



UNIVERSIDADE ESTADUAL PAULISTA  
"JÚLIO DE MESQUITA FILHO"  
Campus de Ilha Solteira

*PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA ELÉTRICA*

“Modelação e Análise da Vida Útil (Metrológica) de  
Medidores Tipo Indução de Energia Elétrica Ativa”

**MARCELO RUBIA DA SILVA**

**Orientador:** Prof. Dr. Carlos Alberto Canesin

Dissertação apresentada à Faculdade de  
Engenharia – UNESP – Campus de Ilha  
Solteira, para obtenção do título de Mestre  
em Engenharia Elétrica.  
Área de Conhecimento: Automação.

**Ilha Solteira - SP**

**Agosto/2010**

## FICHA CATALOGRÁFICA

Elaborada pela Seção Técnica de Aquisição e Tratamento da Informação  
Serviço Técnico de Biblioteca e Documentação da UNESP - Ilha Solteira.

S586m Silva, Marcelo Rubia da.  
Modelação e análise da vida útil (metrológica) de medidores tipo indução de Energia Elétrica Ativa / Marcelo Rubia da Silva. -- Ilha Solteira : [s.n.], 2010.  
165 f. : il.

Dissertação (mestrado) - Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira. Área de conhecimento: Automação, 2010.

Orientador: Carlos Alberto Canesin

1. Medidores eletromecânicos de energia elétrica ativa. 2. Probabilidade de falha. 3. Inteligência artificial. 4. Bases de dados. 5. Regressão stepwise. 6. Árvores de decisão.



UNIVERSIDADE ESTADUAL PAULISTA  
CAMPUS DE ILHA SOLTEIRA  
FACULDADE DE ENGENHARIA DE ILHA SOLTEIRA

### CERTIFICADO DE APROVAÇÃO

**TÍTULO:** Modelação e Análise da Vida Útil (Metrológica) de Medidores Tipo Indução de Energia Elétrica Ativa

**AUTOR:** MARCELO RUBIA DA SILVA

**ORIENTADOR:** Prof. Dr. CARLOS ALBERTO CANESIN

Aprovado como parte das exigências para obtenção do Título de MESTRE em ENGENHARIA ELÉTRICA, Área: AUTOMAÇÃO, pela Comissão Examinadora:

  
Prof. Dr. CARLOS ALBERTO CANESIN  
Departamento de Engenharia Elétrica / Faculdade de Engenharia de Ilha Solteira

  
Prof. Dr. JÚLIO BORGES DE SOUZA  
Departamento de Engenharia Elétrica / Faculdade de Engenharia de Ilha Solteira

  
Prof. Dr. DENIZAR CRUZ MARTINS  
Departamento de Engenharia Elétrica / Universidade Federal de Santa Catarina - Florianópolis/SC

Data da realização: 27 de agosto de 2010.

*À Deus.*  
*À minha família.*

# AGRADECIMENTOS

Agradeço à Deus pela luz e sabedoria que me acompanharam durante esta etapa.

Agradeço à meus pais, Arudale José da Silva e Sueli de Fátima Rubia da Silva, por todo o amor que me proporcionaram. Obrigado por aceitarem minhas decisões e por compreenderem os momentos de ausência.

Ao Prof. Dr. Carlos Alberto Canesin pela oportunidade e confiança.

Ao Prof. Dr. João Onofre Pereira Pinto pelos ensinamentos e conversas. Sem seu auxílio a qualidade do trabalho não seria a mesma.

Ao Me. Luigi Galotto Junior pela troca de conhecimento. Foram longas nossas discussões sobre inteligência artificial e banco de dados.

Ao colegas de trabalho, Me. Leonarndo Poltronieri Sampaio, Prof. Dr. Guilherme de Azevedo e Melo, Me. Moacyr Brito e demais companheiros de laboratório, pelo apoio e amizade.

À Priscila Rulli Meneses, pela atenção e carinho especiais. Obrigado por me apoiar sempre.

Aos meus irmãos, Gabriel Rubia da Silva e Arueli Rubia da Silva por compreenderem os meus esforços.

Aos meus primos, Marlene Figueira Ferreira e Samuel Antônio Figueira Ferreira pela hospitalidade e apoio.

À todos os membros do laboratório BATLAB pela fraternidade.

Ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pelo apoio financeiro, possibilitando o desenvolvimento deste trabalho.

## RESUMO

O estudo da confiabilidade operacional de equipamentos se tornou fundamental para as empresas possuírem o devido controle dos seus ativos, tanto pelo lado financeiro quanto em questões de segurança. O estudo da taxa de falha de equipamentos prevê quando as falhas irão ocorrer possibilitando estabelecer atitudes preventivas, porém, seu estudo deve ser realizado em condições de operação estabelecidas e fixas. Os medidores de energia elétrica, parte do ativo financeiro das concessionárias de energia, são equipamentos utilizados em diversas condições de operação, tanto nas condições do fluxo de energia, tais como presenças de harmônicos, subtensões, sobre-tensões e padrões de consumo distintos, quanto pelo local físico de instalação, tais como maresia, temperatura, umidade, etc. As falhas nos medidores eletromecânicos de energia elétrica são de difícil constatação uma vez que a maioria dos erros de medição, ocasionados principalmente por envelhecimento de componentes, não alteram a qualidade da energia fornecida e nem interrompem o seu fornecimento. Neste sentido, este trabalho propõe uma nova metodologia de determinação de falhas em medidores eletromecânicos de energia elétrica ativa. Faz-se uso de banco de dados de uma concessionária de energia elétrica e do processo de descoberta de conhecimento em bases de dados para selecionar as variáveis mais significativas na determinação de falhas em medidores eletromecânicos de energia elétrica ativa, incluindo no conjunto de falhas a operação com erros de medição acima do permitido pela legislação nacional (2010). Duas técnicas de mineração de dados foram utilizadas: regressão *stepwise* e árvores de decisão. As variáveis obtidas foram utilizadas na construção de um modelo de agrupamento de equipamentos associando a cada grupo uma probabilidade de falha. Como resultados finais, um aplicativo em plataforma amigável foi desenvolvido para a implementação da metodologia proposta e um estudo de caso é apresentado para demonstrar sua utilização.

**Palavras-chave:** *Medidores eletromecânicos de energia elétrica ativa. Probabilidade de falha. Inteligência artificial. Descoberta de conhecimento em bases de dados. Regressão stepwise. Árvores de decisão.*

# ABSTRACT

The operational reliability study of equipments has become primal in order to enterprises have the righteous control over their assets, both by financial side as by security reasons. The study for the hazard rate of equipments allows to foresee the failures for the equipments and to act preventively, but this study must be accomplished under established and fixed operation conditions. The energy meters, for their part, are equipments utilized in several operating conditions so on the utilization manner, like presence of harmonics, under-voltages and over-voltages and distinct consumption patterns, as on the installation location, like swel, temperature, humidity, etc. Failures in electromechanical Wh-meters are difficult to detect once that the majority of metering errors occurred mainly by aging of components do not change the quality of offered energy neither disrupt its supply. In this context, this work proposes a novel methodology to obtain failure determination for electromechanical Wh-meters. It utilizes Wh-databases from an electrical company and of the process of knowledge discovery in databases to specify the most significant variables in determining failures in electromechanical Wh-meters, including in the failure set the operation with metering errors above those permitted by national regulations (2010). Two techniques of data mining were used in this work: stepwise regression and decision trees. The obtained variables were utilized on the construction of a model of clustering similar equipments and the probability of failure of those clusters were determined. As final results, an application in a friendly platform were developed in order to apply the methodology, and a case study was accomplished in order to demonstrate its feasibility.

**Keywords:** *Active electromechanical energy meters. Probability of failure. Artificial intelligence. Knowledge discovery in databases. Stepwise regression. Decision trees.*

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO GERAL.....</b>	<b>17</b>
<b>1.1</b>	<b><i>Medidores eletromecânicos .....</i></b>	<b>19</b>
<b>1.2</b>	<b><i>Medidores eletrônicos.....</i></b>	<b>22</b>
<b>1.3</b>	<b><i>Erros de medição.....</i></b>	<b>24</b>
<b>1.4</b>	<b><i>Legislação sobre erros de medição.....</i></b>	<b>26</b>
<b>1.4.1</b>	<b>Agência Nacional de Energia Elétrica (ANEEL).....</b>	<b>26</b>
<b>1.4.2</b>	<b>Agência Reguladora de Saneamento e Energia do Estado de São Paulo (ARSESP).....</b>	<b>27</b>
<b>1.4.3</b>	<b>Instituto Nacional de Metrologia, Normalização e Qualidade Industrial (Inmetro).....</b>	<b>27</b>
<b>1.5</b>	<b><i>Motivação .....</i></b>	<b>29</b>
<b>1.6</b>	<b><i>Determinação de falha de equipamentos .....</i></b>	<b>30</b>
<b>1.7</b>	<b><i>Estado da arte.....</i></b>	<b>33</b>
<b>1.8</b>	<b><i>Metodologia .....</i></b>	<b>34</b>
<b>2</b>	<b>SELEÇÃO DAS VARIÁVEIS MAIS SIGNIFICATIVAS PARA DETERMINAÇÃO DE FALHAS EM MEDIDORES ELETROMECÂNICOS DE ENERGIA ELÉTRICA ATIVA .....</b>	<b>36</b>
<b>2.1</b>	<b><i>Considerações Iniciais.....</i></b>	<b>36</b>
<b>2.2</b>	<b><i>O banco de dados da concessionária .....</i></b>	<b>36</b>
<b>2.3</b>	<b><i>Descoberta de Conhecimento em Bases de Dados .....</i></b>	<b>38</b>
<b>2.4</b>	<b><i>Seleção.....</i></b>	<b>41</b>
<b>2.5</b>	<b><i>Integração de dados.....</i></b>	<b>45</b>
<b>2.5.1</b>	<b>Agrupamento de modelos .....</b>	<b>48</b>
<b>2.6</b>	<b><i>Pré-processamento e limpeza dos dados .....</i></b>	<b>48</b>
<b>2.6.1</b>	<b>Dados ausentes (<i>missing values</i>).....</b>	<b>49</b>
<b>2.6.2</b>	<b>Dados incorretos .....</b>	<b>50</b>
<b>2.6.3</b>	<b>Dados discrepantes (<i>outliers</i>) .....</b>	<b>52</b>
<b>2.6.4</b>	<b>Dados redundantes .....</b>	<b>55</b>
<b>2.7</b>	<b><i>Transformação.....</i></b>	<b>56</b>
<b>2.7.1</b>	<b>Transformação de variáveis categóricas em numéricas .....</b>	<b>57</b>

2.7.2	Transformação de variáveis contínuas em discretas (discretização)....	58
2.8	<i>Mineração de Dados</i> .....	59
2.8.1	Classificação.....	60
2.8.2	Regressão .....	61
2.8.3	Agrupamento ( <i>Clustering</i> ) .....	62
2.8.4	Associação.....	63
2.8.5	Sumarização ( <i>Summarization</i> ) .....	63
2.8.6	Detecção de desvios (Deviation detection).....	63
2.9	<i>Técnicas de mineração de dados</i> .....	63
2.9.1	Regressão .....	64
2.9.2	Árvores de decisão.....	72
2.10	<i>Pós-processamento</i> .....	82
2.10.1	Modelo final.....	82
2.11	<i>Considerações finais</i> .....	84
3	<b>METODOLOGIA DE CÁLCULO DA PROBABILIDADE DE FALHA DOS EQUIPAMENTOS</b> .....	86
3.1	<i>Determinação dos pesos das variáveis de entrada</i> .....	87
3.2	<i>Criação das classes finais</i> .....	93
3.3	<i>Minimização de erros</i> .....	94
3.3.1	Regressão via kernel.....	95
3.3.2	Peso inicial.....	98
3.4	<i>Considerações finais</i> .....	99
4	<b>APLICATIVO EM PLATAFORMA AMIGÁVEL</b> .....	100
4.1	<i>Tela inicial</i> .....	100
4.2	<i>Importação</i> .....	101
4.2.1	Importação de dados do parque e de consumo dos medidores .....	102
4.2.2	Importação dos dados de ensaio .....	105
4.3	<i>Análise de probabilidade de falhas</i> .....	109
4.3.1	Efetuar análise .....	109
4.3.2	Criar novo modelo utilizando IA.....	114
4.4	<i>Configurações</i> .....	117

<b>4.5</b>	<b><i>Sistema especialista para amostragens</i></b> .....	<b>118</b>
<b>4.6</b>	<b><i>Considerações finais</i></b> .....	<b>122</b>
<b>5</b>	<b>ESTUDO DE CASO</b> .....	<b>124</b>
<b>5.1</b>	<b><i>Introdução</i></b> .....	<b>124</b>
<b>5.2</b>	<b><i>Aplicação da metodologia sobre os equipamentos</i></b> .....	<b>125</b>
<b>5.3</b>	<b><i>Considerações finais</i></b> .....	<b>127</b>
<b>6</b>	<b>CONCLUSÃO E CONTINUIDADE DA PESQUISA</b> .....	<b>128</b>
	<b>REFERÊNCIAS</b> .....	<b>131</b>
	<b>APÊNDICE A – Agrupamento de modelos semelhantes</b> .....	<b>138</b>
	<b>APÊNDICE B – Árvore de decisão completa</b> .....	<b>141</b>

## LISTA DE FIGURAS

<b>Figura 1.1</b>	Lâmpada desenvolvida por Thomas Edison .....	17
<b>Figura 1.2</b>	Medidor químico de energia elétrica desenvolvido por Thomas Edison .....	18
<b>Figura 1.3</b>	Medidor de indução de Shallenberger .....	19
<b>Figura 1.4</b>	Vista explodida do medidor monofásico.....	20
<b>Figura 1.5</b>	Gráfico de confiabilidade para distribuição normal .....	31
<b>Figura 1.6</b>	Curva de taxa de falha típica (curva “da banheira”) .....	33
<b>Figura 2.1</b>	Relacionamento entre as tabelas da concessionária.....	37
<b>Figura 2.2</b>	Relação de DCBD com outras áreas .....	38
<b>Figura 2.3</b>	Processo de DCBD .....	40
<b>Figura 2.4</b>	Exemplo de relatório no padrão ARSESP.....	43
<b>Figura 2.5</b>	Construção da tabela RESULTADOS .....	47
<b>Figura 2.6</b>	Exemplo de dados ausentes na tabela MED_CONS. a) Caso de UC inexistente até dado período. b) Caso de UC desligada.....	49
<b>Figura 2.7</b>	Registro com quebra dos dados em duas linhas .....	51
<b>Figura 2.8</b>	Utilização de sequência de escape. a) String sem uso do caractere de escape. b) Uso de caractere de controle “nova linha” .....	52
<b>Figura 2.9</b>	Descrição de clientes que pagam ou não os empréstimos em função da renda e da quantidade de dívidas .....	60
<b>Figura 2.10</b>	Exemplo de classificação.....	61
<b>Figura 2.11</b>	Exemplo de regressão .....	62
<b>Figura 2.12</b>	Exemplo de agrupamento .....	62
<b>Figura 2.13</b>	Convenção dos modelos.....	64
<b>Figura 2.14</b>	Convenção de matrizes .....	65
<b>Figura 2.15</b>	Fluxograma simplificado de <i>forward stepwise</i> .....	69
<b>Figura 2.16</b>	Valores de RMSE para todas as etapas do algoritmo <i>forward stepwise</i> .....	70
<b>Figura 2.17</b>	Valores de RMSE para todas as etapas do algoritmo <i>backward stepwise</i> .....	71
<b>Figura 2.18</b>	Exemplo de árvore de decisão.....	73

<b>Figura 2.19</b>	Representação dos valores da variável de saída classificados de acordo com as variáveis de entrada .....	76
<b>Figura 2.20</b>	Estrutura inicial da árvore de decisão .....	78
<b>Figura 2.21</b>	Árvore de decisão obtida com os dados da concessionária .....	79
<b>Figura 2.22</b>	Detalhe da árvore para <code>instalacoes="1"</code> e <code>subtipo="M1A"</code> .....	81
<b>Figura 3.1</b>	Efeito de $\sigma$ sobre a regressão kernel. a) Resultado para $\sigma$ pequeno. b) Resultado para $\sigma$ elevado.....	97
<b>Figura 3.2</b>	Efeito do <code>peso inicial</code> sobre a resposta da regressão kernel. a) Com <code>peso inicial = 0</code> . b) Com <code>peso inicial = 1</code> .....	98
<b>Figura 4.1</b>	Tela inicial do aplicativo .....	101
<b>Figura 4.2</b>	Menu de importação .....	102
<b>Figura 4.3</b>	Assistente de importação de dados de parque e consumo.....	103
<b>Figura 4.4</b>	Tela de informações adicionais sobre a importação .....	104
<b>Figura 4.5</b>	Último estágio do assistente de importação dos registos com dados de parque e consumo.....	105
<b>Figura 4.6</b>	Exemplo de arquivo de ensaio com preenchimento de abas.....	106
<b>Figura 4.7</b>	Assistente de importação de dados de ensaio.....	107
<b>Figura 4.8</b>	Escolha da base de dados de parque e consumo durante à importação de dados de ensaio .....	108
<b>Figura 4.9</b>	Último estágio da importação de dados de ensaio .....	108
<b>Figura 4.10</b>	Menu de análise de probabilidade de falhas.....	109
<b>Figura 4.11</b>	Interface de análise de probabilidade de falhas.....	110
<b>Figura 4.12</b>	Abas da seção de filtros da interface de análise de probabilidade de falhas .....	111
<b>Figura 4.13</b>	Histograma da interface de análise de probabilidade de falhas.....	112
<b>Figura 4.14</b>	Mapa de distribuição da probabilidade de falhas em cada escala por localidade.....	112
<b>Figura 4.15</b>	Tabela resumida para análise da probabilidade de falhas .....	113
<b>Figura 4.16</b>	Tabela detalhada para análise da probabilidade de falhas.....	113
<b>Figura 4.17</b>	Cabeçalho de um relatório resumido .....	114
<b>Figura 4.18</b>	Assistente de geração de modelo de probabilidade de falhas .....	115
<b>Figura 4.19</b>	Visualização do ajuste das probabilidades de falhas para os novos dados de ensaio .....	116
<b>Figura 4.20</b>	Menu de configurações.....	117

