

# Localização de Falhas em Sistema de Distribuição com Medidores Inteligentes Utilizando Árvores de Decisão

Marcel A. Araújo\*, Rogério A. Flauzino\*\*, Lucas A. Moraes\*\*, Fabbio A. S. Borges\*\*\*, Fabiano Salvadori\*\*\*\*

\*Unidade Acadêmica do Cabo de Santo Agostinho (UACSA), Universidade Federal Rural de Pernambuco (UFRPE)  
Cabo de Santo Agostinho, PE, (e-mail:marcel.araujo@ufrpe.br)

\*\*Departamento de Engenharia Elétrica e de Computação, Universidade de São Paulo (USP)  
São Carlos, SP, (e-mail: raflauzino@sc.usp.br, lucas.moraes@usp.br)

\*\*\*Departamento de Computação de Piripiri, Universidade Estadual do Piauí (UESPI)  
Piripiri, PI, (e-mail: fabbio@prp.uespi.br)

\*\*\*\* Smart Grid Group, Universidade Federal da Paraíba  
João Pessoa, PB, (e-mail: salvadori.fabiano@gmail.com)

---

**Abstract:** The outages of power supply caused by faults in electric power systems stand out because of the potential damage that they can cause to the constituent elements of the distribution systems, the concessionaires and the consumers, compromising the power quality available. In this context, this work aims to locate faults in distribution systems employing a capacity of monitoring the voltage of the smart meters allocated in different points of the electric network. The proposed approach is based on the use of voltage data and the sensitization of the passage of electric current by the smart meters to indicate the location of occurrence of the fault in relation to the meters through intelligent systems based on Decision Trees. The results obtained demonstrate that the use of redundant information provided by several smart meters minimizes the localization errors, which reduces the search by the place where the fault occurred by the network maintenance teams.

**Resumo:** As interrupções de energia elétrica causadas por faltas nos sistemas elétricos de potência se destacam em virtude do potencial dano que podem causar aos elementos constituintes dos sistemas de distribuição, às concessionárias e aos consumidores, comprometendo a qualidade da energia disponibilizada. Neste contexto, este trabalho objetiva localizar faltas em sistemas de distribuição empregando a capacidade de monitoramento da tensão dos medidores inteligentes alocados em diferentes pontos da rede elétrica. A abordagem proposta baseia-se na utilização de dados de tensão e da sensibilização a passagem de corrente elétrica pelos medidores inteligentes para apontamento do local de ocorrência da falta em relação aos medidores por meio de sistemas inteligentes baseados em Árvores de Decisão. Os resultados obtidos demonstram que o uso de informações redundantes fornecidas por vários medidores inteligentes minimiza os erros de localização, o que reduz a busca pelo local de ocorrência da falta pelas equipes de manutenção da rede elétrica.

**Keywords:** Decision tree, Fault location, Distribution systems, Smart grid, Smart meters.

**Palavras-chaves:** Árvores de decisão, Localização de faltas, Medidores inteligentes, Redes elétricas inteligentes, Sistemas de Distribuição.

---

## 1. INTRODUÇÃO

O crescente e contínuo desenvolvimento tecnológico e econômico atrelado ao aumento do consumo de energia elétrica vem impulsionando a permanente ampliação e modernização dos Sistemas Elétricos de Potência (SEP). Em meio às soluções em desenvolvimento a de maior destaque é a implantação das Redes Elétricas Inteligentes (REIs), as quais preconizam fornecer energia com altos índices de confiabilidade, continuidade e segurança. Para tanto, por meio das mesmas busca-se desenvolver e implantar nas redes elétricas ferramentas de monitoramento, comunicação, informação, controle e medição (Borlase, 2013; Bayindir et al., 2016; Ghosal e Conti, 2019; Kumar et al., 2019).

Um dos grandes desafios do setor elétrico deve-se a interrupções não programadas no fornecimento de energia por desligamentos de partes ou de todo o alimentador de distribuição, causadas por faltas temporárias ou permanentes. A ocorrência de faltas nos Sistemas de Distribuição (SDs) deve-se a diferentes fatores, tais como problemas elétricos, mecânicos, de isolamento, de manutenção e de natureza térmica, entre outros (Anderson, 1995; Glover; Sarma; Overbye, 2011). Devido aos danos físicos para a rede elétrica e seus equipamentos sua correção deve ser realizada por equipes de manutenção de forma rápida e eficiente vislumbrando a operação segura e econômica do SEP (Batista, 2016; Bahmanyar, 2017; Beheshtaein et al., 2019).

Além disso, as interrupções causadas por faltas no sistema elétrico podem ocasionar condições de operação indesejáveis,

riscos aos usuários, transtornos a consumidores residenciais, e interrupção ou falha de processos industriais. Este conjunto de aspectos afeta diretamente os indicadores de qualidade e continuidade da energia, implicando em elevados custos operacionais às concessionárias devido às compensações aos consumidores afetados (Short, 2004; Dugan et al., 2012; Grainger et al., 2015).

Nesta conjuntura, a pesquisa desenvolvida tem como objetivo a localização de faltas em SDs por meio da capacidade de monitoramento da rede e aquisição de sinais proporcionada pela infraestrutura de medição avançada das redes elétricas inteligentes. Para tanto, inicialmente foi feita a aquisição de sinais por diversos medidores inteligentes espalhados em um alimentador de distribuição para formação de um banco de dados de tensão. Em seguida, foi realizada a extração de informações deste banco de dados por meio de sistema inteligentes pautados na aplicação de um algoritmo de Árvores de Decisão (AD). Por fim, cruzando tais informações com a topologia da rede elétrica sob análise foi possível o apontamento da distância entre os medidores inteligentes e o local de ocorrência da falta.

