LII Simpósio Brasileiro de Pesquisa Operacional
João Pessoa - PB, 3 a 5 de novembro de 2020

ESTIMATIVA DE PERDAS TÉCNICAS EM SISTEMAS DE DISTRIBUIÇÃO DE ENERGIA VIA APRENDIZAGEM DE MÁQUINA

Pedro H. M. Nascimento, Tiago R. Alves, Leone A. Campos, Cristina M. B. Castro, Leonardo W. Oliveira, Ivo C. S. Junior
Universidade Federal de Juiz de Fora
Rua José Lourenço Kelmer, s/n - São Pedro, Juiz de Fora
pedro.nascimento@engenharia.ufjf.br,
tiago.alves@engenharia.ufjf.br, leone.andrade@engenharia.ufjf.br,
cristina.castro@ufjf.edu.br, leonardo.willer@ufjf.edu.br,
ivo.junior@ufjf.edu.br

RESUMO

Este artigo apresenta o estudo da aplicação de três metodologias de aprendizagem de máquina, *Random Forest*, *Support Vector Regression* e *Multi-layer Perceptron*, na estimativa de perdas técnicas em sistemas de distribuição de energia elétrica. Para esse estudo são utilizados um alimentador com geração distribuída de natureza eólica e dois alimentadores reais com geração distribuída do tipo pequena central hidrelétrica. Os modelos de estimação são submetidos ao treinamento supervisionado com a utilização da técnica de validação cruzada que garante a generalização do modelo, retirando a influência de certos conjuntos de treinamento. Os resultados dos testes dos modelos passam por análises estatísticas para comparação entre as metodologias. Foram obtidos bons resultados de estimativas das perdas técnicas, com validação das metodologias aplicadas ao problema e a possibilidade de aplicação das técnicas apresentadas.

PALAVRAS CHAVE. Aprendizagem de máquina, Sistema de distribuição, Perdas técnicas, Regressão.

Tópicos: EN&PG, IC

ABSTRACT

This paper presents the study of the application of three machine learning methodologies, Random Forest, Support Vector Regression, and Multilayer Perceptron, in the estimation of technical losses in electricity distribution systems. This study uses, a feeder with wind generation and two real feeders with distributed generation of the central hydroelectric generating type. The models are subjected to supervised training using the cross-validation technique that ensures the generalization of the model, removing the influence of certain training sets. The test results of the models undergo statistical analysis to compare the methodologies. Good results were obtained from loss estimates, validation of the methods used in the problem, and the possibility of applying the presented technique.

KEYWORDS. Technical Losses. Distributed Generation. Computational Intelligence.

Paper topics: EN&PG, IC

Galoá { Este trabalho foi publicado utilizando o **Galoá** proceedings

1. Introdução

Apesar de ainda ser um conceito com múltiplas definições por parte de várias organizações internacionais, como o Instituto de Engenheiros Eletricistas e Eletrônicos (IEEE) e o *Electric Power Research Institute* (EPRI), por exemplo, geração distribuída (GD) pode ser vista como a produção de energia elétrica próxima aos locais de consumo, sendo normalmente ligada ao sistema de distribuição e com capacidade inferior às plantas de geração principais do sistema de potência [Adefarati e Bansal, 2016]. A tecnologia de GD vem sendo cada vez mais empregada na rede elétrica brasileira, apenas as micro e minigeração distribuída (MMGD) já ultrapassa 3GW [ANEEL, 2020], tornando importante o estudo dos impactos causados por estas fontes de energia na rede de distribuição e das perdas a elas associadas, de modo a garantir a confiabilidade e a eficiência da rede [Adefarati e Bansal, 2016]. As GDs podem ser de fontes de geração renováveis, como solar, eólica, hídrica e biomassa; ou não renováveis, como é caso das turbinas de combustão e a gás, microturbinas e motores de combustão interna alternativos [Kayal et al., 2017]

Usualmente é menos custoso instalar GDs em comparação com a implantação de novas linhas de distribuição e transmissão e novas plantas de geração de grande porte [Yuvaraj et al., 2018]. GDs trazem inúmeros benefícios para o sistema de distribuição dentre os quais pode-se citar: (i) aumento da confiabilidade da entrega de energia; (ii) possível solução para lidar com pico de demanda [Tolmasquim e Filho, 2019, p. 42]; (iii) redução da emissão de gases do efeito estufa (GDs de fonte renovável); (iv) fornecimento de suporte de tensão [Dulău et al., 2016]; (v) melhoria da qualidade de energia; (vi) redução de custos com combustível, operação e manutenção [Ali et al., 2017]. Contudo, existem malefícios possíveis na implementação de GDs na rede, dentre esses: (i) aumento do nível de curto circuito; (ii) funcionamento incorreto de relés de proteção [Wahyudi et al., 2019]; (iii) possíveis complicações no perfil de tensão e (iv) aumento das perdas em caso de posicionamento ruim da GD [Seepromting et al., 2018]; (v) problemas de planejamento da operação devido a dificuldades na previsão de geração pelas GDs, visto que a maior parte advém de fontes renováveis intermitentes [Tolmasquim e Filho, 2019, p. 42].

