos métodos estatísticos para lidar com os dados faltantes e garantir a validade da análise. Existem várias técnicas para tratar esses dados, como a imputação média ou mediana, onde os dados faltantes são substituídos pela média ou mediana dos valores disponíveis anteriores, por exemplo. Outras técnicas incluem o *Last Observation Carried Forward* (LOCF) e *Next Observation Carried Backward* (NOCB), que utilizam o último ou o próximo ponto de dados observado para preencher as lacunas, mantendo a integridade temporal nas sequências de dados [128].

Também existem métodos de interpolação linear e *splines*, que estimam valores faltantes usando funções lineares ou polinomiais que são úteis quando os dados exibem tendências. A fórmula para interpolação linear entre dois pontos (x_{t-1}, y_{t-1}) e (x_{t+1}, y_{t+1}) com (y_1) e (y_2) conhecidos e (y) ausente no tempo (t) pode ser expressa por (24):

$$y_t = y_{t-1} + \frac{(y_{t+1} - y_{t-1})}{(x_{t+1} - y_{t-1})} \times (x_t - x_{t-1}) \quad (24)$$

Além disso, técnicas como imputação baseada em modelo estatístico como ARIMA (*Autoregressive Integrated Moving Average*) ou modelos de espaço de estado podem prever valores faltantes, ajustando um modelo aos pontos de dados observados e usando esse modelo para estimar valores faltantes. Abordagens de aprendizado de máquina como *k-Nearest Neighbors* (k-NN) ou arquiteturas de *deep learning* podem prever valores ausentes com base em padrões aprendidos com os dados de treinamento.

Assim como dados nulos ou ausentes, também é importante a verificação de dados horários duplicados. Isso envolve a detecção e remoção de registros duplicados que podem resultar de erros de sincronização ou retransmissões de dados. Ao realizar essas verificações, é possível garantir que os dados agregados mensais sejam mais confiáveis. Entradas duplicadas também podem surgir por erros de medição ou coleta de dados devido a erros do sistema, erro humano, transmissão repetida de pacotes de dados ou anomalias de processamento de dados durante etapas de agregação ou transformação, caso as operações forem executadas de forma repetida ou se os mecanismos de *checkpoint* falharem.

Medições duplicadas podem inflar ou esvaziar artificialmente a variabilidade percebida nos dados, afetando medidas como média, variância e desvio padrão, além de interferir em relações ou tendências que não são reais, dificultando a análise de correlações ou covariâncias, por exemplo. A resolução mais efetiva é a checagem em relação à data e hora que está presente em bancos de dados de séries temporais, e conseqüentemente a remoção de entradas duplicadas.

Para esta dissertação foram aplicadas as técnicas de *Missing Completely at Random* (MCAR), de maneira que, para dados horários faltantes, foram feitas médias das últimas 10 horas equivalentes aos dias anteriores e a remoção de dados duplicados.

O método de agregação mensal pode variar dependendo da natureza dos dados e da finalidade da análise, podendo ser aplicada a soma (dados horários em totais mensais), a média (pontos de dados por hora, benéfica para entender tendências centrais, como velocidade média do vento ou potência média) ou a média ponderada (pesos podem ser atribuídos a pontos de dados por hora antes da média para levar em conta as horas variáveis

de operação ou outros fatores que podem afetar a significância de cada ponto). Neste caso será aplicada a soma, expressa por (25):

$$S_m = \sum_{i=1}^N X_{m,i} \quad (25)$$

onde S_m é a soma do mês M , N é o número de horas do mês e X_m são os pontos de dados por hora.

A geração eólica do complexo de Pindaí, agregada por mês, pode ser visualizada por meios gráficos na Figura 9 e em forma de números absolutos na Tabela 7 (referente ao somatório dos quatro parques).

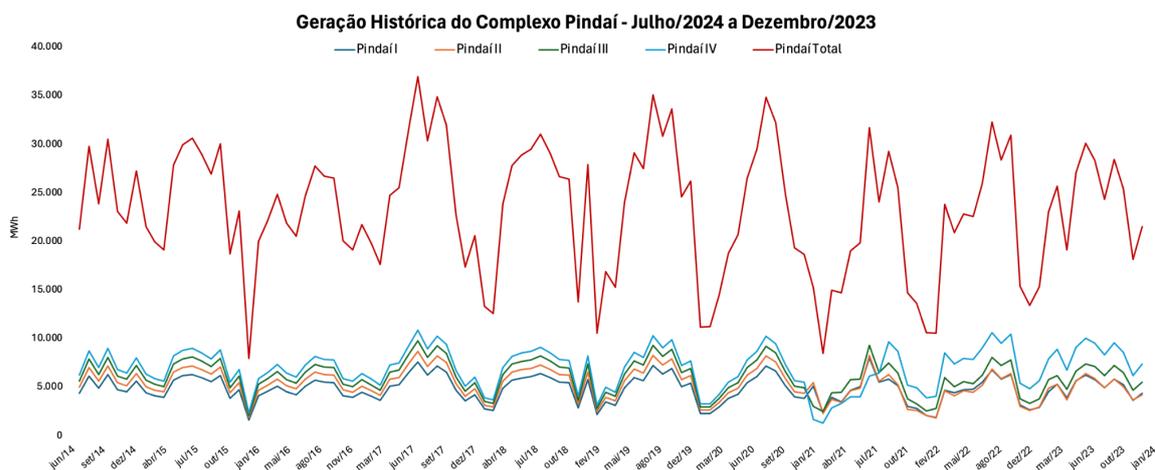


Figura 9 - Geração Mensal Histórica do Complexo Pindaí [MWh]

Tabela 7 - Geração Mensal Histórica do Complexo Pindaí [MWh]

Ano	Janeiro	Fevereiro	Março	Abril	Maior	Junho	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
2014	-	-	-	-	-	-	21.280,05	29.770,34	23.865,30	30.529,83	23.108,27	21.910,84
2015	27.239,98	21.526,29	19.992,34	19.150,17	27.866,82	29.943,98	30.611,95	29.033,59	26.930,78	30.063,55	18.716,82	23.114,69
2016	7.931,91	20.005,69	22.154,25	24.855,85	21.872,55	20.529,45	24.705,03	27.765,70	26.720,26	26.527,57	20.060,94	19.116,99
2017	21.718,35	19.747,18	17.640,80	24.759,97	25.508,24	31.543,69	36.977,56	30.380,12	34.899,79	31.954,49	22.727,51	17.349,70
2018	20.580,09	13.320,06	12.597,85	23.848,18	27.825,89	28.911,27	29.473,54	31.022,99	29.048,36	26.697,73	26.421,86	13.774,35
2019	27.949,26	10.545,63	16.890,62	15.283,62	24.010,89	29.094,91	27.487,89	35.082,71	30.849,61	33.658,89	24.617,90	26.200,77
2020	11.168,09	11.217,73	14.498,21	18.756,44	20.716,74	26.512,13	29.497,55	34.861,09	32.238,02	24.920,82	19.330,54	18.667,91
2021	15.236,45	8.466,07	14.971,92	14.716,70	19.008,71	19.853,35	31.732,53	24.089,98	29.294,89	25.525,59	14.704,22	13.629,71
2022	10.587,72	10.568,53	23.799,91	20.919,92	22.844,20	22.543,00	25.994,05	32.306,06	28.411,11	30.921,91	15.408,20	13.439,00
2023	15.328,97	23.046,24	25.687,46	19.141,45	27.088,00	30.117,77	28.315,33	24.355,70	28.450,51	25.434,09	18.170,29	21.522,13

A análise estatística dos dados de geração ao longo do período de julho de 2014 a dezembro de 2023 é apresentada na tabela 8. A média da geração, calculada em 23.292,29 MWh, fornecendo uma referência central para o desempenho típico do complexo eólico. Essa média é acompanhada por um desvio padrão de 6.532,45 MWh, o que indica uma

variabilidade considerável nos dados, sugerindo que a produção de energia do parque eólico não é constante, mas sujeita a flutuações substanciais ao longo do período observado.

O valor mínimo de geração registrado foi de 7.931,91 MWh, que representa um ponto de baixa significativa na produção de energia. Este valor pode refletir condições adversas, como períodos de baixo vento, manutenção do parque, ou outros fatores operacionais que impactaram negativamente a capacidade de geração. Em contraste, o valor máximo de 36.977,56 MWh marca um pico de produção, possivelmente associado a condições ideais de vento ou a uma maximização da eficiência operacional do parque. A diferença entre esses valores máximo e mínimo, conhecida como amplitude, é de 29.045,65 MWh, destacando a vasta gama de variação na capacidade de geração do parque ao longo do tempo.

Os quartis da distribuição fornecem informações adicionais sobre a dispersão dos dados. O primeiro quartil (25%) situa-se em 19.123,11 MWh, indicando que 25% das observações de geração estão abaixo desse valor. Já o terceiro quartil (75%) é de 27.765,70 MWh, o que significa que 75% das observações de geração estão abaixo desse limiar. A mediana, que se localiza entre esses quartis, oferece uma visão do ponto central da distribuição, em 23.938,09 MWh, indicando uma leve assimetria positiva nos dados.

O desvio padrão é uma medida estatística que indica a quantidade média pela qual os valores de geração de energia eólica diferem da média ao longo do período analisado. Em termos práticos, um desvio padrão dessa magnitude sugere que a geração de energia do complexo experimenta flutuações substanciais mês a mês. Este valor é particularmente relevante porque, em um contexto de operação e planejamento de energia, uma alta variabilidade na produção pode complicar a previsão e o equilíbrio entre oferta e demanda. Um desvio padrão de 6.532,45 MWh significa que em muitos meses, a geração de energia pode estar significativamente acima ou abaixo da média de 23.292,29 MWh. Tal variação pode ser influenciada por fatores sazonais, como variações na velocidade do vento, mudanças climáticas, ou operações de manutenção que afetam a disponibilidade das turbinas.

Além da amplitude de 29.045,65 MWh, que representa a diferença entre o valor máximo de 36.977,56 MWh e o valor mínimo de 7.931,91 MWh, indica a grande variação na geração de energia ao longo do período observado.

Tabela 8 - Estatísticas Descritivas dos Dados de Geração

Estatísticas Descritivas	Valores
Média	23.292,29
Desvio Padrão	6.532,45
Mínimo	7.931,91
1° Quartil (25%)	19.123,11
Mediana	23.938,09
3° Quartil (75%)	28.387,17
Máximo	36.977,56
Amplitude	29.045,65

O conjunto de dados de treinamento influencia a capacidade dos modelos de capturar padrões e dinâmicas dentro da base de dados, consistindo em uma coleção de pares de entrada-saída usados pelo algoritmo de aprendizado para aprender pelo exemplo supervisionado.

Um dos principais riscos durante a fase de treinamento é o *overfitting*, em que um modelo aprende os detalhes e o ruído nos dados de treinamento a ponto de impactar negativamente o desempenho em novas observações, ocorrendo normalmente quando o conjunto de dados é muito pequeno ou quando o modelo é excessivamente complexo. Por outro lado, o *underfitting* ocorre quando um modelo é muito simples em relação à complexidade necessária para capturar os padrões nos dados, quando os dados de treinamento são insuficientemente representativos ou na presença de ruídos, resultando em um modelo com desempenho ruim mesmo no conjunto de dados de treinamento, não conseguindo capturar tendências.

Em relação ao conjunto de dados de validação, é usado para fornecer uma avaliação imparcial do ajuste do modelo para um novo conjunto de dados distinto daquele utilizado para treinamento, possibilitando também o ajuste dos hiperparâmetros do modelo. Nesta etapa, é importante refinar os hiperparâmetros do modelo sem o uso do conjunto de testes, que deve permanecer única e exclusivamente para avaliação final, evitando vazamento de

dados. A função do conjunto de validação é evitar o *overfitting* enquanto otimiza a configuração e os hiperparâmetros, garantindo que o modelo generalize bem para dados novos.

Na etapa de validação cruzada, a matemática envolvida na seleção dos melhores parâmetros baseia-se na formação de uma matriz multidimensional, onde cada dimensão representa um hiperparâmetro específico com seus possíveis valores. No modelo do perceptron de múltiplas camadas, por exemplo, podem ser utilizados seis hiperparâmetros: o número de neurônios na camada oculta (com 3 valores), a taxa de *dropout* (4 valores), a taxa de aprendizado (3 valores), o otimizador (3 valores), a função de ativação (3 valores) e o tamanho de amostras *Batch Size* (3 valores). Isso resulta em uma matriz tridimensional com 972 ($3 \times 4 \times 3 \times 3 \times 3 \times 3$) combinações possíveis. Cada célula dessa matriz corresponde a uma combinação única, na qual o modelo é treinado e avaliado. Durante a validação cruzada, o conjunto de dados é dividido em k partes (*fold*s), e o modelo é treinado k vezes, cada vez utilizando um *fold* diferente como conjunto de validação e os restantes como conjunto de treinamento [129]. O desempenho médio do modelo em cada combinação de hiperparâmetros é calculado e a combinação que resulta na melhor métrica de avaliação média é escolhida como a melhor configuração. Esse processo sistemático garante uma exploração do espaço de hiperparâmetros e ajuda a encontrar a configuração ótima que maximiza a performance do modelo.

À medida que o número de hiperparâmetros aumenta, a grade cresce exponencialmente, o que pode aumentar a carga computacional e tornar o processo de busca por hiperparâmetros ótimos extremamente demorado e ineficiente. Este fenômeno é conhecido como explosão combinatorial de hiperparâmetros, diferentemente da maldição da dimensionalidade, que está relacionada ao número de entradas no conjunto de dados [79]. Um recurso que pode ser aproveitado durante o treinamento do modelo é a parada antecipada (*Early Stopping*). Esse método avalia o desempenho do modelo em um conjunto de validação após cada iteração. Se o desempenho no conjunto de validação não melhorar após um determinado número de iterações consecutivas, o treinamento é interrompido.

O conjunto de dados de teste é projetado para fornecer uma avaliação da capacidade de generalização do modelo para novos dados, servindo como teste definitivo do desempenho do modelo fora das condições controladas das fases de treinamento e validação.

Como visto anteriormente, os dados de teste devem ser completamente independentes dos conjuntos de treinamento e validação, para garantir que as métricas de desempenho avaliadas nos dados de teste reflitam previsões honestas, livres de vieses que poderiam ser introduzidos se o modelo tivesse sido exposto a esses dados durante o treinamento. Esse conjunto deve ser representativo da população geral ou da variedade de cenários que o modelo encontrará em uso real, tendo uma distribuição semelhante de entradas e saídas.

A literatura não fornece uma diretriz definitiva sobre a maneira correta ou mais adequada de separar os conjuntos de dados de treinamento, validação e teste. A maneira ideal depende do contexto específico em que a série temporal está sendo aplicada, incluindo a natureza dos dados e os objetivos da análise. Além disso, a quantidade e a qualidade dos dados disponíveis também servem de insumos para definir a proporção de separação, de maneira que conjuntos de dados maiores podem permitir uma maior alocação para dados de treinamento sem comprometer as fases de validação e teste. Embora a literatura forneça várias recomendações para a divisão de conjuntos de dados, como proporções de 70% para treinamento, 15% para validação e 15% para teste, ou 60% para treinamento, 20% para validação e 20% para teste, essas proporções devem ser vistas apenas como pontos de partida. Esse contexto é abordado em diversos estudos como nos trabalhos de Friedman [69] e Lundberg e Lee [73] onde é enfatizada a necessidade de flexibilidade e ajuste com base em requisitos específicos do modelo e comportamentos dos dados.

Portanto, para este estudo, foi definido que o conjunto de treinamento será delimitado de julho de 2014 a janeiro de 2021, correspondendo a 70% dos dados totais. O conjunto de validação será de fevereiro de 2021 a dezembro de 2022 (20%), e por fim, os dados de teste de janeiro de 2023 até dezembro de 2023 (10%), conforme abordado na seção 5.1 referente aos resultados da análise da série temporal. A composição da série temporal de geração histórica é agregada com dados meteorológicos, abordado na seção seguinte.

3.2 Modelos de Reanálise de Previsão Meteorológica com Alta Resolução

Conforme abordado na seção de referencial teórico, no presente estudo são utilizadas variáveis meteorológicas derivadas de modelos de reanálise, que englobam dados históricos e previstos, adaptados ao horizonte específico relacionado ao Complexo Eólico Pindaí. Os modelos são projetados para capturar a dinâmica atmosférica em diferentes escalas, desde sistemas meteorológicos amplos até fenômenos localizados como tempestades ou microclimas, sendo desenvolvidos por serviços meteorológicos internacionais, como o *Deutscher Wetterdienst* (DWD) na Alemanha, NOAA nos Estados Unidos, *Météo-France* na França, e outros [30], apresentados na Tabela 9.

Tabela 9 - Modelos Meteorológicas e Provedores Internacionais. Fonte: [30]

Modelo meteorológico	Provedor Nacional de Meteorologia	País de origem	Resolução	Duração da previsão	Frequência de atualização
ICON	Serviço Meteorológico Alemão (DWD)	Alemanha	2 - 11 km	7,5 dias	A cada 3 horas
GFS e HRRR	NOAA	Estados Unidos	3 - 25 km	16 dias	A cada hora
ARPEGE E AROME	Météo-France	França	1 - 25 km	4 dias	A cada hora
IFS e AIFS	ECMWF	União Europeia	25 km	7 dias	A cada 6 horas
UKMO	Escritório Met do Reino Unido	Reino Unido	2 - 10 km	7 dias	A cada hora
MSM e GSM	JMA	Japão	5 - 55 km	11 dias	A cada 3 horas
MET Nordic	MET Noruega	Noruega	1 km	2,5 dias	A cada hora
HARMONIE	KNMI	Holanda	2 km	2,5 dias	A cada hora
HARMONIE	DMI	Dinamarca	2 km	2,5 dias	A cada 3 horas
GEM	Serviço meteorológico canadense	Canadá	2,5 km	10 dias	A cada 6 horas
GFS GRAPES	Administração Meteorológica da China (CMA)	China	15 km	10 dias	A cada 6 horas
ACCESS-G	Escritório Australiano de Meteorologia (BOM)	Austrália	15 km	10 dias	A cada 6 horas
COSMO 21 e 5M	AM ARPAE ARPAP	Itália	2 km	3 dias	A cada 3 horas

A extração dos dados é realizada através de API em formato JSON disponibilizada na plataforma do Open-Meteo por meio da ferramenta de código aberto, de maneira que a partir de um algoritmo em linguagem Python é possível realizar a consulta das variáveis exógenas escolhidas e determinar os parâmetros de busca.

A extração inicia-se através das coordenadas de latitude e longitude do ponto de grade, onde o complexo de Pindaí está localizado na latitude de 14° 29' 24" Sul (-14.5) e longitude de 42° 41' 24" Oeste (-42.5), conforme apresentado na Figura 10. A data de consulta para as variáveis históricas será de 01 de julho de 2014 até 31 de dezembro de 2022, apresentado na Tabela 10. As variáveis de busca são disponibilizadas em base horária e consequentemente agrupada posteriormente para base mensal.

Tabela 10 - Parâmetros e Descrição das Variáveis Históricas. Fonte: Adaptado de [30]

Parâmetro	Descrição	Unidade
<i>Latitude e Longitude</i>	Latitude = 14° 29' 24" e longitude = 42° 41' 24".	Graus (°), minutos (') e segundos (")
<i>Temperatura</i>	Temperatura do ar a 2 metros acima do solo	°C (°F)
<i>Umidade Relativa do Ar</i>	Umidade relativa a 2 metros acima do solo	%
<i>Precipitação</i>	Soma total da precipitação (chuva, aguaceiros, neve) da hora anterior. Dados armazenados com precisão de 0,1 mm.	mm
<i>Pressão da Superfície</i>	Pressão do ar atmosférico reduzida ao nível médio do mar (MSL) ou pressão na superfície. Normalmente, a pressão no nível médio do mar é usada em meteorologia, a pressão superficial diminui com o aumento da elevação.	hPa
<i>Velocidade do Vento a 10 metros</i>	Velocidade do vento a 10 metros acima do solo, sendo o nível padrão.	m/s
<i>Velocidade do Vento a 100 metros</i>	Velocidade do vento a 100 metros acima do solo.	m/s
<i>Rajadas de Vento</i>	Rajadas a 10 metros do solo na hora indicada, sendo definidas como as rajadas de vento máximas da hora anterior.	m/s

Para todos os parâmetros da Tabela 10, foi estipulada a combinação dos modelos meteorológicos chamada *Best Match*, reunindo informações dos modelos do Centro Europeu de Previsões Meteorológicas de Médio Prazo (ECMWF) [131], ERA5 (ECMWF *Reanalysis* v5 - quinta geração de reanálises atmosféricas do ECMWF com resolução espacial de 31 km e resolução temporal de uma hora) [132], ERA5 *Land* (reanálise específica apenas em variáveis terrestres com resolução espacial de 9 km) [133], ERA5 *Seamlessly* (dados de

satélite e observações de superfície com modelos atmosféricos e oceânicos) [132] e CERRA (*Copernicus Regional Reanalysis for Europe*) [134]. É possível visualizar o resultado da busca por meios gráficos conforme a Figura 11 ou pelo *output* do algoritmo na Figura 12.