Posto isto, na sequência, a segunda seção apresenta os principais aspectos referentes às Árvores de Decisão. A terceira seção expõe os detalhes sobre a implementação computacional do alimentador e formação do banco de dados de tensão. Na quarta seção são mostrados os resultados da aplicação do algoritmo de Árvores de Decisão sob o banco de dados de tensão para a localização de faltas. Finalmente, as conclusões da pesquisa são apresentadas na quinta seção.

## 2. ÁRVORES DE DECISÃO

A aplicação de técnicas e ferramentas de Aprendizado de Máquina (AM) tem se destacado nos processos de descoberta de conhecimento por meio de algorítmicos, utilizando representações de modelos mais flexíveis e heurísticas para a realização de busca. Em virtude do aumento do volume e variedade de dados, ampliação da potencialidade de processamento computacional com menores custos, e necessidade de armazenamento de grande massa de dados de forma mais acessível, o desenvolvimento e aplicação das técnicas de AM tem ganho destaque. Isto decorre da sua capacidade de inferir modelos automatizados com baixo esforço computacional para análise de grandes e complexos bancos de dados, fornecendo resultados rápidos e precisos, mesmo que em larga escala (Freitas, 2002; Haykin, 2008; Bell, 2014).

O conhecimento formalizado em bases de conhecimento pode ser representado por meio de regras, mapeadas em Árvores de Decisão (ADs) cujo objetivo é criar um modelo viável que preveja o valor de uma variável desconhecida com base no conjunto de variáveis de entrada. Amplamente utilizadas em algoritmos de classificação, as ADs são representações simples do conhecimento e um meio eficiente de construir classificadores que predizem classes baseadas nos valores de atributos de um conjunto de dados. Esta técnica de Aprendizado de Máquina utiliza uma representação em árvore para classificação e predição das amostras desconhecidas, sendo estruturada hierarquicamente em um conjunto de nós

interconectados (Theodoridis e Koutroumbas, 2008; Haykin, 2008; Bell, 2014, Gollapudi, 2016).

Quando se trabalha com a aplicação dos conceitos de AD sobre um conjunto de dados espera-se que ela classifique corretamente exemplos desconhecidos, porém verifica-se que na prática podem ocorrer classificações equivocadas. Uma forma de quantificar o nível de acerto e/ou erro de classificação é por meio da matriz de confusão de um classificador, a qual também fornece informações sobre a acurácia, sensibilidade, especificidade e confiabilidade positiva e negativa.

Além disso, o software WEKA (Weka, 2020), empregado na presente pesquisa, fornece métricas adicionais sobre os classificadores das ADs induzidas, os quais são: Estatística Kappa, Erro Absoluto Médio e Erro Relativo Absoluto.

A Estatística Kappa é uma medida que indica o grau de concordância entre dois classificadores mostrando quanto a concordância observada se aproxima da esperada, considerando para tanto as probabilidades de as concordâncias terem acontecido (Landis e Koch, 1977; McHugh, 2012). Uma das formas de calculá-la é por meio de (1), em que  $k$  é o valor da Estatística Kappa  $P(a)$  é a probabilidade da concordância observada e  $P(b)$  da concordância esperada.

$$k = \frac{P(a)-P(b)}{1-P(b)} \quad (1)$$

Para a avaliação de modelos induzidos, como os classificadores de AD, a Estatística Kappa é calculada entre o classificador e o conjunto de treinamento, este último considerado como um classificador infalível. Segundo o apresentado por (Landis e Koch, 1977) a Tabela 1 fornece uma interpretação para a Estatística Kappa, relacionando seu valor com o grau de concordância dos classificadores.

**Tabela 1: Interpretação da Estatística Kappa (adaptado de Landis e Koch, 1977).**

<b>k</b>	<b>Interpretação</b>
< 0	nenhuma concordância
0 - 2	concordância baixa
0,21 – 0,4	concordância regular
0,41 – 0,6	concordância moderada
0,61 – 0,8	concordância substancial
0,81 - 1	concordância quase perfeita

Para os cálculos dos outros parâmetros pelo WEKA deve-se tomar a seguinte notação: dada a  $i$ -ésima amostra na forma  $(x_i, y_i)$ ,  $x_i$  é o atributo e  $y_i = f(x_i)$  a classe rotulada pelo processo de classificação, em que  $h(\cdot)$  é o classificador ou hipótese, e  $n$  é o número de exemplos (Landis e Koch, 1977; McHugh, 2012). Tendo em vista esta notação, a relação entre os valores exatos e aproximados para os classificadores pode ser estimada pelos seus erros, dentre os quais destaca-se o Erro Absoluto Médio (EAM) por meio de (2):

$$EAM = \frac{1}{n} \sum_{i=1}^n |y_i - h(x_i)| \quad (2)$$

Já o Erro Relativo Absoluto (ERA) pode ser calculado a partir de (3), em que  $\bar{y}$  é a média dos valores para todas as classes.

$$ERA = \frac{\frac{1}{n} \sum_{i=1}^n |y_i - h(x_i)|}{\frac{1}{n} \sum_{i=1}^n |y_i - \bar{y}|} \quad (3)$$

### 3. METODOLOGIA

A ideia geral do método desenvolvido para localização de faltas foi compor um banco de dados a partir das informações fornecidas pelos medidores inteligentes instalados em todos os transformadores de um determinado sistema de distribuição. Deste banco de dados podem ser extraídas informações do sistema elétrico tais como: matrizes admitância e impedância, informações de curto-circuito, tensões em cada uma de suas barras, topologia de rede, entre outras. Sobre estas informações foi realizada a aplicação de ferramentas de Aprendizado de Máquina, especificamente um algoritmo de Árvores de Decisão, para extração e processamento do conhecimento, o qual em conjunto com informações do sistema elétrico fornece o local ou região de possível ocorrência da falta.