Conforme descrito em [SGT-ANEEL, 2020], as perdas no sistema de distribuição podem ser divididas em dois subgrupos: perdas técnicas (PTs) e perdas não técnicas (PNTs). As perdas técnicas são inevitáveis em qualquer sistema de distribuição de energia elétrica devido às características intrínsecas dos componentes da rede. Desta forma, essas perdas são calculadas de acordo com os Procedimentos de Distribuição (PRODIST), documento regulatório da ANEEL para normas da rede de distribuição. PTs causam impacto diretamente no custo da tarifa de energia, já que são parte do custo de operação das concessionárias de energia. PNTs, também chamadas de perdas comercias, advêm principalmente de furtos e erros na medição, sendo calculadas pela diferença entre perdas totais e perdas técnicas, e não serão alvo deste estudo, já que são associadas à gestão comercial das concessionárias.

Cenários variáveis de geração e consumo são realidades dos sistemas elétricos. Em uma sociedade, crescimento econômico e tecnológico, sazonalidade, dias da semana, dentre outros fatores, afetam as características de consumo, enquanto a dinamicidade das fontes primárias de algumas GDs, como eólica, solar e hídrica, afetam os perfis de geração. [Khan et al., 2016] reconhece a importância da previsão de carga para equilibrar demanda e suprimento de energia. Desse modo, fazer previsões corretas acerca das perdas elétricas de um sistema frente a vários cenários de geração é uma das possíveis maneiras de se lidar com o problema da variabilidade de geração.

Problemas muito comuns relacionados às GDs são alocação e dimensionamento destas, uma vez que, se não forem feitos corretamente, podem acarretar efeitos contrários aos benefícios, como aumentar perdas de potência e causar sobretensão [Biswas et al., 2017]. Inúmeros métodos

Galoá { Este trabalho foi publicado utilizando o **Galoá** proceedings

meta-heurísticos, analíticos e híbridos são empregados para resolver esta dificuldade [Bawazir e Cetin, 2020], todavia é importante salientar que o escopo deste trabalho visa lidar com um problema posterior ao de alocação e dimensionamento, isto é, deseja-se mostrar que as técnicas de inteligência artificial utilizadas são capazes de fazer previsões de perdas para vários cenários de geração sem a necessidade de executar um novo fluxo de potência para cada um desses, algo custoso computacionalmente, principalmente para sistemas reais de grandes porte. De forma sucinta, objetiva-se treinar, a partir de um banco de dados de fluxos de potência executados, os modelos para que estes forneçam resultados de perdas, eliminando a necessidade de execuções de fluxo de potência para novos cenários de geração distribuída.

Neste estudo, será utilizado o *Open Distribution System Simulator* (OpenDSS), *software* desenvolvido pelo EPRI, com intuito de simular redes de distribuição para criação de bancos de dados, isto é, executar fluxos de potência. Além disso, são aplicadas três técnicas de inteligência computacional no com objetivo de encontrar uma forma menos custosa computacionalmente, capaz de substituir os fluxos de potência necessários para a previsão de perdas de energia no sistema de distribuição.

2. Metodologias de Aprendizagem de Máquina

Neste trabalho são utilizadas três metodologias de aprendizagem de máquina, *Random Forest* (RF), *Support Vector Regression* (SVR) e *Multilayer Perceptron* (MLP), submetidas ao aprendizado supervisionado, no qual são apresentadas informações com as entradas do modelo e com o valor de saída que este modelo deveria retornar. Essas metodologias são empregadas em inúmeros tipos de problemas, como classificação, *clustering*, regressão e reconhecimento facial. Os métodos de regressão consistem, basicamente, em determinar um valor $f(x)$ por meio de x. Assim, este trabalho faz uso dessas metodologias em sua vertente de regressão, pois o objetivo é que esses modelos deem como resposta um valor de perda técnica para um determinado alimentador.

A RF é um método derivado da Árvore de Regressão, que se baseia em hierarquia e na importância dos seus ramos. Na regressão com RF é utilizado um grande número de árvores de regressão com subconjuntos de dados, e seu resultado é uma combinação dos resultados de todas as árvores, o que torna essa metodologia menos susceptível a *over-fitting* [Breiman, 2001].

O SVR é um aprimoramento do *Support Vector Machine* para problemas de regressão, que foi desenvolvido com base em técnicas de aprendizado estatístico, inicialmente, mais voltado à classificação, contudo, com bons desempenhos quando comparado a outras técnicas de aprendizagem de máquinas, novas derivações dessa metodologia foram surgindo. O SVR se diferencia de outras técnicas no tratamento do erro, pois, ao invés de tentar minimizar o erro observado de treinamento, tenta minimizar o limite do erro generalizado, buscando sempre um desempenho generalizado, além disso, é pouco influenciado por *outliers* [Drucker et al., 1997].

