

Association for Information Systems AIS Electronic Library (AISeL)

Proceedings of the XI Brazilian Symposium on
Information Systems (SBSI 2015)

Brazilian Symposium on Information Systems
(SBIS)

5-2015

Data Mining Techniques Assessment for Shutdowns Prediction of Electric Power Systems

Anderson Trindade Maia
UFPA, andersontrindade@ufpa.br

Jefferson Magalhães de Morais
UFPA, jmorais@ufpa.br

Yomara Pinheiro Pires
UFPA, yomara@ufpa.br

Aldebaro Barreto da Rocha Klautau Júnior
UFPA, aldebaro@ufpa.br

Daniel Augusto Martins
UFPA, odanielmartins@gmail.com

Follow this and additional works at: <http://aisel.aisnet.org/sbis2015>

Recommended Citation

Maia, Anderson Trindade; de Morais, Jefferson Magalhães; Pires, Yomara Pinheiro; Klautau, Aldebaro Barreto da Rocha Júnior; and Martins, Daniel Augusto, "Data Mining Techniques Assessment for Shutdowns Prediction of Electric Power Systems" (2015). *Proceedings of the XI Brazilian Symposium on Information Systems (SBSI 2015)*. 79.
<http://aisel.aisnet.org/sbis2015/79>

This material is brought to you by the Brazilian Symposium on Information Systems (SBIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in Proceedings of the XI Brazilian Symposium on Information Systems (SBSI 2015) by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Avaliação de Técnicas de Mineração de Dados para Predição de Desligamentos em Sistemas Elétricos de Potência

Alternative Title: Data Mining Techniques Assessment for Shutdowns Prediction of Electric Power Systems

Anderson Trindade Maia Faculdade de Computação - ICEN - UFFA Rua Augusto Correa nº 01 – Guamá Belém – Pa - Brasil andersontrindade@ufpa.br	Jefferson Magalhães de Moraes Faculdade de Computação - ICEN - UFFA Rua Augusto Correa nº 01 – Guamá Belém – Pa – Brasil jmorais@ufpa.br	Yomara Pinheiro Pires Faculdade de Engenharia Elétrica - ITEC - UFFA Rua Augusto Correa nº 01 – Guamá Belém – Pa – Brasil yomara@ufpa.br
---	---	---

Aldebaro Barreto da Rocha Klautau Júnior Faculdade de Engenharia Elétrica - ITEC - UFFA Rua Augusto Correa nº 01 – Guamá Belém – Pa – Brasil aldebaro@ufpa.br	Daniel Martins Faculdade de Engenharia Elétrica - ITEC - UFFA Rua Augusto Correa nº 01 – Guamá Belém – Pa – Brasil odanielmartins@gmail.com
--	--

RESUMO

O volume de informações históricas relacionado ao comportamento dos sistemas elétricos de potência, fez aumentar significativamente a dimensão das bases de dados das empresas de energia, sem entretanto contribuir com a melhoria nos aspectos de operação, manutenção e qualidade do serviço, salvo para eventuais consultas do comportamento histórico de variáveis. Diversos trabalhos na literatura vêm sendo propostos para investigar diversos aspectos relacionados este problema. Alguns desses trabalhos abordam o problema da predição de ocorrência de desligamentos em sistemas de potência utilizando técnicas de mineração. No entanto, devido ao grande volume de dados, a busca por metodologias de preparação e seleção de dados torna-se importante. Neste trabalho, é apresentada uma proposta metodológica utilizando técnicas de mineração de dados aplicadas às séries temporais de medição fasorial sincronizada para auxílio da prevenção de desligamentos devido a desvios de tensão em sistemas elétricos de potência. São testadas algumas técnicas de mineração em dados obtidos em uma linha de transmissão de 230 kV que interliga as subestações de Tucuruí, Altamira, Rurópolis, constituindo o sistema elétrico Tramo Oeste do Pará.

Palavras-Chave: PMU's, Sistemas Elétricos de Potência, desligamentos e mineração de dados.

A permissão para fazer cópias digitais ou impressas de todo ou parte deste trabalho para uso pessoal ou de sala de aula é concedida sem custos desde que as cópias não sejam feitas ou distribuídas para o lucro ou vantagem e que as cópias possam suportar este aviso e a citação completa na primeira página. Para copiar o contrário ou republicar, ou ainda, postar em servidores bem como redistribuir as listas, será preciso autorização prévia específica e / ou uma taxa.

SBSI 2015, 26-29 maio de 2015, Goiânia, Goiás, Brasil.
Direitos autorais SBC 2015.

