



UNIVERSIDADE FEDERAL DO PARÁ
CAMPUS UNIVERSITÁRIO DO GUAMÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**METODOLOGIA PARA A CLASSIFICAÇÃO DE PERTURBAÇÕES A
PARTIR DE MEDIÇÕES FASORIAIS SINCRONIZADAS UTILIZANDO
TÉCNICAS DE MINERAÇÃO DE DADOS**

ANDERSON TRINDADE MAIA

UFPA / ICEN / PPGCC
Belém – Pará
2015

Dados Internacionais de Catalogação na Publicação (CIP)
Sistemas de Biblioteca da UFPA

Maia, Anderson Trindade, 1979 -

Metodologia para a Classificação de Perturbações a partir de Medições Fasoriais Sincronizadas Utilizando Técnicas de Mineração de Dados / Anderson Trindade Maia. - 2015.

Orientador: Jefferson Magalhães de Moraes;

Co-Orientadora: Yomara Pinheiro Pires.

Dissertação (Mestrado) – Universidade Federal do Pará, Instituto de Ciências Exatas e Naturais, Programa de Pós-Graduação em Ciência da Computação, Belém, 2015.

1. Mineração de dados (computação). 2. Mineração de dados (computação) - Inovações tecnológicas. 3. Sistemas de energia elétrica. 4. Mineração de dados-Metodologia. 5. Sistemas de medição fasorial sincronizada . I. Título.

METODOLOGIA PARA CLASSIFICAÇÃO DE PERTURBAÇÕES A PARTIR DE MEDIÇÕES FASORIAIS SINCRONIZADAS UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação (área de concentração: Sistemas de Computação), do Instituto de Ciências Exatas e Naturais, da Universidade Federal do Pará, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Jefferson Magalhães de Moraes (PPGCC - UFPA)

Co-Orientadora: Prof.^a Dr.^a Yomara Pinheiro Pires (FACOMP - UFPA)

ANDERSON TRINDADE MAIA

METODOLOGIA PARA CLASSIFICAÇÃO DE PERTURBAÇÕES A PARTIR DE MEDIÇÕES FASORIAIS SINCRONIZADAS UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação (área de concentração: Sistemas de Computação), do Instituto de Ciências Exatas e Naturais, da Universidade Federal do Pará, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação, sob orientação do Prof. Dr. Jefferson Magalhães de Moraes e co-orientação da Prof.^a Dr.^a Yomara Pinheiro Pires.

Data de Avaliação: 21 de dezembro de 2015

BANCA EXAMINADORA

Prof. Dr. Jefferson Magalhães de Moraes – Orientador
Programa de Pós-Graduação em Ciência da Computação
Universidade Federal do Pará

Prof.^aDr.^aYomara Pinheiro Pires – Co-orientadora
Faculdade de Computação (Campus Castanhal)
Universidade Federal do Pará

Prof. Dr. Claudomiro de Souza de Sales Júnior
Programa de Pós-Graduação em Ciência da Computação
Universidade Federal do Pará

Prof. Dr. Ubiratan Holanda Bezerra
Programa de Pós-Graduação em Engenharia Elétrica
Universidade Federal do Pará

UFPA / ICEN / PPGCC
Belém – Pará
2015

A minha esposa e ao meu filho.

AGRADECIMENTOS

Primeiramente, agradeço a Deus por me conceder o dom da vida e do conhecimento para que este empreendimento científico fosse concretizado. De igual forma, manifesto minha profunda gratidão...

A minha esposa Rosilene e ao meu filho Gabriel, minhas inspirações superlativas, pela compreensão, pelo incentivo e por todo amor que me dedicam em todos os momentos da minha vida. O desenho dos meus sonhos e da realidade das minhas ações é efetivado com os traços e com as cores da cumplicidade dos nossos desejos e realizações que nós compartilhamos;

Aos meus pais, Ana e Olivar, esteios da minha formação, principalmente pela simplicidade e pela alegria com que realizam suas vidas. Além de agradecer por trazem ao mundo, deixo registrado o meu carinho, o meu respeito, a minha admiração pelo amor incondicional que sempre dedicaram a mim;

Ao meu irmão Cleison, pelo amor fraternal alicerçado no respeito às nossas diferenças, opiniões e trajetórias de vida. Você é uma das grandes razões da nossa união e da presença de Deus, conjugada na órbita de nossos pais;

A todos que fazem parte da minha grande família (tios, primos, entes queridos), sou grato pela convivência e pela importância que cada um tem nesse universo social, no qual compartilhamos experiências, orientações, carinho e respeito;

Ao meu caro orientador, Prof. Dr. Jefferson Moraes, e a minha co-orientadora, Prof.^a Dr.^a Yomara Pires, por suas imensas e valiosas contribuições, pelo comprometimento e pela sapiência no apoio às escolhas dos melhores caminhos para realização dessa dissertação;

Aos meus amigos: Daniel Martins, pelos seus contributos, pela compreensão e pelos recursos que foram tomados para concepção deste trabalho; Ao Jean, pela nossa amizade, pela ajuda na área em que realizei esta pesquisa; Ao meu grande amigo Manu, pelo incentivo e pela força nos momentos mais difíceis dessa caminhada;

À Eletrobras-Eletronorte, pelo fornecimento da base de dados e pelo suporte financeiro; À Faculdade de Computação (Campus Universitário de Castanhal -UFPA), que sempre incentivou e apoiou a minha trajetória acadêmica;

Ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal do Pará (PPGCC/UFPA), em especial ao corpo docente do Programa, pelo ambiente profícuo e fértil que ofereceram para o progresso da minha formação.

SUMÁRIO

INTRODUÇÃO	15
I. Descrição Geral do Problema	15
II. Motivação	17
III. Objetivos	17
IV. Estado da Arte.....	18
V. Estrutura da Dissertação.....	23
CAPÍTULO 1: FUNDAMENTAÇÃO TEÓRICA	24
1.1. Introdução	24
1.2. Mineração de Dados	24
1.2.1. Tarefa de Mineração de Dados.....	26
1.3. Classificadores	27
1.3.1. Software Weka.....	27
1.3.2. Redes Neurais Artificiais.....	28
1.3.3. Maquinas de Vetores de Suporte.....	29
1.3.4. RandomForest	30
1.3.5. K – Vizinho Mais Próximo.....	31
1.4. Tecnologia de Medição Fasorial.....	31
1.4.1. Sistema de Medição FasorialSincronizada.....	32
1.4.2. Unidade de Medição Fasorial (PMU).....	34
1.4.3. Concentrador de Medições Fasoriais.....	34
1.4.4. Transmissão de Dados em SMFS.....	36
CAPÍTULO 2: MATERIAIS E MÉTODOS	38
2.1. Área de Estudo	38
2.2. Metodologia Proposta.....	39
2.3. Análise e Preparação da Base de Dados.....	39
2.3.1. Seleção das Variáveis de Interesse.....	41
2.3.2. Processo de Rotulação da Base de Dados.....	42

2.4. Classificação.....	43
2.4.1. Seleção Automática do Modelo.....	44
2.4.2. Validação Cruzada.....	44
2.5. Medidas de Desempenho.....	45
CAPÍTULO 3: RESULTADOS E DISCUSSÕES	48
3.1. Introdução	48
3.2. Resultados Para a Rede Neural	49
3.3. Resultados Para o RandomForest.....	51
3.4. Resultados Para o KNN	52
3.5. Resultados Para o Classificador SVM.....	53
3.6. Comparação Entre os Classificadores.....	55
CONCLUSÃO	56
I. Trabalhos Futuros	56
REFERÊNCIAS BIBLIOGRÁFICAS	59
APÊNDICE	62
I. Descrição das Variáveis da PMU.....	62

LISTA DE ABREVIATURAS E SIGLAS

ARFF - *Attribute-Relation file format*

IA - Inteligência Artificial

KNN - *K-nearestneighbors*

MLP - *Multilayerperceptron*

RF - Randomforest

RNA - Rede neural artificial

SVM - *Support vector machine*

UHE - Usina Hidrelétrica

WEKA - *Waikato environment for knowledge analysis*

KDD – Descoberta de Conhecimento em Base de Dados

SIN – Sistema Elétrico Interligado

PCHs – Pequenas Centrais Elétricas

ANNEL – Agência Nacional de Energia Elétrica

SMFS – Sistema de Medição Fasorial Sincronizada

SEP – Sistema Elétrico de Potência

LISTA DE SÍMBOLOS

Σ -Somatório

Φ - Função de estimação

k - Kernel (Núcleo)

λ - Autovalores

ω - vetor peso

x_v - vetor de suporte

E_f - Taxa de erro

a_c - Acurácia

T_s – Taxa de sensibilidade

T_e – Taxa de especificidade

LISTA DE FIGURAS

Figura 01. Exemplo de um Sistema Elétrico de Potência. (Fonte: Morais, 2011)	15
Figura 02. Etapas do processo de descoberta de conhecimento (KDD).(Fonte: adaptado de Fayyad et al. 1996)	25
Figura 03. Configuração de um sistema de Medição Fasorial Sincronizada via GPS.....	34
Figura 04. Estrutura básica de uma PMU.....	35
Figura 05. Estrutura básica funcional de um concentrador de dados fasoriais	35
Figura 06. Esquema do sistema elétrico Tramoeste e posicionamento das PMUs ao longo desses sistemas.....	38
Figura 07. Esquema da metodologia proposta para classificação de perturbações em sistemas elétricos de potência a partir de medições fasoriais sincronizadas utilizando técnicas de mineração de dados.....	39
Figura 08. Exemplo de um log do configurador de estado do SAGE.....	40
Figura 09. RMS da forma de onda de tensão da fase B no momento da ocorrência de uma perturbação.....	43
Figura 10. Modelo de Classificação Adotado.....	44
Figura 11. Exemplo do processo de validação cruzada. (Fonte: Faceli, 2011)	45
Figura 12. Gráfico da evolução da taxa de erro do classificador neural considerando a combinação dos parâmetros da tabela 4	49
Figura 13. Gráfico da evolução da taxa de erro do Random Forest considerando o número de árvores	51
Figura 14. Gráfico da evolução da taxa de erro do KNN de acordo com o número de vizinhos mais próximos definidos no <i>grid</i>	52
Figura 15. Gráfico da evolução da taxa de erro do classificador SVM considerando a combinação dos parâmetros da Tabela 9.....	54
Figura 16. Gráfico que compara as taxas de erro (E_f) dos classificadores adotados no trabalho.....	55

LISTA DE TABELAS

Tabela 1.	Medições típicas realizadas em uma PMU.....	41
Tabela 2.	Variáveis selecionadas conforme indicado por especialistas do setor elétrico.....	42
Tabela 3.	Matriz de confusão para um problema de duas classes (positivo e negativo) ..	46
Tabela 4.	Grid de parâmetros adotado no procedimento de seleção automática do modelo	48
Tabela 5.	Listagem da combinação dos parâmetros de uma rede neural. Valores entre parênteses representam H, L e M, respectivamente	50
Tabela 6.	Taxa de sensibilidade T_s , Taxa de especificidade T_e e Medida F obtida pelo classificador neural, considerando o melhor modelo obtido, isto é, H =, L= e M=	51
Tabela 7.	Taxa de sensibilidade T_s , Taxa de especificidade T_e e Medida F obtida pelo classificador randomforest, considerando 10 árvores	52
Tabela 8.	Taxa de sensibilidade T_s , Taxa de especificidade T_e e Medida F obtida pelo classificador KNN com 1 vizinho mais próximo	53
Tabela 9.	Listagem da combinação dos parâmetros dos classificadores SVM. Valores entre parênteses representam C e G, respectivamente	53
Tabela 10.	Taxa de sensibilidade, Taxa de especificidade T_e e Medida F obtida pelo classificador SVM, considerando o melhor modelo obtido, isto é, C=10 e G=100	55

RESUMO

MAIA, Anderson Trindade. **Metodologia para a classificação de perturbações a partir de medições fasoriais sincronizadas utilizando técnicas de mineração de dados**. 2015. 66 f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Pará, Belém, 2015.

Atualmente, procurando atender requisitos de qualidade cada vez mais rigorosos, as empresas do setor elétrico estão se equipando com tecnologias avançadas para aquisição e armazenamento de informações. Um típico exemplo dessa tecnologia são os equipamentos de oscilografia, e mais recentemente os Sistemas de Medição Fasorial Sincronizada (SMFs). Apesar da tecnologia avançada, estabelecer metodologias eficientes para prever a ocorrência de perturbações nos dados coletados dos SMFs, é uma atividade incipiente no setor elétrico. Assim, o objetivo deste trabalho consiste em propor uma metodologia para classificação de perturbações a partir de medições fasoriais sincronizadas utilizando técnicas de mineração de dados. A metodologia proposta utiliza os dados de medições fasoriais de uma linha de transmissão de 230 KV que interliga as subestações de Altamira, Rurópolis e Tucuruí (Pará), que fazem parte do sistema elétrico Tramoeste. Após tratamento e pré-processamento dos dados coletados, foram aplicadas técnicas de mineração de dados no processo de classificação das perturbações. Mais especificamente, foram avaliados sistematicamente, a partir de um procedimento automático de seleção do modelo, os classificadores Random Forest, Redes Neurais Artificiais, K-Vizinho mais Próximo e Máquinas de Vetores de Suporte. Os Resultados obtidos mostram que os classificadores RadomForest(0.41% de erro) e K-Vizinho mais próximo (1.87% de erro), foram os que tiveram o melhor desempenho, indicando que a metodologia proposta é viável e promissora.

Palavras-chave: Mineração de dados, Sistemas Elétricos de Potência, Sistemas de Medição Fasorial Sincronizada.

ABSTRACT

MAIA, Anderson Trindade. **Metodologia para a classificação de perturbações a partir de medições fasoriais sincronizadas utilizando técnicas de mineração de dados**. 2015. 66 f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Pará, Belém, 2015.

Nowadays, seeking to meet the increasingly stringent quality requirements, the electric power industry has advanced technologies to acquire and store information. A typical example is the the oscillography equipment, and more recently Synchronized Phasor Measurement Systems (SMFs). Despite the advanced technology, establish efficient methods to predict the occurrence of disturbances in data collected from SMFs, it is an incipient activity. Thus, this work propose a methodology for disturbance classification from synchronized phasor measurements using data mining techniques. The proposed methodology uses data from phasor measurements of a 230 kV transmission line linking Altamira, Tucuruí and RurópolisTramoeste substations (state of Pará/BRA). After preprocessing of collected data were applied data mining techniques in the disturbance classification process. Ore specifically, we evaluated systematically the following classifiers by automatic model selection: Random Forest, Artificial Neural Networks, K-Nearest Neighbor and Support Vector Machines. The results show that the Random Forest (0.41% error) and K-Nearest Neighbor (1.87% error) outperform the other methods; indicate that the proposed methodology is feasible and promising.

Keywords: Data Mining, Eletric Power Systems, Synchronized Phasor Measurement Systems.

INTRODUÇÃO

I. Descrições Gerais do Problema

Um Sistema Elétrico de Potência (SEP) típico, conforme ilustra a figura abaixo, é geralmente dividido em três zonas funcionais, a saber: geração, transmissão e distribuição (Morais 2011). Tais zonas funcionais estão sujeitas à ocorrência de distúrbios naturais ou provocados pelo homem. Como consequência disso, as formas de onda de tensão ou corrente sofrem certas perturbações, desviando de seus valores nominais, caracterizando os chamados eventos de qualidade de energia elétrica.

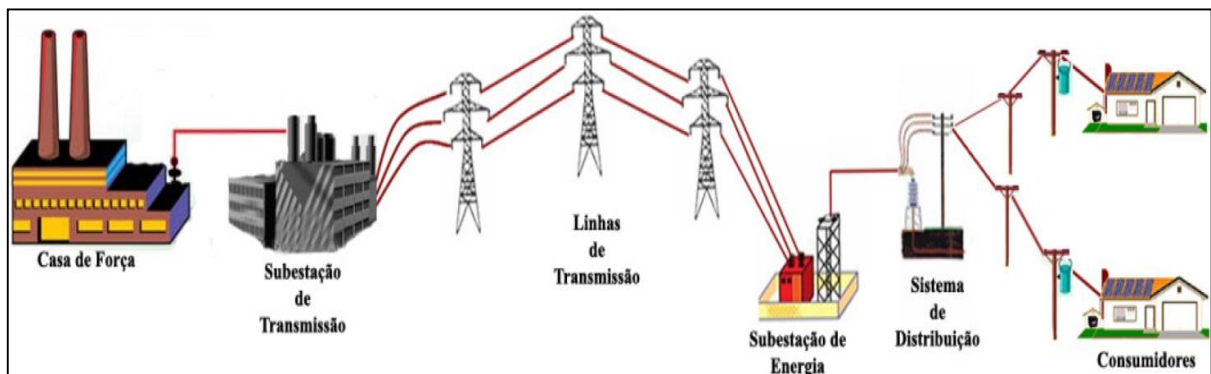


Figura 01: Exemplo de um Sistema Elétrico de Potência. (Fonte: Morais 2011).

Ultimamente, observa-se que o setor elétrico brasileiro vem sofrendo um impacto mais acentuado para conseguir manter a qualidade de energia elétrica que possam atender de forma aceitável e efetiva o aumento da demanda de seus consumidores. Com isso, verificam-se as seguidas paralisações nas obras das hidrelétricas, necessárias ao atendimento da matriz energética do tipo hidráulica, vem 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 os problemas no setor elétrico.

O setor elétrico vem apresentando um aumento da presença de pequenos produtores independentes, principalmente em relação às chamadas Pequenas Centrais Hidrelétricas (PCHs) que são interligadas à rede elétrica de distribuição e sem controle centralizado. Além disso, também existe o aumento na utilização de energia eólica, sabidamente instável e causadora de variações abruptas na geração, no qual vem trazendo à operação do Sistema Elétrico Interligado (SIN) grandes desafios. 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.

Com a reestruturação do setor elétrico, a estabilidade no fornecimento de energia tornou-se fator primordial para o sucesso das empresas distribuidoras de energia. A Agência Nacional de Energia Elétrica (ANEEL), que é o órgão regulador e fiscalizador do sistema elétrico brasileiro, é a responsável por definir padrões de qualidade, bem como as penalidades pelo não atendimento desses padrões. A adoção de novos padrões de desempenho induziu as empresas fornecedoras de energia a reverem suas metodologias de avaliação e atuação no sistema a fim de atender novas exigências.