O alimentador piloto escolhido para avaliação nesta pesquisa faz parte do sistema de distribuição de uma concessionária de energia elétrica do Estado de São Paulo, opera em 13,8 kV e a disposição geográfica da sua rede primária é apresentada na Fig. 1. O presente alimentador de distribuição foi modelado

com 1600 barras, a contar com a subestação, e 505 pontos de medição pela alocação de medidores inteligentes junto ao lado de média tensão de cada um dos transformadores do sistema. Na Fig. 1 a subestação está representada por um círculo em alaranjado, as barras do sistema por anéis na cor ciano, e os medidores inteligentes alocados em cada um dos 505 transformadores por círculos em preto. Além disso, para as simulações considerou-se que antes da ocorrência das faltas as tensões de fase do sistema têm mesmo módulo e estão defasadas de 120°, o sistema opera sem cargas e sem conexão de geradores distribuídos, e são simuladas faltas francas entre a Fase A e a terra em cada uma de suas 1599 barras (não foi simulada falta na barra da subestação).

A partir das simulações, o banco de dados foi formado por 25 tensões, 2 impedâncias e 1 indicação de sensibilização por corrente de falta para cada um dos 505 medidores para as 1599 faltas simuladas, totalizando um conjunto de dados com 22.609.860 parâmetros. Ressalta-se que os medidores e barras do sistema foram numerados consecutivamente a partir da subestação, e ao longo da aplicação da metodologia proposta para localização de faltas foi assumida a mesma numeração inicialmente definida.

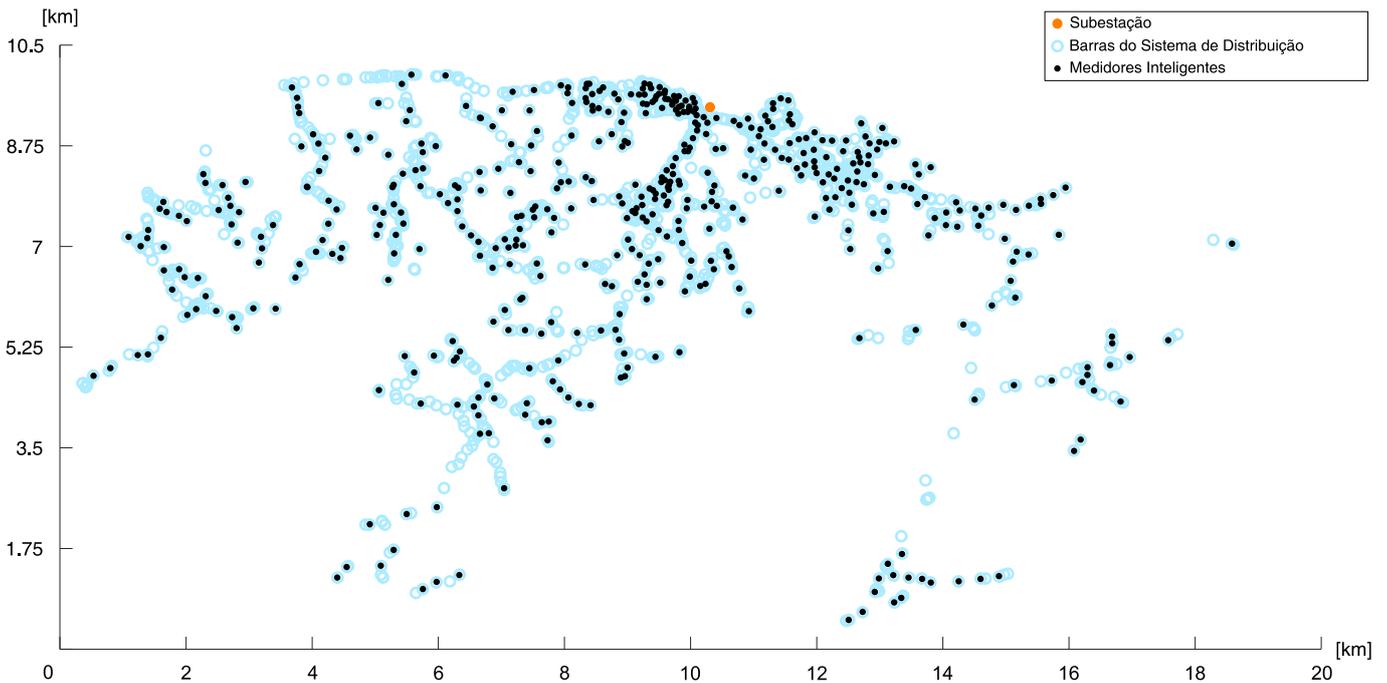


Figura 1 - Disposição geográfica da rede primária do alimentador de distribuição real.

O método desenvolvido baseia-se na análise da propagação de desequilíbrios de tensão pelo sistema elétrico em virtude da ocorrência de faltas, os quais apresentam diferentes variações nos distintos pontos da rede em função do local da falta e das características do sistema de distribuição. A partir da aquisição e avaliação das tensões de fase ( $V_A$ ,  $V_B$  e  $V_C$ ) e cálculo dos módulos das tensões de linha e de sequência separando suas partes reais e imaginárias ( $V_{AB}$ ,  $V_{BC}$ ,  $V_{CA}$ ;  $V_0$ ,  $V_1$ ,  $V_2$ ;  $V_{BR}$ ,  $V_{Blm}$ ,  $V_{CR}$ ,  $V_{Clm}$ ;  $V_{ABR}$ ,  $V_{ABlm}$ ,  $V_{BCR}$ ,  $V_{BClm}$ ,  $V_{CAR}$ ,  $V_{CALm}$ ;  $V_{0R}$ ,  $V_{0lm}$ ,  $V_{1R}$ ,  $V_{1lm}$ ,  $V_{2R}$ ,  $V_{2lm}$ ), busca-se estimar por meio de dois sistemas inteligentes distintos (SI-1 e SI-2) as impedâncias de sequência zero ( $Z_{0k}$ ) e positiva ( $Z_{1k}$ ) entre os “k” medidores e o local da

falta. Posteriormente, são calculadas as distâncias elétricas ( $d_{0k}$  e  $d_{1k}$ ) equivalentes dos “k” medidores inteligentes em relação ao ponto sob falta a partir dos valores das impedâncias de sequência zero e positiva.

