



UFPA

**AVALIAÇÃO DA SEGURANÇA ESTÁTICA DE
SISTEMAS DE POTÊNCIA UTILIZANDO ÁRVORE
DE DECISÃO**

Bernard Carvalho Bernardes

2º Semestre/2010

UNIVERSIDADE FEDERAL DO PARÁ
CAMPUS DE TUCURUÍ
FACULDADE DE ENGENHARIA ELÉTRICA
TUCURUÍ-PARÁ
Turucuí, 08 de dezembro de 2010.

UNIVERSIDADE FEDERAL DO PARÁ
CAMPUS DE TUCURUÍ
FACULDADE DE ENGENHARIA ELÉTRICA

AVALIAÇÃO DA SEGURANÇA ESTÁTICA DE SISTEMAS DE POTÊNCIA
UTILIZANDO ÁRVORE DE DECISÃO

Trabalho submetido ao colegiado do
curso de Engenharia Elétrica para
obtenção do grau de Engenheiro
Eletricista.

Bernard Carvalho Bernardes
Autor

Tucuruí-PA
2º Semestre/2010

Avaliação da Segurança Estática de Sistemas de Potência Utilizando Árvore de Decisão

Este trabalho foi julgado em 08/12/2010 adequado para a obtenção do Grau de Engenheiro Eletricista, e aprovado na sua forma final pela banca examinadora que atribuiu o conceito EXCELENTE.

Prof. Dr. João Paulo Abreu Vieira

(Orientador/UFPA)

Prof. M.Sc. Ivaldo Ohana

(Co Orientador/UFPA)

Prof. Dr. Ubiratan Holanda Bezerra

(MEMBRO DA BANCA EXAMINADORA/UFPA)

Prof. M.Sc. Wellington Alex dos Santos Fonseca

(MEMBRO DA BANCA EXAMINADORA/UFPA)

Prof. M.Sc. Raphael Barros Teixeira

(DIRETOR DA FACULDADE DE ENGENHARIA ELÉTRICA/UFPA)



UNIVERSIDADE FEDERAL DO PARÁ
CAMPUS DE TUCURUÍ
FACULDADE DE ENGENHARIA ELÉTRICA
COORDENAÇÃO DE TRABALHOS DE CONCLUSÃO DE CURSO

AUTOR(ES) / MATRÍCULA(S)

NOME: BERNARD CARVALHO BERNARDES	MATRÍCULA:
NOME:	MATRÍCULA:

ORIENTADOR (NOME/INSTITUIÇÃO/EMPRESA)

JOÃO PAULO ABREU VIEIRA

CO-ORIENTADOR (NOME/INSTITUIÇÃO/EMPRESA)

IVALDO OHANA

TÍTULO DO TRABALHO

AValiação da SEGURANÇA ESTÁTICA DE SISTEMAS DE POTÊNCIA UTILIZANDO ÁRVORE DE DECISÃO

MEMBROS DA BANCA EXAMINADORA (NOME/INSTITUIÇÃO/EMPRESA)

	CONCEITO
1. Ivaldo Ohana:	EXC
2. Ubiratan Holanda Bezerra	EXC
3. JOÃO PAULO ABREU VIEIRA	EXC
4. Wellington Alex dos Santos Fonseca	EXC

DIA DA DEFESA: 08 / 12 / 2010

HORA DA DEFESA: 15:00

João Paulo Abreu Vieira

ORIENTADOR (A)

Ivaldo Ohana:

CO-ORIENTADOR (A)

MEMBRO (A) 1

MEMBRO (A) 2

*Ao meu pai, Severino
Bernardes, e a minha mãe,
Maria da Conceição
Martins Carvalho, pelos
conhecimentos repassados e
pela paciência com a qual
me criaram.*

Agradecimentos

Acima de tudo a Deus por permitir que eu chegasse até aqui com vida e saúde.

Aos professores João Paulo Abreu Vieira e Ivaldo Ohana, pelo empenho e dedicação na orientação deste trabalho, pelas inúmeras conversas que tivemos sempre me incentivando a alcançar os objetivos que traçamos juntos, meus mais profundos e sinceros agradecimentos.

Aos meus amigos, companheiros de todas as horas, pela amizade e momentos felizes que me proporcionaram. Em especial aos amigos do CP-Five pelos longos anos de amizade.

A minha namorada Núbia, pelo companheirismo e cumplicidade que me proporcionou por todos esses dias.

A minha família, por confiarem em mim e me proporcionarem uma boa base familiar que me permitiu alcançar meus objetivos.

A todos os citados meus mais sinceros agradecimentos por fazerem parte da minha vida.

*“Uma pessoa inteligente
resolve um problema, um
sábio o previne.”*

Albert Einstein

Sumário

Lista de Figuras	ix
Lista de Tabelas	xi
Resumo	xii
Capítulo 1 – Introdução	13
1.1 Introdução	13
1.2 Objetivo Geral	14
1.3 Organização do Trabalho	15
Capítulo 2 – Segurança Estática de Sistemas de Potência	16
2.1 Introdução	16
2.2 Segurança Estática de Sistemas de Potência	16
2.2.1 Fluxo de Carga	18
2.2.2 Método de Newton-Raphson Aplicado ao Fluxo de Carga	20
2.3 Estabilidade de Tensão	23
2.3.1 Análise Estática da Estabilidade de Tensão	24
Capítulo 3 – Descoberta de Conhecimento em Banco de Dados	27
3.1 Introdução	27
3.2 Descoberta de Conhecimento em Banco de Dados	27
3.3 Mineração de Dados	29
3.3.1 Tarefas de Mineração de Dados	29
3.4 Árvore de Decisão	32
3.4.1 Algoritmo de Hunt	35
Capítulo 4 – Estudo de Caso	39
4.1 Introdução	39

4.2	Sistema Elétrico de Potência em Estudo	41
4.3	Criação do Banco de Dados de Treino e Teste	42
4.4	Criação e Avaliação da Árvore de Decisão	52
4.4.1	Criação e Avaliação da Árvore de Decisão Utilizando a Mesma Base de Dados	52
4.4.2	Criação e Avaliação da Árvore de Decisão Utilizando Duas Bases de Dados Distintas	55
	Capítulo 5 – Conclusões	58
5.1	Conclusões Gerais	58
5.2	Sugestões para Trabalho Futuros	59
	Referências Bibliográficas	61

Lista de Figuras

2.1	Curva P-V, Adaptada de [ISO09]	25
3.1	Processo de Descoberta de Conhecimento em Banco de Dados, Adaptado de [TAN09]	28
3.2	Mineração de Dados como Intersecção de Muitas Áreas	30
3.3	Uma Árvore de Decisão Criada no <i>Software RapidMiner</i> , Versão 4.2, Para o Problema do Jogo de Tênis	34
3.4	Algoritmo de Hunt Para Induzir Árvores de Decisão	37
4.1	Fluxograma do Processo de Criação e Avaliação da Árvore de Decisão	40
4.2	Diagrama de Impedâncias do Sistema de Nove Barras, Todas as Impedâncias Estão em pu na Base 100 MVA, Adaptado de [AND03]	41
4.3	Fluxograma da Ferramenta de Criação do Banco de Dados das Condições Operacionais do Sistema	43
4.4	Resposta do FluxMining ao fluxo de carga para o sistema de 9 Barras	50
4.5	Resposta do lfdemo ao fluxo de carga para o sistema de 9 barras	51
4.6	Processo de Obtenção de Árvores de Decisão Pelo RapidMiner	52
4.7	Árvore de Decisão Criada Para o Banco de Dados de Entrada	54
4.8	Matriz de Confusão Para o Modelo de Árvore Proposto	54
4.9	Algoritmo da Árvore de Decisão Para o Sistema de 9 barras	55
4.10	Operadores no RapidMiner para Treino e Avaliação da Árvore de Decisão	

Utilizando Bancos de Dados Diferentes	56
4.11 Matriz de Confusão Para a Árvore Proposta Utilizando um Banco de Dados para Treino e Outro para Teste	57

Lista de Tabelas

3.1	Tabela de Decisão do Jogo de Tênis	32
3.2	Conjunto de Treino Para Prever Tomadores de Empréstimos que se Tornarão Inadimplente	36
4.1	Banco de Dados das Barras do Sistema Exemplo de 9 Barras	47
4.2	Banco de Dados das Linhas do Sistema Exemplo de 9 Barras	48
4.3	Banco de Dados das Linhas que Podem ser Retiradas do Sistema	48
4.4	Banco de Dados dos Limites Operacionais das Barras do Sistema	49
4.5	Banco de Dados dos Limites Operacionais das Linhas do Sistema	49

Resumo

Este trabalho apresenta o desenvolvimento de uma metodologia de avaliação da segurança estática de um sistema de potência utilizando mineração de dados, mais especificamente a técnica árvore de decisão, a qual é uma ferramenta supervisionada para solucionar problemas de classificação em grande massa de dados. No método proposto a árvore de decisão foi treinada e testada para estimar e classificar o estado operativo do sistema, levando em conta que o banco de dados de treinamento possui algumas condições operacionais do sistema e as mais prováveis contingências. Como recurso ao sistema de avaliação da segurança estática proposto, simulações realizadas em uma ferramenta criada para esse propósito chamada de FluxMining, permitiram analisar o estado de operação em regime permanente do sistema elétrico. Posteriormente, adequando-se à metodologia em questão, o software livre RAPIDMINER de mineração de dados, o qual inclui os algoritmos de árvores de decisão foi utilizado para criar a árvore de decisão a fim de realizar a tarefa de classificação.

Palavras-chave: Sistemas Elétricos de Potência, Segurança Estática de Sistemas de energia elétrica, Descoberta de Conhecimento em Bases de Dados, Fluxo de Carga, Mineração de dados.

Capítulo 1 - Introdução

1.1 Introdução

Como os modernos sistemas de potência estão operando cada vez mais estressados devido à desregulamentação do setor elétrico e a dificuldade de adição de novas instalações de transmissão, a ocorrência de contingências críticas pode levar a violação dos limites operacionais da segurança e até mesmo iniciar graves problemas de estabilidade. Uma contingência severa pode causar graves impactos sobre o sistema elétrico, determinada principalmente pelo tipo e gravidade da perturbação. Na atualidade, a forma mais prática para avaliar os diferentes problemas relacionados à segurança é realizada por meio de ferramentas computacionais de simulação comerciais que representam adequadamente a modelagem dos componentes do sistema de potência. No entanto, o desenvolvimento de métodos para a avaliação que possibilitem a operação em tempo real ainda é um grande desafio devido à alta não-linearidade e a alta dimensionalidade dos grandes sistemas de potência. O problema da dimensionalidade dos dados, de fato, torna impossível verificar todas as combinações de contingências em um prazo de tempo razoável.

Para se avaliar a segurança estática de sistemas de potência deve-se definir a princípio o ponto de operação do mesmo a fim de determinar se seu estado operacional é seguro ou inseguro. Esse estudo é importante, pois os sistemas de potência operam cada vez mais próximos dos seus limites operacionais, podendo operar de forma insegura com pequenas variações das cargas ligadas ao mesmo ou na ocorrência de contingências previstas ou imprevistas.

Dessa forma, conhecer quais as condições operacionais do sistema de energia elétrica que podem levá-lo a uma provável condição de insegurança torna-se o objetivo de muitos estudos para desenvolvimentos de ferramentas computacionais que possam impedir que o sistema responda dessa forma, ou indicar que um simples re-despacho da geração possa vir a impedir que algum componente do sistema de energia elétrica

(geradores, linhas de transmissão e transformadores de potência, banco de reatores/capacitores, etc.) opere inseguramente.

Diversas técnicas de inteligência computacional, tais como redes neurais, algoritmos genéticos, dentre outras, são empregadas em ferramentas computacionais que objetivam resolver o problema citado. Muitas dessas técnicas apresentam uma etapa de treinamento, para “aprender” o comportamento de um sistema a fim de posteriormente prever esse comportamento. No caso dos sistemas de potência, essa etapa de treinamento é facilitada na medida em que esses sistemas geram grandes quantidades de dados operacionais, os quais são obtidos através dos sistemas de supervisão e controle ou por meio de simulações que representam fielmente o comportamento do sistema elétrico. Assim, é possível a utilização dessas bases de dados em conjunto com alguma técnica para avaliar o comportamento em tempo real do sistema elétrico gerenciado.

Este trabalho se motivou em tentar avaliar a estabilidade estática de sistemas de potência utilizando uma dessas técnicas de inteligência computacional, as árvores de decisão, partindo dessas grandes bases de dados.

1.2 Objetivo Geral

Este trabalho apresenta o desenvolvimento de uma metodologia de avaliação da segurança estática de sistemas de potência em tempo real usando uma árvore de decisão, com o propósito de analisar a segurança pós-contingência, que inclui os problemas da violação da magnitude de tensão (VMT), violação dos limites térmicos (VLT) e a estabilidade estática de tensão (EET). Os principais procedimentos da metodologia proposta incluem: 1) O desenvolvimento de uma ferramenta de criação da base de dados em MATLAB que realiza um conjunto de simulações a fim de rotular os aspectos da segurança estática considerando vários cenários de operação para todas as possíveis N-1 contingências e uma lista pré-selecionada das prováveis N-2 contingências; 2) Criação de uma árvore de decisão utilizando duas técnicas de treino e teste, sendo que a base de dados obtida consta de atributos que podem ser provenientes de variáveis do sistema

elétrico adquiridas de um sistema de supervisão, controle e aquisição de dados (Supervisory Control and Data Acquisition – SCADA) ou de um sistema de unidade de medição fasorial (Phaser Measurement Units – PMU), caracterizando o estado do sistema.

1.3 Organização do Trabalho

Este trabalho está organizado em 5 capítulos.

O capítulo 1 dá uma introdução ao conteúdo deste trabalho e seus objetivos, fornecendo informações sobre a importância do estudo da segurança estática em sistemas de potência.

O capítulo 2 trata sobre a segurança de sistemas de potência, explicando algumas definições e caracterizando o problema estudado.

O capítulo 3 aborda sobre a descoberta de conhecimento em banco de dados apresentando e explicando os principais conceitos e aspectos desse processo. Mais especificamente, este capítulo se concentra no processo de mineração de dados por meio da técnica de classificação utilizando árvores de decisão.