<b>Figura 4.21</b>	Interface de configuração da imagem do parque e modelo de probabilidade de falhas selecionados .....	118
<b>Figura 4.22</b>	Menu do sistema especialista para amostragens .....	119
<b>Figura 4.23</b>	Interface de amostragem de equipamentos.....	120
<b>Figura 4.24</b>	Exemplo de simulação da acurácia esperada.....	121
<b>Figura 4.25</b>	Interface de gerenciamento de amostragens.....	121
<b>Figura 4.26</b>	Exemplo de relatório de amostragem de medidores.....	122

## LISTA DE TABELAS

<b>Tabela 1.1</b>	Distribuição de medidores de energia elétrica de acordo com tipo e região geográfica .....	20
<b>Tabela 1.2</b>	Elementos do medidor monofásico .....	21
<b>Tabela 1.3</b>	Funcionalidades disponíveis nos medidores eletrônicos. ....	23
<b>Tabela 1.4</b>	Distribuição de medidores de acordo com a faixa de idade.....	26
<b>Tabela 2.1</b>	Descrição das variáveis da tabela <code>DADOS</code> .....	44
<b>Tabela 2.2</b>	Descrição das variáveis da tabela <code>CLIMATICOS</code> .....	45
<b>Tabela 2.3</b>	Sequências de escape do MySQL.....	51
<b>Tabela 2.4</b>	Consumo de unidades consumidoras monofásicas.....	53
<b>Tabela 2.5</b>	Quantidade de registros para cada subtipo no depósito de dados e tabela <code>UC_MEDIDOR</code> .....	53
<b>Tabela 2.6</b>	Segmento de um depósito de dados .....	57
<b>Tabela 2.7</b>	Segmento de um depósito de dados após flattening .....	57
<b>Tabela 2.8</b>	Lista de variáveis categóricas utilizadas para MD .....	58
<b>Tabela 2.9</b>	Exemplo de técnicas de mineração de dados e tarefas em que podem ser utilizadas .....	64
<b>Tabela 2.10</b>	Variáveis selecionadas pelo algoritmo <i>forward stepwise</i> .....	70
<b>Tabela 2.11</b>	Variáveis selecionadas pelo algoritmo <i>backward stepwise</i> .....	71
<b>Tabela 2.12</b>	Dados sobre tempo para se jogar tênis .....	74
<b>Tabela 2.13</b>	Ganho de informação da variável de saída utilizando uma variável de entrada.....	77
<b>Tabela 2.14</b>	Taxa de ganho para as variáveis do problema sobre jogo de tênis .....	78
<b>Tabela 2.15</b>	Matriz de confusão para a árvore de decisão .....	80
<b>Tabela 2.16</b>	Matriz de confusão para a árvore de decisão após poda.....	81
<b>Tabela 3.1</b>	Classificação da probabilidade de falha para a variável <code>subtipo equip</code> .....	87
<b>Tabela 3.2</b>	Classificação da probabilidade de falha para a variável <code>instalacoes</code> .....	88
<b>Tabela 3.3</b>	Classificação da probabilidade de falha para a variável <code>media</code> .....	88
<b>Tabela 3.4</b>	Classificação da probabilidade de falha para a variável <code>maximo</code> .....	88

<b>Tabela 3.5</b>	Classificação da probabilidade de falha para a variável dias_de_vida .....	88
<b>Tabela 3.6</b>	Classificação da probabilidade de falha para a variável cod_tipo_local .....	89
<b>Tabela 3.7</b>	Classificação da probabilidade de falha para a variável cod_localidade .....	89
<b>Tabela 3.8</b>	Classificação da probabilidade de falha para a variável cod_classe_principal .....	93
<b>Tabela 3.9</b>	Classificação da probabilidade de falha para a variável precipitacao .....	93
<b>Tabela 3.10</b>	Exemplo de determinação do grupo de risco de um medidor .....	93
<b>Tabela 3.11</b>	Modelo final para determinação de probabilidade de falha nos medidores .....	94
<b>Tabela 4.1</b>	Formatação padrão para geração do arquivo "uc_medidor.csv" ..	102
<b>Tabela 4.2</b>	Formatação padrão para geração do arquivo "med_cons.csv" .....	103
<b>Tabela 4.3</b>	Exemplo de planilha para importação de dados de ensaio.....	105
<b>Tabela 4.4</b>	Escalas de probabilidade de falhas utilizada no aplicativo.....	111
<b>Tabela 5.1</b>	Distribuição de medidores no parque da concessionária de acordo com o tipo de medidor e número de fases, ano 2005 .....	124
<b>Tabela 5.2</b>	Distribuição de medidores no parque da concessionária de acordo com o tipo de medidor eletromecânico e número de fases, ano 2008. ....	125
<b>Tabela 5.3</b>	Grupos de medidores e quantidade de equipamentos para o estudo de caso.....	126
<b>Tabela A</b>	Modelos agrupados utilizados na tabela RESULTADOS .....	138

## LISTA DE ABREVIATURAS E SIGLAS

a.C.	Antes de Cristo
A/D	Analógico para digital
ANEEL	Agência Nacional de Energia Elétrica
APR	All Possible Regression
ARSESP	Agência Reguladora de Saneamento e Energia do Estado de São Paulo
BE	Backward Elimination
CART	Classification and Regression Tree
CSV	Comma-separated Values
Conmetro	Conselho Nacional de Metrologia, Normalização e Qualidade Industrial
CSPE	Comissão de Serviços Públicos de Energia
DCBD	Descoberta de Conhecimento em Bases de Dados
Embrapa	Empresa Brasileira de Pesquisa Agropecuária
FS	Forward Stepwise
ID3	Inductive Decision Tree
Inmetro	Instituto Nacional de Metrologia, Normalização e Qualidade Industrial
kV	Kilovolt
kVAr	Kilovolt-ampère reativo
kWh	Kilowatt-hora
LEP	Laboratório de Eletrônica de Potência
MD	Mineração de Dados
MME	Ministério de Minas e Energia
MTBF	Tempo médio entre falhas
RMSE	Root Mean Squared Error
Sinmetro	Sistema Nacional de Metrologia, Normalização e Qualidade Industrial
TB	Terabytes
TC	Transformador de corrente
TP	Transformador de potencial
UC	Unidade consumidora
WEKA	Waikato Environment for Knowledge Analysis

$R(t)$	confiabilidade
$Pr$	probabilidade do item falhar
$\lambda$	taxa de falha
$X$	matriz de valores das variáveis de entrada
$Y$	matriz de valores das variáveis de saída
$\hat{Y}$	matriz de valores estimados das variáveis de saída
$\beta$	Coefficientes lineares da regressão
$\varepsilon$	resíduo
$q$	valor de entrada para o qual a saída se deseja estimar
$\sigma$	largura de banda da função kernel

# 1 INTRODUÇÃO GERAL

---

---

A energia elétrica é item fundamental no desenvolvimento socioeconômico de um país sendo que sua utilização está diretamente ligada à disponibilidade de tecnologias existentes.

O homem sempre foi instigado para obter este conhecimento, sendo que estudos nesta área se iniciaram na Grécia, no século VI a.C., pelo filósofo Tales de Mileto ao tentar descrever o fenômeno da eletricidade estática, produzida ao se esfregar bastões de âmbar em tecidos e observar a atração de pequenas quantidades de palha pelo objeto energizado. No século XVII, Otto von Guericke estuda os efeitos da eletrificação por atrito. No século XVIII, Alessandro Volta inventa a bateria. Somente no século XIX o ser humano compreende melhor as leis que regem a eletricidade: Michael Faraday descobre a indução eletromagnética e James Clerk Maxwell a formula através de equações matemáticas, conhecidas como as leis de Maxwell [1 e 2].

A invenção da lâmpada elétrica por Thomas A. Edison, em 1879, marca o início do uso comercial da energia elétrica. Em 1880, Edison funda a *Edison Electric Illuminating Company of New York* (Companhia Edison de Iluminação Elétrica de Nova Iorque)<sup>1</sup>, criando a primeira planta de geração elétrica e sistema de distribuição de energia. A figura 1.1 apresenta a lâmpada elétrica de Thomas Edison.

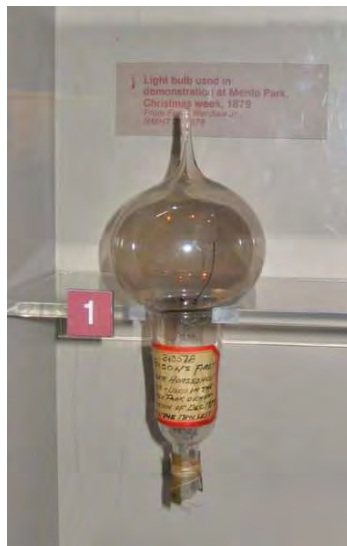
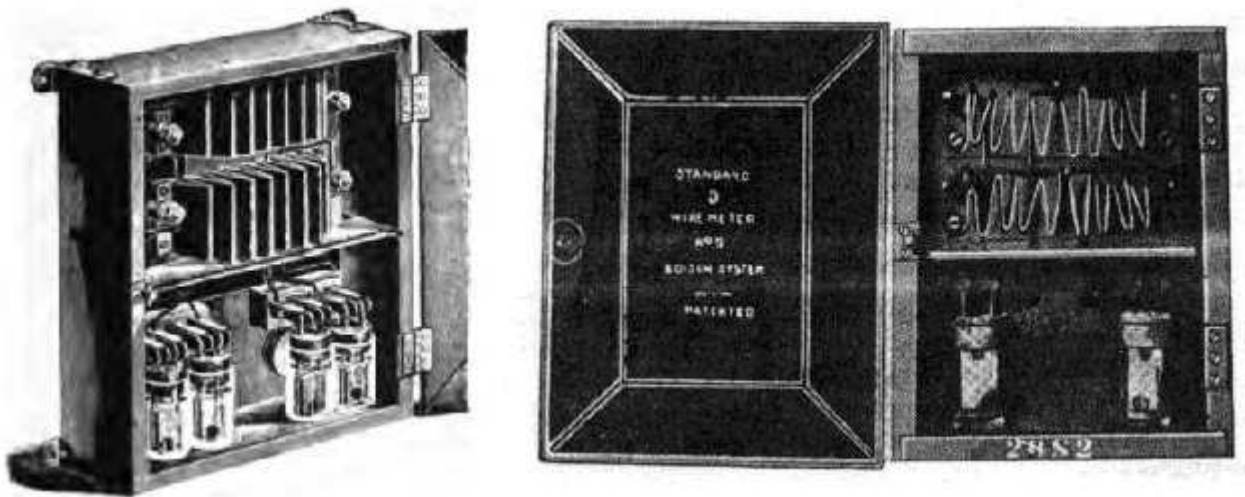


Figura 1.1 – Lâmpada desenvolvida por Thomas Edison. Fonte: [64]

---

<sup>1</sup> Anos mais tarde, após fusão com a Thomson-Houston Electric Company, torna-se General Electric.

A energia elétrica produzida pela empresa de Edison era em corrente contínua. Inicialmente, a cobrança era feita pela quantidade de lâmpadas que o cliente tinha conectado ao sistema [3]. Na tentativa de realizar uma melhor cobrança/faturamento, Edison desenvolveu um medidor de energia elétrica químico: um jarro com placas de zinco imersas em uma solução da mesma substância. As placas eram conectadas em série no terminal de entrada da Unidade Consumidora (UC). A corrente fluindo pelo jarro faz com que a concentração de zinco em uma das placas diminua enquanto a concentração na outra placa aumenta. Todo mês as placas eram pesadas e o pagamento era realizado de acordo com o aumento de peso da placa positiva. A figura 1.2 apresenta o medidor desenvolvido por Edison.



**Figura 1.2 – Medidor químico de energia elétrica desenvolvido por Thomas Edison. Fonte: [3 e 4]**

Em 1886 a *Westinghouse Company* compra as patentes de um gerador de energia alternada e de um transformador de potência desenvolvidos por Nikola Tesla, marcando o início da comercialização de energia elétrica na forma de corrente alternada. A partir daí se inicia uma disputa entre Westinghouse e Edison sobre a melhor forma de distribuição de energia elétrica. A rivalidade entre os dois marcou o período com o título “A Batalha das Correntes”. Os acionistas observaram maior rentabilidade no sistema da Westinghouse que, apesar de ser considerado mais letal, segundo Thomas Edison, do que o sistema em corrente contínua, não possuía a limitação de transmissão a apenas algumas quadras e permitia a utilização de condutores de menores bitolas, beneficiando-se da capacidade de elevação de tensão por meio de transformadores, diminuindo-se perdas, custos e aumentando o lucro.

Em 1888, Oliver Blackburn Shallenberger desenvolve um medidor de energia para corrente alternada, o que passou a ser adotado como padrão na indústria [5]. O medidor de Shallenberger baseia-se num medidor de ampère-hora: o inventor foi responsável por desenvolver uma estratégia de tornar o fluxo do circuito de corrente do medidor em quadratura com a tensão, possibilitando a medição de energia em corrente alternada. A figura 1.3 apresenta o medidor de Shallenberger.

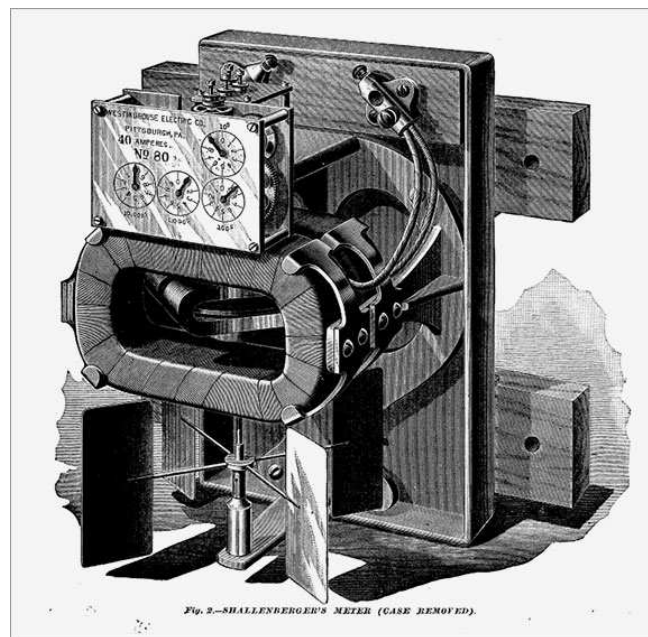


Figura 1.3 – Medidor de indução de Shallenberger. Fonte: [65]

Desenvolvimentos posteriores, elaborados por outros pesquisadores, resultaram em simplificações de projeto, melhoria das características de operação e, conseqüentemente, melhor desempenho [6]. Entretanto, o princípio de funcionamento manteve-se inalterado até o início da década de 1990 quando os medidores eletrônicos de energia elétrica começaram a ser comercializados.

## 1.1 Medidores eletromecânicos

Os medidores de energia elétrica eletromecânicos são os equipamentos em maior utilização no parque metrológico brasileiro atualmente (dados 2009) [7]. O parque de medição nacional, neste período, era constituído de 92,61% de medidores eletromecânicos. A tabela 1.1 apresenta a distribuição de todos medidores (eletromecânicos e eletrônicos), por regiões.

Tabela 1.1 – Distribuição de medidores de energia elétrica de acordo com tipo e região geográfica.

Região Geográfica	Tipo de Medidor	
	Eletromecânico	Eletrônico
Norte	79,09%	20,91%
Nordeste	88,27%	11,73%
Centro-Oeste	96,11%	3,89%
Sudeste	94,60%	5,40%
Sul	97,38%	2,62%
BRASIL	92,61%	7,39%

Nas regiões onde a implantação do sistema de energia elétrica é mais recente (Norte e Nordeste) a quantidade de medidores eletrônicos aumenta. Observa-se claramente a preferência pela instalação de medidores eletrônicos em novos clientes. O parque metrológico das áreas norte e nordeste tem expandido em função de programas de incentivo governamental (*e.g.*, Programa Luz Para Todos) [7].

O medidor eletromecânico tipo indução é constituído de diversas partes, incluindo: base, terminais, bloco de terminais, mostrador, cilindro ciclométrico, núcleo, bobina de corrente, bobina de tensão/potencial, elemento motor, dispositivos de calibração, elemento frenador, tampa do medidor, elemento móvel, catraca, dispositivos de compensação, estrutura/armação e mancais. Portanto, os medidores de energia elétrica, tipo indução, são dispositivos eletromecânicos com partes móveis que se degradam de forma progressiva com o tempo. Um diagrama esquemático para um medidor eletromecânico monofásico é apresentado na figura 1.4 e tabela 1.2.