Além disso, a metodologia proposta também faz uma avaliação da sensibilização dos “k” medidores inteligentes pelas correntes de falta, estimando por meio de outro sistema inteligente independente (SI-3) a indicação da sensibilização (+1) ou não (-1) dos mesmos frente a ocorrência de faltas no sistema de distribuição, como ilustrado na Fig. 2. Para tanto, foram definidas as 25 tensões como os atributos de entrada dos

três sistemas inteligentes, e as impedâncias de sequência zero e positiva como saídas de SI-1 e SI-2, e a indicação de sensibilização como saída de SI-3.

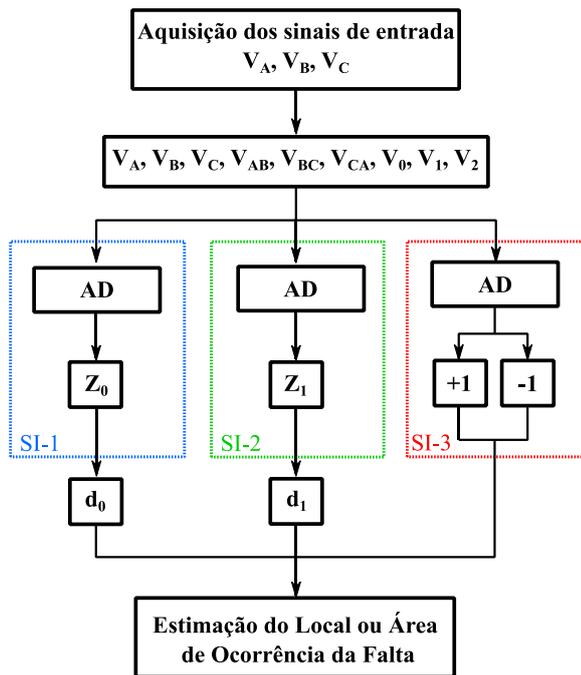


Figura 2 - Ilustração representativa do método proposto.

A fim de configurar um processo de classificação, buscou-se a relação entre os desequilíbrios de tensão e faixas de classificação das impedâncias, e entre os desequilíbrios de tensão e a sensibilização ou não dos medidores inteligentes pela passagem de correntes de falta, empregando a técnica de aprendizagem de máquina denominada Árvore de Decisão.

Conforme ilustrado pela Fig. 2, a partir do cruzamento das informações obtidas com as distâncias elétricas da falta em relação aos medidores e da sensibilização dos mesmos, foi possível então estimar o local ou região de ocorrência da falta no sistema de distribuição por meio de um algoritmo de baixo esforço computacional e resultados confiáveis empregando-se AD.

#### 4. IMPLEMENTAÇÃO DOS SISTEMAS INTELIGENTES

A partir do banco de dados formado com as informações obtidas pela modelagem e simulação computacional do sistema elétrico sob estudo nos softwares MATLAB® e Simulink, deu-se início a implementação dos três sistemas inteligentes distintos (SI-1, SI-2 e SI-3) para estimativa das impedâncias de sequência zero ( $Z_{0k}$ ), impedâncias de sequência positiva ( $Z_{1k}$ ), e da sensibilização dos medidores inteligentes pelas correntes de falta, respectivamente. Para tanto, fez-se uso do algoritmo J48 da técnica de classificação Árvore de Decisão (AD) presente no software livre WEKA (Weka, 2020).

O algoritmo J48 é a implementação na linguagem Java do algoritmo C4.5, o qual constrói um modelo de AD baseado em um conjunto de dados de treinamento para posterior classificação de outras instâncias de um conjunto de teste. Sua função de avaliação, utilizada para escolher qual atributo determinará a ramificação da árvore, baseia-se na análise da

função do ganho de informação do atributo. O algoritmo J48 utiliza técnicas de poda para remover da AD os segmentos que possuem menor ganho de informação, inibindo a ocorrência de *overfitting*. Destaca-se que no presente contexto, *overfitting* se refere a problemas de sobreajuste da árvore a partir de um aprendizado muito específico e da geração de um modelo de baixa generalização.

A fim de evitar o *overfitting* do modelo de AD a ser criado optou-se pela validação cruzada como método de treinamento e teste, pois por meio do mesmo os exemplos são aleatoriamente divididos em “n” partições mutuamente exclusivas de tamanho aproximadamente igual e cada um dos exemplos é utilizado tanto para treinamento quanto para teste. Na literatura correlata, testes extensivos mostram que *10-fold cross-validation*, ou seja, assumir  $n = 10$ , possibilita testes bastante eficientes para a validação cruzada, conforme exposto em (Kohavi, 1995; Witten e Frank, 2005; Delen; Walker; Kadam, 2005), e dessa forma este foi o valor padrão adotado no desenvolvimento da AD desta pesquisa.