O MLP é uma categoria dentro de Redes Neurais Artificiais. O *perceptron* de camada única, ou apenas *perceptron*, é a estrutura mais básica de uma rede neural, capaz de resolver apenas problemas linearmente separáveis, a sua estrutura é composta basicamente do neurônio artificial com sua função de ativação e seus pesos sinápticos ajustáveis. Sendo assim, o MLP generaliza o conceito de *perceptron*, possibilitando a resolução de problemas com vários níveis de complexidade. Os variados tipos de função de ativação, presentes em cada neurônio artificial da rede, possibilitam a utilização desta para vários tipos de problemas, como classificação, regressão e outros. Os pesos sinápticos podem ser ajustados por diversas técnicas, porém a mais comum é baseada na retropropagação dos erros [Haykin, 1994].

3. Técnicas de Avaliação

A fim de avaliar os resultados dos testes dos modelos de previsão, são utilizadas métricas de erro voltadas a modelos de regressão. São elas: coeficiente de determinação (1), erro médio quadrático (2), erro absoluto médio (3) e maior erro percentual absoluto (4), onde y_i é a amostra real, $\hat{y}_i$ é a amostra de saída do modelo e max é uma função que retorna o maior valor.

$$R^2(y_i, \hat{y}_i) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}, \qquad \overline{y} = \sum_{i=1}^{n} y_i \tag{1}$$

$$MSE(y_i, \hat{y}_i) = \frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2 \tag{2}$$

$$MAE(y_i, \hat{y}_i) = \frac{1}{n}\sum_{i=0}^{n-1}|y_i - \hat{y}_i| \tag{3}$$

$$MAX(y_i, \hat{y}_i) = max\left\{100 * \frac{(y_i - \hat{y}_i)}{y_i}\right\} \tag{4}$$

A técnica de validação cruzada foi utilizada neste trabalho para maior confiabilidade dos resultados obtidos. Essa técnica permite avaliar a capacidade de generalização de um modelo de aprendizagem de máquina sem que o treinamento seja influenciado pelo banco de dados utilizado. A validação cruzada consiste no particionamento do banco de dados em N partes, dentre as quais uma ou mais são utilizadas no treinamento e as restantes são utilizadas no teste do modelo, conforme apresentado na Figura 1.

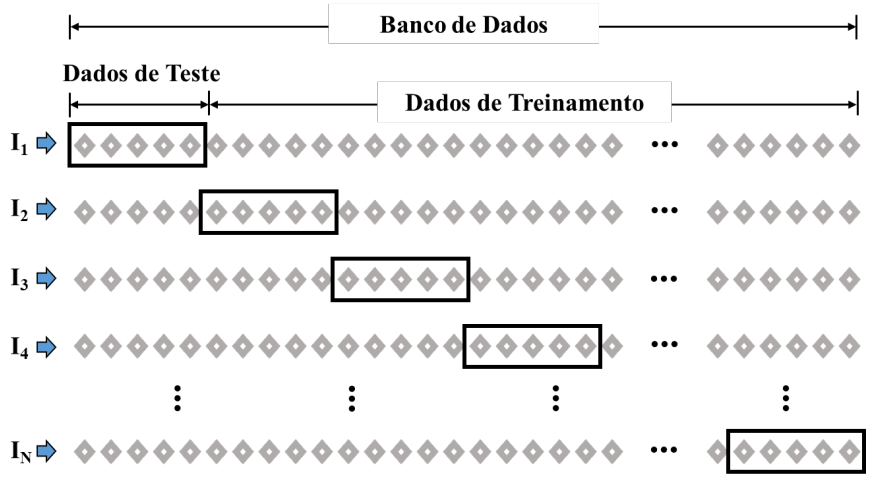


Figura 1: Modelo de avaliação cruzada.

Além da validação cruzada, também são realizadas repetições dos treinamentos dos modelos, partindo-se de sementes aleatórias diferentes. Sendo assim, os cálculos utilizados para avaliação do modelos são realizados com (5), onde M representa as repetições de treinamento, N, o número de partes em que o banco de dados é dividido na validação cruzada e $x(i, j)$, o resultado da métrica de erro para a repetição ij-ésima etapa da validação cruzada.

$$y = \frac{1}{M.N} \sum_{i=1}^{M} \sum_{j=1}^{N} x(i,j) \tag{5}$$

4. Dados do problema

O presente trabalho apresenta o estudo de um modelo de estimação de perdas técnicas realizado com três diferentes alimentadores. O primeiro é o CKT5, desenvolvido pelo EPRI, com características de um sistema de distribuição real e com modificações pertinentes para as simulações, como a inserção de GDs de características eólicas. Os outros dois alimentadores são reais e pertencentes ao banco de dados do "Grupo Energisa S.A.", tais alimentadores já possuem GDs ao longo das suas extensões, e estas são do tipo Central Geradora Hidrelétrica (CGH).

Conforme descrito no PRODIST, o cálculo de perdas técnicas para sistemas de distribuição em média tensão e baixa tensão (MT/BT) deve ser realizado com base em dias típicos de cada mês. Os dias típicos são separados em três: dia útil, sábado e domingo, em que, junto ao domingo, estão inclusos os feriados. Dessa maneira, são executadas três simulações para cada mês do ano e o cálculo das perdas anuais é realizado com a expansão dos resultados dessas simulações para o número de cada tipo de dia existente em cada mês.

