



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
CURSO DE ENGENHARIA ELÉTRICA

GABRIEL EUGÊNIO DE AGUIAR SILVEIRA

**APRENDIZADO DE MÁQUINA APLICADO À PREDIÇÃO DE POTÊNCIA DE
GERAÇÃO DISTRIBUÍDA NA REDE DE DISTRIBUIÇÃO DE MÉDIA TENSÃO**

Fortaleza

2021

GABRIEL EUGÊNIO DE AGUIAR SILVEIRA

APRENDIZADO DE MÁQUINA APLICADO À PREDIÇÃO DE POTÊNCIA DE GERAÇÃO
DISTRIBUÍDA NA REDE DE DISTRIBUIÇÃO DE MÉDIA TENSÃO

Trabalho de Conclusão de Curso apresentado à
Universidade Federal do Ceará, como requisito
parcial à obtenção do título de Bacharel em
Engenharia Elétrica. Orientador: Prof. Ruth Leão
Pastora, PhD. Coorientador: Eng. Camilo Martins
Castelo Branco Camurça, Esp.

Fortaleza

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

S588a Silveira, Gabriel Eugênio de Aguiar.

Aprendizado de máquina aplicado à predição de potência de geração distribuída na rede de distribuição de média tensão / Gabriel Eugênio de Aguiar Silveira. – 2021.

59 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Tecnologia, Curso de Engenharia Elétrica, Fortaleza, 2021.

Orientação: Profa. Dra. Ruth Leão Pastora.

Coorientação: Prof. Esp. Camilo Martins Castelo Branco Camurça.

1. Geração distribuída. 2. Aprendizado de máquina. 3. Inteligência artificial . 4. Setor elétrico. 5. Algoritmos. I. Título.

CDD 621.3

GABRIEL EUGÊNIO DE AGUIAR SILVEIRA

APRENDIZADO DE MÁQUINA APLICADO À PREDIÇÃO DE POTÊNCIA DE GERAÇÃO
DISTRIBUÍDA NA REDE DE DISTRIBUIÇÃO DE MÉDIA TENSÃO

Trabalho de Conclusão de Curso apresentado à
Universidade Federal do Ceará, como requisito
parcial à obtenção do título de Engenheiro
Eletricista.

Aprovada em: 06/04/2021.

BANCA EXAMINADORA

Prof. PhD. Ruth Leão Pastora (Orientador)
Universidade Federal do Ceará (UFC)

Eng. Esp, Camilo Martins Castelo Branco Camurça (Coorientador)
Enel Distribuição Ceará

Prof. MSc. Lucas Silveira Melo Universidade Federal do
Ceará (UFC)

Eng. André Wagner de Barros Silva Universidade Federal
do Ceará (UFC)

Agradecimentos

Agradeço primeiramente a Deus, por ter me dado saúde, força e discernimento para superar todas as dificuldades.

Agradeço à minha família por ter me apoiado de diversas maneiras durante minha graduação, em especial minha mãe Zuila Aguiar e meu pai Giordane Silveira.

Agradeço especialmente minha namorada Amanda Nojoza por ter me ajudado tanto com ideias e sugestões na confecção do trabalho.

Agradeço a todo corpo docente do Departamento de Engenharia Elétrica que, além de formar exímios engenheiros, há também o investimento na formação do cidadão. Faço também um agradecimento especial à professora PhD Ruth Leão que se dispôs a me ajudar e orientar nesse desafio sendo sempre disponível e solícita.

Por fim, agradeço a toda equipe do Planejamento da Rede AT/MT da Enel Distribuição Ceará, por toda disponibilidade em me ensinar e compartilhar desafios.

“A menos que modifiquemos a nossa maneira de pensar, não seremos capazes de resolver os problemas causados pela forma como nos acostumamos a ver o mundo.”

Albert Einstein

RESUMO

Este trabalho apresenta um modelo, que utiliza técnicas de Aprendizado de Máquina, capaz de estimar a capacidade disponível de potência para inserção de Geração Distribuída (GD) em qualquer ponto da rede de distribuição de energia elétrica do Ceará. Foram desenvolvidos algoritmos baseados em quatro modelos distintos de aprendizado de máquina, Floresta Randômica, Máquina de Vetor de Suporte, Redes Neurais Artificiais e XGBoost. Os algoritmos foram testados e a partir da aplicação de métricas de desempenho. Foi selecionado o modelo XGBoost. Os dados de entrada do algoritmo desenvolvido foram obtidos a partir de informações disponibilizadas pela empresa Enel Distribuição Ceará referentes a estudos técnicos realizados pela área de Planejamento da Rede AT/BT e dados operacionais da rede elétrica de concessão da distribuidora. Dada a grande demanda por conexão de GDs nos últimos anos, a contribuição do trabalho é oferecer ao acessante informação transparente sobre a capacidade disponível da rede elétrica para conexão de GD sem violação de restrições operativas. Além de apresentar o diagnóstico preditivo da capacidade da rede para conexão de GDs, o trabalho contribui para a melhor condução das solicitações de acesso à rede para ambos, acessante e distribuidora. A taxa de acerto da predição dada pelo algoritmo foi avaliada para diferentes classes de clientes predominantes em alimentador: residencial, comercial, industrial e rural. A maior taxa de acerto foi encontrada em alimentadores industriais e a menor em alimentadores rurais.

Palavras chaves: Geração Distribuída 1. Aprendizado de Máquina 2. Inteligência Artificial 3. Setor Elétrico 4. Algoritmos 5. Previsão 6.

ABSTRACT

This paper presents a model, which uses Machine Learning techniques, capable of estimating the available power capacity for insertion of Distributed Generation (DG) in any point of the electric power distribution network in Ceará. Algorithms were developed based on four different machine learning models, Random Forest, Support Vector Machine, Artificial Neural Networks and XGBoost. The algorithms were tested and from the application of performance metrics the XGBoost model was selected. The input data of the developed algorithm were obtained from information provided by the company Enel Distribuição Ceará regarding technical studies carried out by the HV / LV Network Planning area and operational data from the distributor's concession electric network. Given the great demand for DG connections in recent years, the contribution of the work is to provide the system user with transparent information about the available capacity of the electrical network for DG connection without violating operational restrictions. In addition to presenting the predictive diagnosis of the network's capacity for connecting DGs, the work contributes to the better handling of network access requests for both accessor and distributor. The accuracy rate of the prediction given by the algorithm was evaluated for different classes of customers predominant in feeder: residential, commercial, industrial and rural. The highest hit rate was found in industrial feeders and the lowest in rural feeders.

Key words: 1. Distributed Generation. 2. Machine Learning. 3. Artificial Intelligence. 4. Electrical Sector. 5. Algorithms. 6. Prediction.

LISTA DE FIGURAS

Figura 1 - Capacidade instalada de geração solar no mundo de 2009 até 2019.....	14
Figura 2- Capacidade instalada de geração eólica no mundo de 2009 até 2019.....	15
Figura 3 - Quantidade anual de conexões no Brasil de GD entre os anos de 2013 a 2020	16
Figura 4 - Quantidade de conexões de GD realizadas no Ceará de 2013 até 2020.	17
Figura 5 - Solicitações de GD feitas à Enel Distribuição Ceará entre 2018 e 2020...	18
Figura 6 - Participação das principais fontes na capacidade instalada de 2005 a 2030.	23
Figura 7 - Potência instalada de GD em megawatt por ano no Brasil até 2020	24
Figura 8 - Representação de uma Árvore de Decisão e a denominação de suas partes.	28
Figura 9 - Representação de separação de duas classes por hiperplano linear.	29
Figura 10 - Representação modelo neural básico.	31
Figura 11 - Exemplo de representação do modelo GMM	36
Figura 12 - Representação dos estudos que foram considerados outliers, a figura evidencia qual Subestação é referente ao valor fora do padrão.....	37
Figura 13 - Correlação dos Dados de Entrada plotadas em um mapa de calor.	39
Figura 14 - Correlação dos dados de entrada dispostos em gráfico de barras.	40
Figura 15- Escores de correlação dos dados após a regressão das variáveis entre de entrada e saída	42
Figura 16 - Representação da busca dos melhores hiperparâmetros do modelo Random Forest a fim minimizar os erros.	45
Figura 17 - Representação da busca dos melhores hiperparâmetros do modelo Support Vector Machine a fim minimizar os erros.....	46
Figura 18 - Representação da busca dos melhores hiperparâmetros do modelo Redes Neurais Artificiais a fim minimizar os erros.....	46
Figura 19 - Representação da busca dos melhores hiperparâmetros do modelo Extreme Gradient Boosting a fim minimizar os erros.....	47
Figura 20 - Representação gráfica dos valores previstos pelo algoritmo Random Forest e os valores originais.....	48
Figura 21 - Representação gráfica dos valores previstos pelo algoritmo Support Vector Machine e os valores originais.	48

Figura 22 - Representação gráfica dos valores previstos pelo algoritmo Redes Neurais Artificiais e os valores originais.	49
Figura 23 - Representação gráfica dos valores previstos pelo algoritmo XGBoost e os valores originais.	49
Figura 24 - Representação dos tempos de treinamento de cada modelo.	50
Figura 25 - Interface do Interplan: software utilizado pela distribuidora, para a realização dos estudos de fluxo de potência.	51
Figura 26 - Representação gráfica do nível de acerto do algoritmo da predição da potência.....	52
Figura 27 - Arquivo html mostrando a quantidade de barras por região.....	54
Figura 28 - Resultado da navegação do usuário pelo arquivo protótipo	55

LISTA DE TABELAS

Tabela 1 - Resumo das métricas dos modelos.	47
--	----

LISTA DE SIGLAS E ABREVIações

REN	Resolução Normativa
TW	Terawatt
ANEEL	Agência Nacional de Engenharia Elétrica
GD	Geração Distribuída
GW	Gigawatt
MW	Megawatt
kW	Kilowatt
FIEC	Federação das Indústrias do Estado do Ceará
AT	Alta Tensão
MT	Média Tensão
BT	Baixa Tensão
CAPEX	<i>Capital Expenditure</i>
ML	<i>Machine Learning</i>
IA	Inteligência Artificial
SEP	Sistema Elétrico de Potência
RNA	Redes Neurais Artificiais
RF	<i>Random Forest</i>
SVM	<i>Support Vector Machine</i>
MLP	<i>Multilayer Perceptron</i>
GMM	<i>Gaussian Mixture Model</i>
MAE	<i>Mean Absolut Error</i>
MSE	<i>Mean Squared Error</i>
MRT	Monofilar com Retorno por Terra
ms	Milissegundo
html	<i>HyperText Markup Language</i>
EPE	Empresa de Pesquisa Energética
Icc3Φ	Curto-Circuito Trifásico Simétrico
FA	Floresta Aleatória
AM	Aprendizado de Máquina
GS	<i>Grid Search</i>
BP	<i>British Pretroleum</i>

SUMÁRIO

1. INTRODUÇÃO	14
1.1 Motivação	18
1.2 Objetivos Gerais	19
1.3 Objetivos Específicos	19
1.4 Estrutura do Trabalho	20
2. REVISÃO BIBLIOGRÁFICA.....	21
2.1 Sistema Elétrico Brasileiro	21
2.2 Geração Distribuída.....	23
3. ALGORITMOS DE APRENDIZADO DE MÁQUINA E FERRAMENTAS COMPUTACIONAIS.....	26
3.1 Inteligência Artificial e Aprendizado de Máquina (<i>Machine Learning</i>).....	26
3.2 Floresta Aleatória (FA).....	27
3.3 Máquina de Vetor de Suporte (Support Vector Machine)	28
3.4 Redes Neurais Artificiais (RNA).....	30
3.5 <i>Extreme Gradient Boosting</i> (XGBoost)	31
3.6 Python e R.....	32
4. METODOLOGIA	33
4.1. Coleta e Estruturação dos Dados	33
4.2. Cálculo de novos parâmetros	35
4.3. Detecção de <i>Outliers</i>	35
4.4 Normalização dos Dados.....	38
4.5 Escolha dos Parâmetros Previsores.....	38
4.5.1 <i>Correlação dos Dados</i>	39
4.5.2 <i>Seleção das Variáveis</i>	41
4.6. Seleção Hiperparâmetros dos Modelos.....	43

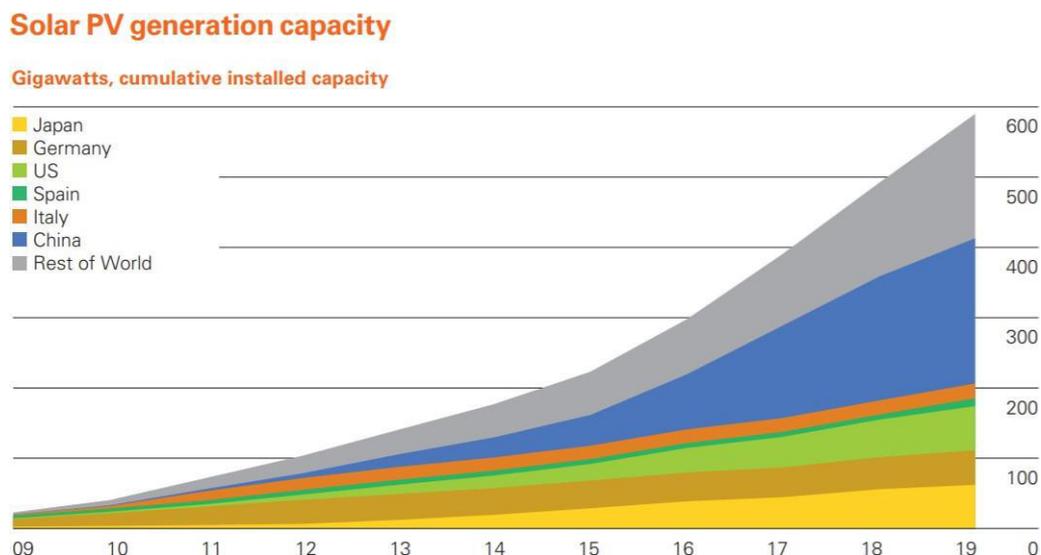
4.6.1	Grid Search	43
4.7.	Métricas de Desempenho	44
4.7.1	<i>Erro Médio Absoluto e Erro Médio Quadrático</i>	45
4.7.2	<i>Tempo de Treinamento</i>	50
5.	RESULTADOS	51
5.2	Protótipo	53
6.	CONCLUSÃO	55

1. INTRODUÇÃO

O setor elétrico tem passado por uma série de mudanças como avanços nas tecnologias de geração e de informação e comunicação, aumento na demanda de energia elétrica, novas medidas regulatórias do setor elétrico, a exemplo das Resoluções Normativas (REN) 482/2012 e 687/2015, mitigação dos impactos ambientais e restrições sobre a construção de novas linhas de transmissão. Todos esses fatores contribuem para a inserção e crescimento da Geração Distribuída (GD) como insumo renovável predominantemente (OLAMAIE; NIKNAM, 2006, p. 1).

Dados apresentados pela empresa *British Petroleum* (BP) no documento *Statistical Review of World Energy* (2020) sobre mercado de energia no mundo, mostram que as fontes renováveis mais populares no mundo são solar fotovoltaica e eólica, que somadas perfazem um total de aproximadamente 1,2 TW de capacidade instalada no mundo. Além disso, verifica-se que há uma tendência de crescimento para os próximos anos, como pode ser observado nas Figuras 1 e 2.

Figura 1 - Capacidade instalada de geração solar no mundo de 2009 até 2019.

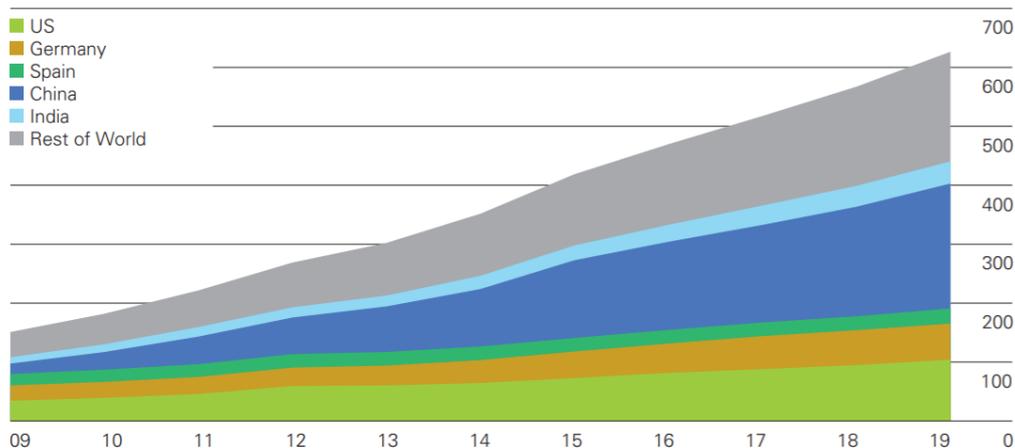


Fonte: [Statistical Review of World Energy, 2020]

Figura 2- Capacidade instalada de geração eólica no mundo de 2009 até 2019.