Diante deste novo paradigma 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 (Bollen et al., 2009). Um típico exemplo dessa tecnologia são os Sistemas de Medição Fasorial Sincronizada (SMFs). (Bakken et al., 2011).

Os SMFS são tecnologias inovadoras para o aprimoramento da operação do SEP. Basicamente, fazem parte desse sistema as unidades de medição fasorial (PMUs - *Phasor Measurements Units*) que detectam e transmitem os valores de tensão, corrente e seus respectivos ângulos de fase sincronizadas no tempo pelos chamados concentradores de medição fasorial (PDC - *Phasor Data Concentrator*), disponibilizando as grandezas medidas para as mais diversas aplicações.

O Avanço tecnológico dos SMFS, quando aplicado a monitoração, por exemplo, pode agregar avanços significativos para o conhecimento do comportamento do setor elétrico. A medição dos chamados sincrofasores já faz com que o estado do SEP seja medido de forma mais exata, rápida e confiável. Logo, essa nova perspectiva trazida à operação pelos SMFS oferece soluções inovadoras do ponto de vista relacionadas à supervisão, ao controle e à proteção dos sistemas de potência. Portanto, surgem novas idéias para o campo da investigação sobre as possíveis aplicações das informações obtidas.

Apesar da tecnologia avançada, estabelecer metodologias eficientes para prever com uma certa antecedência de tempo a ocorrência de perturbações nos dados coletados dos SMFs é uma atividade incipiente no setor elétrico. Uma das razões é a dificuldade no tratamento do grande volume de dados coletados pelos SMFs, que apresentam taxas de amostragem de até a 60 amostras por segundo. Outro problema é que os dados são frequentemente armazenados

sem serem devidamente rotulados, o que complica a aplicação de técnicas de mineração de dados baseadas em aprendizado supervisionado (Witten; Frank, 2005).

II. Motivação

O avanço da tecnologia da informação permitiu que o volume crescente de dados fosse armazenado com menos custo computacional. Contudo, esses dados em seus formatos originais de armazenamento não trazem muito valor agregado para conhecimento e precisam ser tratados para serem úteis e aproveitados. Contudo, uma vez tratados e passados por um processo de mineração de dados, podem revelar conhecimentos relevantes para vários tipos de aplicações como, por exemplo, a determinação de padrões em SEPs.

Durante uma perturbação em um SEP a identificação e o restabelecimento da rede devem ser realizados o mais rápido possível. Para tanto, se faz necessário conhecer a natureza do problema. Um modelo capaz de lidar com as incertezas existentes no problema de classificação se torna extremamente necessário. Assim, o desenvolvimento de uma metodologia baseada em técnicas de mineração de dados que possam apontar com intervalo mínimo de antecedência a ocorrência de uma perturbação, a partir de dados reais oriundos dos SMFs, é a motivação principal desta dissertação.

III. Objetivos

O objetivo geral deste trabalho consiste em propor uma metodologia para classificação de perturbações em linhas de transmissão a partir de medições fasoriais sincronizadas utilizando técnicas de mineração de dados.

Como objetivos específicos, este trabalho visa:

- a) Desenvolver um aplicativo que recupere de forma seletiva dados da medição fasorial de acordo com um intervalo de tempo. Mais especificamente, este aplicativo fará a leitura dos arquivos provenientes das PMUs que possuem um formato específico e são do tipo binário. Além disso, dada à elevada taxa de amostragem da medição fasorial, serão selecionadas amostras apenas dentro de janelas de intervalos de tempo em torno das ocorrências de perturbações de interesse, sem prejuízo da confiabilidade dos dados e sua relevância na descrição dos fenômenos sob análise;

- b) Selecionar as variáveis (parâmetros) de interesse para o problema que poderão incluir, além das variáveis armazenadas pelas PMUs, variáveis calculadas a partir dessas como, por exemplo, a diferença angular de tensões entre barras de instalações diferentes;
- c) Rotular as amostras coletadas e pré-processadas com o auxílio de um especialista para que se possa treinar e testar os classificadores utilizados no trabalho;
- d) Avaliar de forma sistemática alguns dos principais algoritmos de classificação utilizados na literatura. Neste caso, a escolha dos valores que definem os modelos dos classificadores, tais como o número de neurônios na camada escondida de uma rede neural, é feita através de um processo automático denominado de seleção do modelo (*modelselection*);
- e) Comparar o desempenho dos classificadores utilizando métricas de avaliação tais como taxa de acerto, taxa de especificidade, taxa de sensibilidade e outras.

IV. Estado da Arte

Existem inúmeros trabalhos relacionados à proposta do presente estudo. A maioria aborda técnicas de mineração de dados na predição de contextos elétricos, sendo alguns destacados nessa sessão. Uma abordagem observada nessa direção é a análise de perturbações em SEPs com informações oriundas de bases de dados simuladas e extraídas de PMUs, no qual são destacados alguns estudos.

Em Zimmer et al. (2012) foi feito um estudo onde são descritos resultados que fazem detecção automática de perturbações sistêmicas em redes elétricas. Um sistema de medição fasorial instalado na rede de baixa tensão é utilizado para obter dados de perturbações no SIN. A metodologia proposta consiste na utilização de um filtro de média móvel, do cálculo da taxa de variação da frequência e de um algoritmo de detecção e validação da perturbação. Três estudos de casos de ocorrências de perturbações na rede elétrica do SIN são utilizados para demonstrar seu funcionamento, nos quais as alterações foram detectadas, inclusive sinalizando suas respectivas localizações. O estudo apresentou resultados que demonstram que a metodologia utilizada nessa pesquisa pode ser utilizada para extração de dados reais de PMUs para predições de perturbações em SEPs.

Por outro lado, Gómes e Rios (2013) apresentam dois aspectos considerados ideais para o problema da instalação de PMUs. Em primeiro lugar, um modelo de ILP (Integer

Linear Programação) encontra o número de PMU e a sua colocação em fases separadas, e maximiza a observabilidade do sistema em cada período de tempo. O modelo leva em conta para PMUs: o orçamento disponível por cada fase, a expansão do sistema de energia, juntamente com a colocação múltiplos estágios, a redundância na colocação, links de comunicação, definição do usuário para restrições de tempo de atribuição e a injeção sem efeito. Em segundo lugar, é proposta uma metodologia para identificar autocargas a serem observados para a monitorização da estabilidade dinâmica. A metodologia identifica grupos coerentes em grandes sistemas de energia usando uma nova técnica baseada na teoria dos grafos. A metodologia é aplicada sobre o sistema de teste de 14 barras. Os resultados indicam que o modelo de otimização encontra o número ótimo de PMU quanto a colocação por fases, enquanto a observabilidade em cada fase é maximizada. Além disso, mostra-se que os planos de expansão e requisitos particulares de observabilidade pode ser considerado no modelo sem aumentar o número de PMU necessária, e o efeito de injeção de zero, o que reduz o número de PMU, pode ser considerado no modelo.

Zhang (et al, 2013) fez uma pesquisa onde demonstra a classificação de perturbações em SEPs com base de dados simulada. O estudo utilizou um ANFIS (Sistema Artificial de inferência neural-fuzzy) esquema de classificação de perturbações com base no sistema de distribuição de aterramento-efetivo. As correntes transitórias são obtidas por transformada *wavelet* após as perturbações ocorrerem. De acordo com a característica estatística de correntes transitórias em diferentes tipos de perturbações. Os mesmos podem caracterizar os traços do tipo de falta e mostrar um disciplinador diferente em variados tipos de perturbações. Os dados são introduzidos em três ANFIS para obter o tipo de falta. A abordagem proposta necessita apenas as tensões e correntes medidas na subestação, e pode identificar dez tipos de perturbações do tipo curto-circuito com precisão. O modelo de simulação é estabelecida em ambiente chamado PSCAD/EMTDC, e o desempenho da abordagem da pesquisa é analisada. Os resultados mostram que a proposta do estudo tem alta precisão. Além disso, a adaptabilidade das abordagens propostas para o sistema de aterramento e neutro é compensado em diferentes configurações de rede através de simulações.

Liao (2014) fez um estudo que apresenta a localização exata de perturbações em um sistema de distribuição de energia elétrica, no qual é importante para manter a confidencialidade do sistema. Diversos métodos foram propostos no passado em relação a esse contexto, que normalmente têm diferentes pressupostos e, portanto, são aplicáveis às circunstâncias específicas. O estudo utiliza um método geral de localização de perturbações que é aplicável às redes de distribuição com desequilíbrios e multi-fontes, empregando

tensões e correntes na subestação local. O método não requer que o tipo de defeito seja conhecido, e é aplicável a qualquer tipo de perturbações. O experimento é baseado na matriz de impedância que permite o estabelecimento das equações que regem a relação entre as medições e a localização de falhas. Os estudos de avaliação demonstraram a eficácia da proposta e a sua robustez com respeito a possíveis erros de medição e as variações de carga que podem prevenir futuros desligamentos em toda rede.

A publicação de Zin et al. (2015) apresenta um novo algoritmo para detecção de anomalias e classificação de perturbações em linhas de transmissão de energia usando transformada de *wavelet* discreta (DWT) e rede neural *back-propagation* (BPNN) com base na transformação de Clarke sobre a transmissão paralela. No modo Alfa e beta correntes geradas pela transformação de Clarke foram utilizados para converter o sinal de DWT e obter os coeficientes de transformada *wavelet* (WTC) e também coeficiente de energia *wavelet* (WEC). Daubechies4 (DB4) foi usado como uma *wavelet* mãe para decompor os componentes de alta frequência do sinal de erro. A simulação foi realizada utilizando PSCAD/EMTDC para modelagem do sistema de transmissão. A simulação foi realizada em diferentes locais ao longo da linha de transmissão com diferentes tipos de perturbações, resistência de falta, localização de perturbações e ângulo inicial de falha em um determinado modelo de sistema de potência. Quatro métodos estatísticos utilizados foram utilizados para determinar a precisão, detecção e classificação das perturbações. Os resultados mostram que a melhor transformação Clarke ocorreu na configuração de 12-24-48-4, respectivamente. Por exemplo, os erros usando o método de erro médio quadrado, os erros de BPNN, Reconhecimento de Padrões Rede e Fit Rede são 0,03721, 0,13115 e 0,03728, respectivamente. Isto indica que os resultados do BPNN teve o menor erro.

Em Fathabadi (2015) são apresentadas duas abordagens para detectar perturbações do tipo curto-circuito em linhas de transmissão de energia que podem fazer previsões de desligamentos. Os dois métodos propostos abrangem aspectos teóricos e técnicos. A primeira abordagem é um método de computação que usa transformada *wavelet* discreta. A segunda abordagem é um método baseado em *hardware* que utiliza uma nova proposta em duas fases de filtro de resposta de impulso finito com uma frequência de amostragem de 32 kHz, e um tempo muito curto de processo em três amostras de tempo. Os resultados obtidos através da simulação de 230 kV, 50 Hz em linha de transmissão de energia trifásico são dados para validar os resultados teóricos, e verificar explicitamente que a abordagem baseada em filtro tem uma precisão de 100% na presença de 10% de perturbações enquanto a precisão do abordagem baseada transformada *wavelet* é de aproximadamente 97%, entretanto é menos

complexo e com pouco custo de implementação.

Babu e Mohan (2015), realizaram um estudo de classificação de perturbações que podem ocasionar desligamento em linhas de transmissão de energia. Neste Trabalho foi proposto uma classificação dessas perturbações usandoo Modo Empírico de Decomposição (EMD) e Support Vector Machines (SVMs). EMD é usado para decompor tensões de linha de transmissão em funções de modo intrínseco (IMFs). *Huang Hilbert Transform* (HHT) é usado para a extração de características de IMFs. Um modelo SVM múltipla é introduzida para classificar a condição das perturbações entre as dez perturbações no sistema de potência. O algoritmo é validado usando ambiente MATLAB/ SIMULINK. Os resultados demonstram que a combinação de EMD e SVM podem ser um classificador eficiente com níveis aceitáveis de precisão.

Chompoobutrcool e Vanfretti (2015), apresenta um estudo abrangente de informações extraídas de PMUs em SEPs que analisa as oscilações dos sinais dominantes em inter-área (*WADC - wide-area damping control*) e sua aplicação para controle de amortecimento do sistema de energia de área ampla (WADC). A análise, realizada em ambos os pequenos e grandes sistemas de estudo, centra-se nas relações que surgem a partir de características físicas de oscilações inter-área, ou seja, a observação modal de sinais em caminhos dominantes e suas propriedades de sistema de *loop* em controle correspondente (ou seja, estabilidade e robustez). O objetivo é ser capaz de explorar adequadamente os sinais de caminho dominantes para a mitigação das oscilações inter-área. Diretrizes e considerações são fornecidos para facilitar a concepção de WADC usando a abordagem proposta.

Zamani et al. (2015) estudou a aplicação de PMU como medida protetiva especial com segurança avançada em SEPs. Um método híbrido baseado no PMU é proposto para avaliar os limites de estabilidade transitória de um sistema multi-máquina com antecedência de contingências de rede. O método é baseado na técnica única de máquina equivalente (SIME), que combina algoritmos no domínio do tempo com o método direto de critério de área igual, para determinar os limites de potência de estabilidade sobre as grandes linhas de transmissão. Além disso, com base na condição da rede em tempo real, obtidos por medições sincrofasores, uma estratégia é proposto para determinar se eficazmente um sinal de alarme deve ser emitido para a próxima contingência. A eficácia do método de alarme automático baseado em PMU é avaliado e discutido por meio de estudos de simulação realizados no simulador digital em tempo real (RTDS) e ambiente de software MATLAB.

Em Mohammadi e Maryam (2015) apresentam a necessidade de seleção de recursos e redução de dimensão como um passo fundamental na avaliação de sistemas de energia em que

o número de recursos que representa o estado de redes de energia aumenta drasticamente a segurança. Acreditam que essas grandes quantidades de atributos não são apropriados para serem utilizados como entradas para técnicas de inteligência computacional, porque pode levar a um procedimento demorado com resultados insuficientes e eles não são adequados para fins on-line e atualizações.

Os referidos pesquisadores propuseram um método combinado para uma avaliação de segurança on-line de tensão em que a dimensão dos dados de *token* de PMU é reduzido por meio de análise de componentes principais (PCA). Em seguida, os recursos com diferentes índices de estabilidade são colocados em várias categorias e seleção de recursos é feita por análise de correlação em cada categoria. Estas características selecionadas são, então, dado a árvores de decisão (DTS) para classificação e avaliação na segurança de sistemas de energia. O método é aplicado ao sistema de teste de 39 subestações de uma parte da rede de energia do Irã. Os resultados apresentam que as DTs com dados reduzidos têm regras de decomposição mais simples, melhor desempenho na economia de tempo, o erro DT razoável e são mais adequados para atualizações constantes.

Os trabalhos de Chompoobutrgool, Vanfretti(2015), Zamani et al. (2015), Guo et al. (2015) e Mohammadi e Maryam (2015) serviram de referência básica para o estudo, pois utilizam informações armazenadas em PMUs. Apesar de serem pesquisas que abordam estudos relacionados ao uso de base de dados reais, não enfocam a classificação de perturbações de informações extraídas de PMUS. Diferente de Zimmer (2012) que serviu de referência para estudo, pois apresenta uma metodologia enfocando diretamente o emprego de PMUs na classificação de perturbações utilizando filtro de média móvel, cálculo da taxa de variação da frequência e de um algoritmo de detecção e validação da perturbação, servindo como principal referência para o estudo e validando a originalidade da dissertação e sua relevância com uma nova proposta de metodologia que aborda o emprego de técnicas de mineração de dados para esse fim.

V. Estrutura da Dissertação

Esta dissertação é dividida em três sessões. Após esta introdução, no Capítulo 1 são explicados os elementos teóricos utilizados neste trabalho. No Capítulo 2, intitulado “Fundamentação Teórica”, é apresentada a metodologia proposta, enfatizando-se a área de estudo, o processo de composição da base de dados com as medições fasoriais sincronizadas coletadas e o processo de classificação de perturbações. No Capítulo 3 são apresentados os resultados obtidos pelos classificadores como prova de conceito da metodologia aplicada. Na sessão final são apresentadas as conclusões e sugestões de trabalhos futuros.

CAPÍTULO 1

FUNDAMENTAÇÃO TEÓRICA

1.1. Introdução

Este capítulo tem como objetivo fornecer uma descrição conceitual sobre mineração de dados, classificadores e as tecnologias de medição fasorial, que são os principais assuntos de interesse das áreas de estudo que compõem esta dissertação.

1.2. Mineração de Dados

É crescente a quantidade de dados armazenados em um sistema computacional. Uma boa parte desses dados, em particular dados coletados automaticamente por aplicações de monitoração, excede a capacidade de compreensão humana. Conseqüentemente grandes quantidades de dados tornam-se “dados mortos” sem nenhuma utilidade. Entretanto, dependendo do contexto os mesmos podem se configurar como um agente potencial no que se refere ao suporte a tomada de decisão. Por exemplo, em sistemas elétricos de potência, onde os dispositivos de monitoramento possuem elevadas taxas de amostragem, é possível prever o tempo de vida útil de um equipamento de proteção elétrica através da aplicação de técnicas de mineração de dados (Pires, 2009).

Portanto, a mineração de dados vem despertando o interesse de muitos especialistas no setor elétrico devido a ampla disponibilidade e heterogeneidade de dados armazenados eletronicamente em inúmeros equipamentos que assistem os SEPs, e também a necessidade de encontrar novos padrões ou transformar dados ocultos em informações úteis que possam ser aplicados como suporte à tomada de decisão ou a avaliação de resultados.