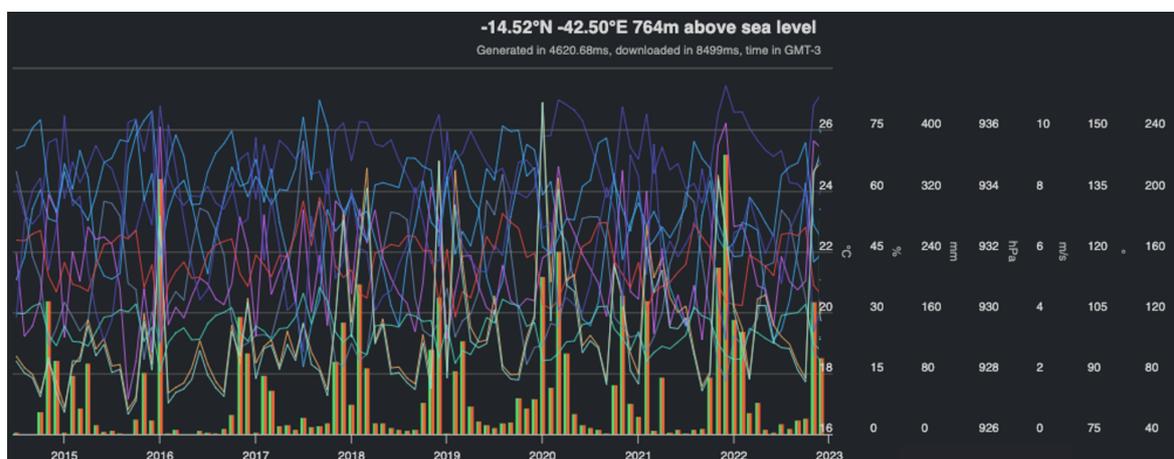


Figura 11 - Variáveis Meteorológicas Históricas Extraídas do Modelo de Reanálise *Best Match*. Fonte: [30]

	date	temperature_2m	relative_humidity_2m	precipitation	rain	surface_pressure	cloud_cover	wind_speed_10m	wind_speed_100m	wind_direction_10m	wind_direction_100m	wind_gusts_10m	direct_radiation
0	2014-07-01 03:00:00+00:00	18.654499	64.831306	0.0	0.0	930.776367	1.5	1.868154	4.418144	105.524178	95.194344	3.7	0.0
1	2014-07-01 04:00:00+00:00	18.004499	68.198761	0.0	0.0	930.502258	0.0	2.158703	5.024938	103.392517	95.710503	4.1	0.0
2	2014-07-01 05:00:00+00:00	17.754499	69.737404	0.0	0.0	929.974304	3.6	2.236068	5.108816	100.304787	93.386394	4.6	0.0
3	2014-07-01 06:00:00+00:00	17.354500	72.703957	0.0	0.0	930.135864	1.8	2.137756	4.904080	100.784266	92.337257	4.7	0.0
4	2014-07-01 07:00:00+00:00	16.854500	76.532448	0.0	0.0	929.994202	1.8	1.941649	4.304649	101.888634	92.662949	4.7	0.0

Figura 12 - Variáveis Meteorológicas Históricas Extraídas por API JSON

Para a etapa de previsão, será necessário atualizar a série temporal com projeções das variáveis meteorológicas no intervalo de 01 de janeiro de 2023 a 31 de dezembro de 2023. Diferente das variáveis realizadas, o sistema de previsão proveniente a partir do API JSON da Open-Meteo disponibiliza previsões a partir de dados diários, e não horários. Os modelos meteorológicos utilizados para a consulta podem ser vistos na Tabela 11.

Tabela 11 - Modelos Meteorológicos Utilizados para Previsão. Fonte: [30]

Modelo Climático	Origem	Executado por	Resolução
CMCC-CM2-VHR4	Itália	Fundação Centro Euro-Mediterraneo de Mudanças Climáticas, Lecce (CMCC)	30 km
FGOALS_g3_H	China	Academia Chinesa de Ciências, Pequim (CAS)	28 km
HiRAM_SIT_HR	Taiwan	Centro de Pesquisa para Mudanças Ambientais, Academia Sinica, Nankang, Taipei (AS-RCEC)	25 km
MRI_AGCM3_2_S	Japão	Instituto de Pesquisa Meteorológica, Tsukuba, Ibaraki (MRI)	20 km
EC_Earth3P_HR	Europa	Consórcio EC-Earth, Rossby Center, Instituto Meteorológico e Hidrológico Sueco/SMHI, Norrköping, Suécia	29 km
MPI_ESM1_2_XR	Alemanha	Instituto Max Planck de Meteorologia, Hamburgo 20146, Alemanha	51 km
NICAM16_8S	Japão	Agência Japonesa para Ciência e Tecnologia Marinha e Terrestre, Kanagawa 236-0001, Japão (MIROC)	31 km

Na Tabela 12, são apresentadas as variáveis extraídas da plataforma de previsão. Diferentemente das extrações de dados realizadas anteriormente, a velocidade do vento a

100 metros acima do nível do solo não está disponível. Portanto, a extração foi realizada para as seguintes características:

Tabela 12 - Parâmetros e Descrição das Variáveis Previstas

Parâmetro	Descrição	Unidade
<i>Temperatura</i>	Temperatura média do ar a 2 metros acima do solo	°C (°F)
<i>Umidade Relativa do Ar</i>	Umidade relativa média a 2 metros acima do solo	%
<i>Precipitação</i>	Soma da precipitação diária (incluindo chuva, aguaceiros e queda de neve).	mm
<i>Pressão ao Nível do Mar</i>	Pressão do ar atmosférico reduzida ao nível médio do mar (MSL) ou pressão na superfície.	hPa
<i>Velocidade do Vento a 10 metros</i>		m/s
<i>Velocidade Máxima do Vento a 10 metros</i>	Velocidade média e máxima do vento 10 metros acima do solo em um dia	
<i>Rajadas de Vento</i>	Rajadas a 10 metros do solo na hora indicada, sendo definidas como as rajadas de vento máximas da hora anterior.	m/s

É possível visualizar o resultado das variáveis de previsão por meios gráficos conforme a Figura 13 ou pelo *output* do algoritmo na Figura 14.

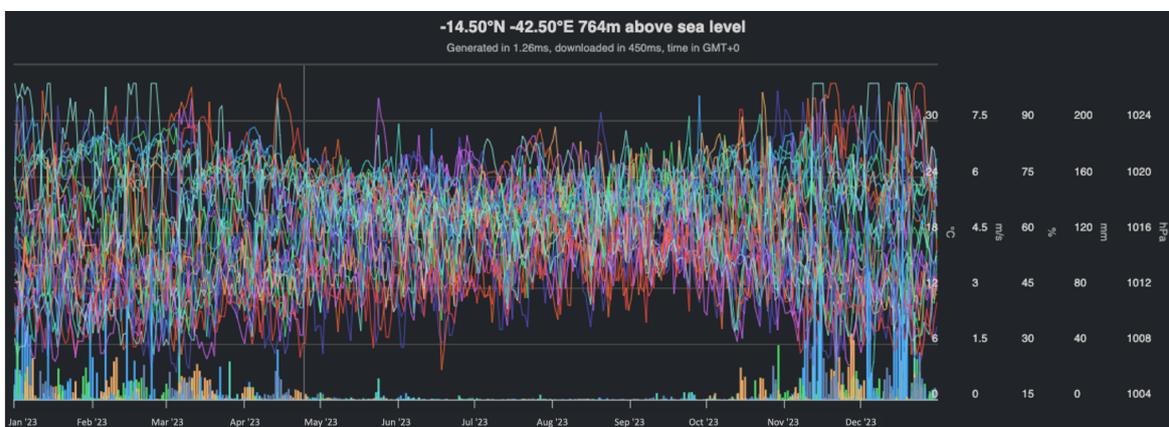


Figura 13 - Variáveis Meteorológicas Previstas Extraídas por API JSON. Fonte: [30]

	date	temperature_2m_mean	wind_speed_10m_mean	wind_speed_10m_max	relative_humidity_2m_mean	precipitation_sum	pressure_msl_mean
0	2023-01-01 00:00:00+00:00	22.434219	2.915140	4.217610	88.961189	4.742554	1016.448486
1	2023-01-02 00:00:00+00:00	22.437826	2.804382	4.049246	88.922531	4.386754	1015.445435
2	2023-01-03 00:00:00+00:00	23.548958	2.752959	4.165493	80.818428	1.172514	1014.343933
3	2023-01-04 00:00:00+00:00	23.910088	2.926465	4.056580	79.714317	1.330283	1013.942566
4	2023-01-05 00:00:00+00:00	25.221216	2.762309	4.173037	68.610214	0.660285	1013.141113
361	2023-12-27 00:00:00+00:00	24.650864	3.205134	4.569157	56.847260	1.527906	1013.768188
362	2023-12-28 00:00:00+00:00	23.246946	3.483828	4.002165	68.874069	3.902505	1012.363586
363	2023-12-29 00:00:00+00:00	23.193029	2.639092	2.647394	82.900871	5.087454	1011.058838
364	2023-12-30 00:00:00+00:00	24.289114	1.571102	2.589175	73.927681	5.084708	1011.854309
365	2023-12-31 00:00:00+00:00	25.935198	1.962467	3.768425	54.954475	0.677598	1012.449646

Figura 14 - Variáveis Meteorológicas Previstas Extraídas por API JSON

Conforme será visualizado na seção de resultados, a correlação entre variáveis exógenas e a produção de energia do complexo eólico destaca o papel significativo da velocidade do vento a 100 metros acima do nível do solo. No entanto, devido à indisponibilidade dessa medição específica para consulta no API de dados futuros, torna-se necessário estimar a velocidade do vento de 100 metros usando dados coletados a 10 metros acima do nível do solo, ajustados através de uma transformação baseada na altura relativa e rugosidade do terreno local.

A transformação da velocidade do vento em altura relativa é uma técnica em estudos de energia eólica que permite a extrapolação das velocidades do vento medidas de uma altura para outra, usando a teoria logarítmica do perfil do vento [104]. A transformação é baseada no perfil logarítmico do vento, que é derivado da lei da parede na dinâmica dos fluidos, sendo representada pela seguinte formulação (26):

$$V_h = V_{h_0} \times \frac{\log\left(\frac{h}{z}\right)}{\log\left(\frac{h_0}{z}\right)} \quad (26)$$

onde V_h é a velocidade do vento de destino a ser calculada (100 metros), V_{h_0} é a velocidade do vento em uma altura referência (10 metros), h é a altura destino, h_0 é a altura referência e z é o comprimento de rugosidade do terreno, característico da textura da superfície local e dos obstáculos.

A rugosidade do terreno é influenciada por características superficiais como vegetação, estruturas construídas e topografia, afetando o perfil do vento próximo ao solo. O comprimento aerodinâmico é um parâmetro da modelagem do comportamento do vento próximo à superfície, levando em conta os efeitos da cobertura da terra sobre os perfis eólicos, de maneira que esse comprimento é representado pela altura acima do solo na qual o perfil de velocidade do vento é ajustado logaritmicamente devido às características da

superfície. Um valor alto indica maior rugosidade superficial, resultando em uma maior extração do vento em altitudes mais baixas. Para realizar o cálculo foi considerado um valor de 0,4 metros (z), de acordo com o modelo de rugosidade do estado da Bahia [135], transformando a velocidade do vento prevista de 10 metros para 100 metros. Com isso, a base de dados meteorológica é complementada e agrupada com a base de geração histórica de cada parque eólico.

3.3 Machine Learning

O trabalho foi desenvolvido em linguagem de programação Python, com destaque para as bibliotecas *Scikit-Learn*, *TensorFlow*, *Keras*, *Skforecast*, *XGBoost*, *pandas*, *NumPy*, *matplotlib*, *seaborn* e *Open-Meteo Requests*, conforme especificado na seção 4.

Para melhor entendimento, a Figura 15 ilustra o fluxograma do processo realizado da aplicação de modelos de aprendizado de máquina para a previsão de energia eólica. O algoritmo começa com a extração de dados e pré-processamento, submetidos a processos de limpeza, normalização e transformação para garantir que estejam em um formato adequado para análise. Em seguida, são inseridas variáveis meteorológicas importantes provenientes dos modelos de reanálise fornecidos pela Open-Meteo. Após a inserção dessas variáveis, realizam-se análises estatísticas e divisão do *dataset* em conjuntos de treinamento, validação e teste. A verificação do horizonte de previsão acontece para determinar se deve retornar a etapa de nova extração de dados ou continuar para a seleção do modelo de aprendizado de máquina.

Na fase de seleção do modelo, diferentes técnicas de aprendizado de máquina são consideradas, incluindo regressão linear rígida, florestas aleatórias, regressão por vetores suporte, perceptron de múltiplas camadas e *eXtreme Gradient Boosting*. Cada modelo é configurado com suas respectivas características e otimizado com hiperparâmetros para melhorar o desempenho dos modelos escolhidos, conforme será abordado na seção seguinte. A avaliação dos modelos é feita utilizando métricas como o coeficiente de determinação (R^2), erro percentual absoluto médio (MAPE), raiz do erro quadrático médio (RMSE) e o erro absoluto médio (MAE). Essas métricas são comparadas para determinar o desempenho

relativo de cada modelo. Com base nessa comparação, é selecionado o melhor modelo de aprendizado de máquina para a previsão de energia eólica.

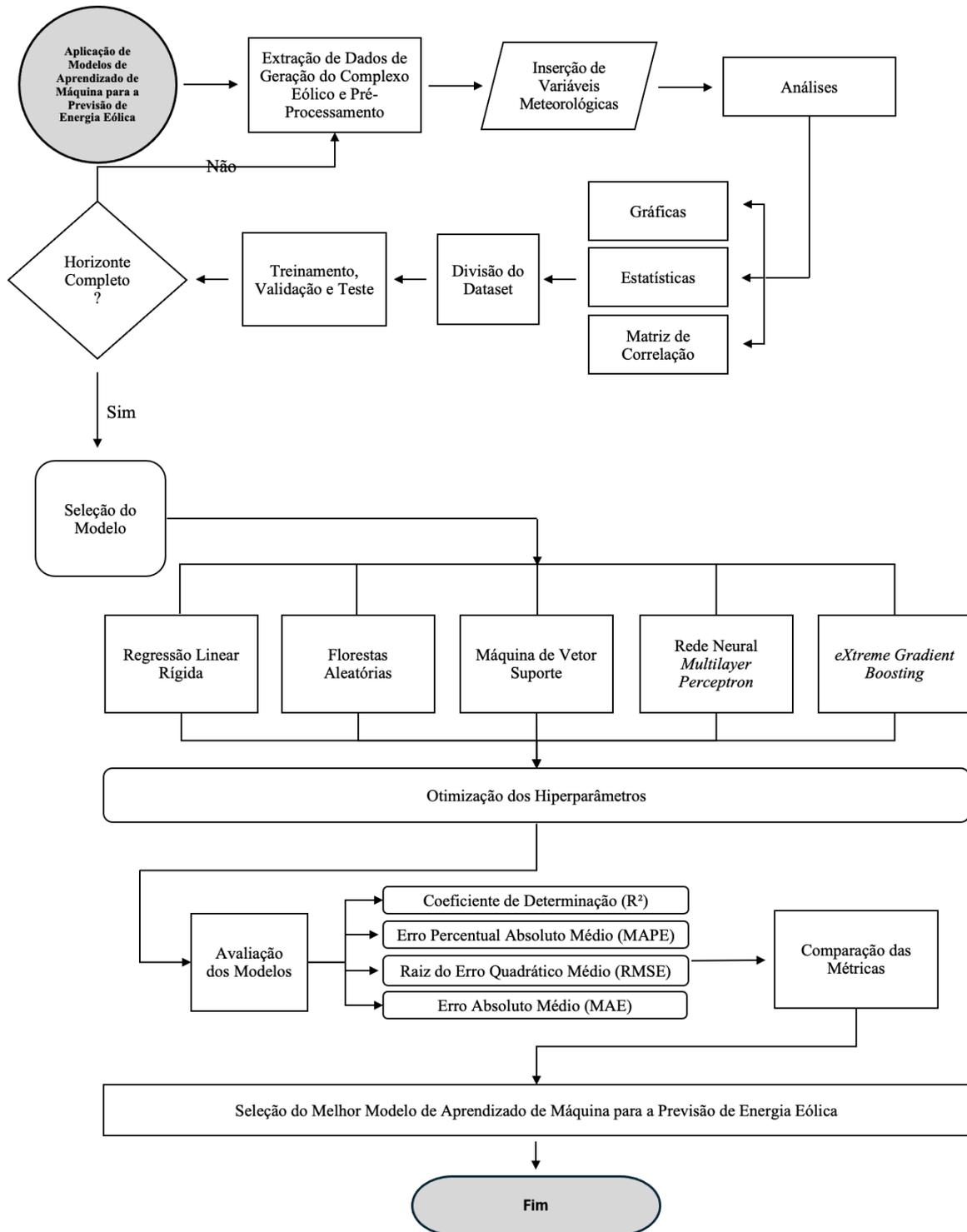


Figura 15 - Fluxograma do Algoritmo de Previsão de Longo Prazo

3.3.1 Regressão Linear Rígida

A Regressão Linear Rígida é implementada usando a classe *Ridge* do módulo *sklearn.linear_model* na biblioteca *Scikit-Learn*. A otimização considera o parâmetro de regularização (*alpha*) para melhorar o condicionamento do problema e reduzir a variância das estimativas, sendo definido como 10.0. O parâmetro *fit_intercept* especifica se uma constante (viés ou interceptação) deve ser adicionada à função de decisão, e neste caso, é definido como *True*. O parâmetro *normalize* é definido como *False*, que combinado com *fit_intercept*, garante a normalização antes da regressão, subtraindo a média e dividindo pela norma (*L2*).

O número máximo de iterações *max_iter*, é definido como *None*, o que significa que será utilizado o método de otimização por gradiente conjugado sem limite de iterações predefinido. O parâmetro *tol*, referente à precisão da solução, é definido como 0.001. O parâmetro *solver* é configurado como *auto*, permitindo que o algoritmo escolha automaticamente a melhor rotina computacional com base no conjunto de dados. Por fim, o parâmetro *random_state* é definido como *None*. Este parâmetro é utilizado para definir a semente do gerador de números aleatórios, mas, neste caso, só seria configurado se o *solver* fosse diferente de *auto*.

Os valores testados para cada um dos hiperparâmetros são ilustrados na Tabela 13, juntamente com seu resultado ótimo. A validação cruzada testou 1.440 combinações diferentes de hiperparâmetros (para cada parque eólico), levando 4 minutos.

Tabela 13 – Validação Cruzada dos Hiperparâmetros - Regressão Linear Rígida

Hiperparâmetros	Valores Testados	Resultado Ótimo
<i>alpha</i>	[0.1, 1, 10, 100]	10
<i>fit_intercept</i>	[True, False]	True
<i>normalize</i>	[True, False]	False
<i>max_iter</i>	[100, 200, 500, None]	None
<i>tol</i>	[0.0001, 0.001, 0.01, 0.1]	0.001

<i>solver</i>	[auto, svd, cholesky, sparse_cg, lsqr, sag]	Auto
<i>random_state</i>	None	None

3.3.2 Florestas Aleatórias

A regressão com *Random Forest* é implementada usando a classe *RandomForestRegressor* do módulo *sklearn.ensemble* na biblioteca *Scikit-Learn*. O número de árvores na floresta, *n_estimators*, foi definido como 100. O critério para otimização, que mede a qualidade das divisões, é a *squared_error*, utilizado para minimizar o erro quadrático médio. A profundidade máxima de cada árvore, *max_depth*, está configurada como *None*, permitindo que as árvores cresçam até que todas as folhas sejam puras ou contenham menos do que o número mínimo de amostras exigido.

O parâmetro *min_samples_split*, que define o número mínimo de amostras necessárias para dividir um nó, é definido como 2, enquanto *min_samples_leaf*, que define o número mínimo de amostras necessárias para estar em uma folha, é configurado como 1. O *min_weight_fraction_leaf*, que representa a fração mínima ponderada da soma total dos pesos das amostras necessárias para estar em uma folha, é definido como 0. O número de características a serem consideradas ao procurar a melhor divisão, *max_features*, é configurado como *auto*, permitindo que o algoritmo escolha automaticamente o número ideal. O *max_leaf_nodes*, que define o número máximo de folhas, está configurado como *None*, permitindo um crescimento ilimitado. O parâmetro *min_impurity_decrease*, que controla a divisão dos nós com base na redução mínima da impureza, é definido como 0.

O parâmetro *bootstrap*, que especifica se as amostras são substituídas durante a construção das árvores, é configurado como *True*. O número de execuções em paralelo, *n_jobs*, é definido como -1, utilizando todos os processadores disponíveis. O *random_state* está definido como *None*, controlando a aleatoriedade do *bootstrapping* e garantindo a reprodutibilidade dos resultados. O parâmetro *verbose* é configurado como 0, desativando a saída de *logs* detalhados. Por fim, o *max_samples*, que define o número máximo de amostras

a serem extraídas para treinar cada árvore, é configurado como *None*, utilizando todas as amostras disponíveis.

Os valores testados para cada um dos hiperparâmetros são ilustrados na Tabela 14, juntamente com seu resultado ótimo. A validação cruzada, utilizando a classe *GridSearchCV* do *Scikit-Learn*, testou 186.624 combinações diferentes de hiperparâmetros (para cada parque eólico), levando 34 minutos.

Tabela 14 – Validação Cruzada dos Hiperparâmetros – Florestas Aleatórias

Hiperparâmetros	Valores Testados	Resultado Ótimo
<i>n_estimators</i>	[100, 600, 700, 1.000]	700
<i>critério</i>	[squared_error, absolute_error]	Squared_error
<i>max_depth</i>	[None, 5]	None
<i>min_samples_split</i>	[1, 2, 3, 5]	2
<i>min_samples_leaf</i>	[1, 3, 5]	1
<i>min_weight_fraction_leaf</i>	[0, 0.01, 0.1]	0
<i>max_features</i>	[Auto, sqrt, 0.1]	Auto
<i>max_leaf_nodes</i>	[None, 20, 50]	None
<i>min_impurity_decrease</i>	[0, 0.001, 0.01]	0
<i>bootstrap</i>	[True, False]	True
<i>n_jobs</i>	[-1]	-1
<i>random_state</i>	[None, 42, 123]	None
<i>verbose</i>	[0]	0
<i>max_samples</i>	[None, 0.5]	None

3.3.3 Regressão por Vetores Suporte

A Regressão por Vetores Suporte foi implementada usando a classe SVR do módulo *sklearn.svm* na biblioteca *Scikit-Learn*. O hiperparâmetro C foi definido como 1.000, que controla a penalidade por erros e regulariza o modelo, determinando o *trade-off* entre um limite de decisão suave e a correta previsão dos exemplos de treinamento. O $coef0$ é um coeficiente independente usado em *kernels* polinomiais e sigmoide, foi definido como 0.1.