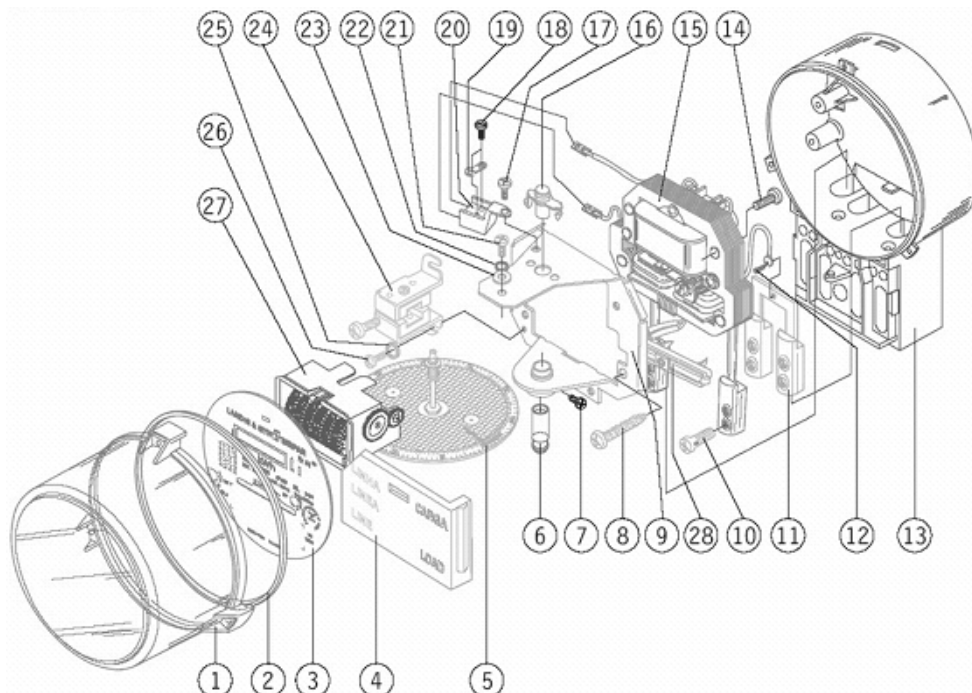


Figura 1.4 – Vista explodida do medidor monofásico. Fonte: [8]

**Tabela 1.2 – Elementos do medidor monofásico.**

1	Tampa do medidor	11	Terminal ligação de neutro	19	Gancho de prova
2	Gaxeta da tampa do Medidor	12	Parafuso fixação do terminal de potencial-ligação de neutro	20	Suporte prova interna
3	Placa de identificação			21	Parafuso fixação do registrador
4	Tampa do bloco terminais	13	Base e bloco do medidor		
5	Elemento Móvel	14	Parafuso fixação do elemento armadura	22	Arruela pressão
6	Mancal inferior (magnético)			23	Arruela lisa
7	Parafuso fixação do mancal inferior	15	Elemento motor	24	Imã frenador
8	Parafuso fixação do elemento motor	16	Mancal superior (pino guia)	25	Arruela pressão
9	Armadura	17	Parafuso fixação superior do terminal de prova	26	Parafuso fixação do conjunto imã
10	Parafuso terminal de corrente	18	Parafuso fixação do gancho de prova	27	Registrador
				28	Suporte da indutiva

Os medidores de indução podem ser classificados segundo as seguintes características [9]:

- Tipo de energia a ser medida:
  - Medidor de energia ativa (kWh);
  - Medidor de energia reativa (kVAr);
- Número de elementos motores/número de fios:
  - Monofásico: um elemento motor;
  - Polifásicos: dois ou três elementos motores;
- Tipo de ligação à rede
  - Medidores diretos: ligados diretamente à rede e à carga;
  - Medidores indiretos: conectados através de transformadores de potencial (TP) e/ou transformadores de corrente (TC);
- Classe de exatidão (utilizada comercialmente no Brasil):
  - Classe 1: medidores que, novos, não apresentam mais de 1% de erro de medição;
  - Classe 2: medidores que, novos, não apresentam mais de 2% de erro de medição.

## 1.2 Medidores eletrônicos

Os medidores eletrônicos realizam a leitura das variáveis através de amostragem em pequenos intervalos de tempo utilizando conversores analógico para digital (A/D). A classe de exatidão do equipamento é determinada pela quantidade de bits nos conversores A/D e precisão de processamento, possibilitando a estes serem mais precisos do que os medidores eletromecânicos.

Os medidores eletrônicos podem realizar simultaneamente várias tarefas em um mesmo equipamento. Pode-se medir energia ativa e reativa, demanda máxima, fator de potência, tensão e corrente, além do uso de memória de massa para registro do consumo com informação de data e hora e sistema antifraude [10]. Um único equipamento eletromecânico é incapaz de realizar, isoladamente, as tarefas mencionadas.

O uso de medidores eletrônicos permite adentrar em uma nova área tecnológica, incluindo leitura através de telemetria. Para realizar as leituras dos equipamentos é utilizado um protocolo de comunicação entre o medidor e um equipamento utilizado por um leitorista para armazenamento dos registros de leitura, ou através de protocolos de comunicação de rede enviando os dados diretamente para um banco de dados da concessionária.

A implementação da medição eletrônica tem ocorrido de forma gradativa no Brasil devido, principalmente, ao alto custo do equipamento. As primeiras unidades consumidoras a receber medição eletrônica foram do grupo A.

O sistema tarifário brasileiro é estruturado em dois grandes grupos de consumidores, grupo A e grupo B. A Agência Nacional de Energia Elétrica (ANEEL) define, por meio da resolução 456 de 2000, que o grupo A é composto por unidades consumidoras atendidas em alta superior ou igual a 2,3 kV, ou ainda unidades consumidoras atendidas com tensão inferior a 2,3 kV a partir de sistema subterrâneo de distribuição, caracterizado por estrutura de tarifação binômica e subdividido nos seguintes subgrupos:

- A1: nível de tensão de 230 kV ou superior;
- A2: nível de tensão de 88 a 138 kV;
- A3: nível de tensão de 69 kV;
- A3a: nível de tensão de 30 a 44 kV;
- A4: nível de tensão de 2,3 a 25 kV;
- AS: sistemas subterrâneos.

As principais funcionalidades disponíveis nos medidores eletrônicos são apresentadas na tabela 1.3.

**Tabela 1.3 – Funcionalidades disponíveis nos medidores eletrônicos. Fonte: [11]**

<b>Funcionalidades</b>	
<b>Características inerentes</b>	Classe de exatidão
	Sensibilidade a pequenas cargas (baixa corrente de partida)
	Perdas técnicas (aprox. 0,5W)
	Auto Diagnóstico
<b>Faturamento</b>	Energia Ativa (kWh)
	Energia Reativa (kVAr)
	Energia nos 4 quadrantes
	Demanda Máxima (kW)
<b>Tarifação</b>	Tarifa Binomial
	Tarifa horária
	Pré-pagamento
<b>Qualidade do serviço</b>	Apuração da duração das interrupções
	Apuração da quantidade (frequência de ocorrência)
<b>Qualidade do produto</b>	Registro do valor de tensão
	Registro do valor da frequência
	Qualidade de Energia
<b>Antifraude</b>	Mecânica (abertura de tampa)
	Detecção eletrônica (abertura de tampa)
	Software (energia reversa)

Unidades consumidoras nesta categoria devem ter aferidos os valores de demanda de potência ativa, energia ativa, fator de potência, e consumo de energia elétrica e demanda reativas quando o fator de potência for inferior a 0,92, além de possuir tarifa horo-sazonal. Estas características e o consumo elevado, típico deste tipo de UCs, tornam esses clientes favoráveis para utilização dos medidores eletrônicos.

O grupo B é composto por unidades consumidoras com tensão de fornecimento inferior a 2,3 kV, caracterizado pela estruturação tarifária monômnia e subdividido nos seguintes subgrupos:

- B1: residencial e residencial de baixa renda;
- B2: rural, abrangendo diversas classes (e.g, agropecuária, indústria rural, serviço público de irrigação);

- B3: demais classes;
- B4: iluminação pública;

As unidades consumidoras nesta categoria devem ter aferidos os valores de consumo de energia elétrica ativa, e, quando aplicável, o consumo de energia elétrica reativa excedente. A medição do fator de potência é facultativa mas, sem esta, não se pode aplicar o consumo de energia elétrica reativa.

As empresas ainda possuem incertezas quanto à utilização dos medidores eletrônicos, especialmente para cliente de baixa tensão (grupo B). Entre as principais dúvidas estão a vantagem de implementação atual, uma vez que os medidores eletrônicos instalados em UCs do grupo B realizam as mesmas tarefas dos medidores eletromecânicos. A falta de informações sobre a vida útil, atualmente estimada em 15 anos, e a insegurança do comportamento em condições climáticas adversas também são motivos para uma implantação lenta.

Em 30 de Abril de 2009, a Agência Nacional de Energia Elétrica (ANEEL) realizou uma consulta pública para “estimular a discussão sobre a implantação de medição eletrônica em UCs de baixa tensão (classe B) e os aspectos relacionados a este tipo de iniciativa, tais como abordagem regulatória, impacto tarifário, funcionalidades agregadas e planos de implantação” [12].

A ANEEL tem apresentado às concessionárias de energia estudos de casos de uso de medidores eletrônicos em outros países, como Itália e Suécia, por exemplo, para ampliar a reflexão a respeito das novas tecnologias aplicadas à medição de energia elétrica [13].

### **1.3 Erros de medição**

O medidor de energia eletromecânico, por ser um equipamento constituído por partes eletromecânicas, pode ter seu funcionamento e precisão influenciados pelas seguintes características externas [9 e 14]:

- Variações de temperatura ambiente: segundo [9], um aumento de temperatura de 10°C pode ocasionar uma redução de 4% de torque motor e redução de 5% de torque frenante;
- Variações da tensão da rede: sobretensões fazem com que o medidor indique um valor de consumo menor do que o real, já subtensões acarretam em medição superior ao valor real;

- Variações da frequência da rede: as pequenas variações de frequência não afetam o funcionamento do medidor, mas para frequência maior que a nominal a rotação do elemento motor tende a diminuir e para frequência menor que a nominal a rotação do elemento motor tende a aumentar;
- Campos magnéticos externos: o princípio de funcionamento dos medidores eletromecânicos é a indução eletromagnética. Campos magnéticos externos intensos, que possuam capacidade de induzir corrente no elemento girante, podem modificar o valor do torque no elemento girante, provocando erros de medição;
- Distorções harmônicas: os medidores baseados no princípio de indução eletromagnética apresentam erros de medição perante distorções harmônicas, podendo estes chegar a valores superiores a 10%.

De acordo com [14], dois itens que podem gerar erros de medição no decorrer dos anos são as engrenagens e o freio magnético. Sob fricção excessiva o elemento móvel tende a girar de maneira mais lenta registrando medição inferior ao valor real. Caso o freio magnético diminua sua intensidade, o medidor tenderá a permanecer girando após diminuição de carga.

De acordo com [15] e [16] os medidores de indução foram designados para medir tensões e correntes senoidais. Submetidos à tensão e/ou corrente não-senoidal, ou seja, que apresentam harmônicos, os medidores de indução apresentam erros devido à característica de torque frente a diversas frequências que constituem a energia. Os estudos realizados em [17], [18], [19] e [20] comprovam a afirmação. No estudo realizado em [17] constatou-se que os erros provocados por harmônicas podem chegar a 10%, em condições usuais das redes de distribuição com carregamento não-linear. Os ensaios em laboratório realizados por [20] confirmam os efeitos dessas distorções em medidores eletromecânicos e eletrônicos com base em análises de distorções obtidas em campo. Dentre os resultados obtidos constatou-se que taxas de distorções harmônicas de tensão de até 5% apresentam influência muito pequena nos erros dos medidores enquanto que as taxas de distorção harmônicas de corrente e os valores de deslocamento instantâneos de tensão e corrente exercem grande influência nos erros de medição.

O tempo em que o medidor está instalado no parque de medição também é um fator que leva aos erros de medição. Os medidores eletromecânicos possuem uma vida útil econômica estimada em 25 anos [9]. A tabela 1.4 apresenta a constituição do parque metrológico brasileiro de acordo com o tempo de vida dos medidores.

**Tabela 1.4 - Distribuição de medidores de acordo com a faixa de idade. Fonte: [7]**

<b>Faixa de idade</b>	<b>Frequência</b>
Estoque	3,12%
0 a 5 anos	29,86%
6 a 10 anos	21,05%
11 a 15 anos	12,98%
16 a 20 anos	8,31%
21 a 25 anos	7,82%
+ de 25 anos	16,86%

## **1.4 Legislação sobre erros de medição**

Os medidores de energia elétrica possuem leis de aprovação de modelos desde 1966.

Em 2006 foi definida uma lei de verificação inicial dos equipamentos, isto é, que verifica se os erros de medição estão dentro dos limites definidos para a classe de exatidão. Em 2008 surgiram normas para verificação após reparo e o regulamento de ensaio. Em 2009 foi determinada a verificação periódica dos medidores de energia elétrica instalados nas áreas de concessão das distribuidoras de energia elétrica.

A seguir serão apresentados os principais órgãos relacionados com as normas e leis sobre medidores de energia elétrica no Brasil.

### **1.4.1 Agência Nacional de Energia Elétrica (ANEEL)**

Definição: a “Agência Nacional de Energia Elétrica (ANEEL), autarquia em regime especial, vinculada ao Ministério de Minas e Energia (MME), foi criada pela Lei 9.427 de 26 de Dezembro de 1996. Tem como atribuições: regular e fiscalizar a geração, a transmissão, a distribuição e a comercialização da energia elétrica, atendendo reclamações de agentes e consumidores com equilíbrio entre as partes e em benefício da sociedade; mediar os conflitos de interesses entre os agentes do setor elétrico e entre estes e os consumidores; conceder, permitir e autorizar instalações e serviços de energia; garantir tarifas justas; zelar pela qualidade do serviço; exigir investimentos; estimular a competição entre os operadores e assegurar a universalização dos serviços.” [21]

Através da resolução 456 de 29 de Novembro de 2000 a ANEEL define as condições gerais de fornecimento de energia elétrica, a serem observadas na prestação e utilização do serviço público de energia elétrica, tanto pelas concessionárias quanto pelos consumidores. Nesta resolução estão as definições de grupo A e grupo B, unidades consumidoras, etc. O art.

38 afirma que a verificação periódica dos medidores deverá ser efetuada segundo critérios estabelecidos na legislação metrológica.

### **1.4.2 Agência Reguladora de Saneamento e Energia do Estado de São Paulo (ARSESP)**

A Agência Reguladora de Saneamento e Energia do Estado de São Paulo (ARSESP) tem como objetivo “regular, controlar e fiscalizar, no âmbito do Estado, os serviços de gás canalizado e, preservadas as competências e prerrogativas municipais, de saneamento básico de titularidade estadual” [22]. Sua criação ocorreu em 1998 a partir da Comissão de Serviços Públicos de Energia (CSPE).

Na área de energia elétrica ela atua na fiscalização das 14 concessionárias de distribuição que atuam no Estado de São Paulo. Entre suas atividades destacam-se a regulamentação, fiscalização, normalização, monitoramento e definição de indicadores e parâmetros relativos aos padrões dos serviços e manutenção das instalações elétricas. É dividida em três gerências:

- Gerência de fiscalização técnica e comercial;
- Gerência de estudos técnicos e apoio à solução de conflitos;
- Gerência de contratos.

A primeira gerência é responsável pelo recebimento de relatórios de desempenho dos medidores de energia elétrica.

### **1.4.3 Instituto Nacional de Metrologia, Normalização e Qualidade Industrial (Inmetro)**

O Instituto Nacional de Metrologia, Normalização e Qualidade Industrial (Inmetro) é “uma autarquia federal, vinculada ao Ministério do Desenvolvimento, Indústria e Comércio Exterior, que atua como Secretaria Executiva do Conselho Nacional de Metrologia, Normalização e Qualidade Industrial (Conmetro), colegiado interministerial, que é o órgão normativo do Sistema Nacional de Metrologia, Normalização e Qualidade Industrial (Sinmetro).” [23]. Entre as atividades do Inmetro destacam-se:

- Executar as políticas nacionais de metrologia e da qualidade;
- Verificar a observância das normas técnicas e legais, no que se refere às unidades de medida, métodos de medição, medidas materializadas, instrumentos de medição e produtos pré-medidos;
- Fomentar a utilização da técnica de gestão da qualidade nas empresas brasileiras;
- Planejar e executar as atividades de acreditação de laboratórios de calibração e de ensaios, de provedores de ensaios de proficiência, de organismos de certificação, de inspeção, de treinamento e de outros, necessários ao desenvolvimento da infra-estrutura de serviços tecnológicos no País;
- Desenvolvimento, no âmbito do Sinmetro, de programas de avaliação da conformidade, nas áreas de produtos, processos, serviços e pessoal, compulsórios ou voluntários, que envolvem a aprovação de regulamentos.

Através da Resolução 11 de 12 de outubro de 1988, do Conmetro, e Portaria 83 de 1 de junho de 1990, fica entendido que os medidores de energia elétrica utilizados para faturamento em UCs devem: corresponder aos modelos aprovados pelo Inmetro, ser aprovados em verificação inicial, nas condições fixadas pelo Instituto, e serem verificados periodicamente.

A Portaria 285 de 11 de agosto de 2008 estabelece as condições mínimas a serem observadas na fabricação, instalação e utilização de medidores de energia elétrica ativa, inclusive reconicionados, baseados no princípio de indução, monofásicos ou polifásicos. Nesta, os erros percentuais para medidores novos, ou reconicionados, classe de exatidão 2 são de  $\pm 2\%$ .

Para medidores em utilização no parque de medição os erros admissíveis para equipamentos classe de exatidão 2 são de  $\pm 4\%$ . De acordo com as normas da concessionária admitida para este estudo, cujo todos medidores eletromecânicos para medição em UCs são classe 2, o erro máximo admissível é  $\pm 3\%$  e este será o valor limite a ser utilizado neste trabalho.

## 1.5 **Motivação**

Os medidores de energia elétrica com funcionamento eletromecânico possuem uma vida útil econômica estimada em 25 anos. Por vida útil econômica se entende o período de tempo em que ocorre a depreciação do equipamento ou, em termos mais práticos, o período de tempo em que é economicamente viável realizar manutenção no equipamento [9].

Os erros de medição são difíceis de serem detectados, uma vez que geralmente não impedem o fornecimento de energia, apresentando apenas alteração sobre o registro da quantidade de energia elétrica consumida. A modificação da quantidade de energia elétrica aferida dificilmente é atribuída a erros de medição sendo usualmente considerada alteração no padrão de consumo: novos equipamentos instalados, feriados, época do ano, etc.