Wind generation capacity

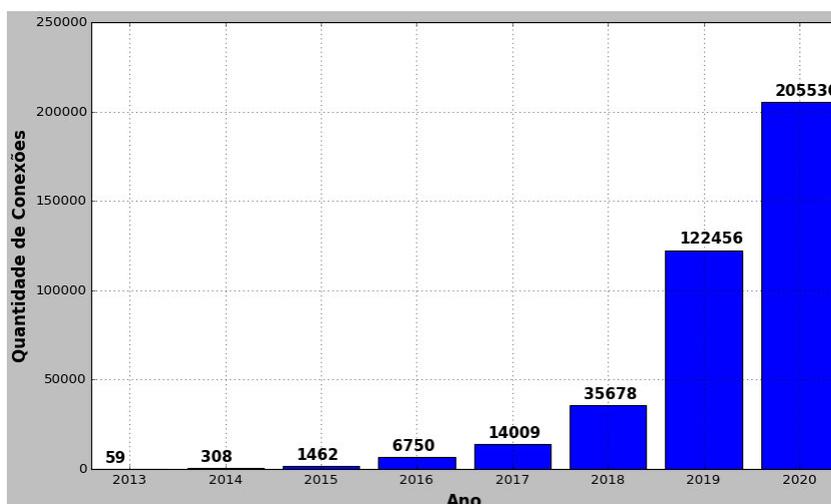
Gigawatts, cumulative installed capacity



Fonte: [Statistical Review of World Energy, 2020]

No Brasil, o cenário não é diferente. De acordo com a Agência Nacional de Engenharia Elétrica (ANEEL), a GD ultrapassou a marca superior a 5 GW em 2020. Tal marca foi alcançada devido ao expressivo potencial das fontes renováveis no Brasil, e o mesmo prevê que o crescimento será vigoroso nos próximos anos. As instalações de GD no Brasil cresceram de 14.009 unidades instaladas em 2017 para 35.678 em 2018, fechando o ano de 2020 com 205.530 unidades.

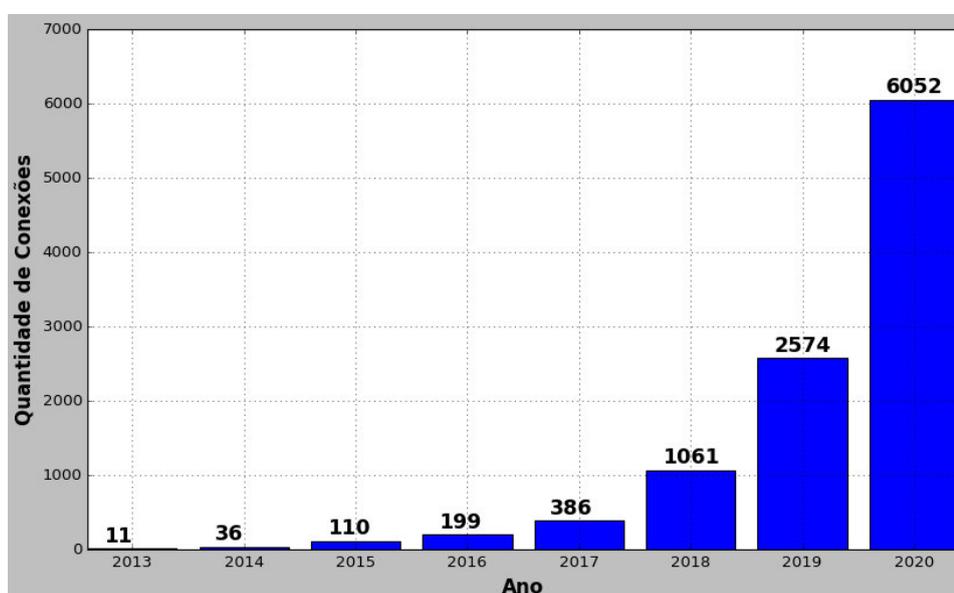
Figura 3 - Quantidade anual de conexões no Brasil de GD entre os anos de 2013 a 2020.



Fonte: [ANEEL, 2021]

Segundo a Federação das Indústrias do Estado do Ceará (FIEC), o Ceará possui grande potencial energético tanto solar quanto eólico devido as condições climáticas favoráveis. A exploração desse potencial tem se intensificado, alinhado com o que vem ocorrendo no Brasil e no mundo. Segundo a ANEEL, a potência instalada de GDs no Ceará é superior a 195 GW, perfazendo um total de 11.829 usinas até junho/2020. O número de conexão de GD tem crescido exponencialmente, com pode ser observado na Figura 4.

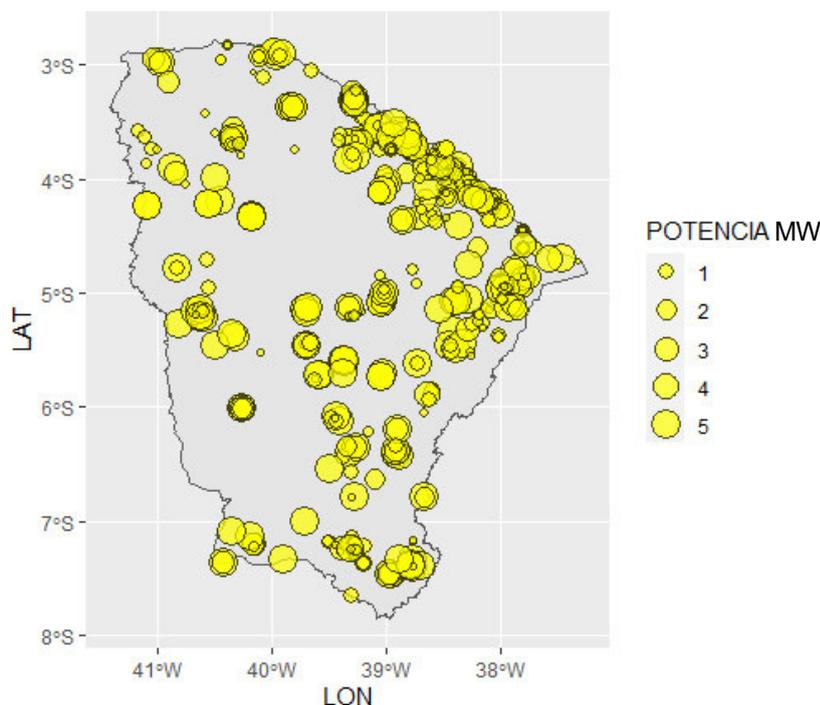
Figura 4 - Quantidade de conexões de GD realizadas no Ceará de 2013 até 2020.



Fonte: [ANEEL, 2021]

Na Figura 5, pode-se observar os pedidos de acesso realizados a Enel Distribuição Ceará, mais especificamente à área de estudos técnicos de Alta e Média Tensão: Planejamento da Rede AT/MT. Nessa figura observa-se a localização e potência escolhidas e solicitadas pelos clientes.

Figura 5 - Solicitações de GD feitas à Enel Distribuição Ceará entre 2018 e 2020.



Fonte: [Próprio Autor]

O acentuado montante de solicitação de conexão de GDs tem causado certa dificuldade às empresas distribuidoras para lidar com esse cenário, fruto das transformações no setor elétrico. Por outro lado, do ponto de vista dos clientes, nunca foi tão atrativo investir em GD desde a publicação das Resoluções Normativas tanto a 482/2012 quanto a 687/2015, pois reduziu a barreira de um pequeno consumidor se conectar à rede (GOMES, 2018, p. 1-2). Além disso, incentivos fiscais e ambientais têm corroborado para a diminuição do custo da tecnologia, principalmente dos empreendimentos fotovoltaicos, a modalidade mais utilizada. A redução do custo foi o catalisador da evolução da GD, possibilitando o aumento na viabilidade técnico-econômica de implantação das usinas (PINHO; GALDINO, 2014, p 3-5).

Esse rápido crescimento tem originado grande demanda nos processos internos das distribuidoras: processos regulatórios, avaliação, projeto, instalação e monitoramento. Um dos pontos de maior atenção desse processo é a emissão de Informações e Pareceres de Acesso, que são documentos emitidos mediante um estudo de viabilidade técnico-econômica da inserção dessa geração no sistema elétrico da concessionária. Porém, a alta demanda de pedidos de conexão tem gerado certo transtorno, desde a entrada dos documentos pelo acessante até a

emissão de parecer pela concessionária, ressaltando que o ponto crítico nesse processo são os estudos, atividade que demanda maior especialização técnica, por isso, onera mais tempo nesse fluxo.

O elevado fluxo de pedidos de conexão, tem demandado esforço e tempo de todo corpo técnico maior que o desejado com a análise do reflexo da conexão de GDs à rede. Além disso, de uma forma geral, não existe uma cultura entre as distribuidoras de disponibilizarem dados da rede elétrica de forma proativa aos clientes. Tal fato, agrava a situação, pois normalmente não há informações disponíveis sobre a capacidade da rede. Fica ao encargo do cliente submeter inúmeros pedidos de análise de GD, a fim de descobrir se existe ou não viabilidade em determinado ponto de conexão pretendido. Isto leva a solicitações desnecessárias ou redundantes, em pontos normalmente inviáveis do ponto de vista de mínimo custo global, gerando sobrecarga ainda maior ao processo.

As distribuidoras, com as escolhas aleatórias dos acessantes por ponto viável de conexão, têm alocado seu capital de investimento, o *Capital Expenditure* (CAPEX), para a conexão dessas gerações, que se torna menos rentável para a empresa e para a população. Para a distribuidora é mais interessante alocar seu CAPEX para a conexão de novas unidades consumidoras tipo carga. Já do ponto de vista da sociedade, os custos decorrentes do processo de análise e reanálise das solicitações de acessantes incorrem sobre a tarifa, não sendo justo para clientes não prosumidores.

Visando beneficiar as partes interessadas envolvidas no contexto de GDs, distribuidora, acessante e sociedade, este trabalho de caráter técnico-computacional, estabelece uma metodologia de previsão da capacidade de conexão de GD sem a necessidade de intervenções, sobre qualquer ponto da rede elétrica, por meio de algoritmo de aprendizado de máquina, tomando como base estudos realizados por engenheiros especialistas do Planejamento da Rede da Enel Distribuição Ceará e dados operacionais do sistema elétrico da concessionária.

1.1 Motivação

Diante do crescimento exponencial da geração distribuída e a tendência de assim se manter nos próximos anos, as empresas distribuidoras de energia

elétrica se veem cada vez mais imersas na tarefa de realizar estudos e análise dos impactos de GDs à rede elétrica. A falta de conhecimento do acessante sobre os pontos de conexão mais adequados tem contribuído no aumento de solicitações pelos acessantes, uma vez que buscam um ponto de conexão que não acarrete em necessidade de intervenção na rede elétrica. Em que cerca de 69% dos pedidos de estudos de viabilidade técnica, realizado entre 2018 e 2020, para GD, tem o caráter prospectivo, ou seja, para identificar se há ou não viabilidade econômica por parte do empreendedor.

Após o cenário apresentado anteriormente sobre crescimento de GDs, as dificuldades enfrentadas pelas empresas distribuidoras em atender com rapidez e qualidade os acessantes, a complexidade nas análises, e por fim, o desejo dos acessantes de informações técnicas sobre a rede de distribuição da concessionária, a principal motivação deste trabalho foi identificar que o crescimento das GDs poderia se tornar uma oportunidade de apresentar uma proposta inovadora com o intuito de contribuir na condução da inserção otimizada da GD na rede elétrica.

1.2 Objetivos Gerais

O objetivo geral deste trabalho é desenvolver um algoritmo baseado em inteligência computacional, que faz uso de dados fornecidos pela Enel Distribuição Ceará, capaz de prever a capacidade de inserção de GD em cada ponto do sistema elétrico de concessão da distribuidora. Com o auxílio do algoritmo computacional, melhores pontos poderão ser escolhidos, fazendo com que a GD cumpra suas funções, beneficiando à rede elétrica e melhorando os processos internos da empresa, postergando intervenções na rede elétrica e a relação com os clientes.

1.3 Objetivos Específicos

- Determinar as variáveis que mais sensibilizem o algoritmo baseado em aprendizado de máquina e os dados operacionais da rede que serão usados no treinamento do algoritmo.
- Determinar o melhor modelo de aprendizado de máquina para a solução do problema.
- Determinar quantitativamente e qualitativamente a precisão e

performance
dos algoritmos de *Machine Learning* (ML).

1.4 Estrutura do Trabalho

O presente trabalho está estruturado em seis capítulos. No Capítulo 1 é apresentado a motivação e objetivos do trabalho. No Capítulo 2 é feita uma revisão bibliográfica abordando alguns temas centrais: Setor Elétrico e GD, tendo como principal objetivo apresentar estudos de outros autores e suas contribuições, apontando semelhanças, diferenças e contribuições com esse trabalho apresentado. No Capítulo 3, os conceitos de Inteligência Artificial (IA) e ML são apresentados, bem como alguns algoritmos de ML, métodos matemáticos e ferramentas computacionais, com o intuito de fornecer uma visão geral teórica, para embasamento da metodologia adotada neste trabalho. No Capítulo 4 é descrita a metodologia e os modelos desenvolvidos. No Capítulo 5 são apresentados os resultados dos modelos. Por fim, o Capítulo 6 apresenta a conclusão.

2. REVISÃO BIBLIOGRÁFICA

2.1 Sistema Elétrico Brasileiro

A energia elétrica representa um fator determinante para o desenvolvimento econômico e social dos seres humanos. O grau de desenvolvimento de um país pode ser avaliado pela infraestrutura disponível que permite o acesso de toda a população à energia elétrica (ANEEL, 2008).

Silva (2011, p. 6) divide o desenvolvimento das questões energéticas no Brasil em sete fases, porém, para o estudo em pauta os mais relevantes são o surgimento do uso da eletricidade no Brasil em 1887 até o final da República Velha em 1930; e a reestruturação do setor elétrico iniciada em 1996 que permanece até os dias atuais.

As reformas ocorridas desde a década de 1990 possuem impacto significativo no desenvolvimento do setor até a configuração atual. Esse período foi marcado pelas privatizações das concessionárias, processo iniciado no Brasil em 1996. Os principais objetivos dessa nova configuração foram promover investimentos do setor privado para expandir o sistema elétrico (SILVA, 2011, p. 81).

A fim de cumprir sua principal função de suprir a demanda dos consumidores, sejam eles residenciais, comerciais, rurais ou industriais, o Sistema Elétrico de Potência (SEP) transmite a energia elétrica da fonte geradora por meio das linhas de transmissão e distribuição, elevando e abaixando os níveis de tensão por meio das subestações de transformação até alcançar o usuário final (MOHAN, 2016).

A configuração tradicional e ainda predominante nos SEPs é denominada de geração centralizada. A geração centralizada é caracterizada por grandes centrais de produção de energia elétrica, que no Brasil, majoritariamente, são usinas hidrelétricas ou termelétricas com grandes capacidades instaladas. Essas usinas localizam-se longe dos grandes centros de consumo, necessitando de longas e custosas linhas de transmissão (MARTÍN-MARTÍNEZ, 2017, p. 1-3).