O capítulo 4 mostra a aplicação da árvore de decisão para avaliação da segurança estática de um sistema de potência. Nesse capítulo será apresentado tanto o sistema de potência que será avaliado como a ferramenta utilizada para obter o banco de dados e a respectiva árvore de decisão do processo.

O capítulo 5 apresenta a conclusão deste trabalho e algumas sugestões de trabalhos futuros que possam vir realizados para dar continuidade no mesmo.

Capítulo 2 - Segurança de Sistemas de Potência

2.1 Introdução

A segurança de sistemas de potência pode ser definida como a habilidade do sistema em suportar qualquer distúrbio, de uma lista de fenômenos prováveis, sem que haja danos sérios à operação. Desta forma, o sistema é dito seguro se, após um determinado distúrbio na rede, este não é levado a um estado de emergência. O nível de segurança de um sistema está fortemente vinculado à sua robustez, depende das condições operativas e também da probabilidade de ocorrência de distúrbios.

A avaliação dessa segurança refere-se às análises e métodos utilizados na determinação do estado de segurança (seguro ou inseguro) de um sistema, tendo como base critérios pré-definidos. Dependendo da modelagem e das técnicas utilizadas, esta avaliação pode ser classificada como estática ou dinâmica. Esta última pode englobar uma série de outros aspectos relativos à estabilidade do sistema, tais como estabilidade transitória, estabilidade a pequenas perturbações e estabilidade de tensão. Já a segurança estática, tema foco do presente trabalho, avalia as condições do sistema em regime permanente no que diz respeito aos níveis de tensão e aos limites térmicos dos circuitos de transmissão, além da estabilidade estática de tensão.

2.2 Segurança Estática de Sistemas de Potência

A segurança estática de um sistema de potência está ligada a um dado ponto de equilíbrio da rede, para o qual são avaliados os níveis de tensão e carregamentos [CAP93]. Normalmente, uma ferramenta para analisar a segurança estática se resume a algoritmos de fluxo de potência. Ou seja, para um dado ponto de operação, o sistema é considerado estaticamente seguro se os níveis de tensão em todas as barras da rede estão dentro da faixa aceitável e se os fluxos nos circuitos de transmissão e nos

transformadores se encontram dentro dos seus limites térmicos. O efeito das contingências deve ser levado em consideração de forma que os critérios de segurança de regime permanente sejam sempre respeitados, mesmo quando houver perda de algum dispositivo do sistema, como a saída de uma geração por exemplo.

O problema da segurança estática pode surgir quando ocorrem contingências ou variações de patamares de cargas no sistema de potência, e que pode ser separado em três estudos paralelos, sendo eles a violação da magnitude de tensão (VMT), a violação dos limites térmicos das linhas de transmissão e transformadores (VLT) e a estabilidade estática de tensão (EET).

Um sistema de potência, operando em regime estacionário, está sujeito a restrições de carga e de operação. As restrições de carga representam as injeções de potências ativa e reativa nas barras de carga (PQ), bem como as injeções de potência reativa nas barras de geração (PV). Entre as restrições de operação estão os limites impostos às magnitudes das tensões nodais (quando esse limite é ultrapassado tem-se a VMT), aos fluxos de potência aparente nas linhas de transmissão e transformadores (quando ultrapassado, caracteriza a VLT) e às injeções de potências ativa e reativa nas barras de geração, segundo [MON83].

Os limites de operação do sistema e de equipamentos, referidos anteriormente, em muitos casos não são rígidos, no sentido de que podem ser violados por algum tempo. Em situações de emergência, eles são, em geral, mais abertos. Assim, a capacidade térmica de uma linha de transmissão ou o carregamento máximo de um transformador é diferente quando se consideram as sobrecargas possíveis para, por exemplo, três minutos ou uma hora, segundo [MON83].

No estudo dessas restrições deve ser levada em consideração a lista de contingências da qual o sistema está sujeito. Essa lista deve conter as contingências simples e certo número limitado de contingências múltiplas. Não existem critérios para a elaboração dessa lista, podendo ser levado em conta à experiência prévia do sistema e as probabilidades de ocorrência de falhas.

Conforme já dito, a segurança estática em sistemas de potência é avaliada por ferramentas de fluxo de carga, que determinam o “ponto de operação” do sistema para uma determinada condição de operação.

2.2.1 Fluxo de Carga

Um sistema elétrico de potência pode operar com um imenso número de estados operacionais, e ainda suprir a potência demandada, sendo função do engenheiro de sistemas de potência selecionar o melhor estado possível dentre os existentes, e esta seleção é baseada no estudo de fluxo de carga [MON83].

O fluxo de carga é a ferramenta que obtém as condições de operação (tensões, defasagem angular, fluxos de potência, etc.) de um sistema elétrico em função de sua topologia e dos níveis de demanda e geração. O ponto de operação de um sistema de potência é definido pelas magnitudes e defasagem angular das tensões das barras da rede elétrica. Com todas as tensões definidas, pode-se então conhecer os fluxos de potências ativa e reativa no sistema elétrico como um todo.

No estudo do fluxo de carga, a modelagem do sistema é estática, significando que o sistema elétrico é representado por um conjunto de equações e inequações algébricas, ou seja, é um estudo em regime permanente. O cálculo do fluxo de carga é, em geral, realizado utilizando-se métodos computacionais desenvolvidos especificamente para a resolução do sistema de equações e inequações algébricas que constituem o modelo estático do sistema elétrico, segundo [MON83].

Cada barra do sistema elétrico está associada a quatro variáveis, sendo que duas delas entram no problema, para ajudar a resolver o fluxo de carga, como dados e as outras duas como incógnitas, são elas:

V_i - magnitude da tensão nodal da barra i

δ_i - ângulo da tensão nodal

P_i - geração líquida de potência ativa na barra i ; e

Q_i - geração líquida de potência reativa

Em que P_i e Q_i são dados por:

$$P_i(V, \delta) = V_i \sum_{k=1}^n (G_{ik} \cos \delta_{ik} + B_{ik} \sin \delta_{ik}) V_k \quad (2.1)$$

$$Q_i(V, \delta) = V_i \sum_{k=1}^n (G_{ik} \sin \delta_{ik} - B_{ik} \cos \delta_{ik}) V_k \quad (2.2)$$

Onde: G_{ik} e B_{ik} são elementos da matriz admitância nodal.

Para a solução numérica destas equações, em geral, assume-se que o sistema trifásico é balanceado, com as linhas de transmissão representadas por suas impedâncias de seqüência positiva. Em forma matricial, a equação da tensão nodal complexa pode ser escrita como:

$$I = Y_{bus} V \quad (2.3)$$

Onde:

I - vetor de injeções de corrente

Y_{bus} - matriz admitância nodal

V - vetor das tensões nodais

O incremento das potências ativa e reativa nas barras é dada por:

$$\Delta P_i = P_i^{esp} - V_i \sum_{k=1}^n (G_{ik} \cos \delta_{ik} + B_{ik} \sin \delta_{ik}) V_k \quad (2.4)$$

$$\Delta Q_i = Q_i^{esp} - V_i \sum_{k=1}^n (G_{ik} \sin \delta_{ik} - B_{ik} \cos \delta_{ik}) V_k \quad (2.5)$$

Os cálculos dos fluxos de carga são em geral estabelecidos de uma forma tal que as tensões complexas são determinadas para que as potências ativas, P^{esp} , e reativas, Q^{esp} , das cargas e de determinados geradores sejam mantidas em valores especificados. Além disso, também é desejado que a magnitude da tensão, V^{esp} , e a potência ativa, P^{esp} , em certas barras de geração também sejam mantidas em valores especificados.

Dependendo de quais variáveis são os dados e quais são consideradas como incógnitas, definem-se três tipos de barras:

1. Barra Swing ou de Referência ou, ainda, barra Slack: Neste tipo de barra, a magnitude V_i e o ângulo δ_i da tensão são especificados e as injeções de potência ativa P^{esp} e reativa Q^{esp} são calculadas.
2. Barra PV ou de Tensão Controlada ou, ainda, barra de Geração: Neste tipo de barra, a injeção de potência ativa P^{esp} e a magnitude V_i da tensão são especificadas e o ângulo δ_i da tensão nodal e a injeção de potência reativa Q^{esp} são calculados.
3. Barra PQ ou de Carga: Neste tipo de barra, a qual é geralmente predominante sobre os outros dois tipos do sistema, as injeções de potência ativa P^{esp} e reativa Q^{esp} são especificadas e a magnitude V_i e o ângulo δ_i da tensão nodal são as variáveis calculadas.

2.2.2 Método de Newton-Raphson Aplicado ao Fluxo de Carga

Newton-Raphson é um método geral para a determinação de raízes reais de equações não lineares. Para o caso do fluxo de carga, trabalha-se com o vetor $g(x) = 0$, definido nas equações (2.6) e (2.7):

$$g_i^P = P_i^g - P_i^c - V_i \sum_{k \neq i} V_k (G_{ik} \cos \delta_{ik} + B_{ik} \sin \delta_{ik}) = 0 \quad (2.6)$$

$$g_i^Q = Q_i^g - Q_i^c - V_i \sum_{k \neq i} V_k (G_{ik} \sin \delta_{ik} - B_{ik} \cos \delta_{ik}) = 0 \quad (2.7)$$

As equações de iteração serão dadas por:

$$\left[\frac{\delta^{k+1}}{V^{k+1}} \right] = \left[\frac{\delta^k}{V^k} \right] + \left[\frac{\Delta \delta^k}{\Delta V^k} \right] \quad (2.8)$$

$$J = \begin{bmatrix} H & N \\ M & L \end{bmatrix} \quad (2.9)$$

Onde J é uma matriz quadrada, chamada de matriz Jacobiana.

As submatrizes que compõem a matriz jacobiana, J , dada em (2.9), são geralmente representadas por:

$$H = \frac{\partial P}{\partial \delta}; \quad N = \frac{\partial P}{\partial V};$$

$$M = \frac{\partial Q}{\partial \delta}; \quad L = \frac{\partial Q}{\partial V};$$

A equação geral do método de Newton-Raphson aplicado ao problema de fluxo de carga pode ser expressa na forma compacta como:

$$\begin{bmatrix} \Delta \delta^k \\ \Delta V/V^k \end{bmatrix} = - \begin{bmatrix} H & N \\ M & L \end{bmatrix}^{-1} \begin{bmatrix} \Delta P^k \\ \Delta Q^k \end{bmatrix} \quad (2.10)$$

As componentes das submatrizes jacobianas H, N, M, L são dadas por:

$$H_{ii} = \frac{\partial P_i}{\partial \delta_i} = -V_i^2 B_{ii} - V_i \sum_{i \in k} (G_{ik} \text{sen} \delta_{ik} - B_{ik} \text{cos} \delta_{ik}) V_k \quad (2.11)$$

$$H_{ik} = \frac{\partial P_i}{\partial \delta_k} = V_i (G_{ik} \text{sen} \delta_{ik} - B_{ik} \text{cos} \delta_{ik}) V_k \quad (2.12)$$

$$N_{ii} = \frac{\partial P_i}{\partial V_i} = V_i G_{ii} + \sum_{i \in k} (G_{ik} \text{cos} \delta_{ik} + B_{ik} \text{sen} \delta_{ik}) V_k \quad (2.13)$$

$$N_{ik} = \frac{\partial P_i}{\partial V_k} = V_i (G_{ik} \text{cos} \delta_{ik} + B_{ik} \text{sen} \delta_{ik}) \quad (2.14)$$

$$M_{ii} = \frac{\partial Q_i}{\partial \delta_i} = -V_i^2 G_{ii} + V_i \sum_{i \in k} (G_{ik} \text{cos} \delta_{ik} + B_{ik} \text{sen} \delta_{ik}) V_k \quad (2.15)$$

$$M_{ik} = \frac{\partial Q_i}{\partial \delta_k} = -V_i (G_{ik} \text{cos} \delta_{ik} + B_{ik} \text{sen} \delta_{ik}) V_k \quad (2.16)$$

$$L_{ii} = \frac{\partial Q_i}{\partial V_i} = -V_i B_{ii} + \sum_{i \in k} (G_{ik} \text{sen} \delta_{ik} - B_{ik} \text{cos} \delta_{ik}) V_k \quad (2.17)$$

$$L_{ik} = \frac{\partial Q_i}{\partial V_k} = V_i (G_{ik} \text{sen} \delta_{ik} - B_{ik} \text{cos} \delta_{ik}) \quad (2.18)$$

Os elementos $H_{ii}, N_{ii}, M_{ii}, L_{ii}$, deduzidos nas equações de (2.10) a (2.18) podem ser colocados em função das injeções de potência ativa e reativa na barra i , conforme abaixo:

$$H_{ii} = -Q_i - V_i^2 B_{ii} \quad (2.19)$$

$$N_{ii} = V_i^{-1} (P_i + V_i^2 G_{ii}) \quad (2.20)$$

$$M_{ii} = P_i - V_i^2 G_{ii} \quad (2.21)$$

$$L_{ii} = V_i^{-1} (Q_i - V_i^2 B_{ii}) \quad (2.22)$$

O processo de solução do fluxo de carga consiste, então, no seguinte algoritmo:

- a. Fazer $k = 1$, estimar valores iniciais para δ^k e V^k (normalmente $\delta = 0$ e $V = 1,0 pu$);
- b. Formar os vetores de incremento ΔP^k e ΔQ^k , usando as equações (2.4) e (2.5);
- c. Verificar se os critérios de convergência foram atingidos. Se eles foram atingidos, o processo termina, caso contrário o processo avança para o item d;
- d. Formar a matriz jacobiana, usando as equações de (2.11) a (2.22), e invertê-la;
- e. Calcular os vetores de correção $\Delta \delta^k$ e ΔV^k , usando a expressão (2.10);
- f. Obter os novos valores δ^{k+1} e V^{k+1} , usando a expressão (2.8), e fazer $k = k+1$ e voltar ao item b.