O grau do polinômio $degree$, foi configurado como 2, indicando que um *kernel* polinomial de segunda ordem será utilizado. O parâmetro $epsilon$, que especifica a largura do intervalo onde erros não são penalizados na função de perda, está configurado como 0.01. O parâmetro $gamma$ está definido como *scale*, o que significa que seu valor é calculado automaticamente pelo algoritmo com base nas características dos dados, ajustando a influência de cada ponto de treinamento.

O *kernel* utilizado foi configurado como *poly*, indicando o uso de um *kernel* polinomial. O parâmetro $shrinking$ foi definido como *True*, que acelera o processo de otimização do modelo, tornando o treinamento mais rápido.

Os valores testados para cada um dos hiperparâmetros são ilustrados na Tabela 15, juntamente com seu resultado ótimo. A validação cruzada, utilizando a classe *GridSearchCV* do *Scikit-Learn*, testou 35.000 combinações diferentes de hiperparâmetros (para cada parque eólico), levando 29 minutos.

Tabela 15 - Validação Cruzada dos Hiperparâmetros – Regressão por Vetores Suporte

Hiperparâmetros	Valores Testados	Resultado Ótimo
C	[0.1, 1, 10, 100, 300, 500, 1000]	1.000
$coef0$	[0.0, 0.1, 0.5, 0.7, 1.0]	0.1
$degree$	[0, 1, 2, 3, 4]	2
$epsilon$	[0.001, 0.01, 0.1, 0.2, 0.3]	0.01
$gamma$	[<i>scale</i> , <i>auto</i> , 0.1, 0.01, 0.001]	<i>scale</i>

<i>kernel</i>	[<i>rbf</i> , <i>linear</i> , <i>poly</i> , <i>sigmoide</i>]	<i>poly</i>
<i>shrinking</i>	[<i>True</i> , <i>False</i>]	<i>True</i>

3.3.4 Perceptron de Múltiplas Camadas

A regressão com Perceptron de Múltiplas Camadas foi implementada utilizando a biblioteca *Scikit-Learn* e *TensorFlow Keras* para construção, compilação e treinamento do modelo de rede neural.

A camada de entrada é definida pela dimensão dos dados de entrada, que neste caso é composta por oito características, tais como a geração histórica, temperatura a 2 metros, umidade Relativa a 2 metros, precipitação, pressão na superfície, velocidade do vento a 10 metros, velocidade do vento a 100 metros e rajada de vento a 10 metros. O modelo possui uma camada oculta com 128 neurônios, utilizando a função de ativação *Rectified Linear Unit* (ReLU). Após a camada oculta, uma camada de *dropout* com uma taxa de 0.5. A camada de saída do modelo tem um único neurônio, pois o objetivo é prever um valor contínuo (regressão), utilizando a função de ativação linear, adequada para tarefas de regressão.

A compilação é configurada com o otimizador *Adaptive Moment Estimation* (Adam) com uma taxa de aprendizado de 0.001. A função de perda utilizada é o erro quadrático médio. Durante o treinamento, o tamanho do lote (*batch size*) foi definido como 32, determinando o número de amostras propagadas pela rede antes de atualizar os parâmetros. O parâmetro *early stopping* foi implementado para interromper o treinamento se o desempenho no conjunto de validação não melhorasse após 1.000 épocas consecutivas, monitorando a perda de validação (*val_loss*).

Os valores testados para cada um dos hiperparâmetros são ilustrados na Tabela 16, juntamente com seu resultado ótimo. A validação cruzada, utilizando a classe *GridSearchCV* do *Scikit-Learn*, testou 11.520 combinações diferentes de hiperparâmetros (para cada parque eólico), levando 55 minutos.

Tabela 16 - Validação Cruzada dos Hiperparâmetros - Perceptron de Múltiplas Camadas

Hiperparâmetros	Valores Testados	Resultado Ótimo
Neurônios na Camada de entrada	[8]	8
Neurônios na Camada oculta	[5, 30, 64, 128]	128
Neurônios na Camada de saída	[1]	1
Função de ativação	[ReLU, tangente hiperbólica, sigmoide]	ReLU
Tamanho do lote	[32, 64, 128]	32
Coefficiente L2 de regularização	[0.001, 0.01, 0.1, 1]	0.01
Otimizador	[AdaGrad, Adam, RMSprop, SGD]	Adam
Taxa de aprendizado	[0.0001, 0.001, 0.01, 0.1]	0.001
Taxa de dropout	[0, 0.3, 0.4, 0.5, 0.6]	0.5
Early Stopping	1.000 épocas sem redução de erro de 0.01	1.000

3.3.5 eXtreme Gradient Boosting

O núcleo do modelo de previsão do XGBoost é baseado na classe *ForecasterAutoreg* da biblioteca *Skforecast*, para realizar previsões autorregressivas. Adicionalmente, o modelo utiliza o *XGBRegressor* da *Scikit-Learn*, inicializado com *random_state* definido como 123 para garantir a reprodutibilidade dos resultados. Este modelo é configurado com 36 defasagens, indicando a consideração de 36 valores passados na série temporal para prever a geração eólica futura. A otimização do modelo é realizada através de *grid search*, com o hiperparâmetro *n_estimators* definido como 1000, o que implica um conjunto de 1000

árvores para mitigar o *overfitting* enquanto aumenta a precisão do modelo. O parâmetro *max_depth* foi configurado no valor 3, para evitar que o modelo se torne excessivamente complexo. A taxa de aprendizado ou *learning_rate* foi fixada em 0,01, equilibrando a velocidade de convergência e o risco de ultrapassar as soluções ótimas.

Os parâmetros *colsample_bytree* e *subsample* são definidos como 0,05 e 1, respectivamente, controlando a fração de características e amostras utilizadas, o que afeta tanto o viés quanto a variância do modelo. O *min_child_weight* é definido como 1, estabelecendo a soma mínima dos pesos das observações exigidas para uma divisão; e *gamma* é configurado como 0, especificando a redução mínima de perdas necessária para realizar uma divisão, ajustável conforme a função de perda. Para reduzir o *overfitting*, o termo de regularização *L2* (*lambda*) é definido como 1, enquanto o termo de regularização *L1* (*alpha*) é configurado como 0, apropriado para cenários de alta dimensionalidade. O modelo também utiliza todos os núcleos de CPU disponíveis para computação paralela, através do parâmetro *n_jobs* configurado como -1, aumentando a eficiência considerando os dados extensos e cálculos complexos envolvidos. O parâmetro *scale_pos_weight* é definido como 1, útil em casos de desequilíbrio de classe alta para uma convergência mais rápida.

A exploração das defasagens é realizada através do parâmetro *lags_grid*, testando combinações de [1, 2, 3, 20], com o objetivo de identificar a estrutura temporal que melhor capta a dinâmica da geração eólica no complexo do Pindaí. Isso é importante para entender como os valores passados influenciam as previsões futuras e para capturar padrões sazonais ou outras dependências temporais. Finalmente, o uso de *bootstrapping* com 500 iterações auxilia na estimativa dos intervalos de confiança em torno dos valores previstos, oferecendo uma medida da incerteza preditiva do modelo.

Os valores testados para cada um dos hiperparâmetros são ilustrados na Tabela 17, juntamente com seu resultado ótimo. A validação cruzada, a partir da função *grid_search_forecaster* do *skforecast.model_selection*, testou 451.584 combinações diferentes de hiperparâmetros (para cada parque eólico), sendo as configurações mencionadas acima identificadas como os melhores resultados para o modelo XGBoost, minimizando o *overfitting* e maximizando a precisão. O processo de validação cruzada levou 1 hora e 33 minutos, rodando em um MacBook Air chip M1, CPU de 8 núcleos (4 de desempenho e 4 de eficiência), GPU de 7 núcleos e *Neural Engine* de 16 núcleos.

Tabela 17 - Configuração dos Hiperparâmetros – *eXtreme Gradient Boosting*

Hiperparâmetros	Valores Testados	Resultado Ótimo
<i>n_estimators</i>	[100, 500, 600, 700, 1.000, 1.300, 1.500]	1.000
<i>max_depth</i>	[1, 3, 5, 7]	3
<i>learning_rate</i>	[0.001, 0.005, 0.01, 0.02, 0.03]	0.01
<i>colsample_bytree</i>	[0.01, 0.05, 0.1, 0.4]	0.05
<i>subsample</i>	[0.2, 0.5, 1]	1
<i>min_child_weight</i>	[1]	1
<i>gamma</i>	[0]	0
<i>reg_lambda</i>	[0, 1]	1
<i>reg_alpha</i>	[0, 1]	0
<i>n_jobs</i>	[1]	-1
<i>lags_grid</i>	[1, 2, 3, 5, 10, 20]	[1, 2, 3, 20]
<i>refit</i>	[<i>True</i> , <i>False</i>]	<i>True</i>
<i>random_state</i>	[42, 123]	123
<i>lags</i>	[12, 24, 36]	36

3.4 Critérios de Avaliação dos Resultados

A avaliação do desempenho do modelo corresponde a métricas de acurácia e aplicabilidade dos modelos preditivos. Nesse sentido, foram utilizados quatro métricas para avaliar a qualidade dos modelos de regressão: o Coeficiente de Determinação (R^2), o Erro Percentual Absoluto Médio (MAPE), a Raiz do Erro Quadrático Médio (RMSE) e o Erro Absoluto Médio (MAE) [14].

O coeficiente de determinação (R^2) mede estatisticamente o grau em que as variações na variável dependente podem ser atribuídas às variações nas variáveis independentes [14]. O coeficiente é calculado conforme a expressão (27):

$$R^2 = 1 - \frac{\text{Soma dos quadrados de resíduos (SSR)}}{\text{Somar total de quadrados (SST)}} \quad (27)$$

onde a soma total dos quadrados (*Total Sum of Squares* - SST) mede a variância total dos dados observados e é calculada como a soma dos quadrados das diferenças entre cada valor observado e a média geral dos dados. A soma de quadrados de resíduos (*Sum of Squares of Residuals* - SSR) é a soma de erros quadrados de previsão, medindo a variância nos e aplicado como a soma dos quadrados das diferenças entre cada valor observado e seu valor previsto. Dessa forma, um coeficiente alto (próximo a 1) indica que uma maior proporção da variância é capturada pelo modelo com uma forte explicabilidade, enquanto valores baixos de coeficiente significam que o modelo não tem a capacidade explicativa de variabilidade dos dados de resposta em torno de sua média [14].

Existe uma variação da métrica, sendo chamada de R^2 ajustado, modificando a fórmula original para dar conta do número de preditores no modelo, fornecendo uma medida mais equilibrada que penaliza o uso excessivo de preditores irrelevantes. A expressão é apresentada em (28).

$$R^2 \text{ Ajustado} = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1} \quad (28)$$

onde n é o número de observações e k é o número de preditores.

O erro percentual absoluto médio (MAPE) é uma medida estatística que consiste em avaliar a precisão em contextos que a comparação da magnitude dos erros como porcentagens, é significativa [14]. É quantificada a precisão a partir do cálculo da média dos erros percentuais absolutos pelos quais as previsões se desviam dos valores reais por meio da expressão (29):

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (29)$$

onde y_i é o valor real, \hat{y}_i representa o valor previsto e n é o número de observações.

A métrica do MAPE é expressa como uma porcentagem, de maneira que um valor de, por exemplo, 3%, significa que o erro de previsão está dentro de 3% dos valores reais, sendo fácil sua interpretação para avaliar a qualidade da previsão. Um problema identificado é que o MAPE trata de superestimativas e subestimações de forma assimétrica, na qual os maiores valores reais terão um erro percentual menor em comparação com valores reais menores para o mesmo erro absoluto, distorcendo a avaliação ao lidar com dados que variam significativamente em magnitude [14].

A Raiz do Erro Quadrático Médio (RMSE) mede a magnitude das previsões de um modelo em relação aos valores observados, proporcionando uma compreensão intuitiva da precisão do modelo. A métrica é calculada em três etapas, sendo a primeira referente ao cálculo da diferença dos erros, de maneira que para cada ponto de dados, calcula-se a diferença entre o valor observado y_i e o valor predito pelo modelo \hat{y}_i . Esta diferença é chamada de erro residual. Em seguida, cada erro residual é elevado ao quadrado, para evitar que os erros positivos e negativos se anulem, e para penalizar maiores discrepâncias mais severamente [14]. A média desses erros quadráticos é então calculada, e finalmente, tira-se a raiz quadrada dessa média, resultando no RMSE, conforme fórmula apresentada em (30):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (30)$$

O Erro Absoluto Médio (MAE) é uma métrica estatística amplamente utilizada no contexto da modelagem preditiva para avaliar a precisão de modelos de previsão. Essa métrica mensura a média dos erros absolutos entre as previsões e os valores observados empiricamente [14]. A expressão matemática do MAE é dada por (31):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (31)$$

onde n representa o número total de observações, y_i é o valor real observado e \hat{y}_i é o valor previsto. A métrica mede a magnitude média dos erros sem levar em conta a sua direção, ou seja, trata de forma igual tanto os erros positivos quanto os negativos.

3.5 Considerações Gerais

Este capítulo apresentou todas as etapas do desenvolvimento dos modelos de aprendizado de máquina para a previsão de geração de energia eólica para o ano de 2023 do Complexo Eólico de Pindaí. Inicialmente, foi descrito o complexo em estudo, destacando as especificidades do local, as condições climáticas predominantes e as características operacionais que o tornam adequado para este tipo de análise preditiva. A compreensão dessas particularidades foi utilizada como critério para a escolha e configuração dos modelos de previsão, assegurando que as técnicas aplicadas estivessem alinhadas com as dinâmicas atmosféricas e operacionais do parque eólico. Em seguida, foi discutido o conjunto de dados utilizado no estudo, composto tanto por dados históricos de geração de energia quanto por variáveis meteorológicas derivadas de modelos de reanálise de alta resolução. A inclusão dessas variáveis de previsão meteorológica, com alto detalhamento espacial e temporal, garantiu que as previsões capturassem os padrões médios de vento e as flutuações de curto e médio prazo.

Na parte dedicada aos modelos, foram implementadas e testadas cinco abordagens distintas, tais como Regressão Linear Rígida, Florestas Aleatórias, Regressão por Vetores Suporte, Perceptron de Múltiplas Camadas e *eXtreme Gradient Boosting*. Um aspecto diferencial deste trabalho foi o teste exaustivo de hiperparâmetros para cada um desses modelos, utilizando técnicas de validação cruzada para encontrar as melhores combinações de parâmetros para cada técnica. Este processo consiste na contribuição para a lacuna de pesquisas existentes, através da exploração a fundo da diversidade de combinações de hiperparâmetros em diferentes modelos.

No caso da Regressão Linear Rígida, o parâmetro fundamental a ser ajustado é o coeficiente de regularização (α). Este coeficiente controla o grau de penalização imposto sobre os coeficientes do modelo, influenciando diretamente o *trade-off* entre viés e variância. O ajuste neste estudo foi tratado como um problema de otimização, de maneira que valores baixos desse parâmetro podem gerar um modelo com coeficientes grandes e, conseqüentemente, suscetível ao *overfitting*, pois o modelo se ajustaria excessivamente aos dados de treinamento, capturando ruídos. Por outro lado, valores muito altos restringem em

demasia os coeficientes, tornando o modelo incapaz de capturar as variações necessárias para previsões acuradas. Para determinar o valor ideal, foi empregada a validação cruzada em um *grid* de valores que variaram de 10^{-1} a 10^2 . Além disso, a escolha de um modelo de regularização L_2 (regressão rígida) foi feita com base na literatura, onde essa regularização considerada em problemas de multicolinearidade, como os enfrentados ao lidar com múltiplas variáveis meteorológicas.

Para o modelo de Florestas Aleatórias, a análise focou-se em três hiperparâmetros principais: o número de árvores na floresta, a profundidade máxima das árvores e o número mínimo de amostras por folha. O número de árvores define quantas árvores de decisão independentes serão geradas. A literatura mostra que um aumento no número de árvores melhora a estabilidade do modelo ao reduzir o erro de generalização, pois a média das previsões das árvores tende a ser mais robusta. No entanto, mais árvores também aumentam o custo computacional. Portanto, o valor ideal de árvores foi escolhido testando valores entre 100 e 1000, com a observação de que, após certo ponto (geralmente em torno de 500), o ganho marginal de performance foi pouco significativo. A profundidade máxima das árvores, por sua vez, influencia diretamente a complexidade do modelo. Árvores muito profundas tendem a criar regras muito específicas que não generalizam bem para dados novos. A profundidade ótima foi determinada testando várias configurações, com a validação cruzada indicando que uma profundidade moderada, em torno de 5 níveis, era suficiente para capturar as complexidades sem *overfitting*.

No caso da Regressão por Vetores Suporte, o objetivo é encontrar um hiperplano que minimize o erro em uma margem de tolerância (ϵ), com dois parâmetros principais a serem ajustados: o parâmetro de regularização (C) e a largura da margem (ϵ). O parâmetro de regularização controla o compromisso entre maximizar a margem e minimizar o erro de previsão para os pontos que estão fora da margem. Um valor alto permite que o modelo seja mais rígido, forçando a minimizar os erros ao custo de menor margem, o que pode gerar *overfitting*. Valores baixos, por outro lado, aumentam a margem, mas podem resultar em previsões menos precisas. Foram uma faixa ampla, de 10^{-1} a 10^3 , com a validação cruzada revelando que valores altos em torno de 10^3 produzem o melhor equilíbrio. Em relação a margem de tolerância, sua escolha é fundamental para determinar a zona de insensibilidade

ao erro. Foram testados valores entre 10^{-3} a 3×10^3 , encontrando que valores próximos a 10^{-2} resultavam na melhor performance.

O modelo de Perceptron de Múltiplas Camadas, uma forma de rede neural *feedforward*, exige um ajuste fino de diversos hiperparâmetros devido à sua natureza altamente não linear e flexível. Os hiperparâmetros principais incluem o número de camadas ocultas, o número de neurônios por camada, a taxa de aprendizado e a função de ativação. O número de camadas e neurônios determinam a capacidade da rede de aprender representações complexas. Camadas adicionais podem permitir ao modelo capturar interações mais profundas entre as variáveis, mas aumentam o risco de *overfitting*. Foi implementada 1 camada oculta, e neurônios variando de 5 a 128 nesta camada. Os melhores resultados foram obtidos com aproximadamente 128 neurônios. A taxa de aprendizado controla o tamanho dos passos dados na otimização dos pesos da rede. Uma taxa muito alta pode levar à oscilação do modelo em torno de mínimos locais, enquanto uma taxa muito baixa pode tornar o processo de treinamento ineficientemente lento. A taxa de aprendizado ideal foi encontrada através da otimização em uma escala logarítmica, variando de 10^{-4} a 10^{-1} , com o melhor desempenho observado em torno de 10^{-3} . Entre as funções de ativação testadas, a função ReLU se destacou, proporcionando a convergência mais rápida e estabilidade durante o treinamento, em comparação com a função sigmoide ou tangente hiperbólica.

Finalmente, no modelo de *eXtreme Gradient Boosting*, o número de árvores foi otimizado em uma faixa entre 100 e 1.500 árvores, com o uso de *boosting* permitindo que o XGBoost faça ajustes finos. Foi verificado que aproximadamente 1.000 árvores ofereciam o equilíbrio suficiente entre precisão e complexidade computacional. A taxa de aprendizado controla a magnitude das atualizações dos coeficientes ao longo das iterações, taxas altas levam a oscilações e falta de convergência, enquanto taxas muito baixas podem exigir um número excessivo de iterações. A melhor performance foi encontrada com taxas entre 10^{-2} .

Adicionalmente, a avaliação de desempenho dos modelos foi realizada com base em critérios de erro que permitiram a análise quantitativa da precisão de cada técnica. Essas métricas serviram para estabelecer comparações com trabalhos semelhantes e validar os resultados obtidos, assegurando que os modelos desenvolvidos atingiram níveis satisfatórios de precisão preditiva.

Capítulo 4 - Resultados

Os resultados obtidos deste trabalho foram divididos em quatro níveis, exemplificadas nas seções deste capítulo. Os resultados incluem as análises das séries temporais, análises das etapas de treinamento dos modelos, previsão da geração eólica dos quatro parques referentes ao Complexo de Pindaí para o ano de 2023 e a discussão referente aos resultados obtidos.

4.1 Resultados da Análise da Série Temporal

Conforme abordado na seção 4.1 referente ao conjunto de dados da série temporal, o levantamento da geração histórica do complexo eólico foi extraído em uma escala temporal de 10 anos, de julho de 2014 até dezembro de 2023. A série temporal bruta foi coletada em base horária e totalizando 228 mil observações, sendo 57 mil horas para cada parque eólico. Após o tratamento e normalização dos dados, utilizando a transformação para base mensal, a série temporal de cada parque foi convertida para 114 meses de dados, conforme mostrado na Figura 16. As características estatísticas são apresentadas na Tabela 18.