Os erros de medição podem ser considerados de dois tipos:

- Positivos: quando o valor real é maior que o valor aferido;
- Negativos: quando o valor real é menor que o valor aferido.

A necessidade de se localizar medidores com erros de medição pode ser vista de dois pontos distintos. Pelo lado do consumidor de energia elétrica, medidores com erros negativos apresentam cobrança indevida de consumo, tendo o consumidor que pagar por uma quantidade de energia que ele não utilizou. Pelo lado da concessionária de energia, medidores com erros positivos apresentam redução de lucro ao faturar uma quantidade menor de energia do que o fornecido. Pela legislação, a manutenção dos medidores de energia elétrica é de responsabilidade das concessionárias.

Dada a dificuldade de se determinar quando o equipamento apresenta erros de medição superiores ao limite estipulado pela legislação específica e a necessidade de se localizar tais equipamentos, torna-se necessária uma metodologia de predição de falha de medição para os medidores de energia elétrica. Definindo-se falha como qualquer avaria que o leve ao defeito operacional e ou ao erro de medição maior do que aquele estabelecido pela legislação em vigor, em função de sua classe de exatidão.

Antes de apresentar a proposta deste trabalho, é necessário apresentar o conceito em que ele está baseado. A seguir serão apresentadas as metodologias existentes para determinação de falhas em equipamentos. A falha é uma interrupção permanente da habilidade do sistema de executar uma função requerida sob uma especificada condição de operação [24].

Portanto, no contexto dos medidores de energia elétrica, apresentado neste trabalho, falha fica entendida como a falta de capacidade dos medidores de realizar a aferição dos valores de consumo das UCs dentro dos limites estipulados pela legislação vigente.

## **1.6 Determinação de falha de equipamentos**

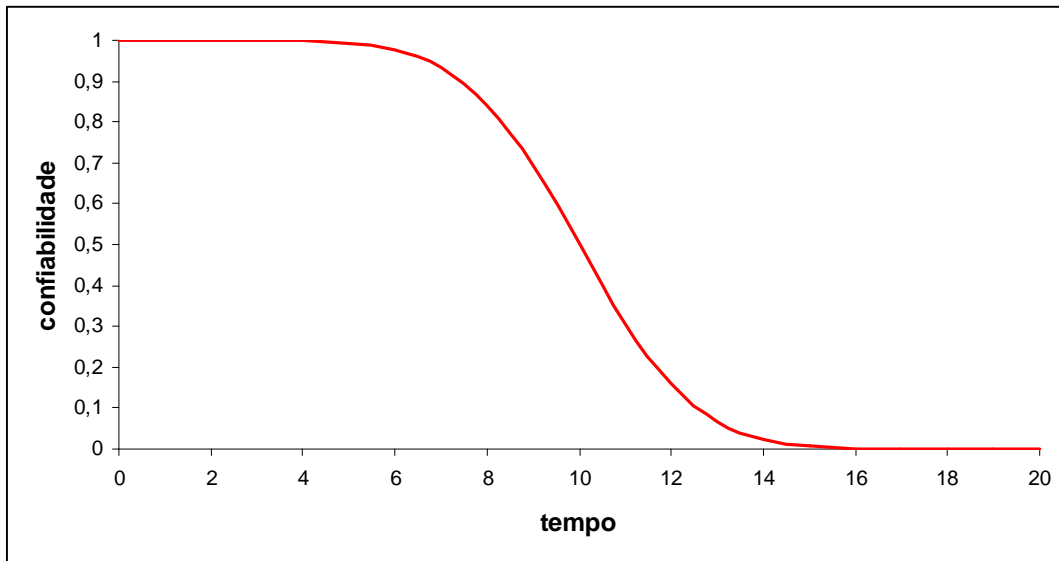
A confiabilidade expressa a quantidade de equipamentos que não falham durante um determinado período. O estudo da confiabilidade é utilizado para definir planos de manutenção preventiva/corretiva de equipamentos, melhoria da qualidade e determinação de prazo de garantia. Exemplo de utilização do estudo da confiabilidade em sistemas de segurança crítica são os reatores nucleares e aviões [24]. Seu estudo influi diretamente sobre o tempo em que um equipamento irá funcionar sem falhar.

Existem diversas definições para confiabilidade. Apresentaremos neste trabalho a definição dada pelo *US Military Handbook* [25]: “A probabilidade que um item irá executar sua função sobre determinadas condições e durante um determinado período de tempo”. A confiabilidade geralmente é utilizada como uma função temporal  $R(t)$ :

$$R(t) = \Pr(T > t) \quad (1.1)$$

Onde  $\Pr$  é a probabilidade do item falhar em um tempo aleatório  $T$  maior do que o tempo em análise  $t$ .

A figura 1.5 apresenta a curva da função de confiabilidade para uma distribuição normal.



**Figura 1.5 – Gráfico de confiabilidade para distribuição normal.**

A palavra confiabilidade provém do inglês “*reliability*” que foi utilizada pela primeira vez em estudos militares na década de 50 [25].

No caso de medidores de energia elétrica, a confiabilidade é a função que define a probabilidade do medidor medir valores com erros inferiores aos máximos determinados por legislação específica durante um determinado período de tempo.

Outra variável relacionada com o estudo da confiabilidade é a taxa de falha. A taxa de falha é definida como “o número total de falhas (percentual) em uma amostra, dividido pelo tempo de análise, durante um intervalo de medição específico sob condições definidas” [26].

Desta forma, a taxa de falha, representada pela letra grega  $\lambda$ , é representada pela equação:

$$\lambda(t) = \frac{\text{falhas}}{\text{amostra} \cdot \Delta t} \quad (1.2)$$

A taxa de falha pode ser expressa em função da confiabilidade como mostra a equação 1.3.

$$\lambda(t) = \frac{R(t) - R(t + \Delta t)}{\Delta t \cdot R(t)} \quad (1.3)$$

Onde  $\Delta t$  é o intervalo de tempo em que se deseja analisar a taxa de falha.

Fazendo o intervalo de tempo muito pequeno, tem-se:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{R(t) - R(t + \Delta t)}{\Delta t \cdot R(t)} \quad (1.4)$$

$$\lambda(t) = -\lim_{\Delta t \rightarrow 0} \frac{R(t + \Delta t) - R(t)}{\Delta t \cdot R(t)} = -\frac{dR(t)}{dt} \cdot \frac{1}{R(t)} \quad (1.5)$$

A curva de taxa de falha representa a probabilidade de o componente falhar no tempo  $t$ . Na figura 1.6 está representada a curva da taxa de falha conhecida como “curva da banheira”. Ela representa grande parte de equipamentos e é dividida em três partes:

- Mortalidade infantil: a taxa de falha é elevada no início da vida e decai para um valor pequeno. Esta característica se deve a equipamentos com defeito de fabricação;
- Vida normal: a taxa de falha é constante e pequena neste período. As causas das falhas são aleatórias;
- Envelhecimento: a taxa de falha cresce até que todos os equipamentos falhem. O motivo para este crescimento é o envelhecimento dos componentes.

Outro termo importante no estudo da confiabilidade é o tempo médio entre falhas, do inglês “*mean time between failure*” (MTBF), que é o inverso da taxa de falha. O MTBF informa o tempo existente entre a ocorrência de duas falhas, informação útil para se determinar os prazos para planos de manutenção de equipamentos. Para taxa de falha constante, o MTBF é dado pela equação 1.6.

$$MTBF = \frac{1}{\lambda} \quad (1.6)$$

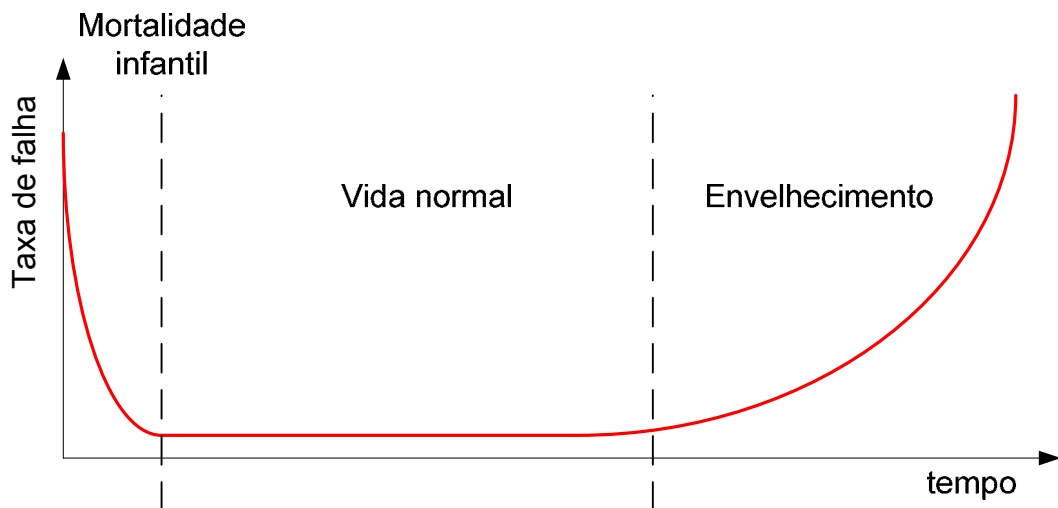


Figura 1.6 – Curva de taxa de falha típica (curva “da banheira”).

Os estudos de confiabilidade são utilizados em alguns campos de mercado, que representam alto risco ou perdas elevadas devido a falhas, tais como: usinas nucleares, aeronáutica, setor petrolífero, setores da indústria química, setor automotivo [27].

## 1.7 Estado da arte

Na referência [28] é apresentado um modelo de predição para falência de empresas, utilizando taxa de falha. Neste contexto, falha representa a saída do grupo de empresas consideradas saudáveis.

Em [29] são comentadas aplicações de inteligência artificial para análise de falha de aviões, sistema especialista para auxílio na tomada de decisão da manutenção programada de gerador da Taiwan Power Company, diagnóstico de transformadores de potência utilizando redes neurais, sistema especialista para manutenção preditiva de motores diesel.

Estudos de confiabilidade de compressores utilizados em sistema de refrigeração são apresentados em [25].

Em [30] foi desenvolvido um algoritmo para agendamento de manutenções utilizando a confiabilidade como uma das variáveis.

Em [31] foi realizada a análise da taxa de falha para cubículos classe 15 kV. O autor revela que o estudo da confiabilidade em sua conotação clássica é inválido quando utilizado em espaços amostrais e períodos de observação diferentes, não permitindo a utilização dos métodos de distribuição estatísticos clássicos como Weibull, Log-normal e Poisson.

Em [32] foi apresentada uma metodologia para automação do processo de aferição de medidores visando aumentar a eficiência dos planos de inspeção. No trabalho, o autor utilizou um banco de dados com informações dos medidores cadastrados no parque de medição de uma empresa. Os medidores foram agrupados em “famílias”, utilizando como critério de agrupamento o modelo do medidor. Foi desenvolvido um critério para se determinar se uma dada família necessita de calibração, através de amostragem: estabelecendo-se que, uma vez que um equipamento da família apresente falha, todos os equipamentos daquela família com data de instalação anterior devem ser aferidos. Um sistema computacional foi desenvolvido para a edição dos dados e apresentação dos resultados.

## **1.8 Metodologia**

A metodologia proposta em [32] fornece um direcionador para elaboração de planos de amostragem, porém não é fornecida no texto a base científica utilizada para geração das regras. Ainda assim, o agrupamento utilizando unicamente o modelo do medidor como variável de agrupamento pode ser uma estratégia pouco eficiente, tendo em vista que os medidores de energia estão sujeitos às características próprias dos equipamentos, do meio ambiente e das suas condições gerais de utilização.

Considerando que o medidor eletromecânico utilizado na medição de energia elétrica ativa é um equipamento que permanecerá, ainda, em uso por bastante tempo (superior a 10 anos) no mercado brasileiro, este trabalho concentrará seu foco somente neste tipo de medidor.

Neste trabalho uma nova metodologia para determinação de probabilidades de falhas dos equipamentos é proposta. Utilizando um banco de dados de uma concessionária de energia, serão identificadas, dentre um leque de variáveis, as que são mais significativas na determinação de falhas dos equipamentos. Com o uso destas variáveis uma metodologia de agrupamento de equipamentos em classes de probabilidades de falhas é proposta.

Como resultado, um aplicativo é desenvolvido em plataforma amigável, utilizando a metodologia proposta e os dados do parque de medição da concessionária para determinar as probabilidades de falhas dos equipamentos.

Neste contexto, a dissertação foi estruturada em capítulos, sendo que o capítulo um aqui desenvolvido foi destinado à introdução geral, revisão bibliográfica, motivação e metodologia para o trabalho.

O capítulo dois trata do problema de escolha das variáveis mais significativas para a determinação de falhas em medidores de energia, apresentando o banco de dados da concessionária, o leque de variáveis, a metodologia utilizada para a identificação das variáveis mais significativas na determinação de falhas nos medidores eletromecânicos de energia elétrica ativa e as variáveis que compõem o modelo de agrupamento.

No capítulo três é construído e apresentado o modelo estatístico de agrupamento dos equipamentos em classes de probabilidades de falhas.

O capítulo quatro apresenta o aplicativo desenvolvido para a análise das probabilidades de falhas dos equipamentos e suas funcionalidades.

Um estudo de caso é apresentado no capítulo cinco, sendo que as conclusões finais e proposta de continuidade da pesquisa são apresentadas no capítulo seis.

## 2 SELEÇÃO DAS VARIÁVEIS MAIS SIGNIFICATIVAS PARA DETERMINAÇÃO DE FALHAS EM MEDIDORES ELETROMECCÂNICOS DE ENERGIA ELÉTRICA ATIVA

---

---

### 2.1 *Considerações Iniciais*

Determinar um modelo matemático para os erros de medição é um problema NP-difícil [33], *i.e.*, a modelação é impossível ou muito complexa. Como alternativa, deve-se utilizar técnicas de inteligência artificial, como heurísticas, metaheurísticas, lógica *fuzzy* ou outra metodologia.

A Descoberta de Conhecimento em Bases de Dados (DCBD) é uma metodologia que visa determinar padrões existentes em bases de dados. DCBD é um processo que faz uso de várias ferramentas estatísticas e de aprendizagem de máquina para obter os resultados. Neste trabalho, DCBD será utilizada para determinação das variáveis mais significativas na ocorrência de falhas em medidores eletromecânicos de energia elétrica ativa.

Este capítulo aborda todo o processo de descoberta de conhecimento em bases de dados, apresentando as ferramentas e objetivos de cada etapa. Inicialmente será apresentada a imagem do banco de dados fornecido pela concessionária de energia elétrica, objeto de estudo deste trabalho. Serão apresentadas as variáveis utilizadas na pesquisa e seus significados.

O objetivo deste capítulo é descobrir quais são as variáveis mais significativas, dentre as utilizadas no processo de DCBD, para determinação dos casos de medidores que apresentam falhas.

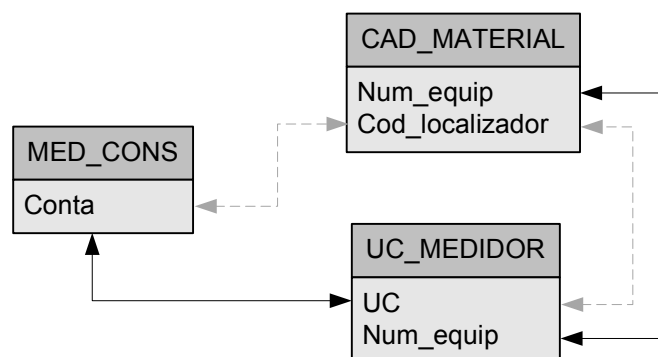
### 2.2 *O banco de dados da concessionária*

A empresa disponibilizou uma imagem de seus dados no mês de junho de 2008, os quais foram reproduzidos no Laboratório de Eletrônica de Potência (LEP) utilizando o banco de dados MySQL 5.1 Community Server.

O banco de dados recebido constitui-se de várias tabelas que agregam informações sobre diversas áreas, desde o serviço de atendimento ao consumidor até o consumo mensal das UCs. As tabelas de interesse para esta pesquisa são:

- **CAD\_MATERIAL:** Tabela com informações de cadastro de equipamentos destinados para medição de energia elétrica: medidores de energia elétrica, transformadores de corrente, registradores. A tabela contém informações detalhadas dos medidores de energia, como quantidade de elementos motores, constante de fabricação, tipo de mancal, presença de elo e localização do equipamento em campo;
- **MED\_CONS:** Tabela com informações de consumo mensal de várias unidades consumidoras no período de seis anos e seis meses (de janeiro de 2002 a junho de 2008);
- **UC\_MEDIDOR:** Esta tabela contém informações sobre cada instalação e retirada de medidores de energia elétrica em unidades consumidoras até junho de 2008. Cada registro (linha) da tabela representa uma instalação de medidor de energia em uma unidade consumidora. A duração da instalação é indicada pelas variáveis `data_inst Equip` e `data_retirada Equip`, caso o equipamento ainda esteja instalada `data_retirada Equip` assume valor 0 (e.g., "0000-00-00"). A tabela apresenta informações detalhadas sobre as unidades consumidoras (e.g., nome cliente, endereço, tensão nominal) e sobre os equipamentos (e.g., ano de fabricação, modelo do equipamento e número de identificação).

As tabelas descritas acima estão relacionadas por uma, ou mais variáveis, como mostra a figura 2.1. A variável `cod_localizador` pode apresentar a informação das variáveis `conta` e `UC` quando uma outra variável, que indica o tipo de local em que o equipamento se encontra, assumir um valor específico.



**Figura 2.1 – Relacionamento entre as tabelas da concessionária.**

Os dados recebidos da concessionária já compõem um depósito de dados, mas ainda é necessário eliminar variáveis não necessárias no âmbito do trabalho.