Os critérios de convergência (parada das interações) são:

$$\Delta P_i \leq \varepsilon_p \quad (2.23)$$

$$\Delta Q_i \leq \varepsilon_q \quad (2.24)$$

Onde ε_p e ε_q são tolerâncias (erros) pré-especificadas.

2.3 Estudo da Estabilidade de Tensão

A Estabilidade de tensão é a capacidade de um sistema em manter as tensões dentro de limites operacionais aceitáveis quando submetido a contingências ou aumento de carga. Um sistema é considerado instável quando o aumento da demanda de carga e/ou alterações nas condições de operação produzem uma progressiva e incontrolável queda de tensão, conduzindo o sistema ou parte deste a um perfil de tensões extremamente baixo [KUN94].

A instabilidade de tensão é um fenômeno essencialmente local, mas suas conseqüências podem ampliar de maneira significativa o problema. Uma destas conseqüências é um fenômeno bem mais complexo, chamado colapso de tensão, que é uma seqüência de eventos sucessivos, que resultam em sub-tensões acentuadas em todas as barras ou em uma região significativa do sistema.

Assim sendo, é adequado classificar a estabilidade de tensão em duas categorias: estabilidade de tensão para grandes perturbações e estabilidade de tensão para pequenas perturbações. Esta subdivisão, essencialmente, separa os fenômenos que são examinados por meio de análises dinâmicas não-lineares, dos fenômenos que podem ser examinados usando análises em regime permanente [KUN94].

A estabilidade de tensão para grandes perturbações está relacionada com a habilidade do sistema em controlar as tensões das barras após a ocorrência de grandes perturbações no sistema, tais como curto-circuito, perda de carga ou perda de geração, por exemplo. Determinar essa forma de estabilidade requer uma análise do desempenho dinâmico do sistema em um tempo suficiente para capturar as interações de dispositivos, tais como Transformadores com Variação de Tap ou *Transformers Under Load Tap Changers* (ULTC) e limitadores de sobre-corrente de campo dos geradores do sistema ou *Over Excitation Limiter* (OXL). A estabilidade de tensão para grandes perturbações pode ser estudada usando simulações não-lineares no domínio do tempo, que inclui a modelagem adequada, além disso, pode ser subdividida com foco no curto-prazo (transitória) e longo prazo (de segundos a muitos minutos).

A estabilidade de tensão a pequenas perturbações (pequenos sinais) está relacionada com a habilidade do sistema em controlar as tensões das barras após pequenas perturbações, como um crescimento gradual das cargas do sistema. Esta forma de estabilidade pode ser efetivamente estudada com técnicas em regime permanente, que usam a linearização das equações dos sistemas dinâmicos para um dado ponto de operação.

Após uma perturbação, as tensões do sistema, muitas vezes, podem não retornar ao seu valor original. Portanto, é necessário definir a região dos níveis de tensão considerada aceitável. O sistema é, então, dito ter “estabilidade finita”, ou seja, é estável dentro da região determinada para os níveis de tensão.

2.3.1 Análise Estática da Estabilidade de Tensão

A estabilidade de tensão que pode ser analisada por ferramentas que visam o estudo da estabilidade estática do sistema de potência, no que diz respeito à capacidade do sistema em manter seus níveis de tensão, dentro de valores aceitáveis, quando ocorrem pequenas perturbações no sistema. Um exemplo disso é tentar atender as cargas do sistema quando essas superam a capacidade conjunta da transmissão e da geração. Ou seja, é um fenômeno que ocorre quando há um excesso de carga atendida pelo sistema, acima do valor que ele pode suportar. Quando a carga em um barramento do sistema aumenta, a tensão deste barramento tende a diminuir até atingir o seu limite (chamado de condição de limítrofe de tensão). A partir desse valor, se a carga aumentar ainda mais ocorrerá uma forte degeneração da tensão caracterizando a instabilidade do sistema. Quando esse fenômeno é observado, haverá uma série de eventos interligados neste processo havendo a deterioração do perfil de tensão, e certamente o sistema pode evoluir para um cenário de blecaute.

A observação do comportamento das tensões nodais, considerando o aumento gradativo do carregamento do sistema é uma maneira de se estudar a estabilidade de tensão. Nesse caso, a análise pode ser considerada como um problema linear.

Graficamente, pode-se ilustrar a situação do comportamento da carga em relação à tensão em uma barra do sistema para se determinar o ponto em que surge a chamada instabilidade de tensão. Trata-se da figura 2.1, chamada por [TAY94] como curva P-V. Esse procedimento visa ilustrar os valores críticos da potência ativa e da tensão na carga usando algum processo de busca. A diferença entre a potência ativa crítica e a potência ativa inicial é tomada como sendo o máximo acréscimo possível para a carga na referida barra e o valor encontrado para a potência crítica é tomado como sendo o máximo valor de carregamento, ou seja, se a carga for superior a esse valor na barra em estudo tem-se um indicativo de que sua tensão levará o sistema a um colapso. O ponto extremo, N, é conhecido como “nariz” da curva P-V.

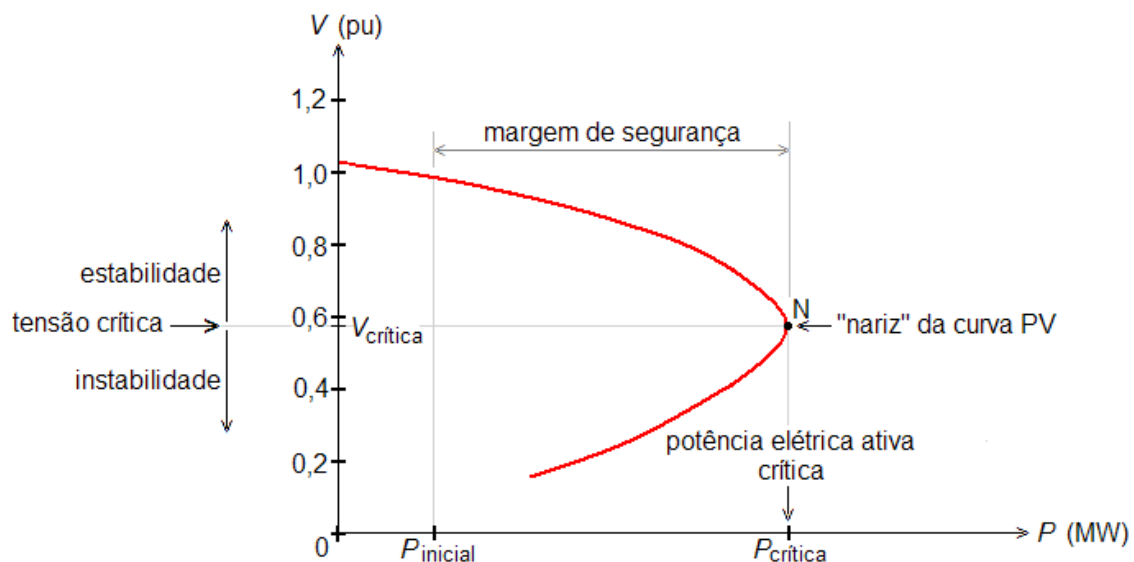


Figura 2.1 – Curva P-V, adaptada de [ISO09]

A análise de estabilidade estática de tensão fornece o ponto de máximo carregamento do sistema, condição na qual ocorre um mau condicionamento ou singularidade na matriz Jacobiana utilizada no processo de solução do fluxo de potência. Esta condição ocorre em sistemas em que se aplica um aumento gradual na carga e a solução não converge com o uso do fluxo de carga convencional. Este é o ponto de máximo carregamento, e para maiores valores de carga, não existe solução real para as equações de fluxo de potência. A distância em MW de um particular estado de operação até o ponto de carregamento máximo permite que se avalie o grau de segurança.

A existência de soluções com bifurcações tipo sela-nó nestas equações indica a aproximação do ponto de colapso do sistema. Estas bifurcações ocorrem quando o sistema se aproxima do ponto de colapso e um pequeno acréscimo de carga que ultrapassa o limite estático de estabilidade de tensão, desaparecendo o último ponto de equilíbrio estável. A inexistência de solução para estas equações indica que o sistema não tem condições físicas de atender a certo perfil de carregamento. Mudanças no despacho de geração, topologia de rede e compensação de reativos são então necessárias para que exista um novo e seguro estado de operação. O traçado das curvas P-V, através do Fluxo de Potência Continuado (FPC), com a parametrização do crescimento da carga na região próxima do colapso onde ocorre a singularidade da matriz Jacobiana, permite encontrar o ponto onde uma bifurcação tipo sela-nó ocorre.

Capítulo 3 - Descoberta de Conhecimento em Banco de Dados

3.1 Introdução

Rápidos avanços na tecnologia de coleta e armazenamento de dados permitiram que as organizações acumulassem uma vasta quantidade dos mesmos. Essas grandes quantidades de dados armazenadas em banco de dados podem vir a ter informações interessantes e extraí-las tem provado ser uma tarefa extremamente desafiadora.

Nesse paradigma surgiu a idéia de descobrir conhecimento nessas bases de dados existentes, um processo de extração de padrões, processo conhecido como descoberta de conhecimento em banco de dados (*Knowledge Discovery in Databases – KDD*). O processo de KDD é um processo iterativo porque o conhecimento descoberto apresentado ao usuário deve ser usado como base para ser avaliado e aprimorado. O KDD é interdisciplinar e envolvem diversas áreas, tais como a estatística, matemática, banco de dados, aprendizado de máquinas, sistemas especialistas e reconhecimento de padrões. O processo KDD combina técnicas, algoritmos e definições de todas essas áreas com o objetivo principal de extrair conhecimento de grandes bases de dados

3.2 Descoberta de Conhecimento em Banco de Dados

Na realização da KDD os dados de entrada podem ser apresentados em uma infinidade de formatos, desde arquivos simples até planilhas ou Tabelas relacionais, e podem estar armazenados em repositórios centrais de dados (*Data Warehouse*) ou distribuídos em múltiplos locais. O processo de KDD é constituído de três etapas principais, conforme a Figura 3.1.

A etapa de **pré-processamento** tem por objetivo transformar os dados de entrada brutos em um formato apropriado para as análises subseqüentes. Nessa etapa pode ocorrer a junção de dados de múltiplas fontes, a limpeza de dados para remover os ruídos, a seleção de registros, dentre outras transformações importantes que são de

relevância para a etapa de mineração de dados, essa etapa pode tomar até 80% do tempo necessário para todo o processo, devido às dificuldades de integração de bases de dados heterogêneas [MAN96]. Os dados pré-processados podem passar ainda por outra transformação, para serem armazenados corretamente, visando à facilitação do uso das técnicas de mineração de dados. Nesse contexto as utilizações de *Data Warehouses* expandiram-se consideravelmente, pois, nessas estruturas, as informações estão alocadas da maneira mais eficiente. *Data Warehouse* é um depósito central de dados extraído de dados operacionais, em que a informação é orientada a assuntos, não volátil e de natureza histórica [ADR96]. Devido a essas características, esses depósitos tendem a se tornar grandes repositórios de dados extremamente organizados, o que facilita a aplicação da mineração de dados.

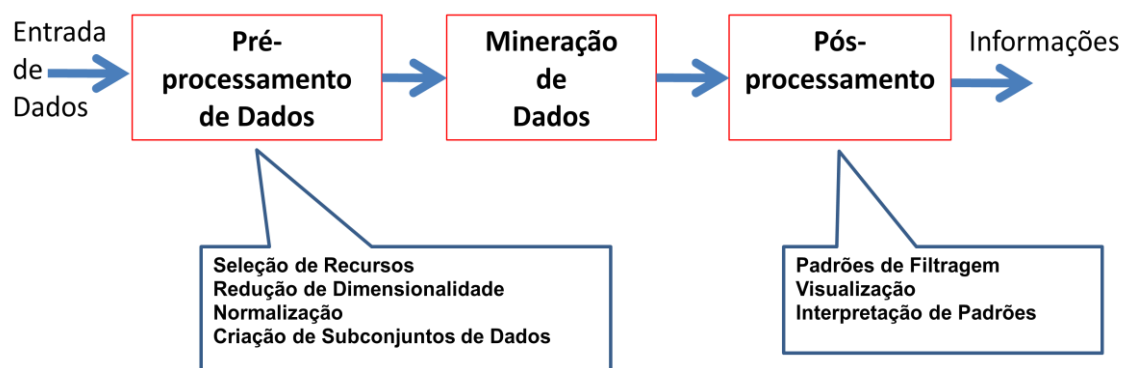


Figura 3.1 – Processo de Descoberta de Conhecimento em Banco de Dados, adaptado de [TAN09]

A etapa central do processo é a Mineração de Dados. Nessa etapa o objetivo principal é a aplicação de técnicas de mineração nos dados pré-processados, o que envolve o uso de métodos inteligentes de extração de padrões ou conhecimentos dos dados.

A etapa final do processo de KDD, que integra os resultados da mineração de dados com os sistemas de tomada de decisão, é o **Pós-processamento**. Esta etapa assegura que apenas resultados relevantes, válidos e úteis, sejam incorporados a esses sistemas. Medições estatísticas e métodos de teste de hipóteses também podem ser aplicados no pós-processamento para filtrar os resultados não legítimos da mineração de dados.

3.3 Mineração de Dados (*Data Mining*)

A mineração de dados é o processo de descoberta automática de informações úteis em grandes depósitos de dados [TAN09]. Parte integrante no processo de *KDD*, que é o processo geral de conversão de dados brutos em informações úteis. As técnicas de mineração de dados foram organizadas para atuar sobre grandes bases de dados com o intuito de descobrir padrões úteis que poderiam de outra forma, permanecer ignorados. Essas técnicas também têm a capacidade de prever o resultado em observações futuras.

A mineração de dados é uma tecnologia que surgiu da intersecção de quatro grandes áreas: estatística clássica, inteligência artificial, aprendizado de máquina e reconhecimento de padrões, conforme a Figura 3.2, além de uma grande quantidade de outras áreas. Em especial os sistemas de bancos de dados, que são necessários para fornecer suporte ao armazenamento, indexarem e processar consultas. Outras técnicas, como as de computação paralela, de alto desempenho, são muitas vezes importantes para tratar certos conjuntos de dados volumosos, assim como as técnicas de computação distribuídas que podem auxiliar a abordar a questão do tamanho e são essenciais quando os dados não podem ser juntados em um único local.