Além disso, a Figura 16 destaca os padrões sazonais que influenciam a geração de energia eólica, os quais estão intrinsecamente ligados às variações climáticas ao longo do ano. Durante o período úmido, que se estende de novembro a abril, observa-se um aumento significativo na precipitação, fenômeno este associado ao deslocamento para o sul da Zona de Convergência do Atlântico Sul. Esse deslocamento provoca uma alteração nas condições atmosféricas, resultando em uma diminuição das velocidades dos ventos na região. Essa correlação é evidenciada na coloração azulada do gráfico, que marca os meses de maior precipitação. A diminuição da intensidade dos ventos durante este período contribui para uma menor eficiência das turbinas eólicas, refletindo-se, conseqüentemente, em menores índices de geração de energia.

Em contrapartida, o período seco, que ocorre de maio a outubro, é caracterizado pelo recuo da Zona de Convergência para o norte, o que modifica substancialmente o regime de ventos. Durante esses meses, destacados em vermelho no gráfico, há um aumento

significativo nas velocidades dos ventos, favorecendo a operação das turbinas eólicas. Essa alteração sazonal resulta em uma maior eficiência na conversão da energia eólica em energia elétrica, elevando os níveis de geração durante este período. Tal comportamento sazonal é de extrema relevância para o planejamento e a otimização da operação dos parques eólicos, uma vez que permite antecipar variações na produção de energia com base nos padrões climáticos.

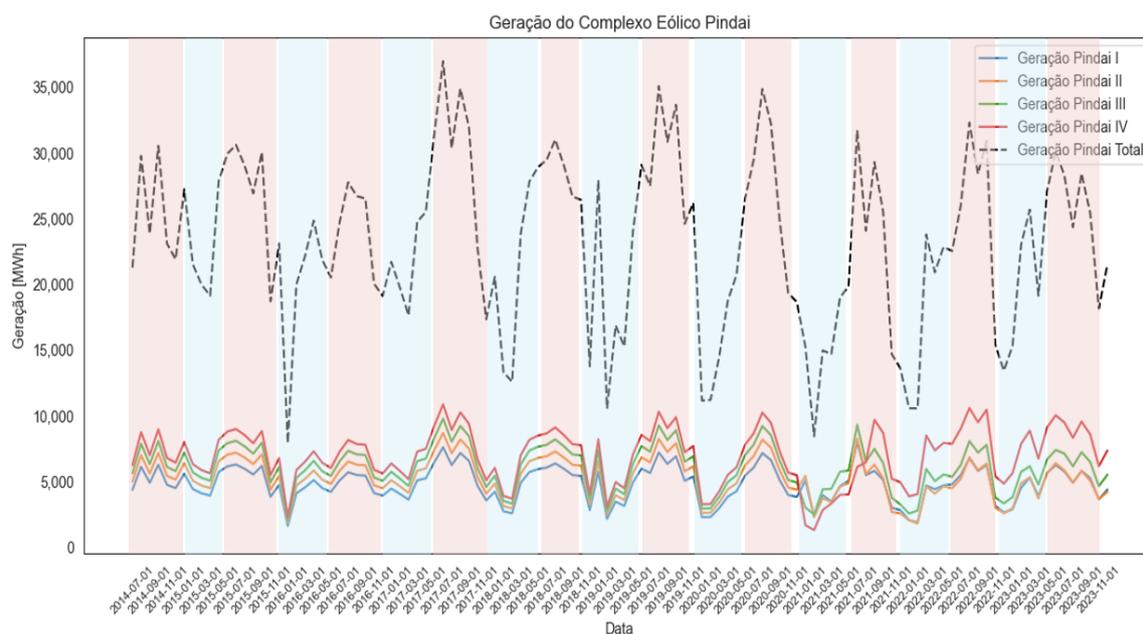


Figura 16 - Geração Mensal Histórica do Complexo Pindaí [MWh]

Tabela 18 - Estatística da Geração Mensal [MWh]

Parque Eólico	Geração Média [MWh]	Mediana [MWh]	Desvio Padrão [MWh]
Pindaí I	4.851,81	5.005,91	1.345,95
Pindaí II	5.429,75	5.392,45	1.573,94
Pindaí III	6.105,39	6.214,45	1.735,44
Pindaí IV	7.005,35	7.296,43	2.109,65

O Parque Eólico Pindaí I apresenta uma geração média de 4.851,81 MWh e uma mediana de 5.005,91 MWh, indicando que metade dos dados observados está abaixo desse

valor, sugerindo uma tendência central que se opõe aos *outliers*. O desvio padrão, em 1.345,95 MWh, aponta para um nível moderado de variabilidade na geração, possivelmente influenciado pela flutuação das condições de vento. Em comparação, o Parque Eólico Pindaí II tem uma geração média aumentada para 5.329,75 MWh, refletindo maior eficiência em condições de vento em relação ao Pindaí I. A mediana, ligeiramente superior à média, é de 5.392,45 MWh, indicando uma distribuição enviesada para valores mais altos. A variabilidade, indicada por um desvio padrão de 1.573,94 MWh, também é maior, sugerindo que o Pindaí II pode ser mais suscetível a variações ambientais que afetam a geração eólica.

O Parque Eólico Pindaí III aumenta ainda mais a potência média para 6.105,39 MWh, com uma mediana maior de 6.214,45 MWh, mostrando uma distorção positiva consistente na distribuição dos valores de geração. Este parque apresenta o maior desvio padrão entre os três primeiros, com 1.735,44 MWh, indicando oscilações significativas na geração de energia, possivelmente devido a velocidades de vento altamente variáveis. Por fim, o Parque Eólico Pindaí IV supera todos em termos de geração média, com 7.005,35 MWh, e possui a maior mediana, com 7.296,43 MWh. Essa distribuição inclinada sugere uma frequência considerável de valores de geração mais altos, provavelmente refletindo a capacidade do parque em aproveitar a energia eólica de forma otimizada. No entanto, também é registrado o maior desvio padrão de 2.109,65 MWh, resultado da dispersão nos dados e flutuações substanciais que podem ser atribuídas a fatores ambientais e técnicos.

Outro tipo de representação gráfica é através de *boxplots*, ilustrado na Figura 17. Esses gráficos apresentam a distribuição histórica da geração dos parques eólicos, revelando tendências sazonais e variabilidades na produção de energia alinhadas com os padrões climáticos regionais no nordeste do Brasil. Conforme abordado na seção 2.2 sobre as condições climatológicas da região da Bahia, a sazonalidade da geração eólica é nitidamente dividida entre o período úmido, de novembro a abril, e o período seco, de maio a outubro. Esse padrão é impulsionado pelas condições meteorológicas regionais influenciadas pela Zona de Convergência Intertropical e características topográficas locais.

Durante o período chuvoso, a região experimenta um aumento da precipitação e da cobertura de nuvens devido ao deslocamento para o sul da ZCIT. Essa mudança resulta na diminuição da velocidade do vento, pois as precipitações interrompem os gradientes de pressão atmosférica que impulsionam o fluxo do vento. Conseqüentemente, os parques

eólicos apresentam números de geração mais baixos durante esses meses, já que as velocidades reduzidas do vento limitam a capacidade das turbinas de operar na capacidade ideal, não atingindo as velocidades necessárias para converter energia de forma eficiente.

Por outro lado, o período seco é caracterizado pelo recuo da ZCIT para o norte, diminuindo a precipitação e a cobertura de nuvens, o que leva a céus mais claros e a um aquecimento solar mais pronunciado da superfície terrestre. Este aquecimento reforça os gradientes de pressão locais, aumentando o fluxo de vento em toda a região. Durante este período, as velocidades do vento excedem os limites operacionais da maioria das turbinas, atingindo níveis que maximizam a eficiência mecânica e operacional dos equipamentos.

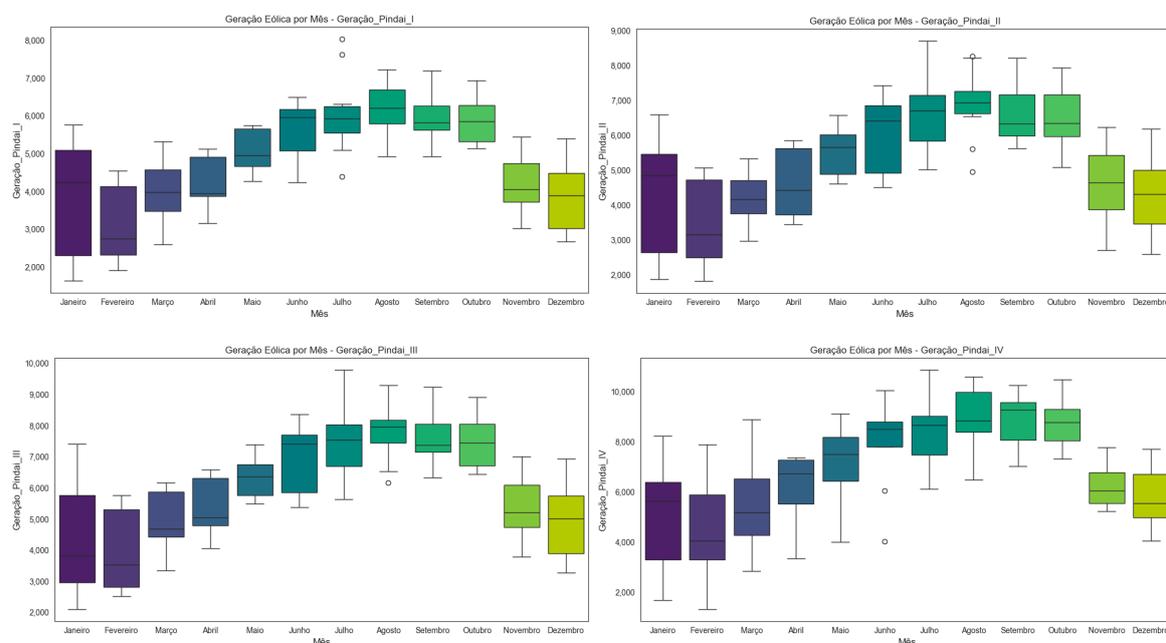


Figura 17 - Boxplots da Distribuição Histórica de Geração no Período de 2014 a 2023 em MWh. Pindaí I (Superior Esquerdo), Pindaí II (Superior Direito), Pindaí III (Inferior Esquerdo) e Pindaí IV (Inferior Direito)

Para o Parque Pindaí I, os dados indicam um aumento gradual na geração eólica mediana de maio a outubro, com um pico durante a estação seca de julho a agosto, atingindo uma geração mediana de 6.000 MWh no terceiro trimestre. Em contraste, observa-se um desempenho abaixo da média nos meses de novembro a abril, característico das condições climáticas úmidas do nordeste do Brasil, que prevalecem no primeiro semestre do ano.

De maneira semelhante, os Parques Pindaí II e III apresentam flutuações sazonais ainda mais pronunciadas. Por exemplo, Pindaí II inicia com uma geração média de 4.900 MWh em janeiro, escalando para um pico de 6.900 MWh em agosto. O *boxplot* de Pindaí III mostra um aumento considerável na mediana e na faixa de geração entre junho e agosto, com a mediana subindo de 7.000 MWh em junho para 7.900 MWh em agosto.

Pindaí IV exhibe as mudanças mais abruptas, com os valores médios de geração subindo acentuadamente para 8.400 MWh em agosto, acompanhados por uma expansão na faixa de valores, especialmente no quartil superior. Em todos os parques, os valores dos quartis mais baixos aumentam gradualmente durante a estação seca. Isso indica que, mesmo em dias com velocidades de vento relativamente baixas, a geração de energia não cai tanto quanto durante a estação chuvosa. Em outras palavras, durante a estação seca, mesmo os dias menos favoráveis para a geração de energia eólica ainda produzem mais energia do que os dias semelhantes na estação chuvosa. Esse impacto sazonal é confirmado pela queda na geração de novembro a março, coincidindo com a estação chuvosa, quando as velocidades do vento são geralmente mais baixas e inconsistentes.

A partir da análise de correlação entre as variáveis exógenas meteorológicas e a geração de energia eólica no complexo eólico de Pindaí, é possível entender como os fatores ambientais influenciam a produção de energia, apresentado na Figura 18. Entre as variáveis, a velocidade do vento a 100 metros apresenta a maior correlação positiva com a geração de energia nos quatro parques, com Pindaí I apresentando a correlação mais forte com 90,33%, seguido por Pindaí III com 89,95%, Pindaí II com 88,94% e Pindaí IV com 85,96%. Essa forte correlação exemplifica a importância majoritária que os ventos de alta altitude desempenham na maximização da produção de energia. Da mesma forma, velocidade do vento a 10 metros e rajadas de vento a 10 metros também apresentam correlações positivas significativas. Por exemplo, a velocidade do vento a 10 metros tem uma correlação de 87,93% em Pindaí III e porcentagens ligeiramente menores nos outros parques, indicando que mesmo em altitudes mais baixas a velocidade do vento continua sendo um determinante da eficácia da geração.

Por outro lado, são perceptíveis correlações negativas com temperatura, umidade relativa do ar e precipitação. Particularmente, a precipitação apresenta a correlação negativa mais forte, especialmente em Pindaí I, onde atinge -69,53%. Isso porque a precipitação inibe

a eficiência das turbinas, por meio de impactos diretos na mecânica da turbina ou efeitos indiretos, como mudanças na densidade do ar e nos padrões de vento. A umidade relativa do ar e a temperatura também apresentam correlações negativas em todos os parques, com a temperatura em Pindaí I apresentando o impacto negativo com -36,26%. Essa variável afeta negativamente a geração eólica devido à influência na densidade e pressão do ar, e consequentemente, afetam a velocidade e a estabilidade do vento. A pressão da superfície apresenta uma correlação neutra com a geração eólica, variando marginalmente entre os parques. Isso sugere que embora a pressão atmosférica impacte a formação e o comportamento do vento, sua influência direta na geração eólica pode ser menos pronunciada em comparação com outros fatores. As correlações de pressão superficial variam de 16,42% em Pindaí II a 21,69% em Pindaí IV.

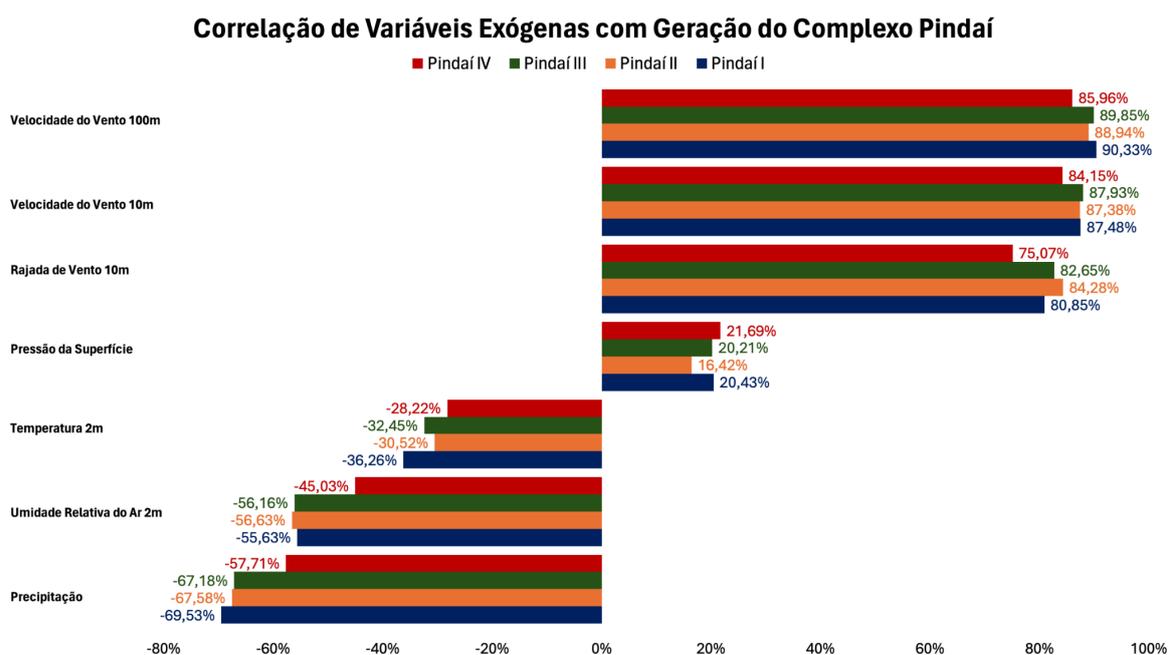


Figura 18 - Correlação de Variáveis Exógenas com a Geração do Complexo de Pindaí

De maneira semelhante, a distribuição dos pesos atribuídos às variáveis exógenas detalha as variáveis que exercem maior influência sobre os resultados, ilustrado na Figura 19. A variável com maior peso é a velocidade do vento a 100m (0.38), indicando que a performance do vento em alturas elevadas é determinante na geração de energia, seguida pela velocidade do vento a 10m (0.32). As rajadas de vento a 10m, precipitação, umidade

relativa do ar a 2m, temperatura a 2m e pressão da superfície, com menores pesos: 0.17, 0.11, 0.012, 0.007 e 0.001, respectivamente.

Pesos Atribuídos as Variáveis Exógenas

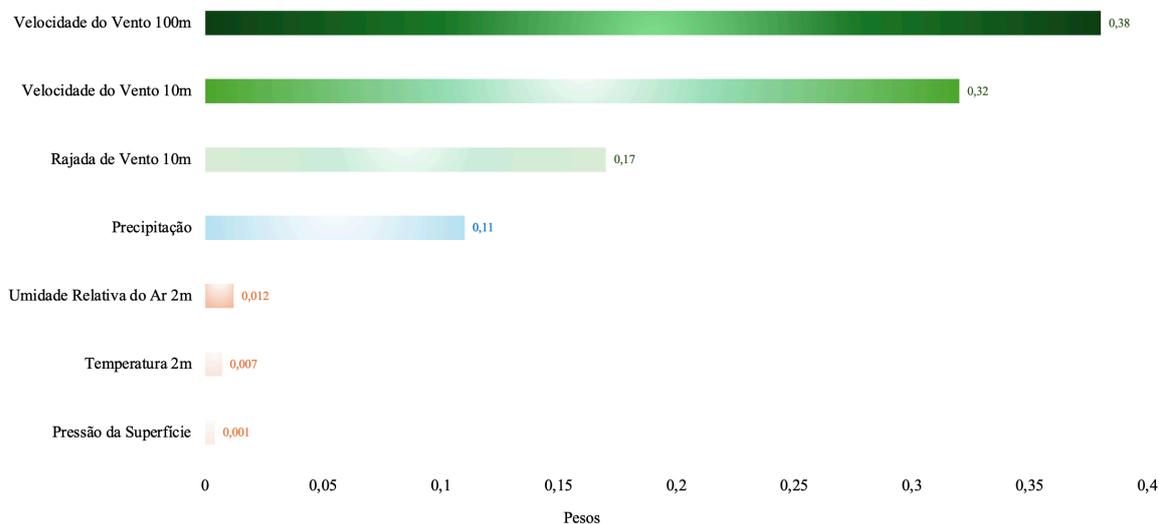


Figura 19 - Peso Atribuído às Variáveis Exógenas

A correlação entre a geração e a velocidade do vento a 100 metros também pode ser vista nos gráficos de dispersão, na Figura 20. Cada ponto representa um par observado de valores: velocidade do vento a 100 metros e a geração de energia correspondente. A distribuição dos pontos aborda a variabilidade e as tendências dentro dos dados, enquanto a reta de regressão ilustra o efeito médio da velocidade do vento na geração em toda a faixa observada. A inclinação positiva da reta corrobora a forte correlação positiva previamente quantificada, em relação à maior energia cinética disponível em massas de ar em movimento mais rápido, na qual as turbinas convertem de forma mais eficiente em energia elétrica.

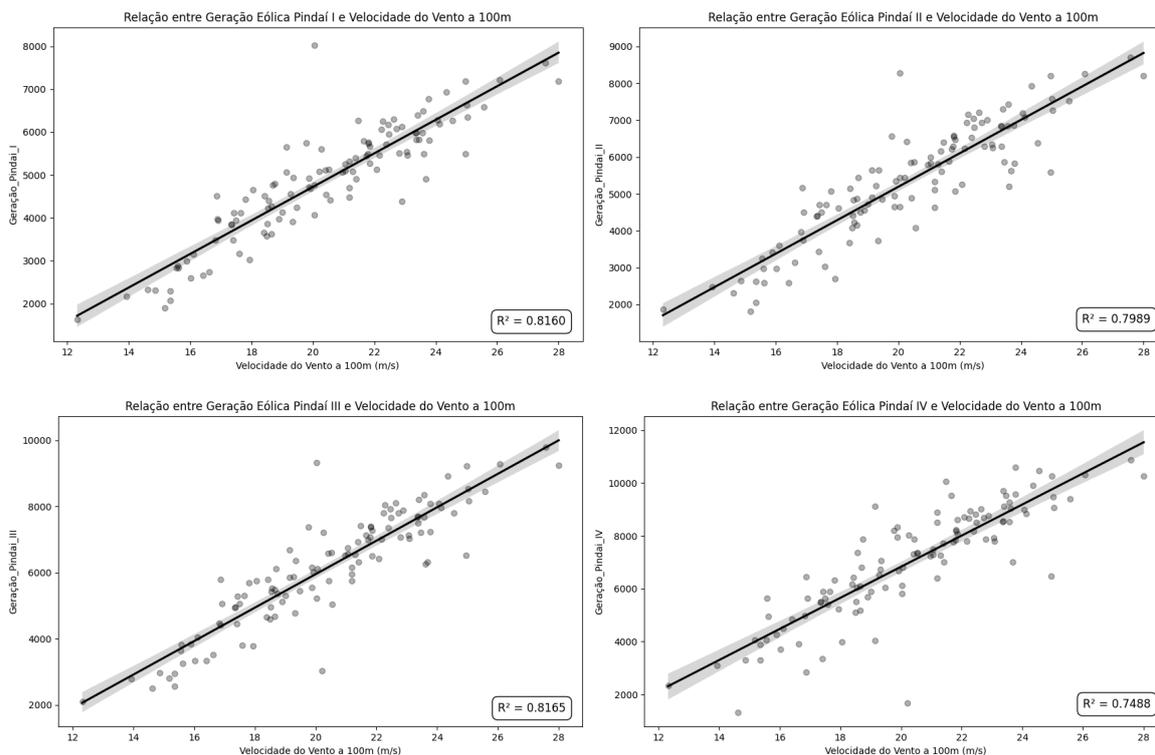


Figura 20 - Dispersão entre a Geração Histórica e a Velocidade do Vento a 100m. Pindaí I (Superior Esquerdo), Pindaí II (Superior Direito), Pindaí III (Inferior Esquerdo) e Pindaí IV (Inferior Direito)

O ajuste da linha de regressão, medido pelo coeficiente de determinação (R^2), revela a proporção da variância na geração de energia que pode ser explicada pelas variações na velocidade do vento, demonstrando a previsibilidade e a confiabilidade da velocidade do vento como um determinante da potência de saída. Os valores de R^2 para cada parque eólico indicam a proporção da variação na geração de energia que é previsível a partir da velocidade do vento a 100 metros.