Objetivando o desenvolvimento de um modelo capaz de estimar as perdas de um alimentador baseado nas penetrações das GDs a ele acopladas, fez-se uso das metodologias de aprendizagem de máquinas apresentadas anteriormente e de técnicas de avaliação dos resultados para validar tais modelos.

Os modelos são submetidos a uma aprendizagem supervisionada, ou seja, é apresentado aos modelos um banco de dados para treinamento com o intuito de que estes ajustem suas variáveis internas. Logo, foram criados bancos de dados para cada um dos três alimentadores mencionados. Os bancos de dados foram construídos com simulações realizadas com o *software* OpenDSS para duzentos diferentes cenários anuais de penetração das GDs. Portanto, foram realizadas ao todo 7200 simulações por alimentador, uma vez que cada cenário anual consiste em 36 simulações, sendo três dias típicos para cada mês.

O banco de dados utilizado para o treinamento e teste dos modelos de aprendizagem de máquina foi desenvolvido levando-se em consideração a penetração da GD ao longo de 24 horas, o tipo de dia típico, o mês e as perdas técnicas que representam a saída dos modelos.

O treinamento dos modelos de aprendizagem de máquina é mais eficaz quando as entradas são normalizadas e, devido a essa característica, algumas formas de tratamento desses dados foram utilizadas. A primeira modificação diz respeito à geração de cada uma das GDs, que foi normalizada com relação à própria capacidade máxima de geração da GD, logo, seus valores pertencem ao intervalo de zero a um. A segunda modificação é devido à existência de variáveis categóricas nominais, representadas pelo tipo de dia e meses, portanto, combinações binárias foram utilizadas para substituí-las. Os dias típicos foram substituídos por uma sequência de três dígitos binários, na qual "001" representa os domingos e os feriados, "010", os sábados e "100", os dias úteis. A mesma lógica segue para as substituições de meses, em que janeiro é representado por "1000000000000", fevereiro, por "010000000000" e assim sucessivamente até o mês de dezembro. A Figura 2 apresenta a configuração de entradas e saída do modelo para um alimentador com duas GDs acopladas.

5. Estudo de Casos

O estudo de casos foi realizado empregando a linguagem de programação Python, que apresenta bibliotecas robustas de aprendizagem de máquina, dentre as quais foram utilizadas a

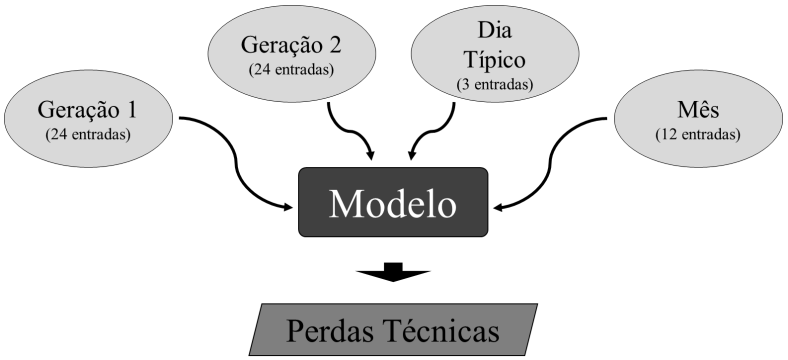


Figura 2: Modelo de estimação de perdas.

Scikit Learn com os métodos *Random Forest* e *Support Vector Regression* e a *Tensoflow/Keras* para *Multi-Layer Perceptron*.

Os modelos foram submetidos a uma análise de sensibilidade com relação a alguns de seus parâmetros internos, para que estes fossem definidos com valores únicos para todos os casos analisados neste trabalho. O número de árvores ou estimadores ficou definido como 300 e a métrica de erro médio quadrático, *mean square error* - MSE, foi utilizada na avaliação das divisões dos ramos, sendo os demais parâmetros deste modelo recebendo os valores padrões da biblioteca. No SVR, a função *kernel* foi definida como "*rbf*", o coeficiente do núcleo (γ) como "Scale"o que torna esse coeficiência adaptado ao número de amostras, o parâmetro de regularização (C) que está relacionado a penalização das funções, recebendo valor unitário, e o valor de distância (ϵ) igual a $0,001$, onde nenhuma penalidade é aplicada caso a função de perda do treinamento apresente valores inferiores ao de ϵ. No modelo MLP, foram definidos: 128 neurônios artificiais por camada oculta, 2 camadas ocultas, funções de ativação do tipo "ReLU", otimizador de parâmetros "Adam", máximo de 100 épocas de treinamento, 20% dos dados de treinamento separados para a validação do treinamento (ocorre durante a etapa de treinamento) e treinamento realizado em lotes de 10 amostras processadas antes da atualização das variáveis internas.

O banco de dados para cada caso consiste de 200 cenários de geração, ou seja, 7200 amostras para treinamento e teste. Em todos os treinamentos dos modelos de aprendizagem de máquina foi utilizada a validação cruzada, isto é, o banco de dados foi dividido em quatro partes, sendo apenas uma dessas partes utilizada para o treinamento e as outras três restantes utilizadas no teste. Isso se deve ao objetivo de se alcançar um modelo bem treinado e com a utilização de poucos cenários de geração.