3.3.1 Tarefas de Mineração de Dados

As tarefas de mineração de dados são geralmente divididas em duas categorias principais: Tarefas Preditivas e Tarefas Descritivas. As tarefas preditivas buscam prever o valor de um determinado atributo baseado nos valores dos outros atributos (classificação e regressão). O atributo a ser previsto é conhecido como **variável dependente** ou **alvo**, enquanto que os atributos usados para realizar essa previsão são conhecidos por **variáveis independentes** ou **explicativas**.

As tarefas descritivas têm por objetivo derivar padrões (correlações, grupos, trajetórias e anomalias) que possam resumir os relacionamentos subjacentes nos dados.

Essas tarefas, na mineração de dados, são muitas vezes exploratórias em sua natureza e requerem técnicas de pós-processamento, freqüentemente, para validar e explicar os resultados encontrados.

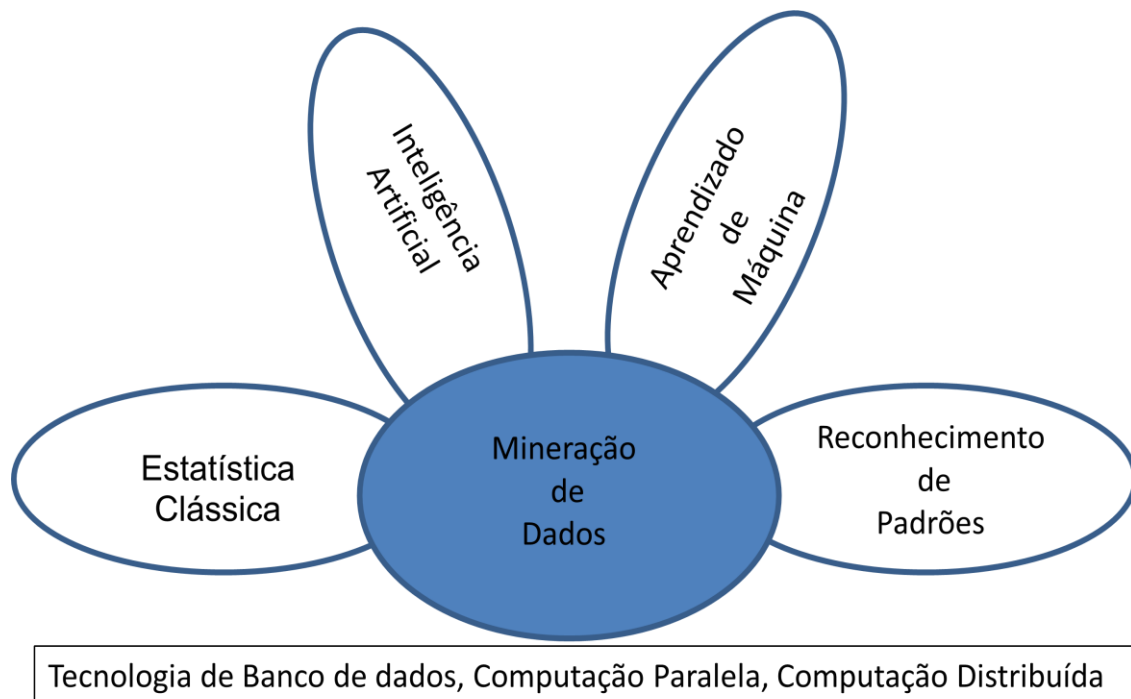


Figura 3.2 – Mineração de Dados como Intersecção de Muitas Áreas

Para encontrar respostas ou extrair conhecimentos interessantes, existem diversos métodos disponíveis na literatura. Contudo, para que o processo de mineração de dados seja relevante, é importante que estabeleçamos metas bem definidas. Podemos alcançar essas metas por meio dos métodos da mineração de dados. Segundo [TAN09] os métodos centrais da mineração de dados são: **Análise de grupos**, **Análise de Associação**, **Detecção de Anomalias** e **Modelagem de Previsão**. Essas técnicas se baseiam em métodos já existentes em outras áreas, como as de aprendizado de máquina, reconhecimento de padrões e estatística, indo desde as tradicionais estatísticas multivariadas, como análise de agrupamentos e regressões, até modelos atuais de aprendizagem, como redes neurais, lógica difusa e algoritmos genéticos.

A **análise de grupos** procura encontrar grupos de observações intimamente relacionadas, nos bancos de dados, de modo que observações que pertençam ao mesmo grupo sejam mais semelhantes entre si do que as pertencentes à outros grupos. Para

tanto, utiliza os algoritmos de K-means, algoritmo de agrupamento hierárquico aglomerativo e DBSCAN, por exemplo. As análises de grupos têm sido usadas para juntar conjuntos de dados de clientes relacionados, para compactar dados, para agrupar documentos, dentre outros.

As **análises de associação** são utilizadas para descrever padrões que apresentem características altamente associadas dentro dos dados analisados. A procura de associações entre variáveis é, freqüentemente, um dos propósitos das pesquisas que buscam dados relevantes e convenientes. O algoritmo FP-Grow é o utilizado para encontrar essa associação entre variáveis. Medidas estatísticas como correlação e teste de hipóteses apropriadas revelam a freqüência de uma regra no universo dos dados minerados. Aplicações úteis da análise por regras de associação incluem a descoberta de genes que possuam funcionalidade associada ou a identificação de páginas Web que sejam acessadas juntas, por exemplo.

A **detecção de anomalias** é uma tarefa que busca identificar observações cujas características sejam significativamente diferentes do resto dos dados minerados. O objetivo de um algoritmo que realize essa tarefa é descobrir as anomalias verdadeiras e evitar que se rotulem erroneamente objetos normais como anômalos. Ou seja, um bom algoritmo de detecção de anomalias deve ter uma alta taxa de detecção e uma baixa taxa de alarme falso. Como exemplos de aplicações dessa tarefa citam-se as que incluem a detecção de fraudes em cartões de crédito, intromissões numa dada rede de computadores e padrões incomuns de doenças.

A **modelagem de previsão (Regressão e Classificação)** se refere à tarefa de construir um modelo de previsão para a variável alvo como uma função das variáveis explicativas. As tarefas de modelagem de previsão se dividem em dois tipos, sendo eles a **regressão**, que é usada para variáveis alvo contínuas, e a **classificação**, usada para variáveis alvo do tipo nominal. Por exemplo, prever o preço futuro de um produto é tarefa de regressão, por que a variável alvo é do tipo contínua (preço). Já prever se um cliente vai ou não comprar determinado produto é uma tarefa de classificação, pois, a variável alvo é do tipo nominal (compra, Sim ou Não). Ambas as tarefas têm o objetivo de aprender um modelo que minimize o erro entre os valores previsto e real da variável alvo.

Classificação é a tarefa de aprender uma função alvo que mapeie cada conjunto de atributos x para um dos rótulos de classes y pré-determinados. Para isso podem ser utilizados classificadores baseados em árvores de decisão, baseado em regras, classificadores de vizinho mais próximo, classificadores Bayesianos, Rede Neural Artificial, Support Vector Machine (SVM) ou Método de Grupos por exemplo.

3.4 Árvore de Decisão

Os classificadores baseados em árvores de decisão são um dos ramos na área de inteligência artificial (AI). Pertencentes ao sub-campo de aprendizagem de máquina mais especificamente, pois, possuem habilidade de “aprender” através de registros em uma base de dados.

Para ilustrar como a classificação por árvore de decisão funciona, utiliza-se, neste trabalho, o exemplo do jogo de tênis, cujo problema central desse exemplo é determinar se haverá ou não, jogo partindo de três atributos predefinidos.

Suponha que um espectador deseje saber que haverá ou não jogo de tênis em seu clube num determinado dia. Uma forma de se descobrir é levantar uma série de questões sobre as características do dado dia. A primeira questão que foi levantada é como está o tempo nesse dia (ensolarado, nublado ou chuvoso). A próxima questão é a de como está a umidade desse dia (alta ou normal). E, por fim, foi levantada a questão da velocidade do vento (fraco ou forte). Com base nesses três atributos o espectador fez uma pesquisa que levou em consideração os estados desses atributos por quatorze dias e montou a Tabela 3.1

Tabela 3.1 – Tabela de Decisão do Jogo de Tênis

Condição				Ação
Dia	Tempo	Umidade	Vento	Jogo de Tênis
D1	Ensolarado	Alta	Fraco	NÃO
D2	Ensolarado	Alta	Forte	NÃO
D3	Nublado	Alta	Fraco	SIM

D4	Chuvoso	Alta	Fraco	SIM
D5	Chuvoso	Normal	Fraco	SIM
D6	Chuvoso	Normal	Forte	NÃO
D7	Nublado	Normal	Forte	SIM
D8	Ensolarado	Alta	Fraco	NÃO
D9	Ensolarado	Normal	Fraco	SIM
D10	Chuvoso	Normal	Fraco	SIM
D11	Ensolarado	Normal	Forte	SIM
D12	Nublado	Alta	Forte	SIM
D13	Nublado	Normal	Fraco	SIM
D14	Chuvoso	Alta	Forte	NÃO

A Tabela 3.1 mostra o resultado da pesquisa realizada por nosso espectador e ilustra como podemos resolver um problema de classificação fazendo uma série de questionamentos cuidadosamente organizados sobre os atributos do registro de teste. Cada vez que é recebida uma resposta, a questão seguinte é feita até que chega-se a uma conclusão sobre o rótulo da classe do registro. Essa série de questões e respostas possíveis pode ser organizada na forma de uma árvore de decisão, como sua estrutura hierárquica constituída de nodos e setas direcionadas. A Figura 3.3 ilustra a árvore de decisão para o problema do jogo de tênis. A árvore tem três tipos de nodos:

- Um **nodo raiz** não possui setas chegando. Mas possui duas ou mais setas saindo;
- **Nodos internos**, cada um com uma seta chegando e duas ou mais saindo;
- **Nodos folha** ou **terminais**, cada um dos quais possuindo exatamente uma seta chegando e nenhuma saindo.

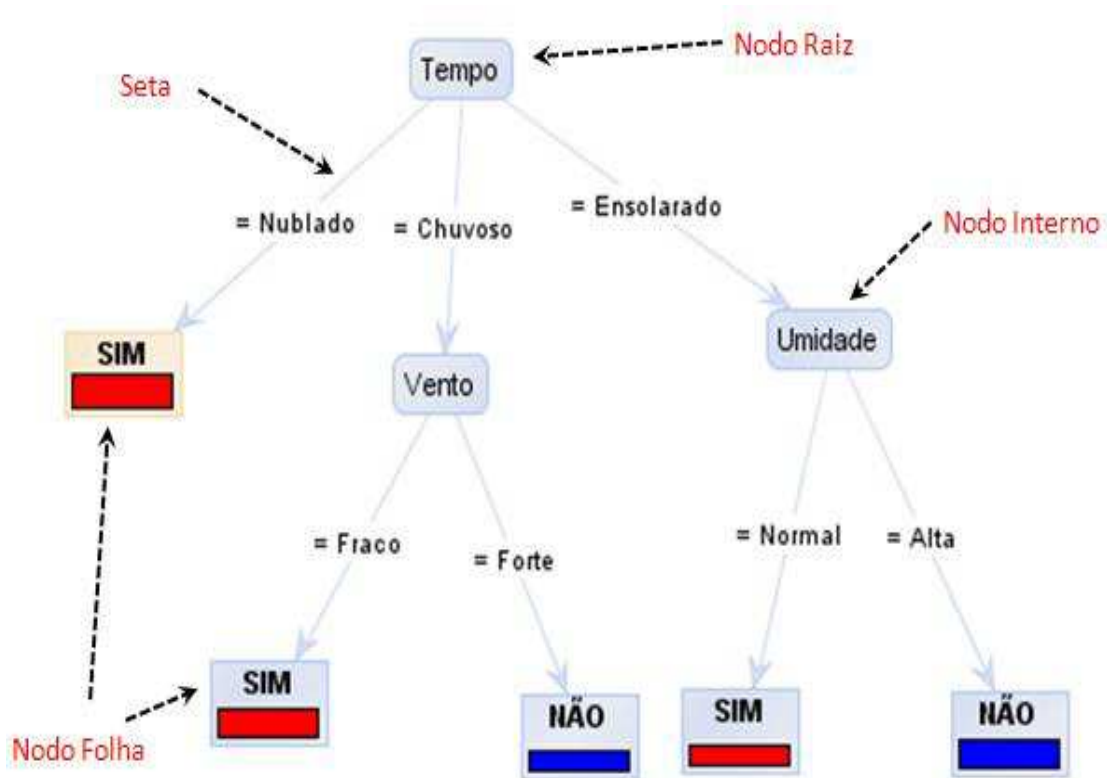


Figura 3.3 – Uma Árvore de Decisão Criada no *Software RapidMiner*, versão 4.2, Para o Problema do Jogo de Tênis.

Em uma árvore de decisão, cada nodo folha recebe um rótulo de classe. Os nodos não internos, isso incluem o nodo raiz e os nodos internos, contêm condições de testes de atributos para separar registros que possuam características diferentes. Por exemplo, o nodo raiz mostrado na figura 3.3 usa o atributo “tempo” para determinar se vai ou não haver jogo de tênis no clube, sendo que se este atributo for igual a nublado haverá jogo, em conformidade com o banco de dados de teste, chegando ao nodo folha. Contudo, se atributo “tempo” for igual a chuvoso ou ensolarado mais testes serão realizados para determinar o nodo raiz. Se o tempo for chuvoso então o teste deve ser feito no atributo vento, se este novo atributo apresentar o valor “forte” então não haverá jogo, caso contrário haverá. Por outro lado, se tempo for ensolarado o novo teste deve ser feito no atributo “umidade”, se este atributo for igual a “Normal” haverá jogo, caso contrário não haverá. Exibindo o problema do jogo de tênis em laços condicionais, resultaria:

```

INICIO
|   Se TEMPO for igual a NUBLADO, Então, SIM, haverá jogo
|   Se TEMPO for igual a CHUVOSO, Então
|   |   Se VENTO for igual a FRACO, Então, SIM, haverá jogo
|   |   Se VENTO for igual à FORTE, Então NÃO haverá jogo
|   Se TEMPO for igual a ENSOLARADO, Então
|   |   Se UMIDADE for igual à ALTA, Então NÃO haverá jogo
|   |   Se UMIDADE for igual à NORMAL, Então, SIM, haverá jogo
|   FIM SE
FIM

```

Uma vez criada à árvore de decisão, a tarefa de classificar um registro de teste é realizado de forma direta. Começando pelo nodo raiz, aplica-se a condição de teste ao registro e seguimos a ramificação apropriada baseados no resultado do teste. Isso pode levar tanto para o nodo folha, terminando e classificando o processo, como para um nodo interno, no qual pode-se aplicar uma nova condição de teste até alcançar um nodo folha. O rótulo da classe associada ao nodo folha é então atribuído ao registro.