ABSTRACT

The volume of historical information related to the behavior of electric power systems, has significantly increased the size of the energy companies databases, but without contributing to the improvement in the aspects of operation, maintenance and quality of service, except for any queries historical behavior of variables. Several studies in the literature have been proposed to investigate various aspects this problem. Some of these works address the problem of disconnections of occurrence prediction in power systems using mining techniques. However, due to the large volume of data, the search for the preparation of methodologies and data selection becomes important. In this work, a methodology using data mining techniques applied the synchronized phasor measurement time series to aid the prevention of shutdowns due to voltage deviations in electrical power systems is presented. Are tested some mining techniques on data obtained in a 230 kV transmission line linking the substations of Tucuruí, Altamira, Rurópolis, constituting the electrical system Tramo western Pará.

Categories and Subject Descriptors

H.2.8 [Database Applications]
I.2.1 [Applications and Expert Systems]

General Terms

Algorithms

Keywords: PMUs, Electric Power Systems, shutdowns and data mining.

1. INTRODUÇÃO

O setor elétrico brasileiro ultimamente vem sofrendo um impacto mais acentuado para conseguir manter as necessidades de geração, transmissão e distribuição de energia elétrica (Figura 1) que possam atender de forma aceitável e efetiva o aumento da demanda de seus consumidores em energia elétrica.

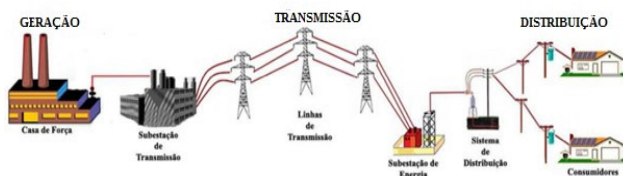


Figura 1: Exemplo de Geração, Transmissão, Interligação e Distribuição em um Sistema Elétrico de Potência[1].

Como não bastasse ainda, as seguidas paralisações nas obras das hidrelétricas, necessárias ao atendimento da natural matriz energética do tipo hidráulica, vêm obrigando o governo a investir em energia térmica, de construção mais simples e rápida, porém, de agressão ambiental talvez superior à das hidráulicas devido à alta emissão de carbono na atmosfera aumentando ainda mais os problemas no setor elétrico.

O setor elétrico também vem apresentando um aumento da presença de pequenos produtores independentes, principalmente as Pequenas Centrais Hidrelétricas - PCHs e, portanto, sem controle centralizado, interligadas à rede elétrica de distribuição além do aumento na utilização de energia eólica, sabidamente instável e causadora de variações abruptas na geração vem trazendo à operação do Sistema Elétrico Interligado - SIN brasileiro um desafio sem precedentes.

Um outro fator que deve ser analisado dentro desse contexto, diz respeito à inevitável política de interligação dos diversos sistemas elétricos regionais em um único sistema elétrico de forma a criar uma colaboração das bacias hidrográficas. Este tipo de cenário vem apresentando uma maior propagação de distúrbios elétricos, antes localizados, agora atingindo diversas áreas ou regiões simultaneamente, aumentando muito o risco de grandes blecautes como os que se tem presenciado recentemente em algumas regiões do Brasil. Do ponto de vista de estabilidade de sistemas elétricos, as variações de tensão e frequência no SIN, vem se tornando cada vez mais frequentes, provocando a atuação de equipamentos de proteção e os consequentes desligamentos.

Diante deste cenário e procurando soluções para essas questões, as empresas do setor elétrico estão buscando utilizar tecnologias avançadas para aquisição e armazenamento de informações, as quais incluem os chamados distúrbios ou eventos de interesse[2]. Um típico exemplo dessa tecnologia são os equipamentos de oscilografia e, mais recentemente as PMUs¹ (Figura 2), que detectam e transmitem os valores de tensão, corrente e seus respectivos ângulos de fase. Essas informações são recebidas,

1 PMUs - Phasor Meter Units, são equipamentos de medição fasorial sincronizada que Esses equipamentos são capazes de amostrar grandezas elétricas sincronamente em subestações geograficamente afastadas, a taxas de amostragem de até 60 amostras por segundo [3].

organizadas temporalmente, armazenadas e disponibilizadas para uso pelo concentrador de medição fasorial PDC (*Phasor Data Concentrator*).

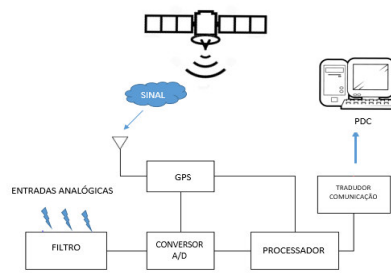


Figura 2 – PMU: Estrutura Básica

Os centros de controle de sistemas elétricos de potência (Figura 3), por sua vez, tratam e armazenam ordenadamente em banco de dados históricos, uma quantidade enorme de informação que, na maioria esmagadora das vezes, fica simplesmente armazenada e raramente observada no futuro. Entretanto, esta grande massa de informação, se sabe, guarda informações extremamente úteis e definitivamente necessárias para uma compreensão holística dos fenômenos. Se faz necessário, assim, minar estes dados na busca por características, relações, associações, agrupamentos e etc., que tragam informações mais úteis sobre a operação do sistema elétrico de potência.