Por ser considerado multidisciplinar, existem várias definições para o termo mineração de dados. A seguir, são destacados os conceitos mais utilizados:

- a) Fayyadet al. (1996): “Mineração de dados é um passo no processo de Descoberta de Conhecimento que consiste na realização da análise dos dados e na aplicação de algoritmos de descoberta que, sob certas limitações computacionais, produzem um conjunto de padrões de certos dados.”;

- b) Hanet al. (2011): “Mineração de dados é análise de grandes conjuntos de dados a fim de encontrar relacionamentos inesperados e de resumir os dados de uma forma que eles sejam tanto úteis quanto compreensíveis ao dono dos dados.”;
- c) Hanet al. (idem): “Mineração de dados é análise de grandes conjuntos de dados a fim de encontrar relacionamentos inesperados e de resumir os dados de uma forma que eles sejam tanto úteis quanto compreensíveis ao dono dos dados.”;
- d) Cabenaet al. (1998): “Mineração de dados é um campo interdisciplinar que junta técnicas de máquinas de conhecimentos, reconhecimento de padrões, estatísticas, banco de dados e visualização, para conseguir extrair informações de grandes bases de dados.”.

Alguns autores definem mineração de dados como uma etapa dentro do processo de descoberta de conhecimento em base de dados (*Knowledge Discovery in Databases-KDD*) (Fayyadet al., 1996; Krzysztofet al., 2007) e outros como sinônimo de KDD (Wang2005; Han et al., 2011). No entanto, todos concordam que o processo de mineração deve ser iterativo, iterativo e dividido em fases. A seguinte ilustração apresenta as etapas do processo de KDD:

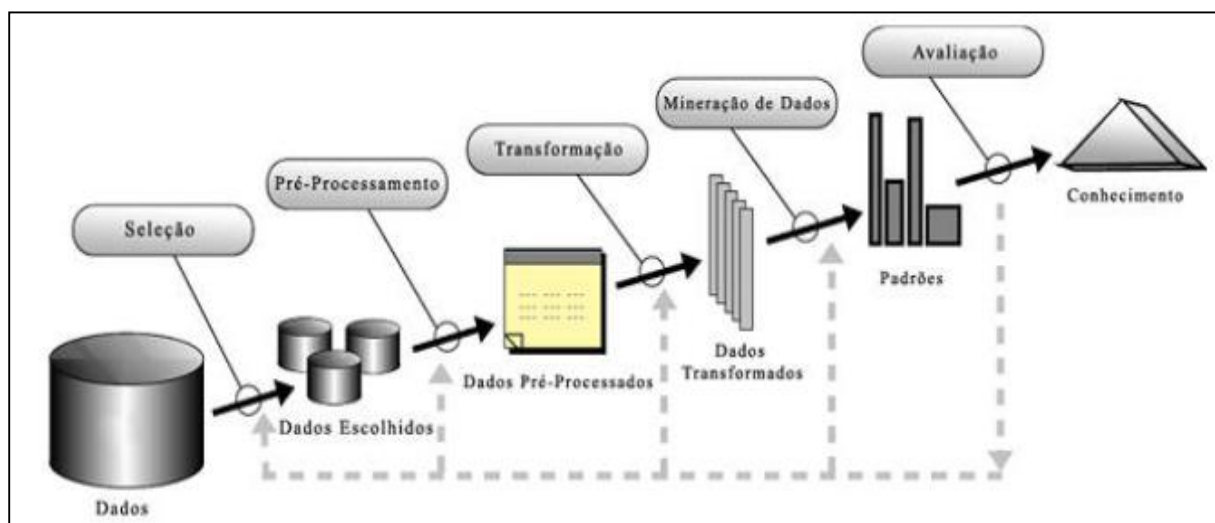


Figura 02: Etapas do processo de descoberta de conhecimento (KDD). (Fonte: adaptado de Fayyadet al., 1996).

Como se pode observar na figura acima, existem várias etapas no processo de KDD para que se possa chegar à resposta desejada. Hamrouniet al. 2015 coloca que a fase que compõem o Pré-Processamento de dados (limpeza, integração, seleção e transformação dos dados)são etapas que tem uma relevância significativa dentro de todo o processo, pois são procedimentos essenciais para que algoritmos de mineração de dados possam retornar

resultados mais precisos. A seguir é apresentada uma breve descrição das etapas apresentadas na Figura 2:

- a) Seleção: Fase que são escolhidos os atributos mais relevantes para o domínio do problema;
- b) Pré-processamento: consiste na limpeza (retirada de ruídos e informações desnecessárias) e na integração de dados;
- c) Transformação: transforma os dados para um formato apropriado quando necessário;
- d) Mineração de dados: aplicação de algoritmos inteligentes baseados em aprendizado de máquina para extração de padrões de comportamento nos dados;
- e) Interpretação e avaliação dos dados: etapa onde são avaliados e identificados os padrões de interesse para o domínio do problema, os quais darão suporte à tomada de decisões humanas.

1.2.1. Tarefas da Mineração de Dados

A mineração de dados geralmente é categorizada pela sua capacidade em realizar determinadas tarefas. São exemplos de tarefas da mineração de dados (Hanet al., 2011):

- a) Associação: consiste na descoberta de quais atributos estão relacionados. Apresentam a forma: SE atributo X ENTÃO atributo Y;
- b) Sumarização: é a tarefa utilizada para encontrar uma descrição simples e compacta do conjunto de dados. A sumarização pode ser aplicada em diferentes níveis de abstração e pode ser vista a partir de diferentes ângulos;
- c) Agrupamento (*clustering*): tarefa que tem como objetivo identificar e aproximar dados que possuem características semelhantes. Um agrupamento (*cluster*) é uma coleção de registros similares entre si, porém diferentes dos registros dos outros grupos;
- d) Classificação: tarefa que consiste na busca de uma função (mapeamento) que permita associar cada registro a um rótulo categórico denominado de classe. Uma vez determinada em uma fase de treino, essa função pode ser aplicada em uma fase de teste a novos registros, de forma a prever as classes nas quais tais dados melhor se enquadram;
- e) Regressão: é uma tarefa muito similar a classificação, porém é usada quando a saída desejada é dada por um valor numérico e não um categórico;

Neste trabalho a tarefa aplicada é a classificação. Mais especificamente, considerando as amostras das medições fasoriais sincronizadas obtidas, a tarefa ao longo desse trabalho consiste em classificar esses dados em uma das 3 perturbações possíveis rotuladas de acordo com a metodologia apresentada na Seção 3.2: normal, pré-perturbação e perturbação.

1.3. Classificadores

Um *classificador* F é um mapeamento $F: \mathbb{R}^k \rightarrow \{1, \dots, Y\}$, onde K é a dimensão do vetor de entrada $x \in \mathbb{R}^k$ e o *rótulo* $y \in \{1, \dots, Y\}$ é a classe. Quando se treina um classificador computacional usando aprendizado supervisionado, é dado um *conjunto de treinamento* $T = \{(x_1, y_1), \dots, (x_n, y_n)\}$ contendo N exemplos de (x, y) .

Para tarefa de classificação, foi utilizado o software de mineração de dados WEKA que será discutido na próxima seção. As seções seguintes listarão os principais classificadores utilizados. Como cada um desses classificadores poderia render material para uma dissertação inteira, ao invés de discuti-los em profundidade, busca-se prioritariamente ilustrar como os mesmos são usados no WEKA.

1.3.1. Software Weka

O WEKA (*Waikato Environment for Knowledge Analysis*) é reconhecido por profissionais da área acadêmica e científica como uma ferramenta de referência em aprendizado de máquina e mineração de dados. A ferramenta apresenta um conjunto de implementações de algoritmos que utilizam diversas técnicas de inteligência computacional. Esta ferramenta foi implementada na linguagem de programação Java, tornando-o acessível nas principais plataformas computacionais (Witten & Frank 2005; Hall et al, 2009).

O WEKA inclui algoritmos de regressão, classificação, agrupamento, regras de associação e seleção de parâmetros (atributos). Atualmente está na versão 3.6.13, sendo organizado em três módulos de operação. No primeiro, *Simple Command Line Interface* (CLI), a interação do usuário com WEKA ocorre através de linhas de comando. O Explorer é considerado o principal módulo, pois executa a interface gráfica para execução dos algoritmos de aprendizado de máquina suportados pelo WEKA. E o terceiro, é o módulo *Experimenter* no qual o usuário, também por meio de interface gráfica, executa testes estatísticos em diferentes algoritmos simultaneamente a fim de avaliar os resultados obtidos.

Primeiramente, antes de utilizar qualquer pacote do software WEKA, os dados devem ser convertidos para um formato de arquivo que a ferramenta consiga entender. Nesta dissertação o formato adotado é próprio do WEKA denominado de arff (*Attribute-RelationFileFormat*). Um arquivo no formato arff é um arquivo ASCII dividido em três partes. A primeira parte, denominada relação indicada pelo marcador “@relation” que ocupa a primeira linha do arquivo e identifica o nome da relação. A segunda parte, iniciada com o marcador “@attribute” contém a lista de todos os atributos (ou parâmetros), onde é necessário definir o tipo e os valores que eles podem assumir. Ao utilizar os valores categóricos para os parâmetros é importante ressaltar que estes devem permanecer entre chaves e separados por vírgula. Por último, a terceira parte, encontra-se logo após a linha que apresenta o marcador “@data” e consiste dos exemplos (*instances*), isto é, cada linha representa uma entrada para os algoritmos do WEKA. Vale ressaltar que um arquivo no formato arff não especifica qual parâmetro é a classe, permitindo desta maneira que o arquivo seja mais flexível. Portanto, o que o WEKA entende por *instance*, reúne o que a notação que será adotada chama de *instancez* mais o rótulo *y*.

O WEKA possui diversos algoritmos de classificação, sendo que apenas alguns dos principais, tais como Rede Neural Artificial (RNA) multicamadas treinadas com algoritmo *backpropagation*, Random Forest (RF), máquinas de vetores de suporte (*Support Vector Machine* - SVM) e K-vizinho mais próximo (*K-NearestNeighbor*(KNN)), foram utilizados. O motivo para escolha destes algoritmos, é que eles representam alguns dos principais paradigmas de aprendizado de máquina.

1.3.2. Redes Neurais Artificiais

As Redes Neurais Artificiais são sistemas paralelamente divididos em camadas interligadas por pesos sinápticos e compostas por neurônios artificiais que calculam certa função matemática (usualmente não linear). Tais unidades são dispostas em uma ou mais camadas e interligadas pelos chamados pesos sinápticos (Haykin et al., 2009).

Neste trabalho foi utilizada uma Rede Neural do tipo *perceptron* de múltiplas camadas (*MultiLayerPerceptron* - MLP) treinada com o algoritmo *backpropagation* também chamado de regra delta generalizada (Haykin et al., 2009). Mais especificamente, foi utilizada a classe *MultiLayerPerceptron* do WEKA. Os principais parâmetros dessa classe são:

- a) -L: corresponde à taxa de aprendizado utilizada pelo algoritmo *backpropagation* e este valor deve ser entre 0 e 1 (Padrão é 0.3);

- b) -M: taxa de momento para o algoritmo backpropagation e este valor deve ser entre 0 e 1 (Padrão é 0,2);
- c) -N: este parâmetro corresponde ao número de épocas para treinamento da rede e o padrão é 500;
- d) -H: Corresponde à quantidade de camadas ocultas que podem ser criadas na rede. Por exemplo -H 3, 2 cria duas camadas intermediárias com 3 e 2 neurônios, respectivamente.

Outra forma de representar pode ser através do uso de letras: a opção corresponde “a” (números de atributos + número de classes)/2; as outras opções são: “i” (número de atributos), o (número de classes) e, “t” (números de atributos + número de classes).

1.3.3. Máquinas de Vetores de Suporte

As máquinas de vetores de suporte constituem uma técnica de aprendizado de máquina fundamentada pela teoria da aprendizagem estatística. São geralmente aplicadas usando um conjunto de treinamento selecionado aleatoriamente e classificado com antecedência. As SVMs usam um dispositivo chamado de mapeamento do *kernel* para mapear dados no espaço amostral. A finalidade de um classificador SVM está em encontrar um hiperplano (superfície de decisão) que maximize a separação no espaço de classes (Faceli, 2011; Vapnik, 2013).

Um ponto importante no algoritmo de aprendizagem por vetor de suporte é a função *kernel* ou função núcleo. Na literatura, várias possibilidades de *kernel* SVM são apresentadas em aplicações envolvendo reconhecimento de padrões, tais como: linear, polinomial, sigmóide e funções de base radial (*Radial BasisFunction*- RBF).

As SVMs, e outros métodos *kernel*, podem ser caracterizados como uma função de estimação φ definida a seguir:

$$\frac{1}{V} \sum_{v=1}^V \cdot (x_v, x_v) + \lambda \varnothing_{H_k}^2$$

Equação1

Onde:

- H_k corresponde ao espaço euclidiano gerado pelo kernel K ;

- $\varphi = h + b$, onde h corresponde ao produto do vetor (ω) pelo vetor de suporte (x_v) com $h \in H_k$;
- b corresponde ao bias, $b \in \mathbb{R}$;
- $L(\varphi(x_v), y_v)$ correspondente a função perda (risco fundamental);
- λ corresponde a auto valores;
- V corresponde ao número de exemplos de treino;

No WEKA, o classificador SVM é implementado na forma de um módulo extra. Este classificador implementa uma versão da LibSVM otimizada para uso com o WEKA (El-Manzalawy; Honavar, 2005). Para lidar com o problema de multi-classes, as SVMs são organizadas na LibSVM no esquema *one-versus-one* (ou all-pairs). (Rifkin; Klautau, 2004). No escopo desse trabalho, foi utilizada a função kernel RBF. Assim, os principais parâmetros utilizados foram o -C, parâmetro de penalidade do termo de erro, e o -G, largura da função dos kernels.

1.3.4. Random Forest

O classificador Random Forest utiliza mecanismos de dividir para conquistar para resolver problemas de decisão de modo que cada árvore depende dos valores de um vetor aleatório disposto de forma independente e com a mesma distribuição para todas as árvores da floresta. A idéia principal é que problemas mais complexos possam ser divididos em problemas mais simples de forma recursiva. Essas soluções podem ser combinadas em forma de árvore para gerar uma solução do problema maior. A proposta é que dividindo o espaço de exemplos em subespaços que 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 (Faceli, 2011).

No WEKA, o Random Forest tem inúmeros parâmetros, sendo que os principais são (Zhou et al., 2014):

- a) - Depth: responsável pela profundidade da árvore.
- b) -K: número de recursos em cada iteração.
- c) -I: número de árvores do modelo.

1.3.5. K–Vizinho Mais Próximo

Os classificadores vistos anteriormente são caracterizados pelo fato de utilizarem os dados de treinamento para construir um modelo de classificação, o qual, uma vez encontrado e testado, estará pronto para testar qualquer padrão novo. Diferentemente desses classificadores, o classificador KNN (K-vizinho mais próximo) utiliza os próprios dados de treinamento como modelo de classificação, isto é, para cada novo padrão que se quer classificar, utiliza-se os dados do treinamento para verificar quais são os exemplos nessa base de dados que são mais próximos do padrão em análise. A cada novo padrão a ser classificado faz-se uma varredura nos dados de treinamento, o que provoca um grande esforço computacional.

Considerando um conjunto de treinamento com M exemplos e seja $z = (z_1, \dots, z_k)$ uma nova *instance*, ainda não classificada, a fim de classificá-la, calcula-se as distâncias, através de uma medida de similaridade, entre z e todos os exemplos do conjunto de treinamento e considera-se os K exemplos mais próximos (com menores distâncias) em relação z . Verifica-se então, qual a classe que aparece com mais frequência, entre os K vizinhos encontrados. O padrão z será classificado de acordo com a classe y mais frequente dentre os K exemplos encontrados (Witten & Frank, 2005; Faceli 2011).

O KNN no WEKA é implementado na classe IBK e seus principais parâmetros são:

- a) -N: número de centros (ou K);
- b) -S: esta opção gera aleatoriamente os centros.

1.4. Tecnologia de Medição Fasorial

Nos Sistemas Elétricos de Potência, no que se refere à área de operação, observa-se que no decorrer dos últimos anos vários trabalhos de pesquisa têm abordado a possibilidade de capturar informações fasoriais, enfatizando o desenvolvimento de sistemas de medição fasorial sincronizada baseado em equipamentos como as PMUs. Nesse tipo de sistema todas as medidas são realizadas com base no mesmo instante de tempo, a partir da fonte de sincronização fornecida por um sinal de GPS (*Global Positioning System*), que também controlam medidas de magnitude e ângulo das tensões nas barras, permitindo analisar o estado real de sistemas elétricos (Singh et al., 2011).

Diante deste cenário, segundo Tochetto (2014), novas tecnologias baseada em medição fasorial sincronizada estão sendo propostas por grupos de pesquisa e também por empresas do setor elétrico. Isto é possível, pois os sistemas de medição fasorial sincronizada viabilizam a medição de grandezas em subestações geograficamente distantes, a uma taxa que varia de 10 a 60 medidas por segundo e com precisão angular adequada aos parâmetros das aplicações que utilizam essa grandeza. Essas características quando aliadas as necessidades tecnológicas atuais potencializam o desenvolvimento de novas ferramentas para supervisão e controle do sistema em tempo real.

Neste sentido, para que um sistema de medição fasorial sincronizada seja compreendido melhor, faz-se necessário apresentar alguns conceitos sobre o funcionamento e os componentes que fazem parte deste sistema. Tais conceitos serão apresentados nas próximas seções.

1.4.1. Sistema de Medição Fasorial Sincronizada

Sistema de medição fasorial sincronizada é um sistema de medição simultânea de grandezas elétricas, que na maioria das vezes são coletadas em unidades de instalações geograficamente distantes umas das outras, usando os equipamentos de medição fasorial (PMU), conectadas a um concentrador de dados fasoriais PDC. Este último equipamento, normalmente é uma unidade lógica que coleta os dados fasoriais e os dados de eventos registrados por uma PMU (Zamoraet al., 2014).

De acordo com Benmouyalet al. (2002) um fasor “é um número complexo associado a uma onda senoidal. A magnitude do fasor é a mesma que a magnitude da onda senoidal. O ângulo de fase do fasor é a fase da onda para $t = 0$ ”. Considere um sinal senoidal apresentado na Equação abaixo:

$$x(t) = X_m \cos(\omega t + \phi)$$

Equação 2

Onde:

- ω é a frequência do sinal em radianos por segundo;
- ϕ é a fase em radianos;
- X_m é a amplitude máxima do sinal e

- $\frac{X_m}{\sqrt{2}}$ é o valor RMS (Root Mean Square) ou valor eficaz do sinal.