## 2.3 *Descoberta de Conhecimento em Bases de Dados*

A automação da tarefa de armazenamento de informações através de computadores possibilitou ao ser humano armazenar quantidades de dados muito além da capacidade humana. Estimava-se, em 2005, que a maior base de dados do mundo possuía 222 TB (terabytes) [34].

A quantidade de dados armazenados tem crescido em dois aspectos: a quantidade de registros em uma base de dados; a quantidade de variáveis armazenadas nos bancos de dados. Segundo [35], bancos de dados com  $10^9$  registros e/ou  $10^3$  variáveis já eram comuns em 1996. Segundo [36], a quantidade de dados no mundo duplica a cada ano, enquanto a quantidade de informações nos dados diminui drasticamente.

Infelizmente, não existe uma relação direta entre a quantidade de dados e a quantidade de informações. Analisar grande quantidade de dados e obter informações de forma manual é uma tarefa dispendiosa, lenta e altamente subjetiva.

Como consequência da grande quantidade de dados armazenada, e da falta de informações nos mesmos, surgiu a necessidade de automatizar o processo de análise e obtenção de informações dos dados. A necessidade de se trabalhar os dados aparece através da frase em [37]: “Somos ricos em dados, porém, pobres em informação”. Outro exemplo claro da necessidade de se obter conhecimento dos dados é a frase de John Naisbett [38]: “Nós estamos afogando em informações, mas famintos por conhecimento”.

A Descoberta de Conhecimento em Bases de Dados é um processo que visa obter padrões a partir de grandes bancos de dados. Ela integra diversas áreas de conhecimento como estatística, aprendizagem de máquina, sistemas especialistas, reconhecimento de padrões e banco de dados para extrair conhecimento de alto nível a partir de dados de baixo nível no contexto de grandes bases de dados [35]. A figura 2.2 apresenta as áreas que compartilham conhecimento com DCBD.



**Figura 2.2 – Relação de DCBD com outras áreas.**

O interesse de DCBD é extrair padrões que possam se tornar conhecimento a partir de dados. Utiliza um processamento de dados, métodos e ferramentas de mineração de dados e o resultado pode ser: definição de modelos; predição de comportamentos futuros; reconhecimento de padrões; determinação de relacionamentos entre variáveis; classificação de dados em grupos (*clustering*); etc. Exemplos de áreas de utilização de DCBD são [35]: publicidade e propaganda, finanças, detecção de fraudes, telecomunicações, indústria. Casos da utilização de DCBD na ciência são apresentados em [39].

A maior barreira existente no processo de DCBD está na limitação dos dados. São raras as vezes em que os dados são coletados com o objetivo de mineração de conhecimento. De um modo geral, as informações existem como subproduto de outras tarefas. Exemplo: no caso de definição de tendências de compras, onde a venda de um produto está relacionado com a venda de outro, a existência de registros de venda de ambos os produtos foi armazenada como informação fiscal, e não para o objetivo de mineração. Caso o objetivo fosse a mineração de conhecimento poderiam ser armazenados dados com informações mais valiosas.

O processo de DCBD consiste de diversas etapas sendo a etapa de Mineração de Dados (*Data Mining*) uma etapa de destaque no processo [35]. Nesta, informações são extraídas dos dados. Pelo fato de ser o núcleo do processo de DCBD, Mineração de Dados (MD) é a etapa que possui maior quantidade de referências na literatura específica. Muitos autores, como [40], confundem os termos DCBD e MD, utilizando o nome do segundo para referenciar o primeiro. Conseqüentemente, ambos os termos estão sendo aceitos, mas utilizaremos DCBD neste trabalho.

Ressalta-se que apesar de MD possuir a maior quantidade de referências e de ser de elevada importância, as outras etapas do processo não podem ser desconsideradas. Em verdade, elas garantem a confiabilidade dos resultados obtidos na etapa de MD. Em torno de 80% do tempo do processo de DCBD está relacionado com a preparação dos dados, enquanto que os 20% restantes estão relacionados com a MD [36 e 41].

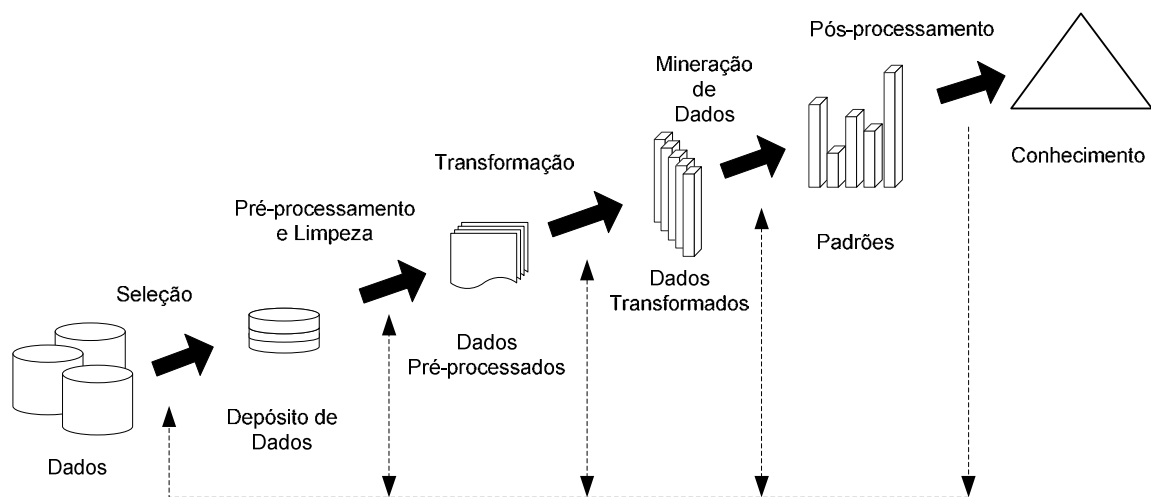
São as demais etapas do processo que diferenciam DCBD das diversas áreas em que DCBD está envolvida. Enquanto estas áreas estão preocupadas apenas com as técnicas de obtenção do conhecimento, DCBD se preocupa com todo o processo de pré-processamento e limpeza dos dados, garantindo melhores resultados. Pode-se dizer que DCBD é um processo que se preocupa desde o modo como os dados são armazenados até como o conhecimento é obtido e validado. As demais áreas apresentadas na figura 2.1 estão incluídas no processo de descoberta de conhecimento.

Na década de 60, a aplicação de técnicas de MD sem o devido tratamento dos dados foi chamada, de forma pejorativa de *data fishing* ou *data dredging*, uma vez que os padrões obtidos podem não apresentar significado verdadeiro.

A partir de 1990, o nome Mineração de Dados foi utilizado no ramo comercial para designar DCBD. O termo Descoberta de Conhecimento em Bases de Dados foi apresentado pela primeira vez em 1989, utilizado pela comunidade de aprendizagem de máquina e inteligência artificial.

Outros nomes utilizados são: *Data Archaeology*, *Information Harvesting*, *Information Discovery*, *Knowledge Extraction*.

As etapas do processo de DCBD e sua ordem são: seleção, pré-processamento e limpeza, transformação, mineração de dados e pós-processamento. A figura 2.3 apresenta um diagrama do processo de DCBD.



**Figura 2.3 – Processo de DCBD. Fonte: [35]**

O processo de DCBD é interativo e iterativo. Interativo porque é possível interagir nas etapas do processo; iterativo porque o processo pode ocorrer mais de uma única vez para um mesmo conjunto de dados, permitindo o retorno para uma etapa anterior do processo. Por exemplo: durante a transformação dos dados pode-se ter notado que uma variável contém dados redundantes, sendo necessário retornar a etapa de pré-processamento.

Em [35], DCBD é definido como: “o processo não-trivial, interativo e iterativo, para identificação de novos padrões, válidos, potencialmente úteis e altamente compreensíveis”.

## 2.4 Seleção

Os dados geralmente estão em bancos de dados que agregam informações de diversas áreas de conhecimento e estão sujeitos a alterações.

Na etapa de seleção os dados são armazenados em um Depósito de Dados (*Data Warehouse*), local em que os dados se encontram de forma consolidada, disponíveis para análise.

A seleção de variáveis ocorre de maneira empírica sendo recomendado que um especialista atue no processo uma vez que a remoção de dados acarreta em perda de informação. É necessário eliminar somente variáveis desnecessárias, isto é, variáveis não-relacionadas com o conhecimento que se pretende atingir. Eliminação de dados resulta em eliminação de conhecimento.

A utilização de muitas variáveis para a mineração eleva o tempo computacional e a utilização de variáveis desnecessárias pode tornar o conhecimento obtido confuso ou levar a aprendizagem de ruído.

O resultado da etapa de seleção pode ser um banco de dados, a junção de bases de dados distintas ou um subconjunto de um banco de dados.

No banco de dados da concessionária a análise da tabela `CAD_MATERIAL` demonstra que ela se destina a descrever os medidores de energia elétrica pertencentes ao ativo da concessionária. A tabela apresenta dados construtivos dos equipamentos e a localização (referente a junho de 2008) dos mesmos. Apesar das características construtivas dos equipamentos serem relevantes para determinação de falhas nos mesmos a tabela foi desconsiderada no estudo, uma vez que a variável `subtipo equip`, existente tanto na tabela `CAD_MATERIAL` quanto `UC_MEDIDOR`, agrega as características construtivas do equipamento. A idéia é formular a solução do problema de maneira simplificada, de modo a possibilitar a construção de um sistema especialista.

A tabela `UC_MEDIDOR` contém várias informações das instalações dos medidores nas UCs, dados sobre as UCs e dados sobre os medidores. As variáveis selecionadas desta tabela para o processo de DCBD são organizadas em três grupos:

- Dados referentes ao medidor de energia elétrica:
  - `subtipo equip`: modelo do equipamento.
  - `ano_fabric`: ano de fabricação do equipamento.

- `cod_tipo_fase`: quantidade de fases que o medidor é conectado (e.g. "1" = monofásico).
- `contabil`
- Dados referentes à UC:
  - `cod_tipo_local`
  - `cod_localidade`: município.
  - `cod_classe_principal`: classes para ação tarifária. O valor de cada classe é definido em pela resolução ANEEL nº 456 de 29 de novembro de 2000.
  - `cod_classe_consumo`: subclasses para ação tarifária.
  - `grupo_tensao_fornecimento`: tensão no ponto de entrada da UC.
- Dados referentes ao tempo de existência do medidor:
  - `data_inst Equip`: data em que o equipamento foi instalado na UC.
  - `data_retirada Equip`: data em que o equipamento foi retirado da UC. Valor "0" representa que o medidor não foi retirado.

Os dados contidos na tabela `UC_MEDIDOR` são referentes às instalações de todos os medidores do ativo da concessionária. Para que os dados representem somente os medidores eletromecânicos de energia ativa, instalados em unidades consumidoras, foi necessário realizar um processo de filtragem.

As informações para a filtragem são obtidas das variáveis `contabil`, `cod_tipo_espec` e `tipo Equip`. O valor da variável `cod_tipo_espec` igual a "CON" indica que o equipamento é utilizado para medição de energia elétrica nas unidades consumidoras. A utilização desta variável é importante dada a existência de medidores que são utilizados para outros fins. Um exemplo é o medidor de fronteira, utilizado para medir a energia transmitida por um pedaço do sistema de transmissão ou distribuição (fronteira). A quantidade de medidores de fronteira é baixa, sendo possível desenvolver planos de inspeções que verifiquem os estados de todos equipamentos. A variável `tipo Equip` representa o tipo de equipamento, o valor "MD" significa que o equipamento é um medidor de energia elétrica. A variável `contabil` é uma aglutinação de dados de outras variáveis sobre os equipamentos, algumas inexistentes nas tabelas recebidas. Através da variável é possível obter o número de fases que o equipamento deve ser conectado (e.g., monofásico), determinar o tipo de medição a que se destina (e.g., medição indireta), se o equipamento é eletrônico ou eletromecânico, a tensão de operação e o tipo de equipamento (e.g., transformador de corrente).

As variáveis selecionadas da tabela MED\_CONS são o número da unidade consumidora (necessário para relacionar os dados desta tabela com os dados de UC\_MEDIDOR), fornecida pelo campo conta, e todos os valores de consumo mensal, em kWh.

Uma vez analisados os dados percebe-se que as tabelas da concessionária não apresentam o conhecimento necessário para a classificação dos equipamentos, *i.e.*, não existe no depósito de dados informações sobre o valor do erro de medição dos medidores de energia elétrica ou variável similar, que informe se o erro de medição do equipamento está dentro dos limites permitidos pelas leis dos órgãos de regulamentação. A única variável que representa algo mais próximo disto é motivo\_retirada que indica o motivo da retirada do equipamento da UC em análise (registro). Se a retirada do equipamento foi devido a defeito no mesmo, então motivo\_retirada apresentará o valor "RETIRADA DO EQUIPAMENTO COM DEFEITO". Apesar de a variável apresentar informações sobre medidores defeituosos ela não é válida para a análise dos dados uma vez que a empresa informou que medidores podem ser retirados por outros motivos (*e.g.*, motivo\_retirada = "FALTA DE PAGAMENTO"), serem inspecionados e se constatar defeito no equipamento, sem a alteração da variável motivo\_retirada, isto gera uma incerteza sobre os medidores definidos como "dentro das conformidades".

Com o intuito de se obter respostas precisas sobre quais medidores estavam ou não com erros metrológicos dentro dos limites estabelecidos pelas leis dos órgãos de regulamentação recorreu-se aos ensaios realizados anualmente para a ARSESP. Foram utilizados os dados de ensaios dos anos de 2004 a 2008. A figura 2.4 apresenta um exemplo de uma tabela de ensaio de medidores para a ARSESP.

Fabricante		Data		Quantidade		Modelo		
GE		01/12/2008		382		D58C		
Tensão (V)	Kd	Rr	Fios	Fases	Elementos	I peq (A)	I Máx (A)	
120	10,8	92X16/27	4	3	3	15	100	
Número do Medidor	Leitura do Medidor	Número do lacre	Teste Dielétrico	Conjunto			Erro Médio	Observações
				C. Nominal	C. Indutiva	C. Pequena	C. Nom / C. Peq.	
			1 KV	100% - FP1,0	100% - FP0,5	10% - FP1,0	100%/10% FP1,0	
3T113174	1338	POTEM 2005	APROVADO	0,928	1,500	1,635	1,282	
3T1120212	6046	L24	APROVADO	4,975	5,266	5,817	5,396	
3T114019	0569	L00	APROVADO	1,824	2,204	5,824	3,824	
3T112669	4033	L18	APROVADO	-0,926	-0,466	-2,786	-1,856	
3T114004	0001	L63	APROVADO					Disco enroscando.
3T1121464	2834	L51	APROVADO	0,600	1,091	-0,457	0,072	
3T111626	2755	L01	APROVADO	-0,679	0,093	-2,506	-1,593	
3T1139398	2251	L02	APROVADO	0,193	0,806	1,897	1,045	
3T1118323	9607	L35	APROVADO	15,590	17,760	15,510	15,550	
3T109108	5623	L12	APROVADO	0,675	1,820	0,231	0,453	
3T110660	7422	L12	APROVADO	-0,233	0,195	-2,977	-1,605	
3T1093509	4867	L35	APROVADO	0,511	2,244	6,586	3,549	
3T113876	5369	L12	APROVADO	-1,428	-2,082	-2,033	-1,731	
3T1105167	1697	L54	APROVADO	1,168	1,868	2,022	1,595	
3T1098195	1128	L54 SEM 1 LACRE	APROVADO	-1,382	-1,386	-0,026	-0,704	
3T1141201	2204	L54	APROVADO	0,055	0,640	-0,717	-0,331	
3T110459	6973	L11	APROVADO	1,102	1,748	1,317	1,210	
3T111962	1283	L63	APROVADO	0,097	1,500	0,588	0,343	
3T1095919	7759	L11	APROVADO	-0,157	-0,542	-17,300	-8,729	Limalha no imã.

Figura 2.4 – Exemplo de relatório no padrão ARSESP.

Os dados de ensaios para a ARSESP foram acrescentados ao depósito de dados em uma nova tabela, denominada DADOS, com três campos. A tabela 2.1 apresenta o descritivo de DADOS. A tabela DADOS contém 11.068 registros.

**Tabela 2.1 – Descrição das variáveis da tabela DADOS.**

<b>Variável</b>	<b>Tipo</b>	<b>Descrição</b>
num equip	Alfanumérica	nº de identificação do equipamento
ano	Numérica	ano do ensaio
defeito	Booleana	Determina se o equipamento apresentou erros de medição fora dos limites permitidos por lei (verdadeiro) ou não (falso).

## **Dados climáticos**

Dados climáticos, como temperatura e precipitação média, são variáveis que inicialmente parecem influenciar a vida útil dos equipamentos. O conjunto temperatura, precipitação média e altitude podem influenciar na velocidade do desgaste dos equipamentos, afetando o erro de medição. Esses desgastes podem ser devido à:

- Salinidade, para localidades em que a altitude seja baixa;
- Umidade, para localidades que apresentem alta precipitação média;
- Dilatações térmicas em localidades com temperatura média elevada.

O banco de dados não apresenta informações sobre essas três variáveis (altitude, precipitação e temperatura). Para obter tais dados, foi utilizada a base de dados meteorológicos fornecida pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa), através do Banco de Dados Climáticos do Brasil<sup>1</sup>. No sítio são apresentados os dados históricos de temperatura e precipitação por município, informando o período de análise e a média dos valores para cada um dos doze meses. Também são apresentadas as coordenadas geográficas e a altitude do município.

As informações climáticas de algumas localidades não estão disponíveis no sítio. Para garantir informações de todas as localidades, utilizou-se a metodologia de regressão kernel nos dados da Embrapa para estimar os valores de temperatura média anual, precipitação média mensal e altitude para todas as localidades.