Para a avaliação do desempenho dos modelos de treinamento, o estudo foi dividido em três casos, A, B e C, onde cada alimentador é trabalhado separadamente, com apresentação de suas características e adaptações quando existirem.

5.1. Caso A

O alimentador CKT5, da EPRI, apresenta, por padrão, curvas de carga anual, logo, modificações foram necessárias para que as simulações fossem realizadas com base no que é apresentado no PRODIST. Assim, para cada mês foram definidas três curvas de carga características para cada dia típico. Além disso, nesse alimentador não há a presença de geração distribuída em suas barras de média tensão, então, foram inseridas duas gerações eólicas em dois pontos do alimentador, conforme apresentado na Figura 3.

Galoá { Este trabalho foi publicado utilizando o **Galoá** proceedings

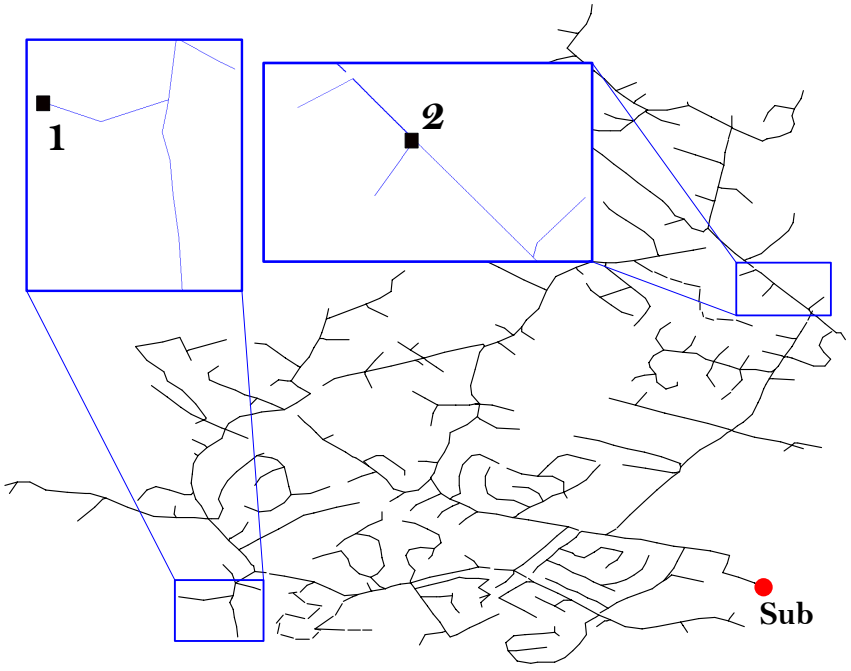


Figura 3: Localização das GDs eólicas no alimentador CKT5.

As duas GDs eólicas inseridas nesse alimentador apresentam as mesmas características, sendo 1MW de potência instalada, conexão trifásica e com fator de potência unitário, ou seja, sem injeção de reativo na rede de distribuição. Com relação às localizações das alocações, tem-se a eólica 1 inserida na barra 1127424 e a eólica 2 na barra 6578. Também são características desse alimentador: média e baixa tensão iguais a 12,47kV e 0,24kV, respectivamente; 2998 barras; 1379 consumidores, 591 transformadores e 2420 linhas.

A Tabela 1 apresenta os valores das quatro métricas de erro apresentadas na seção 3 e calculadas de acordo com a equação (5), referente à validação cruzada. Desta forma, esta análise para o alimentador em questão, permite verificar que o SVR foi o modelo que obteve melhor desempenho nas previsões, em consequência de melhores valores obtidos para todas as quatro métricas de erro.

Tabela 1: Avaliação das metodologias para o caso A.

	R^2	MSE	MAE	MAX
RF	0,997621	3,21E-05	0,002940	9,89
SVR	0,999739	3,52E-06	0,001236	3,61
MLP	0,998595	1,89E-05	0,002962	4,61

A Figura 4 apresenta os histogramas contendo os valores de erro percentual do banco de testes (150 cenários ou 5400 amostras), onde o eixo vertical representa a densidade de probabilidade. Em cada um dos gráficos são apresentados a média (m) e o desvio padrão (std), assim, pode-se concluir que, afastando-se da média uma distância igual a três vezes o desvio padrão, é possível afirmar que $99,6\%$ das amostras apresentaram erro no intervalo entre $-2,14\%$ e $2,06\%$ com a utilização do SVR.