A representação do fasor desse sinal senoidal é dado por:

$$x(t) = \frac{X_m}{\sqrt{2}} e^{j\phi} = \frac{X_m}{\sqrt{2}} (\cos \phi + j \sin \phi)$$

Equação 3

O sinal senoidal e sua representação fasorial dado pelas equações 2 e 3 são representados na figura seguinte:

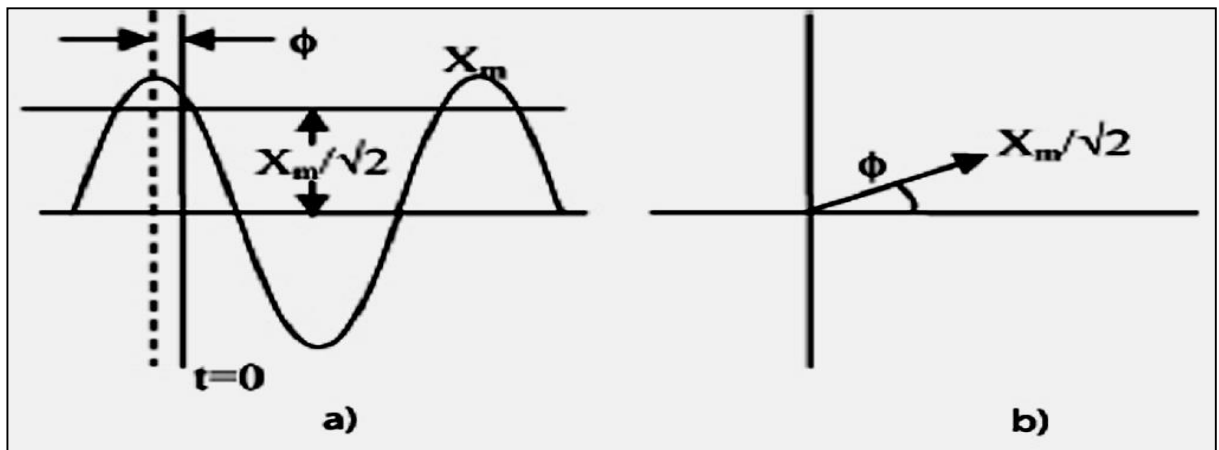


Figura 2: Representação fasorial de um sinal senoidal: a) Sinal senoidal; b) Representação fasorial.
(Fonte: Singh et al., 2011).

A próxima figura apresenta a configuração de um sistema de medição fasorial sincronizada. Como se pode observar na figura, as informações são coletadas a partir das PMUs localizadas em pontos estratégicos do sistema, de maneira que os dados (fasores) são sincronizados via GPS recebendo um selo de tempo e posteriormente encaminhados aos concentradores de dados fasoriais (PDC), que organizam as informações recebidas de forma a agrupar aquelas que correspondem a um mesmo instante de amostragem.

Além da PMU e do PDC outros equipamentos participam do funcionamento ideal desse tipo de sistema. Por exemplo, os equipamentos de GPS permitem amostragens das grandezas elétricas de forma sincronizada no tempo, isto é, tomar amostragens simultâneas ainda que os instrumentos de medição não estejam próximos ou nem na mesma instalação.

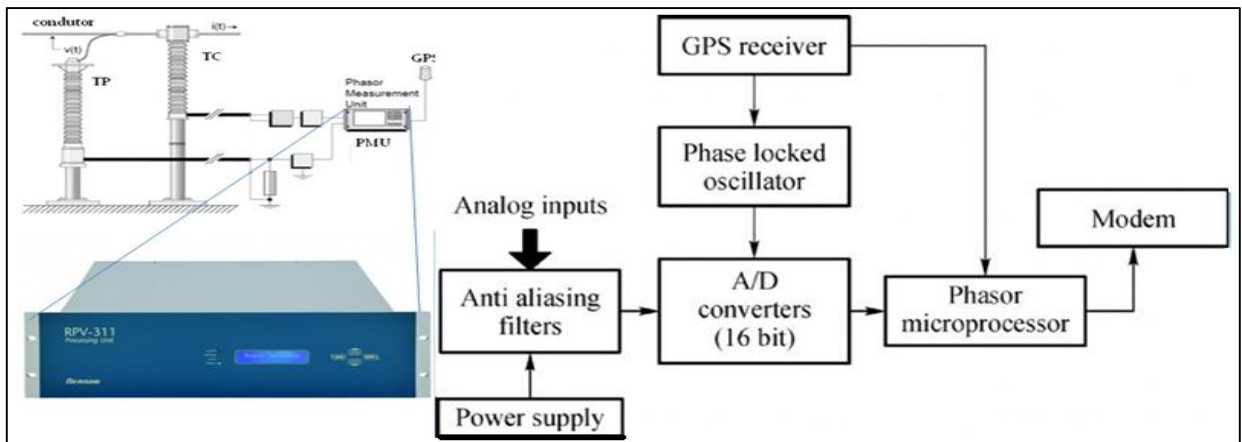


Figura 03: Configuração de um sistema de Medição Fasorial Sincronizada via GPS.

1.4.2. Unidade de Medida Fasorial

A unidade de medida fasorial (PMU) é fundamentalmente responsável pela medição de tensões e correntes alternadas na forma de fasores, com adição de uma *tag* de tempo, obtida do sistema de sincronização. A PMU obtém utilizando a sincronização de um GPS e uma cadência que pode variar entre 10 até 60 amostras por segundo, as amplitudes e fases dos dados selecionados no exato instante da passagem pelo milissegundo exato.

A estrutura de uma PMU pode ser compreendida de modo simplificado a partir da Figura4 (abaixo). As entradas analógicas registradas pela PMU passam inicialmente por um filtro. Depois de filtrado, o sinal de entrada (tensão e corrente) é transformado em valores discretos através de um conversor analógico-digital (A/D). A etapa seguinte refere-se ao processamento matemático das amostras a partir de um microprocessador que realiza o cálculo dos fasores usando Transformada Discreta de Fourier [*Discrete Fourier Transform – DF*](Fechine, 2010). Por fim, os dados recebem um selo de tempo via GPS e são enviados a outro equipamento, os concentradores de medição fasorial(Singhet al., 2011).

1.4.3. Concentrador de Medições Fasoriais

O concentrador de dados fasoriais tem como funcionalidade receber informações dos fasores coletados pelas PMU através dos canais de comunicação, organizando de maneira assíncrona, para em seguida disponibilizá-los de acordo com as aplicações solicitadas, tanto em tempo real quanto para estudos *off-line*. Além disso, ele deve ser capaz de fazer um

tratamento de erros de transmissão, solicitar as informações perdidas e, principalmente, ter operação contínua em tempo real.

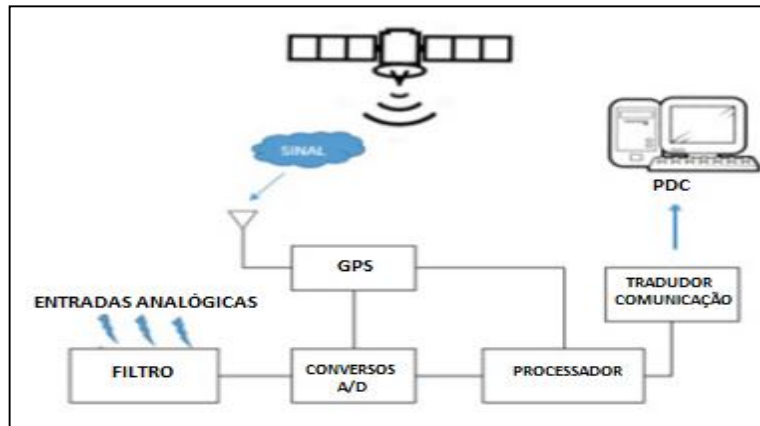


Figura 04: Estrutura Básica de uma PMU.

Este processo é realizado diariamente, gerando assim um fluxo de dados capaz de mapear o estado do sistema de forma bastante próxima à situação em tempo real, e que pode ser ainda armazenado em memória ou apresentado em monitores nos centros de operação. A Figura 5 apresenta a estrutura básica funcional de um PDC:

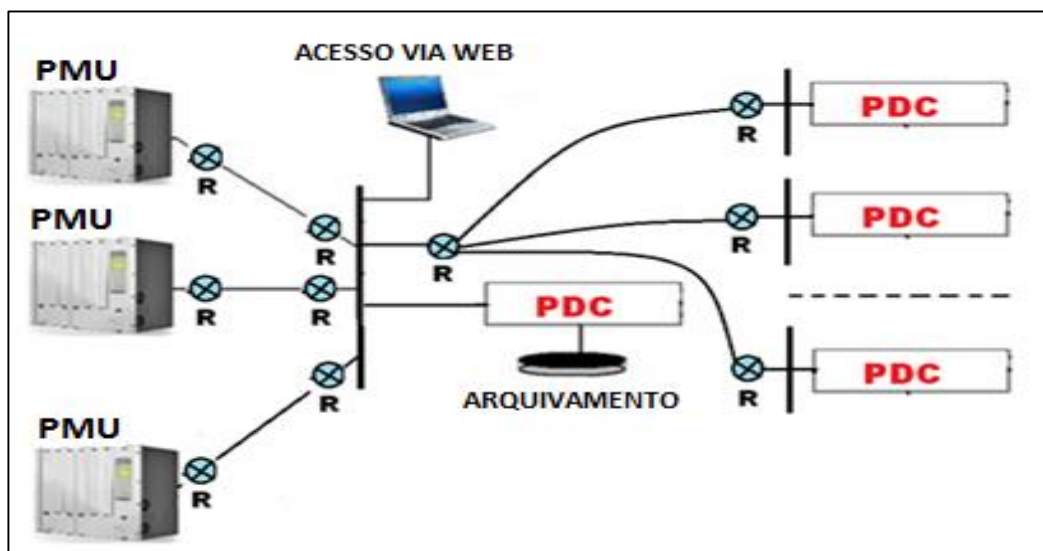


Figura 05: Estrutura básica funcional de um concentrador de dados fasoriais.

Estas funções do PDC podem ser executadas normalmente, usando computadores pessoais e, portanto, é responsabilidade do concentrador fazer o direcionamento dessas informações para os devidos terminais.

Para conseguir atender todos esses requisitos, o PDC deve estar preparado para trabalhar com um custo computacional alto. Tal requisito decorre da necessidade de continuidade da operação em tempo real, de eficiência para armazenamento de dados, de alta disponibilidade e confiabilidade, da capacidade de comunicação eficiente, de fácil integração e de atendimento a diversos tipos de aplicação. Através dessas características, o concentrador deve ainda verificar junto às PMUs a existência de perturbações no sistema elétrico e, caso alguma medição aponte para isto, os dados que serão recebidos na sequência devem ficar armazenados durante um período de tempo (normalmente 2 minutos). Assim, o PDC é um dos equipamentos mais complexos do SMFS. Este equipamento requer atenção prioritária e um alto investimento no desenvolvimento de aplicativos que atendam de forma adequada o SEP.

1.4.4. Transmissão de Dados em SMFS

Para a transmissão segura dos dados fasoriais, é necessário estabelecer uma rede de comunicação (canais de comunicação) que seja capaz de viabilizar a transferência de dados entre as PMUs e o PDC, assim como permitir a troca de dados entre os mais diversos PDCs que fazem parte do SEP em diferentes áreas. Os meios de comunicação mais utilizados no momento são: a internet e as redes privadas das próprias empresas.

Atualmente, os canais de comunicação considerados aptos para o funcionamento correto de um SMFS incluem opções de estruturas que são capazes de operar através de meios específicos, ou ainda, através de tecnologia sem fio (*wireless*). Ao se escolher este último meio implica utilização de canais de micro-ondas ou de sistemas de satélites.

Por outro lado, ao escolher um meio que utilize uma tecnologia com fio, diversas são as escolhas possíveis, entre essas se pode citar o uso das linhas telefônicas, cabo de fibra ótica ou mesmo a rede de baixa e média tensão, o que pressupõe o emprego da tecnologia PLC [*Power Line Communication*](Singh et al., 2011).

Um SMFS deve ser suportado por uma infraestrutura de comunicação com uma velocidade capaz de alinhar e agrupar os dados medidos pela PMU. Naturalmente, alguns sistemas de potência não apresentam a comunicação adequada. Contudo, deve-se levar em consideração os benefícios trazidos pelas PMU, que a partir de agora podem justificar a instalação de uma grande estrutura de comunicação.

Poucas pesquisas foram até hoje realizadas sobre a questão dos atrasos de comunicação envolvendo a transmissão dos dados dos sistemas de medição fasorial. No

entanto, sabe-se que a velocidade para o envio de informações está associada com o tipo de meio de comunicação escolhido. Pesquisas apontam para um padrão do IEEE 1344, mais recentemente atualizado para IEEE C37.118 que define fornece um formato para os dados transmitidos por uma PMU (IEEE, WorkingGroup H-8).

Os principais requisitos para os sistemas de comunicação são:

- a) Alta imunidade a ruídos;
- b) Expansibilidade;
- c) Envio de dados contínuos em tempo real;
- d) Segurança;
- e) Envio de dados perdidos;
- f) Padronização bem definida;
- g) Facilidade de Integração.

Como algumas aplicações já fazem uso desse tipo de sistema (controle e operação em tempo real, por exemplo) podem requerer altas taxas de atualizações de dados, é possível estabelecer uma estreita relação entre a aplicação escolhida e a eficiência do meio de comunicação. As vantagens e desvantagens do meio de comunicação devem ser levadas em consideração de modo a se escolher o meio mais adequado, ou seja, que compense requisitos técnicos e financeiros, determinados pelas características particulares de cada sistema de potência.

CAPÍTULO 2

MATERIAIS E MÉTODOS

2.1. Áreas de Estudo

Desde a instalação dos primeiros geradores e linhas de transmissão até os dias de atuais, os Sistemas Elétricos de Potência têm se tornado mais extensos e interligados, procurando atender uma demanda cada vez maior. A velocidade dessa demanda, aliada a situações como a desregulamentação do setor juntamente com a incorporação de novas tecnologias de equipamentos, vem aumentando e tornando complexo a operação dos sistemas elétricos. Diante desse cenário, surge então a necessidade de um aprimoramento de instrumentos e métodos dedicados a monitoração e controle em tempo real.

O principal objetivo dos sistemas de potência é gerar, transmitir e entregar energia a seus clientes atendendo a demanda requisitada, obedecendo a critérios como, por exemplo, qualidade, confiabilidade e economia, dentre outros (Baltazar, 2007). Para atender tal objetivo se faz necessário no setor elétrico à execução de diversos aprimoramentos que, normalmente, são caracterizados como de operação e expansão.

É neste contexto, conforme mencionado no capítulo introdutório, que este trabalho insere-se, buscando uma solução para um problema recorrente no setor elétrico. Como o estudo é experimental, foi preciso definir um escopo para aplicação da metodologia proposta. Assim, a área de estudo e investigação do presente trabalho compreende os dados obtidos da medição fasorial sincronizada correspondente às subestações de Tucuruí, Altamira e Rurópolis que estão interligadas por uma linha de transmissão de 230 kV. Este trecho do sistema interligado nacional é denominado Tramoeste de acordo com a ilustração abaixo:

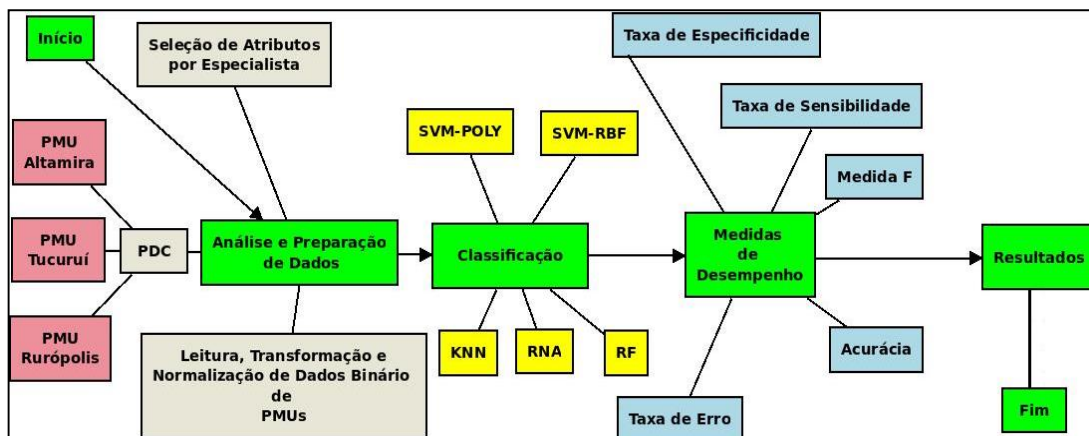


Figura 06: Esquema do sistema elétrico Tramoeste e posicionamento das PMUs ao longo desse sistema.

O sistema Tramoeste foi escolhido pelos seguintes motivos: (a) por se tratar de um circuito radial e sujeito a um crescimento de demanda significativa, apresentou um número de ocorrências de desligamentos maior que o normal no período em que os dados de medição fasorial foram coletados; (b) porque tal sistema, em 230 KV, contempla apenas quatro instalações supervisionadas: Tucuruí, Altamira, Transamazônica e Rurópolis; (c) e porque no período das medições já estava dotado de medição fasorial nas tensões de barra de cada subestação e correntes nas linhas de transmissão. Cada um dos três medidores disponíveis é mostrado na Figura 6 como PMU.

2.2. Metodologia Proposta

A metodologia proposta, para classificação de perturbações em sistemas elétricos de potência a partir de medições fasoriais sincronizadas utilizando técnicas de mineração de dados, pode ser visualizada na Figura 7. Como podemos observar na figura, a metodologia envolve várias etapas, desde a análise e preparação de dados até aplicação dos algoritmos de classificação adotado neste trabalho e suas avaliações de desempenho. Todas as etapas da metodologia proposta serão detalhadas nas próximas seções.