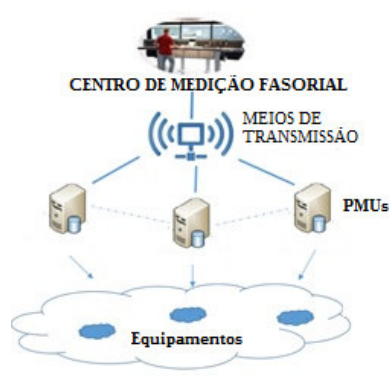


Figura 3 – Arquitetura Básica de um Sistema de Supervisão e Controle de um SEP.

Apesar da tecnologia avançada, minar tais dados para inferir, por exemplo, relações de causa e efeito é ainda uma atividade incipiente [4]. Uma das razões é a falta de ações para relacionar os eventos com dados adicionais que poderiam ajudar a inferir a causa. Estabelecer esta relação é, hoje, uma das grandes preocupações nas indústrias de geração, transmissão e distribuição de energia elétrica.

Neste cenário cada vez mais competitivo, a tecnologia da informação, quando bem utilizada, torna-se um importante diferencial entre as empresas que primam por excelência na qualidade de serviço prestado. Neste contexto, surgem as aplicações e técnicas de mineração de dados cujo objetivo é descobrir padrões de comportamento e de novos conhecimentos sobre as informações armazenadas. Portanto, a gestão aleatória baseada principalmente em uma simples intuição vai dando lugar a inteligência do negócio.

Alguns trabalhos relacionados já abordaram metodologias para diagnóstico de desligamento em SEPs utilizando técnicas tais como interpolação em algoritmo simplista [5], Redes Neurais Artificiais (RNA) [6, 9], Lógica Fuzzy [7], Redes Bayesianas [8], *Random Forest*, *Support Vector Machine* (SVM) e *K-NearestNeighbor* (KNN) [1, 4].

Devido à dificuldade no tratamento dessas informações, ainda é difícil muitas das vezes adotar uma metodologia que seja capaz de mostrar com um tempo mínimo de antecedência a probabilidade de ocorrência de um evento de desligamento bem como também, verificar se este acontecimento está associado a fatores locais aparentes.

Em outra situação, por exemplo, pode não haver uma causa aparente, onde é comum apontar para uma falta transitória, como motivo de interrupção. Um exemplo típico é o desligamento por curto-circuito.

Após a ocorrência de um curto-circuito, se este não é permanente, o operador do sistema elétrico restabelece o sistema. Neste momento, o operador pode erroneamente interpretar esse evento (causa de interrupção) como uma ação de proteção. No entanto, na realidade, o curto-circuito e a atuação de proteção podem ser apenas consequências de algum evento anterior que ocorreu na rede, tais como condutores esgarçados, isoladores danificados ou poluídos, e muitos outros fatores que podem interromper o fornecimento de energia. A utilização correta da metodologia a ser adotada para preparação dos dados a serem minerados, é tão importante quanto os resultados da mineração no conhecimento do fenômeno, pois muitas vezes permite identificar a recorrência de um conjunto de desligamentos que podem estar originando a degradação da rede.

Com base nesse contexto, o objetivo deste artigo sinaliza para a necessidade de se investigar alternativas que permitam criar uma metodologia cujo objetivo seja a preparação dos dados das informações históricas de sistemas elétricos utilizando um processo de identificação, transformação, seleção e validação de informações para posterior mineração de dados em base de dados de sequência de eventos (SOE – *SequentialOfEvents*) e Medição Fasorial Sincronizada de forma a auxiliar os especialistas na obtenção de diagnóstico e especificamente na prevenção de desligamentos em sistemas elétricos de Potência.

2. METODOLOGIA

2.1 Area de Estudo

A área de estudo e investigação da pesquisa compreende os dados obtidos da medição fasorial sincronizada correspondente as subestações de Tucuruí, Altamira e Rurópolis que estão interligadas por uma linha de transmissão de 230 kV. Este trecho do Sistema Interligado Nacional – SIN, é denominado Tramoeste do Pará conforme ilustra a Figura 4.

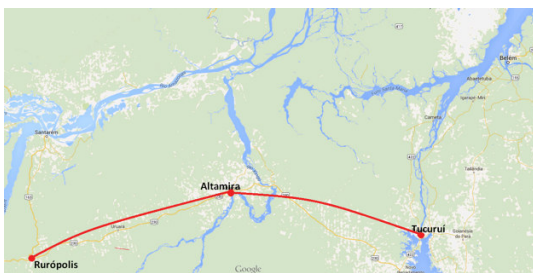


Figura 4 – Região do Tramoeste.

Considerando que o fenômeno em observação é do tipo elétrico em alta tensão (230kV), alguns procedimentos devem ser adotados para extrair as informações do sistema elétrico.