Para indução da AD no software WEKA pelo algoritmo J48 foi necessário deixar o comando “*unpruned*” marcado como “*False*” para que fosse realizada a poda da árvore pelo algoritmo, e o comando “*subtreeRaising*” foi assumido como “*True*” para que ocorresse a poda em cada sub-árvore criada. Foi tomada especial atenção com os parâmetros de entrada, “*confidenceFactor*” e “*minNumObj*”, os quais interferem diretamente nas regras criadas no processo de indução da AD, e, por conseguinte, nos resultados de sua aplicação para classificação. Após vários testes observou-se que para os classificadores de maior acerto dos três sistemas inteligentes (SI-1, SI-2 e SI-3) o valor dos parâmetros “*confidenceFactor*” e “*minNumObj*” deveria ser 0,2 e 0,35, 0,2 e 0,3 e 0,25 e 0,3 respectivamente.

Trabalhando com todo o banco de dados procurou-se definir a melhor relação entre o número de divisões de classes das impedâncias e a taxa de acerto dos classificadores. Para que isto fosse feito foi proposta a divisão das impedâncias de sequência zero e positiva em 5 conjuntos de classes, da seguinte forma: conjunto 1 - 2 classes; conjunto 2 - 3 classes; conjunto 3 - 5 classes; conjunto 4 - 7 classes e conjunto 5 - 10 classes.

Além disso, no processo de treinamento dos sistemas inteligentes foi utilizado o maior número de dados possível desde que cada uma das classes tivessem o mesmo número de dados, e para validação da metodologia foi empregado o restante dos dados. Destaca-se que este processo de formação de classes com a mesma quantidade de dados dentro de um mesmo conjunto visa selecionar os dados de maior informação e também reduzir o conjunto de dados para treinamento dos sistemas inteligentes.

Posto isto, realizando testes com os 5 conjuntos de classes de impedâncias tanto para  $Z_0$  quanto para  $Z_1$ , nota-se por meio da Fig. 3 que os acertos dos classificadores para ambas as impedâncias são maiores para a divisão dos dados em 3 classes diferentes. Além disso, a medida que se dividem os dados em um número de classes maior, o acerto diminui para ambos os casos.



Figura 3 - Taxa de acerto dos classificadores das impedâncias  $Z_0$  e  $Z_1$ .

Na Fig. 4 os índices da Estatística Kappa seguem padrões semelhantes aos dos acertos dos classificadores a medida que se aumenta o número de classes. Destaca-se que para os classificadores de maior acerto (3 classes), a Estatística Kappa indica concordância substancial entre os classificadores dos modelos induzidos e os do conjunto de treinamento, corroborando com o maior acerto verificado para os mesmos.

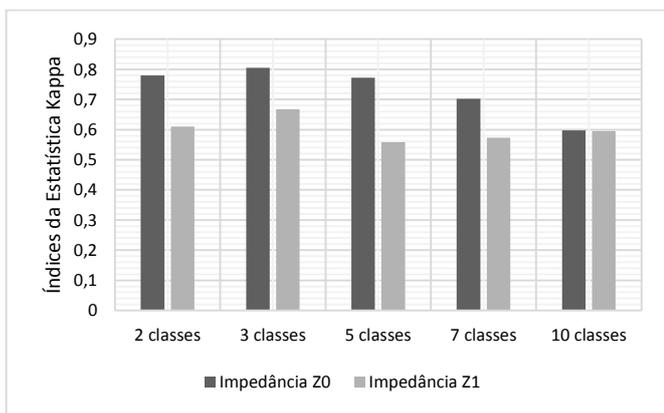


Figura 4 - Índices da Estatística Kappa para os classificadores das impedâncias  $Z_0$  e  $Z_1$ .

Na Fig. 5 também se confirma que os melhores classificadores para as impedâncias  $Z_0$  e  $Z_1$  são os que utilizam a divisão das impedâncias em 3 classes, uma vez que possuem menor erro relativo absoluto. Já na Fig. 6 a taxa do erro absoluto médio é menor para a divisão em 5, 7 e 10 classes para a impedância  $Z_0$  e em 10 classes para a impedância  $Z_1$ . Contudo, apesar deste último parâmetro discordar dos anteriores, pode-se concluir que os melhores classificadores para as impedâncias  $Z_0$  e  $Z_1$  são os que utilizam a divisão das impedâncias em 3 classes e, portanto, foram utilizados no processo de classificação das mesmas, as quais posteriormente servirão de base para o cálculo equivalente das distâncias elétricas. Ressalta-se que a adoção da divisão das impedâncias em 3 classes com igual quantidade de dados implica no emprego de 4,56% e 2,51% do conjunto total de dados para treinamento dos classificadores de  $Z_0$  e  $Z_1$ .

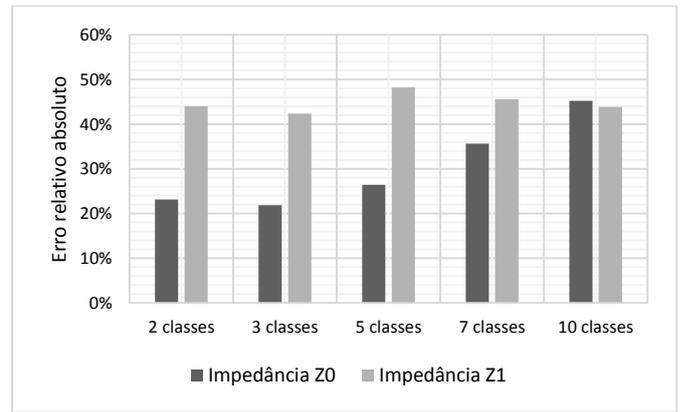


Figura 5 - Taxa de erro relativo absoluto para os classificadores das impedâncias  $Z_0$  e  $Z_1$ .