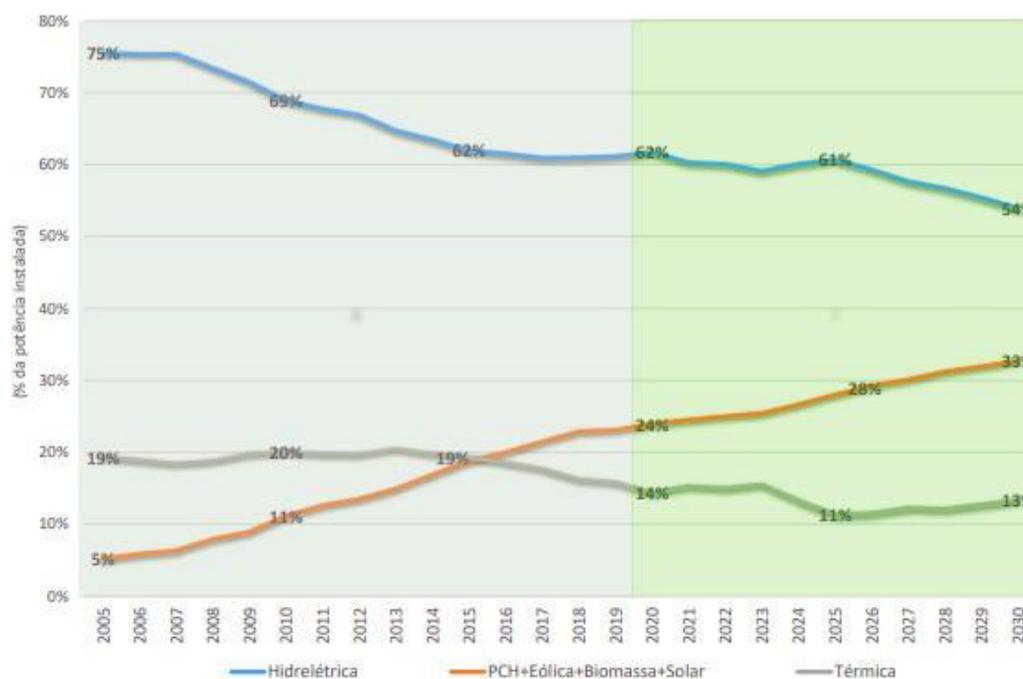
Essa configuração é predominante até hoje, porém até metade do século XX ela era quase hegemônica. Muitos fatores que explicam o sucesso até hoje dessa centralização ou configuração tradicional: economias de escala em

instalações de usinas de geração de grande capacidade instalada; viabilidade técnica e econômica de transportar grandes quantidades de energia a longas distâncias, menor necessidade de investimento em geração e maior confiabilidade por meio do compartilhamento de unidades de geração em diferentes regiões (MARTÍN-MARTÍNEZ, 2017, p. 1-3).

No entanto, este paradigma tradicional foi perturbado pela inserção paulatina de GD. Um catalisador dessa inserção foi devido à adesão de muitos governos em investir em tecnologias renováveis, incluindo cogeração, para a luta contra as mudanças climáticas (MARTÍN-MARTÍNEZ, 2017, p. 2). No Brasil e no mundo, outros dois fatores intensificaram esse movimento: incentivos fiscais e diminuição do custo da tecnologia.

A Figura 6 é mostrado a crescente relevância das principais fontes GD no Brasil: eólica e solar, as quais vem aumentando sua participação na matriz energética brasileira desde 2005, e segundo a previsão da Empresa de Pesquisa Energética (EPE), essas fontes continuarão aumentando sua participação até 2030, como mostra o Plano Decenal 2030.

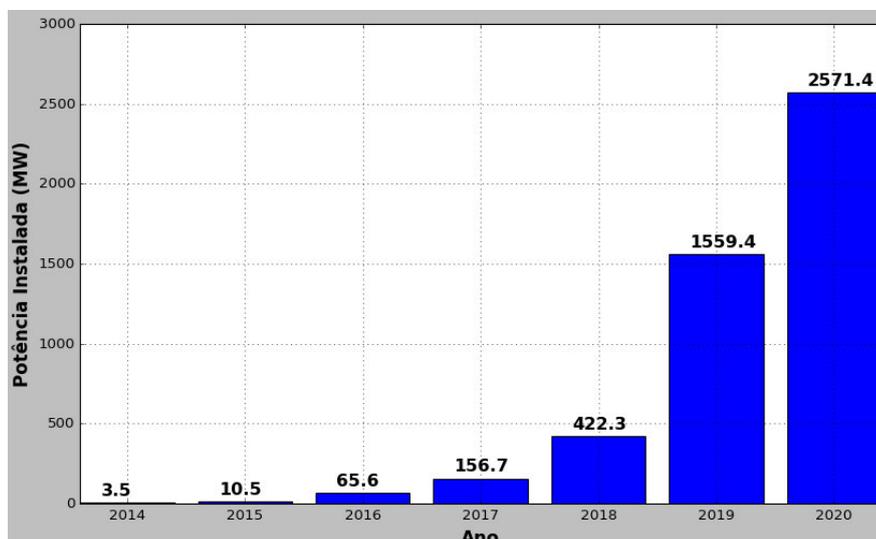
Figura 6 - Participação das principais fontes na capacidade instalada de 2005 a 2030.



Fonte: [Plano Decenal de Expansão de Energia 2030, 2020]

De acordo com dados públicos fornecidos pela ANEEL, verifica-se na Figura 7 a forte evolução da potência instalada de geração distribuída no Brasil

Figura 7 - Potência instalada de GD em megawatt por ano no Brasil até 2020.



Fonte: [ANEEL, 2020]

2.2 Geração Distribuída

De acordo com anexo da Resolução Normativa N° 674/2015, de 11 agosto de 2015, GD são:

Centrais geradoras de energia elétrica, de qualquer potência, com instalações conectadas diretamente no sistema elétrico de distribuição ou através de instalações de consumidores, podendo operar em paralelo ou de forma isolada e despachadas - ou não - pelo Operador Nacional do Sistema Elétrico - ONS.

Em contraponto da geração centralizada, a GD tem por característica principal ser instalada próximo aos usuários finais, fornecendo energia elétrica. Os benefícios da geração distribuída podem ser de natureza econômica, técnica e ambiental, e dependem do tipo de tecnologia, da capacidade e da localização da GD (HA; HUY; RAMACHANDARAMURTHY, 2017, p. 1).

A inserção de GD pode ser muito benéfica tanto ao sistema elétrico quanto aos consumidores, mitigando alguns problemas inerentes ao sistema de distribuição radial: tensões baixas devido à queda de tensão ao longo dos alimentadores de distribuição, redução de perdas técnicas ao diminuir a circulação de corrente no alimentador uma vez que a GD está próxima à carga, bem como a

prevenção de altos custos de investimentos para atualizações da rede elétrica. (KAYAL; CHANDA, 2013, p. 1).

Além da localização e potência da GD, outra característica que se mostra decisiva na viabilidade da inserção de GD é a modalidade escolhida pelo solicitante no que tange ao modelo de negócio de geração, que pode ser compartilhado, autoconsumo remoto e condomínio solar (Res.no.687/2015, ANEEL). Dentre essas modalidades, a que mais tem influência é a de autoconsumo remoto, que segundo o órgão regulador, possibilita que os créditos de energia gerados por um sistema fotovoltaico em um local X possam ser usados para reduzir o valor da conta de luz de um local Y completamente diferente, desde que dentro da rede da mesma distribuidora. O maior impacto na viabilidade técnico-econômica é devido a ausência, em muitas vezes, de carga no local da instalação da usina, pois o há a ocorrência de fluxo reverso, ou seja, corrente elétrica no sentido da carga para a fonte. Já no caso de haver uma carga no local da GD, haverá o consumo local dessa geração, diminuindo a possibilidade de fluxo reverso, além de diminuir nas perdas técnicas no alimentador de distribuição. Esses benefícios e malefícios serão mais ou menos agravados dependendo da magnitude da GD.

Assim, as GDs desempenham papel importante nos sistemas de energia em todo o mundo, com impacto significativo na confiabilidade do sistema, no fluxo de energia, no perfil de tensão, na estabilidade e na qualidade de energia. GDs baseadas em fontes de energia renovável podem aliviar a pobreza energética, utilizando eólica e solar fotovoltaica disponíveis na região (KAYAL; CHANDA, 2013, p. 1).

Para obter os benefícios com a inserção de GDs no sistema de distribuição, é importante verificar a localização e a potência da GD (KAYAL; CHANDA, 2013, p. 1-2).

KAYAL e CHANDA (2013, p. 7-8) construíram uma metodologia baseada em redes neurais artificiais a fim de alocar de forma ótima e com a potência adequada uma GD no sistema de distribuição. Os autores concluíram que uma GD inserida na rede elétrica com tais requisitos atendidos, obtém-se a melhora no perfil de tensão tanto no alimentador quanto no barramento, além disso, constataram que uma GD instalada em um local não otimizado acarreta no aumento das perdas técnicas.

O modelo e o algoritmo aqui abordados também visam determinar a alocação adequada para GDs de modo que sua inserção resulte em benefícios e não incorra em problemas para a rede ou mesmo em necessidade de investimentos para correção dos problemas decorrentes da conexão. Diferenças entre o trabalho aqui proposto e o de Kayal e Chanda (2013) começa na motivação. Para Kayal e Chanda (2013), a motivação do estudo estava relacionado aos impactos na qualidade da energia (tensão e perdas) devido a inserção de GDs na rede. Os autores usaram estudos de fluxo de carga, testando de forma massiva, diferentes pontos de conexão de GDs e avaliando o impacto sobre o perfil de tensão. O resultado dessas avaliações de fluxo de potência foi armazenado e posteriormente treinados por uma Rede Neural Artificial (RNA), de forma que a saída do modelo computacional fosse uma potência que cumprisse duas premissas: que o perfil de tensão no local da GD de pelo menos 0,95 p.u. e que a tensão em todas as outras barras não fossem inferiores a 0,90 p.u.

No presente trabalho, a principal motivação é realizar a previsão da potência máxima de GD sem a necessidade de adequações. Potência gerada de forma rápida e automática através de algoritmo de aprendizado de máquina sem a necessidade da realização de um estudo de viabilidade.

Outra diferença, a principal, foi na metodologia aplicada, apesar de ambos os estudos utilizarem dados como fontes de entrada e métodos utilizando inteligência computacional. O trabalho de KAYAL e CHANDA (2013) utiliza como base para as Redes Neurais Artificiais (RNA) treinarem os dados de tensão e a potência da GD. Já no trabalho aqui apresentado, tanto dados operacionais da rede elétrica como tensão, corrente, perdas técnicas, curto-circuito, dentre outros, quanto às 640 solicitações dos clientes, estudadas pelos engenheiros especialistas da área, foram utilizados como entrada para treinamento do modelo.

3. ALGORITMOS DE APRENDIZADO DE MÁQUINA E FERRAMENTAS COMPUTACIONAIS

Este capítulo tem como objetivo apresentar quatro modelos distintos de Aprendizado de Máquina, a saber, Floresta Randômica, Máquina de Vetor de Suporte, Redes Neurais Artificiais e XGBoost. Os modelos foram aplicados na predição da potência disponível em diferentes pontos da rede elétrica de distribuição da Empresa Enel Distribuição Ceará.

3.1 Inteligência Artificial e Aprendizado de Máquina (*Machine Learning*)

Define-se Inteligência Artificial (IA) como qualquer dispositivo que percebe seu ambiente, aprenda e tome decisões que maximizam sua chance de sucesso em algum objetivo, após se especializarem, podendo ser através de dados, modelo mais comum, ou não (ABBAS; NASSER; EL AHMAD, 2015, p. 4).

Apesar de estudada desde os anos 50, a inteligência artificial (IA) progrediu bastante nas últimas décadas, e uma força motriz para esse desenvolvimento foram a junção de alguns fatores: aprimoramentos recentes em aprendizado profundo (*Deep Learning*), disponibilidade de grandes volumes de dados e melhoria nas tecnologias de *hardware*, possibilitando maior poder de processamento (SAMEK; WIEGAND; MÜLLER, 2017, p. 1-2).

Sendo uma porção da IA, o aprendizado de máquina, é um conjunto de métodos matemáticos e estatísticos associados a programação de computadores, a fim de solucionar problemas através de classificação, regressão e segmentação. Esse aprendizado é parametrizável, além disso, pode ser feito através de dados, treinamento supervisionado, ou de experiências passadas, chamado de treinamento por reforço (MOHRI, 2018, p. 1-2).

Os sistemas baseados em aprendizado de máquina, hoje se destacam em uma série de tarefas complexas como detecção de objetos em imagens e vídeos, compreensão de linguagens naturais e até o processamento de sinais da fala. O sucesso dos sistemas de IA, especialmente o aprendizado profundo, *Deep Learning*, mostram o caráter revolucionário desta tecnologia, que terá um grande impacto além do mundo acadêmico e também dará origem a mudanças disruptivas

nas indústrias e sociedades (HE, 2016, p 1-2; CHO, 2014, p. 1; SAMEK; WIEGAND; MÜLLER, 2017, p. 1).

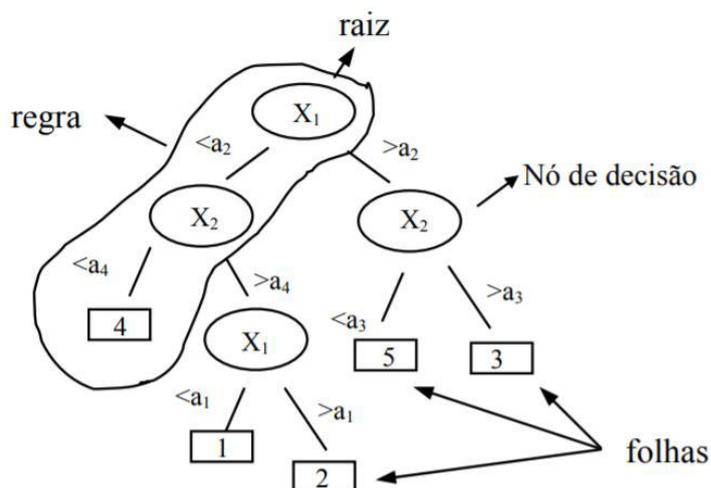
3.2 Floresta Aleatória (FA)

Para o melhor entendimento de Florestas Aleatórias (*Random Forest*), que são diversas árvores de decisão, a apresentação do conceito de Árvore de Decisão (*Decision Tree*) inicialmente se faz necessário.

Árvores de decisão são modelos estatísticos que utilizam um treinamento supervisionado para a classificação ou regressão dos dados a fim de obter novos dados. Os dados são divididos obedecendo regras definidas como parâmetros. Para GAMA (2004, p. 1-2), as árvores de decisão são modelos que utilizam a estratégia de dividir para conquistar. Um problema complexo será decomposto em subproblemas mais simples e recursivamente esta técnica é aplicada a cada sub-problema.

A Figura 8 apresenta os componentes que compõe uma árvore de decisão: raiz, nó, folhas e a regra. A raiz ou nó raiz, o qual é o começo da estrutura hierárquica da árvore, onde a primeira decisão é realizada e a árvore será dividida em subconjuntos, o nó é a parte em que as variáveis são divididas obedecendo uma regra, a folha é o último estágio, no qual é alcançada a resposta prevista. Por fim, a regra é a associação desses componentes em que cada observação dos dados será analisada com base nos critérios criados pela árvore. Desse procedimento uma resposta (folha) será alcançada (MOISEN, 2008, p. 2-4).

Figura 8 - Representação de uma Árvore de Decisão e a denominação de suas partes.



Fonte: [GAMA, 2004]

Feitas as considerações iniciais, viabiliza-se o aprofundamento no modelo de Floresta Aleatória (FA), um dos métodos utilizados neste trabalho. Esse modelo cria várias árvores de decisão, uma floresta, cada uma treinada em uma amostra inicial dos dados de treinamento, e pesquisa um subconjunto aleatoriamente das variáveis de entrada para determinar uma divisão para cada nó desses (GISLASON; BENEDIKTSSON; SVEINSSON, 2006, p. 2). Cada árvore gerada é única, e ao final da construção do modelo há uma grande variedade de resultados distintos. Assim, o uso das FAs tende a ser mais preciso, devido ser um voto de pluralidade (ponderada) ou média da previsão de todas as árvores da coleção (MOISEN, 2008, p. 6).