Em Pindaí I, o R^2 de 0,8160 indica que aproximadamente 81,60% da variação na geração de energia pode ser explicada pela velocidade do vento, sendo um valor alto e refletindo uma forte relação linear. Da mesma forma, Pindaí III apresenta um R^2 de 0,8165, também indicando uma forte influência da velocidade do vento, semelhante à observada em Pindaí I. Por outro lado, Pindaí II e Pindaí IV apresentam valores de R^2 ligeiramente menores, sendo 0,7989 e 0,7488, respectivamente. Essa redução sugere que outros fatores,

como a eficiência das turbinas, características geográficas ou variáveis ambientais adicionais podendo impactar na influência da geração de energia nesses locais.

No contexto da estacionariedade em séries temporais, embora não seja necessário transformar a série temporal em estacionária para a aplicação dessas metodologias, o teste de Dickey-Fuller Augmented foi utilizado como uma medida confirmatória para avaliar a estacionariedade. O teste confirma a estacionariedade do parque Pindaí I com uma estatística de teste de -4,293, superando os valores críticos e acompanhada por um p-valor baixo em 0,000456. Isso indica uma forte rejeição da hipótese nula de uma raiz unitária, afirmando que a série temporal não exibe estruturas dependentes do tempo, como tendências ou padrões sazonais, o que poderia complicar a modelagem preditiva. Em contraste, os resultados iniciais para Pindaí II, III e IV revelam características não estacionárias, com suas séries originais não passando no teste de ADF. Por exemplo, Pindaí II apresentou inicialmente uma estatística de teste de -1,529 com um p-valor de 0,5189, indicando a presença de uma unidade de raiz. Padrões semelhantes foram observados em Pindaí III e IV, onde as estatísticas do teste inicial e os valores de p não atingiram o limiar para rejeitar a hipótese nula, caracterizando a influência de elementos não estacionários, como tendências ou flutuações cíclicas nos dados de geração eólica.

Portanto, foi aplicada a técnica de diferenciação de primeira ordem, de maneira que ao diferenciar essas séries uma vez, observa-se a exibição de estacionariedade pronunciada. Isso é evidenciado por estatísticas de testes negativos que estão abaixo dos valores críticos e valores de p próximos de zero. Especificamente, a série diferenciada de Pindaí II mostra uma estatística de teste de -7,392, a de Pindaí III de -7,847 e a de Pindaí IV de -7,502, todas indicando estacionariedade pós-diferenciação, conforme apresentado na Tabela 19.

Tabela 19 - Testes Augmented Dickey-Fuller para o Complexo Eólico Pindaí

Parque	Tipo do Teste	Estatística do Teste	p-value	Valores Críticos (1%, 5%, 10%)	Status de estacionariedade
<i>Pindaí I</i>	Original	-4.2933	0.000456	-3.4942, -2.8895, -2.5817	Estacionária
	Original	-1.5295	0.5189	-3.4961, -2.8903, -2.5821	Não Estacionária

<i>Pindaí II</i>	Diferenciação de primeira ordem	-7.3919	7.96e-11	-3.4961, -2.8903, -2.5821	Estacionária
<i>Pindaí III</i>	Original	-1.8443	0.3587	-3.4961, -2.8903, -2.5821	Não Estacionária
	Diferenciação de primeira ordem	-7.8471	5.73e-12	-3.4961, -2.8903, -2.5821	Estacionária
<i>Pindaí IV</i>	Original	-1.9270	0.3195	-3.4975, -2.8909, -2.5824	Não Estacionária
	Diferenciação de primeira ordem	-7.5022	4.23e-11	-3.4955, -2.8900, -2.5820	Estacionária

A série diferenciada, agora estacionária, é utilizada como entrada para o modelo de regressão linear rígida garantindo que as pressuposições de estacionariedade sejam atendidas, permitindo que o modelo capture de forma mais eficaz as relações lineares entre as variáveis preditoras e a variável dependente. Uma vez que o modelo de regressão foi treinado e fez as previsões usando a série diferenciada, o passo final é reverter a diferenciação para obter previsões na escala original da série temporal. Este processo de reversão é importante porque os valores diferenciados não são diretamente interpretáveis em termos do problema original. Somente após essa reversão as previsões podem ser comparadas diretamente aos dados históricos e interpretadas em seu contexto original.

A próxima etapa refere-se a à separação do conjunto de dados, conforme discutido na seção 4.1. O conjunto de treinamento abrange de julho de 2014 a janeiro de 2021, ilustrado na Figura 21. Este período foi selecionado para fornecer aos modelos de aprendizado de máquina um contexto histórico significativo, com variações sazonais, mudanças operacionais e diferentes anomalias de condições da atmosfera, como os períodos de El Niño e La Niña, afetando a condição de ventos pelas precipitações. O conjunto de validação abrange o período subsequente, de fevereiro de 2021 a dezembro de 2022, representando 20% dos dados. Esse segmento funciona como uma ponte entre a compreensão hipotética derivada dos dados de treinamento e a aplicação no mundo real avaliada durante a fase de teste, ajustando os hiperparâmetros dos modelos e avaliando o

desempenho do modelo para garantir a mitigação do *overfitting*. Os 10% restantes dos dados, de janeiro de 2023 a dezembro de 2023, fazem parte do conjunto de testes e serão utilizados para realizar previsões com as metodologias propostas.

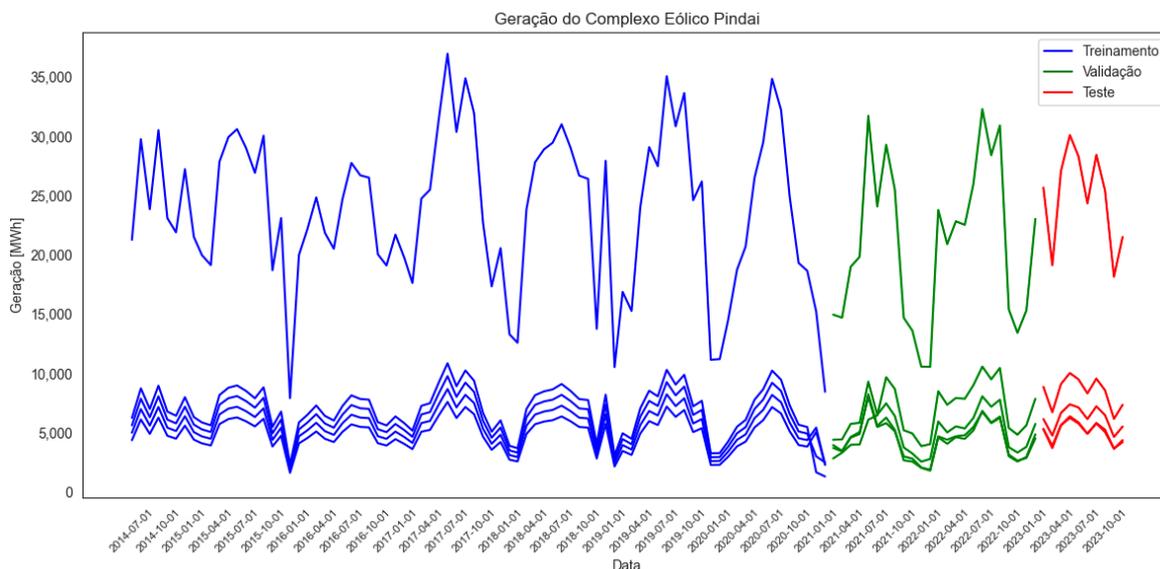


Figura 21 - Separação do Conjunto de Dados em Treinamento, Validação e Teste

4.2 Resultados da Etapa de Treinamento e Teste dos Modelos

As etapas de treinamento e teste das metodologias propostas foram analisadas graficamente para melhor compreender o desempenho dos modelos. Nos modelos de regressão linear rígida, florestas aleatórias e *eXtreme Gradient Boosting*, foram apresentadas as curvas de aprendizado e as curvas de validação. As curvas de aprendizado, ilustradas na Figura 22, mostram como o desempenho do modelo evolui à medida que o tamanho do conjunto de treinamento aumenta, utilizando duas métricas: o desempenho no conjunto de treinamento e o desempenho na validação cruzada. Isso permite identificar se o modelo sofre de viés alto (*underfitting*) ou variância alta (*overfitting*), conforme as curvas convergem para um valor alto e próximo uma da outra.

Para o XGBoost, a curva de aprendizado do parque Pindaí I mostra que o *score* de treinamento começa em 1.00 com nove exemplos e diminui gradualmente para 0.85 com 81

exemplos. Isso indica que o modelo inicialmente se ajusta bem aos dados de treinamento, mas torna-se mais complexo conforme mais exemplos são adicionados. Na validação cruzada, o *score* começa em 0.47 com nove exemplos e aumenta significativamente para 0.87 com 81 exemplos, sugerindo uma ótima capacidade de generalização para dados não vistos.

Na regressão linear rígida, o *score* de treinamento começa em 0.46 com nove exemplos e aumenta para 0.85 com 81 exemplos, enquanto na validação cruzada o *score* aumenta de 0.33 para 0.81, indicando uma capacidade de generalização.

Para o *Random Forest*, o *score* de treinamento é alto, começando em 0.81 e aumentando para 0.97 com 81 exemplos, mostrando um ajuste excessivo aos dados de treinamento, possivelmente causando *overfitting*. Na validação cruzada, o *score* aumenta de 0.28 para 0.79 à medida que mais exemplos são adicionados, perdendo capacidade de generalização conforme o modelo se ajusta demais aos dados de treinamento.

Os padrões observados para os parques eólicos Pindaí II e Pindaí III são semelhantes aos de Pindaí I em termos de comportamento das curvas de aprendizado para os diferentes modelos. No caso de Pindaí IV, o *score* de treinamento começa em 0.89 e aumenta para 0.91, enquanto na validação o *score* aumenta de 0.41 para 0.86. Isso indica um ajuste inicial eficaz aos dados de treinamento, mas com uma leve deterioração conforme mais exemplos são adicionados, mantendo uma boa capacidade de generalização na validação.

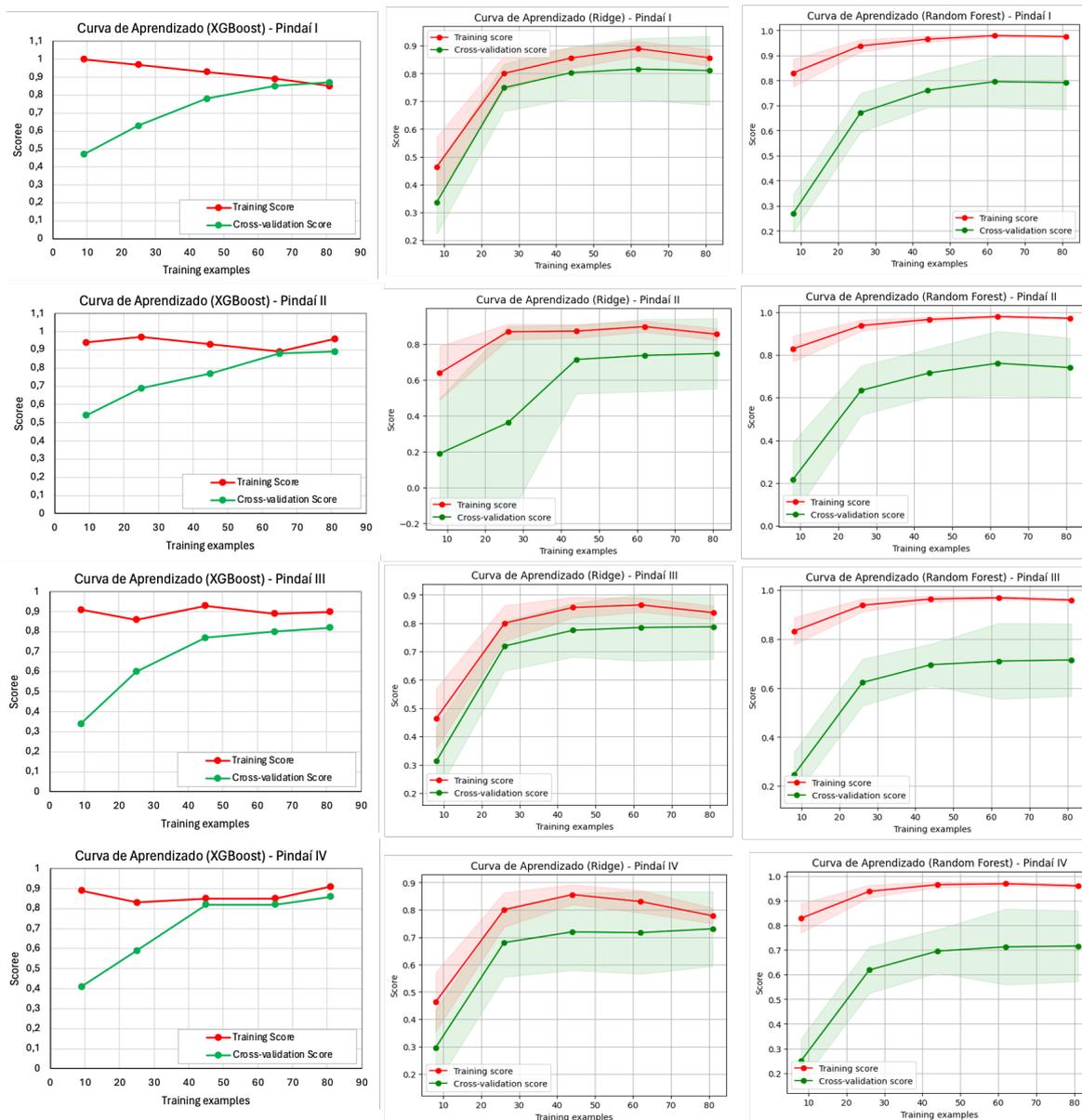


Figura 22 - Curvas de Aprendizagem Referente aos Modelos – XGBoost, Regressão Rígida e Florestas Aleatórias

As curvas de validação apresentadas na Figura 23 são usadas para avaliar o desempenho do modelo em relação a diferentes hiperparâmetros ou configurações. Por exemplo, ao ajustar a profundidade de uma árvore de decisão ou o parâmetro de regularização em uma regressão linear, essas curvas fornecem uma visão sobre o desempenho para diferentes valores dos hiperparâmetros. Semelhante às curvas de

aprendizado, o eixo x representa os valores dos hiperparâmetros ajustados, enquanto o eixo y representa a métrica de desempenho.

Para o XGBoost, observa-se que uma taxa de aprendizado em torno de 0.01 a 0.02 resulta em valores ideais da métrica R^2 para todos os parques do complexo. Na regressão linear rígida os resultados indicam que a variação do hiperparâmetro α não tem um impacto significativo no desempenho do modelo. No caso das florestas aleatórias, não há um aumento significativo no desempenho à medida que o número de estimadores aumenta.

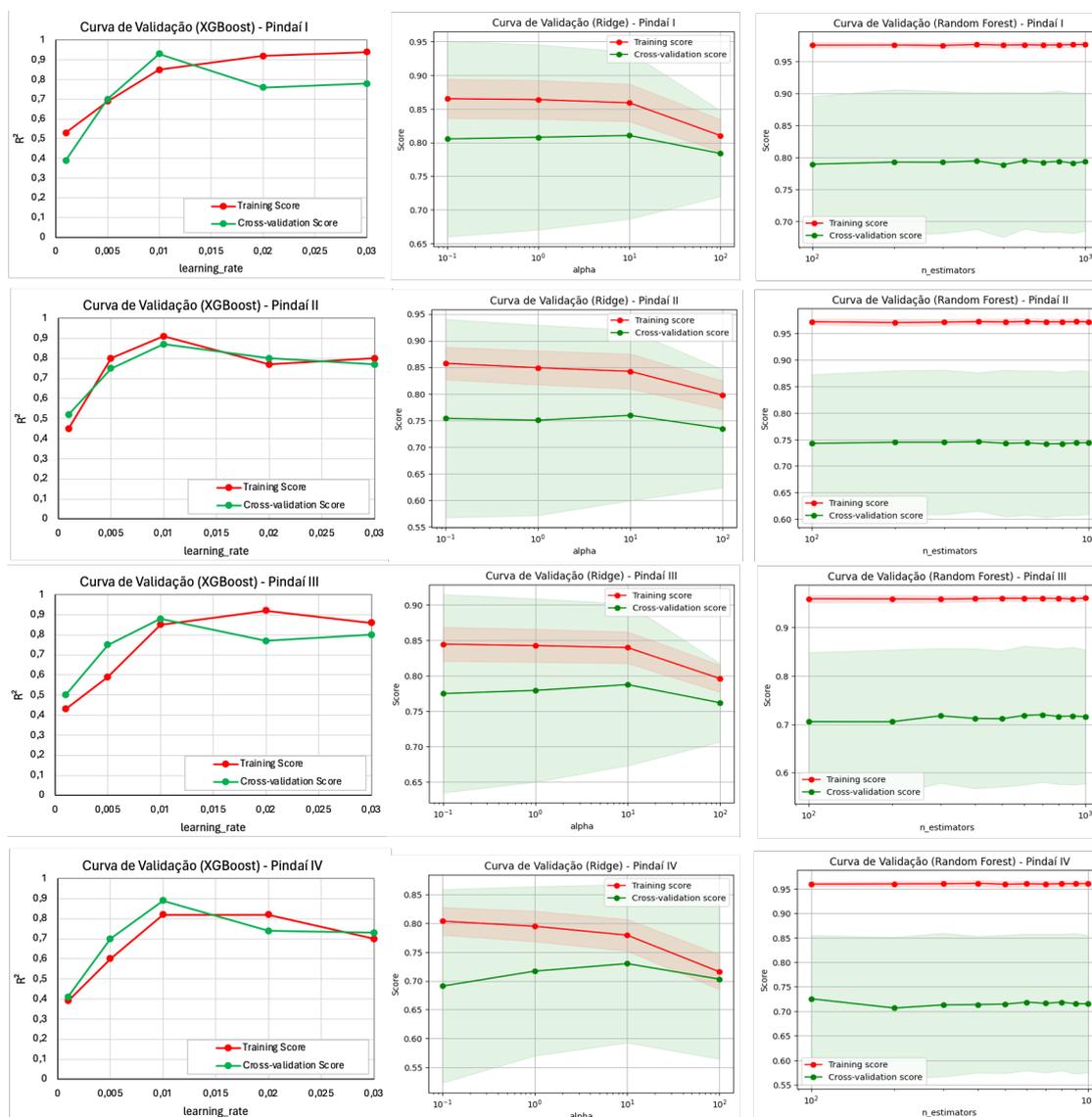


Figura 23 - Curvas de Validação Referente aos Modelos – XGBoost, Regressão Rígida e Florestas Aleatórias

No caso do modelo de Regressão por Vetores Suporte, a análise centrou-se na evolução do R^2 em função do hiperparâmetro C . Esse hiperparâmetro, conforme abordado na seção 3.3, controla o equilíbrio entre a fidelidade ao conjunto de treinamento e a suavidade da função de regressão (penalização dos erros de predição). A Figura 24 apresenta o gráfico da evolução, onde no eixo x o hiperparâmetro C varia entre os valores 0.1, 1, 10, 100, 300, 500 e 1000, enquanto no eixo y estão os valores de R^2 correspondentes a cada um desses valores de C .

Para Pindaí I, o valor de R^2 aumenta consistentemente com o aumento de C , começando em 0.51 para C em 0.1 e alcançando 0.87 para C em 1000. Isso indica uma melhoria contínua na capacidade preditiva do modelo conforme aumenta o valor do hiperparâmetro, sugerindo que o modelo inicialmente sofre de *underfitting* e melhora significativamente seu ajuste aos dados. Em Pindaí II, o valor de R^2 também aumenta com o aumento de C , no entanto a taxa de incremento diminui após C atingir 100, sugerindo que valores muito altos de C não proporcionam ganhos significativos adicionais no desempenho do modelo. Para Pindaí III e IV observa-se um padrão de aumento no valor de R^2 para valores iniciais de C e uma estabilização para valores superiores a 100, indicando que a capacidade de generalização do modelo pode ser atingida antes de C em 1000.