2.2 Metodologia Proposta

A metodologia adotada trata inicialmente das informações provenientes das PMUs que não foram projetadas para um fim específico, mas que apresentam em seus dados eventos de desligamento. Assim, durante a fase de pré-processamento, apesar do PDC oferecer formas de onda de tensão e corrente (fases A, B e C), em sistemas de elevada tensão, como é o caso da aplicação desta pesquisa, as três fases são equilibradas, isto é, seus valores RMS (*Root Mean Squared*) são praticamente equivalentes, e as defasagens entre os fasores se mantêm razoavelmente próxima a 120 graus. Desta forma, foi utilizada apenas uma das fases, como por exemplo a fase B.

No caso do problema da predição do evento de desligamento, as instâncias (exemplos) são compostas por um conjunto de informações reais de maior relevância presentes na base para verificação da ocorrência ou não do evento, cujo resultado corresponde à classe dessas instâncias. Os campos selecionados para aplicação da metodologia são apresentados de forma simplificada na Tabela 1:

Tabela 1 – Campos da base de dados utilizados para mineração.

ID_Variável	Data_hora	Valor Medido
-------------	-----------	--------------

Onde cada campo é descrito como segue:

ID_Variável: Número de identificação da variável.

Data_hora :Valores de tensão da fase B medidos pelos PMUs de Tucuruí na barra 2, de Rurópolis na barra 1 e pelo de Altamira na barra 1.

DIF_FASE: Valores da diferença de fase entre a fase B medida pelo PMUs de Tucuruí na barra 2 e a fase B medida pelo PMU de Rurópolis na barra 1. De forma análoga, obteve-se também os Valores da diferença de fase entre a fase B medida pelo PMUs de Tucuruí na barra 2 e a fase B medida pelo PMU de Altamira na barra 1.

Vale ressaltar, que os campos TENSÃO e DIF_FASE são um conjunto de valores reais fornecidos pela Eletrobras-Eletronorte. Esses valores são medições realizadas continuamente no decorrer do mês de agosto de 2013. Este período foi escolhido, pois apresentou uma incidência maior de desligamentos quando comparados a outros meses do mesmo ano. Diante desse cenário e com objetivo de gerar um arquivo histórico passível de ser analisado por um especialista, os campos TENSÃO e DIF_FASE, foram divididos em vários subcampos, sendo um subcampo para cada medição pertencente à um determinado intervalo de tempo.

Portanto, para aplicar as técnicas de mineração de dados na tarefa de predição de ocorrências de desligamento em SEPs, foi composta uma base a partir dos valores RMS da tensão da fase extraídas de uma PMU, considerando uma janela de tempo de 60 segundos antes da ocorrência do evento. Assim, os dados coletados resultaram em uma base com 392 instâncias (exemplos), sendo 77% registros relacionados a eventos de desligamento e 22% de registros de não desligamento. A Figura 5 apresenta as etapas desenvolvidas desde o processo de transformação até a

aplicação das técnicas de mineração de dados nas medições fasoriais coletadas.

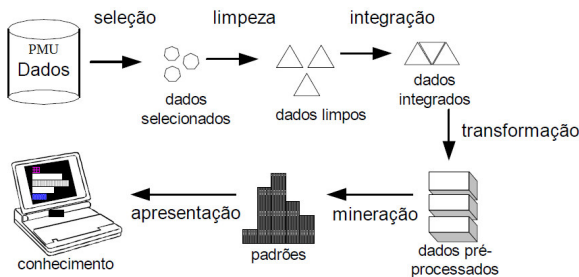


Figura 5 – Etapas do processo de Mineração (Ribeiro, 2004).

Três scripts (BinaryReader.java, CsvFormat.java, e AlarmFormat.java) foram desenvolvidos para etapa de seleção e pré-processamento dos dados. Já as etapas de transformação, os dados extraídos dos equipamentos são em formato binário. E na etapa de mineração dos dados foi utilizado o software WEKA(WaikatoEnvironmentKnowledgeAnalysis) que será mais detalhado nas próximas seções.

- BinaryReader.java: foi o primeiro script implementado, este arquivo realiza a leitura de arquivos binários provenientes da PMU. Estes arquivos contém as medições fasoriais das estações de Rurópolis, Altamira e Tucuruí.
- CsvFormat.java: Este script realiza a etapa de pré-processamento dos dados selecionados. O arquivo de entrada é um arquivo no formato .csv, exportado pelo script BinaryReader.java, e o arquivo de saída é um arquivo no formato .arff. O arquivo de saída do CsvFormat já está preparado para posterior transformação e mineração por meio da ferramenta WEKA.

O algoritmo possui um atributo atualizável via código que determina o limite T de tempo de até T segundos antes do desligamento que será possível de se coletar os dados. Outro atributo importante do algoritmo é o windowSize, que define a janela de tempo que irá ser utilizada pelo software de mineração de dados para tentar classificar determinado evento. Um exemplo segue na Figura 6.