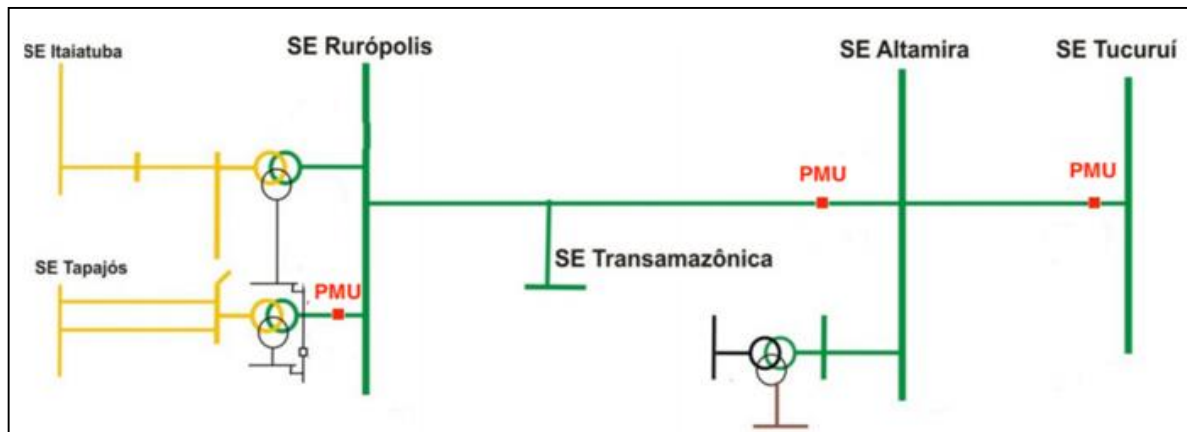


Figura 07: Esquema da metodologia proposta para classificação de perturbações em sistemas elétricos de potência a partir de medições fasoriais sincronizadas utilizando técnicas de mineração de dados

2.3 Análise e Preparação da Base de Dados

A base de dados utilizada no estudo foi fornecida pela Eletrobrás-Eletronorte com informações fasoriais contidas no concentrador de dados das PMUs do sistema Tramoeste. Com intuito de ter uma base com informações que permitam a aplicação da metodologia

proposta por este trabalho, foram selecionados os registros fasoriais do dia 25/12/2014, já que este período apresentou uma grande quantidade de ocorrências, inclusive desligamentos.

Como a taxa de amostragem das PMUs do sistema Tramoeste é de 60 amostras por segundo, um dia de medição corresponde aproximadamente 5 milhões de registros, isto é, um volume de dados muito elevado para aplicação dos algoritmos de classificação. Este fato poderia inviabilizar a metodologia proposta, por isso, foram selecionadas amostras apenas dentro de janelas de intervalos de tempo em torno das ocorrências de desligamento de interesse, selecionadas previamente dos arquivos de log do configurador de estado.

O configurador de estados, parte integrante do sistema de análise de redes (SAR) do SAGE (*SupervisoryControlAnd Data Acquisition*), que abrange desde a instalação de distribuição até um centro de operações, produzindo um arquivo diário, onde são registradas todas as alterações topológicas que ocorrem no sistema elétrico. A composição do nome do arquivo de configuração permite identificar o dia em que foram realizadas as alterações. A Figura 8 mostra, por exemplo, o extrato de um arquivo do configurador de estado. A linha em vermelho é de interesse particular para o estudo porque ela indica um desligamento de uma linha de transmissão. Ela se caracteriza pelo texto “LTR” na terceira coluna. A coluna 5 informa qual linha sofreu desligamento. Este nome, ou melhor, mnemônico, segue um padrão conhecido pelos especialistas da Eletrobras-Eletronorte. Por exemplo, a coluna 5 será lida como: “linha de transmissão número 5 (-05), que interliga a Usina Hidrelétrica de Tucuruí (TU) à subestação de 500 KV (7)”.

10:40:10.230	Atuado	IGQA1AUT	IGQSACA	Automatismo Serv Auxiliar
10:40:10.580	Desatuado	IGQA1AUT	IGQSACA	Automatismo Serv Auxiliar
10:40:10.390	Atuado	IGQA1AUT	IGQSACA	Automatismo Serv Auxiliar
10:40:10.700	Desatuado	IGQA1AUT	IGQSACA	Automatismo Serv Auxiliar
10:40:43.545	Alarme	RUUPD1AX_E_4112_4624	ATRU-LT6-01	TRANSMISSAO BLOQUEIO ALTAMIRA
10:40:43.547	Desligamento	RUUPD1BY_E_40112_44208	TARU-LI5-01	21 DIST ZONA TERRA FASE V
10:40:43.562	Partida	TMUOD1AX_EVT_12300	TMTF6-01/02	SOBRECORRENTE TEMP NEUT ENR 1
10:40:43.574	Partida	ATUPD2DY_EVT_12300	ATT6-01/04	SOBRECORRENTE TEMP NEUT ENR 1
10:10:45.775	Alarme	ATUPD1AX_E_4112_4624	TCAT-LT6-01	TRANSMISSAO BLOQUEIO TUCURUI
10:40:45.925	Normal	ATUPD1AX_E_4112_4624	TCAT-LT6-01	TRANSMISSAO BLOQUEIO TUCURUI

Figura 08: Exemplo de um log do configurador de estado do SAGE.

Com base nas informações da própria empresa e dos dados do configurador, foram identificadas alterações na rede elétrica no dia 25/12/2014 às 09:37:23, com indicação de desligamento da linha de transmissão do trecho compreendido entre Altamira e Rurópolis. Assim, como prova de conceito do estudo, foram coletadas as amostras em torno dessa

ocorrência considerando uma janela de tempo total de 6 minutos (das 9:33:00 até 9:39:00), o que representa aproximadamente um total de 21.600 registros.

Para realizar a leitura e coleta dos dados, foram desenvolvidos dois scripts, já que os dados extraídos dos equipamentos (PMUs) são em formato binário: (a) BinaryReader: tem como objetivo realizar a leitura do arquivos binários provenientes das PMU (Rurópolis, Altamira e Tucuruí); (b) CsvFormat: converte o arquivo em formato .csv, exportado pelo BinaryReader, em um arquivo no formato .arff do WEKA.

2.3.1. Seleção das variáveis de interesse

A Tabela seguinte mostra os tipos de medidas fornecidas por um modelo típico de PMU em um dado instante de tempo, considerando um circuito trifásico (A, B e C) mais o terra:

TIPO DE MEDIDA	NÚMERO DE MEDIÇÕES
Magnitude de corrente sequencial	3
Ângulo de fase de corrente sequencial	3
Magnitude de corrente de fase	3
Ângulo de fase da corrente de fase	3
Magnitude de tensão sequencial	3
Ângulo de fase de tensão sequencial	3
Magnitude de tensão de fase	3
Ângulo de tensão de fase	3
Magnitude de corrente terra	1
Ângulo de fase de corrente terra	1
Frequência	1
Frequência Delta	1

Tabela 1: Medições típicas realizadas em uma PMU.

As PMUs utilizadas no sistema Tramoeste medem um total de 112 variáveis (vide Apêndice A). Em sistemas de elevada tensão, como é caso do Tramoeste, as três fases são equilibradas, isto é, os valores das magnitudes de suas tensões medidos em RMS (*Root Mean Square*) são praticamente iguais, e as defasagens entre os fasores se mantêm próximas a 120 graus. Logo, foram utilizados apenas os dados da fase B no estudo.

Além disso, foram selecionadas apenas as variáveis consideradas mais relevantes pelos especialistas da Eletronorte e que pudessem refletir algum tipo de padrão para predição

de perturbações em sistemas elétricos de potência. Neste sentido, a Tabela 2 apresenta uma descrição sucinta das variáveis de interesse selecionadas, perfazendo um total de 8 variáveis:

RMS_Tens_FN_B_132kV_B230_Rur	Valor RMS tensão fase-neutro fase B (132 kV) na barra de 230
Ângulo_Cor_Tuc_Rur	Ângulo em graus da corrente da linha Tucurui-Rurópolis.
Ângulo_Tens_B_230kV_Rur	Ângulo em graus da tensão da fase B na barra de 230 kV de Rurópolis.
Corrente_B_Alt_Rur	Corrente da fase B, em amperes, da linha Altamira-Rurópolis
RMS_Tens_FF_BV_230kv_B230_Alt	Valor RMS Tensão fase-fase Fase BV (230kV) na barra de 230 em Altamira
Ângulo_Tens_B_230_Alt	Ângulo em graus de tensão da fase B na barra de 230 kV de Altamira
Ângulo_Cor_Tuc_Alt	Ângulo em graus da corrente da linha Tucurui-Altamira.
Corrente_B_Tuc_Alt	Corrente da fase B, em amperes, da linha Tucurui-Altamira.

Tabela 2: Variáveis selecionadas conforme indicado por especialistas do setor elétrico.

As diferentes variáveis selecionadas possuem eventualmente distintas faixas de valores. Os algoritmos de classificação podem ser completamente ofuscados por parâmetros (variáveis) que possuem grandes faixas dinâmicas quando comparados às faixas de outros. Por exemplo, no caso, se lida com valores de tensões em kilovolts e correntes em Ampères. Diante disso, tipicamente aplica-se um processo de normalização nos dados de entrada para que, as formas de onda de tensão e corrente fiquem nas mesmas faixas de valor, neste caso no intervalo [0, 1].

2.3.2. Processo de Rotulação da base de dados

Nesta etapa, foi necessário também aplicar um processo de rotulação dos dados provenientes das PMUs, considerando três estados possíveis: normal, pré-perturbação e perturbação. Este processo contou com a ajuda de um especialista e foi realizado de forma empírica. Mais especificamente, foi realizada uma análise gráfica de uma mostra das formas de onda de tensão e corrente obtidas pelas PMUs, tomando como base uma janela de

aproximadamente 2 minutos e meio antes depois da ocorrência registrada às 09:37:23. Nesta análise foram identificados claramente os três estados conforme ilustra a Figura 9:

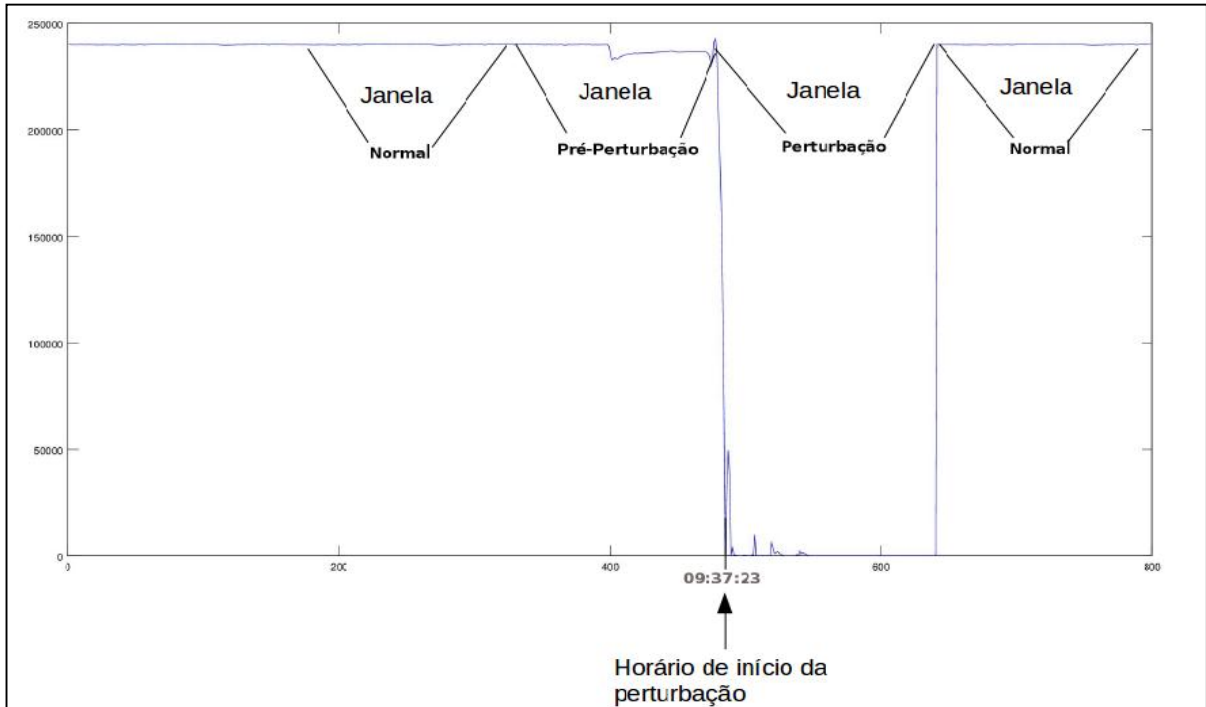


Figura 09: RMS da forma de onda de tensão da fase B no momento da ocorrência de uma perturbação.

De acordo com a figura acima, observa-se que antes da ocorrência, é possível verificar uma oscilação no valor de tensão, característica padrão em todas as perturbações analisadas no estudo que provocam ou não desligamento. Essa característica também foi vista nas outras variáveis selecionadas.

Na tentativa de balancear o número de amostras entre as três classes (Normal, pré-perturbação e perturbação) foi fixada uma janela de 160 amostras (aproximadamente 2,5 minutos) totalizando 480 registros. Estes foram então utilizados como entrada para os classificadores.

2.4. Classificação

Na etapa de classificação foram utilizados os algoritmos apresentados na Seção 2.3. O modelo de classificação de perturbações adotado é apresentado na seguinte ilustração:

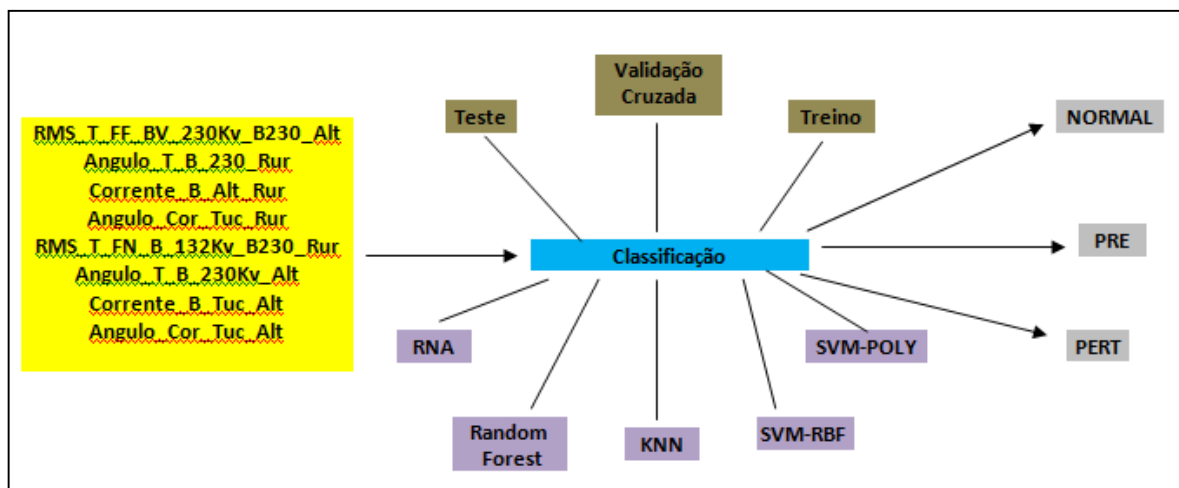


Figura 10: Modelo de Classificação adotado.

2.4.1. Seleção Automática do Modelo

Frequentemente o melhor desempenho de um classificador em relação a uma base de dados específica só pode ser alcançado ajustando-se exaustivamente os parâmetros do algoritmo. Esta tarefa é chamada de seleção automática do modelo e corresponde, por exemplo, a escolher os parâmetros tais como o número de neurônios na camada escondida de uma RNA.

Uma estratégia popular para seleção automática do modelo e que foi adotada neste trabalho é a validação-cruzada (*cross-validation*) considerando 10-*folds* implementada na classe CVParameterSelection do WEKA (Witten; Frank 2005).

2.4.2. Validação Cruzada

A validação dos modelos de classificação deste trabalho foi feito através da validação cruzada (*cross-validation*) com a qual os dados são fragmentados em dois subconjuntos, denominados conjunto de treinamento e conjunto de teste.

De modo geral a técnica de validação cruzada divide a base de dados em r partes (*r-foldcross-validation*). Destas, $r-1$ são utilizadas para o treinamento e uma serve como base de teste. O processo é repetido r vezes, de forma que cada parte é utilizada uma vez como conjunto de teste. Ao final, a correção total é calculada pela média aritmética e desvio padrão dos resultados obtidos em cada etapa, obtendo-se assim uma estimativa da qualidade do modelo de conhecimento gerado e permitindo análises estatísticas(Santos, 2009). Tal processo é verificável a seguir:

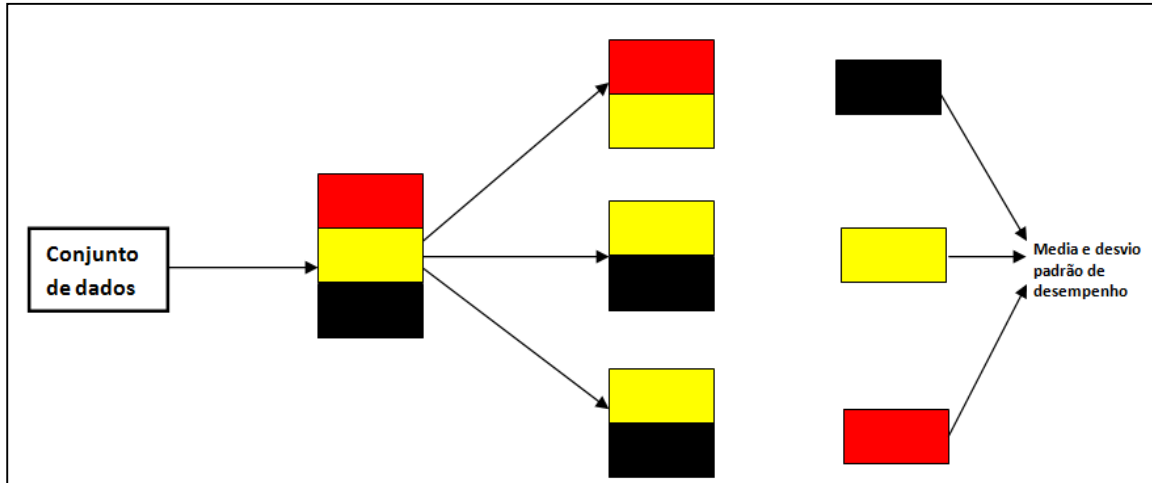


Figura 11: Exemplo do processo de validação cruzada. (Fonte: Faceli, 2011).