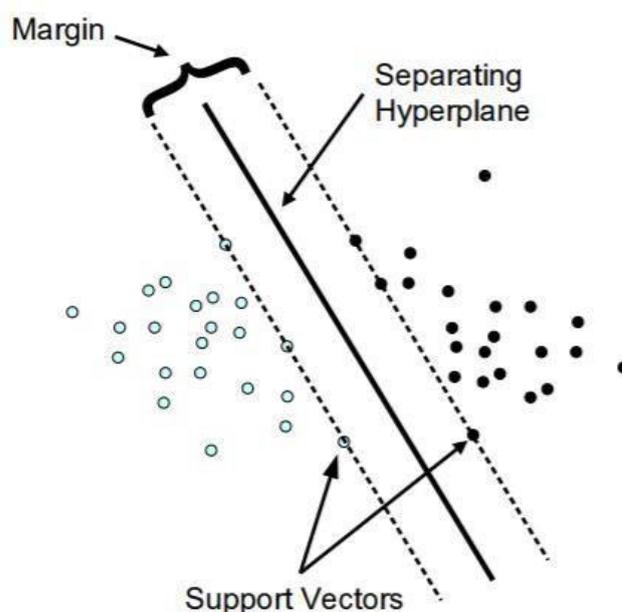
No caso do trabalho aqui apresentado, as árvores geradas entregam uma potência cada. No resultado final do modelo é utilizado a média dos resultados da potência que cada árvore gerou. A vantagem das FA sobre o modelo de árvore de decisão é que são mais precisas e robustas aos ruídos (RODRIGUEZ-GALIANO, 2012, p. 3-4).

3.3 Máquina de Vetor de Suporte (Support Vector Machine)

A máquina de vetores de suporte (SVM – *Support Vector Machine*) implementa a seguinte ideia: mapeia o vetor de entradas, com “n” variáveis, em um espaço de recursos de dimensão z, por meio de algum mapeamento não linear.

Neste espaço, um hiperplano de separação ideal é construído, ou seja, com os dados plotados, será construído um plano para generalizar ao máximo os dados. A Figura 9 evidencia a descrição mencionada anteriormente.

Figura 9 - Representação de separação de duas classes por hiperplano linear.



Fonte: [MEYER, 2015]

Basicamente, procura-se o hiperplano de separação ideal entre as duas classes, maximizando a margem entre os pontos mais próximos das classes. Os pontos situados nos limites são chamados de vetores de suporte, e entre as margens haverá o ponto ótimo de separação dos dados, o hiperplano (MEYER, 2015, p. 1).

Quando não conseguimos encontrar um separador linear, os pontos de dados são projetados em um espaço de dimensão superior onde os pontos de dados efetivamente se tornam linearmente separáveis (MEYER, 2015, p. 1-2).

A descrição acima se referiu a um problema binário, em que a resposta do algoritmo decide entre duas variáveis: 0 e 1, por exemplo. O algoritmo SVM também pode ser aplicado a problemas de regressão, ou seja, em que a resposta do modelo seja uma variável contínua, como por exemplo 5,5. Para isso, basta adicionar função de perda. A função de perda deve ser modificada para incluir uma distância e medir, o resultado dessa medição é o valor numérico de resposta do modelo (GUNN, 1998, p. 29-30).

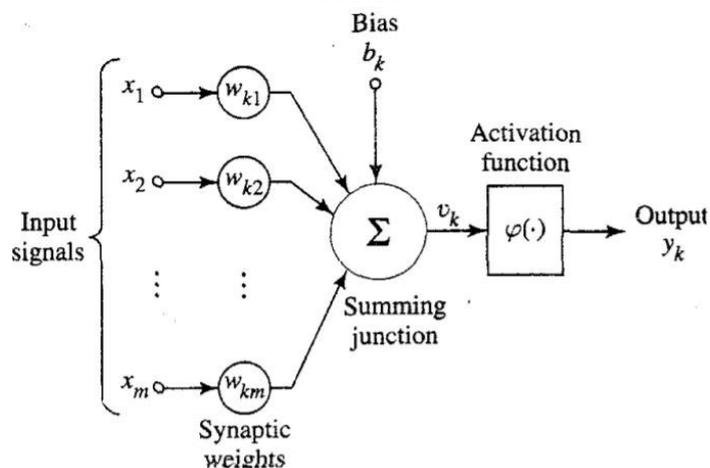
3.4 Redes Neurais Artificiais (RNA)

As Redes Neurais Artificiais foram selecionadas para o teste devido à complexidade do problema e seu poder de processamento elevado. As RNA são técnicas computacionais inferidas da estrutura neural fisiológica, e que adquirem conhecimento através de fatos e ensinamentos prévios, nesse caso, os dados (HAYKIN, 2007, p. 28-29).

O modelo de rede neural implementado foi o Perceptron de multicamadas (MLP – *Multlayer Perceptron*). A rede MLP é um algoritmo de aprendizado supervisionado, que aprende uma função não linear por treinamento de um conjunto de dados X de entrada, de dimensão m , para alcançar um alvo Y de saída, de dimensão k , como ilustrado na Figura 10. A função não linear mapeia um conjunto de entradas em um conjunto de dados de saídas (NETO; BONINI; BONINI, 2010, p. 3-4).

Uma MLP consiste em três tipos de camadas de nós: uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada camada tem uma função específica. A camada de saída recebe os estímulos da camada intermediária e constrói o padrão que será a resposta. As camadas intermediárias funcionam como extratoras de características, seus pesos são uma codificação de características apresentadas nos padrões de entrada e permitem que a rede crie sua própria representação, mais rica e complexa do problema.

Figura 10 - Representação modelo neural básico.



Fonte: [HAYKIN,2007]

O treinamento da rede MLP aplicada à solução do problema de estimação de potência disponível na rede para conexão de GDs pode ser descrito como: os sinais de entrada sendo as variáveis (parâmetros da rede elétrica), a junção aditiva realiza o somatório desses valores, aplica-se uma função de ativação e obtêm-se uma saída, a potência de GD (NETO; BONINI; BONINI, 2010, p. 3).

3.5 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) é uma biblioteca de código aberto que fornece uma implementação eficiente e eficaz do algoritmo de aumento de gradiente (*Gradient Boosting*).

Gradient Boosting se refere a uma classe de conjunto de algoritmos de aprendizado de máquina que podem ser usados para problemas de modelagem preditiva de classificação ou regressão.

Os conjuntos são construídos a partir da combinação entre dois modelos: árvore de decisão e de impulso (*Boosting*). Esse último modelo tem a capacidade de ponderar as variáveis de entrada de acordo com a força de sua correlação com a saída do algoritmo. Ou seja, parâmetros previsores mais promissores recebem um impulso, sendo assim, preponderantes nas decisões do modelo. As árvores são adicionadas uma de cada vez ao conjunto e ajustadas para corrigir os erros de previsão feitos pelos modelos anteriores. Esse é um tipo de modelo de aprendizado de máquina de conjunto conhecido como boosting. (DHALIWAL; NAHID; ABBAS, 2018, p 2-3).

Os algoritmos de árvore de decisão levam em consideração os atributos do conjunto de dados. Os atributos dos dados de entrada podem ser chamados de recursos ou colunas e, então, esses mesmos recursos atuam como o nó condicional ou o nó interno. Agora, correspondendo à condição no nó raiz, a árvore se divide em galhos ou arestas (DHALIWAL; NAHID; ABBAS, 2018, p 3-4).

Para ZHENG (2017, p. 1-2), o aumento do gradiente constrói as árvores aumentadas para obter de forma inteligente as pontuações de recursos, indicando assim a importância de cada recurso para o modelo de treinamento. Quanto mais um recurso é usado para tomar decisões importantes com árvores aumentadas, maior será sua pontuação.

3.6 Python e R

Diante de inúmeras linguagens de programação computacionais disponíveis atualmente, algumas possuem ferramentas que auxiliam e facilitam a construção de projetos na área em que se baseia o presente trabalho: Ciência de Dados (VANDERPLAS, 2016).

Python foi a linguagem selecionada por alguns motivos, tais como, é uma linguagem de alto nível, código aberto, fácil aprendizado, comandos enxutos e grande quantidade de bibliotecas. Essa linguagem possui diversas estruturas de dados (listas, *arrays*, tuplas, matrizes, dicionários, *dataframes*) que facilitam a manipulação de qualquer tipo de dado e grandes massas de dados. Além disso, possui inúmeras bibliotecas disponíveis para aprendizado de máquina e inteligência artificial. Por sua característica de código fonte aberto, gigantes da tecnologia (Google, Microsoft) criam diversos *frameworks* e disponibilizam para uso (BORGES, 2014, p. 14-16).

Além de python, foi utilizada a linguagem R, amplamente utilizada no meio acadêmico por sua facilidade na modelagem matemática e estatística, e também na construção de mapas. Alguns gráficos e mapas foram feitos nessa linguagem.

4. METODOLOGIA

A fim de mitigar os problemas da crescente conexão de GD, que têm contribuído para alta demanda nos processos internos das distribuidoras, foi desenvolvido um modelo de predição da potência máxima de GD sem que demande adequações na rede. O modelo baseia-se em algoritmos de aprendizado de máquina supervisionado, que usa dados históricos de estudos realizados por engenheiros especialistas da área de Planejamento da Rede AT/MT como insumos de treinamento.

O desenvolvimento do trabalho foi realizado de forma a identificar de maneira proativa e disponibilizar informações quanto à capacidade da rede elétrica de distribuição em absorção de GD.

Da problematização até a construção final do modelo, foram várias as etapas percorridas:

1. Coleta e estruturação dos dados: estudos realizados pela área e parâmetros operacionais da rede elétrica;
2. Cálculos de novos parâmetros a partir dos dados coletados;
3. Detecção de *outliers*;
4. Normalização dos dados;
5. Escolha dos parâmetros previsores;
6. Seleção dos Hiperparâmetros dos Modelos;
7. Métricas de Desempenho.

4.1. Coleta e Estruturação dos Dados

Nessa fase, ainda exploratória, foram coletados o maior número possível de parâmetros da rede de distribuição de Média Tensão, a fim de determinar quais parâmetros serão os mais promissores, ou seja, que possuem maior relação com a variável de saída.

Os dados referentes aos estudos feitos nos anos de 2018, 2019 e 2020 pela área de estudo de Planejamento da Rede foram de fácil acesso, estando bem estruturados e organizados. Porém os dados operacionais da rede, foram coletados em diversas bases de dados diferentes concentradas em uma única grande área:

Operação e Manutenção da Rede.

Os dados coletados foram armazenados e estruturados em uma única base, no formato de tabela (planilha) no formato .xlsx. A seguir são listados os dados coletados:

1. **ICC_15:** Curto-Circuito Trifásico Simétrico ($I_{cc3\Phi}$), em kiloampéres (kA) na barra MT (15kV) das Subestações de Distribuição (SED);
2. **ICC:** Curto-Circuito Trifásico Simétrico ($I_{cc3\Phi}$), em kA, no ponto solicitado pelo cliente para a conexão de sua GD;
3. **CARGA (DCC):** Corrente Máxima do Alimentador (A);
4. **TRAFOS:** Quantidade Total de Trafos no alimentador estudado;
5. **KM_MT:** Comprimento do alimentador MT em quilômetros;
6. **KM_BT:** Comprimento do alimentador BT em quilômetros;
7. **REL:** Quantidade de equipamentos de manobra: religadores e chaves seccionadoras com acionamento remoto por alimentador;
8. **kW_MTBT:** Demanda, em quilowatts (kW), dos transformadores de distribuição que atendem aos clientes BT;
9. **kW_MT:** Demanda, em quilowatts (kW), dos transformadores particulares que atendem aos clientes MT;
10. **kW_TOTAL:** Soma da Demanda de clientes BT e MT por alimentador;
11. **Perdas_Tecnicas:** Perdas Técnicas em kWh no período diário por alimentador;
12. **CientesBT:** Quantidade de Clientes BT;
13. **CientesMT:** Quantidade de Clientes MT;
14. **Carregamento (%):** Relação entre a corrente máxima e a ampacidade do cabo de saída do alimentador;
15. **P.U.:** Nível de corrente na sua curva característica diária em que ao meio-dia o carregamento se encontra, em relação ao máximo: 1 p.u.;
16. **VALOR:** Valor monetário das obras necessárias de adequação para a conexão da GD, valor esse previamente estudado

pelos engenheiros a fim de viabilizar a conexão de uma GD;

Todos esses dados foram coletados para cada estudo de conexão de GD associado. Ao final foi obtido uma base de dados, em que para cada estudo realizado as variáveis elétricas de cada ponto da rede estudado foram coletadas, a fim de buscar uma explicação causal da viabilidade de inserção de GD com os parâmetros elétricos intrínsecos à rede de distribuição.

4.2. Cálculo de novos parâmetros

Mesmo de posse da grande quantidade de parâmetros já listados, com uma breve análise dos mesmos, verifica-se que novas informações podem ser calculadas a fim de obter mais parâmetros e poder testá-los.

Os valores calculados foram:

1. **Delta:** valor em kA, referente a diferença entre o $I_{cc3\Phi}$ na barra da SED e o do ponto solicitado pelo cliente;
2. **$I_{cc3\Phi}/km$:** valor em kA que o nível de curto-circuito se degrada por quilômetro; esse valor é obtido através da relação entre o delta e a distância entre a SED e o ponto solicitado pelo cliente;
3. **Km Total:** a soma da extensão de rede MT com a de rede BT do alimentador;
4. **KW Total:** a soma da demanda de clientes BT e clientes MT;
5. **Total de Clientes:** a soma dos clientes de BT e MT;

Os valores calculados Delta e $I_{cc3\Phi}/km$ são formas de evidenciar algumas inferências: qualidade e robustez da rede e distância elétrica entre fonte e GD.

4.3. Detecção de *Outliers*

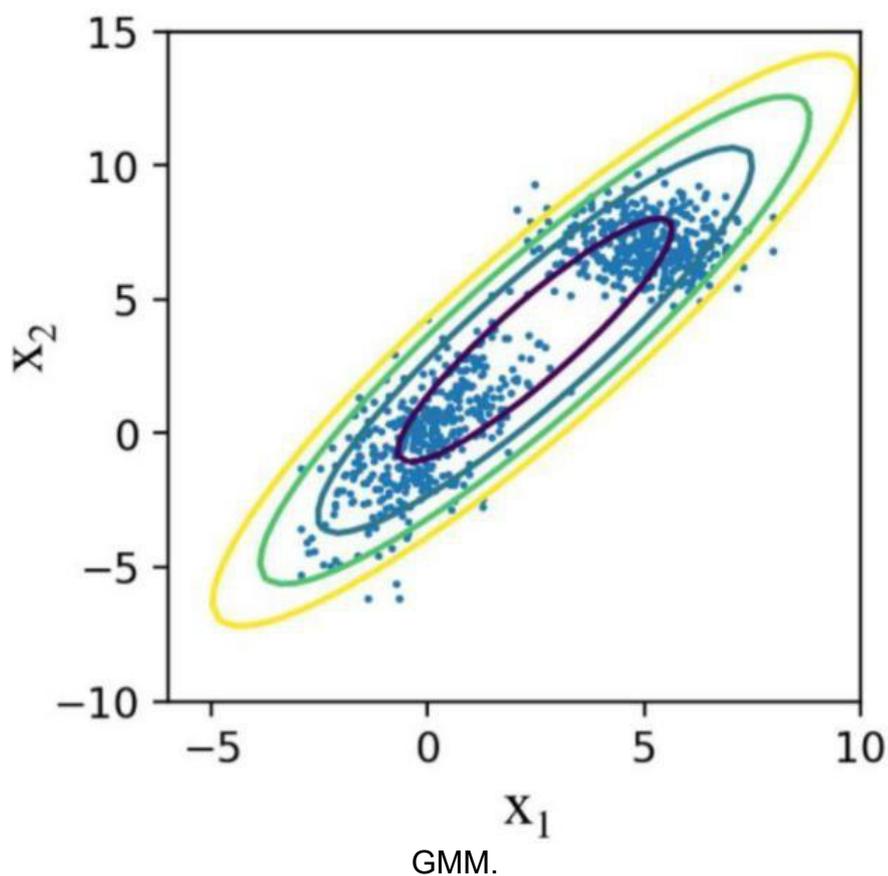
Após a coleta e estruturação dos dados, foram identificados os pontos que não seguem o padrão da maioria dos dados, ou seja, detecção de *outliers* (ROUSSEEUW, 1990, p. 1).