---

<sup>1</sup> <http://www.bdclima.cnpm.embrapa.br>

As informações foram obtidas por localidade e armazenadas no banco de dados em uma nova tabela, CLIMATICOS. A tabela 2.2 apresenta uma descrição dos campos de CLIMATICOS.

**Tabela 2.2 – Descrição das variáveis da tabela CLIMATICOS.**

Variável	Tipo	Descrição
localidade	Numérica	Código do município.
temperatura	Numérica	Temperatura média.
precipitacao	Numérica	Precipitação média mensal.
altitude	Numérica	Altitude do município.

## 2.5 Integração de dados

“Integração de dados é o problema de combinar dados que residem em diferentes fontes, e prover ao usuário uma visão unificada desses dados” [42].

A integração de dados trata da formulação das consultas (*queries*) no banco de dados, garantindo que os dados sejam combinados de modo correto em uma mesma tabela. Para o caso em estudo, algumas tabelas apresentam informações sobre os equipamentos de forma indireta e serão modificadas.

Na tabela UC\_MEDIDOR existe o histórico do medidor, porém a forma de armazenamento utilizada, através de vários registros, não apresenta facilidade para obtenção de dados históricos, tornando-os inutilizáveis. A transformação dos dados históricos, armazenados através de vários registros, em variáveis permite fácil acesso às informações sobre o histórico do medidor além de diminuir a quantidade de registros existentes.

O dado histórico considerado importante para o problema é o tempo de vida do medidor no parque, porém tal informação não pode ser obtida com os dados disponíveis. Como alternativa, calculou-se a quantidade de instalações do medidor de energia e a quantidade de dias que o medidor permaneceu na última unidade consumidora (a qual o medidor estava quando retirado para ensaio). Uma instalação é definida como a contagem de registros com mesmo valor do campo num equip na tabela UC\_MEDIDOR.

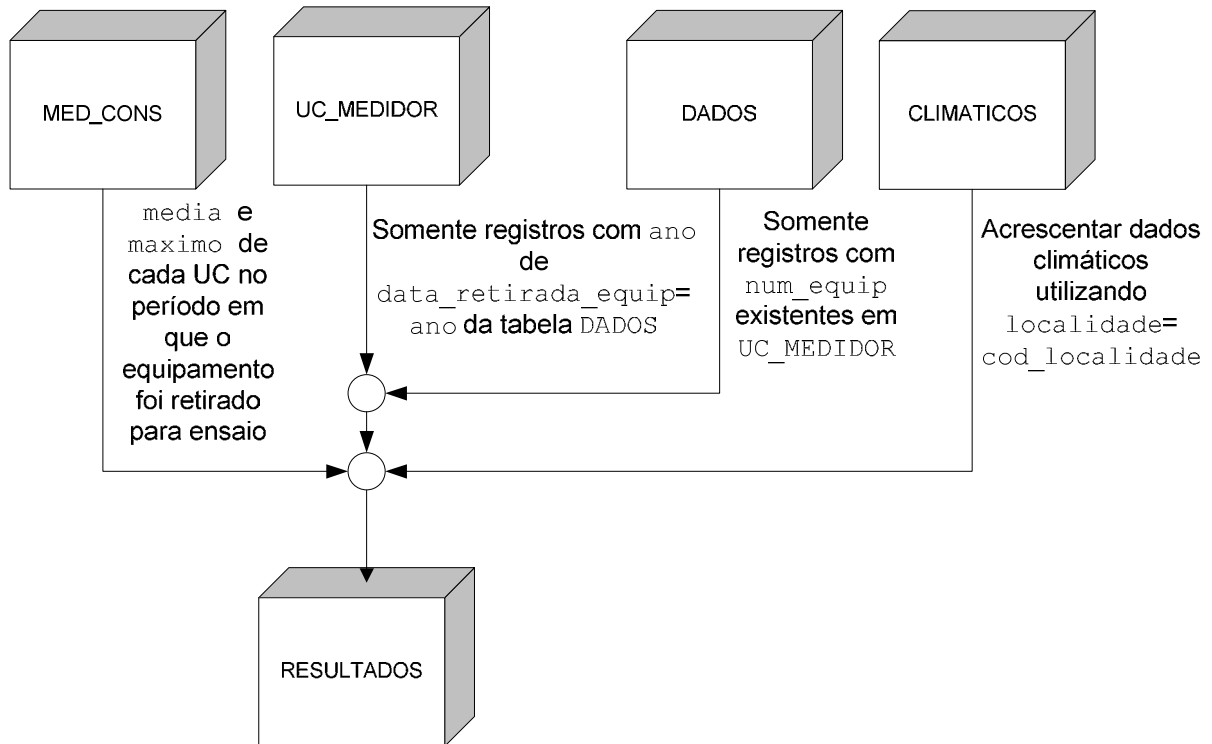
Para a integração com os dados obtidos dos relatórios para a ARSESP, utilizou-se o campo data\_retirada equip para obter a quantidade de instalações de acordo com os dados da tabela DADOS. Somente foram utilizados para determinação da quantidade de instalações os registros que possuem o valor do ano de data\_retirada equip inferior ou igual ao valor de

ano em DADOS. A informação do número de instalações do equipamento foi armazenada em um novo campo, nomeado `instalacoes`.

O cálculo do tempo de permanência em uma UC é facilmente obtido através da diferença entre as variáveis `data_inst equip` e `data_retirada equip` da tabela UC\_MEDIDOR. A restrição é que este cálculo seja realizado somente nos registros em que o medidor foi retirado para ensaio, isto é realizado cruzando as informações de UC\_MEDIDOR com a variável ano em DADOS.

Como já mencionado, a tabela MED\_CONS contém o histórico do consumo mensal das UCs. Esta informação está relacionada com a UC, estando relacionada indiretamente com o equipamento. Duas maneiras foram utilizadas para relacionar os dados desta tabela com os medidores: uma maneira, bem simples, é a determinação do máximo valor de consumo mensal da UC. A outra maneira é o cálculo do consumo médio mensal. Ambos os métodos utilizam somente os dados de consumo mensais da UC em que o medidor se encontrava quando retirado para ensaio ARSESP. Este período de tempo é calculado do mesmo modo que o tempo de permanência em uma UC. A informação do consumo máximo foi armazenada em uma variável denominada `maximo` enquanto a informação do consumo médio foi armazenada em uma variável denominada `media`.

De acordo com a metodologia os dados devem ser agrupados em uma única tabela. A tabela RESULTADOS foi criada para armazenamento de todas variáveis de interesse. Os dados da tabela final são preenchidos através de uniões condicionais entre as tabelas. A figura 2.5 apresenta de maneira simplificada como os dados foram agrupados.



**Figura 2.5 – Construção da tabela RESULTADOS.**

As variáveis que compõem a tabela RESULTADOS, e que serão utilizadas no processo de mineração de dados, são:

- Dados referentes ao medidor de energia elétrica:
  - subtipo\_equip
  - ano\_fabric
  - cod\_tipo\_fase
  - contabil
- Dados referentes à UC:
  - media: consumo médio mensal calculado no período em que o equipamento foi instalado até sua retirada para ensaio.
  - maximo: consumo máximo observado no período em que o equipamento foi instalado até sua retirada para ensaio.
  - cod\_tipo\_local
  - cod\_localidade
  - cod\_classe\_principal
  - cod\_classe\_consumo
  - grupo\_tensao\_fornecimento
  - temperatura
  - precipitação

- altitude
- Dados referentes ao tempo de existência do medidor:
  - instalacoes: armazena a quantidade de instalações que o medidor teve até retirada para ensaio.
  - dias\_de\_vida: tempo (em dias) que o medidor permaneceu na UC antes de ser retirado para ensaio.
- Dados de indicação de erro acima do permitido:
  - defeito

### 2.5.1 Agrupamento de modelos

Os equipamentos no banco de dados recebido possuem uma grande quantidade de modelos – mais de 200 – o que dificulta a utilização da variável. Em consulta com especialistas da concessionária verificou-se que existem modelos que possuem mais de um nome na variável `subtipo Equip`, a criação de nomes fictícios ajuda os profissionais que não possuem facilidade no uso de banco de dados a encontrarem grupos específicos de equipamentos. Como exemplo destaca-se os valores "C1X412010" e "C1X424010" da variável `subtipo Equip` que representam um mesmo modelo, o "C1X4". Modelos "C1X412010" são utilizados em tensões da classe 127V enquanto os modelos "C1X424010" são utilizados em tensões da classe 220V.

Os modelos que podem ser agrupados e foram utilizados neste trabalho são apresentados no anexo A.

## 2.6 *Pré-processamento e limpeza dos dados*

A validade do conhecimento obtido depende fortemente da qualidade dos dados utilizados. Dados que apresentam informações discrepantes e campos com valores ausentes ou incorretos aumentam a possibilidade de aprendizagem de ruído. Um modelo construído a partir de dados com ruídos “aprende” o ruído, tornando a representação de novos dados ineficiente. Portanto, é necessário que haja uma etapa de pré-processamento e limpeza dos dados.

Em [43] cita-se os tipos de erros que um depósito de dados pode apresentar e as formas de tratamento, neste trabalho destacam-se:

### 2.6.1 Dados ausentes (*missing values*)

Erros deste tipo são: campos vazios ou variáveis não preenchidas nas bases de dados. A justificativa pode estar em erros no programa de inserção dos dados, anomalias no sistema ou erro computacional (*bug*).

Variáveis sem nenhuma informação (vazias) devem ser removidas do Depósito de Dados. Valores de campos vazios podem ser estimados ou então todo o registro ser removido. Métodos para predição dos dados ausentes são apresentados em [44].

#### No banco de dados

A análise de dados ausentes ocorre através da verificação de variáveis com valor nulo. Para a tabela `MED_CONS`, campos não preenchidos em um determinado registro representam a inexistência do histórico da UC no mês. O motivo para a inexistência do histórico pode ser a inexistência da UC naquele mês (caso para UCs novas) ou o desligamento da UC (à pedido ou por falta de pagamento). Exemplo para ambos os casos são apresentados na figura 2.6.

Neste caso específico, os dados ausentes não são devido a anomalias nem erros computacionais, e sim consequência da configuração do banco de dados. Justamente por causa destas configurações, os registros nulos pertencem a datas fora dos períodos considerados para consumo, não sendo selecionados para compor os cálculos. Esta propriedade específica da tabela `MED_CONS` inviabiliza a busca por campos com dados ausentes (registros nulos em períodos onde existe consumo). Como medida de segurança para cálculos (que tornam o resultado de uma operação matemática nula quando existe um valor nulo), valores nulos são convertidos para valor igual a zero.

78	555169	0181	UB	B	TR	466	128	15	2	TR	466	128	15	25	69	172	236	186	93			
79	555177	0181	UB	B	BI	10	0	20	50	7	BI	10	0	20	50	70	110	160	180	150	2	
80	555185	0181	UB	B	BI					7	BI	163	177	157	25	0	0	0				
81	555193	0181	UB	B	BI	66	39	44	60		BI	66	39	44	60	67	92	78	85	181	86	6
82	555207	0181	UB	B	TR	110	160	70	1		TR	110	160	70	180	110	80	80	70	120		
83	555215	0181	UB	B	BI	148	211	163			BI	148	211	163	192	220	184	192	206			

(a)

(b)

Figura 2.6 – Exemplo de dados ausentes na tabela `MED_CONS`. a) Caso de UC inexistente até dado período.  
b) Caso de UC desligada.

A variável `ano_fabric` após importação dos dados da concessionária apresentou 11.184 registros nulos. Optou-se por tratar este registro somente após a criação da tabela RESULTADOS.

Analisando a tabela final, apenas um registro nulo foi encontrado. A análise dos dados originais da empresa apresenta a primeira instalação do equipamento em 1998. Na análise dos dados originais os demais equipamentos pertencentes ao mesmo modelo agrupado ("FYT201") também possuem registro vazio para a variável `ano_fabric` e datas de primeira instalação em uma faixa grande de datas, tornado impossível estimar o valor do registro nulo. O registro foi descartado. Este era o único registro do modelo "FYT201" na tabela RESULTADOS.

## 2.6.2 Dados incorretos

Dados incorretos são dados presentes em um campo mas que não pertencem ao grupo de valores daquele campo. Um exemplo claro de dado incorreto é a presença do valor 5 no campo que indica o número de fases de uma instalação elétrica a qual um medidor de energia é conectado. Atualmente trabalha-se com instalações elétrica de 1 (monofásica), 2 (bifásica) ou 3 (trifásica) fases. O número 5 como quantidade de fases em uma instalação é uma informação inválida, incorreta.

A importação de dados externos nos bancos de dados pode resultar em quantidade de campos menor do que o esperado no banco de dados. Se não for utilizado um mecanismo de proteção, as informações no banco de dados podem apresentar os dados de uma variável em um campo diferente, alterando o conhecimento existente nos dados.

### No banco de dados

Foram observados dois problemas de dados incorretos. O primeiro, observado durante a importação dos dados, foi a quebra dos registros em duas ou mais linhas, *i.e.*, um caractere de nova linha estava presente no meio dos dados de importação. A figura 2.7 apresenta um exemplo de dados com quebra dos registros em duas linhas.

```

10 0001|MD|00000091|CVAD58|1001|04|EQUIPAMENTO SUCATEADO|24-JAN-02|1|1990|1|1|120|2|2|P||ALX|A
11 0001|MD|00000140|CVAD58|1001|04|EQUIPAMENTO SUCATEADO|24-JAN-02|1|1990|1|1|120|2|2|P||ALX|A
12 0001|MD|00000171|CVAD58|1001|04|EQUIPAMENTO SUCATEADO|24-JAN-02|1|1990|1|1|120|2|2|P||ALX|A
13 0001|MD|00000325|CVAD58|1001|04|EQUIPAMENTO
14 SUCATEADO|24-JAN-02|1|1990|1|1|120|2|2|P||ALX|AL.001-2500000.0001|01|1|1|CON|100000|1|1|KW
15 0001|MD|00000351|98AV120|1000|04|EQUIPAMENTO SUCATEADO|12-JUN-03|1|98|1|1|120|3|1|2|ALX|AL.C
16 0001|MD|00000357|CVAD58|1000|04|EQUIPAMENTO SUCATEADO|12-JUN-03|1|1990|1|1|120|2|2|P||ALX|A
17 0001|MD|00000364|98JV120|1001|05|EQUIPAMENTO COM DEFEITO|30-JUN-08|1|98|1|1|120|3|1|2|POS|EM
18 0001|MD|00000421|98JX240|1000|04|EQUIPAMENTO SUCATEADO|01-FEB-08|1|98|1|1|240|3|1|2|ALX|AL.C
19 0001|MD|00000422|CVAD58|1001|04|EQUIPAMENTO SUCATEADO|24-JAN-02|1|1990|1|1|120|2|2|P||ALX|A
20 0001|MD|00000425|98AX240|1000|04|EQUIPAMENTO SUCATEADO|21-DEC-05|1|98|1|1|240|3|1|2|ALX|AL.C
21 0001|MD|00000431|CVAD58|1001|04|EQUIPAMENTO SUCATEADO|24-JAN-02|1|1990|1|1|120|2|2|P||ALX|A
22 0001|MD|00000584|CVAD58|1001|04|EQUIPAMENTO SUCATEADO|24-JAN-02|1|1990|1|1|120|2|2|P||ALX|A
23 0001|MD|00004433|CVAD58|1001|04|EQUIPAMENTO SUCATEADO|24-JAN-02|1|1990|1|1|120|2|2|P||ALX|A

```

Figura 2.7 – Registro com quebra dos dados em duas linhas.

O segundo problema foi a utilização do caractere \ antes do separador dos campos de dados. Esse caractere é considerado como caractere de escape pelo MySQL.

Um caractere de escape é o primeiro caractere de uma sequência de escape, que por sua vez é um conjunto de caracteres que possuem a habilidade de enviar caracteres de controle (e.g., tecla "ESC") [45 e 46]. Sempre que o caractere de escape é utilizado, o caractere subsequente é analisado. Se a sequência de escape formar um caractere de controle, este é enviado. Caso contrário o caractere subsequente ao caractere de escape é considerado caractere de texto. A tabela 2.3 apresenta as sequências de escape do MySQL. A figura 2.8 apresenta o resultado da utilização do caractere de escape em conjunto com o caractere "n" que apresenta o caractere de controle "nova linha".

Tabela 2.3 – Sequências de escape do MySQL.

Sequência	Descrição
\0	Caractere ASCII 0 (nulo)
\'	Um caractere de marcação simples
\"	Um caractere marcação duplas
\b	Caractere <i>backspace</i>
\n	Caractere nova linha
\r	Caractere <i>carriage return</i>
\t	Caractere de tabulação
\z	Caractere ASCII 26 (Ctrl+Z)
\\	Caractere barra invertida
\%	Caractere %
\_	Caractere _

<pre>mysql&gt; select "canto" as "Exemplo"; +-----+   Exemplo   +-----+   canto     +-----+ 1 row in set (0.00 sec)</pre>	<pre>mysql&gt; select "ca\nto" as "Exemplo"; +-----+   Exemplo   +-----+   ca to         +-----+ 1 row in set (0.00 sec)</pre>
(a)	(b)

Figura 2.8 – Utilização de sequência de escape. a) *String* sem uso do caractere de escape. b) Uso do caractere de controle "nova linha".

O sistema de importação do MySQL interpreta o caractere separador de campos, quando utilizado subsequentemente ao caractere de escape, como parte de um campo, *i.e.*, o caractere separador perde a função de caractere especial se comportando como um caractere normal, presente no campo. Como resultado ocorre a mescla de variáveis subjacentes e falta de informação nas últimas variáveis importadas.