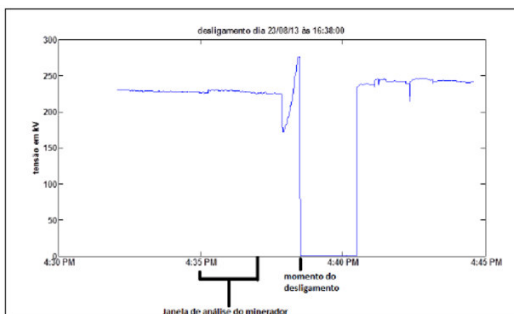


Figura 6 – Gráfico da tensão de uma estação durante o momento de um desligamento.

O arquivo de saída .arff (Attribute-Relation File Format) do CsvFormat.java está organizado da seguinte forma:

- Um cabeçalho – que contém informações sobre quais são os atributos deste conjunto de dados.
- Instâncias – que possui várias linhas, cada linha representa uma instância, com seus valores para cada atributo.

Esta é a organização aceita pelo programa de mineração WEKA. De cada janela de análise é sintetizada uma instância e cada medição fasorial de uma determinada janela compõe um campo de atributo de uma instância. A quantidade de atributos das instâncias depende do tamanho da janela de análise windowSize. Se uma janela de 60 segundos há 500 amostras, cada instância irá possuir 500 atributos.

No caso da mineração efetuada nesta pesquisa, cada instância possui 100 atributos. Estes atributos são nomeados r0,r1,r2,...,r99. Os primeiros 50 atributos (r0 à r49) referenciam 50 amostras de tensão igualmente espaçadas no tempo, enquanto que as últimas 50 amostras (r50 à r99) referenciam as amostras de fase medidas ao mesmo tempo que as amostras de tensão. Essas amostras são coletadas por uma PMU de uma determinada estação. A taxa de amostragem de uma PMU, na época da medição, é de 10 Hz, ou seja, 10 amostras por Segundo.

Como um conjunto muito grande de atributos pode prejudicar a mineração de dados, foi realizada uma subamostragem das amostras coletadas pela PMU, uma variável chamada samplePerInstance foi criada para armazenar o valor da subamostragem, seu valor foi configurado para 50 (50 amostras de tensão e 50 amostras de fase).

A partir das variáveis *samplesPerInstance*, *samplingFrequency*(10Hz) e *windowSize* (tamanho da janela em segundos) pode-se criar um fator de subamostragem definido pela fórmula:

$$X = \frac{WS * SF}{SP}$$

Onde,

WS = Tamanho da Janela

SF = Frequência de amostragem do PMU

SP = Quantidade de atributos de cada Instância.

A cada x amostras, uma amostra é copiada para o arquivo no formato aceito pelo weka e o restante é descartado.

As etapas (Figura 7) seguintes tratam efetivamente da avaliação sistemática dos classificadores adotados para prevenção de desligamentos.

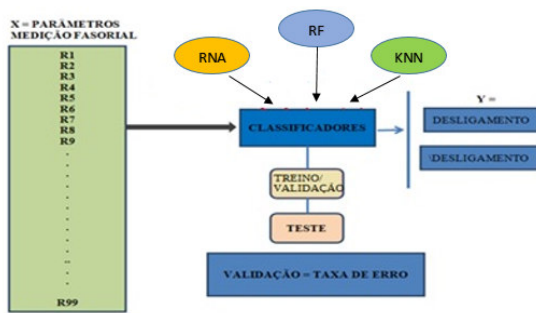


Figura 7 – Modelo de classificação adotado.

3. CLASSIFICADORES

Um *classificador* F é um mapeamento $F : R^k \rightarrow \{1, \dots, Y\}$, onde K é a dimensão do vetor de entrada $z \in R^k$ e o rótulo $y \in \{1, \dots, Y\}$ é a classe. Quando se treina um classificador usando aprendizado supervisionado, é dado um conjunto de treinamento $T = \{(z_1, y_1), \dots, (z_v, y_v)\}$ contendo V exemplos de (z, y) [10].

Um conjunto de teste $\{(z_1, y_1), \dots, (z_r, y_r)\}$ contendo R exemplos e disjuncto do conjunto de treino pode ser usado para calcular a taxa de erro de classificação

$$E_r = \frac{1}{R} \sum_{i=1}^R \tau(F(x_i) \neq y_i)$$

onde τ é a função indicador, que é um (1) caso o argumento seja verdadeiro e zero (0) caso contrário. O erro E_r é uma estimativa da capacidade de generalização do classificador [11].

Para classificação este trabalho utilizou os algoritmos do *software WEKA (Waikato Environment for Knowledge Analysis)*, software livre de domínio público assistido pela licença GPL (*General Public License*), e desenvolvido na Universidade de Waikato na Nova Zelândia, sendo disponibilizado em <http://www.cs.waikato.ac.nz/ml/weka/>. Assim, as próximas seções discutirá brevemente este software e as seções seguintes listarão os principais classificadores utilizados.