A identificação de outliers é uma parte importante da análise de dados, pois caso não sejam expurgadas do modelo, irão dificultar o treinamento e aprendizagem do algoritmo, uma vez que dados anômalos serão ensinados ao

modelo. Diante desses fatos, verifica-se a importância de aplicar uma metodologia na identificação desses pontos indesejáveis.

Para detecção dos dados anômalos foi usado o modelo de misturas Gaussianas (Gaussian Mixture Model - GMM) que é uma função de densidade de probabilidade paramétrica representada como uma soma ponderada de Gauss das densidades de componentes. Quando certo ponto está próximo a uma alta densidade dos dados de entrada, ele possui baixa probabilidade de ser uma anomalia, caso contrário, a probabilidade aumenta ao se afastar da densidade de pontos (TAX, 2005, p. 1). Na Figura 11 é ilustrado o modelo GMM.

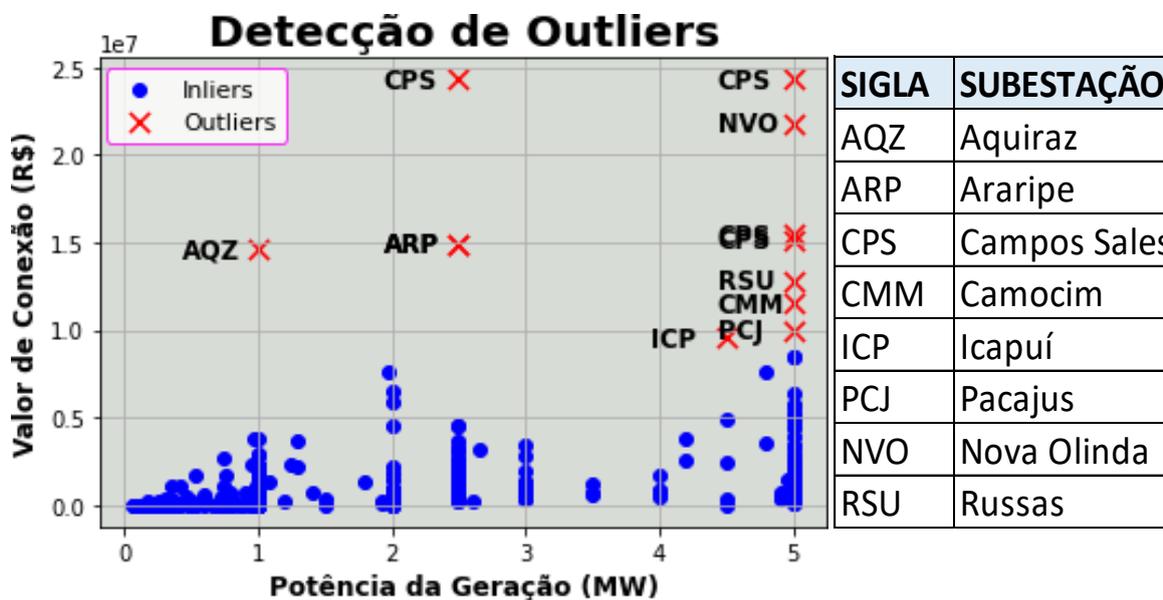
Figura 11 - Exemplo de representação do modelo



Fonte: [SHIMIZU, 2020]

Contextualizando os conceitos de *outliers* para o trabalho em questão, seriam estudos feitos em que obras de adequações foram discrepantes para um dado valor de potência. A Figura 12 evidencia o resultado após os dados serem modelados pelo detector de anomalias GMM.

Figura 12 - Representação dos estudos que foram considerados outliers, a figura evidencia qual Subestação é referente ao valor fora do padrão.



Fonte: [Próprio Autor]

Existem alguns possíveis motivos pelos quais esses estudos foram considerados *outliers* do ponto de vista técnico:

- Localização escolhida pelo cliente tecnicamente inviável para a respectiva potência;
- Existência de pelo menos mais uma GD no mesmo alimentador;
- Alimentadores com a tensão mais elevada naturalmente;
- Locais onde não há mais a possibilidade de novos alimentadores, apenas viabilidade técnica em Alta Tensão;

Diante dessas situações atípicas, existe uma maior probabilidade da ocorrência de um número maior de adequações na rede elétrica a fim de viabilizar tecnicamente a conexão da GD, ou seja, maior valor financeiro previsto para viabilizar a conexão da GD com tais características citadas.

O que o modelo de detecção de anomalias apontou como variáveis com alta probabilidade de serem *outliers*, foram excluídas do treinamento.

4.4 Normalização dos Dados

O objetivo dessa normalização dos dados numéricos previsores é colocá-los em uma escala comum, evitando discrepância nos intervalos. Por exemplo, considerando dois parâmetros do conjunto de dados desse estudo como $I_{cc3\Phi}$ e Corrente Máxima do Alimentador, ambos valores estão em escala diferentes, o curto-circuito varia entre 0,056 e 7,7 kA, já a Corrente Máxima do Alimentador de 6 a 458 A. Observa-se que os intervalos e as escalas são bem diferentes, ou seja, dependendo do modelo de aprendizado de máquina selecionado, nesse caso, a Corrente Máxima do Alimentador sensibilizará mais o valor previsto, não porque o parâmetro é mais relevante, mas por que sua escala e intervalo são maiores. (GÉRON, 2019, p. 68)

. O método de normalização utilizado no trabalho consistiu em atribuir o valor zero para o menor valor e 1 para o maior valor do parâmetro como expresso em (4.1).

$$X_n = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (4.1)$$

Em que:

- X_n : novo valor após a normalização
- X : valor original do parâmetro
- $\min(X)$: mínimo valor do parâmetro a ser normalizado
- $\max(X)$: máximo valor do parâmetro a ser normalizado

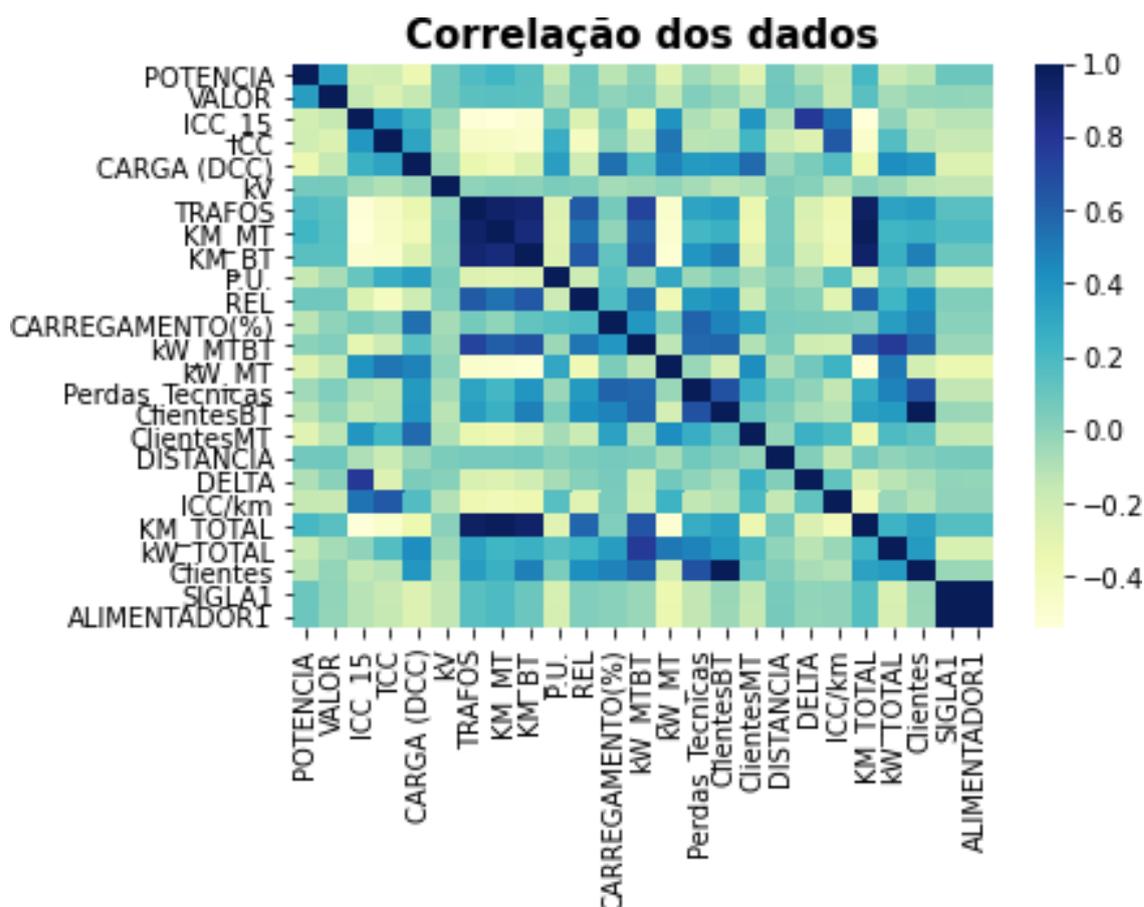
Ao final desse pré-processamento dos dados de entrada, tais parâmetros estarão na mesma escala, permitindo assim, que o modelo de aprendizado de máquina não tenha *performance* comprometida.

4.5 Escolha dos Parâmetros Previsores

No processo de escolha dos parâmetros previsores, os quais serão utilizados no modelo com intuito de construí-lo com alta confiabilidade, alguns quesitos são avaliados, mas antes, uma primeira análise visual foi feita verificando a

correlação linear entre os dados. Para essa correlação visual foi utilizado um mapa de calor (*heatmap*), no qual o nível de correlação zero refere-se a dados que não possuem nenhuma correlação, caso contrário será 1, conforme mostrado na Figura 13.

Figura 13 - Correlação dos Dados de Entrada plotadas em um mapa de calor.



Fonte: [Próprio Autor]

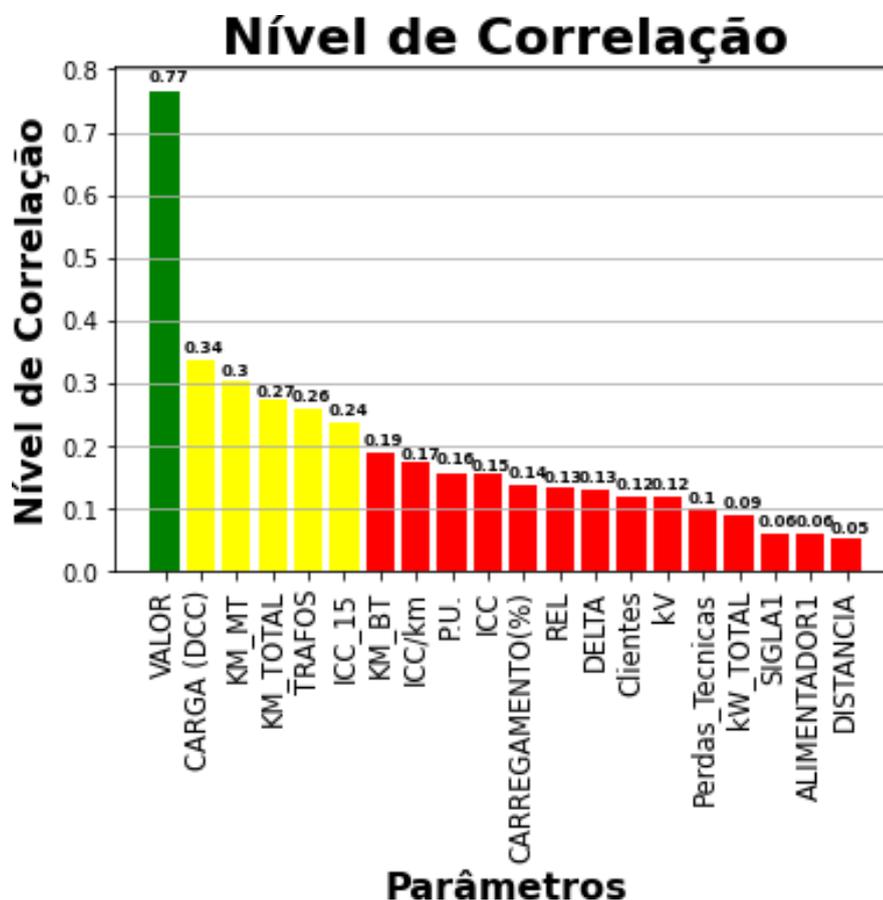
4.5.1 Correlação dos Dados

Com uma análise da Figura 13, verifica-se algumas correlações mais fortes entre parâmetros previsores, porém sem mais inferências precisas.

A correlação em valor absoluto entre 0 e 1 dos parâmetros de entrada em relação ao valor a ser previsto, que é a potência da GD, é mostrado em gráfico de barras na Figura 14. A correlação da Figura 14 quantifica os melhores possíveis parâmetros a serem selecionados no treinamento em ordem decrescente de

correlação.

Figura 14 - Correlação dos dados de entrada dispostos em gráfico de barras.



Fonte: [Próprio Autor]

Essa correlação é encontrada através do Método de Correlação de Spearman, o qual representa uma medida estatística da força de relação monotônica entre dados. Quando o valor absoluto se aproxima de 1, trata-se de uma relação forte. A relação Monotônica é uma função em que as variáveis tendem a mover-se na mesma direção relativa, mas não necessariamente a uma taxa constante, por isso, o coeficiente de Spearman identifica tanto relações lineares quanto não lineares entre os dados.

A partir dos coeficientes de correlação apresentados, já é possível notar quem são os melhores possíveis parâmetros de entrada do modelo: os mostrados em verde e amarelo. Porém, alguns testes ainda devem ser feitos para a escolha com maior confiança.

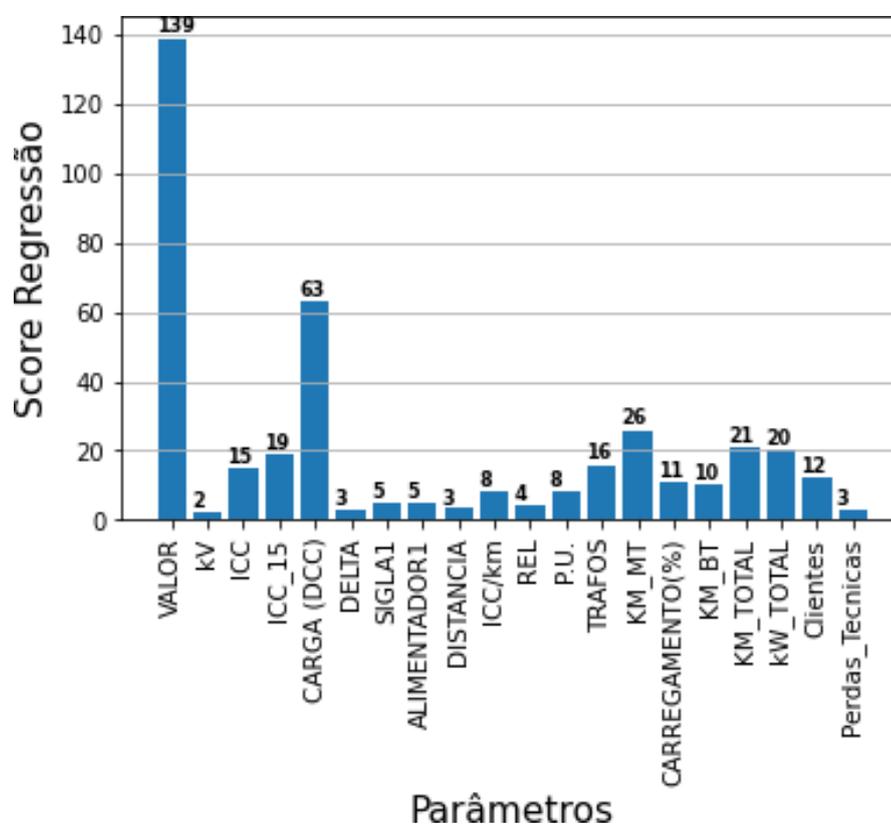
4.5.2 Seleção das Variáveis

A seleção das variáveis é o processo de redução do número de previsores de entrada ao desenvolver um modelo preditivo. Nem todas as variáveis de entrada devem ser usadas, pois aquelas com baixa relevância dificultam a aprendizagem durante o treinamento. Além disso, quanto maior o número de variáveis, maior o esforço computacional, tanto para armazenar, quanto para processar o modelo. Seguindo a trajetória de exploração dos parâmetros de entrada, no tópico anterior através do coeficiente de correlação de Spearman foi possível mapear alguns parâmetros que possuem um bom nível de correlação com a variável alvo, porém nada ainda decisivo.