2.4.3. Medidas de Desempenho

Alguns classificadores são capazes de prover *scores* de confiança $f_i(\mathbf{z})$ 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 $(f_1(\mathbf{z}), \dots, f_Y(\mathbf{z}))$ com Y *scores*. Se o classificador naturalmente não retornar *scores* de confiança, o vetor de *scores* é criado com um *score* unitário $f_j(\mathbf{z}) = 1$ para a classe j sugerida pelo classificador, enquanto que os *scores* das outras classes são zero: $f_i(\mathbf{z}) = 0, i \neq j$. Assim, a decisão final é sempre baseada no valor máximo dos *scores* (chamada regra Max-Wins), vide equação abaixo:

$$F(\mathbf{z}) = \arg \max_{i=1,\dots,Y} f_i(\mathbf{z}).$$

Equação 4

Um *conjunto de teste* $\{(z_1, y_1), \dots, (z_R, y_R)\}$ contendo R exemplos e disjunto do conjunto de treino pode ser usado para calcular a taxa de erro de classificação (Equação abaixo):

$$E_f = \frac{1}{R} \sum_{r=1}^R x(F(z_r) \neq y_r)$$

Equação 5

Onde I é a função indicadora, 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 (Witten; Frank, 2005).

A taxa de erro (E_f) varia entre 0 e 100, e valores próximos a 0 são considerados os melhores índices nas previsões dos classificadores. Em Faceli (2011) afirma-se que a acurácia (taxa de acerto $-a_c$), é o complemento da taxa de erro (vide equação):

$$a_c = 1 - E_f$$

Equação 6

A taxa de acerto também varia entre 0 a 1, entretanto ao contrário da taxa de erro, valores próximos a 1 apresentam melhores previsões nas classificações. Um ponto relevante dentro desse contexto é que tanto a taxa de erro ou acerto, entre outras medidas de desempenho de um classificador podem ser obtidas a partir da matriz de confusão, cuja dimensão corresponde ao número de classes existentes em um determinado conjunto de exemplos. Na diagonal principal estão a quantidade de acertos em cada classe e os elementos fora desta, correspondem a quantidade de erros.

Na matriz de confusão (M_c), as linhas representam as classes verdadeiras, e as colunas, as classes preditas pelo classificador. Logo, cada elemento m_{ij} de uma M_c apresenta o número de exemplos da classe i classificados como pertencentes a classe j . Para Y classes, M_c tem a dimensão $Y \times Y$ (Faceli, 2011). A Tabela 3 apresenta um exemplo de matriz de confusão para um problema de duas classes:

	Classe	Predita
	+	-
+	VP	FN
-	FP	VN

Tabela 3: Matriz de confusão para um problema de duas classes (positivo e negativo)

Onde:

- a) VP: Verdadeiros positivos equivalem a todos os exemplos classificados corretamente;

- b) Verdadeiros negativos equivalem a todos os exemplos classificados como verdadeiros incorretamente;
- c) FP: Falsos positivos equivalem a todos os exemplos classificados corretamente como negativos;
- d) FN: Falsos negativos equivalem a todos os exemplos classificados incorretamente como negativos.

Da mesma forma que se podem extrair as taxas de erro e acerto a partir da matriz de confusão, outras medidas também podem ser calculadas. A taxa de sensibilidade T_s (Equação 7) é a proporção de verdadeiro-positivos corretamente identificados e a taxa de especificidade T_e (Equação 8) é a quantidade de verdadeiros negativos corretamente identificados. Já a medida F (*F-measure*) é definida na Equação 9 e é uma medida harmônica obtida através das taxas de precisão e sensibilidade:

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

Equação 7

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

Equação 8

$$F = \frac{2T_e T_s}{T_e + T_s}$$

Equação 9

Para validar a metodologia apresentada, o próximo capítulo deste trabalho evidenciará os resultados obtidos no processo de classificação de perturbações.

CAPÍTULO 3

RESULTADOS E DISCUSSÕES

3.1. Introdução

Neste capítulo serão descritos os experimentos realizados e os resultados obtidos em relação à aplicação da metodologia proposta para classificação de perturbações a partir de medições fasoriais sincronizadas. Para isso, foi utilizada a base de dados descrita no capítulo anterior, a qual possui um total de 480 exemplos, estando devidamente balanceada entre as três classes adotadas: normal, pré-perturbação e perturbação. Realizado todo o processo de pré-processamento da base, os dados foram convertidos para formato (ARFF) padrão do WEKA, possibilitando o uso dos algoritmos de classificação deste software.

Para encontrar o melhor modelo de cada classificador, foi utilizado um procedimento automático de seleção implementado no WEKA pela classe *CVParameterSelection*. Assim, foi composto um *grid* (Tabela 4) de variação dos principais parâmetros de cada classificador. A escolha deste *grid* foi baseada em experimentos anteriores simulados em laboratório.

Classificador	Parâmetros	Valores do Grid	#Pontos do Grid
RNA	H	5, 11, 17, 43, 75, 107 e 139	63
	L	0.1, 0.3 e 0.5	
	M	0.1, 0.3 e 0.5	
RF	I	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, e 30	12
SVM-RBF	G	0.1, 1, 10, e 100	16
	C	0.1, 1, 10, e 100	
KNN	K	1, 3, 5, 7, 9, 11, 13 e 15	8

Tabela 4: Grid de parâmetros adotado no procedimento de seleção automática do modelo.

A ideia nesse processo de seleção do modelo é obter os melhores valores para os parâmetros dos classificadores. Neste sentido, as próximas seções descrevem para cada classificador os resultados deste processo baseado nas métricas de desempenho: taxa de acerto (A_c), taxa de erro (E_f), a taxa de sensibilidade (T_s), taxa de especificidade (T_e) e a medida F .

3.2. Resultado para a Rede Neural

Após vários testes a arquitetura mais robusta para RNA foi com apenas uma camada escondida. O número de neurônios nessa única camada escondida H , a taxa de aprendizado L e taxa de *momentum* M variaram de acordo com o *grid* da Tabela 4, sendo que o número de iterações N foi fixado em 1000. Como podemos observar na tabela 4, a combinação dos valores desses parâmetros gera 63 pontos (possibilidades) no *grid*. A Tabela 4 apresenta uma listagem de todas as combinações dos parâmetros da RNA, onde os valores entre parenteses representam H , L e M , respectivamente.

Na figura seguinte a taxa de erro E_f obtida a partir da execução do processo de seleção do modelo do classificador neural, considerando a combinação dos valores (*grid*) apresentada na tabela 4. O eixo x representa os pontos do *grid*. Por exemplo, o valor 3 no eixo x corresponde ao modelo (5,0.1,0.5), isto é, os valores dos parâmetros H , L e M neste modelo neural são, respectivamente, 5, 0.1 e 0.5:

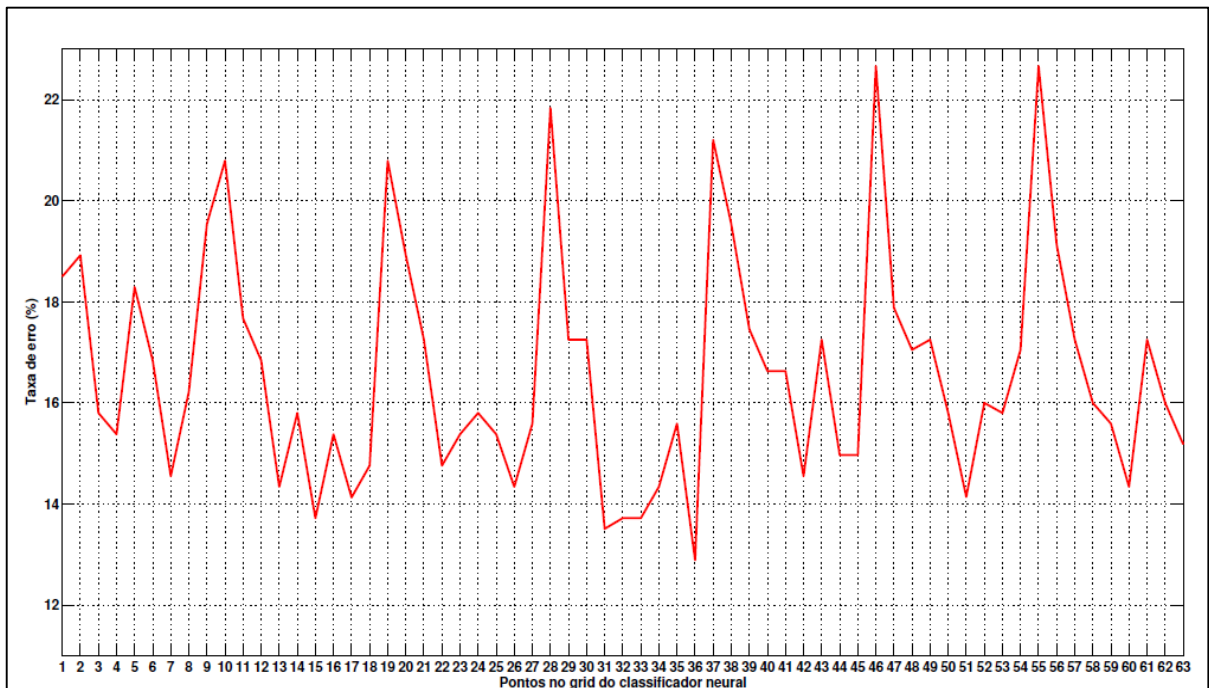


Figura 12: Gráfico da evolução da taxa de erro do classificador neural considerando a combinação dos parâmetros da tabela 4.

Ao observarmos a figura acima, vemos que o melhor modelo da rede neural obtido (taxa de erro $E_f = 12.89\%$) foi com o ponto do *grid* 36, onde o número de neurônios na camada escondida é 43, taxa de aprendizado é 0.5 e de *momentum* 0.5. A Tabela 5 apresenta

os resultados obtidos pela rede neural utilizando outras métricas desempenho, considerando esses parâmetros:

Id	Combinações no grid (H, L e M)	Id	Combinações no grid (H, L e M)
1	(5, 0.1, 0.1)	33	(43, 0.3, 0.5)
2	(5, 0.1, 0.3)	34	(43, 0.5, 0.1)
3	(5, 0.1, 0.5)	35	(43, 0.5, 0.3)
4	(5, 0.3, 0.1)	36	(43, 0.5, 0.5)
5	(5, 0.3, 0.3)	37	(75, 0.1, 0.1)
6	(5, 0.3, 0.5)	38	(75, 0.1, 0.3)
7	(5, 0.5, 0.1)	39	(75, 0.1, 0.5)
8	(5, 0.5, 0.3)	40	(75, 0.3, 0.1)
9	(5, 0.5, 0.5)	41	(75, 0.3, 0.3)
10	(11, 0.1, 0.1)	42	(75, 0.3, 0.5)
11	(11, 0.1, 0.3)	43	(75, 0.5, 0.1)
12	(11, 0.1, 0.5)	44	(75, 0.5, 0.3)
13	(11, 0.3, 0.1)	45	(75, 0.5, 0.5)
14	(11, 0.3, 0.3)	46	(107, 0.1, 0.1)
15	(11, 0.3, 0.5)	47	(107, 0.1, 0.3)
16	(11, 0.5, 0.1)	48	(107, 0.1, 0.5)
17	(11, 0.5, 0.3)	49	(107, 0.3, 0.1)
18	(11, 0.5, 0.5)	50	(107, 0.3, 0.3)
19	(17, 0.5, 0.3)	51	(107, 0.3, 0.5)
20	(17, 0.1, 0.3)	52	(107, 0.5, 0.1)
21	(17, 0.1, 0.5)	53	(107, 0.5, 0.3)
22	(17, 0.3, 0.1)	54	(107, 0.5, 0.5)
23	(17, 0.3, 0.3)	55	(139, 0.1, 0.1)
24	(17, 0.3, 0.5)	56	(139, 0.1, 0.3)
25	(17, 0.5, 0.5)	57	(139, 0.1, 0.5)
26	(17, 0.5, 0.3)	58	(139, 0.3, 0.1)
27	(17, 0.5, 0.5)	59	(139, 0.3, 0.3)
28	(43, 0.1, 0.1)	60	(139, 0.1, 0.1)
29	(43, 0.1, 0.3)	61	(139, 0.5, 0.1)
30	(43, 0.1, 0.5)	62	(139, 0.5, 0.3)
31	(43, 0.3, 0.1)	63	(139, 0.5, 0.5)
32	(43, 0.3, 0.3)		

Tabela 5: Listagem da combinação dos parâmetros de uma rede neural. Valores entre parênteses representam *H*, *L* e *M*, respectivamente.

Classe	T_s	T_e	F
Normal	0.37	0.15	0.44
Pré-perturbação	0.60	0.31	0.54
Perturbação	0.88	0.11	0.84

Tabela 6: Taxa de sensibilidade T_s , Taxa de especificidade T_e e Medida F obtida pelo classificador neural, considerando o melhor modelo obtido, isto é, $H =$, $L =$ e $M =$.

3.3 Resultados para o Random Forest

Para o classificador Random Forest a seleção de modelo levou em consideração o principal parâmetro deste classificador: o número de árvores a ser gerada I . A Figura 4.2 apresenta a taxa de erro de classificação do randomforest variando-se o parâmetro I , eixo x, conforme os valores definidos no *grid* da Tabela 6.

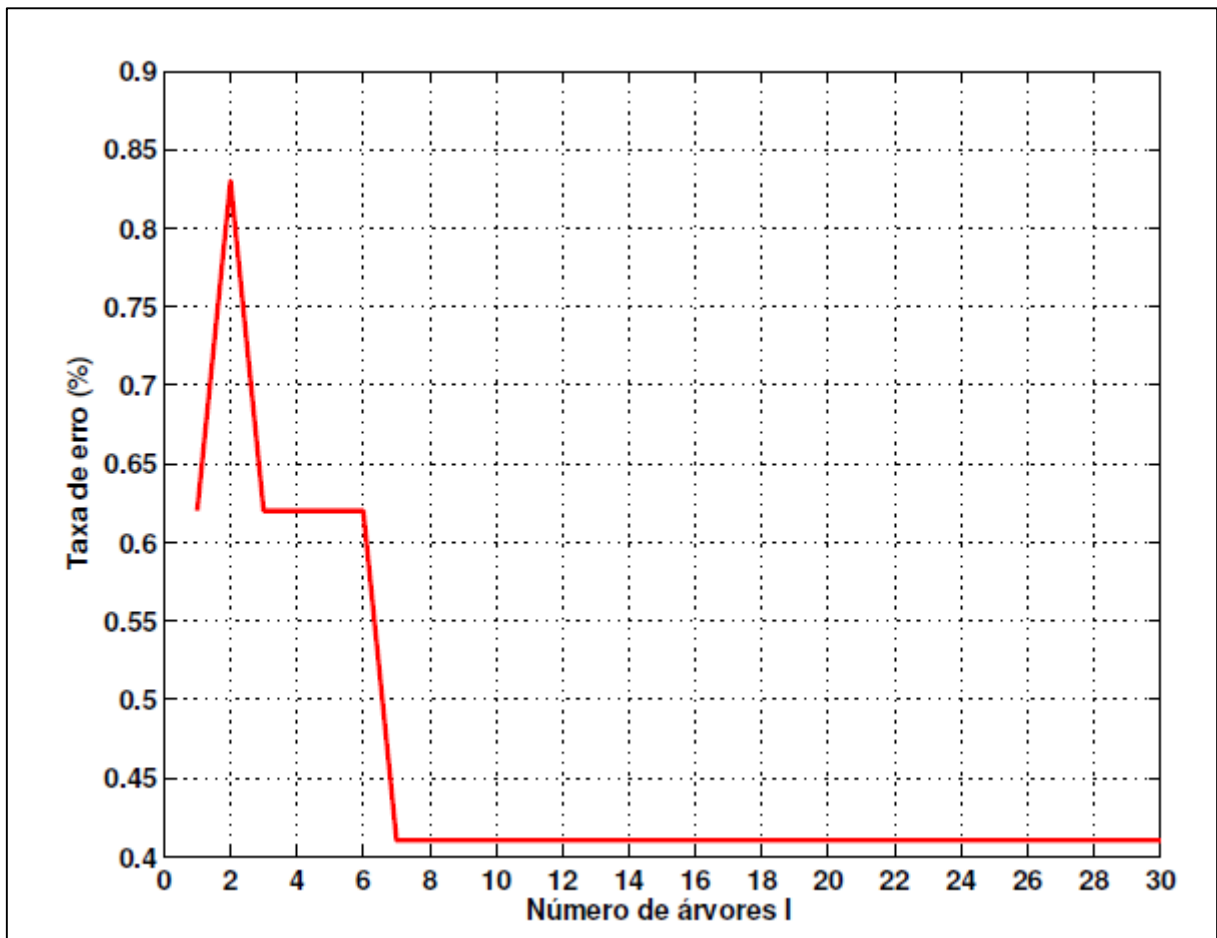


Figura 13: Gráfico da evolução da taxa de erro do Random Forest considerando o número de árvores.

Analisando a Figura 13 observa-se claramente que a partir de 7 árvores o algoritmo Random Forest converge para uma taxa de erro de 0.41%. A Tabela 7 apresenta os resultados do classificador randomForest considerando as outras medidas de desempenho, o número de árvores igual a 10 e as três classes adotadas na metodologia.

Classe	T_s	T_e	F
Normal	0.988	0	0.994
Pré-perturbação	1	0.01	0.99
Perturbação	1	0	0.1

Tabela 7: Taxa de sensibilidade T_s , Taxa de especificidade T_e e Medida F obtida pelo classificador randomforest, considerando 10 árvores.