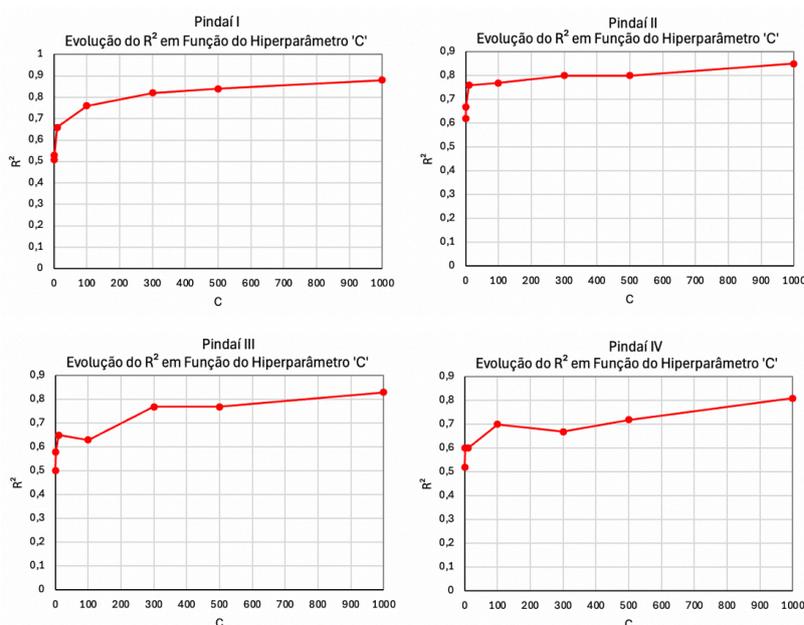


Figura 24 - Evolução do Coeficiente de Determinação em Função do Hiperparâmetro C - Regressão por Vetores Suporte

Para o modelo de Perceptron de Múltiplas Camadas, os gráficos ilustram a evolução e variações do coeficiente de determinação ao longo das épocas de treinamento, apresentado na Figura 25. Foi utilizado o critério de *Early Stopping* para o treinamento do modelo quando não há uma variação maior que 0.01 no R^2 após 1.000 épocas consecutivas. Isso ajuda a evitar o *overfitting*, economizando tempo computacional e garantindo que o modelo não se ajuste excessivamente aos dados de treinamento, o que poderia comprometer sua capacidade de generalização.

Os gráficos na primeira coluna mostram a evolução do R^2 ao longo das épocas de treinamento. Para o Parque Pindaí I, o *Early Stopping* foi acionado em 5.500 épocas ao atingir um R^2 de 0.86. No caso do Parque Pindaí II, o treinamento levou 15.700 épocas para alcançar um R^2 de 0.87. Já o Parque Pindaí III atinge um R^2 de 0.89 após 17.300 épocas e o Parque Pindaí IV alcança um R^2 de 0.85 em 10.700 épocas. Esses valores indicam a quantidade de tempo necessário para que o modelo se ajuste adequadamente aos dados de cada parque, refletindo diferentes níveis de complexidade e características dos dados dos parques eólicos.

O segundo gráfico mostra as variações do R^2 ao longo do tempo, indicando as flutuações do desempenho do modelo conforme as épocas progridem. Inicialmente, observa-se uma alta variabilidade no R^2 , que vai diminuindo gradativamente até estabilizar próximo ao final das épocas necessárias para cada parque. Esse comportamento é esperado, pois o modelo ajusta seus pesos e *bias* durante o treinamento, melhorando sua capacidade preditiva e reduzindo o erro.

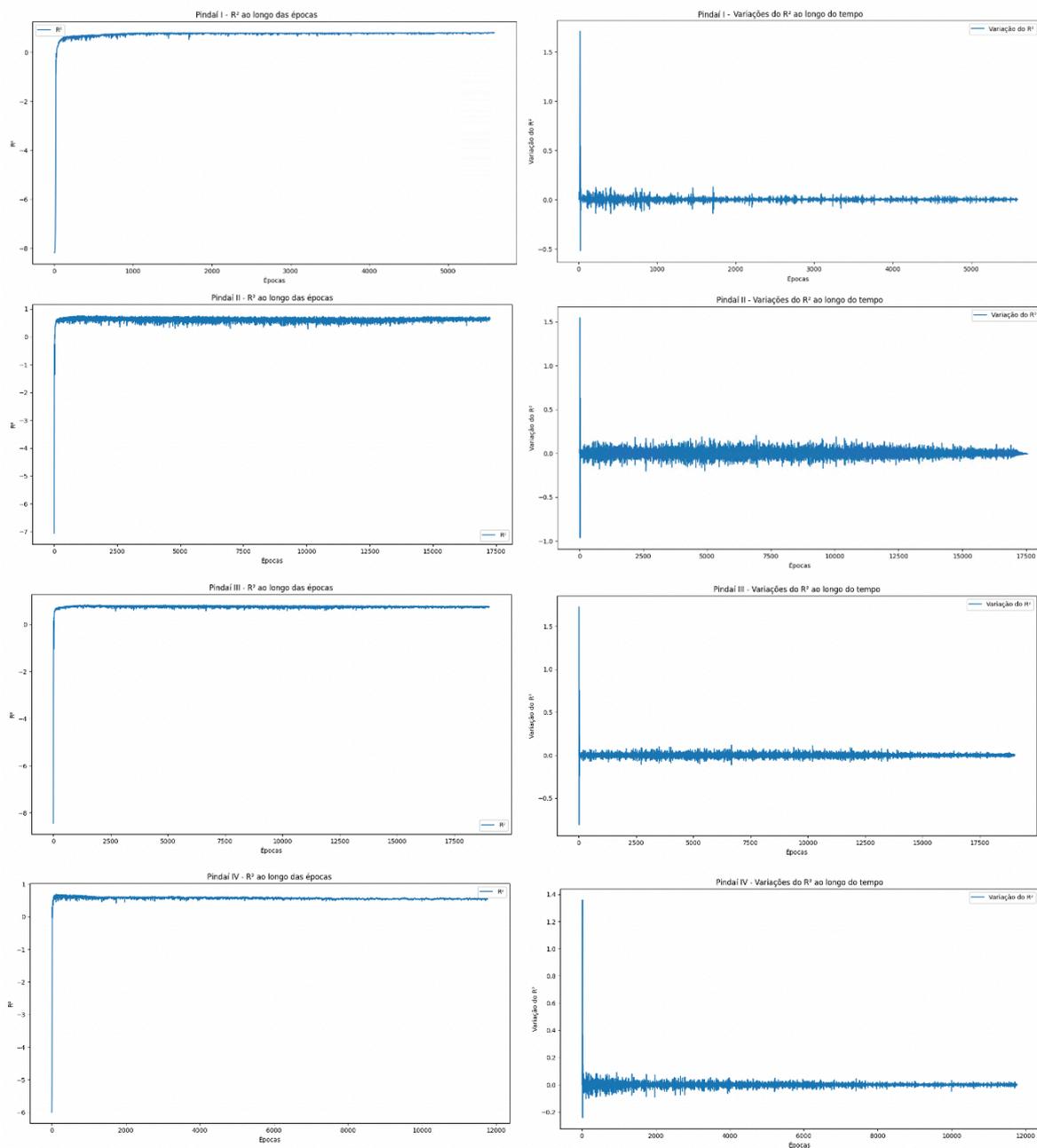


Figura 25 - Evolução e Variações do Coeficiente de Determinação ao Longo das Épocas - MLP

4.3 Resultado da Previsão do Ano de 2023

Na estrutura deste estudo, será realizada a fase de previsão relativo ao período de teste de janeiro a dezembro de 2023. A comparação dos modelos será realizada através de métricas de avaliação do Coeficiente de Determinação R^2 , Erro Percentual Absoluto Médio (MAPE), a Raiz do Erro Quadrático Médio (RMSE) e o Erro Absoluto Médio (MAE), para identificar a metodologia mais eficaz. Para essa previsão prospectiva, serão utilizados os dados preditivos referente às informações meteorológicas provenientes da API de dados em formato JSON disponibilizados pela Open-Meteo, através dos modelos de reanálise de alta resolução. Inicialmente, a base de dados foi extraída em MWh, oferecendo uma visão da energia total produzida em um determinado intervalo de tempo. A apresentação dos resultados será convertida para MWmed, permitindo uma comparação direta da capacidade média de geração de energia ao longo do ano, dividindo os valores mensais em MWh pelo número de horas no respectivo mês. Os meses de janeiro, março, maio, julho, agosto, outubro e dezembro possuem 744 horas. Os meses abril, junho, setembro e novembro com 720 horas. E o mês de fevereiro com 672 horas.

Os resultados da previsão em comparação com a geração realizada podem ser vistos através das Figuras 26, 27, 28 e 29 e da Tabela 20, para cada um dos parques eólicos.

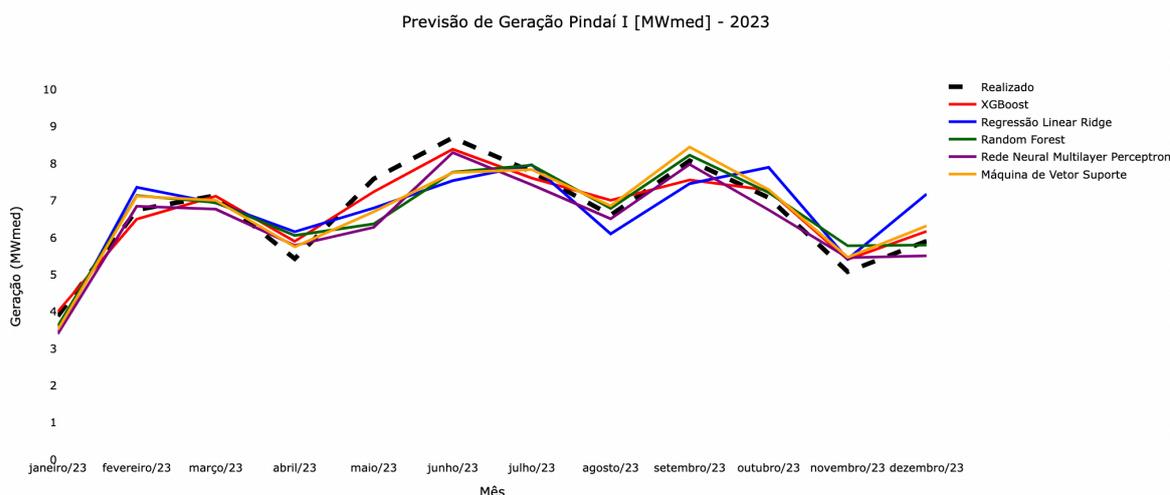


Figura 26 – Previsão de Geração de Pindaí I - 2023

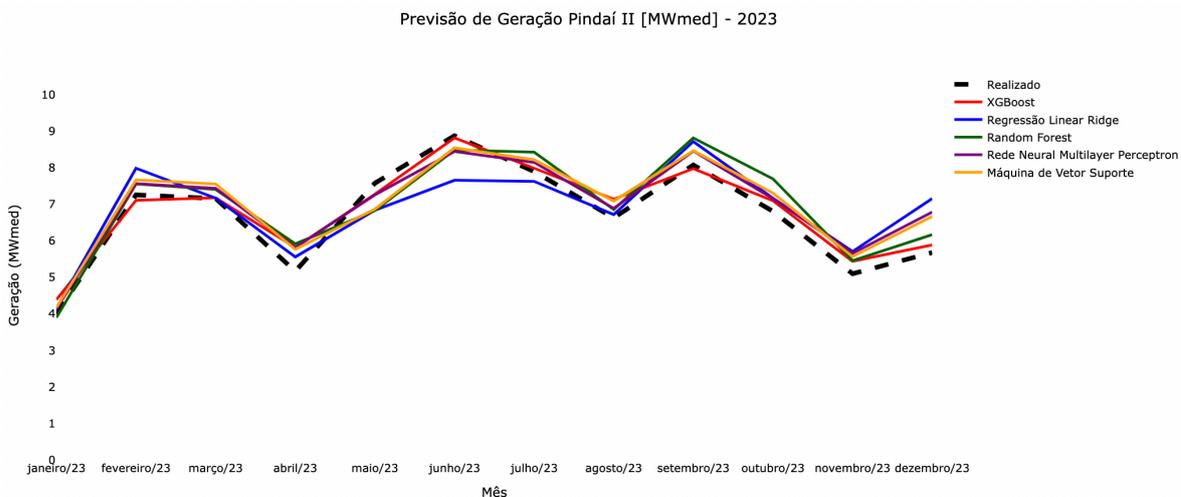


Figura 27 - Previsão de Geração de Pindaí II - 2023

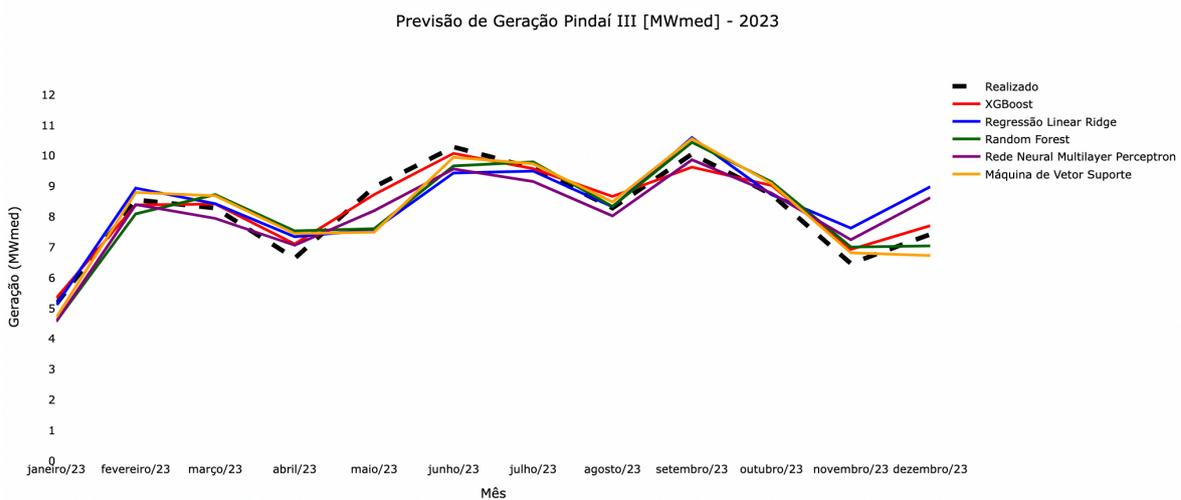


Figura 28 - Previsão de Geração de Pindaí III - 2023

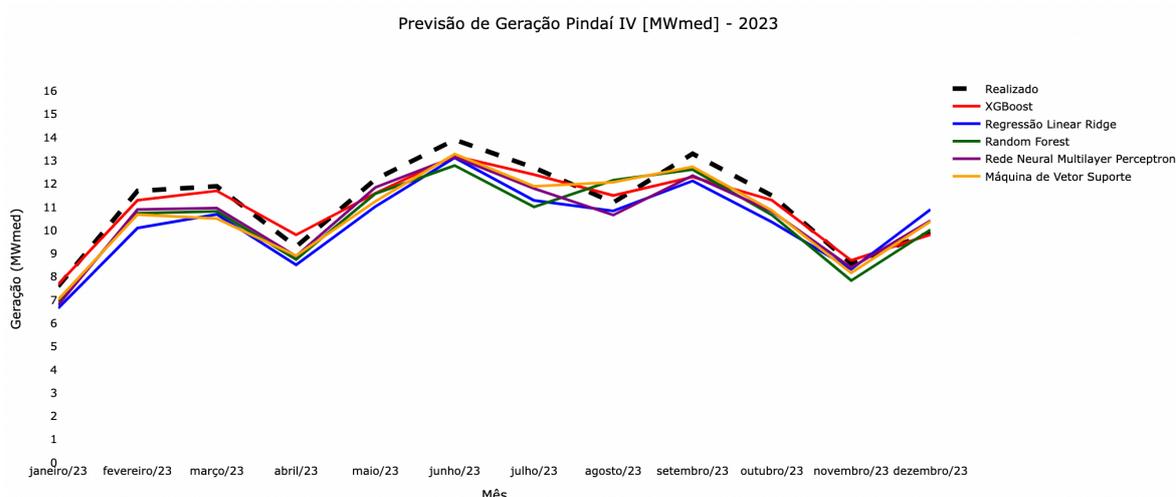


Figura 29 - Previsão de Geração de Pindaí IV – 2023

Tabela 20 - Geração Prevista e Realizada para o Complexo Eólico Pindaí - 2023

[MWmed]

Parque	Geração Realizada	Regressão Rígida	Floresta Aleatória	Regressão por Vetores Suporte	Rede Neural MLP	XGBoost
<i>Pindaí I</i>	[3.88, 6.75, 7.14, 5.43, 7.59, 8.70, 7.78, 6.61, 8.08, 7.08, 5.08, 5.91]	[3.44, 7.36, 6.94, 6.16, 6.80, 7.54, 7.97, 6.10, 7.46, 7.90, 5.42, 7.18]	[3.61, 7.14, 6.95, 6.06, 6.37, 7.77, 7.96, 6.79, 8.23, 7.22, 5.78, 5.80]	[3.51, 7.12, 7.01, 5.75, 6.70, 7.76, 7.84, 6.86, 8.45, 7.30, 5.47, 6.32]	[3.40, 6.85, 6.77, 5.79, 6.28, 8.30, 7.43, 6.51, 7.98, 6.75, 5.46, 5.51]	[4.00, 6.50, 7.12, 5.90, 7.24, 8.39, 7.61, 7.01, 7.56, 7.28, 5.41, 6.17]
<i>Pindaí II</i>	[4.00, 7.26, 7.16, 5.17, 7.58, 8.88, 7.90, 6.63, 8.08, 6.81, 5.10, 5.68]	[4.15, 7.99, 7.17, 5.56, 6.84, 7.66, 7.63, 6.72, 8.72, 7.17, 5.71, 7.16]	[3.90, 7.56, 7.41, 5.92, 6.83, 8.49, 8.43, 6.86, 8.82, 7.70, 5.46, 6.17]	[4.18, 7.67, 7.56, 5.77, 6.87, 8.55, 8.22, 7.09, 8.47, 7.31, 5.57, 6.67]	[4.14, 7.57, 7.44, 5.82, 7.26, 8.45, 8.15, 6.89, 8.46, 7.16, 5.66, 6.79]	[4.39, 7.11, 7.18, 5.82, 7.28, 8.82, 7.99, 7.14, 7.98, 7.10, 5.44, 5.89]
<i>Pindaí III</i>	[5.14, 8.56, 8.29, 6.64, 8.98, 10.29, 9.60, 8.29, 10.06, 8.74, 6.47, 7.43]	[5.12, 8.94, 8.43, 7.35, 7.58, 9.44, 9.50, 8.33, 10.60, 8.73, 7.63, 8.99]	[4.58, 8.10, 8.73, 7.54, 7.61, 9.67, 9.80, 8.34, 10.44, 9.15, 7.01, 7.05]	[4.71, 8.80, 8.69, 7.45, 7.50, 9.95, 9.74, 8.49, 10.56, 9.09, 6.82, 6.73]	[4.57, 8.40, 7.95, 7.07, 8.20, 9.57, 9.16, 8.03, 9.87, 8.78, 7.25, 8.63]	[5.34, 8.39, 8.42, 7.11, 8.73, 10.08, 9.58, 8.67, 9.63, 9.03, 6.93, 7.71]
	[7.59, 11.7, 11.9, 9.3, 11.9, 9.3, 11.9, 9.3, 11.9, 9.3, 11.9, 9.3]	[6.65, 10.10, 10.69, 8.52, 10.69, 8.52, 10.69, 8.52, 10.69, 8.52, 10.69, 8.52]	[6.95, 10.74, 10.82, 8.75, 10.82, 8.75, 10.82, 8.75, 10.82, 8.75, 10.82, 8.75]	[7.03, 10.68, 10.51, 8.91, 10.51, 8.91, 10.51, 8.91, 10.51, 8.91, 10.51, 8.91]	[6.79, 10.90, 10.96, 8.89, 10.96, 8.89, 10.96, 8.89, 10.96, 8.89, 10.96, 8.89]	[7.68, 11.3, 11.7, 9.81, 11.7, 9.81, 11.7, 9.81, 11.7, 9.81, 11.7, 9.81]

<i>Pindaí</i>	12.2, 13.9,	11.03, 13.13,	11.58, 12.79,	11.25, 13.29,	11.85, 13.16,	11.6, 13.2,
<i>IV</i>	12.7, 11.2,	11.29, 10.84,	11.01, 12.16,	11.90, 12.07,	11.80, 10.66,	12.4, 11.5,
	13.3, 11.5,	12.13, 10.36,	12.62, 10.66,	12.74, 10.86,	12.36, 10.82,	12.3, 11.3,
	8.58, 9.91]	8.33, 10.90]	7.84, 10.03]	8.17, 10.38]	8.40, 10.42]	8.71, 9.82]

A Tabela 21 apresenta os resultados das métricas de erro referentes à previsão dos quatro parques eólicos. Para o parque Pindaí I, o modelo XGBoost apresenta o melhor desempenho, com um MAPE de 4.36%, um R^2 de 0.94, um MAE de 0.28 e um RMSE de 0.32. Esses valores indicam que o XGBoost é altamente eficaz na previsão dos dados deste parque, conseguindo minimizar os erros e capturar bem a variabilidade dos dados. A Regressão por Vetores Suporte também performa bem, com um MAPE de 6.01%, R^2 de 0.87, MAE de 0.39 e RMSE de 0.47, seguido pela Rede Neural MLP, que tem um MAPE de 6.14%, R^2 de 0.86, MAE de 0.38 e RMSE de 0.49. O modelo *Random Forest*, com um MAPE de 6.52%, R^2 de 0.83, MAE de 0.42 e RMSE de 0.55, mostra um desempenho intermediário, enquanto a Regressão Rígida apresenta o pior desempenho relativo, com um MAPE de 9.83%, R^2 de 0.71, MAE de 0.63 e RMSE de 0.72.