Para maior segurança, uma segunda análise nos dados foi implementada que trata de um modelo linear para testar o efeito individual de cada uma das variáveis de entrada na variável alvo. Esta é uma função de pontuação a ser usada em um procedimento de seleção de parâmetro. Esse processo é feito em 3 etapas:

1. O parâmetro de interesse e os dados são ortogonalizados por regressores constantes;
2. A correlação cruzada entre os dados e os regressores é calculada;
3. Essa correlação é convertida em um escore;

Figura 15- Escores de correlação dos dados após a regressão entre as variáveis.



Fonte: [Próprio Autor]

Com a análise, os parâmetros a serem considerados foram os com nível de correlação, apresentados na Figura 14, maiores ou iguais a 0.2. Porém, essa escolha é uma escolha inicial e durante o processo de treinamento parâmetros ainda podem ser excluídos, pois caso uma variável de entrada deteriore a performance do modelo, a mesma será excluída.

Por fim, os parâmetros iniciais a serem utilizados no treinamento do modelo serão os seguintes: VALOR, CARGA (DCC), ICC_15, KM_MT, KM_TOTAL, kW_TOTAL e TRAFOS, mas durante a fase de avaliação parâmetros podem ser alterados, excluídos ou incluídos.

4.6. Seleção Hiperparâmetros dos Modelos

São vários os modelos supervisionados de Aprendizado de Máquina (AM) como por exemplo: Regressão Linear, Regressão Logística, Árvore de Decisão (*Tree Decision*), Florestas Aleatórias (*Random Forest*), Vetor de Máquina de Suporte (Support Vector Machine), Gradiente Descendente (Gradient Descent), Redes Neurais, dentre outros. Essa variedade de modelos causa certa indecisão ao selecionar o modelo que melhor se adaptará ao problema do presente trabalho.

Tendo em vista a complexidade inerente a resolução da proposta desse trabalho, modelos menos complexos tendem a não atender com níveis de erros apropriados, por isso devem ser testados modelos mais robustos a fim de prover menores erros ao modelo.

4.6.1 Grid Search

Os aspectos analíticos de um projeto de ciência de dados pode ser dividido em três partes: 1) A primeira parte diz respeito à coleta dos dados necessários e a seleção dos recursos de entrada; 2) A segunda parte gira em torno da escolha do modelo certo de aprendizado de máquina; 3) A última parte é sobre como encontrar os hiperparâmetros de cada algoritmo AM. Cada algoritmo AM tem uma grande quantidade de parâmetros internos inerentes a cada modelo. Esses parâmetros controlam a precisão do modelo. Portanto, os hiperparâmetros são particularmente importantes em um projeto de ciência de dados.

Definir os valores ótimos dos parâmetros pode ser desafiador e exigir muitos recursos. Para contornar a dificuldade de seleção dos hiperparâmetros de cada modelo AM, foi usado um pacote da biblioteca Scikit-learn, denominado Grid Search (GS). Pesquisa de rede ou Grid Search é uma técnica de ajuste que tenta calcular os valores ótimos de hiperparâmetros. É uma busca exaustiva realizada sobre os valores de parâmetros específicos de um modelo. O modelo também é conhecido como estimador. No processo de execução do algoritmo AM, os parâmetros internos são combinados e o erro médio absoluto é monitorado. Ao final de diversas iterações, tem-se o modelo com menor erro médio absoluto e o retorno dos parâmetros obtidos (PEDREGOSA et al., 2011, p. 4).

Essa busca automática, além de economizar bastante tempo, também

minimiza o risco que o modelo fique enviesado sofrendo *overfitting*, que acontece quando o modelo tem bom desempenho apenas no conjunto de treinamento, mas não nos novos valores a serem preditos, que não foram vistos no treinamento (GÉRON, 2019, p. 27-28).

4.7. Métricas de Desempenho

Métricas de desempenho, enquanto medidas de erro, podem ser definidas como uma construção lógica e matemática projetada para medir a fidelidade dos resultados obtidos com o resultado previsto. (BOTCHKAREV, 2019, p. 1-2)

As métricas mais comumente utilizadas são o erro médio absoluto (MAE – *mean absolute error*) e o erro quadrático médio (MSE – *mean squared error*). Na regressão do aprendizado de máquina, as métricas de desempenho são usadas para comparar as previsões do modelo treinado com os dados do conjunto de teste, que são dados reais, ou seja, comparar a realidade com a saída do algoritmo. (BOTCHKAREV, 2018, p. 3).

Os resultados dessas comparações de métricas influenciam a escolha de um modelo de AM em detrimento de outros com menor performance, foi o que foi feito neste trabalho e os resultados serão mostrados a seguir.

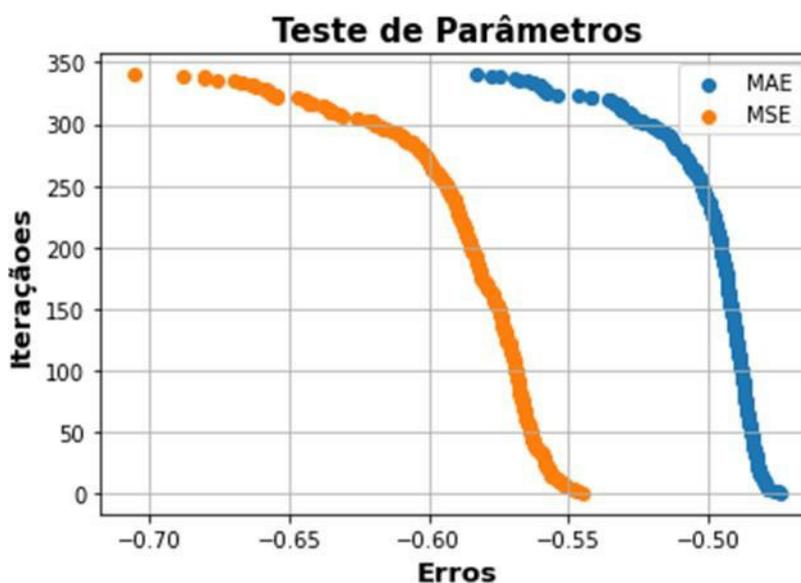
Após o treinamento e testes de desempenho dos quatro modelos de algoritmos diferentes utilizados, foi possível obter alguns avaliadores para a comparação: Erro Médio Absoluto, Erro Médio Quadrático e Tempo de Treinamento.

- Erro Médio Absoluto: É o valor absoluto da diferença entre o valor previsto e o valor real. Essa métrica representa o tamanho do erro que podemos esperar da previsão em média, tanto para mais como menos.
- Erro Médio Quadrático: A definição é quase a mesma do MAE, porém o erro absoluto é elevado ao quadrado, a fim de evidenciar, dar mais peso a diferenças maiores.
- Tempo de Treinamento: Tempo em milissegundos (ms) que o algoritmo levou para treinar com os dados de treinamento.

4.7.1 Erro Médio Absoluto e Erro Médio Quadrático

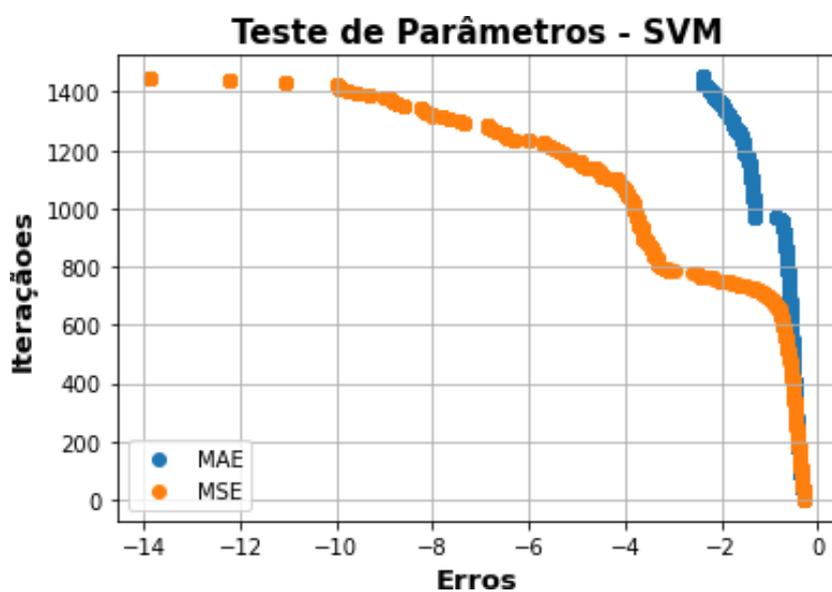
Conforme dito, utilizando o pacote GS, em que automatizou a busca pelos melhores hiper-parâmetros de cada algoritmo, também foi possível registrar essa busca. Com o GS foi possível resgatar a trajetória de busca do menor erro possível. As Figuras 16 a 19 mostram os quatro modelos treinados: RF, SVM, RNA e XGBoost, respectivamente, e os valores de erro médio absoluto e quadrático com as iterações.

Figura 16 - Representação da busca dos melhores hiperparâmetros do modelo Random Forest a fim minimizar os erros.



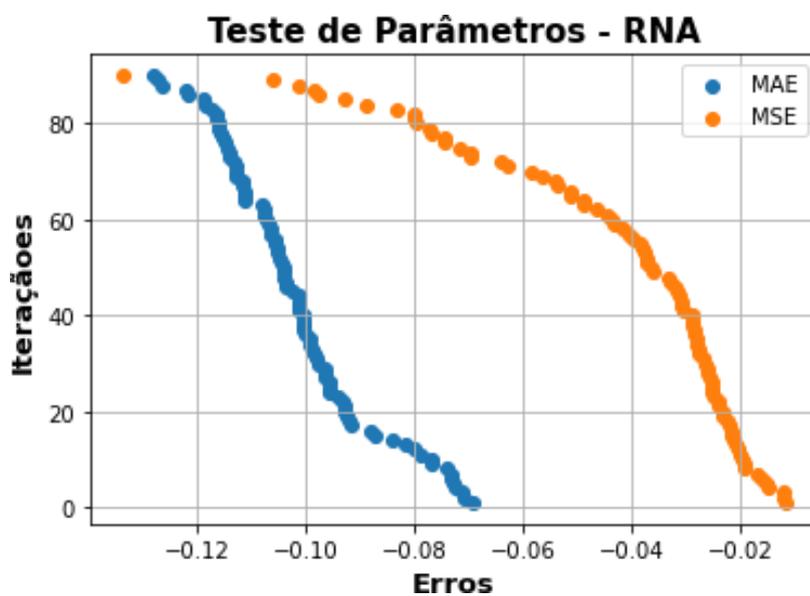
Fonte: [Próprio Autor]

Figura 17 - Representação da busca dos melhores hiperparâmetros do modelo Support Vector Machine a fim minimizar os erros.



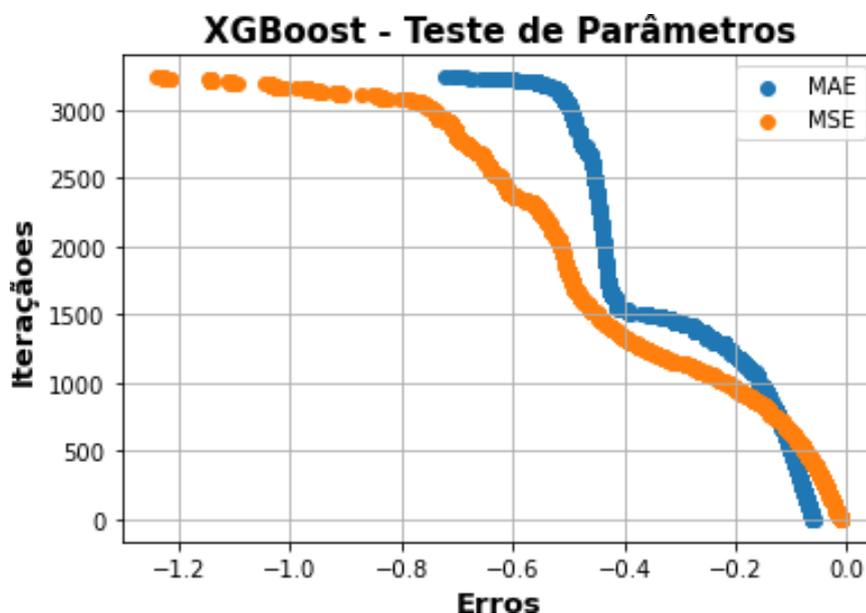
Fonte: [Próprio Autor]

Figura 18 - Representação da busca dos melhores hiperparâmetros do modelo Redes Neurais Artificiais a fim minimizar os erros.



Fonte: [Próprio Autor]

Figura 19 - Representação da busca dos melhores hiperparâmetros do modelo Extreme Gradient Boosting a fim minimizar os erros.



Fonte: [Próprio Autor]

Na Tabela 1 são apresentados o erro final MAE e MSE para cada algoritmo de aprendizado de máquina:

Tabela 1 - Resumo das métricas dos modelos.

Modelo	MAE	MSE
Random Forest	0,480	0,589
Support Vector Machine	0,129	0,135
Redes Neurais Artificiais	0,068	0,0168
XGBoost	0,049	0,0036

Fonte: [Próprio Autor]

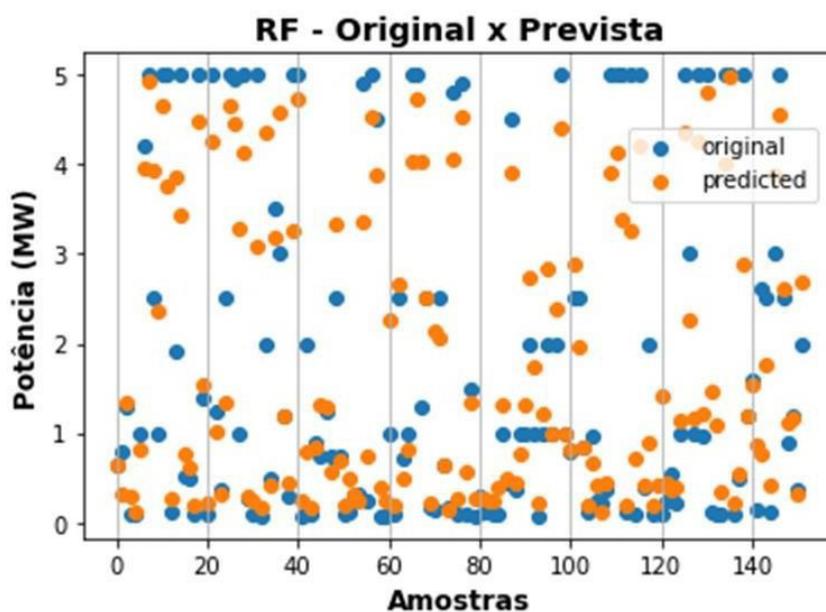
Pode-se observar a partir das curvas das Figuras 16, 17, 18 e 19 e a Tabela 1 que os melhores algoritmos são as Redes Neurais e o *XGBoost*, avaliando apenas as métricas.

Para materializar o resultado, no caso do *XGBoost*, algoritmo com menores erros, esse algoritmo em média erra 0,049 MW de Potência de GD a cada resposta, ou seja, 49 kW para mais ou para menos.