3.4 Resultados para o KNN

Para o classificador K-vizinho mais próximo (KNN), o principal parâmetro avaliado foi o número de vizinhos mais próximos. Este parâmetro variou de acordo com o grid da Tabela 16 e os resultados obtidos podem ser visualizados no gráfico:

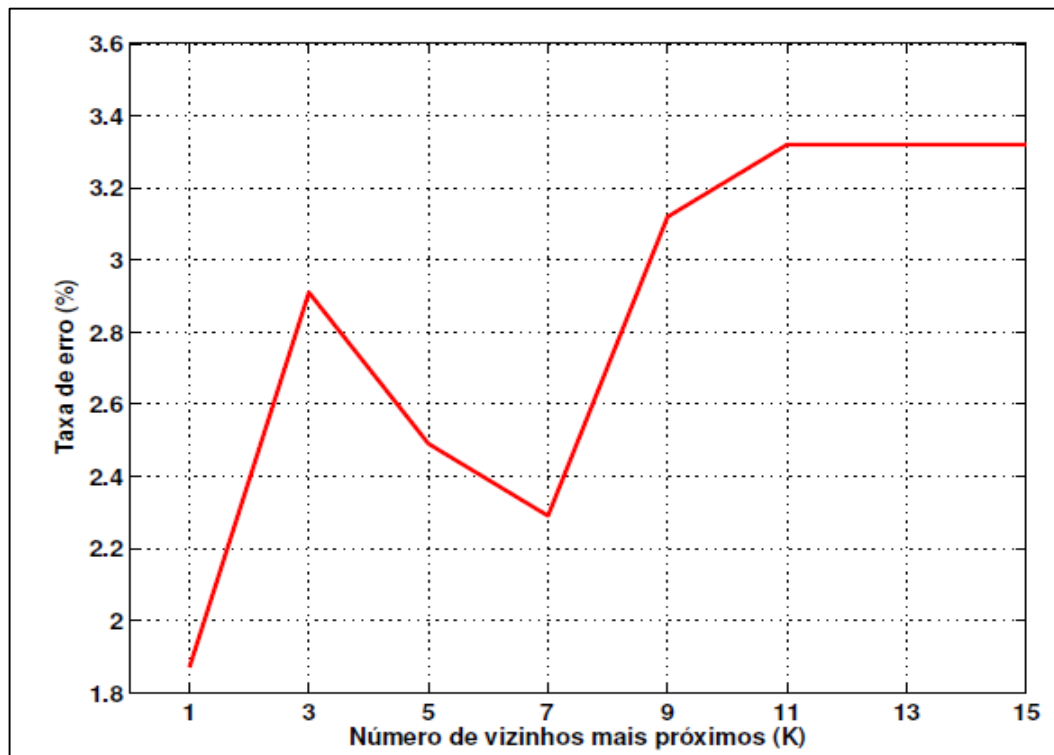


Figura 14: Gráfico da evolução da taxa de erro do KNN de acordo com o número de vizinhos mais próximos definidos no *grid*.

Como podemos observar na figura acima, o melhor resultado obtido para este classificador é com 1 vizinho mais próximo ($E_f=1.87\%$). É possível visualizar que o algoritmo tende a convergir a partir de 11 vizinhos. A Tabela 8 mostra o desempenho do classificador utilizando as outras medidas de desempenho.

Classe	T_s	T_e	F
Normal	0.97	0.02	0.97
Pré-perturbação	0.97	0.01	0.98
Perturbação	0.99	0	0.99

Tabela 8: Taxa de sensibilidade, Taxa de especificidade T_e e Medida F obtida pelo classificador KNN com 1 vizinho mais próximo.

3.5. Resultados para o Classificador SVM

Para o classificador Máquina de Vetores de Suporte (SVM), os parâmetros otimizados foram γ e penalidade do erro C . Tais parâmetros variaram de acordo com *grid* da Tabela 18. Vale ressaltar que a SVM adotada neste trabalho utiliza as funções de base radial (RBF).

Conforme podemos observar na Tabela 18, o número de pontos no *grid* é igual 16. Assim, a Tabela 9 apresenta uma listagem de todas as combinações dos parâmetros do classificador SVM, onde os valores entre parênteses representam C e γ , respectivamente:

Id	Combinação no grid	Id	Combinação no grid
1	(0.1, 0.1)	9	(10, 0.1)
2	(0.1, 1)	10	(10, 1)
3	(0.1, 10)	11	(10, 10)
4	(0.1, 100)	12	(10, 100)
5	(1, 0.1)	13	(100, 0.1)
6	(1, 1)	14	(100, 1)
7	(1, 10)	15	(100, 10)
8	(1, 100)	16	(100, 100)

Tabela 9: Listagem da combinação dos parâmetros dos classificadores SVM. Valores entre parênteses representam C e γ , respectivamente.

A próxima figura mostra a taxa de erro E_f obtida a partir da execução do processo de seleção do modelo para o classificador SVM adotado, considerando a combinação dos valores *grid* apresentada na Tabela 18. Nota-se que o eixo x representa os pontos do *grid*. Por exemplo, o valor 2 no eixo x corresponde ao modelo (0.1, 1), isto é, os valores dos parâmetros C e G são 0.1, e 1, respectivamente:

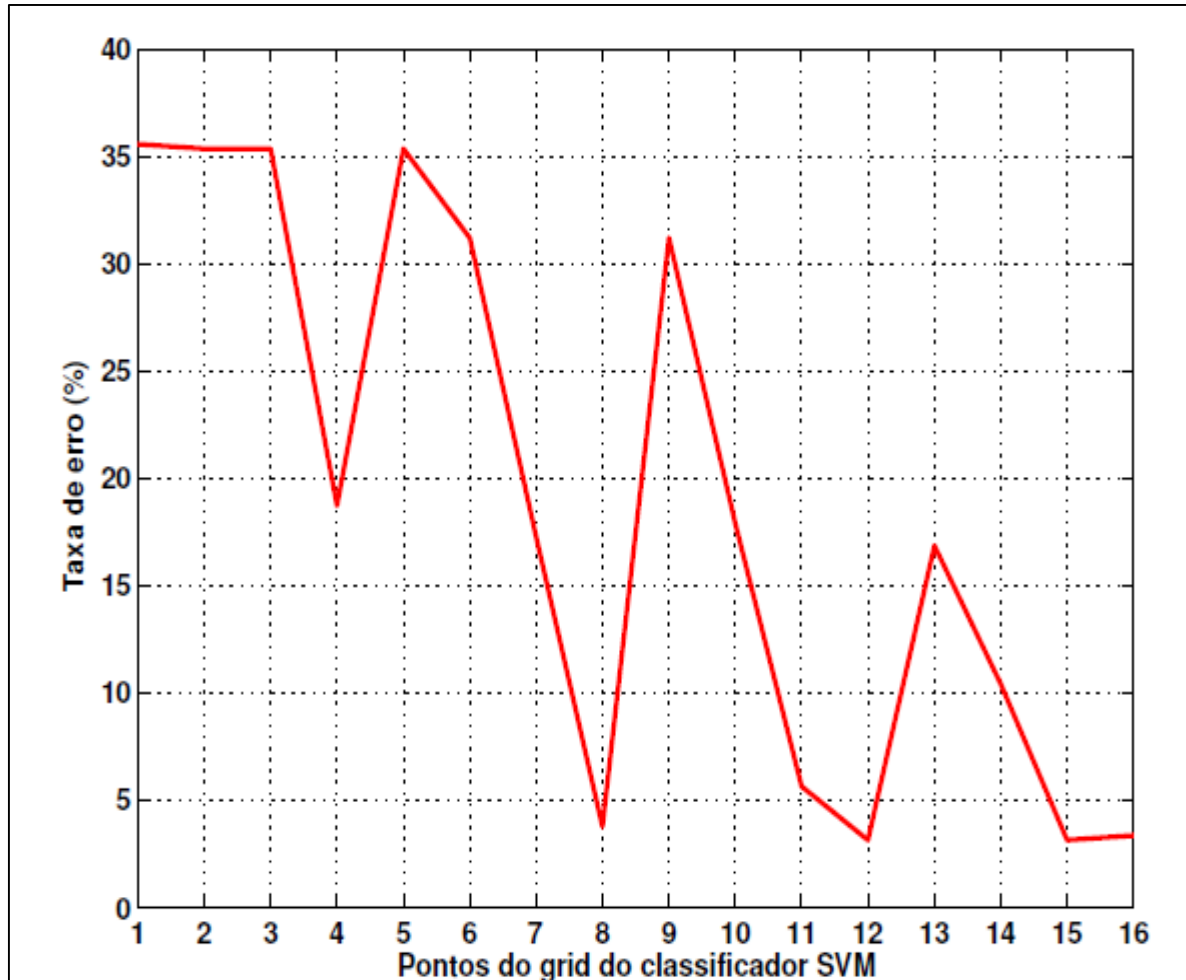


Figura 15: Gráfico da evolução da taxa de erro do classificador SVM considerando a combinação dos parâmetros da Tabela 9.

Analisando o gráfico acima podemos observar que os melhores resultados (taxa de erro $E_f = 3.11\%$), para o classificador SVM, foram considerando os parâmetros $C=10$ e $G=100$. A Tabela a seguir apresenta os resultados obtidos pela SVM utilizando outras métricas de desempenho, considerando esses melhores parâmetros:

Classe	T_s	T_e	F
Normal	0.	0	0
Pré-perturbação	0.96	0.51	0.64
Perturbação	0.98	0.03	0.96

Tabela 10: Taxa de sensibilidade, Taxa de especificidade T_e e Medida F obtida pelo classificador SVM, considerando o melhor modelo obtido, isto é, $C=10$ e $G=100$.

3.6. Comparação Entre Os Classificadores

Para efeito de comparação entre os classificadores, a Figura 4.5 mostra as taxas de erro destes considerando seus melhores modelos. Analisando a figura podemos observar que os classificadores Randomforest, KNN e SVM, nesta ordem, foram os que apresentaram os melhores resultados. O Classificador neural, apesar do desempenho inferior aos demais, apresentou uma taxa de erro considerada satisfatória dada a complexidade do problema. No caso específico deste classificador seria interessante realizar uma investigação mais aprofundada, já que existe um grau de liberdade elevado no que se refere à definição dos valores dos seus parâmetros.

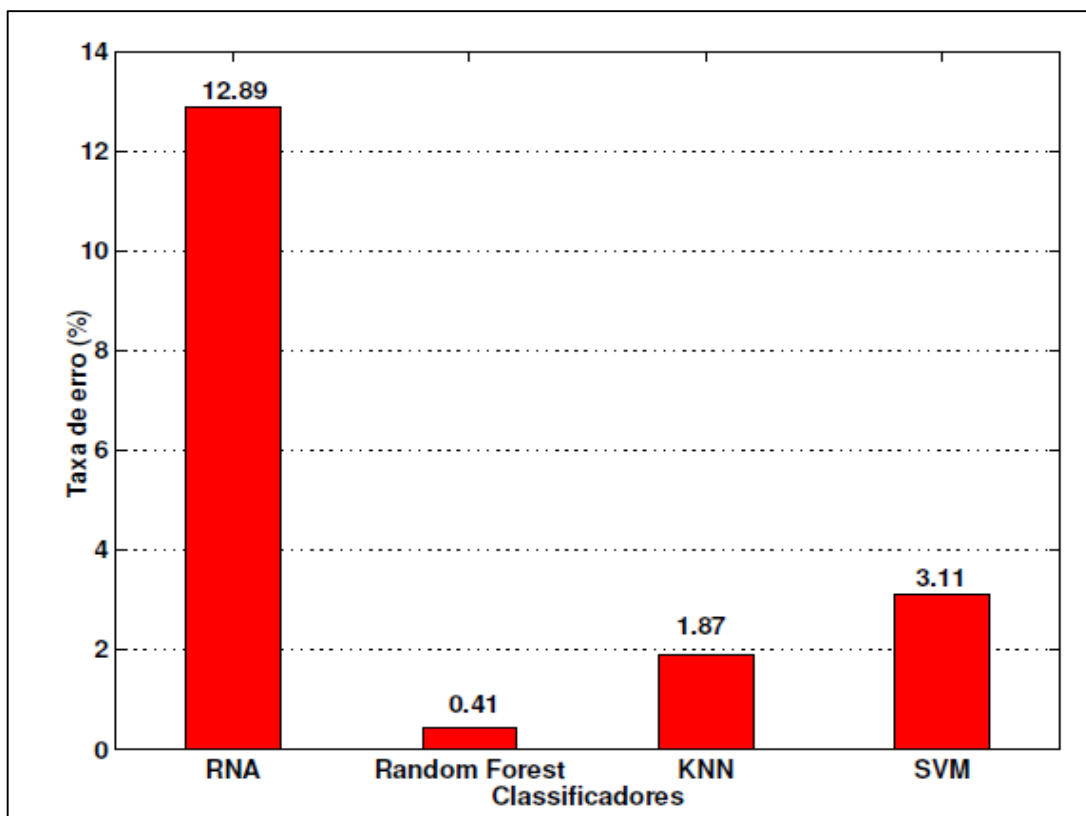


Figura 16: Gráfico que compara as taxas de erro (E_f) dos classificadores adotados no trabalho.

CONCLUSÃO

A monitoração nos sistemas elétricos de potência tem como objetivo garantir uma maior segurança e qualidade no fornecimento de energia elétrica. Através do monitoramento e o armazenamento dos dados, pode-se obter informação de interesse para uma melhor avaliação das causas, ocorrências, variações e tendências que possam ser relevantes ao processo de tomada de decisão. Um exemplo desse direcionamento seria o acionamento de algum dispositivo de proteção para balanceamento de carga diante de um possível corte no fornecimento de energia. Sendo assim, medidas preventivas, avaliação de dispositivos de proteção, identificação e classificação de perturbações elétricas podem auxiliar de forma mais objetiva e contribuir para um melhor entendimento do comportamento do sistema.

Diante desse cenário, esta pesquisa adotou uma metodologia que procura identificar e classificar perturbações em sistemas elétricos de potência. A metodologia utilizada emprega técnicas e algoritmos de mineração de dados baseados em valores de tensão, corrente e diferença de ângulo entre as tensões extraídas diretamente das PMUS de Altamira, Rurópolis e Tucuruí que fazem parte do Tramoeste no Estado do Pará.

Um ponto relevante na classificação de perturbações em sistemas elétricos de potência é que a maioria das bases de dados utilizada para este fim trabalha com dados simulados. Nessa direção, aproveitar dados reais dentro desse cenário de descoberta do conhecimento com a extração de informações contidas em PMUs diferencia o presente estudo em relação a outras abordagens.

O processo de construção da base de dados utilizada nesta pesquisa partiu do princípio de que a utilização de dados reais tem um caráter mais significativo, em relação a utilização de dados simulados. Então, em parceria e em comum acordo com a Eletrobras-Eletronorte, foi possível fazer um estudo em conjunto com especialista da empresa para extração dos dados reais que estavam disponíveis nas PMUs das subestações que fazem parte Tramoeste. Durante esse período foram obtido informações do dia 25/12/2015 com três ocorrências de perturbações elétricas, no qual duas destas provocaram desligamento no fornecimento de energia. Formando assim, uma base de dados para o emprego na Mineração de Dados.

Os resultados encontrados após o processo de classificação em perturbações elétricas, fazendo uso de dados reais extraídos de PMUs, mostraram perspectivas animadoras para construção de artefatos tecnológicos, como modelos que utilizam técnicas de mineração de dados que favoreçam a gestão operacional em sistemas elétricos de potência. O modelo adotado no estudo apresentou um bom desempenho, uma vez que, as taxas de erro atingiram

valores aceitáveis mesmo em um cenário real de perturbações elétricas. Todavia, classificar essas alterações elétricas com técnicas de mineração de dados é somente uma dentre as várias opções que podem ser empregadas para este fim.

Quanto ao desempenho dos classificadores, os resultados apontaram que Random Forest e KNN apresentam um melhor desempenho com relação aos outros classificadores. Quando o objetivo é alcançar um bom nível de generalização, o classificador KNN é mais atrativo. Entretanto, a partir do momento que o objetivo é tentar minimizar o custo computacional, é interessante adotar algoritmos mais simples como as árvores de decisão. Por isso, deve ser feito um estudo mais aprofundado no processo de treinamento da árvore de decisão para evitar problemas tais como *overfitting*.

A respeito da seleção de atributos a ajuda dos especialistas foi de fundamental importância para reduzir a dimensionalidade de informações de não interesse para o domínio do problema investigado no estudo. Essa postura favoreceu para que os algoritmos de mineração de dados tivessem um melhor desempenho, pois diminui o custo computacional na classificação das perturbações elétricas.

Conclui-se que a aplicação de mineração de dados em sistemas elétricos de potência pode ser uma alternativa viável para auxiliar na tomada de decisão na gestão operacional. Em especial, quando se utiliza dados reais extraídos de PMUs na classificação de perturbações elétricas.

Certamente ainda existe a necessidade de se investigar mais alternativas que possam fazer estimações mais precisas sobre perturbações em redes elétricas. Entretanto, o objetivo do trabalho foi justamente diminuir essa realidade, dando subsídios de partida para que futuros estudos possam surgir aprimorando as técnicas adotadas nessa dissertação na mesma linha de pesquisa.

I. Trabalhos Futuros

Como sugestões para trabalhos futuros relacionados com as contribuições apresentadas neste trabalho, é possível citar:

- a) Construir uma base de dados integrando informações das PMUs com outras fontes de dados para que se possam estabelecer relações de causa e efeito;
- b) Aplicar técnicas de seleção automática de atributos nos dados provenientes das PMUs instaladas no sistema tramoeste utilizando abordagens do tipo filtro e *wrapper* (Hall et al., 2009);

- c) Como as perturbações em sistemas elétricos de potência podem ser influenciadas também pela variação de tensão, corrente, diferença de ângulo entre outros, seria relevante investigar aspectos relacionados a aumentar o intervalo de tempo de coleta de dados das PMUs, de forma que o operador possa ter maior tempo para uma eventual tomada de decisão;
- d) Investigar a possibilidade de aprimorar a metodologia proposta do estudo para assistir e fazer previsões de perturbações em outros SEPs;
- e) Avaliar outros algoritmos de mineração de dados de baixo custo computacional para classificação de perturbações, tais como as Redes Bayesianas e Sistemas Fuzzy.