No Parque Pindaí II, o XGBoost novamente se destaca com um MAPE de 4.50%, R^2 de 0.95, MAE de 0.25 e RMSE de 0.32, indicando uma alta precisão e capacidade de generalização. A Rede Neural MLP segue com um MAPE de 6.73%, R^2 de 0.87, MAE de 0.42 e RMSE de 0.49, demonstrando também uma boa performance. A Regressão por Vetores Suporte e o *Random Forest* apresentam desempenhos similares, com MAPE de 7.52% e 7.25%, R^2 de 0.86 e 0.85, MAE de 0.48 e RMSE de 0.52 e 0.54. A Regressão Rígida novamente mostra o desempenho mais baixo, com MAPE de 8.42%, R^2 de 0.74, MAE de 0.55 e RMSE de 0.70.

Para o Parque Pindaí III, o modelo XGBoost se sobressai com um MAPE de 3.55%, R^2 de 0.96, MAE de 0.27 e RMSE de 0.31, reafirmando sua superioridade em termos de precisão e robustez na previsão dos dados. A Rede Neural MLP e a Regressão por Vetores Suporte mostram desempenhos competitivos, com MAPE de 6.46% e 6.30%, R^2 de 0.84 e 0.83, MAE de 0.49 e RMSE de 0.59 e 0.61, respectivamente. O *Random Forest* apresenta um desempenho intermediário, com MAPE de 6.75%, R^2 de 0.83, MAE de 0.52 e RMSE de 0.62. A Regressão Rígida, com um MAPE de 7.25%, R^2 de 0.72, MAE de 0.57 e RMSE de 0.79, novamente fica atrás dos outros modelos.

No Parque Pindaí IV, o XGBoost lidera com um MAPE de 3.20%, R^2 de 0.94, MAE de 0.37 e RMSE de 0.46, confirmando sua consistência e eficácia na previsão de séries temporais de dados eólicos. A Rede Neural MLP segue com um MAPE de 5.83%, R^2 de 0.86, MAE de 0.64 e RMSE de 0.69, indicando um bom ajuste aos dados. A Regressão por Vetores Suporte, com MAPE de 6.46%, R^2 de 0.83, MAE de 0.72 e RMSE de 0.78, e o *Random Forest*, com MAPE de 7.40%, R^2 de 0.76 MAE de 0.83 e RMSE de 0.91, apresentam desempenhos inferiores, mas ainda consideráveis. A Regressão Rígida, com MAPE de 8.81%, R^2 de 0.67, MAE de 0.98 e RMSE de 1.05, tem o desempenho mais fraco entre os modelos avaliados.

Tabela 21 – Métricas de Avaliação para a Previsão do Complexo Eólico Pindaí

Parque	Modelo	MAPE	R^2	RMSE	MAE
<i>Pindaí I</i>	Regressão Rígida	9.83%	0.71	0.72	0,63
	Floresta Aleatória	6.52%	0.83	0.55	0,42
	Regressão por Vetores Suporte	6.01%	0.87	0.47	0,39
	Rede Neural MLP	6.14%	0.86	0.49	0,38
	XGBoost	4.36%	0.94	0.32	0,28
<i>Pindaí II</i>	Regressão Rígida	8.42%	0.74	0.70	0,55
	Floresta Aleatória	7.25%	0.85	0.54	0,48
	Regressão por Vetores Suporte	7.52%	0.86	0.52	0,48
	Rede Neural MLP	6.73%	0.87	0.49	0,42
	XGBoost	4.50%	0.95	0.32	0,25
	Regressão Rígida	7.25%	0.72	0.79	0,57

<i>Pindaí III</i>	Floresta Aleatória	6.75%	0.83	0.62	0,52
	Regressão por Vetores Suporte	6.30%	0.83	0.61	0,49
	Rede Neural MLP	6.46%	0.84	0.59	0,49
	XGBoost	3.55%	0.96	0.31	0,27
<i>Pindaí IV</i>	Regressão Rígida	8.81%	0.67	1.05	0,98
	Floresta Aleatória	7.40%	0.76	0.91	0,83
	Regressão por Vetores Suporte	6.46%	0.82	0.78	0,72
	Rede Neural MLP	5.83%	0.86	0.69	0,64
	XGBoost	3.20%	0.94	0.46	0,37

A Tabela 22 ilustra o resultado médio das métricas de avaliação para todos os parques. A regressão rígida apresentou o maior erro percentual absoluto médio de 8.58%, com um coeficiente de determinação de 0.71, um erro absoluto médio de 0.68 e RMSE de 0.82. Estes resultados indicam uma performance relativamente fraca em capturar a variabilidade dos dados e em fornecer previsões precisas, refletindo limitações intrínsecas do modelo para este tipo de problema.

Por outro lado, o modelo de florestas aleatórias mostrou uma melhoria significativa com um MAPE de 6.98%, R^2 de 0.82, MAE de 0.56 e RMSE de 0.65. Este modelo beneficia-se da sua capacidade de manejar relações não-lineares e de reduzir a variância ao combinar múltiplas árvores de decisão, resultando em uma maior precisão e capacidade de generalização comparada à regressão rígida.

O desempenho da Regressão por Vetores Suporte foi ainda melhor, com um MAPE de 6.57%, R^2 de 0.85, MAE de 0.52 e RMSE de 0.59. A SVR demonstrou uma eficácia superior em encontrar o hiperplano de separação ótimo no espaço de características, manejando bem as complexidades dos dados e fornecendo previsões mais precisas.

Perceptron de Múltiplas Camadas apresentou um desempenho robusto com um MAPE de 6.29%, R^2 de 0.86, MAE de 0.48 e RMSE de 0.56. Com sua estrutura de camadas ocultas, a MLP mostrou-se eficaz em capturar padrões não-lineares complexos nos dados, resultando em previsões precisas e uma alta capacidade de generalização.

Entretanto, o modelo que mais se destacou foi o XGBoost, com um MAPE significativamente mais baixo de 3.90%, R^2 de 0.95, MAE de 0.29 e RMSE de 0.35. Este modelo de *Gradient Boosting* demonstrou uma capacidade excepcional de ajustar-se aos dados de treinamento ao mesmo tempo que mantém uma excelente capacidade de generalização. A combinação de seu algoritmo de *boosting* com técnicas avançadas de regularização e otimização proporcionou uma performance superior, tornando-o o mais adequado para a previsão de geração eólica mensal entre os modelos avaliados.

Em resumo, os resultados destacam a superioridade do XGBoost em termos de precisão e robustez na previsão de geração eólica, seguido de perto pela rede neural MLP e pela Regressão por Vetores Suporte. As florestas aleatórias também mostraram uma performance sólida, enquanto a regressão rígida ficou aquém das demais técnicas, refletindo suas limitações em capturar a complexidade dos dados de geração eólica.

Tabela 22 - Resultado Médio de todos os Parques – Métricas de Avaliação

Resultado Médio de todos os Parques – Métricas de Avaliação				
Modelo	MAPE	R^2	RMSE	MAE
Regressão Rígida	8.58%	0.71	0.82	0.68
Floresta Aleatória	6.98%	0.82	0.65	0.56
Regressão por Vetores Suporte	6.57%	0.85	0.59	0.52
Rede Neural MLP	6.29%	0.86	0.56	0.48
XGBoost	3.90%	0.95	0.35	0.29

4.4 Resultados e Discussão

Os resultados da análise, conforme mencionado anteriormente, destacam o desempenho superior do XGBoost em todas as métricas e parques eólicos, demonstrando o maior coeficiente de determinação R^2 , significando sua capacidade de explicar a variância nos dados de geração eólica em relação aos modelos de linha de base. Por exemplo, no parque Pindaí I, o XGBoost apresentou um MAPE de 4.36%, um R^2 de 0.94, um desvio absoluto de 0.28 e RMSE de 0.32, superando significativamente os Vetores Suporte (MAPE de 6.01%, R^2 de 0.87, RMSE de 0.47) e a Rede Neural MLP (MAPE de 6.14%, R^2 de 0.86, RMSE de 0.49). O *Random Forest* e a Regressão Rígida apresentaram desempenhos ainda inferiores, com MAPE de 6.52% e 9.83%, respectivamente, e menores valores de R^2 e RMSE.

A tendência de superioridade do XGBoost se manteve nos outros parques eólicos. No parque Pindaí II, o XGBoost obteve um MAPE de 4.50%, R^2 de 0.95, RMSE de 0.32 e desvio absoluto de 0.26, enquanto os outros modelos, como a Rede Neural MLP e a Regressão por Vetores Suporte, tiveram desempenhos consideravelmente menores. Para o parque Pindaí III, o XGBoost novamente liderou com um MAPE de 3.55%, R^2 de 0.96, RMSE de 0.31 e desvio absoluto de 0.27, destacando-se pela precisão e robustez. No parque Pindaí IV, o XGBoost reafirmou sua superioridade com um MAPE de 3.20%, R^2 de 0.94, RMSE de 0.46 e desvio absoluto de 0.38, confirmando sua consistência na previsão de séries temporais de dados eólicos. Comparativamente, os outros modelos, incluindo *Random Forest* e Regressão Rígida, ficaram atrás em todas as métricas avaliadas.

A análise comparativa permite o destaque dos pontos relativos de cada modelagem, como a variabilidade no desempenho dependendo das características específicas dos dados de cada parque eólico, com superioridade para o XGBoost em relação a MLP, SVR, Regressão Rígida e Florestas Aleatórias.

A suposição da Regressão Linear Rígida é da existência uma relação linear entre os preditores e a variável resposta, impondo uma limitação significativa, especialmente no contexto de previsão de séries temporais, em que a complexidade e a não linearidade dos dados são prevalentes. A flexibilidade limitada da Regressão Rígida surge da sua natureza

essencialmente linear. Ao forçar uma relação linear, o modelo não consegue capturar a verdadeira complexidade das interações entre variáveis das séries temporais, em que padrões não lineares, dependências temporais e efeitos sazonais frequentemente desempenham papéis significativos. A incapacidade de acomodar estas dinâmicas não lineares resulta em previsões insatisfatórias, particularmente em cenários onde as relações entre as variáveis preditoras e a variável alvo são mediadas por interações complexas e por fatores externos que não seguem uma distribuição linear. Adicionalmente, ao aplicar uma penalização aos coeficientes, a Regressão Rígida pode obscurecer a importância de variáveis preditoras que, apesar de possuírem uma contribuição não linear significativa, são mal representadas sob a suposição de linearidade. Essa abordagem, embora útil para evitar *overfitting* em modelos com muitas variáveis, pode, paradoxalmente, subestimar ou eliminar preditores que, sob uma análise mais flexível e não linear, teriam um impacto substancial na previsão.

Em relação a metodologia das Florestas Aleatórias, embora robusta na construção de modelos preditivos, apresenta desafios significativos em termos de complexidade computacional. A essência das Florestas Aleatórias está na construção de múltiplas árvores de decisão a partir de diferentes subconjuntos de dados, o que requer um volume substancial de cálculos, especialmente quando o número de árvores é elevado e os dados são volumosos. Esse problema de complexidade computacional se agrava quando o modelo é aplicado a grandes conjuntos de dados com muitas variáveis exógenas, particularmente quando essas variáveis são estocásticas, como é o caso das variáveis climáticas. Tais variáveis aumentam o número de preditores e introduzem uma incerteza adicional, exigindo que o modelo avalie múltiplos cenários e possíveis interações. O resultado é um aumento exponencial no tempo de processamento e nos recursos computacionais necessários para treinar o modelo de forma eficaz.

Além disso, a necessidade de treinar uma quantidade significativa de árvores individuais dentro da floresta implica em múltiplos ciclos de cálculo, em que cada árvore deve ser construída, testada e ajustada. Esse processo não é trivial, especialmente quando se lida com grandes volumes de dados que precisam ser processados em paralelo para otimizar o desempenho do modelo. A gestão dessa complexidade exige infraestrutura computacional avançada, como *clusters* de alto desempenho ou computação em nuvem, o que pode não ser viável em todos os contextos.

A Regressão por Vetores de Suporte distingue-se das metodologias anteriores, como a Regressão Rígida e as Florestas Aleatórias, por sua eficiência em espaços de alta dimensão e sua capacidade de lidar com complexidades não lineares nos dados. A força da SVR em espaços de alta dimensão decorre de sua base matemática, a partir de técnicas de otimização convexa para encontrar a solução de um problema de maximização da margem. Esse enfoque permite que o modelo funcione bem mesmo quando o número de preditores é maior do que o número de observações. Além disso, a SVR é menos suscetível ao problema da maldição da dimensionalidade, uma vez que a complexidade do modelo depende principalmente do número de suportes vetores, que são os pontos de dados mais críticos para a construção do modelo, e não de todos os preditores presentes no conjunto de dados.

Outra característica que confere à SVR uma vantagem significativa é o uso de funções de *kernel*. Através de funções de *kernel*, a SVR pode mapear os dados originais para um espaço de características de maior dimensão, em que relacionamentos não lineares se tornam linearmente separáveis. A flexibilidade das diferentes opções de *kernel* permite que o modelo se adapte aos cenários, capturando variações nos dados que poderiam ser ignoradas por modelos mais rígidos.

O uso eficaz de *kernels* também permite a melhor generalização em novos dados, uma vez que o modelo pode capturar e aprender padrões complexos durante o treinamento. Essa capacidade de generalização é importante em previsão de séries temporais, levando em consideração que as condições subjacentes podem mudar com o tempo, e um modelo que consiga capturar a essência dos relacionamentos não lineares tende a ter um desempenho superior na previsão de eventos futuros. Entretanto, o processo de ajuste do modelo, que inclui a escolha do *kernel* adequado e a determinação dos parâmetros de regularização, pode ser computacionalmente intensivo e requer um entendimento profundo das características dos dados.

O Perceptron de Múltiplas Camadas apresenta diversas vantagens técnicas que o tornam uma escolha eficiente, especialmente na consideração de variáveis climáticas complexas. A capacidade de modelar relações não lineares entre variáveis de entrada e saída e capturar essas interações complexas é um diferencial significativo em relação a modelos mais simples, como a regressão linear. Além disso, o MLP é flexível em termos de sua arquitetura. A capacidade de customizar o número de camadas ocultas e o número de

neurônios em cada camada permite ajustar o modelo ao nível de complexidade dos dados em questão. Isso é útil na modelagem com múltiplas variáveis defasadas, em que diferentes camadas podem ser responsáveis por capturar diferentes aspectos dos padrões temporais e sazonais presentes nos dados. Outra vantagem é sua capacidade de generalização quando treinado adequadamente. Através de técnicas como regularização, que inclui métodos como regularização $L2$ e *dropout*, o MLP se mantém eficiente na captura dos padrões relevantes nos dados de treinamento, enquanto preserva a capacidade de fazer previsões em novos conjuntos de observações. Além disso, o MLP pode integrar múltiplas variáveis temporais, permitindo a modelagem simultânea de diversos fatores climáticos que influenciam a geração eólica. Ao incorporar essas variáveis em diferentes estágios do processo de modelagem, consegue-se capturar as interações mais complexas que podem se manifestar ao longo do tempo. Essa capacidade é relevante na previsão de séries temporais em que a interação entre múltiplos fatores é determinante para a precisão das previsões.

A superioridade da metodologia do *eXtreme Gradient Boosting* pode ser elencada por diversos fatores. Um exemplo disso é relativo à estrutura do algoritmo de aprendizado, que captura padrões não lineares complexos nos dados típicos da dinâmica da geração eólica. O aumento de gradiente permite refinamentos iterativos nas previsões do modelo, tornando-o adequado para conjuntos de dados onde a relação entre variáveis de entrada e saídas é complexa e em camadas.

A técnica do *Gradient Boosting* é projetada para aprimorar iterativamente modelos individuais, geralmente árvores de decisão de baixa profundidade. Embora essas árvores não sejam robustas, o mecanismo de aprendizado iterativo permite uma adaptação contínua, identificando e corrigindo erros residuais em cada iteração subsequente. Esse ajuste fino progressivo é essencial para melhorar o ajuste do modelo aos dados de treinamento, resultando em previsões mais precisas. A eficácia dessa abordagem é evidenciada por sua capacidade de alcançar altos coeficientes de determinação (R^2), superando redes neurais e regressão por vetores suporte em termos de precisão. Além disso, o XGBoost se destaca por sua versatilidade, permitindo o uso de diferentes funções objetivas e a personalização de funções de perda. Isso possibilita a adaptação do modelo a várias aplicações de regressão, ajustando funções de perda específicas para priorizar erros em áreas críticas dos dados ou para lidar com diferentes escalas de erro.

Outra característica é sua capacidade de fornecer pontuações de importância de recursos, que contribuem para a interpretabilidade e compreensão do modelo. O cálculo dessas pontuações é baseado no ganho de informação, que mede a melhoria na qualidade das divisões de dados proporcionada por cada recurso durante a construção das árvores de decisão. Ao sumarizar esses ganhos em todas as árvores, o XGBoost gera uma pontuação de importância que destaca as características mais influentes na previsão. Essa análise facilita a seleção de recursos e o pré-processamento, permitindo uma priorização eficiente de variáveis-chave, orientando estratégias de tratamento de dados mais eficazes.

O XGBoost também implementa o aprendizado paralelizado. Isso permite que o treinamento do modelo seja distribuído entre múltiplos núcleos de processamento (*Central Processing Unit* - CPU), unidades de processamento gráfico (*Graphics Processing Unit* - GPU) e *clusters* de redes de computadores, otimizando significativamente o tempo de execução. O *design* do algoritmo inclui otimizações para maximizar o uso eficiente do *hardware* disponível, como a exploração do cache de memória. Essa otimização inclui a alocação de *buffers* internos em cada *thread* de execução para armazenar estatísticas de gradientes, reduzindo a necessidade de acesso frequente à memória principal, que é mais lenta, e acelerando o processo de treinamento.

A arquitetura do XGBoost suporta a execução em *clusters* de computadores, permitindo escalabilidade praticamente ilimitada. Isso é alcançado através da paralelização distribuída, onde o treinamento do modelo pode ser dividido entre diferentes máquinas, cada uma lidando com uma parte do processo de construção das árvores. Essa capacidade acelera o treinamento e permite a manipulação de volumes de dados que seriam impraticáveis de processar em uma única máquina.

Com base no *benchmark* de trabalhos semelhantes discutido na seção 3.1, os resultados deste estudo evidenciam a superioridade das metodologias aplicadas, com destaque para a Rede Neural MLP e o XGBoost. A revisão de estudos similares na área de previsão de geração eólica indica que valores de RMSE abaixo de 0,5, R^2 superiores a 0,90, MAPE abaixo de 6% e MAE inferiores a 0,40 são considerados indicadores de um desempenho preditivo robusto. Os resultados obtidos neste trabalho, alinhados com esses valores de referência, confirmam a alta precisão dos modelos aplicados, destacando sua eficácia na modelagem e previsão de dados eólicos.

Capítulo 5 - Conclusões e Trabalhos Futuros

Esta dissertação teve como principal finalidade o estudo e desenvolvimento de métodos de previsão para o auxílio na tomada de decisões estratégicas no âmbito do empreendimento eólico, especificamente para o Complexo Eólico Pindaí e sua operação comercial. No trabalho, foram integradas técnicas de análise de *Big Data*, engenharia de dados e computação por meio de algoritmos de aprendizado de máquina, com o objetivo de realizar prospecções futuras sobre a geração de energia do complexo no ano de 2023. Para tal, foram modeladas e implementadas cinco técnicas de aprendizado de máquina supervisionada: *eXtreme Gradient Boosting*, Perceptron de Múltiplas Camadas, Regressão por Vetores Suporte, Regressão Linear Rígida e Florestas Aleatórias. Essas técnicas foram fundamentadas em suas respectivas considerações matemáticas e estruturais, e os dados históricos de geração dos parques eólicos do Complexo Pindaí, combinados com dados meteorológicos realizados e previstos de alta resolução, acessados por meio da Open-Meteo API, forneceram a base empírica para a modelagem.

O estado da arte sobre a previsão de geração de energia eólica tem explorado diversas abordagens de aprendizado de máquina para lidar com a complexidade estocástica dos dados meteorológicos, integrando técnicas avançadas de análise de séries temporais com algoritmos supervisionados de alta precisão. Cada técnica apresenta um enfoque específico na otimização da previsão, seja para maximizar a eficiência operacional dos parques eólicos, minimizar erros de previsão ou garantir a robustez das estimativas em condições climáticas variáveis. Dentre as metodologias mais recorrentes na literatura, as técnicas de aprendizado supervisionado, como o *eXtreme Gradient Boosting*, as Redes Neurais Perceptron de Múltiplas Camadas e a Regressão por Vetores Suporte, foram selecionadas por sua capacidade de capturar padrões complexos e não lineares nos dados eólicos. A validação dessas metodologias foi realizada por meio da aplicação de métricas de desempenho, como o Coeficiente de Determinação, o Erro Percentual Absoluto Médio, o Erro Médio Absoluto e a Raiz do Erro Quadrático Médio. A análise de sensibilidade permitiu avaliar a robustez dos modelos em relação às variáveis climáticas, e os resultados indicaram uma superioridade

significativa das previsões, especialmente no que diz respeito à capacidade de antecipar com precisão as variações sazonais e interanuais do vento, fundamentais para a operação e planejamento estratégico do complexo eólico. A implementação dos modelos foi ajustada para garantir a maximização da acurácia das previsões, permitindo uma maior confiabilidade nas decisões operacionais e contratuais relacionadas à geração de energia.

A aplicação de uma análise detalhada dos hiperparâmetros para cada modelo, ajustados por meio de validação cruzada, foi um dos diferenciais desta pesquisa. Esse processo otimizou o desempenho de cada técnica e preencheu lacunas presentes na literatura, especialmente no que se refere à exploração exaustiva de combinações de hiperparâmetros em modelos de aprendizado de máquina aplicados à previsão de geração de energia eólica.