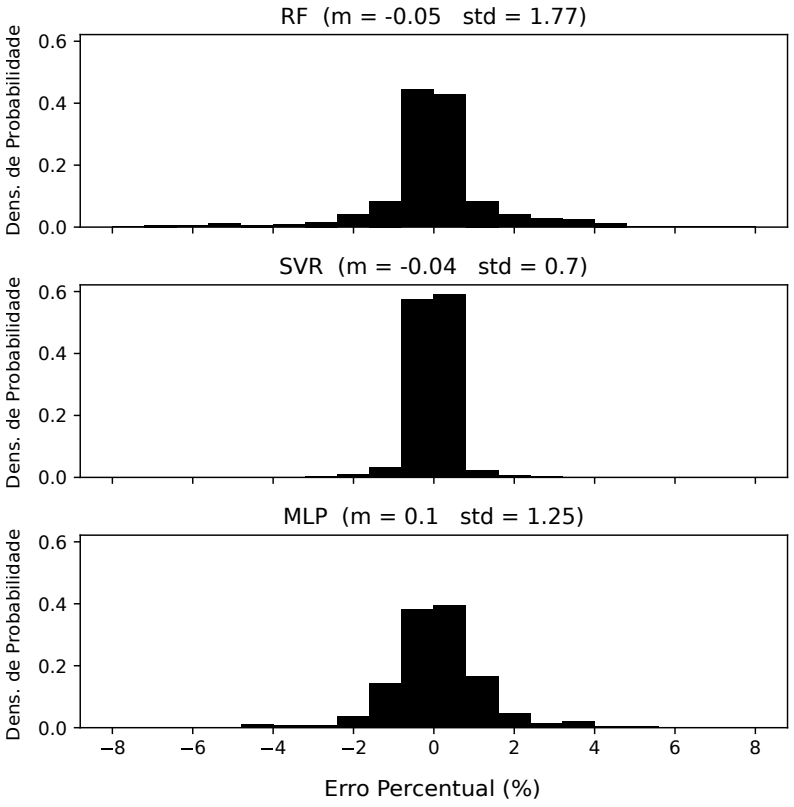


Figura 4: Distribuição de erros percentuais para o caso A.

5.2. Caso B

No caso B, foi utilizado um alimentador real, localizado na região sudeste do Brasil. Nesse alimentador existe a presença de uma CGH, com conexão trifásica em média tensão de 11,4kV e potência máxima de geração de 1MW. O alimentador possui como características: tensão de média e baixa iguais a 11,4kV e 0,22kV, respectivamente; 4533 barras; 4018 consumidores, 416 transformadores e 2053 linhas.

A análise das métricas de erro realizada a partir da Tabela 2 permite inferir que o SVR apresentou melhor desempenho frente às outras metodologias, contudo a métrica R^2 das metodologias SVR e MLP são iguais até a terceira casa decimal e o erro percentual absoluto máximo entre ambas é ligeiramente diferente.

Tabela 2: Avaliação das metodologias para o caso B.

	R^2	MSE	MAE	MAX
RF	0,998378	2,96E-05	0,001789	27,35
SVR	0,999872	2,34E-06	0,000932	8,62
MLP	0,999206	1,45E-05	0,002204	8,78

Nos histogramas da Figura 5, retirada de uma das simulações que mais generaliza os resultados, pode-se observar o bom desempenho dos três métodos, visto que suas densidades de

probabilidade são maiores quando próximas à média. Com base em três desvios padrão, pode-se inferir que $99,6\%$ das amostras testadas para o SVR se encontram no intervalo entre $-1,6\%$ e $1,76\%$ de erro e, no caso do MLP, entre $-1,71\%$ e $2,07\%$ de erro.

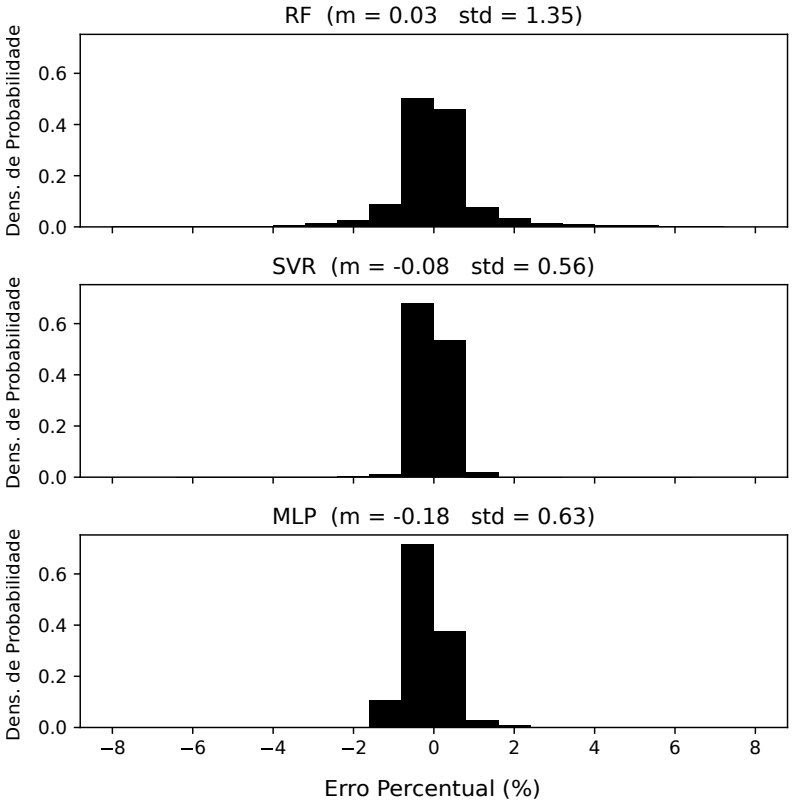


Figura 5: Distribuição de erros percentuais para o caso B.