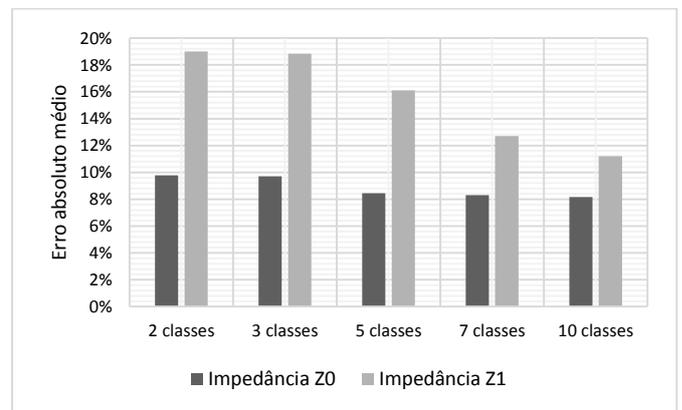


Figura 6 - Taxa de erro absoluto médio para os classificadores das impedâncias  $Z_0$  e  $Z_1$ .

Analisando agora as matrizes de confusão dos classificadores de  $Z_0$  e  $Z_1$  para as impedâncias divididas em 3 classes conforme apresentado na Tabela 2 e na Tabela 3 confirma-se o grau de acurácia total (ou taxa de acerto) e de precisão de cada uma de suas classes. A partir dos dados da Tabela 2 verifica-se que a acurácia do classificador da impedância  $Z_0$  é 87,02%, e a precisão das classes A, B e C é 70,03%, 90,56% e 96,83%, respectivamente. Já a partir dos dados da Tabela 3 averigua-se que a acurácia do classificador da impedância  $Z_1$  é 77,92%, e a precisão das classes A, B e C é 62,56%, 80,20% e 90,41%, respectivamente.

Tabela 2: Matriz de confusão para o melhor classificador de  $Z_0$ .

	A	B	C
A	7304	3012	113
B	866	11796	363
C	87	336	12921

Tabela 3: Matriz de confusão para o melhor classificador de  $Z_1$ .

	A	B	C
A	4139	1475	1002
B	439	5448	906
C	74	589	6247

Por fim, cabe destacar que após treinamento e teste o classificador de sensibilização dos medidores inteligentes frente a passagem de corrente de falta obteve taxa de acerto de 94,06%.

## 5. RESULTADOS

A fim de validar a metodologia proposta foram empregados os classificadores implementados e ajustados, conforme exposto na seção anterior, sob os dados de simulação de faltas sobressalentes do processo de treinamento.

Após a indução das árvores de decisão dos três sistemas inteligentes foram repassados novos valores de tensão de fase, de linha e de sequência para os classificadores, a partir dos quais foram estimadas as impedâncias  $Z_0$  e  $Z_1$  entre os medidores e o local da falta por meio dos processos de classificação das ADs. Em seguida, foram calculadas as distâncias elétricas equivalentes, em quilômetros, entre os medidores inteligentes e o ponto sob falta no sistema de distribuição sob análise.

Na Fig. 7 é apresentada a distribuição do erro relativo entre os valores esperados e os estimados para a distância  $d_0$ , por meio dos quais observa-se que o classificador estima a distância  $d_0$  para todo o espectro de valores esperados. Destaca-se que 21,84% das distâncias foram estimadas corretamente sem erro, 54,05% possuem erro entre -5 e +5 km, 77,39% têm erro entre -10 e +10 km, e 89,04% possuem erro entre -15 e +15 km. Com os resultados obtidos desprende-se também que entre 0 e +5 km registra-se a faixa com o maior número de ocorrências do erro para a distância  $d_0$ , com cerca de 38,12% do total dos erros.

Para fins de localização de falta, ainda avaliando o erro de estimação da distância  $d_0$  foi calculado seu erro absoluto médio, o qual para os dados de validação é 8,33 km. Cabe destacar que para o alimentador de distribuição analisado esta faixa de valores de erro da distância poderia auxiliar a localização do ponto em que ocorreu a falta. Contudo, devido a necessidade de correção rápida e eficiente de uma falta no SEP, a busca pelo local exato da falta em um trecho de até 8,33 km por equipes de manutenção é considerada demorada e onerosa pelas concessionária de energia.

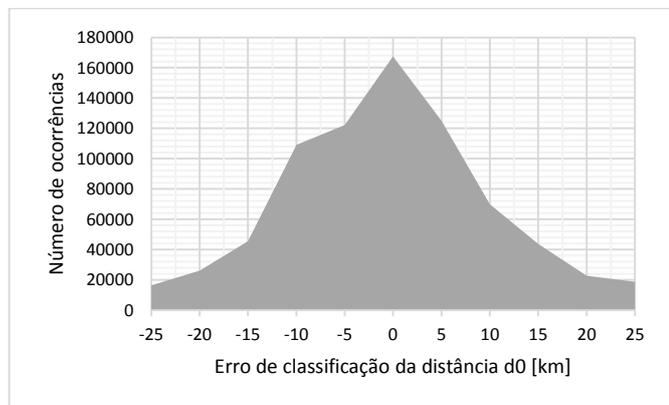


Figura 7 - Erro de estimação da distância da falta em função da impedância  $Z_0$  para o conjunto de dados de validação.

Já na Fig. 8 é exposta a distribuição do erro relativo entre os valores esperados e os estimados para a distância  $d_1$ , por meio dos quais também se nota que o classificador estima esta distância para todo o espectro de valores esperados. Ressalta-se que 17,44% das distâncias foram estimadas corretamente sem erro, 46,98% possuem erro entre -5 e +5 km, 70,37% têm erro entre -10 e +10 km, e 83,94% possuem erro entre -15 e +15 km. Com os resultados obtidos é possível observar ainda que entre 0 e +5 km registra-se a faixa com o maior número de ocorrências do erro para a distância  $d_1$ , com cerca de 32,63% do total dos erros.