Apesar do exemplo acima ter se mostrado preciso para todos os dados do registro de teste, as árvores de decisão não necessariamente precisam ser 100% precisas. Muitas árvores de decisão podem ser construídas a partir de um conjunto de atributos. Algumas mais precisas do que outras. Contudo, encontrar a árvore ótima é computacionalmente inviável por causa do tamanho exponencial do espaço de pesquisa. Apesar dessa dificuldade, muitos algoritmos eficientes têm sido desenvolvidos para induzir uma árvore de decisão razoavelmente precisa em uma quantidade razoável de processamento e tempo. Um desses algoritmos é o **algoritmo de Hunt**, que é a base de muitos algoritmos de indução de árvores de decisão existentes, incluindo ID3, C4.5 e CART.

3.4.1 O Algoritmo de Hunt

No algoritmo de Hunt uma árvore de decisão cresce de uma forma recursiva pelo particionamento dos registros de treino em sucessivos subconjuntos mais puros [TAN09]. Suponha-se que D_t seja o conjunto de registros de treino que está associado

ao nodo t e $y = \{y_1, y_2, y_3, \dots, y_c\}$ sejam os rótulos das classes. A seguir será apresentada uma definição recursiva do algoritmo de Hunt, referenciado por [TAN09].

- **Passo 1:** Se todos os registros em D_t pertencem à mesma classe y_t , então t é um nodo da folha rotulado como y_t .
- **Passo 2:** Se D_t contiver registros que pertençam a mais de uma classe uma **condição de teste atribuído** é selecionada para particionar os registros em subconjuntos menores. Um nodo filho é criado para cada resultado da condição de teste e os registros de D_t são distribuídos para os filhos baseados nos resultados. O algoritmo é então aplicado recursivamente a cada nodo filho.

Para ilustrar como o algoritmo de Hunt funciona, considerar-se o problema de prever se um solicitante de empréstimo quitará sua dívida ou se o mesmo se tornará inadimplente. Para esse problema apresenta-se abaixo, na Tabela 3.2, a Tabela de decisão do mesmo:

Tabela 3.2 – Conjunto de Treino Para Prever Tomadores de Empréstimos que se Tornarão Inadimplente.

Tid	Casa Própria	Estado Civil	Renda Anual	Inadimplente
1	Sim	Solteiro	125K	Não
2	Não	Casado	100K	Não
3	Não	Solteiro	70K	Não
4	Sim	Casado	120K	Não
5	Não	Divorciado	95K	Sim
6	Não	Casado	60K	Não
7	Sim	Divorciado	220K	Não
8	Não	Solteiro	85K	Sim
9	Não	Casado	75K	Não
10	Não	Solteiro	90K	Sim

A árvore inicial, da figura 3.4(a), apresenta um único nodo como rótulo de classe, Inadimplente = Não, isso significa que a maioria dos tomadores de empréstimos pagará seus débitos. Como no nodo raiz há registros de ambas as classes, Inadimplente = Não e Inadimplente = Sim, os registros são subseqüentemente divididos em

subconjuntos menores baseados nos resultados das outras condições. Como o atributo “Casa Própria” somente pode apresentar dois valores distintos, Sim ou Não, essa divisão tem início nesse atributo, Figura 3.4(b). O algoritmo de Hunt é então aplicado recursivamente a cada filho do nó raiz. Pela observação da Tabela 3.2 nota-se que todos os tomadores de empréstimos que possuem casa própria pagaram suas dívidas. Logo, o filho à esquerda do nó raiz é um nó folha rotulado como Inadimplente = Não, (Figura 3.4(b)). Para o outro filho é necessário continuar aplicando o passo recursivo do algoritmo de Hunt até que todos os registros pertençam à mesma classe. As árvores resultantes de cada passo recursivo são mostradas nas Figuras 3.4(c) e 3.4(d).

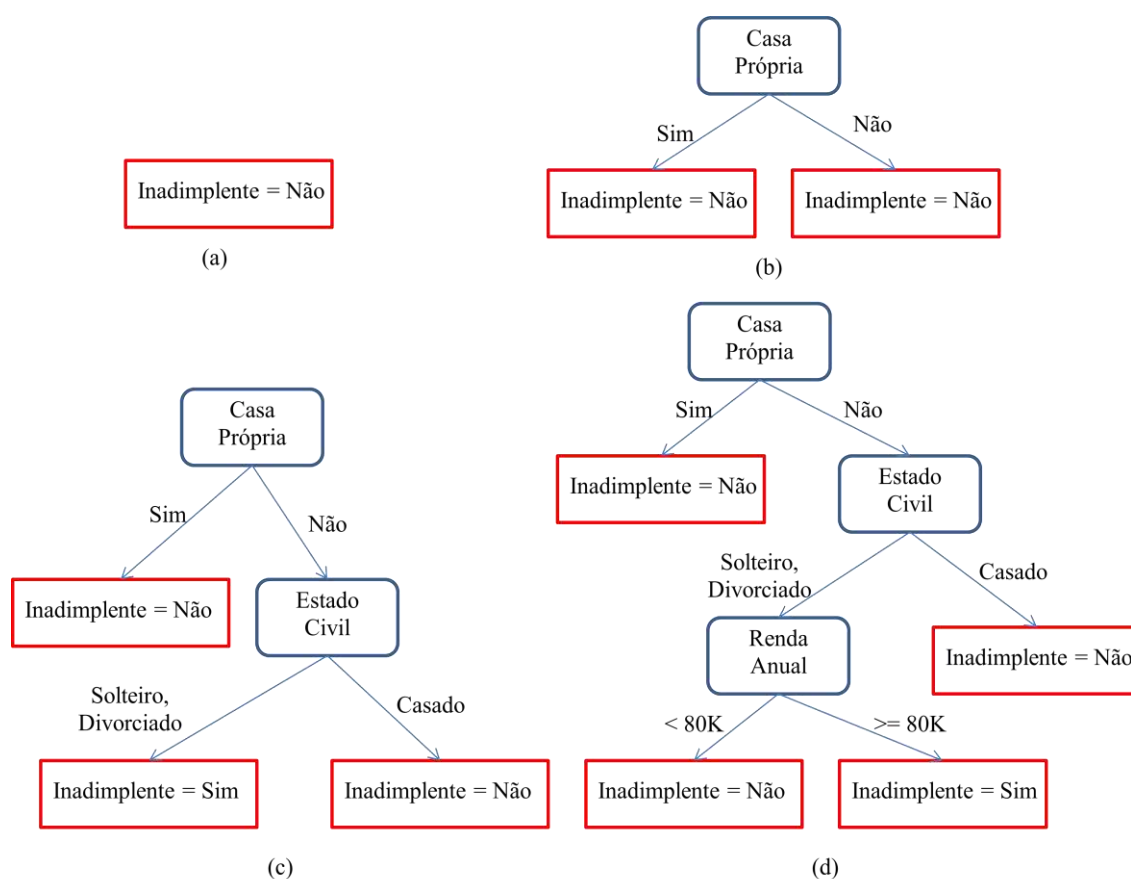


Figura 3.4 – Algoritmo de Hunt Para Induzir Árvores de Decisão

O algoritmo funcionará se toda combinação de valores de atributos estiver presente nos dados de treino e cada combinação possui um único rótulo de classe. Algumas condições especiais podem acontecer levando a criação de ações adicionais [TAN09]. Por exemplo:

- É possível que alguns dos nodos filhos criados no Passo 2 estejam vazios, ou seja, não existam registros associados a estes nodos. Isto pode acontecer se nenhum dos registros de treino tiver a combinação de valores de atributos associada a tais nodos. Neste caso, o nodo é declarado um nodo folha com o mesmo rótulo de classe da classe majoritária dos registros de treino associados ao seu nodo pai;
- No Passo 2, se todos os registros associados a D_t possuírem valores atribuídos idênticos, exceto pelo rótulo da classe. Então não é possível dividir mais estes registros. Neste caso, o nodo é declarado um nodo folha com o mesmo rótulo de classe da classe majoritária dos registros de treino associados a este nodo.

Capítulo 4 - Estudo de Caso

4.1 Introdução

Diversas ferramentas computacionais vêm sendo desenvolvidas para avaliar a segurança operacional de sistemas de energia elétrica, as quais representam adequadamente a modelagem dos componentes que compõem o mesmo. No entanto, o desenvolvimento de métodos para avaliação da segurança que possibilitem a operação em tempo real ainda é um grande desafio, devido à alta não-linearidade e alta dimensionalidade dos grandes sistemas de potência. A utilização de algoritmos de inteligência computacional se encontra em constante crescimento e atualização, e podem fornecer uma solução promissora no desenvolvimento de ferramentas de avaliação da segurança de sistemas de potência fornecendo rápidos resultados de predição em casos de contingências críticas, auxiliando a operação em tempo real, no sentido de evitar falhas em cascatas e interrupção de energia de longo prazo para consumidores de uma grande área, como os blackouts.

Dentre os algoritmos de inteligência computacional, se destacam aqueles que avaliam o comportamento do sistema no passado para, assim, tentar prever novos métodos de controle e de tomada de decisão no futuro, com o menor erro de classificação possível sobre os estados em que o sistema venha a atuar para evitar sua operação fora de certos padrões de qualidade determinados por órgãos competentes. Uma dessas ferramentas é a árvore de decisão, técnica de classificação utilizada para descobrir conhecimentos em bases de dados.

Para que se construa uma árvore de decisão, sobre qualquer processo que se deseje estudar, deve-se ter um banco de dados de entrada, para a construção da árvore, e outro banco de dados para avaliar a árvore criada, de forma que se obtenha o erro de classificação da árvore, que deve estar em uma faixa aceitável, para que, posteriormente, uma ferramenta que tenha incorporada no seu código fonte a árvore de decisão criada, com erro dentro da faixa, possa ser integrada, no caso do presente trabalho, no ambiente operacional de tempo real das concessionárias de energia elétrica,

para auxílio aos operadores e engenheiros de planejamento, na tomada de decisão a respeito da segurança do sistema elétrico em operação.

Os sistemas de potência operam dentro de condições consideradas seguras quase que na sua totalidade de tempo. Assim, conseguir uma base de dados, suficientemente grande, que leve em consideração todos os possíveis estados operacionais dentro de todas as possíveis contingências, previsíveis ou não, a que o sistema está sujeito, podem demandar anos e anos de coleta de dados. Tendo em vista esse problema, foi necessária a criação de uma ferramenta computacional que simule a operação do sistema dentro de todas essas possíveis condições e que avalie o estado em que o sistema está operando (seguro ou inseguro), resultando em bases de dados extensas, que possam ser utilizadas nas ferramentas de criação de árvores de decisão.

Para a criação da árvore de decisão foi utilizado o software RapidMiner que, além das árvores de decisão, pode utilizar todas as tarefas de mineração de dados, com seus respectivos algoritmos. A figura 4.1 apresenta o fluxograma desse processo.

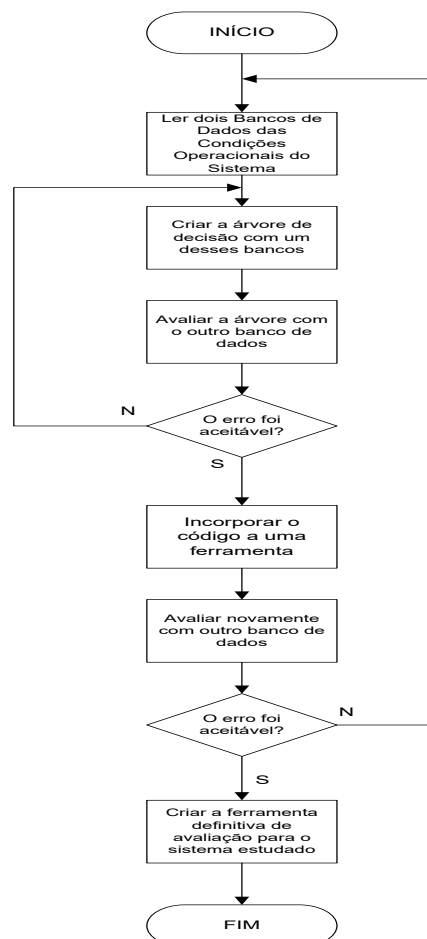


Figura 4.1 – Fluxograma do Processo de Criação e Avaliação da Árvore de Decisão

4.2 Sistema Elétrico de Potência Estudado

Para exemplificar a aplicação de árvores de decisão na avaliação da segurança estática de sistemas de potência foi utilizado um sistema clássico no estudo de estabilidade, trata-se do sistema de 9 barras, das quais 3 são barras de geração e as 6 restantes são barras de carga, além de constar com 6 linhas de transmissão e 3 transformadores, como apresentado na figura 4.2. O sistema é representado aqui por seu diagrama de impedâncias de seqüência positiva (unifilar), o qual é utilizado para formar os dados de entrada das barras, e dos ramos (linhas de transmissão e transformadores), a fim de resolver o problema do fluxo de potência dentro das inúmeras situações que, também, são consideradas. Os formatos desses dados de entrada serão explicados mais adiante, tais como todos os procedimentos que levaram a criação do banco de dados, o qual será utilizado para a construção da árvore de decisão para esse sistema elétrico.