5.3. Caso C

O segundo alimentador real é utilizado no caso C e também está localizado na região sudeste do Brasil. Duas CGHs estão presentes nesse alimentador, com potência máxima de geração iguais a 1,4MW e 0,51MW, sendo ambas trifásicas conectadas em barras de média tensão a 11,4kV. O alimentador apresenta as seguintes características: média e baixa tensão iguais a 11,4kV e 0,22kV, respectivamente; 1416 barras; 724 consumidores; 218 transformadores e 726 linhas.

Analisando a Tabela 3, na qual estão expostos os valores das métricas de erro para cada metodologia, observa-se que o MLP apresentou resultados superiores aos demais em três das quatro métricas. Os desempenhos dos modelos para este alimentador foram os piores quando comparados aos resultados dos casos A e B.

Tabela 3: Avaliação das metodologias para o caso C.

	R^2	MSE	MAE	MAX
RF	0,956175	3,09E-05	0,000536	42,18
SVR	0,972500	1,96E-05	0,000721	33,72
MLP	0,989871	6,80E-06	0,000778	17,03

Na Figura 6, onde são expostos os histogramas com os erros em porcentagem relativos a cada modelo, observa-se ainda a tendência de resultados mais concentrados próximos ao valor da média. Novamente, com base em três vezes o desvio padrão, pode-se inferir que o SVR apresenta $99,6\%$ de suas estimações com erros entre $-4,31\%$ e $4,51\%$, e o MLP entre $-2,65\%$ e $3,23\%$, apresentando estimações com menores erros.

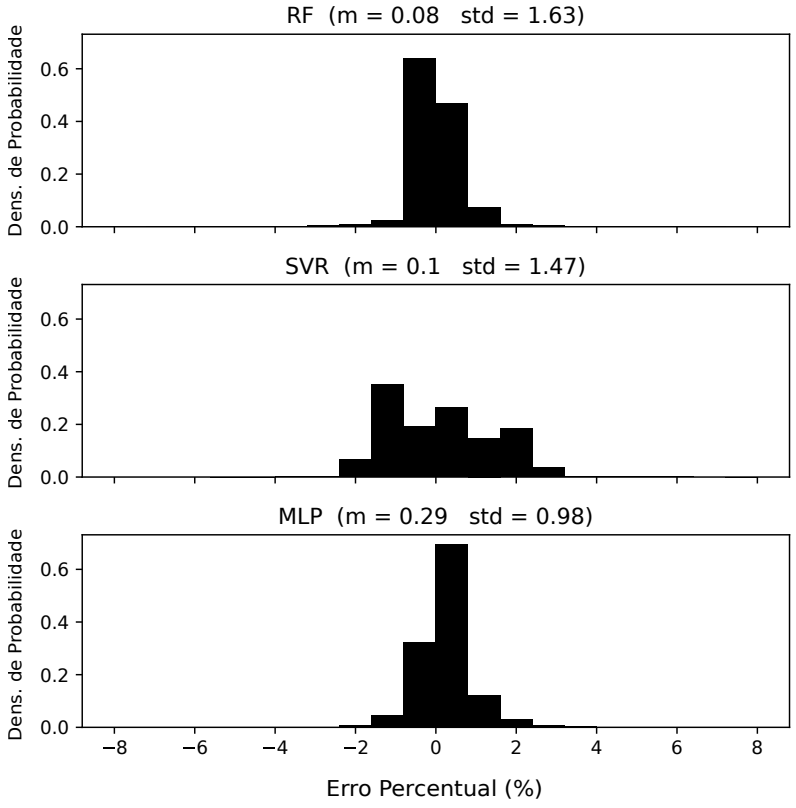


Figura 6: Distribuição de erros percentuais para o caso C.

6. Conclusão

O presente artigo apresentou um estudo da aplicação de três metodologias de aprendizagem de máquina, *Random Forest*, *Support Vector Regression* e *Multi-layer Perceptron*, na estimativa de perdas técnicas em sistemas de distribuição de energia elétrica.

Estas estimativas são importantes para as concessionárias de energia elétrica, uma vez que dependendo do nível de penetração proveniente de GD, da localização destas fontes de geração (níveis de tensão) e da potência injetada na rede pelas mesmas, as perdas técnicas podem sofrer variações significativas. Sendo assim, as GDs de natureza renováveis trazem para as perdas técnicas a estocasticidade característica de suas fontes, uma vez que as potências injetadas na rede dependem de ciclos (e.g., anos com maiores níveis de afluência, irradiação solar e vento).

Diante dos fatos apresentados, pode-se verificar através dos resultados que os modelos baseados em aprendizagem de máquina apresentaram-se eficientes nas estimações de perdas técnicas em sistemas de distribuição. Com a análise das métricas de erro aplicadas aos resultados

de cada modelo, pode-se afirmar que as três metodologias utilizadas apresentaram bons desempenhos. Contudo os modelos baseados em SVR e MLP se destacaram com resultados melhores e muito próximos, e com os gráficos de desvio padrão apresentados, pode-se observar que ambas metodologias obtiveram com erro percentual inferior a 5%, em módulo.