Outro diferencial desta pesquisa foi a utilização de modelos meteorológicos de reanálise de alta resolução, integrando variáveis meteorológicas como exógenas no processo de construção dos algoritmos de aprendizado de máquina. A incorporação dessas variáveis climáticas, como velocidade do vento, pressão atmosférica, temperatura, rajadas de vento e precipitação, forneceu uma base realista para as previsões de geração de energia, permitindo que os modelos capturassem com maior fidelidade as dinâmicas atmosféricas que influenciam diretamente a produção eólica. Este enfoque preenche uma lacuna significativa na literatura, na qual muitas abordagens ainda não exploram suficientemente o impacto das condições meteorológicas detalhadas na geração eólica, otimizando a acurácia das previsões e reforçando a capacidade de antecipar variações climáticas sazonais e interanuais.

As implicações financeiras e operacionais desta pesquisa revelam-se significativas, uma vez que a precisão das previsões de geração eólica permite aos operadores e investidores otimizar suas estratégias de mercado. A capacidade de antecipar a produção energética com maior confiabilidade possibilita uma gestão mais eficiente dos recursos, reduzindo custos operacionais e melhorando a alocação de investimentos. Previsões precisas de longo prazo facilitam a negociação de contratos de venda de energia em condições mais vantajosas, garantindo maior estabilidade financeira e maximizando o retorno sobre o investimento no setor de energia renovável.

Tendo em vista os resultados obtidos nesta dissertação e levando em consideração a lacuna existente na literatura quanto à utilização de variáveis meteorológicas de alta resolução como insumos, a aplicação de técnicas de aprendizado de máquina e as análises

comerciais dos impactos financeiros que prospecções futuras possibilitam, sugere-se algumas propostas de trabalhos futuros para dar continuidade a esta linha de pesquisa, tais como:

- Expandir o escopo da pesquisa para abranger diferentes escalas temporais, com ênfase nas previsões de curto prazo, que projetam a geração de energia eólica em intervalos de horas à frente. Essa abordagem torna-se mais crítica com a introdução dos preços horários de energia no setor elétrico brasileiro, por meio do modelo de otimização Dessem.
- Desenvolver cenários meteorológicos probabilísticos para aprimorar a compreensão da variabilidade em fatores críticos como a velocidade do vento e a precipitação. Ao invés de depender exclusivamente de um cenário único de projeção futura, a criação de cenários dinâmicos permite simular os potenciais impactos das condições climáticas flutuantes na produção de energia eólica. Utilizando curvas que representem cenários superiores (otimistas) e inferiores (pessimistas) para variáveis meteorológicas, é possível atribuir probabilidades a esses cenários, permitindo a identificação de riscos e oportunidades, a fim de aprimorar a tomada de decisões estratégicas, considerando as incertezas intrínsecas ao ambiente climático.
- Aplicação de técnicas de aprendizado profundo, como Redes Neurais Recorrentes (*Recurrent Neural Network* - RNN) e Redes de Memória de Longo e Curto Prazo (*Long Short-Term Memory* - LSTM).
- A aplicação e o teste de funções de ativação não lineares na metodologia do Perceptron de Múltiplas Camadas.
- Condução de estudos que examinem a curva futura de geração eólica no portfólio de uma comercializadora de energia, com foco na avaliação do impacto financeiro das previsões, visando mitigar riscos e maximizar o retorno sobre o investimento.

Bibliografia

- [1] ONS, "Capacidade Instalada de Geração - Conjunto de dados". [Online]. Disponível: <https://dados.ons.org.br/dataset/capacidade-geracao>. [Acesso em outubro 2023].
- [2] P. E. Sorensen, A. D. Hansen, H. Bindner, "Estimation of the Wind Energy Production in Stochastic Wind Conditions," *Wind Energy*, 2005.
- [3] M. M. W. Lewis, M. Leithead, S. Galloway, "Optimal Maintenance Scheduling for Offshore Wind Farms," *Renewable Energy*, 2018.
- [4] M. A. Hossain, K. Mahmud, S. M. Muyeen, J. H. Khan, M. Padmanaban, "Application of Big Data and Machine Learning in Smart Grid, and Associated Security Concerns: A Review," *IEEE Access*, 2019.
- [5] N. Cristianini, J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.
- [6] A. J. Smola, B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, 2004.
- [7] S. Haykin. *Neural Networks and Learning Machines*, 3rd ed. Prentice Hall, 2009.
- [8] T. Chen, C. Guestrin, "XGBoost: A scalable tree boosting system", in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016.
- [9] S. Marsland, "Machine Learning: An Algorithmic Perspective", Chapman & Hall/CRC, 2014.
- [10] C.M. Bishop, "Pattern Recognition and Machine Learning", Springer, 2006.
- [11] T.M. Mitchell, "Machine Learning", McGraw Hill, 1997.
- [12] E. Alpaydin, "Introduction to Machine Learning", 4th ed., MIT Press, 2020.
- [13] I. Goodfellow, Y. Bengio, A. Courville, "Deep Learning", MIT Press, 2016.
- [14] T. Hastie, R. Tibshirani, J. Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Springer, 2009.
- [15] A.C. Rencher, G.B. Schaalje, "Linear Models in Statistics", John Wiley & Sons, 2008.
- [16] V. Singh, V.K. Asari, R. Rajasekaran, "A Deep Neural Network for Early Detection and Prediction of Chronic Kidney Disease", *Diagnostics*, 2022.

- [17] S. Geman, E. Bienenstock, R. Doursat, "Neural Networks and the Bias/Variance Dilemma", *Neural Computation*, 1992.
- [18] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso", *Journal of the Royal Statistical Society*, 1996.
- [19] C.C. Aggarwal, C.K. Reddy, "Data Clustering: Algorithms and Applications", Chapman & Hall/CRC, 2013.
- [20] I.T. Jolliffe, "Principal Component Analysis", Springer Series in Statistics, Springer, 2002.
- [21] J. Xie, R. Girshick, A. Farhadi, "Unsupervised Deep Embedding for Clustering Analysis", in *Proc. Int. Conf. Machine Learning*, 2016.
- [22] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative Adversarial Nets", in *Advances in Neural Information Processing Systems*, 2014.
- [23] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, Cambridge, MIT Press, 1998.
- [24] C. J. C. H. Watkins, "Learning from delayed rewards," Ph.D. dissertation, King's College, Cambridge, U.K, 1989.
- [25] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, 2015.
- [26] J. Moody, M. Saffell, "Learning to trade via direct reinforcement," *IEEE Transactions on Neural Networks*, 2001.
- [27] Liu, H., Zhang, X., Zhang, Y., Wang, J., Zhang, K. "Deep reinforcement learning for time-varying demand response management in smart grids." *IEEE Transactions on Smart Grid*, 2019.
- [28] A. G. Barto, S. Mahadevan, "Recent advances in hierarchical reinforcement learning," *Discrete Event Dynamic Systems*, 2003.
- [29] R. J. Hyndman, G. Athanasopoulos, "Forecasting: Principles and Practice", 2nd ed, 2018.

- [30] Open-Meteo.com, "Docs". [Online]. Disponível: <https://open-meteo.com/en/docs>. [Acesso em maio de 2024].
- [31] J. D. Hamilton, "Time Series Analysis". Princeton, NJ, USA: Princeton University Press, 1994.
- [32] G. E. P. Box, G. M. Jenkins. "Time Series Analysis: Forecasting and Control". Revised ed., San Francisco, CA, 1970.
- [33] Camelo, H.; Lucio, P.; Junior, J.; Carvalho, P. "Time Series Forecasting Methods and Hybrid Modeling both Applied to Monthly Average Wind Speed for Northeastern Regions of Brazil", 2017.
- [34] M. Paula, W. Casaca, M. Colnago, J. R. da Silva, K. Oliveira, M. A. Dias, R. Negri, "Predicting Energy Generation in Large Wind Farms: A Data-Driven Study with Open Data and Machine Learning," *Inventions*, 2023.
- [35] T. Magesh, S. Franklin, P. S. Santhi, M. Thiyagesan. "Machine Learning-Driven Wind Energy Forecasting for Sustainable Development," *MATEC Web of Conferences*, 2024.
- [36] S.M. Stigler, "Francis Galton's Account of the Invention of Correlation", *Statistical Science*, 1989.
- [37] G. James, D. Witten, T. Hastie, R. Tibshirani, "An Introduction to Statistical Learning", Springer, 2013.
- [38] D. C. Montgomery, E. A. Peck, G. G. Vining, "Introduction to Linear Regression Analysis", 5th ed, 2012.
- [39] D. Shumway, D. Stoffer, "Time Series Analysis and Its Applications: With R Examples", 4th ed., Springer, 2017.
- [40] M.H. Kutner, C.J. Nachtsheim, J. Neter, W. Li, "Applied Linear Statistical Models", 5th ed., 2005.
- [41] D.S. Moore, G.P. McCabe, B.A. Craig, "Introduction to the Practice of Statistics", 9th ed., W.H. Freeman, 2020.
- [42] J. Durbin, G.S. Watson, "Testing for Serial Correlation in Least Squares Regression I", *Biometrika*, 1950.

- [43] S.S. Shapiro, M.B. Wilk, "An analysis of variance test for normality (complete samples)", *Biometrika*, 1965.
- [44] A.E. Hoerl, R.W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems", *Technometrics*, 1970.
- [45] R. Penrose, "A generalized inverse for matrices," *Proceedings of the Cambridge Philosophical Society*, 1955.
- [46] H. Zou, T. Hastie, "Regularization and Variable Selection via the Elastic Net", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2005.
- [47] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, I. Loris, "Sparse and Stable Markowitz Portfolios", *Proceedings of the National Academy of Sciences*, 2009.
- [48] H. Zou, "The Adaptive Lasso and Its Oracle Properties", *Journal of the American Statistical Association*, 2006.
- [49] L. Breiman, "Random Forests", *Machine Learning*, 2001.
- [50] A. Sezer, A.M. Ozbayoglu, "Forecasting with Random Forests: The Effects of Temporal Aggregation on Stock Market Data", in *IEEE Access*, 2018.
- [51] V. N. Vapnik, A. Y. Chervonenkis, "A note on one class of perceptrons," *Automation and Remote Control*, 1964.
- [52] V. Vapnik, "The Nature of Statistical Learning Theory". Springer, 1995.
- [53] C. Cortes, V. Vapnik, "Support-Vector Networks," *Machine Learning*, 1995.
- [54] J. Platt, "Fast Training of Support Vector Machines using Sequential Minimal Optimization," in *Advances in Kernel Methods - Support Vector Learning*, 1998.
- [55] F. E. H. Tay, L. Cao, "Application of support vector machines in financial time series forecasting," *Omega*, 2001.
- [56] J. A. K. Suykens, J. Vandewalle, "Least Squares Support Vector Machine Classifiers," *Neural Processing Letters*, 1999.
- [57] C.-J. Hsieh, K.-W. Chang, C.-J. Lin, S. S. Keerthi, S. Sundararajan, "A dual coordinate descent method for large-scale linear SVM," in *Proceedings of the 25th International Conference on Machine Learning*, 2008.
- [58] A. J. Smola e B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, 2004.

- [59] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, V. Vapnik, "Support vector regression machines," *Advances in Neural Information Processing Systems*, 1997.
- [60] V. N. Vapnik, "Statistical Learning Theory". New York, 1998.
- [61] B. Schölkopf, A. J. Smola, "Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond". Cambridge, 2002.
- [62] W. S. McCulloch, W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, 1943.
- [63] P. J. Werbos, "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences," Ph.D. dissertation, Harvard University, Cambridge, 1974.
- [64] I. Goodfellow, Y. Bengio, A. Courville, "Deep Learning". Cambridge, MIT Press, 2016.
- [65] D. E. Rumelhart, G. E. Hinton, R. J. Williams, "Learning representations by back-propagating errors", 1986.
- [66] Y. LeCun, L. Bottou, G. B. Orr, K. R. Müller, "Efficient backprop," in *Neural Networks: Tricks of the Trade*, Springer, Berlin, 2012.
- [67] A. Y. Ng, "Feature selection, L1 vs. L2 regularization, and rotational invariance," in *Proceedings of the twenty-first international conference on Machine learning*, 2004.
- [68] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, 2014.
- [69] J.H. Friedman, "Greedy function approximation: A gradient boosting machine", *The Annals of Statistics*, 2001.
- [70] C. Krauss, X.A. Do, N. Huck, "Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500", *European Journal of Operational Research*, 2017.
- [71] J. Wang, Z. Chen, Y. Shi, e S. Wang, "Electric Load Forecasting with Recurrent Neural Networks and XGBoost," *IEEE Access*, 2019.
- [72] C. Bojer, "Predicting retail demand using eXtreme Gradient Boosting", *Journal of Retail Analytics*, 2021.

- [73] S.M. Lundberg, S.I. Lee, "A Unified Approach to Interpreting Model Predictions", in *Advances in Neural Information Processing Systems*, 2017.
- [74] R.J. Hyndman, G. Athanasopoulos, "Forecasting: principles and practice", 2nd ed., OTexts, 2018.
- [75] S.J. Taylor, B. Letham, "Forecasting at scale", *The American Statistician*, 2018.
- [76] G. Bontempi, S.B. Taieb, Y.A. Le Borgne, "Machine learning strategies for time series forecasting", in *Business Intelligence*, Springer, 2013.
- [77] M. Abbasi, "Enhancing Smart Grid Efficiency via XGBoost", 2019.
- [78] L. Breiman, "Bagging Predictors", *Machine Learning*, 1996.
- [79] J. Bergstra and Y. Bengio, "Random Search for Hyper-Parameter Optimization", *Journal of Machine Learning Research*, 2012.
- [80] J. Snoek, H. Larochelle, R.P. Adams, "Practical Bayesian Optimization of Machine Learning Algorithms", in *Advances in Neural Information Processing Systems*, 2012.
- [81] J. R. Holton, "An Introduction to Dynamic Meteorology", 4th ed., Netherlands: Academic Press, 2004.
- [82] M.A.V. Silva, "Meteorologia e Climatologia", 2nd ed., Recife, 2006.
- [83] J.M. Wallace, P.V. Hobbs, "Atmospheric Science: An Introductory Survey", Elsevier, 2006.
- [84] E.N. Lorenz, "The Nature and Theory of the General Circulation of the Atmosphere", World Meteorological Organization, 1967.
- [85] W.R. Cotton, R.A. Pielke Sr. "Human Impacts on Weather and Climate", Cambridge University Press, 2007.
- [86] J.R. Holton, "An Introduction to Dynamic Meteorology", Elsevier, 2004.
- [87] E.N. Lorenz, "Deterministic Nonperiodic Flow", *Journal of the Atmospheric Sciences*, 1963.
- [88] D.J. Stensrud, "Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models", Cambridge University Press, 2009.
- [89] J.S. Katz, A.H. Murphy, "Economic value of weather and climate forecasts", Cambridge University Press, 2013.

- [90] R.B. Stull, "An Introduction to Boundary Layer Meteorology", Kluwer Academic Publishers, 1988.
- [91] D. Hartmann, "Global Physical Climatology", 2016.
- [92] INMET, "O que é Zona de Convergência Intertropical?", Ministério da Agricultura e Pecuária. [Online]. Disponível: <https://portal.inmet.gov.br/>. [Acesso em fevereiro 2024].
- [93] E.A. Mathez, "Climate Change: The Science of Global Warming and Our Energy Future", Columbia University Press, 2009.
- [94] NOAA, "In Watching for El Niño and La Niña, NOAA Adapts to Global Warming". [Online]. Disponível: <https://www.climate.gov>. [Acesso em março 2024].
- [95] Tradener. [Online]. Disponível: <https://www.tradener.com.br/>. [Acesso em novembro 2023].
- [96] T. Burton, N. Jenkins, D. Sharpe, E. Bossanyi. "Wind Energy Handbook", 2nd ed. Chichester, 2011.
- [97] S. El Mouhsine, K. Oukassou, M.M. Ichenial, B. Kharbouch, A. Hajraoui, "Aerodynamics and structural analysis of wind turbine blade", Procedia Manufacturing, 2018.
- [98] M. Tan, "Investigating the Effective Design of the Modern Wind Turbine", Medium, 2023. [Online]. Disponível: <https://medium.com>. [Acesso em janeiro 2024].
- [99] E. Hau, "Wind Turbines: Fundamentals, Technologies, Application, Economics", 2nd ed. Berlin, Germany: Springer, 2013
- [100] Camargo-Schubert. [Online]. Disponível: <https://www.camargo-schubert.com/>. [Acesso em janeiro 2024].
- [101] J. Jain, "Wind energy engineering", New York: McGraw-Hill, 2011.
- [102] J.D. Anderson, "Fundamentals of Aerodynamics", McGraw-Hill, 2010.
- [103] T. Burton, N. Jenkins, D. Sharpe, E. Bossanyi, "Wind Energy Handbook", John Wiley & Sons, 2011.
- [104] J. Manwell, J. McGowan, A. Rogers, "Wind Energy Explained: Theory, Design and Application", 2nd ed., John Wiley & Sons, 2009.

- [105] V. Mendoza. "Aerodynamic Studies of Vertical Axis Wind Turbines using the Actuator Line Model", Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology, 2018.
- [106] S.S. Johannes, "Rotor adapted to be driven by wind or flowing water", US Patent, 1929.
- [107] D.G.J. Marie, "Turbine having its rotating shaft transverse to the flow of the current", US Patent, 1931.
- [108] I. Prowell. "An Experimental and Numerical Study of Wind Turbine Seismic Behavior". University of California, San Diego, 2011.
- [109] E. Hau, "Wind Turbines", Springer, Berlin, Germany, 2006.
- [110] S. Buragohain, "Analysis of Performance and Optimization Strategies for Horizontal Axis Wind Turbine (HAWT)", International Journal of Innovative Science and Research Technology, 2023.
- [111] F. Barlas, D. Charalampous, "Fundamentals of Wind Farm Aerodynamic Layout Design", 1st ed. London, UK: Academic Press, 2021.
- [112] S. Heier, "Grid Integration of Wind Energy: Onshore and Offshore Conversion Systems", 3rd ed. Chichester, UK: John Wiley & Sons, 2014.
- [113] Scikit-Learn. "scikit-learn: Machine Learning in Python," [Online]. Disponível: <https://scikit-learn.org/>. [Acesso em maio 2024].
- [114] TensorFlow. "TensorFlow: An end-to-end open-source machine learning platform," [Online]. Disponível: <https://www.tensorflow.org/>. [Acesso em maio 2024].
- [115] Keras. "Keras: The Python Deep Learning API," [Online]. Disponível: <https://keras.io/>. [Acesso em maio 2024].
- [116] Skforecast. "Skforecast: Time Series Forecasting with Scikit-Learn Regressors," [Online]. Disponível: <https://github.com/JoaquinAmatRodrigo/skforecast>. [Acesso em maio 2024]
- [117] XGBoost. "XGBoost: Scalable and Flexible Gradient Boosting," [Online]. Disponível: <https://xgboost.ai/>. [Acesso em maio 2024]
- [118] Pandas. "Python Data Analysis Library," [Online]. Disponível: <https://pandas.pydata.org/>. [Acesso em maio 2024].

- [119] NumPy. "NumPy: The fundamental package for scientific computing with Python," [Online]. Disponível: <https://numpy.org/>. [Acesso em maio 2024].
- [120] Matplotlib. "Matplotlib: Python plotting — Matplotlib 3.4.2 documentation," [Online]. Disponível: <https://matplotlib.org/>. [Acesso em maio 2024].
- [121] Seaborn. "Seaborn: Statistical data visualization," [Online]. Disponível: <https://seaborn.pydata.org/>. [Acesso em maio 2024].
- [122] Open-Meteo Requests. "Open-Meteo API: Weather forecasts for any location," [Online]. Disponível: <https://open-meteo.com/>. [Acesso em maio 2024].
- [123] Enercon E-92 - 2,35 MW - Wind turbine. [Online]. Disponível: <https://en.wind-turbine-models.com/turbines/131-enercon-e-92>. [Acesso em abril 2024].
- [124] Wobben E92/2350 - Manufacturers and turbines - Online access - The Wind Power. [Online]. Disponível: https://www.thewindpower.net/turbine_en_1019_wobben_e92-2350.php. [Acesso em abril 2024].
- [125] SCADA System Explained, [Online]. Disponível: <https://scada-international.com/>. [Acesso em abril 2024].
- [126] G. Bontempi, S. B. Taieb, Y. A. Le Borgne, "Machine Learning Strategies for Time Series Forecasting," European Business Intelligence Summer School, 2013.
- [127] D.B. Rubin, "Inference and Missing Data", *Biometrika*, 1976.
- [128] R.J.A. Little, D.B. Rubin, "Statistical Analysis with Missing Data in Probability and Statistics ", Wiley, 2019.
- [129] D.S. Moore, G.P. McCabe, B.A. Craig, "Introduction to the Practice of Statistics", 9th ed., W.H. Freeman, 2020.
- [130] Google Maps. [Online]. Disponível: <https://www.google.com/maps>. [Acesso em agosto 2024].
- [131] European Centre for Medium-Range Weather Forecasts, "ECMWF," [Online]. Disponível: <https://www.ecmwf.int/>. [Acesso em maio 2024].
- [132] Copernicus Climate Change Service, "ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate," [Online]. Disponível: <https://climate.copernicus.eu/climate-reanalysis>. [Acesso em maio 2024].

- [133] Copernicus Climate Change Service, "ERA5-Land monthly averaged data from 1981 to present," [Online]. Disponível: <https://cds.climate.copernicus.eu> [Acesso em maio 2024].
- [134] Copernicus Climate Change Service, "CERRA - Copernicus Regional Reanalysis for Europe," [Online]. Disponível: <https://cds.climate.copernicus.eu/>. [Acesso em maio 2024].
- [135] Atlas Eólico: Bahia. [Online]. Disponível: <http://www.infraestrutura.ba.gov.br/> [Acesso em maio 2024].