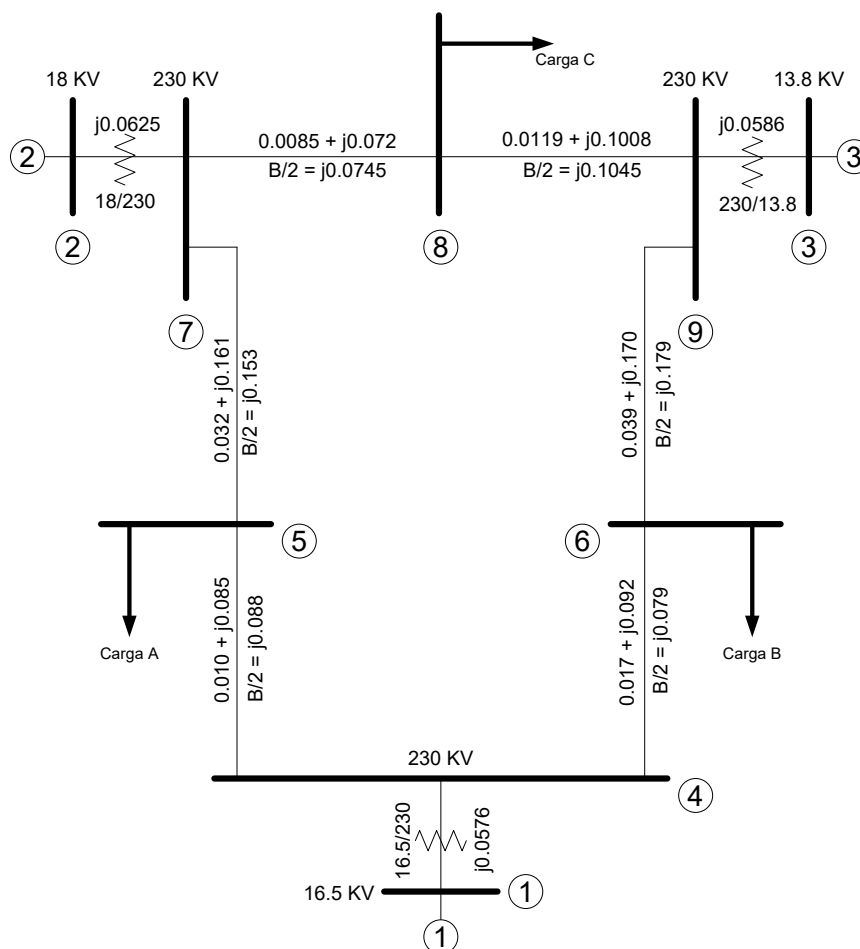


Figura 4.2 – Diagrama de Impedâncias do Sistema de nove barras, todas as impedâncias estão em p.u. na base 100 MVA, Adaptado de [AND03]

4.3 Ferramenta Para Criação do Banco de Dados das Condições Operacionais do Sistema de Potência

No processo de construção dos bancos de dados, foi desenvolvida uma ferramenta computacional que resolve o problema do fluxo de carga para os diferentes cenários de operação que o sistema de potência em estudo pode assumir, levando em consideração tanto as variações de carga quanto algumas contingências simples (N-1), como a saída de uma linha ou de uma geração, e algumas contingências múltiplas (N-2), como a saída de uma linha e de uma geração, por exemplo. Para tanto, a ferramenta em questão, desenvolvida e nomeada no presente trabalho de “FluxMining”, utiliza o método de Newton-Raphson para solucionar o fluxo de carga em todos esses cenários de operação e levando em conta os limites operacionais das barras (limites superior e inferior da magnitude da tensão, da injeção de potências ativa e reativa geradas) e além dos limites das linhas transmissão e transformadores (limites térmicos) para determinar o estado do sistema, estando este operando de forma segura ou insegura.

O FluxMining foi desenvolvido utilizando o software MatLab. Por sua facilidade em lidar com operações matemáticas e matriciais e sua ampla utilização no meio acadêmico o MatLab tornou-se a linguagem apropriada para o desenvolvimento dessa ferramenta.

O processo para obtenção do banco de dados dos estados operacionais dos sistemas tem início com a leitura dados das barras e das linhas de transmissão e transformadores do sistema. Posteriormente, é lido o banco de dados das contingências das linhas e o usuário deve entrar com o valor da variação das cargas, em porcentagem, e com o número de situações de carga diferentes, dentro da variação desejada, que o sistema deve operar. Numa próxima etapa o usuário deve escolher quais das variáveis que o banco de dados de saída, deve conter como atributo objetivo, tal como, se o estado do sistema deve ser apresentado como seguro e inseguro, ou se deve conter, quando inseguro, a causa da insegurança do sistema, podendo ser VMT (violação da magnitude de tensão nas barras), VLT (violação do limite térmico nas linhas de transmissão e transformadores ou das potências ativas e reativas geradas), VMT e VLT ou EET (estabilidade estática de tensão). O fluxograma do processo pode ser visto na figura 4.3.

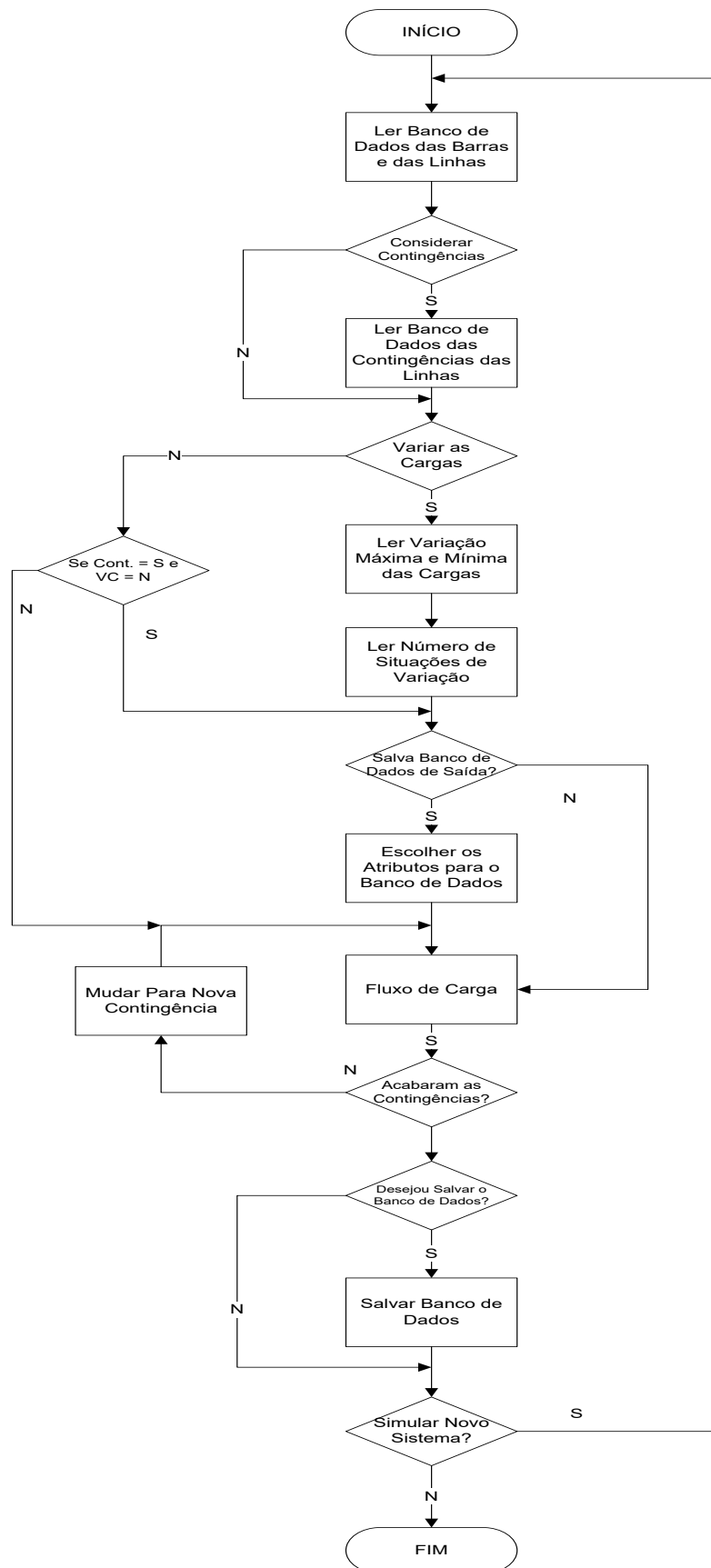


Figura 4.3 – Fluxograma da Ferramenta de Criação do Banco de Dados das Condições Operacionais do Sistema

A estrutura dos registros dos bancos de dados das barras, das linhas de transmissão e transformadores, das linhas de transmissão que serão retiradas (contingências), e dos limites operacionais das barras e das linhas, se apresenta conforme abaixo:

- Banco de Dados das Barras: Possuem 9 atributos para cada barra do sistema sendo elas:
 1. Número da barra;
 2. Tipo da barra: 2 para barra de referência ($V\theta$);
1 para barra de geração (PV); ou
0 para barra de carga (PQ);
 3. Tensão na barra, em p.u., sendo que se a barra for uma barra de carga, o valor desse campo será o utilizado inicialmente para a resolução do fluxo de carga;
 4. Ângulo da barra, em graus, sendo que se a barra for de geração (tensão controlada, PV) ou de carga, o valor desse campo será o utilizado inicialmente para a resolução do fluxo de carga;
 5. Potência ativa gerada pela barra, em p.u., sendo que na barra de referência e nas barras de carga, este campo deve ter valor igual a zero;
 6. Potência reativa gerada pela barra, em p.u., este campo deve ser igual a zero para qualquer tipo de barra, pois as potências reativas são geradas conforme a necessidade das barras de carga;
 7. Potência ativa consumida pela carga da barra, em p.u.;
 8. Potência reativa consumida pela carga da barra, em p.u.; e
 9. Susceptância Shunt da barra (em p.u.).

- Banco de dados das Linhas de Transmissão e Transformadores: Possuem 7 atributos para cada linha do sistema, sendo eles:
 1. Número da Barra de saída da linha de transmissão ou transformador;
 2. Número da Barra de chegada da linha de transmissão ou transformador;
 3. Resistência, em p.u., da linha de transmissão ou transformador;
 4. Reatância, em p.u., da linha de transmissão ou transformador;

5. Susceptância Shunt, em p.u., da linha de transmissão ou transformador;
 6. TAP do transformador elevador ou abaixador, se a linha não contiver um transformador desse tipo o valor desse atributo deve ser igual a zero ou a um; e
 7. Ângulo de defasagem do transformador, se a linha não contiver um transformador defasador o valor desse atributo deve ser igual a zero.
- Banco de dados das contingências, retirada de linhas de transmissão: Nesse banco de dados cuidados especiais devem ser tomados, tais como: não retirar um trecho que contenha um transformador e, também, não deixar que o sistema de potência entre na chamada condição de “ilhamento”, ou seja, garantir que nenhuma barra ou um conjunto de barras, fique isolado do restante. Apresenta apenas 2 atributos, sendo:
 1. Número da barra de saída da linha de transmissão; e
 2. Número da barra de chegada da linha de transmissão.
 - Banco de dados dos limites operacionais das Barras: Possuem 7 atributos para cada barra do sistema, sendo eles:
 1. Número da barra;
 2. Maior valor da magnitude de tensão na barra;
 3. Menor valor da magnitude de tensão na barra;
 4. Maior valor da potência reativa na barra;
 5. Menor valor da potência reativa na barra;
 6. Maior valor da potência ativa na barra; e
 7. Menor valor da potência reativa na barra.
 - Banco de dados dos limites operacionais das linhas de transmissão e transformadores: possuem 3 atributos, sendo:
 1. Número da barra de saída da linha;
 2. Número da barra de chegada da linha; e
 3. Limite térmico, potência aparente, na linha ou transformador.

O FluxMining permite como entrada tanto bancos de dados externos, no formato “txt”, desde que padronizados conforme as estruturas dadas acima, quanto a criação desses bancos em tempo de execução, facilitando assim a utilização da ferramenta por usuários que desconhecem as estruturas dos bancos de dados ou que tenham, externamente, apenas alguns desse bancos citados, podendo criar os registros restantes com a ferramenta em operação. Além de ser possível converter os bancos de dados das barras e das linhas dos dados utilizados pela ferramenta *lfdemo*, do pacote *PSTv3 (Power System Toolbox, Version 3)*, um pacote de ferramentas com aplicações nos estudos de fluxo de carga e estabilidade de sistemas de energia elétrica, validado pela comunidade acadêmica.

Os registros de saída gerados pelo FluxMining podem apresentar diversos grupos de atributos em sua estrutura que podem formar o banco de dados para a criação da árvore, sendo eles:

1. Magnitudes das Tensões (V) em cada barra V_{θ} e PV;
2. Ângulos de Fase da Tensão (θ) em cada barra V_{θ} e PV;
3. Potências ativas geradas (P_g) em cada barra V_{θ} e PV;
4. Potências reativas geradas (Q_g) em cada barra do tipo V_{θ} e PV;
5. Potencia ativa da linha que liga a barra "i" a barra "j", P_{ij} ;
6. Potencia reativa da linha que liga a barra "i" a barra "j", Q_{ij} ;
7. Potencia ativa da linha que liga a barra "j" a barra "i", P_{ji} ;
8. Potencia reativa da linha que liga a barra "j" a barra "i", Q_{ji} ;
9. Diferença angular entre as barras "i" e "j", θ_{ij} ;
10. Magnitudes da Tensão nas barras de carga, V_c ;
11. Somatório das potências ativas das cargas do sistema, P_{cargas} ; e
12. Somatório das potências reativas das cargas do sistema, Q_{cargas} .

Além desses grupos de atributos, o atributo objetivo final será, obrigatoriamente, o estado do sistema que pode ser definido, pelo usuário, de duas maneiras. Na primeira ele será definido apenas por seguro ou inseguro, e na segunda será definido por seguro ou o motivo que o levou a insegurança (VMT, VLT, VMT e VLT ou EET), conforme já citado.