Outra forma de visualizar esse resultado é comparando os valores reais com os valores previstos. Os valores reais foram estudos reais de viabilidade de

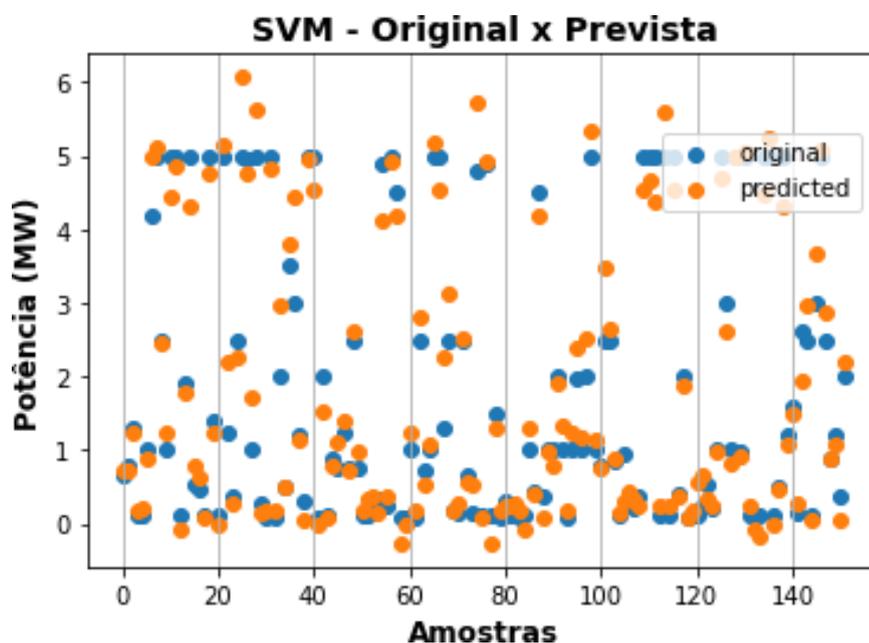
GD realizados pelos engenheiros da distribuidora. Já os valores previstos são dados de resposta do algoritmo com os mesmos parâmetros dos dados estudados pelos engenheiros. Tais gráficos evidenciam quão preciso é o algoritmo.

Figura 20 - Representação gráfica dos valores previstos pelo algoritmo *Random Forest* e os valores originais.



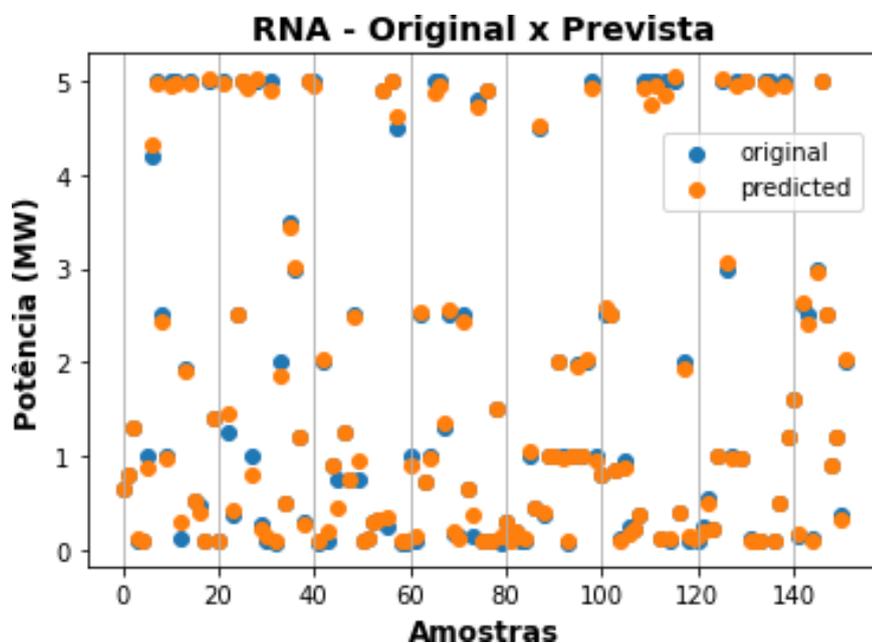
Fonte: [Próprio Autor]

Figura 21 - Representação gráfica dos valores previstos pelo algoritmo Support Vector Machine e os valores originais.



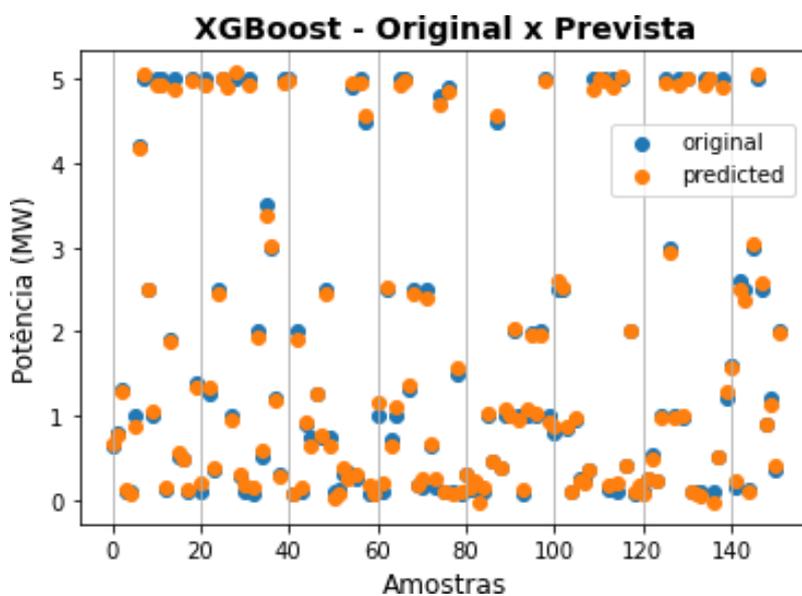
Fonte: [Próprio Autor]

Figura 22 - Representação gráfica dos valores previstos pelo algoritmo Redes Neurais Artificiais e os valores originais.



Fonte: [Próprio Autor]

Figura 23 - Representação gráfica dos valores previstos pelo algoritmo XGBoost e os valores originais.

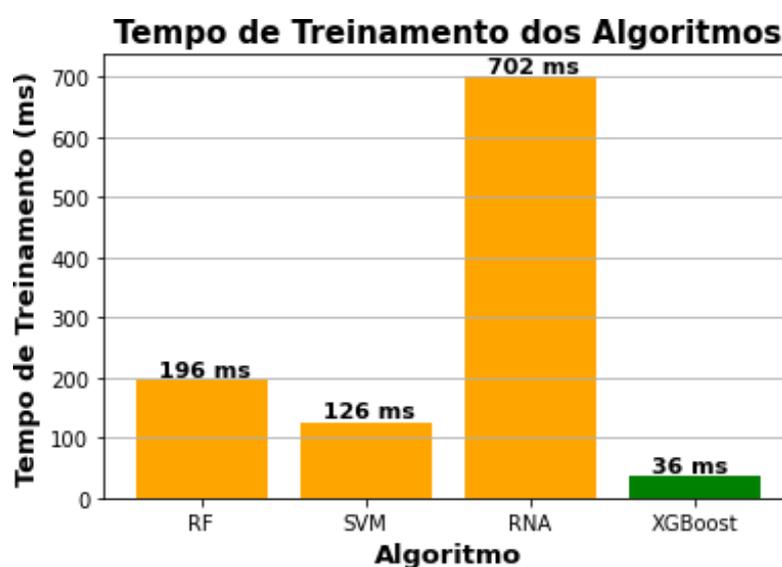


Fonte: [Próprio Autor]

4.7.2 Tempo de Treinamento

Analisando os erros, verifica-se que as RNA e o *XGBoost* possuem os menores erros, além disso, são bem parecidos. Verificando apenas essas métricas, não seria tão diferente a escolha de um desses modelos. Porém, uma última métrica foi avaliada: o Tempo de Treinamento. Na Figura 24 são discriminados os tempos de treinamento de cada modelo.

Figura 24 - Representação dos tempos de treinamento de cada modelo.



Fonte: [Próprio Autor]

O algoritmo *XGBoost* além de possuir os menores erros, ainda é quase vinte vezes mais veloz que as RNA. Esse fator é bastante importante, pois apesar de nesse trabalho não serem usados grandes volumes de dados, em casos em que os dados de treinamentos cheguem à marca de milhões, esse tempo poderia ser prioridade na escolha do algoritmo, chegando ao ponto de inviabilizar sua utilização.

Então, como mostra a literatura sobre a boa performance de algoritmos do tipo impulso (*Boosting*), com o algoritmo *XGBoost* não foi diferente, teve excelente resultado em comparação com outros tipos de algoritmos.

5. RESULTADOS

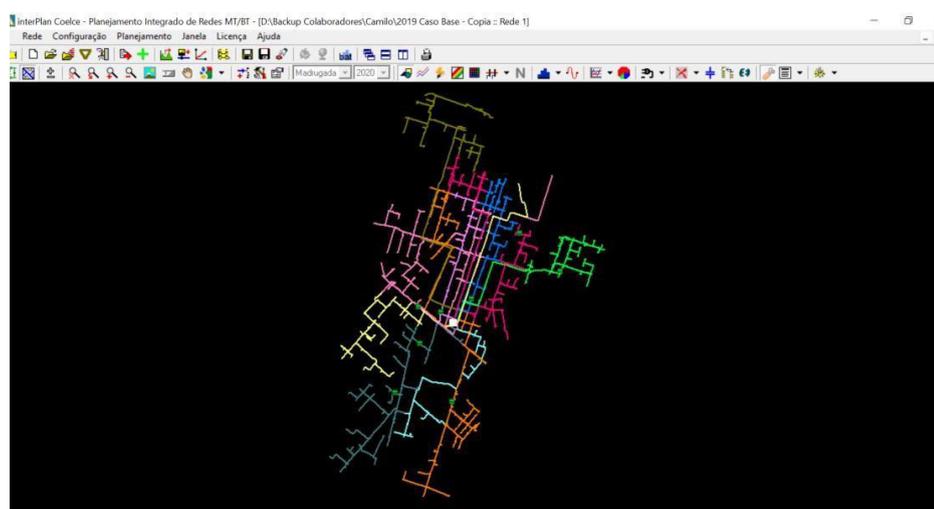
5.1 Avaliação do Algoritmo

Para realizar a validação do modelo, mais especificamente dos seus resultados entregues, o algoritmo foi aplicado em alguns pontos da rede elétrica de MT selecionados aleatoriamente através de um *script* feito na linguagem *python*. O *script* selecionou um total de 80 pontos, sendo 20 pontos de cada tipo de alimentador: comercial, residencial, industrial e rural.

Feita a seleção dos pontos, as características dos mesmos foram coletadas e aplicadas ao modelo a fim de obter a potência máxima de GD sem que seja necessário adequações à rede.

A validação foi feita utilizando um *software* de fluxo de potência especializado, denominado Interplan. O Interplan é um software de análise de redes e estudos de planejamento de sistemas de distribuição de energia elétrica, desenvolvido pela empresa nacional Daimon Especialistas em Energia. O Interplan é o software usado pela Enel Distribuição Ceará na análise de redes de distribuição MT/BT. Na Figura 26 é mostrado um exemplo da interface gráfica do *software* que os especialistas da área de estudos utilizam.

Figura 25 - Interface do Interplan: software utilizado pela distribuidora, para a realização dos estudos de fluxo de potência.

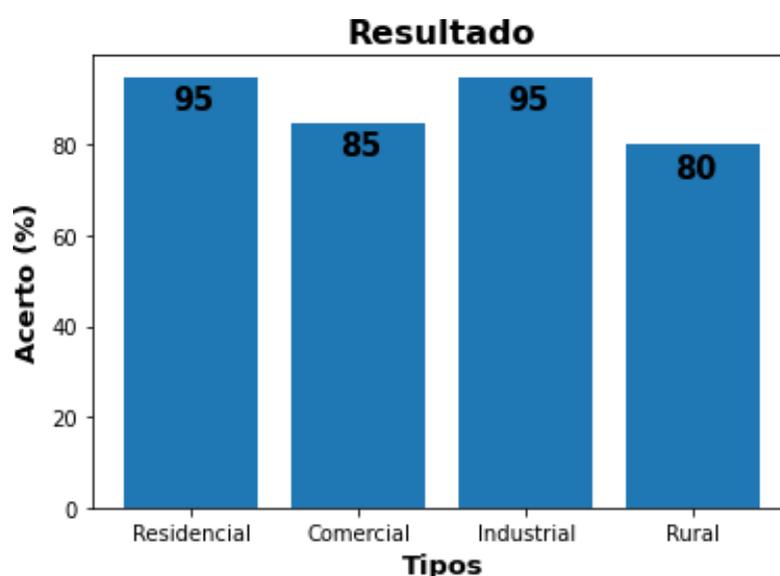


Fonte: [Próprio Autor]

Nas simulações, foram avaliados os impactos das GD conectadas à rede elétrica. Os fatores avaliados foram: tensão, corrente e flexibilidade operacional. Em que a flexibilidade operacional é a simulação de contingência do alimentador em que a GD foi simulada, sendo transferida para outro alimentador de distribuição em que há equipamento de manobra (chaves, religadores). Ou seja, a GD foi simulada em um alimentador em condição normal e outro em situação de falta do mesmo. Com a conexão da GD são verificados se nenhum dos fatores acima violados, em que a violação é determinada por valores regulados pela ANEEL. Quando a tensão não excede valores regulatórios, são verificados os outros dois fatores: a corrente e flexibilidade operacional.

Diante dessa avaliação foi possível diagnosticar a taxa de erro para o modelo segmentado pela característica de clientes predominante no alimentador. Para a determinação dessa taxa considerou-se que o algoritmo acertou a previsão, quando a potência de resposta entregue pelo mesmo não necessitou de obras ou adequações na rede elétrica. Caso contrário, se houve a necessidade de alguma intervenção, considerou-se que o algoritmo errou. Na Figura 26 é mostrada a taxa de acerto segmentada por tipo de alimentador.

Figura 26 - Representação gráfica do nível de acerto do algoritmo da previsão da potência.



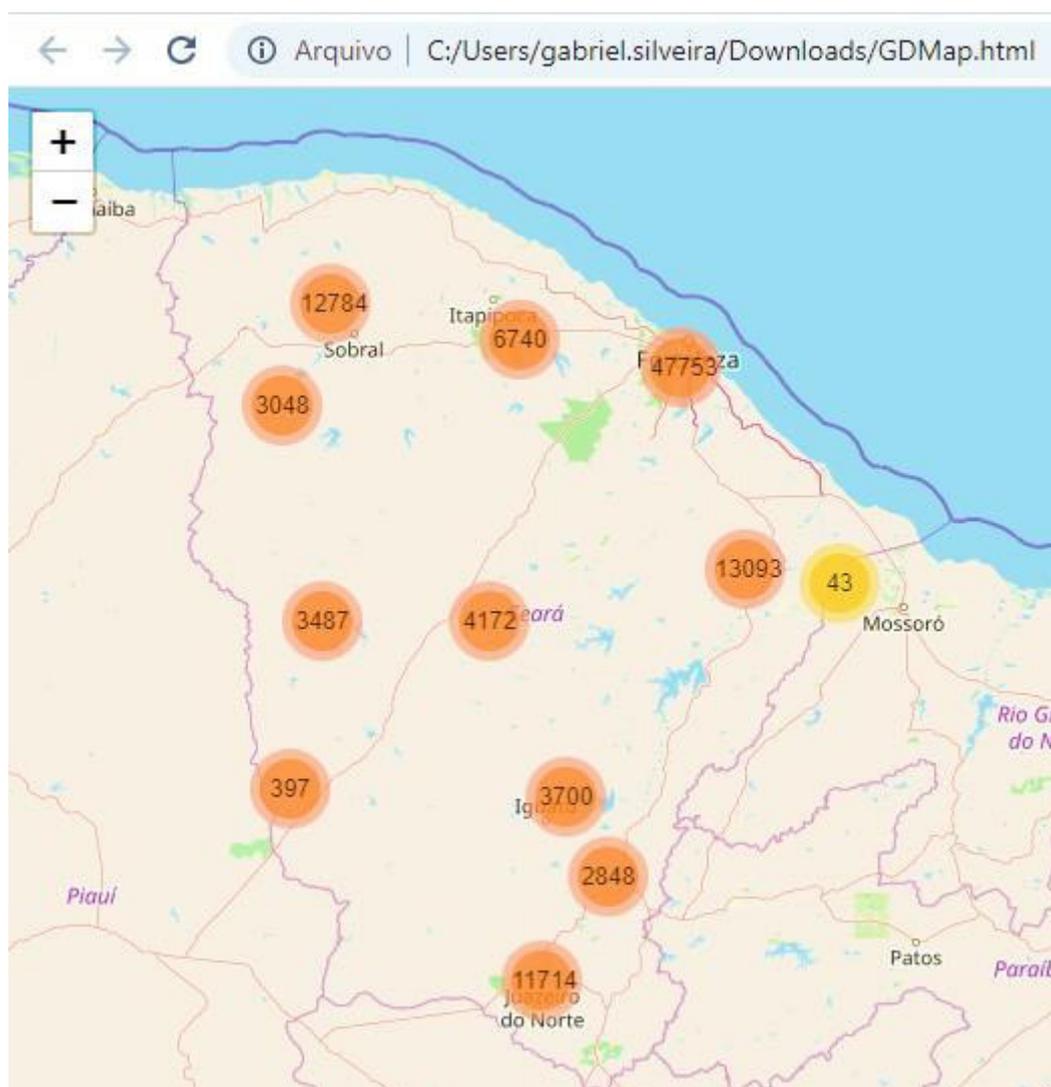
Fonte: [Próprio Autor]

O tipo com menor taxa de acerto foram os alimentadores do tipo rural. Uma hipótese para a menor taxa de acerto no alimentador predominantemente é devido à presença de trechos em condutor monofilar com retorno pela terra (MRT). Embora os sistemas MRT visem atender as necessidades de energia elétrica de consumidores rurais com baixa densidade de carga, com redução de custo de implantação da ordem de 20 a 30% em relação ao sistema fase-fase, tais sistemas mal condicionam as redes de distribuição. Pode-se citar como desvantagem que a potência do ramal MRT é limitada à potência nominal do transformador isolamento e o nível de correntes de curto-circuito disponíveis nas linhas MRT é limitado com a inserção da impedância do transformador de isolamento. Outro fator que pode ter contribuído para a menor taxa de acerto é o comprimento dos alimentadores rurais, com poucos equipamentos de manobra, diminuindo assim sua flexibilidade operativa. Para os outros tipos de alimentadores, teoricamente mais robustos, o nível de acerto foi maior.