Novamente, para fins de localização de falta, avaliando o erro de estimação da distância  $d_1$  foi calculado seu erro absoluto médio, o qual para os dados de validação é 9,59 km. Erro este que é maior que o obtido para a distância  $d_0$ , logo apesar de auxiliar no processo de localização da falta é considerado muito alto pelas concessionária de distribuição de energia.

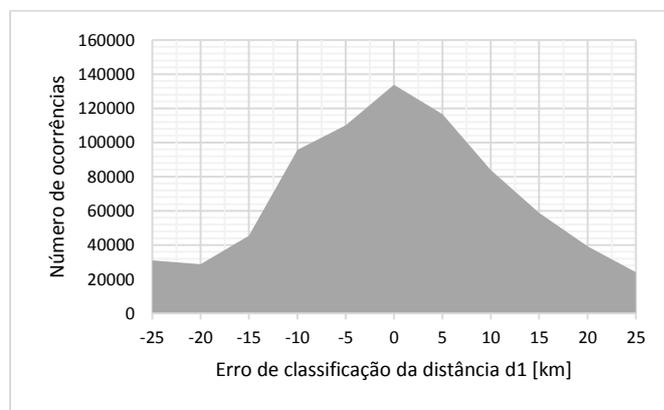


Figura 8 - Erro de estimação da distância da falta em função da impedância  $Z_1$  para o conjunto de dados de validação.

Ainda em relação aos resultados apresentados nas Fig. 7 e Fig. 8, deve ser frisado que já eram esperados erros maiores entre os valores esperados e os estimados para a distância  $d_1$  em relação à distância  $d_0$ , tendo em vista que a AD induzida para a classificação das impedâncias  $Z_0$  possuía acurácia 9,1% maior que a AD induzida para a classificação das impedâncias  $Z_1$ .

Posteriormente, para obter o terceiro parâmetro utilizado pela presente metodologia no processo de localização de faltas foi empregado o conjunto de dados de validação e o classificador do SI-3 para estimação da sensibilização dos medidores inteligentes a passagem da corrente de falta. Os resultados obtidos e apresentados na Fig. 9 mostram o elevado grau de acerto entre o estado real e o estimado cuja taxa de erro relativo absoluto foi de 6,94%.

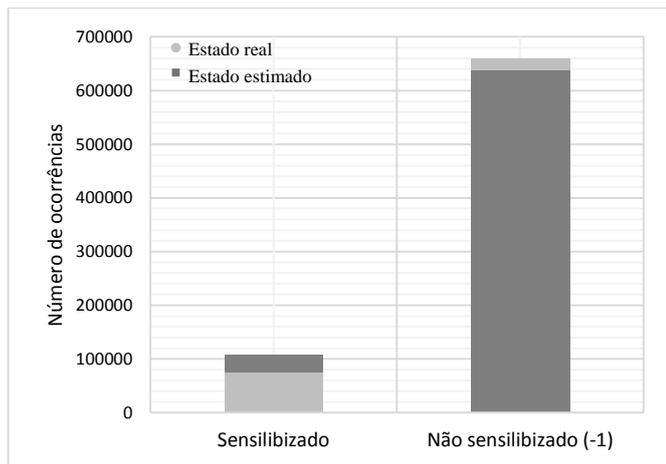


Figura 9 - Número de ocorrência de sensibilização ou não por correntes de falta dos medidores inteligentes para o conjunto de dados de validação.

Por fim, cruzando as informações das estimativas dos 3 sistemas inteligentes e fazendo a intersecção entre as regiões de concentração das distâncias obtidas, chega-se ao erro da distância de localização da falta conforme apresentado na Fig. 10. A partir desta figura, destaca-se que em 28% das simulações testadas a localização ocorre sem erro, 62,01% possuem erro entre -5 e +5 km, 83,47% têm erro entre -10 e +10 km, e 90,67% possuem erro entre -15 e +15 km.

Com os resultados obtidos depreende-se que o cruzamento das informações auxilia no processo de localização da falta, uma vez que o erro de localização foi reduzido quando comparado com os erros alcançados individualmente para  $d_0$  e  $d_1$  devido a uma maior concentração de distâncias com valores menores de erro. Para as faixas de distâncias com maior relevância para o processo de localização da falta houve um aumento de 6,16% e de 10,56% de distâncias estimadas corretamente sem erro em relação a  $d_0$  e a  $d_1$ , e ainda 7,96% e 15,03% de aumento de distâncias com erros entre -5 e +5 km, respectivamente.

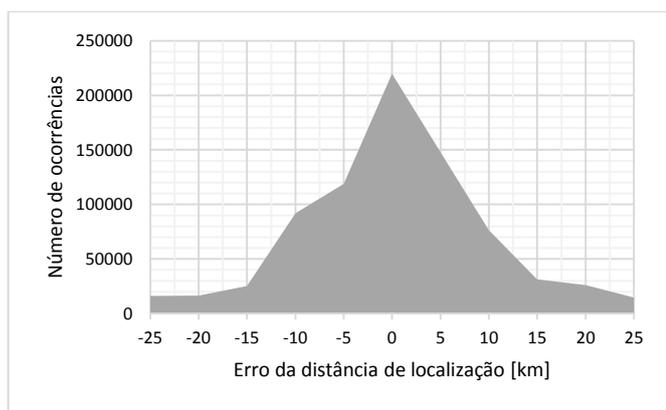


Figura 10 - Erro da distância de localização da falta para o conjunto de dados de validação.

## 6. CONCLUSÃO

A partir dos resultados apresentados com a validação da metodologia proposta foi salientada a relevância das etapas de cruzamento e análise das informações disponibilizadas por  $d_0$ ,

$d_1$  e o estado de sensibilização dos medidores. Este cruzamento visou trabalhar com a redundância de informações disponibilizadas pelos medidores inteligentes a fim de simplificar e agilizar o processo de localização da falta.