3.1 Weka

O WEKA é um *software* de referência reconhecido pela comunidade científica em aprendizado de máquina e mineração de dados [11-13]. O sistema é formado por um conjunto de algoritmos de aprendizado de máquina, implementado em linguagem Java, que o torna acessível nos principais sistemas operacionais.

O WEKA inclui algoritmos de regressão, classificação, agrupamento, regras de associação e seleção de parâmetros (atributos). Atualmente está na versão estável 3.6 que usa o formato de arquivo com extensão *arff* [10]. O mesmo possui diversos algoritmos de Inteligência Computacional de classificação e este trabalho usa apenas alguns dos principais, tais como Rede Neural Artificial (RNA) multicamadas treinadas com algoritmo *backpropagation*, Random Forest (RF) e K-nearestneighbor (KNN).

3.2 Redes Neurais Artificiais

As RNA são sistemas paralelamente divididos em camadas interligadas por pesos sinápticos e compostas por neurônios artificiais que calculam uma certa função matemática (usualmente não-linear). O algoritmo utilizado neste trabalho para treinamento da RNA foi o *backpropagation* também chamado de regra delta generalizada. Nesta pesquisa foi utilizada a classe *Multilayer Perceptron* que é o padrão do WEKA e os principais parâmetros dessa implementação são: (-L) - corresponde à taxa de aprendizado utilizada pelo algoritmo *backpropagation* e este valor deve ser entre 0 e 1 (Padrão é 0,3); (-M) - taxa de momento para o algoritmo *backpropagation* e este valor deve ser entre 0 e 1 (Padrão é 0,2); (-N) - este parâmetro corresponde ao número de épocas para treinamento da rede e o padrão é 500; (-H) - corresponde à quantidade de camadas ocultas que podem ser criadas na rede [10,14].

3.3 Random Forest

O RF é a referência algorítmica em árvores de decisão. A idéia desse algoritmo é dividir o espaço de instâncias em subespaços, no qual os mesmos são ajustados em diferentes modelos, sendo formalmente estruturado em um grafo acíclico em que cada nó, ou é um nó de divisão com dois ou mais sucessores dotado de um teste condicional, ou um nó folha rotulado com uma nova classe. No WEKA o RF tem inúmeros parâmetros, sendo os principais são: (-depth) - responsável pela profundidade da árvore; (K) - número de recursos em cada iteração; I - número de árvores do modelo [10,15].

3.4 K-nearest neighbor

O classificador KNN utiliza os próprios dados de treinamento como modelo de classificação, isto é, para cada novo padrão que se quer classificar, usa os dados do treinamento para verificar quais são os exemplos nessa base de dados que são mais próximos do padrão em análise, ou seja, baseado nas distâncias. A cada novo padrão a ser classificado faz-se uma varredura nos dados de treinamento, o que provoca um grande esforço computacional. O KNN no WEKA é implementado na classe *IBK* e seus principais parâmetros são: (-N) que é igual ao número de centros (ou K) e (-S) que gera aleatoriamente os centros [10,14].

4. RESULTADOS

As métricas utilizadas para avaliar o desempenho dos classificadores na prevenção de desligamentos em sistemas elétricos de potência, levou em consideração a variação dos valores nos parâmetros dos classificadores, assim como a acurácia, taxa de erro, taxa de especificidade, taxa de sensibilidade e medida F . Sendo que nas seções posteriores é apresentado os resultados do estudo com mais detalhes.

4.1 Resultado da Seleção do Modelo

Os valores que definem os modelos dos classificadores, tais como o número de neurônios na camada escondida de uma rede neural, por exemplo, foram escolhidos por meio de um procedimento de seleção automática de modelo. A Tabela 2 apresenta o conjunto de valores testados (*grid*) adotado nesse procedimento. É importante destacar que os valores escolhidos para compor o *grid* foram baseados em experimentos anteriores. Para validação dos

resultados foi adotado validação-cruzada (cross-validation) considerando 10-folds[11].

Tabela 2: Grid de seleção do modelo.

Classificador	Parâmetros	ValoresGrid	Melhor valor
RNA	H	3, 5, 7, 9 e 11	5
RF	I	5, 10, 15, 20 e 25	10
KNN	K	1, 3, 5, 7 e 9	7

Como pode-se observar na Tabela 2 em Valores Grid, foram feitos cinco experimentos para cada classificador com valores diferentes do parâmetro apresentado. A Rede Neural obteve uma melhor taxa de classificação com cinco neurônios em uma camada escondida (H), no qual foram feitos experimentos com 3, 5, 7, 9 e 11 neurônios respectivamente. Random Forest utilizou uma faixa de 5 a 25 árvores aleatórias no parâmetro (I), sendo que o valor ótimo foi com 10 árvores. E por fim o KNN no parâmetro K obteve uma melhor performance de classificação com valor de sete vizinhos mais próximos, sofrendo uma variação entre 1 e 9 dos cinco experimentos.