Como pesquisas futuras, pretende-se analisar a influência do número de cenários necessários para se obter um treinamento eficiente e com a necessidade de poucos cenários de treinamento. Também serão alvo de pesquisa, as variações no nível de carregamento do sistema de distribuição, novas técnicas de aprendizagem de máquina, além de modificações das técnicas utilizadas, com intuito de potencializar o desempenho dos modelos.

7. Agradecimentos

Os autores agradecem o suporte financeiro da ANEEL/ENERGISA Minas Gerais (Projeto PD-06585- 1802 / 2018), do CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil, ao INERGE - Instituto Nacional de Energia Elétrica e ao PPEE/UFJF - Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Juiz de Fora.

Referências

Adefarati, T. e Bansal, R. C. (2016). Integration of renewable distributed generators into the distribution system: a review. *IET Renewable Power Generation*, 10(7):873–884.

Ali, E. S., Abd Elazim, S. M., e Abdelaziz, A. Y. (2017). Ant lion optimization algorithm for optimal location and sizing of renewable distributed generations. *Renewable Energy*, 101:1311 – 1324. ISSN 0960-1481. URL http://www.sciencedirect.com/science/article/pii/S0960148116308126.

ANEEL (2020). Relatórios sobre as unidades consumidoras com geração distribuída. https://app.powerbi.com/view?r=eyJrIjoiZjM4NjM0OWYtN2IwZS00YjViLTllMjItN2E5MzBkN2ZlMzVkIiwidCI6IjQwZDZmOWI4LWVjYTctNDZhMi05MmQ0LWVhNGU5YzAxNzBlMSIsImMiOjR9. Acessado: 2020-06-01.

Bawazir, R. O. e Cetin, N. S. (2020). Comprehensive overview of optimizing pv-dg allocation in power system and solar energy resource potential assessments. *Energy Reports*, 6:173 – 208. ISSN 2352-4847. URL http://www.sciencedirect.com/science/article/pii/S2352484719302756.

Biswas, P. P., Mallipeddi, R., Suganthan, P., e Amaratunga, G. A. (2017). A multiobjective approach for optimal placement and sizing of distributed generators and capacitors in distribution network. *Applied Soft Computing*, 60:268 – 280. ISSN 1568-4946. URL http://www.sciencedirect.com/science/article/pii/S1568494617304052.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., e Vapnik, V. (1997). Support vector regression machines. In *Advances in neural information processing systems*, p. 155–161.

Dulău, L. I., Abrudean, M., e Bică, D. (2016). Optimal location of a distributed generator for power losses improvement. *Procedia Technology*, 22:734 – 739. ISSN 2212-0173. URL http:

//www.sciencedirect.com/science/article/pii/S2212017316000335. 9th International Conference Interdisciplinarity in Engineering, INTER-ENG 2015, 8-9 October 2015, Tirgu Mures, Romania.

Haykin, S. (1994). *Neural networks: a comprehensive foundation*. Prentice Hall PTR.

Kayal, P., Chanda, S., e Chanda, C. K. (2017). An analytical approach for allocation and sizing of distributed generations in radial distribution network. *International Transactions on Electrical Energy Systems*, 27(7):e2322. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/etep.2322. e2322 ETEP-15-0864.R3.

Khan, A. R., Mahmood, A., Safdar, A., Khan, Z. A., e Khan, N. A. (2016). Load forecasting, dynamic pricing and dsm in smart grid: A review. *Renewable and Sustainable Energy Reviews*, 54:1311 – 1322. ISSN 1364-0321. URL http://www.sciencedirect.com/science/article/pii/S136403211501196X.

Seepromting, K., Chatthaworn, R., Khunkitti, P., Kruesubthaworn, A., Siritaratiwat, A., e Surawanitkun, C. (2018). Optimal grid-connected with multi-solar pv placement and sizing for power loss reduction and voltage profile improvement. In *2018 18th International Symposium on Communications and Information Technologies (ISCIT)*, p. 479–483. IEEE.

SGT-ANEEL (2020). Perdas de energia. https://www.aneel.gov.br/metodologia-distribuicao/-/asset_publisher/e2INtBH4EC4e/content/perdas/654800. Acessado: 2020-06-01.

Tolmasquim, M. T. e Filho, M. M. (2019). *Recursos energéticos distribuídos e suas potencialidades*. Synergia, Rio de Janeiro, 1 edition. ISBN 9786550100148.

Wahyudi, C. A. D., Hariyanto, N., e Ganjavi, R. (2019). Adaptive protection of distribution systems with ders considering consumer and generation profiles. In *2019 International Conference on Smart Energy Systems and Technologies (SEST)*, p. 1–6. IEEE.

Yuvaraj, T., Devabalaji, K. R., e Ravi, K. (2018). *Optimal Allocation of DG in the Radial Distribution Network Using Bat Optimization Algorithm*, p. 563–569. Springer Singapore, Singapore. ISBN 978-981-10-4394-9. URL https://doi.org/10.1007/978-981-10-4394-9_55.