Os resultados também evidenciam que a maior parte dos erros da distância de localização se concentram na faixa entre -5 e +5 km, e o erro absoluto médio da metodologia de localização foi de 7,26 km. Com este erro é possível obter uma região de ocorrência da falta que apesar de não ser estritamente reduzida tem comprimento representando 14,52% do comprimento total do alimentador sob estudo, o qual possui aproximadamente 50 km de linhas de distribuição de energia. Deste modo, pode-se concluir que a estimativa obtida auxiliaria no processo de localização da falta e deslocamento das equipes de manutenção, porém está acima do usual para bons localizadores de falta, os quais trabalham com áreas de menos de 1 km de comprimento. Ressalta-se que o erro absoluto médio de 7,26 km decorre de duas situações principais, a dificuldade de generalização dos classificadores frente ao conjunto de dados apresentados para treinamento e teste e o aprendizado insuficiente no processo de indução das ADs.

A fim de melhorar o processo de estimativa das distâncias pelos classificadores, e conseqüentemente aumentar o acerto do método de localização de faltas, propõem-se para o prosseguimento da pesquisa a análise de clusterizações dos dados de entrada dos sistemas inteligentes implementados.

## AGRADECIMENTOS

Os autores gostariam de agradecer à Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) e ao Laboratório Multiusuário de Geração Distribuída (LMGD) da UACSA/UFRPE pela disponibilização da infraestrutura e demais serviços realizados neste trabalho, bem como amparo e auxílio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

## REFERÊNCIAS

- Anderson, P. M. Analysis of Faulted Power Systems: Paul M. Anderson. IEEE Press Power Systems Engineering Series, New York, 1995, p. 513.
- Bahmanyar, A.; Jamali, S.; Estebsari, A.; Bompard, E. "A comparison framework for distribution system outage and fault location methods". Electric Power Systems Research, vol. 145, pp. 19-34, April, 2017.
- Batista, O. E.; Flauzino, R. A.; Araújo, M. A.; Moraes, L. A.; Silva, I. N. "Methodology for information extraction from oscillograms and its application for high-impedance faults analysis". International Journal of Electrical Power & Energy Systems. vol. 76, pp. 23-34, March, 2016.
- Bayindir, R.; Colak, I.; Fulli, G.; Demirtas, K. "Smart grid technologies and applications". Renewable and Sustainable Energy Reviews, vol. 66, pp. 499-516, December, 2016.
- Beheshtaein, S.; Cuzner, R.; Savaghebi, M. and Guerrero, J. M. "Review on microgrids protection," in IET Generation, Transmission & Distribution, vol. 13, no. 6, pp. 743-759, 2019.

- Bell, J. Machine Learning: Hands-on for developers and technical professionals. 1st Edition, Wiley, 2014, 408 p.
- Borlase, S. Smart Grids: Infrastructure, Technology, and Solutions. CRC Press, Taylor & Francis Group, New York, USA, 2013, 591 p.
- Delen, D.; Walker, G.; Kadam, A. "Predicting breast cancer survivability: a comparison of three data mining methods". Artificial Intelligence in Medicine, vol. 34, nº 2, pp. 113–127, 2005.
- Dugan, R. C.; Mcgranaghan, M. F.; Santoso, S.; Beaty, H. W. Electrical power systems quality. 3ª Ed. New York: McGraw-Hill, 2012, 580 p.
- Fang, X.; Misra, S.; Xue, G.; Yang, D. "Smart Grid – The New and Improved Power Grid: A Survey". IEEE Communications Surveys & Tutorials, vol. 14, nº 4, Fourth Quarter, 2012.
- Freitas, A. A. Data Mining and Knowledge Discovery with Evolutionary Algorithms, Springer, 2002, 265 p.
- Ghosal, A. and Conti, M. "Key Management Systems for Smart Grid Advanced Metering Infrastructure: A Survey," in IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2831-2848, 2019.
- Glover, J. D.; Sarma, M. S.; Overbye, J. D. Power System Analysis and Design, 5ª Edição, Editora Cengage Learning, 2011, 848 p.
- Gollapudi, S. Practical Machine Learning. Packt Publishing, 2016, 468 p.
- Grainger, J. Jr.; Stevenson, W. D.; Chang, G. W. Power Systems Analysis. 2ª Ed. McGraw-Hill Education, 2015.
- Haykin, S. Neural Networks and Learning Machines. 3ª Edição, Editora Prentice-Hall; New Jersey; 2008, 936 p.
- Kohavi, R. "A study of cross-validation and bootstrap for accuracy estimation and model selection". In: International Joint Conference on Artificial Intelligence (IJCAI), Montreal, Quebec, Canada, pp. 1137-1143, 1995.
- Kumar, P.; Lin, Y.; Bai, G.; Paverd, A.; Dong, J. S. and Martin, A. "Smart Grid Metering Networks: A Survey on Security, Privacy and Open Research Issues," in IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2886-2927, 2019.
- Landis, J. R.; Koch, G. G. The measurement of observer agreement for categorical data. Biometrics. vol. 33, nº 1, pp. 159–174, March, 1977.
- McHugh, M. L. "Interrater reliability: the kappa statistic", Biochemia Medica, vol. 22, nº 3, pp. 276-282, 2012.
- Short, T. A. Electric Power Distribution Handbook. CRC Press LLC. 2004. 762 p.
- Theodoridis, S.; Koutroumbas, K. Pattern Recognition. 4ª Edition, Academic Press, 2008, 984 p.
- Weka. The Waikato University. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka>. Acessado em 18/06/2020.
- Witten, I. H.; Frank, E. Data mining: practical machine learning tools and techniques. 2nd Ed. Elsevier, 2005, 558 p.