4.2 Resultado da Acurácia, Taxa de Erro, Taxa de Especificidade, Taxa de Sensibilidade e Medida F

Outros elementos relevante para mensurar o desempenho do classificador é a acurácia (taxa de acerto) e taxa de erro. Alguns classificadores são capazes de prover scores de confiança $f_i(x)$ para cada classe $i = 1, \dots, Y$, tais como a probabilidade de distribuição sobre y . Por conveniência, pode-se assumir que todo classificador retorna um vetor $(f_1(x), \dots, f_y(x))$ com Y scores. Se o classificador naturalmente não retornar scores de confiança, o vetor de scores é criado com um score unitário $f_j(x)=1$ para a classe j sugerida pelo classificador, enquanto que os scores das outras classes são zero: $f_i(x) = 0, i \neq j$. Assim, a decisão final é sempre baseada no valor máximo dos scores (chamada regra *max-wins*).

Um conjunto de teste $\{(x_1, y_1, \dots, x_R, y_R)\}$ contendo R exemplos e disjunto do conjunto de treino pode ser usado para calcular a taxa de erro de classificação (Equação 4).

$$E_f = \frac{1}{R} \sum_{i=1}^R \tau(F(x_i) \neq y_i)$$

onde τ é a função indicador, que é um (1) caso o argumento seja verdadeiro e zero (0) caso contrário. O erro E_f é uma estimativa da capacidade de generalização do classificador [11,12].

A taxa de erro varia entre 0 e 100, e valores próximos a 0 são considerados os melhores índices nas predições dos

classificadores. Em [16] afirma-se que a acurácia (taxa de acerto), conforme a Equação 5, serve como complemento da taxa de erro.

$$ac(f) = 1 - E_f$$

[17,18] colocam que a taxa de acerto também varia entre 0 a 100, entretanto ao contrário da taxa de erro, valores próximos a 100 apresentam melhores predições nas classificações. Foram adotadas também medidas de desempenho na predição dos eventos de desligamentos, como por exemplo: a taxa de precisão (P), a taxa de sensibilidade (T_s) e a taxa de especificidade (T_e) e a medida F . O autor também chama atenção que tanto a acurácia como a taxa de acerto, assim como a P , T_s , e T_e e a medida F podem ser extraídas da matriz de confusão, pois as linhas da mesma contem valores² que representam as classes verdadeiras, e as colunas, as classes preditas pelo classificador (Equações 6, 7, 8 e 9).

$$P(f) = \frac{VP}{VP+FP}$$

$$T_s = \frac{VP}{VP+FN}$$

$$T_e = \frac{VN}{VN+FP}$$

$$medida - F(f) = 2 \left[\frac{precisao(f) * sens(f)}{precisao(f) + sens(f)} \right]$$

A Tabela 2 apresenta os resultados dos classificadores utilizados no estudo com as métricas adotadas na predição dos eventos de desligamentos no sistema TRAMOESTE.

Tabela 2: Resultado da Taxa de acerto (acurácia), Taxa de erro, Taxa de sensibilidade e Taxa de especificidade, Medida F dos classificadores utilizados no estudo

Classificador	Acurácia(%)	Taxa de Erro(%)	Ts(%)	Te(%)	MedidaF
RNA	87,50	12,50	88,52	91,12	90,35
RF	81,25	18,50	84,30	86,75	85,82

(4)

2 FN: Falsos negativos: valor de exemplos classificados incorretamente como negativos; FP: Falsos positivos: exemplos classificados corretamente como negativos; VN: Verdadeiros negativos: exemplos classificados como verdadeiros incorretamente e VP: Verdadeiros Positivos: correspondem a todos a todos os exemplos classificados corretamente [17].

KNN	84,37	15,63	86,34	88,70	87,11
-----	-------	-------	-------	-------	-------

De acordo com a Tabela 2 observa-se que a RNA obteve o melhor desempenho em relação aos demais classificadores com acurácia e taxa de erro com valores de 87,2% e 12,2% respectivamente. Enquanto que o KNN foi o classificador com menor desempenho com 84,37% e 81,25% respectivamente na mesma métrica.

5. CONSIDERAÇÕES FINAIS

O estudo utilizou uma metodologia baseada em algoritmos de Inteligência Computacional, como Redes Neurais Artificiais, KNN e Random Forest, auxiliado por mecanismos de seleção automática de modelo na predição de eventos de distúrbios em linhas de transmissão de energia elétrica na rede Tramoeste como prevenção de desligamentos em sistemas elétricos de potência.

As métricas adotadas de acurácia (taxas de acerto), taxa de erro, taxa de especificidade, taxa de sensibilidade e medida F foram utilizadas para validar os resultados obtidos. Entre as técnicas computacionais utilizadas, a Rede Neural Artificial foi a que teve um desempenho mais satisfatório na predição de eventos de distúrbios em linhas de transmissão de energia elétrica na rede Tramoeste.