Por causa do grande número de barras e linhas que os sistemas de potência geralmente contêm, o banco de dados de saída tende a apresentar várias colunas,

- Banco de dados das linhas de transmissão e transformadores:

Tabela 4.2 – Banco de Dados das Linhas do Sistema Exemplo de 9 Barras

Da Bar	P/ Bar	Resistência	Reatância	Susc. Shunt	TAP	δ
1	4	0.0	0.0576	0.0	1.0	0.0
4	5	0.01	0.085	0.0880	1.0	0.0
5	7	0.032	0.161	0.1530	1.0	0.0
4	6	0.017	0.092	0.0790	1.0	0.0
6	9	0.039	0.17	0.1790	1.0	0.0
7	8	0.0085	0.072	0.0745	1.0	0.0
3	9	0.0	0.0586	0.0	1.0	0.0
8	9	0.0119	0.1008	0.1045	1.0	0.0
2	7	0.00	0.0625	0.0	1.0	0.0

- Banco de dados das linhas que serão retiradas: Para que o sistema não tenha ilhamentos não são retiradas os trechos 1-4, 3-9 e 2-7:

Tabela 4.3 – Banco de Dados das Linhas que Podem ser Retiradas do Sistema

Da Barra	Para Barra
4	5
5	7
4	6
6	9
7	8
8	9

- Banco de dados dos limites operacionais das barras: Esse banco de dados foi retirado do pacote *matpower4.0b2 (A Matlab Power System Simulation Package)*, um pacote formado por um conjunto de ferramentas utilizadas pelos acadêmicos em estudos de fluxo de carga de sistemas de energia elétrica. Nas barras de carga os limites superiores e inferiores das potências ativas e reativas geradas foram mantidos iguais a 999 e a -999, respectivamente. Os bancos de dados são mostrados abaixo:

Tabela 4.4 – Banco de Dados dos Limites Operacionais das Barras do Sistema

Nº Bar	V _{MAX}	V _{MIN}	Q _{MAX}	Q _{MIN}	P _{MAX}	P _{MIN}
1	1.1	0.9	3.0	-3.0	2.50	0.010
2	1.1	0.9	3.0	-3.0	3.00	0.010
3	1.1	0.9	3.0	-3.0	2.70	0.010
4	1.1	0.9	999	-999	999	-999
5	1.1	0.9	999	-999	999	-999
6	1.1	0.9	999	-999	999	-999
7	1.1	0.9	999	-999	999	-999
8	1.1	0.9	999	-999	999	-999

- Banco de dados dos limites operacionais das linhas de transmissão e transformadores: Esse banco de dados também foi retirado do pacote *matpower4.0b2* e se apresenta conforme abaixo:

Tabela 4.5 – Banco de Dados dos Limites Operacionais das Linhas do Sistema

Da Barra	Para a Barra	S _{MAX}
1	4	2.50
4	5	2.50
5	6	1.50
3	6	3.00
6	7	1.50
7	8	2.50
8	2	2.50
8	9	2.50
9	4	2.50

Em um primeiro momento foi rodado apenas o fluxo de carga nas condições normais do dado sistema no FluxMining e comparada a resposta desse fluxo de carga com a resposta dada por um software considerado válido pela comunidade acadêmica, neste caso utilizou-se o *lfdemo* do pacote *PSTv3*, a fim de validar os valores encontrados pelo FluxMining. Nas figuras 4.4 e 4.5 são apresentadas as respostas do

fluxo de carga para o sistema de 9 barras em estudo, tanto o encontrado pelo FluxMining quanto o encontrado pelo lfdemo, respectivamente.

Pela observação das figuras pode-se observar que todos os valores do relatório, sejam eles nas barras ou nas linhas, tanto do FluxMining quanto do lfdemo, são iguais. Logo, esse procedimento, valida os resultados encontrados pela ferramenta desenvolvida para que assim, se possa ter garantia de precisão nos resultados obtidos num estudo mais amplo do sistema de potência em questão, que leve em consideração várias das condições em que o sistema estudado pode operar, tais como variações de carga, ocorrências de contingências nas linhas de transmissão, saída de uma geração, dentre outras.

Tendo em vista que a ferramenta desenvolvida (FluxMining) apresenta os valores esperados na resolução do problema do fluxo de carga, foi realizado o teste mais amplo que levou em consideração todas as ocorrências citadas anteriormente.

DADOS DE BARRA						
Barra	Tensão	Ângulo	Pgen	Qgen	Pcarga	Qcarga
1.0000	1.0400	0	0.7164	0.2705	0	0
2.0000	1.0250	9.2800	1.6300	0.0665	0	0
3.0000	1.0250	4.6648	0.8500	-0.1086	0	0
4.0000	1.0258	-2.2168	0	0	0	0
5.0000	0.9956	-3.9888	0	0	1.2500	0.5000
6.0000	1.0127	-3.6874	0	0	0.9000	0.3000
7.0000	1.0258	3.7197	0	0	0	0
8.0000	1.0159	0.7275	0	0	1.0000	0.3500
9.0000	1.0324	1.9667	0	0	0	0

DADOS DE LINHA				
D/Barra	P/Barra	Plinha	Qlinha	S_Aparente
1.0000	4.0000	0.7164	0.2705	0.7658
4.0000	5.0000	0.4094	0.2289	0.4690
5.0000	7.0000	-0.8432	-0.1131	0.8508
4.0000	6.0000	0.3070	0.0103	0.3072
6.0000	9.0000	-0.5946	-0.1346	0.6097
7.0000	8.0000	0.7638	-0.0080	0.7638
3.0000	9.0000	0.8500	-0.1086	0.8569
8.0000	9.0000	-0.2410	-0.2430	0.3422
2.0000	7.0000	1.6300	0.0665	1.6314

Figura 4.4 – Resposta do FluxMining ao fluxo de carga para o sistema de 9 barras

BUS	VOLTS	ANGLE	GENERATION		LOAD	
			REAL	REACTIVE	REAL	REACTIVE
1.0000	1.0400	0	0.7164	0.2705	0	0
2.0000	1.0250	9.2800	1.6300	0.0665	0	0
3.0000	1.0250	4.6648	0.8500	-0.1086	0	0
4.0000	1.0258	-2.2168	0	0	-0.0000	-0.0000
5.0000	0.9956	-3.9888	0	0	1.2500	0.5000
6.0000	1.0127	-3.6874	0	0	0.9000	0.3000
7.0000	1.0258	3.7197	0	0	0.0000	-0.0000
8.0000	1.0159	0.7275	0	0	1.0000	0.3500
9.0000	1.0324	1.9667	0	0	0.0000	-0.0000

LINE	FROM BUS	TO BUS	LINE FLOWS	
			REAL	REACTIVE
1.0000	1.0000	4.0000	0.7164	0.2705
2.0000	4.0000	5.0000	0.4094	0.2289
3.0000	5.0000	7.0000	-0.8432	-0.1131
4.0000	4.0000	6.0000	0.3070	0.0103
5.0000	6.0000	9.0000	-0.5946	-0.1346
6.0000	7.0000	8.0000	0.7638	-0.0080
7.0000	3.0000	9.0000	0.8500	-0.1086
8.0000	8.0000	9.0000	-0.2410	-0.2430
9.0000	2.0000	7.0000	1.6300	0.0665

Figura 4.5 – Resposta do lfdemo ao fluxo de carga para o sistema de 9 barras

Considerou-se que as variações máximas e mínimas em todas as barras foram de 25% do valor nominal e também foram consideradas 27 situações de cargas diferentes dentro dessa variação, ou seja, 13 situações acima, 13 situações abaixo do valor nominal das cargas e o próprio valor nominal (caso base) das cargas. Além de 10 outras situações diferentes de potência ativa gerada nas barras PV.

Foi obtido um banco de dados de saída, o mesmo que será utilizado no treino e avaliação da árvore de decisão, por meio do software RapidMiner, com 10395 registros distintos e, utilizando nesse banco de dados as opções de grupos de atributos 3, 5, 6 e 9 citadas anteriormente, sendo as potências ativas geradas, as potências ativas nas linhas, as potências reativas nas linhas e as diferenças angulares de tensão entre as barras, respectivamente, totalizando um banco de dados com 31 atributos referente ao sistema elétrico de 9 barras e 3 geradores, sendo que desses, 30 são atributos independentes e o restante é o atributo alvo.

4.4 Criação e Avaliação da Árvore de Decisão

4.4.1 Criação e Avaliação da Árvore de Decisão Utilizando a Mesma Base de Dados

Na etapa de criação e avaliação do banco de dados criado pelo FluxMining foi utilizado um software de mineração de dados e aprendizagem de máquinas chamado de RapidMiner, uma ferramenta que apresenta mais de 500 operadores para realização das tarefas de mineração de dados.

Para criar e avaliar previamente a árvore de decisão o RapidMiner faz uso de alguns operadores, são eles o Root, Example Source, Simple Validation, Decision Tree, Applier Chain, Test e Classification Performance, dispostos da forma apresentada pela figura 4.6. Cada um desses operadores possui uma função específica dentro do processo, conforme explicado:

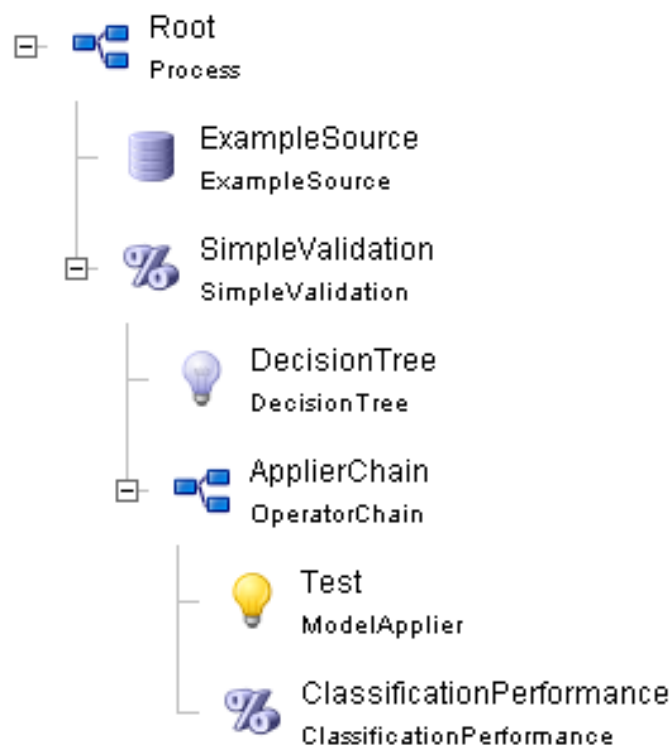


Figura 4.6 – Processo de Obtenção de Árvore de Decisão Pelo RapidMiner

- **Root:** Este operador é o corpo principal dos processos que serão realizados;
- **Example Source:** Neste operador é definido o banco de dados de entrada com suas características, tais como a forma como os atributos estão separados (por vírgula, ponto-e-vírgula, espaço ou por tabulação), se a primeira linha do banco de dados é o nome dos atributos ou não, o tipo de cada atributo (nominal, numérico, inteiro, real, string, binominal ou polinomial) e qual desses atributos será o atributo alvo (para o RapidMiner criar a árvore é obrigatório que o atributo alvo seja do tipo nominal, quando a tarefa é classificação);
- **Simple Validation:** Esse operador divide o banco de dados de entrada em duas partes, uma que será usada para treinar a árvore e outra que será usada para testar a árvore. Essa quebra é feita uniformemente. No caso do banco de dados do sistema em estudo, essa quebra foi de 70% do banco de dados para treino e 30% do banco para teste. Assim, no caso do estado do sistema ser VMT, o Simple Validation contou todos os registros que contêm essa variável e separou 70% delas, ou algo em torno disso, para treinar essa condição e o restante para validar;
- **Decision Tree:** Esse é o operador que ensina a árvore de decisão. Para tanto ele utiliza um algoritmo semelhante ao C4.5 ou CART, que são algoritmos baseados no algoritmo de Hunt. Como já citado, esse operador exige que o atributo alvo seja do tipo nominal. Aplicando o banco de dados do problema proposto nesse operador obteve-se a árvore da figura 4.6 e as regras da referida árvore que serão utilizadas para a criação do sistema de avaliação final do modelo;
- **Applier Chain:** Esse é o operador que funciona como uma sub-rotina, ou um sub-processo dentro do processo principal, Root;
- **Test:** Esse operador utiliza o restante do banco de dados, neste caso, os 30% que não foram utilizados na etapa de treino da árvore de decisão criada;
- **Classification Performance:** Esse operador apresenta a matriz de confusão, matriz essa resultante da aplicação de um banco de dados no modelo da árvore criada, com seus respectivos erros de classificação, conforme a figura 4.7 mostra.