5.2 Protótipo

Com o intuito de disponibilizar com transparência uma visão geral da capacidade de conexão de GD, tanto para colaboradores da distribuidora quanto para clientes externos, foi desenvolvido um arquivo protótipo em linguagem python e html (*HyperText Markup Language*), linguagem de marcação utilizada na construção de páginas *web*. Esse arquivo interativo é o resultado da resposta, para toda a rede MT, do algoritmo de AM desenvolvido. O plano de fundo desse arquivo é o mapa do Ceará, área de concessão da distribuidora; e o conteúdo sobre o mapa são as barras dos alimentadores de MT, onde o usuário tem a liberdade de navegar por essas barras e verificar em cada barra qual a potência ótima de GD, ou seja, a potência máxima sem a necessidade de adequações. Na Figura 25 é mostrado a visão geral do arquivo.

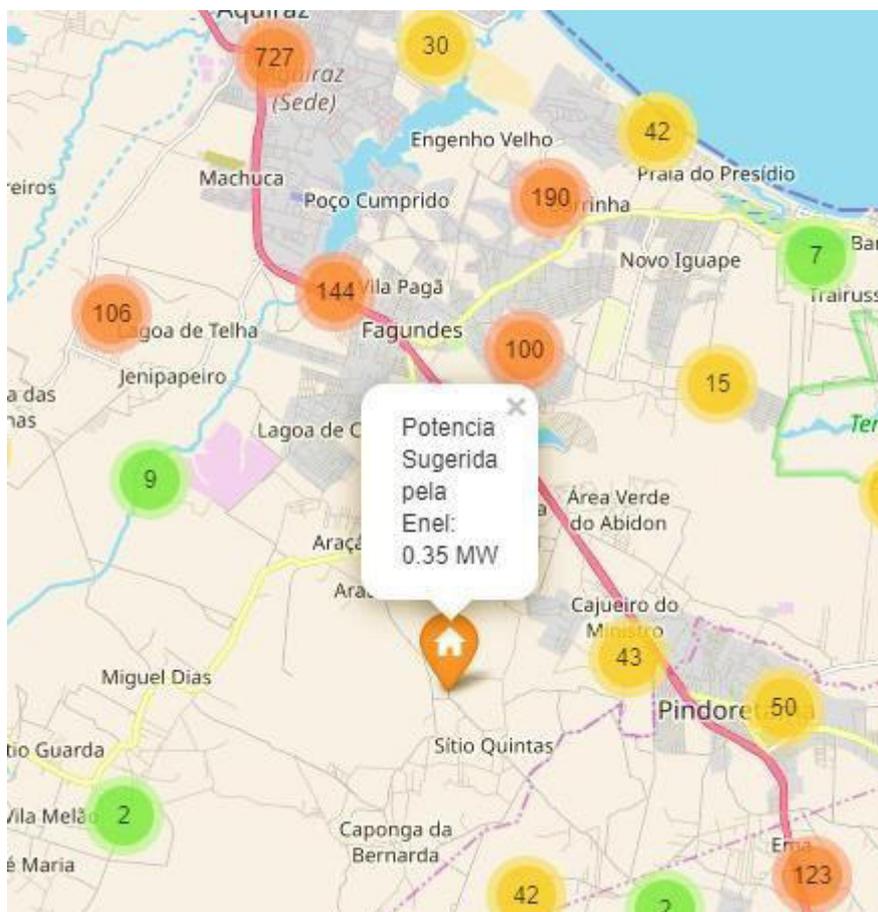
Figura 27 - Arquivo html mostrando a quantidade de barras por região.



Fonte: [Próprio Autor]

A partir da navegação, o usuário pode identificar os pontos da rede elétrica mais interessantes para o seu propósito. Isso permite ao acessante o rastreamento das condições de pontos na rede elétrica antes de solicitar um estudo de viabilidade técnico-econômico à Enel Distribuição Ceará. A Figura 26 mostra o exemplo de uma navegação e o resultado obtido.

Figura 28 - Resultado da navegação do usuário pelo arquivo protótipo.



Fonte: [Próprio Autor]

6. CONCLUSÃO

O presente trabalho apresentou uma proposta de utilização e implementação de um algoritmo utilizando técnicas de aprendizado de máquina capaz de contribuir com as distribuidoras de energia elétrica e acessantes que buscam conexão de unidades de geração distribuída. A aplicação desenvolvida visa disponibilizar informações de alto nível acerca da capacidade da rede elétrica na conexão de GD, a fim de mitigar transtornos de ordem interna e externa à distribuidora.

A partir das verificações feitas nos resultados, o algoritmo mostrou potencial para ser utilizado por futuros acessantes, pois provou-se uma taxa de acerto bem adequada à proposta do trabalho, na qual o cliente pode ter um grau elevado de confiança na resposta da plataforma.

Espera-se que com a disponibilização dessa plataforma, que os solicitantes de conexão de GD tenham a transparência necessária sobre dados técnicos da rede elétrica e maior poder de decisão acerca da viabilidade de conexão à rede elétrica. Do ponto de vista da distribuidora, a aplicação contribuirá para reduzir o número de resubmissão feita pelos acessantes na tentativa de selecionar um ponto que ofereça condições técnicas favoráveis de inserção de GD.

Alguns pontos podem ser verificados para uma melhor modelo como por exemplo a inclusão de outras variáveis de grande relevância para determinar a potência disponível, fato este que pode ser objeto de estudo de avaliação futura. Ainda como sugestão de trabalhos futuros sugere-se a aplicação da mesma metodologia para predição de acesso de carga industrial, comercial e serviços.

REFERÊNCIAS

ABBAS, Nadine; NASSER, Youssef; EL AHMAD, Karim. Recent advances on artificial intelligence and learning techniques in cognitive radio networks. **EURASIP Journal on Wireless Communications and Networking**, v. 2015, n. 1, p. 1-20, 2015.

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. Resolução Normativa 482, de 17 de abril de 2012. Disponível em: < <http://www.aneel.gov.br/cedoc/ren2012482.pdf>>.

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. Resolução Normativa 674, de 11 de agosto de 2015. Disponível em: < <http://www2.aneel.gov.br/cedoc/ren2015674.pdf>>

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. Resolução Normativa 687, de 24 de novembro de 2015.

Disponível em: < <http://www2.aneel.gov.br/cedoc/ren2015687.pdf>>.

ANEEL - Agência Nacional de Energia Elétrica. Atlas de energia elétrica do Brasil. 3 ed. Brasília/DF, 2008.

ANEEL. Brasil ultrapassa marca de 1GW em geração distribuída. <https://www.aneel.gov.br>, 2019. Disponível em: <https://www.aneel.gov.br/sala-deimprensa-exibicao/-/asset_publisher/XGPXSqdMFHrE/content/brasil-ultrapassamarca-de-1gw-em-geracaodistribuida/656877#:~:text=O%20Brasil%20ultrapassou%20a%20marca,2012%20e%20687%2F2015>. Acesso em: 15 de nov. de 2019.

ANEEL. Geração Distribuída – Planilha de Dados: SISGD. Disponível em: <<https://app.powerbi.com/view?r=eyJrIjoiZjM4NjM0OWYtN2IwZS00YjVlTlIIMjltN2E5MzBkN2ZiMzVkIiwidCI6IjQwZDZmOWI4LWVjYTctNDZhMi05MmQ0LWVhNGU5YzAxNzBIMSIsImMiOiR9>>. Acesso em: 15 de mar. de 2021.

BOTCHKAREV, Alexei. A NEW TYPOLOGY DESIGN OF PERFORMANCE METRICS TO MEASURE ERRORS IN MACHINE LEARNING REGRESSION ALGORITHMS. **Interdisciplinary Journal of Information, Knowledge & Management**, v. 14, 2019.

BOTCHKAREV, Alexei. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. **arXiv preprint arXiv:1809.03006**, 2018.

Ceará tem potencial de geração de energia solar 30 vezes maior que a capacidade instalada no Brasil, 2019. Disponível em: <https://www.portalsolar.com.br/blogsolar/energia-solar/ceara-tem-potencial-de-geracao-de-energia-solar-30-vezesmaior-que-a-capacidade-instalada-no-brasil.html> >. Acesso em: 15 de out. de 2020.

CHO, Kyunghyun et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. **arXiv preprint arXiv:1406.1078**, 2014.

DHALIWAL, Sukhpreet Singh; NAHID, Abdullah-Al; ABBAS, Robert. Effective intrusion detection system using XGBoost. **Information**, v. 9, n. 7, p. 149, 2018.

Empresa de Pesquisa Energética. Plano Decenal 2030. Disponível em: <<https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/plano-decenalde-expansao-de-energia-2030>>. Acesso em: 16 de mar. de 2021.

GAMA, João et al. Concept drift in decision-tree learning for data streams. In: **Proceedings of the Fourth European Symposium on Intelligent Technologies and their implementation on Smart Adaptive Systems, Aachen, Germany, Verlag Mainz**. 2004. p. 218-225.

GÉRON, Aurélien. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems**. O'Reilly Media, 2019.

GISLASON, Pall Oskar; BENEDIKTSSON, Jon Atli; SVEINSSON, Johannes R. Random forests for land cover classification. **Pattern recognition letters**, v. 27, n. 4, p. 294-300, 2006.

GOMES, P. Vilaça et al. Technical-economic analysis for the integration of PV systems in Brazil considering policy and regulatory issues. **Energy policy**, v. 115, p. 199-206, 2018.

GUNN, Steve R. et al. Support vector machines for classification and regression. **ISIS technical report**, v. 14, n. 1, p. 5-16, 1998.

HA, Mahmoud Pesaran; HUY, Phung Dang; RAMACHANDARAMURTHY, Vigna K. A review of the optimal allocation of distributed generation: Objectives, constraints, methods, and algorithms. **Renewable and Sustainable Energy Reviews**, v. 75, p. 293-312, 2017.

HAYKIN, Simon; NETWORK, N. A comprehensive foundation. **Neural networks**, v. 2, n. 2004, p. 41, 2004.

HE, Kaiming et al. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. 2016. p. 770778.

KAYAL, Partha; CHANDA, Chandan Kumar. A simple and fast approach for allocation and size evaluation of distributed generation. **International Journal of Energy and Environmental Engineering**, v. 4, n. 1, p. 1-9, 2013.

MARTÍN-MARTÍNEZ, Francisco et al. Centralized vs distributed generation. A model to assess the relevance of some thermal and electric factors. Application to the Spanish case study. **Energy**, v. 134, p. 850-863, 2017.

MEYER, David; WIEN, FH Technikum. Support vector machines. **The Interface to libsvm in package e1071**, v. 28, 2015.

MOHAN, N. **Sistemas elétricos de Potência – um curso introdutório**. Rio de Janeiro: LTC, 2016.

MOHRI, Mehryar; ROSTAMIZADEH, Afshin; TALWALKAR, Ameet. Foundations of machine learning. MIT press, 2018.

MOISEN, G. G. Classification and regression trees. In: **Jørgensen, Sven Erik; Fath, Brian D.(Editor-in-Chief). Encyclopedia of Ecology, volume 1. Oxford, UK: Elsevier. p. 582-588.**, p. 582-588, 2008.

NETO, Alfredo BONINI; BONINI, Carolina dos Santos Batista. REDES NEURAS ARTIFICIAIS: APRESENTAÇÃO E UTILIZAÇÃO DO ALGORITMO PERCEPTRON EM BIOSISTEMAS/ARTIFICIAL NEURAL NETWORKS: INTRODUCTION AND USE OF PERCEPTRON ALGORITHM IN BIOSYSTEMS. **Revista Brasileira de Engenharia de Biosistemas**, v. 4, n. 2, p. 87-95, 2010.

OLAMAIE, J.; NIKNAM, T. Daily Volt/Var control in distribution networks with regard to DGs: A comparison of evolutionary methods. In: **2006 IEEE Power India Conference**. IEEE, 2006. p. 6 pp.

PEDREGOSA, Fabian et al. Scikit-learn: Machine learning in Python. **the Journal of machine Learning research**, v. 12, p. 2825-2830, 2011.

PINHO, João Tavares; GALDINO, Marco Antonio. Manual de engenharia para sistemas fotovoltaicos. **Rio de Janeiro**, v. 1, p. 47-499, 2014.

RODRIGUEZ-GALIANO, Victor Francisco et al. An assessment of the effectiveness of a random forest classifier for land-cover classification. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 67, p. 93-104, 2012.

ROUSSEEUW, Peter J.; VAN ZOMEREN, Bert C. Unmasking multivariate outliers and leverage points. **Journal of the American Statistical association**, v. 85, n. 411, p. 633-639, 1990.

SAMEK, Wojciech; WIEGAND, Thomas; MÜLLER, Klaus-Robert. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. **arXiv preprint arXiv:1708.08296**, 2017.

SHIMIZU, Naoto; KANEKO, Hiromasa. Direct inverse analysis based on Gaussian mixture regression for multiple objective variables in material design. **Materials & Design**, v. 196, p. 109168, 2020.

SILVA, Bruno Gonçalves da. **Evolução do setor elétrico brasileiro no contexto econômico nacional: uma análise histórica e econométrica de longo prazo**. 2011. Tese de Doutorado. Universidade de São Paulo.

British Petroleum, Statistical Review of World Energy 2020. <https://www.bp.com>, 2020. Disponível em: <https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review2020-renewable-energy.pdf>>. Acesso em: 20 de nov. de 2020.

TAX, David MJ; DUIN, Robert PW. Outlier detection using classifier instability. In: **Joint IAPR international workshops on statistical techniques in pattern recognition (SPR) and structural and syntactic pattern recognition (SSPR)**. Springer, Berlin, Heidelberg, 1998. p. 593-601.

VANDERPLAS, Jake. **Python data science handbook: Essential tools for working with data**. " O'Reilly Media, Inc.", 2016.

WICKHAM, Hadley; GROLEMUND, Garrett. **R for data science: import, tidy, transform, visualize, and model data**. " O'Reilly Media, Inc.", 2016.

ZHENG, Huiting; YUAN, Jiabin; CHEN, Long. Short-term load forecasting using EMDLSTM neural networks with a Xgboost algorithm for feature importance evaluation. **Energies**, v. 10, n. 8, p. 1168, 2017.

ZINTGRAF, Luisa M. et al. Visualizing deep neural network decisions: Prediction difference analysis. **arXiv preprint arXiv:1702.04595**, 2017.