Sendo que a grande variação sazonal climática na área de estudo podem provocar distúrbios elétricos, assim como outras causas dessas alterações no sistema elétrico interfere em resultados mais precisos. Essa precisão também é afetada com a limitação de recursos estruturais e humanos.

Portanto, a aplicação de técnicas de inteligência computacional nas predições de eventos de distúrbios em linhas de transmissão de energia elétrica na rede Tramoeste como prevenção de desligamentos em Sistemas Elétricos de Potência pode se transformar em uma contribuição objetiva na administração dos recursos elétricos. Possibilitando pesquisas futuras que possam utilizar outras técnicas, além de Redes Neurais Artificiais, KNN e Random Forest para auxílio de melhor gerenciamento e prevenção em sinistros do sistema elétrico.

AGRADECIMENTOS

Os autores agradecem a Eletrobras-Eletronorte por ter fornecido as bases de dados e pelo suporte financeiro.

REFERÊNCIAS

[1] Morais, J. M. (2011). Avaliação de Desempenho de Classificadores de Falhas em Sistemas Elétricos de Potência. Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica. Universidade Federal doPará.

[2] Bollen, M., (2009). Bridging the gap between signal and power. IEEE Signal Processing Magazine, 26:12, July.

[3] Ree, J. D. L., Centeno, V., Thorp, J. S., and Phadke, A. G. (2010). Synchronized phasor measurement applications in power systems. IEEE Trans. Smart Grid, 1(1):20–27.

[4] Morais, J. M., Pires, Y.P., Cardoso, C., Klautau, A. B. R. (2010). A Framework for Evaluating Automatic Classification of Underlying Causes of Disturbances and Its Application to Short-Circuit Faults. IEEE Transactions on Power Delivery, v. 25, p. 2083-2094.

[5] Krishnanand, K. R., Dash, P. K., Naeem, M. H. (2015). Detection, classification, and location of faults in power transmission lines. Journal of Eletrica Power & Energy System, v. 67, p. 76-86.

[6] He, Z., Lin, S., Deng, Y., Li, X., Qian, Q. (2014). A rough membership neural network approach for fault classification in transmission lines. Journal of Eletrica Power & Energy System, v. 61, p. 429-439.

[7] Liu, Y., Schulz, N.N. (2002). Intelligent system applications in distribution outage management. Power Engineering Society Winter Meeting. IEEE, Vol. 2, pp. 833 – 837.

[8] Tronchoni, A. B. (2008). Identificação de Causas de Desligamentos Não Programados em Redes de Distribuição. Dissertação de Mestrado, PPGE, PUCRS.

[9] Pretto, C. O., Rancich, G.V., Lemos, F.A.B., Rosa, M.A. (2003). Forced Outages Information Treatment System and Cause Identification Based on Mobile Computing and Neural Networks. IEEE Bolonha Power Tech, Bolonha, pp. 1-6.

[10] Arouche Freire, J. C., Santos, Y. S. R. N., Morais, J. M., Oliveira, T. F., Hauser-Davis, R. A., Saraiva, A. C. F. (2014b). Avaliação de Desempenho de Classificadores de Ciclos Hidrológicos em Reservatórios de Água na Região Amazônica. In: XXXIV Congresso da Sociedade Brasileira de Computação, WCAMA, Brasília-DF.

[11] Witten, I. H., Frank, E. (2005). Data mining: practical machine learning tools and techniques with java implementations. Edition Morgan Kaufmann.

[12] Hall, M.; Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H. (2009). The weka data mining software: an update. SIGKDD Explor. Newsl. 11, 10-18.

[13] Guandaline, V. H., Merschmann, L. H. de C. 1. (2013). Avaliação de Técnicas de mineração de dados para predição de ocorrência do mexilhão. TrilhasTécnicas - SBSI, 2014.

[14] Haykin, S. (1998). Neural Networks: A comprehensive Foundation. Prentice Hall, 2 edition, July 1998. p. 900.

[15] Zhou, Q., Zhou, H., Zhou, Q., Yang, F., and Luo, L. (2014). Structure damage detection based on random forest recursive feature elimination. Mechanical Systems and Signal Processing.

- [16] Faceli, k., Lorena, A. C., Gama, J., Carvalho, A. C. P. L. F. *Inteligência artificial: uma abordagem de aprendizado de máquina*. Editora LTC, 2011.
- [17] Arouche Freire, J. C., Lobato, T. C., Moraes, J. M., Oliveira, T. F., Hauser-Davis, R. A., Saraiva, A. C. F. (2014a). Support Vector Machine and Artificial Neural Networks for Hydrological Cycles Classifications of a Water Reservoir in the Amazon. *Journal of Communication and Computer*, v. 11, p. 111-117.
- [18] Riflin, R., Klautau, A. (2004). In defense of one-vs-all classification. *Machine Learning Research* 5, 101-141.