Outros dados incorretos existem devido a conversões incorretas do banco de dados utilizado. A existência de datas de fabricação e datas de instalação de medidores de energia com valor acima do ano de criação do depósito de dados (*e.g.*, ano\_fabric = "2069") ilustram o problema: conversão automática de datas de dois dígitos para quatro dígitos. As correções foram realizadas após a importação, através da subtração de cem anos dos valores acima de 2008.

### 2.6.3 Dados discrepantes (*outliers*)

São dados com valores extremos ou atípicos, ou no caso de variáveis categóricas são valores com pouquíssimos casos. Para variáveis numéricas pode-se analisar a variância, média e desvio padrão dos dados para determinar quais fazem parte da amostra e quais são considerados ruído. Descrição e referências de várias técnicas de detecção de dados discrepantes podem ser encontradas em [47].

Um exemplo de dado discrepante é apresentado na tabela 2.4. São apresentados os valores de consumo de diversas unidades consumidoras monofásicas.

Tabela 2.4 – Consumo de unidades consumidoras monofásicas.

UC	Consumo (kWh)
102013	107
102340	88
102543	201
109305	54
110241	15100
110532	187
...	...
<b>Média</b>	<b>108,7</b>
<b>3º Quartil</b>	<b>280</b>

Analisando a tabela observa-se que o consumo da UC "110241" é muito superior ao consumo médio. O 3º quartil também é muito inferior ao valor desta UC, o que caracteriza este dado como um valor atípico no conjunto de dados analisado.

### No banco de dados

O depósito de dados criado apresenta casos de subtipos de medidores com poucos registros sobre a condição de defeito. A principal causa é a baixa quantidade de registros de ensaio desses modelos resultando em redução da quantidade do modelo após aplicação de comandos de agrupamentos condicionais (`join... where`) A tabela 2.5 apresenta a relação entre a quantidade de subtipos no depósito de dados e na tabela UC\_MEDIDOR. A última coluna indica a quantidade de registros no depósito de dados que representam medidores com falhas. Uma observação deve ser feita para os valores de subtipos apresentados. Nesta seção os subtipos são apresentados de forma agrupada.

Tabela 2.5 – Quantidade de registros para cada subtipo no depósito de dados e tabela UC\_MEDIDOR.

Subtipo	Depósito de Dados (%)	Tabela UC_MEDIDOR (%)	Qtd registros defeito (%)
B4C3V	1 (0,01%)	37129 (1,67%)	0 (0,00%)
C1X4	1 (0,01%)	6092 (0,27%)	1 (100,00%)
D58	378 (3,89%)	84567 (3,80%)	136 (35,98%)
D58C	1548 (15,93%)	79772 (3,58%)	208 (13,44%)
D8L	798 (8,21%)	71229 (3,20%)	34 (4,26%)
F72	2 (0,02%)	70104 (3,15%)	0 (0,00%)
F72G	722 (7,43%)	8144 (0,37%)	9 (1,25%)
FV201	429 (4,41%)	80917 (3,63%)	108 (25,17%)

FV202	1 (0,01%)	18993 (0,85%)	0 (0,00%)
FX221	572 (5,89%)	285781 (12,84%)	14 (2,45%)
M1A	570 (5,87%)	82051 (3,69%)	90 (15,79%)
M1AG	1 (0,01%)	57858 (2,60%)	0 (0,00%)
M1AT	1 (0,01%)	272945 (12,26%)	0 (0,00%)
M8C	1 (0,02%)	19472 (0,87%)	1 (100,00%)
M8L	652 (6,71%)	42298 (1,90%)	189 (28,90%)
MF79G	710 (7,31%)	63866 (2,87%)	16 (2,25%)
MFB120G	1 (0,01%)	29661 (1,33%)	0 (0,00%)
MFT120G	291 (2,99%)	10717 (0,48%)	46 (15,81%)
MV202	619 (6,37%)	279456 (12,55%)	74 (11,95%)
MY202	1005 (10,34%)	45665 (2,05%)	37 (3,68%)
PN5DG	427 (4,39%)	353425 (15,87%)	50 (11,71%)
PN5T-G	351 (3,61%)	77496 (3,48%)	0 (0,00%)
SL1621	362 (3,73%)	90795 (4,08%)	10 (2,76%)
T4C3V	1 (0,01%)	2472 (0,11%)	0 (0,00%)
T8L	274 (2,82%)	55564 (2,50%)	39 (14,23%)
<b>TOTAL</b>	<b>9718 (100,00%)</b>	<b>2226469 (100,00%)</b>	<b>1062 (10,93%)</b>

Analisando a tabela 2.5 pode-se fazer algumas observações sobre os dados:

- Quando a quantidade de registros de medidores com falha é pequena (menor que 3%) comparada com a quantidade de registros de medidores sem falha, não existe informação suficiente a ser aprendida desses dados. Valores percentuais em até 2% podem ser considerados casos atípicos ou incomuns, portanto devem ser descartados. Situação idêntica é encontrada quando a quantidade de registros de medidores sem falha é pequena (menor que 3%) comparada com a quantidade registros de medidores com falha;
- Outra observação é feita quando se analisa o percentual de registros no depósito de dados e na tabela UC\_MEDIDOR. Para que o depósito represente adequadamente os dados do banco de dados, o percentual para cada subtipo de medidor em relação ao total de equipamentos deve ser próximo em ambas as tabelas. Um valor percentual menor de registros no depósito de dados leva à crença de falta de representatividade dos dados da tabela MED\_CONS. Um valor percentual maior leva a uma idéia oposta. O problema do excesso de representatividade está na seguinte premissa: se o percentual no depósito de dados é maior para um subtipo, então existe pelo menos um subtipo com percentual menor;

Através da primeira observação, os subtipos "B4C3V", "C1X4", "F72", "F72G", "FV202", "FX221", "M1AG", "M1AT", "M8C", "MF79G", "MFB120G", "PN5T-G", "SL1621" e "T4C3V" são considerados dados discrepantes.

A segunda observação pode ser desconsiderada, pois o depósito de dados não é uma amostra da tabela UC\_MEDIDOR, e sim uma junção desta com outras tabelas (*e.g.*, MED\_CONS e os dados ARSESP).

## 2.6.4 Dados redundantes

A utilização de diversas fontes de dados na geração do depósito de dados pode resultar em dados redundantes. Estes podem ocorrer devido a nomeações distintas para uma mesma variável ou uso de variáveis distintas para representar um mesmo conhecimento. Exemplo: variável referente aos alunos de uma universidade, em um banco utiliza-se a definição estudante e em outro a definição aluno.

Para determinação de variáveis redundantes faz-se a análise da correlação entre as variáveis. Campos com nomeação distinta para uma mesma variável possuem correlação unitária enquanto campos distintos que representam o mesmo conhecimento possuem módulo dos valores de correlação próximos a 1. O especialista deve decidir qual o limite de correlação e semelhança entre os dados para classificá-los como redundantes.

Outro tipo de redundância é a repetição de um registro na base de dados, problema típico de bancos de dados com tabelas não-transacionais acessadas por 2 ou mais usuários simultaneamente. Ressalta-se que dois registros idênticos em um subconjunto de uma base de dados não é caracterizado diretamente como um erro de redundância, para tal é necessário que os registros estejam repetidos no banco de dados original e que este banco seja caracterizado por conter informações únicas. Desta forma, a análise pelo especialista é fundamental.

Apesar de existirem algoritmos de mineração de dados que não são influenciados por variáveis redundantes deve-se considerar a remoção destas variáveis quando a quantidade de informação a ser verificada for elevada, podendo ultrapassar a capacidade de memória utilizada pelos programas de mineração de dados [44].

## No banco de dados

As redundâncias existentes no banco de dados estão nas variáveis utilizadas para relacionar as diversas tabelas. As variáveis `uc` e `conta`, respectivamente de `UC_MEDIDOR` e `MED_CONS`, contem a mesma informação (número da unidade consumidora).

Existem campos que apresentam aglutinações de informações de dois ou mais campos, ou aglutinação de informações de um campo com novas informações que não são apresentadas por nenhum campo na base de dados. É o caso da variável `contabil` que informa a quantidade de fases a que o medidor deve ser instalado, a tensão de ligação do equipamento à instalação elétrica, se é utilizado para medição direta ou indireta, e se é do tipo eletrônico ou indução eletromagnética. A quantidade de fases para conexão é fornecida simultaneamente por `contabil` e por `cod_tipo_fase` enquanto as demais informações não são apresentadas por nenhum campo da tabela.

## 2.7 Transformação

A etapa de transformação consiste em fazer a adequação dos dados para a etapa de mineração. Nesta etapa pode ocorrer a conversão de variáveis contínuas para discretas (discretização) e conversão de variáveis texto para numéricas (*e.g.*, "A" → "1"). As variáveis discretas não podem ser convertidas em contínuas, uma vez que a quantidade de informação que uma variável contínua contém é superior a de uma variável discreta.

Outro tipo de transformação comum é a transformação de uma variável com N valores possíveis em N variáveis binárias. Essa transformação é denominada *flattening* e é utilizado para melhorar o relacionamento dos dados [36].

Exemplo: um depósito de dados contendo o nome do cliente e o tipo de produto que ele comprou de uma empresa de telecomunicações: para um mesmo cliente podem existir vários registros, cada um contendo um serviço contratado. A tabela 2.6 apresenta um segmento do depósito de dados.

**Tabela 2.6 – Segmento de um depósito de dados.**

<b>Cliente</b>	<b>Produto</b>
...	...
20375	Telefone
20375	Internet
20375	TV por assinatura
20398	Telefone
20398	Identificador de chamadas
...	...

Os dados da variável `Produto` podem ser transformados utilizando *flattening* para que haja a redução de registros e melhor visualização dos dados. A tabela 2.7 mostra o resultado da aplicação de *flattening*. Registros com valor "1" informam que o cliente contratou aquele serviço enquanto valor "0" significa que o serviço não foi contratado.

**Tabela 2.7 – Segmento de um depósito de dados após *flattening*.**

<b>Cliente</b>	<b>Telefone</b>	<b>Internet</b>	<b>TV por assinatura</b>	<b>Identificador de chamadas</b>
20375	1	1	1	0
20398	1	0	0	1

A normalização de variáveis é uma técnica que delimita os valores de variáveis numéricas em um mesmo intervalo de busca. Geralmente os valores são normalizados no espaço 0 a 1, ou -1 a 1. Esta técnica de transformação é útil quando se utiliza de medidas de distâncias entre os dados (*e.g.*, agrupamento). Sem a normalização, uma variável pode apresentar maior efeito de distância sobre as demais, tornando impossível comparar a dispersão dos dados devido à variável não normalizada.

Os tipos de transformações a serem realizadas nos dados dependerão da tarefa de MD e do tipo de algoritmo utilizado para este fim.

### 2.7.1 Transformação de variáveis categóricas em numéricas

A técnica de mineração de dados *stepwise* não pode trabalhar com variáveis que não sejam numéricas. A idéia inicial é converter variáveis tipo texto em numéricas, porém as variáveis, ao serem convertidas, apresentarão uma informação de distância falsa. Variáveis categóricas não possuem informação de distância, e a conversão em valores numéricos

apresenta uma informação falsa sobre a distância. Exemplo: considere uma variável que aceite o nome de frutas, "banana", "maçã" e "laranja". A conversão desta variável em numérica recebe os valores "0", "1" e "2" que representam "banana", "maçã" e "laranja", respectivamente. Em questões numéricas pode-se afirmar que  $2 > 1$ , porém não se pode dizer que "laranja" > "maçã".

De acordo com a literatura todas variáveis categóricas devem ser convertidas em numéricas através de *flattening*. A tabela 2.8 apresenta a listagem de variáveis categóricas do banco de dados e a quantidade de valores que elas assumem.

**Tabela 2.8 – Lista de variáveis categóricas utilizadas para MD.**

Variável	Qtd de valores possíveis
subtipo_equip	11
cod_tipo_fase	3
contabil	3
cod_tipo_local	2
cod_localidade	210
cod_classe_principal	8
cod_classe_consumo	160
grupo_tensao_fornecimento	2

As variáveis `cod_localidade` e `cod_classe_consumo` não passaram pelo processo de transformação. A enorme quantidade de valores que elas podem assumir dificulta a análise após a transformação, gerando muitas colunas. Por se tratar de variáveis categóricas com valores numéricos elas são aceitas pelo algoritmo de regressão. A única implicação é que a regressão tratará os valores destas variáveis como numérico, ou seja, a regressão assume que `cod_classe_consumo=3132` é menor do que `cod_classe_consumo=6019` ou que `cod_localidade=93` é maior que `cod_localidade=71`.

## 2.7.2 Transformação de variáveis contínuas em discretas (discretização)

Uma variável contínua pode ser vista como um conjunto de valores linearmente ordenados, limitados em uma faixa de valores. O processo de discretização consiste

basicamente em subdividir essa faixa em pelo menos duas sub-faixas. A escolha dos pontos de discretização deve ser realizada de modo a prover informações de classificação úteis para os exemplos que compõem as classes [48].

Os primeiros algoritmos de indução para árvores de decisão não trabalhavam com variáveis contínuas, porém existem vários algoritmos de indução que trabalham com variáveis contínuas, discretizando-as em tempo de execução. Estes algoritmos utilizam a discretização em dois subgrupos, utilizando um valor limite. Quando o valor de um elemento da variável for menor do que o limite ele vai para o ramo esquerdo, e quando o valor é maior do que o limite ele vai para o ramo direito.

A discretização em tempo de execução gera problemas para validação cruzada. O algoritmo de indução busca o ponto ótimo de discretização de acordo com os dados utilizados para treino. Como na validação cruzada são realizados vários treinos, com segmentos distintos de dados, cada segmento pode apresentar pontos de discretização distintos.

A alternativa é realizar a discretização antes da execução do algoritmo, tratando todas as variáveis contínuas como discretas. Existe uma diversidade de técnicas de discretização, em [49] faz-se uma revisão sobre várias técnicas.

No banco de dados utilizou-se uma metodologia simples de discretização. Os pontos foram escolhidos de modo a garantir uma quantidade de dados próxima em todas as sub-faixas. As variáveis `ano_fabric`, `media`, `maximo`, `dias_de_vida`, `temperatura`, `precipitacao` e `altitude` foram discretizadas para aplicação da técnica de mineração árvores de decisão.

## **2.8 Mineração de Dados**

A Mineração de Dados é utilizada para obter padrões de interesse, a partir de dados, em uma ou mais formas representativas (*e.g.*, árvores de decisão, regressão). É considerada a etapa central de DCBD.

A utilização da Mineração de Dados em bases de dados sem tratamento prévio foi nomeada de modo pejorativo como *data fishing* ou *data dredging*, uma vez que os padrões obtidos podem não apresentar significado verdadeiro.

Em [35] são classificados dois tipos de objetivos almejados com a utilização de DCBD, verificação e descoberta.

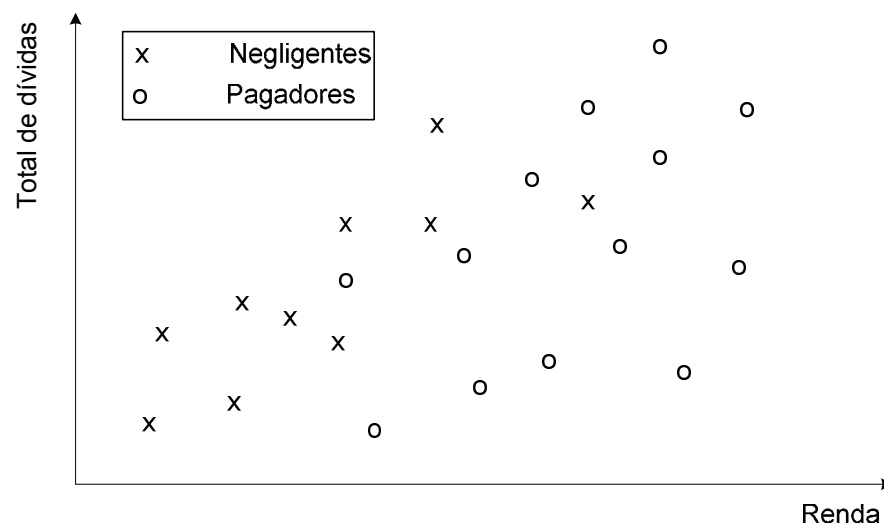
- Verificação: o processo de DCBD é utilizado para confirmar uma hipótese inicial;
- Descoberta: a ferramenta é utilizada para descobrir novos padrões nos dados.

Os dois objetivos principais de MD são predição e descrição.

- Predição: descobre padrões e relacionamentos para determinar valores desconhecidos ou futuros de uma ou mais variáveis de interesse.
- Descrição: descobre padrões que descrevem os dados em uma forma compreensível para humanos.

Nota-se que, em alguns casos, o resultado da MD satisfaz ambos os objetivos.

Os objetivos descritos anteriormente são alcançados utilizando algoritmos para obter uma tarefa de MD. As tarefas descritas a seguir são baseadas em [34] e [38] e utilizam o exemplo apresentado na figura 2.9. É apresentada de forma gráfica, os dados de um banco com informações sobre clientes que pagaram um empréstimo (o) ou não (x). O eixo da abscissa contém a renda da pessoa, e no eixo da ordenada é apresentado o total de dívidas da mesma.