REFERÊNCIAS BIBLIOGRÁFICAS

- Bakken, David E; Anjan Bose, Carl H Hauser, David E Whitehead & Gregory C Zweigle (2011), **‘Smart generation and transmission with coherent, real-time data’**, Proceedings of the IEEE 99(6), 928–951.
- Benmouyal, Gabriel, EO Schweitzer & A Guzmán (2002), **Medição fasorial sincronizada dos relés de proteção para controle, proteção e análise de sistemas elétricos de potência**, em ‘29th Annual Western Protective Relay Conference Spokane, Washington’.
- Bollen, Math HJ, Irene YH Gu, Surya Santoso, Mark F McGranaghan, Peter A Crossley, Moisés V Ribeiro, Paulo F Ribeiro, Fredrik Gustafsson, Klaus Witrisal, Geert Leus et al. (2009), **‘Bridging the gap between signal and power’**, IEEE SIGNAL PROCESSING MAGAZINE .
- Cabena, Peter, Pablo Hadjinian, Rolf Stadler, Jaap Verhees & Alessandro Zanasi (1998), **Discovering data mining: from concept to implementation**, Prentice-Hall, Inc.
- Morais, Jefferson Magalhães de (2011), **Avaliação de desempenho de classificadores de faltas em sistemas elétricos de potência**, Tese de doutorado, Universidade Federal do Pará.
- Baltazar, Antonio Carlos dos Santos (2007), **Qualidade da energia no contexto da reestruturação do setor elétrico brasileiro**, Tese de doutorado, Universidade de São Paulo - USP.
- Santos, Luciano Drosda M. dos; Renê Mikami, Ana Cristina B Kochem Vendramin & Celso Antônio A Kaestner (2009), **‘Procedimentos de validação cruzada em mineração de dados para ambiente de computação paralela’**.
- Faceli, Katti (2011), **Inteligência artificial: uma abordagem de aprendizado de máquina**, Grupo Gen-LTC.
- Fathabadi, Hassan (2015), **‘Two novel proposed discrete wavelet transform and filter based approaches for short-circuit faults detection in power transmission lines’**, *Applied Soft Computing* 36, 375–382.
- Fayyad, Usama, Gregory Piatetsky-Shapiro & Padhraic Smyth (1996), **‘From data mining to knowledge discovery in databases’**, *AI magazine* 17(3), 37.
- Gómez, Oscar & Mario A Ríos (2013), **‘IIP-based multistage placement of pmus with dynamic monitoring constraints’**, *International Journal of Electrical Power & Energy Systems* 53, 95–105.
- Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann & Ian H Witten (2009), **‘The weka data mining software: an update’**, *ACM SIGKDD Explorations newsletter* 11(1), 10–18.
- Hamrouni, Tarek, Sarra Slimani & Faouzi Ben Charrada (2015), **‘A data mining correlated patterns-based periodic decentralized replication strategy for data grids’**, *Journal of Systems and Software* 110, 10–27.

Han, Jiawei, Micheline Kamber & Jian Pei (2011), *Data mining: concepts and techniques: concepts and techniques*, Elsevier.

Haykin, Simon S, Simon S Haykin, Simon S Haykin & Simon S Haykin (2009), *Neural networks and learning machines*, Vol. 3, Pearson Education Upper Saddle River.

Krzysztof, J Cios, W Swiniaski Roman & Lukasz Kurgan (2007), '**Data mining: Knowledge discovery approach**'.

Liao, Yuan (2014), '**A novel method for locating faults on distribution systems**', *Electric Power Systems Research* 117, 21–26..

Mohammadi, Hanieh & Maryam Dehghani (2015), '**Pmu based voltage security assessment of power systems exploiting principal component analysis and decision trees**', *International Journal of Electrical Power & Energy Systems* 64, 655–663.

Oppenheim, Alan V, Ronald W Schafer, John R Buck et al. (1989), *Discrete-time signal processing*, Vol. 2, Prentice-hall Englewood Cliffs.

Pires, Yomara Pinheiro (2009), **Mineração de dados aplicada a Sistemas Elétricos: Classificação de faltas de curto-circuito em linhas de transmissão**, Tese de doutorado, Universidade Federal do Pará.

Rifkin, Ryan & Aldebaro Klautau (2004), '**In defense of one-vs-all classification**', *The Journal of Machine Learning Research* 5, 101–141.

Singh, Bindeshwar, NK Sharma, AN Tiwari, KS Verma & SN Singh (2011), '**Applications of phasor measurement units (pmus) in electric power system networks incorporated with facts controllers**', *International Journal of Engineering, Science and Technology* 3(3).

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.

Tochetto, André Pagani (2014), Utilização de medição fasorial sincronizada no monitoramento em tempo real da segurança da operação uhe-itaipu 60hz, Dissertação de mestrado, Universidade Federal de Itajubá.

Vapnik, Vladimir (2013), *The nature of statistical learning theory*, Springer Science & Business Media.

Wang, John (2005), *Encyclopedia of data warehousing and mining*, IGI Global.

Witten, Ian H & Eibe Frank (2005), *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann.

Zamani, M Amin, Robert Beresh & Stephen L Cress (2015), '**A pmu-augmented stability power limit assessment for reliable arming of special protection systems**', *Electric Power Systems Research* 128, 134–143.

Zamora-Cárdenas, EA, BA Alcaide-Moreno & CR Fuerte-Esquivel (2014), ‘**State estimation of flexible ac transmission systems considering synchronized phasor measurements**’, *Electric Power Systems Research* **106**, 120–133.

Zhang, J, ZY He, S Lin, YB Zhang & QQ Qian (2013), ‘**An anfis-based fault classification approach in power distribution system**’, *International Journal of Electrical Power & Energy Systems* **49**, 43–252.

Zhou, Qifeng, Hao Zhou, Qingqing Zhou, Fan Yang & Linkai Luo (2014), ‘**Structured damage detection based on random forest recursive feature elimination**’, *Mechanical Systems and Signal Processing* **46**(1), 82–90.

Zimmer, Valmor, Thiago Jeremias, Ildemar C Decker & Marcelo Nagostini (2012), ‘**Detecção de perturbações sistêmicas usando dados de medição fasorial sincronizada**, em ‘XIX Congresso Brasileiro de Automática, Campina Grande, Paraíba, Brasil’.

Zin, Abdullah Asuhaimi Mohd, Makmur Saini, Mohd Wazir Mustafa, Ahmad Rizal Sultan et al. (2015), ‘**New algorithm for detection and fault classification on parallel transmission line using dwt and bpnn based on clarke’s transformation**’, *Neurocomputing*.

Lewis, W. M. (2000). **Basis for the protection and management of tropical lakes. Lakes and Reservoirs: Research and Management** 5, 35-48.

Fechine, J.M (2010), **A Transformada de Fourier e Suas Aplicações**. Grupo PET Computação.

APÊNDICE

I. Descrição das Variáveis da PMU

Descrição sucinta das variáveis medidas pelas PMUs e armazenadas no concentrador de dados (PDC) do sistema Tramoeste.

- 1 - Tu-98013 Frequency
- 2 - Tu-98013 Frequency Delta (dF/dt)
- 3 - Tu-98013 Status Flags
- 4 - Tu-98013 VA1 BARRA-2TPBI Positive SequenceVoltage Magnitude
- 5 - Tu-98013 VA1 BARRA-2TPBI Positive SequenceVoltagePhaseAngle
- 6 - Tu-98013 VB1 BARRA-2TPBI Positive SequenceVoltage Magnitude
- 7 - Tu-98013 VB1 BARRA-2TPBI Positive SequenceVoltagePhaseAngle
- 8 - Tu-98013 VC1 BARRA-2TPBI Positive SequenceVoltage Magnitude
- 9 - Tu-98013 VC1 BARRA-2TPBI Positive SequenceVoltagePhaseAngle
- 10 - Tu-98013 VA1 TCCA-TP Positive SequenceVoltage Magnitude
- 11 - Tu-98013 VA1 TCCA-TP Positive SequenceVoltagePhaseAngle
- 12 - Tu-98013 VB1 TCCA-TP Positive SequenceVoltage Magnitude
- 13 - Tu-98013 VB1 TCCA-TP Positive SequenceVoltagePhaseAngle
- 14 - Tu-98013 VC1 TCCA-TP Positive SequenceVoltage Magnitude
- 15 - Tu-98013 VC1 TCCA-TP Positive SequenceVoltagePhaseAngle
- 16 - Tu-98013 IA1 TCAT-2TCA Positive SequenceCurrent Magnitude
- 17 - Tu-98013 IA1 TCAT-2TCA Positive SequenceCurrentPhaseAngle
- 18 - Tu-98013 IB1 TCAT-2TCA Positive SequenceCurrent Magnitude
- 19 - Tu-98013 IB1 TCAT-2TCA Positive SequenceCurrentPhaseAngle
- 20 - Tu-98013 IC1 TCAT-2TCA Positive SequenceCurrent Magnitude

- 21 - Tu-98013 IC1 TCAT-2TCA Positive SequenceCurrentPhaseAngle
- 22 - Tu-98013 IA1 TCCA-2TCB Positive SequenceCurrent Magnitude
- 23 - Tu-98013 IA1 TCCA-2TCB Positive SequenceCurrentPhaseAngle
- 24 - Tu-98013 IB1 TCCA-2TCB Positive SequenceCurrent Magnitude
- 25 - Tu-98013 IB1 TCCA-2TCB Positive SequenceCurrentPhaseAngle
- 26 - Tu-98013 IC1 TCCA-2TCB Positive SequenceCurrent Magnitude
- 27 - Tu-98013 IC1 TCCA-2TCB Positive SequenceCurrentPhaseAngle
- 46 - Tu-98013 RP V-310 Frequency
- 47 - Tu-98013 RP V-310 Frequency Delta (dF/dt)
- 48 - At-98012 Frequency
- 49 - At-98012 Frequency Delta (dF/dt)
- 50 - At-98012 Status Flags
- 51 - At-98012 VA1 BARRA1-3TPBI Positive Sequence Voltage Magnitude
- 52 - At-98012 VA1 BARRA1-3TPBI Positive Sequence Voltage Phase Angle
- 53 - At-98012 VB1 BARRA1-3TPBI Positive Sequence Voltage Magnitude
- 54 - At-98012 VB1 BARRA1-3TPBI Positive Sequence Voltage Phase Angle
- 55 - At-98012 VC1 BARRA1-3TPBI Positive Sequence Voltage Magnitude
- 56 - At-98012 VC1 BARRA1-3TPBI Positive Sequence Voltage Phase Angle
- 57 - At-98012 IA1 TCAT-3TCAX Positive Sequence Current Magnitude
- 58 - At-98012 IA1 TCAT-3TCAX Positive Sequence Current Phase Angle
- 59 - At-98012 IB1 TCAT-3TCAX Positive Sequence Current Magnitude
- 60 - At-98012 IB1 TCAT-3TCAX Positive Sequence Current Phase Angle
- 61 - At-98012 IC1 TCAT-3TCAX Positive Sequence Current Magnitude
- 62 - At-98012 IC1 TCAT-3TCAX Positive Sequence Current Phase Angle
- 63 - At-98012 IA1 ATRU-3TCAY Positive Sequence Current Magnitude

- 64 - At-98012 IA1 ATRU-3TCAY Positive Sequence Current Phase Angle
- 65 - At-98012 IB1 ATRU-3TCAY Positive Sequence Current Magnitude
- 66 - At-98012 IB1 ATRU-3TCAY Positive Sequence Current Phase Angle
- 67 - At-98012 IC1 ATRU-3TCAY Positive Sequence Current Magnitude
- 68 - At-98012 IC1 ATRU-3TCAY Positive Sequence Current Phase Angle
- 69 - At-98012 IA1 ATAA-TCB Positive Sequence Current Magnitude
- 70 - At-98012 IA1 ATAA-TCB Positive Sequence Current Phase Angle
- 71 - At-98012 IB1 ATAA-TCB Positive Sequence Current Magnitude
- 72 - At-98012 IB1 ATAA-TCB Positive Sequence Current Phase Angle
- 73 - At-98012 IC1 ATAA-TCB Positive Sequence Current Magnitude
- 74 - At-98012 IC1 ATAA-TCB Positive Sequence Current Phase Angle
- 93 - At-98012 RP V-310 Frequency
- 94 - At-98012 RP V-310 Frequency Delta (dF/dt)
- 95 - Ru-98012 Frequency
- 96 - Ru-98012 Frequency Delta (dF/dt)
- 97 - Ru-98012 Status Flags
- 98 - Ru-98012 VA1 BARRA1-3TPBI Positive Sequence Voltage Magnitude
- 99 - Ru-98012 VA1 BARRA1-3TPBI Positive Sequence Voltage Phase Angle
- 100 - Ru-98012 VB1 BARRA1-3TPBI Positive Sequence Voltage Magnitude
- 101 - Ru-98012 VB1 BARRA1-3TPBI Positive Sequence Voltage Phase Angle
- 102 - Ru-98012 VC1 BARRA1-3TPBI Positive Sequence Voltage Magnitude
- 103 - Ru-98012 VC1 BARRA1-3TPBI Positive Sequence Voltage Phase Angle
- 104 - Ru-98012 IA1 ATRU-3TCAX Positive Sequence Current Magnitude
- 105 - Ru-98012 IA1 ATRU-3TCAX Positive Sequence Current Phase Angle
- 106 - Ru-98012 IB1 ATRU-3TCAX Positive Sequence Current Magnitude

- 107 - Ru-98012 IB1 ATRU-3TCAX Positive Sequence Current Phase Angle
- 108 - Ru-98012 IC1 ATRU-3TCAX Positive Sequence Current Magnitude
- 109 - Ru-98012 IC1 ATRU-3TCAX Positive Sequence Current Phase Angle
- 110 - Ru-98012 IA1 TARU-3TCBY1 Positive Sequence Current Magnitude
- 111 - Ru-98012 IA1 TARU-3TCBY1 Positive Sequence Current Phase Angle
- 112 - Ru-98012 IB1 TARU-3TCBY1 Positive Sequence Current Magnitude
- 113 - Ru-98012 IB1 TARU-3TCBY1 Positive Sequence Current Phase Angle
- 114 - Ru-98012 IC1 TARU-3TCBY1 Positive Sequence Current Magnitude
- 115 - Ru-98012 IC1 TARU-3TCBY1 Positive Sequence Current Phase Angle
- 116 - Ru-98012 IA1 IBRU-3TCCY1 Positive Sequence Current Magnitude
- 117 - Ru-98012 IA1 IBRU-3TCCY1 Positive Sequence Current Phase Angle
- 118 - Ru-98012 IB1 IBRU-3TCCY1 Positive Sequence Current Magnitude
- 119 - Ru-98012 IB1 IBRU-3TCCY1 Positive Sequence Current Phase Angle
- 120 - Ru-98012 IC1 IBRU-3TCCY1 Positive Sequence Current Magnitude
- 121 - Ru-98012 IC1 IBRU-3TCCY1 Positive Sequence Current Phase Angle
- 140 - Ru-98012 RP V-310 Frequency
- 141 - Ru-98012 RP V-310 Frequency Delta (dF/dt)
- 142 - Ru-98012 RP V-310 Status Flags
- 143 - Ru-98012 VA1 BARRA1-3TPBI RP V-310 Positive Sequence Voltage Magnitude
- 144 - Ru-98012 VA1 BARRA1-3TPBI RP V-310 Positive Sequence Voltage Phase Angle
- 145 - Ru-98012 VB1 BARRA1-3TPBI RP V-310 Positive Sequence Voltage Magnitude
- 146 - Ru-98012 VB1 BARRA1-3TPBI RP V-310 Positive Sequence Voltage Phase Angle
- 147 - Ru-98012 VC1 BARRA1-3TPBI RP V-310 Positive Sequence Voltage Magnitude
- 148 - Ru-98012 VC1 BARRA1-3TPBI RP V-310 Positive Sequence Voltage Phase Angle
- 149 - Ru-98012 IA1 ATRU-3TCAX RP V-310 Positive Sequence Current Magnitude

- 150 - Ru-98012 IA1 ATRU-3TCAX RP V-310 Positive Sequence Current Phase Angle
- 151 - Ru-98012 IB1 ATRU-3TCAX RP V-310 Positive Sequence Current Magnitude
- 152 - Ru-98012 IB1 ATRU-3TCAX RP V-310 Positive Sequence Current Phase Angle
- 153 - Ru-98012 IC1 ATRU-3TCAX RP V-310 Positive Sequence Current Magnitude
- 154 - Ru-98012 IC1 ATRU-3TCAX RP V-310 Positive Sequence Current Phase Angle
- 155 - Ru-98012 IA1 TARU-3TCBY1 RP V-310 Positive Sequence Current Magnitude
- 156 - Ru-98012 IA1 TARU-3TCBY1 RP V-310 Positive Sequence Current Phase Angle
- 157 - Ru-98012 IB1 TARU-3TCBY1 RP V-310 Positive Sequence Current Magnitude
- 158 - Ru-98012 IB1 TARU-3TCBY1 RP V-310 Positive Sequence Current Phase Angle
- 159 - Ru-98012 IC1 TARU-3TCBY1 RP V-310 Positive Sequence Current Magnitude
- 160 - Ru-98012 IC1 TARU-3TCBY1 RP V-310 Positive Sequence Current Phase Angle
- 161 - Ru-98012 IA1 IBRU-3TCCY1 RP V-310 Positive Sequence Current Magnitude
- 162 - Ru-98012 IA1 IBRU-3TCCY1 RP V-310 Positive Sequence Current Phase Angle
- 163 - Ru-98012 IB1 IBRU-3TCCY1 RP V-310 Positive Sequence Current Magnitude
- 164 - Ru-98012 IB1 IBRU-3TCCY1 RP V-310 Positive Sequence Current Phase Angle
- 165 - Ru-98012 IC1 IBRU-3TCCY1 RP V-310 Positive Sequence Current Magnitude
- 166 - Ru-98012 IC1 IBRU-3TCCY1 RP V-310 Positive Sequence Current Phase Angle