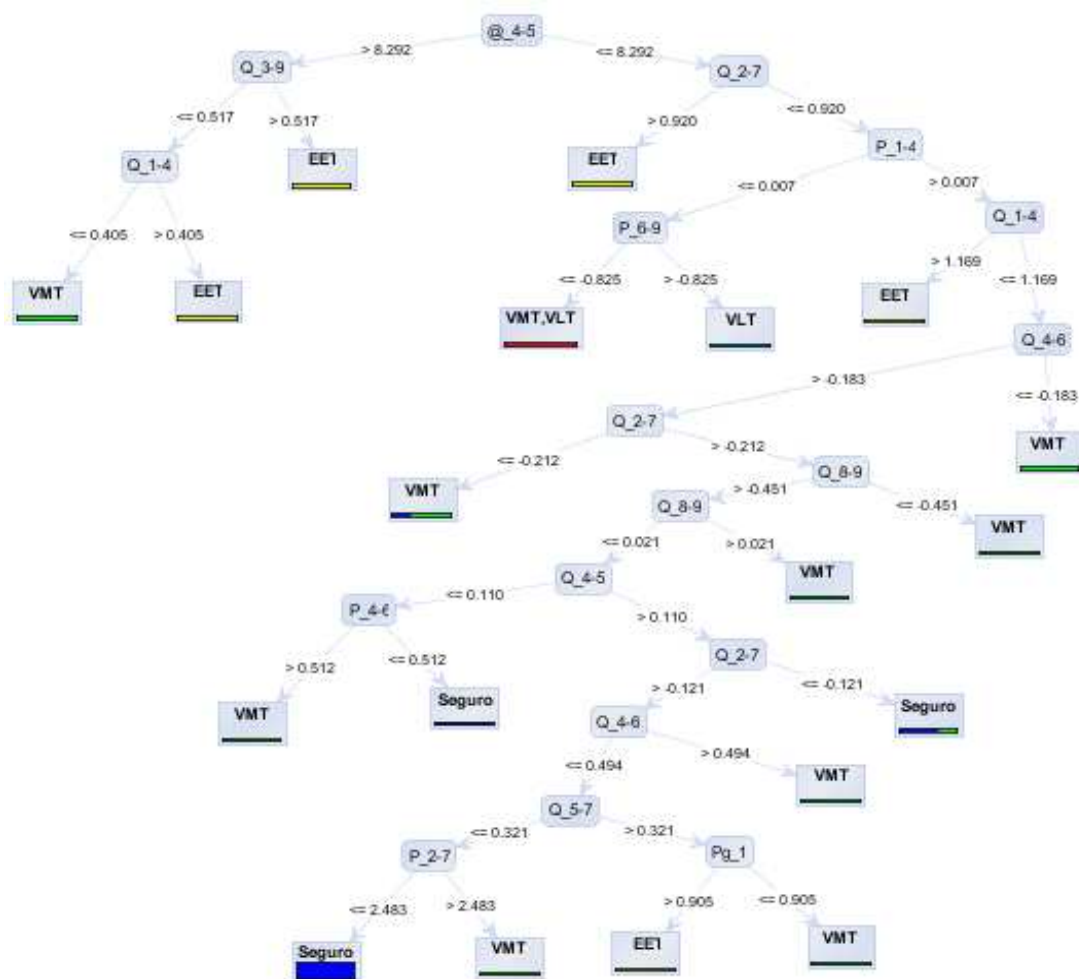


Figura 4.7 – Árvore de Decisão Criada Para o Banco de Dados de Entrada

classification_error: 3.21%						
	true Seguro	true VLT	true VMT	true EET	true VMT,VLT	class precision
pred. Seguro	2037	0	58	0	0	97.23%
pred. VLT	0	86	0	0	3	96.63%
pred. VMT	35	0	453	0	0	92.83%
pred. EET	0	0	4	436	0	98.09%
pred. VMT,VLT	0	0	0	0	7	100.00%
class recall	98.31%	100.00%	87.96%	100.00%	70.00%	

Figura 4.8 – Matriz de Confusão Para o Modelo de Árvore Proposto

A árvore criada apresentou um erro de classificação de 3.21%, como pode ser observado na figura 4.8 da matriz de confusão. O erro apresentado é considerado baixo, podendo assim considerar que o modelo dessa árvore de decisão pode vir a ser o definitivo para a avaliação da segurança estática para o sistema específico em estudo. Para avaliar, definitivamente, o modelo da árvore de decisão criado é necessário o

desenvolvimento de uma ferramenta que tenha incorporado ao seu código-fonte o algoritmo gerado pelo RapidMiner, juntamente com a árvore de decisão, esse algoritmo pode ser visto na figura 4.9 abaixo:

```

@_4-5 <= 8.292
|_ Q_2-7 <= 0.920
| | P_1-4 <= 0.007
| | | P_6-9 <= -0.825: VMT,VLT
| | | P_6-9 > -0.825: VLT
| | P_1-4 > 0.007
| | | Q_1-4 <= 1.169
| | | | Q_4-6 <= -0.183: VMT
| | | | Q_4-6 > -0.183
| | | | | Q_2-7 <= -0.212: VMT
| | | | | Q_2-7 > -0.212
| | | | | | Q_8-9 <= -0.451: VMT
| | | | | | Q_8-9 > -0.451
| | | | | | | Q_8-9 <= 0.021
| | | | | | | | Q_4-5 <= 0.110
| | | | | | | | | P_4-6 <= 0.512: Seguro
| | | | | | | | | P_4-6 > 0.512: VMT
| | | | | | | | Q_4-5 > 0.110
| | | | | | | | | Q_2-7 <= -0.121: Seguro
| | | | | | | | | Q_2-7 > -0.121
| | | | | | | | | | Q_4-6 <= 0.494
| | | | | | | | | | | Q_5-7 <= 0.321
| | | | | | | | | | | | P_2-7 <= 2.483: Seguro
| | | | | | | | | | | | P_2-7 > 2.483: VMT
| | | | | | | | | | | | Q_5-7 > 0.321
| | | | | | | | | | | | | Pg_1 <= 0.905: VMT
| | | | | | | | | | | | | Pg_1 > 0.905: EET
| | | | | | | | | | | | | Q_4-6 > 0.494: VMT
| | | | | | | | | | | | | Q_8-9 > 0.021: VMT
| | | | | | | | | | | | | Q_1-4 > 1.169: EET
| | | | | | | | | | | | | Q_2-7 > 0.920: EET
@_4-5 > 8.292
|_ Q_3-9 <= 0.517
| | Q_1-4 <= 0.405: VMT
| | | Q_1-4 > 0.405: EET
| | Q_3-9 > 0.517: EET

```

Figura 4.9 – Algoritmo da Árvore de Decisão Para o Sistema de 9 barras

4.4.2 Criação e Avaliação da Árvore de Decisão Utilizando Duas Bases de Dados Distintas

Outra forma de se avaliar a árvore de decisão, no próprio RapidMiner, é utilizando um banco de dados diferente daquele utilizado na etapa de treino.

Para essa outra avaliação os operadores são um pouco modificados em relação aos apresentados na figura 4.6, originando a figura 4.10.

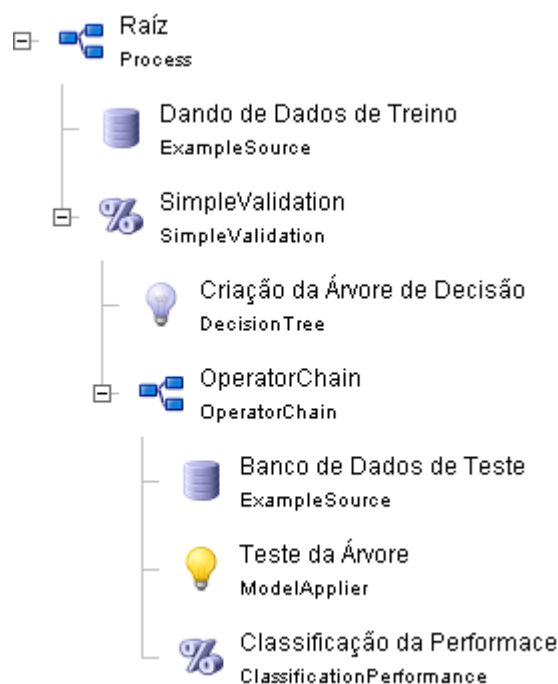


Figura 4.10 – Operadores no RapidMiner para Treino e Avaliação da Árvore de Decisão Utilizando Bancos de Dados Diferentes

Como se pode observar na figura 4.10, nenhum operador novo foi adicionado, a mudança fica por conta do operador *Example Source* dentro do sub-processo *Operator Chain*, o que significa que haverá a entrada de duas bases de dados distintas ao invés de apenas uma, conforme o método utilizado anteriormente. Outra mudança ocorrida foi no parâmetro “*split ratio*” do operador *Simple Validation*, no primeiro processo seu valor era 0.7, o que indicava que 70% do banco de dados na entrada deveria ser de treino e o restante de teste da árvore, nesse novo processo seu valor foi elevado para 1, o que indica que 100% do banco de dados declarado acima dele será utilizado para treino e o banco de dados de teste será o localizado antes do operador *Model Applier*.

Para o outro banco de dados, o número de variações das cargas, foram alterados para 47 em relação aos 27 utilizados no banco de treino anterior.

A árvore de decisão criada foi a mesma da figura 4.7. Porém, o percentual de erro aumentou de 3.21% para 4.44%, conforme figura 4.11, que ainda é um percentual de erro aceitável. O erro apresentou esse pequeno aumento de valor em virtude da

variação das cargas, que também aumentou. Contudo, essa nova forma de avaliação, mostrou que a árvore de decisão criada é bastante precisa em determinar as possíveis operações seguras e inseguras para o sistema de potência estudado. Pois, mesmo utilizando uma base de dados com outras condições de carga não previstas no banco de dados de treino, ainda assim o erro apresentado foi pequeno, o que valida a robustez da árvore de decisão criada para fins de avaliação da segurança estática de sistemas de potência em tempo real.

classification_error: 4.44%						
	true Seguro	true VLT	true VMT	true EET	true VMT,VLT	class precision
pred. Seguro	11935	22	462	0	0	96.10%
pred. VLT	0	374	0	0	22	94.44%
pred. VMT	220	0	2464	55	0	89.96%
pred. EET	0	0	22	2486	0	99.12%
pred. VMT,VLT	0	0	0	0	33	100.00%
class recall	98.19%	94.44%	83.58%	97.84%	60.00%	

Figura 4.11 – Matriz de Confusão Para a Árvore Proposta Utilizando um Banco de Dados para Treino e Outro para Teste

Capítulo 5 - Conclusões

5.1 Conclusões Gerais

Neste trabalho foi apresentada uma nova metodologia para a análise da segurança estática em sistemas de potência. Trata-se de uma metodologia baseada na utilização de árvores de decisão, cujos treinamento e avaliação foram realizados pelo software RapidMiner, realizados a partir de bases de dados geradas por uma ferramenta desenvolvida em MatLab chamada de FluxMining. Essa ferramenta realiza os cálculos das matrizes da rede elétrica, o cálculo do fluxo de potência, assim como os das outras grandezas necessárias para atender a necessidade desse estudo. A variação da carga em pequenos intervalos (designando uma variação pseudo-aleatória) e as contingências adotadas permitiram estabelecer os possíveis estados operativos reais para o sistema que se deseje estudar.

Como ilustração da metodologia foi apresentada a árvore de decisão para os resultados da análise da segurança estática de sistemas de energia elétrica, considerando-se um sistema elétrico de potência de 9 barras, 9 trechos e 3 máquinas síncronas, referenciado por Anderson & Fouad, 2003 [AND03]. Ressalta-se que o sistema referido foi utilizado para uma melhor compreensão da criação e avaliação da árvore de decisão desse sistema, considerando os limites e as contingências impostas.

A árvore de decisão criada e avaliada apresentou um pequeno percentual de erro o que comprova que o método é considerado válido e pode ser implementado na prática em tempo real usando medidas do sistema SCADA ou PMU. O sistema foi avaliado de duas maneiras diferentes. Na primeira o percentual de erro da árvore foi de 3,21%, nessa etapa a árvore de decisão foi treinada e testada pelo mesmo banco de dados, treinada com 70% e avaliada com os 30% restantes. Na segunda o erro foi de 4,44%, porém, nessa etapa a árvore foi treinada por um banco de dados e testada em outro banco de dados, o que diferiu um banco de dados do outro foram os números de situações diferentes das cargas do sistema. O bando de dados de treino teve 27 condições de cargas diferentes enquanto que o de teste teve 35.

Tendo em vista as características de desempenho da metodologia proposta, pequeno percentual de erro, rápido treinamento da árvore de decisão levando em conta vários patamares de carga e as mais prováveis contingências, pode-se empregar o método proposto de análise da segurança estática em qualquer sistema de energia elétrica, em especial aqueles que necessitam de melhorias imediatas, a fim de evitar interrupção do fornecimento de energia aos consumidores.

5.2 Sugestões para Trabalhos Futuros

Os resultados da pesquisa realizada nesse trabalho podem ser considerados satisfatórios, conforme já argumentado anteriormente. Porém, sugerem-se algumas melhorias e procedimentos que podem ser implementados com o propósito de tornar essa metodologia mais eficiente e atrativa ao usuário, propondo, ainda, outros tipos de aplicação na ferramenta desenvolvida:

1. Desenvolvimento de uma interface gráfica, para o FluxMining, amigável ao usuário, podendo ser realizada no próprio MatLab, no Toolbox “GUI”, ou em uma outra ferramenta visual, como o C/C++ ou Delphi, por exemplo;
2. Implementação, no FluxMining, dos limites e controles nos cálculos de fluxo de potência;
3. Implementação dos cálculos de fluxo de potência continuado, para melhor tratar do problema da estabilidade de tensão;
4. Ampliação das condições de contingências consideradas nos bancos de dados;
5. Implementação das árvores de decisões criadas em sistemas visuais com interface amigável que auxiliem o meio acadêmico e as empresas do setor elétrico no estudo da segurança estática de sistemas de potência;

6. Desenvolvimento de metodologias de controle preventivo para a melhoria da segurança dos sistemas elétricos, tais como o re-despacho da geração.

Referências Bibliográficas

- [CAP93] CARPENTIER, J. *Static Security Assessment and Control: a Short Survey*, Proceedings of the Athens Power Tech, 1993, Joint International Power Conference, Vol. 1, September, 1993.
- [ISO09] ISODA, L. Y. *Análise da estabilidade estática de tensão de sistemas elétricos de potência usando uma rede neural baseada na teoria da ressonância adaptativa*. Brasil: São Paulo, Ilha Solteira, 2009.
- [MON83] MONTICELLI, A. *Fluxo de Carga em Redes de Energia Elétrica*. São Paulo: Edgar Blücher. 1983.
- [BRE00] BRETAS, N. G. *Estabilidade transitória em sistemas eletroenergéticos*. São Paulo: EESC/USP. 2000.
- [KUN94] KUNDUR, P. *Power system stability and control*. United States of America: McGraw-Hill, 1994.
- [TAN09] TAN, P. N.; STEINBACH, M.; KUMAR, V. *Introdução ao Data Mining*. Brasil: Rio de Janeiro, Ciência Moderna, 2009.
- [MAN96] MANNILA, H. *Data mining: machine learning, statistics and databases*. International Conference on Statistic and Scientific Database Management, Sweden: Stockholm,8, 1996.
- [ADR96] ADRIANS, P.; ZANTINGE, D. *Data Mining*. England: Addison-Wesley, 1996.
- [AND03] ANDERSON, P. M. ; FOUAD, A. A. *Power system control and stability*. 2.ed. New York: IEEE Series on Power Engeneering, 2003.