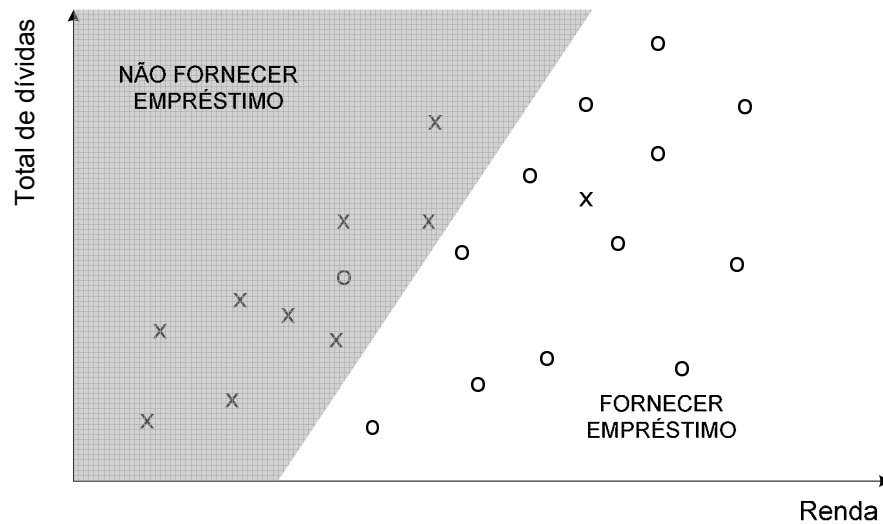


**Figura 2.9 – Descrição de clientes que pagam ou não os empréstimos em função da renda e da quantidade de dívidas.**

### 2.8.1 Classificação

A classificação consiste em aprender uma função que mapeia (classifica) os itens de um banco de dados em uma das várias classes predefinidas. Exemplos de aplicação: sistema

que classifica aeronave através dos valores de suas dimensões e tipo de propulsão, classificação de clientes fraudadores utilizando dados históricos de consumo. A figura 2.10 mostra um mapeamento dos dados de empréstimo em duas regiões, o banco pode querer utilizar classificação para automaticamente decidir quando um empréstimo futuro deve ser fornecido ou negado.



**Figura 2.10 – Exemplo de classificação.**

## 2.8.2 Regressão

A regressão consiste em aprender uma função que mapeia os dados de uma variável do banco de dados em uma função das outras variáveis do banco. A regressão pode ser linear ou não-linear, simples ou multivariada. Exemplos de aplicação de regressão: determinação de biomassa presente em uma floresta utilizando dados coletados através de sensores; prever a demanda por um produto em função de gastos com propaganda. A figura 2.11 apresenta uma regressão simples onde o total de dívidas é uma função linear de renda.

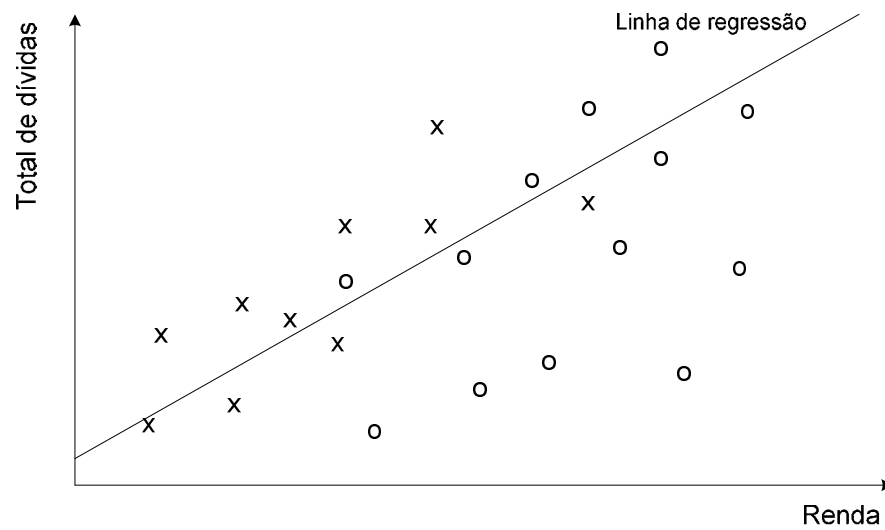


Figura 2.11 – Exemplo de regressão.

### 2.8.3 Agrupamento (*Clustering*)

Agrupamento é uma tarefa descritiva, que consiste em agrupar os dados em um número finito de grupos. Os grupos podem ser mutuamente exclusivos, ou possuírem sobreposição. Exemplos de utilização: definir grupos de consumidores em uma base de dados de mercado. A figura 2.12 apresenta o agrupamento para o exemplo de empréstimo. Observe-se que na análise de agrupamento não existe distinção entre os valores de classe, ou seja, não se sabe durante a aplicação de agrupamento os valores dos grupos.

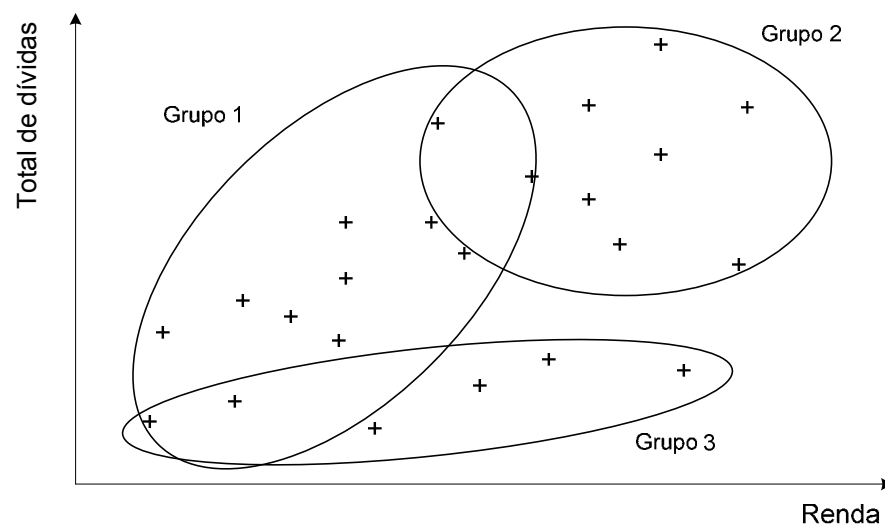


Figura 2.12 – Exemplo de agrupamento.

## 2.8.4 Associação

A tarefa de associação visa encontrar relacionamentos entre variáveis do banco de dados. Tarefas de associação possuem duas etapas: redução das variáveis e geração de regras através da análise de suporte e confiança. Suporte é definido como a quantidade de vezes que uma relação entre duas ou mais variáveis do banco de dados é verdadeira, isto é, que para a mesma relação seja apresentado o mesmo valor da variável de saída [50]. Exemplos de aplicação: análise de relacionamentos de venda de produtos em supermercados (*e.g.* pão e leite).

## 2.8.5 Sumarização (*Summarization*)

A sumarização mapeia os dados em subclasses com descrições simples. Exemplos de utilização: geração de resumos de textos.

## 2.8.6 Detecção de desvios (*Deviation detection*)

A detecção de desvios observa alterações nos dados. Exemplos de aplicação: detecção de fraudes.

## 2.9 *Técnicas de mineração de dados*

As técnicas utilizadas para MD provém das diversas áreas que se relacionam com DCBD. Exemplos de técnicas são:

- Estatística: redes bayesianas, regressão, análise multivariada;
- Aprendizagem de máquina: árvores de decisão, redes neurais, k-means.

As técnicas de mineração de dados não são específicas para uma tarefa, podendo uma técnica ser utilizada em duas ou mais tarefas. A tabela 2.9 apresenta exemplos de técnicas e as tarefas de MD que elas podem realizar.

Tabela 2.9 – Exemplo de técnicas de mineração de dados e tarefas em que podem ser utilizadas.

	Classificação	Regressão	Agrupamento	Associação	Sumarização
Regressão	•	•			
Árvores de Decisão	•	•		•	
Redes neurais	•	•	•		
Estatística Clássica				•	•

## 2.9.1 Regressão

A ideia da regressão é modelar o comportamento de uma variável em função de outras variáveis. Assim, é possível prever o comportamento de uma variável de saída em função de uma ou mais variáveis de entrada.

### Considerações iniciais

As regressões utilizam um conjunto de variáveis de entrada para gerar uma variável de saída. Para facilitar as explicações, será utilizada a mesma convenção apresentada em [26]. No trabalho, é definido um modelo para representação da regressão. O modelo é composto por  $N$  entradas e  $M$  saídas, como mostra a figura 2.13.

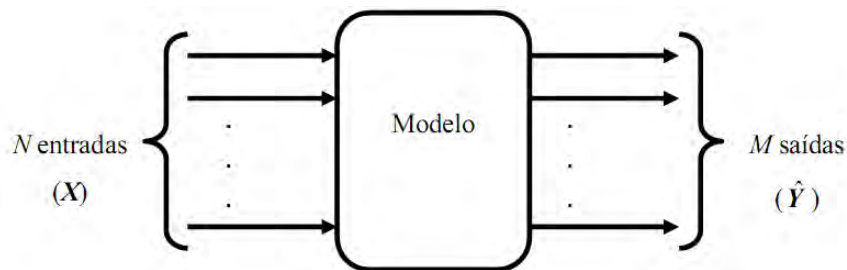


Figura 2.13 – Convenção dos modelos. Fonte: [26]

Defini-se  $x_i$  como a  $i^{\text{ésima}}$  das  $N$  entradas,  $y_j$  a  $j^{\text{ésima}}$ , e  $\hat{y}_j$  a estimaco da  $j^{\text{ésima}}$  sada. Nota-se que  $x_i$ ,  $y_j$  e  $\hat{y}_j$  so vetores, cada um com  $n$  observaes, como mostra a figura 2.14.  possvel definir matrizes de entrada e sada para o modelo. A matriz  $\mathbf{X}$   a matriz de entrada e

as matrizes  $\mathbf{Y}$  e  $\hat{\mathbf{Y}}$  são as matrizes de saída verdadeira (real) e saída do modelo (estimada). A matriz  $\mathbf{X}$  possui dimensão  $n \times N$ , e as matrizes  $\mathbf{Y}$  e  $\hat{\mathbf{Y}}$  possuem dimensões  $n \times M$ .

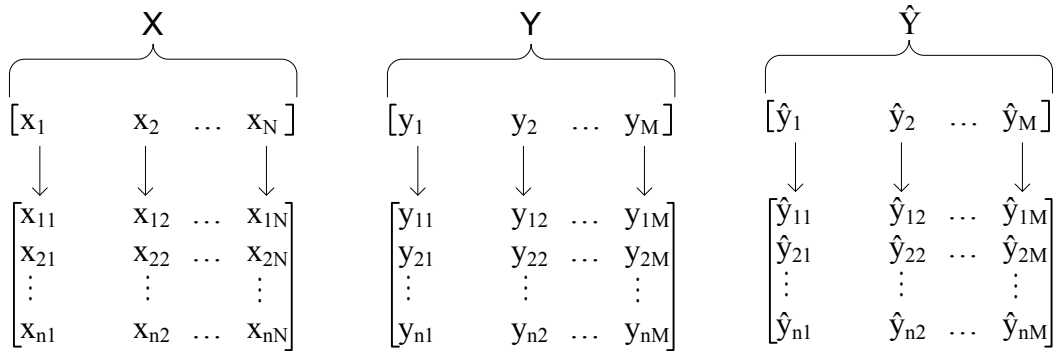


Figura 2.14 – Convenção de matrizes. Fonte: [26]

Como existe somente uma saída de interesse neste trabalho,  $M$  apresenta valor unitário.

## Regressão linear

Na regressão linear, a saída é considerada uma função linear da entrada.

$$\hat{Y} = X \times \beta + \beta_0 \quad (2.1)$$

Onde  $\beta_0$  é um vetor constante. Deseja-se obter a matriz  $\beta$  que soluciona esta equação, assim será possível criar um modelo capaz de prever a saída  $\hat{\mathbf{Y}}$  para qualquer valor de  $\mathbf{X}$ .

A matriz  $\beta$  é obtida através da equação 2.2.

$$\beta = X^{-1} \times (\hat{Y} - \beta_0) \quad (2.2)$$

O fato da matriz  $\mathbf{X}$  não ser quadrática, ou ser quadrática, mas não invertível, impede a utilização da equação 2.2. Como alternativa utiliza-se o conceito de pseudo-inversa [51]. A pseudo-inversa também é conhecida como inversa de Moore-Penrose, em homenagem a E. H. Moore e R. Penrose que formularam de modo independente o conceito [52]. Quando a matriz é invertível, a pseudo-inversa é idêntica à inversa. No caso de matriz não-invertíveis, a

pseudo-inversa apresenta uma matriz com algumas propriedades da inversa, possuindo funcionalidade para a regressão. O uso da pseudo-inversa para obtenção da matriz  $\beta$  é apresentado na equação 2.3. Uma observação importante é que a pseudo-inversa não existe quando a quantidade de variáveis de entrada for maior do que a quantidade de observações. Na condição de  $N > n$ , o processo de pseudo-inversa realiza uma expansão de dimensões para formar a matriz quadrada, o que resulta em uma matriz quadrada singular ou próxima a singularidade, e esta não é invertível. Na condição de  $n > N$ , o processo realiza uma compressão de dimensões, o que geralmente resulta em uma matriz quadrada invertível.

$$\beta = (X^T \times X)^{-1} \times X^T \times (\hat{Y} - \beta_0) \quad (2.3)$$

Os valores da matriz  $\beta$  indicam os pesos das variáveis de entrada para obtenção das variáveis de saída.

Para obtenção de  $\beta$ , várias técnicas podem ser utilizadas. A mais comum é a de mínimos quadrados. Esta técnica utiliza uma medida de erro entre a saída real e a estimada pela regressão, dada por  $\varepsilon$ , para encontrar o valor de  $\beta$ . Este erro é denominado resíduo. A equação 2.4 apresenta a equação do resíduo.

$$Y - \hat{Y} = \varepsilon \quad (2.4)$$

Onde  $\varepsilon$  é um vetor com os valores de erro entre a saída real e estimada

A regressão linear pode ser escrita em função da saída real através da substituição da equação 2.1 em 2.4.

$$Y - (X \times \beta + \beta_0) = \varepsilon \quad (2.5)$$

$$Y = X \times \beta + \beta_0 + \varepsilon \quad (2.6)$$

Nessas condições, o valor de  $\beta$  é dado por:

$$\beta = (X^T \times X)^{-1} \times X^T \times (Y - \beta_0 - \varepsilon) \quad (2.7)$$

## Regressão *Stepwise*

Na mineração de dados podemos ter como objetivo a determinação de um subconjunto de variáveis que melhor represente a saída de interesse. Existem vários algoritmos estatísticos propostos para apresentar subconjuntos de qualidade de forma automática. Alguns algoritmos populares são [53 e 54]: *All Possible Regression* (APR), *Forward Selection* (FS), *Backward Elimination* (BE) e *Stepwise* (ou seleção de Efroymsen).

O algoritmo *stepwise* é uma evolução do método FS. O algoritmo FS inicializa com nenhuma variável, a saída é uma constante. É acrescentada à lista a variável que obtiver o maior valor em um teste de significância da análise de variância (teste F), desde que o resultado do teste seja superior a um limite pré-definido. A segunda variável adicionada também é aquela que satisfaz o mesmo critério de escolha. O processo se repete até que todas as variáveis sejam adicionadas ou um critério de parada for atingido (*e.g.*, nenhuma variável obteve resultado do teste superior ao limite pré-definido).

A diferença de *stepwise* para FS é que a cada passo, pode-se adicionar ou remover uma variável. Com isto uma variável anteriormente incluída no modelo pode ser retirada. O critério utilizado para análise do subconjunto irá depender do critério de adequação dos dados e do objetivo da regressão. Os programas computacionais para cálculo de *stepwise* geralmente utilizam o teste-F para verificar se a diferença de variância da saída, causada pela inserção (ou remoção) de uma variável, possui um valor significativo. Outras técnicas apresentadas em [54] são: coeficiente de correlação múltipla quadrática, erro quadrático total, CP de Mallow, critério de informação de Akaike.

Entre os parâmetros apresentados em cada etapa da regressão o *root mean squared error* (RMSE) apresenta uma informação importante para o usuário. Ele define a variação dos valores estimados em relação aos valores observados no conjunto de dados, isto é, determina a média dos valores quadráticos de resíduo e extrai a raiz, conforme equação 2.8. O resíduo é considerado a diferença entre o ponto estimado e seu valor real, quanto maior a variação do resíduo, menor a adequação do modelo aos dados de treinamento. Um modelo perfeitamente ajustado aos dados de treinamento possui resíduos nulos e portanto o valor do RMSE é igual a 0. Deve-se ressaltar que um modelo perfeitamente ajustado aos dados de treino está sujeito a aprender os ruídos existentes nos dados, o que não é desejável.

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (\hat{Y}_i - Y_i)^2} \quad (2.8)$$

A equação do RMSE se assemelha à equação do desvio padrão, com a diferença que no desvio padrão se utiliza a diferença entre os valores dos dados e o valor médio. Em casos em que o valor médio dos resíduos é nulo, o RMSE apresenta o mesmo valor que o desvio padrão dos resíduos.

Como o algoritmo *stepwise* permite a adição ou remoção de variáveis, uma a cada etapa, pode-se dizer que é um algoritmo de busca em vizinhança [33 e 55]. Definindo-se  $Q$  como a quantidade de variáveis existentes no modelo e  $N$  como o total de variáveis de entrada, um vizinho é definido através de:

- Se  $1 < Q < N$ 
  - Inserção de uma variável ao modelo;
  - Remoção de uma variável do modelo;
- Se  $Q = 1$ 
  - Inserção de uma variável ao modelo.

O critério de parada ocorre quando não existe vizinho com erro dos quadrados dos resíduos menor do que a configuração atual ou então quando o algoritmo realizou uma quantidade predefinida de iterações [54]. A figura 2.15 apresenta o fluxograma do algoritmo *stepwise*.

Uma ou mais variáveis podem ser forçadas ao modelo, ou seja, variáveis que não devem ser consideradas para remoção. O modelo inicial contém o grupo de variáveis forçadas e estas não serão eliminadas em nenhuma iteração do algoritmo.

A regressão *stepwise* trabalha somente com variáveis numéricas, não podendo ser utilizada para variáveis categóricas. Como consequência os valores categóricos devem ser convertidos em variáveis numéricas via *flattening*.

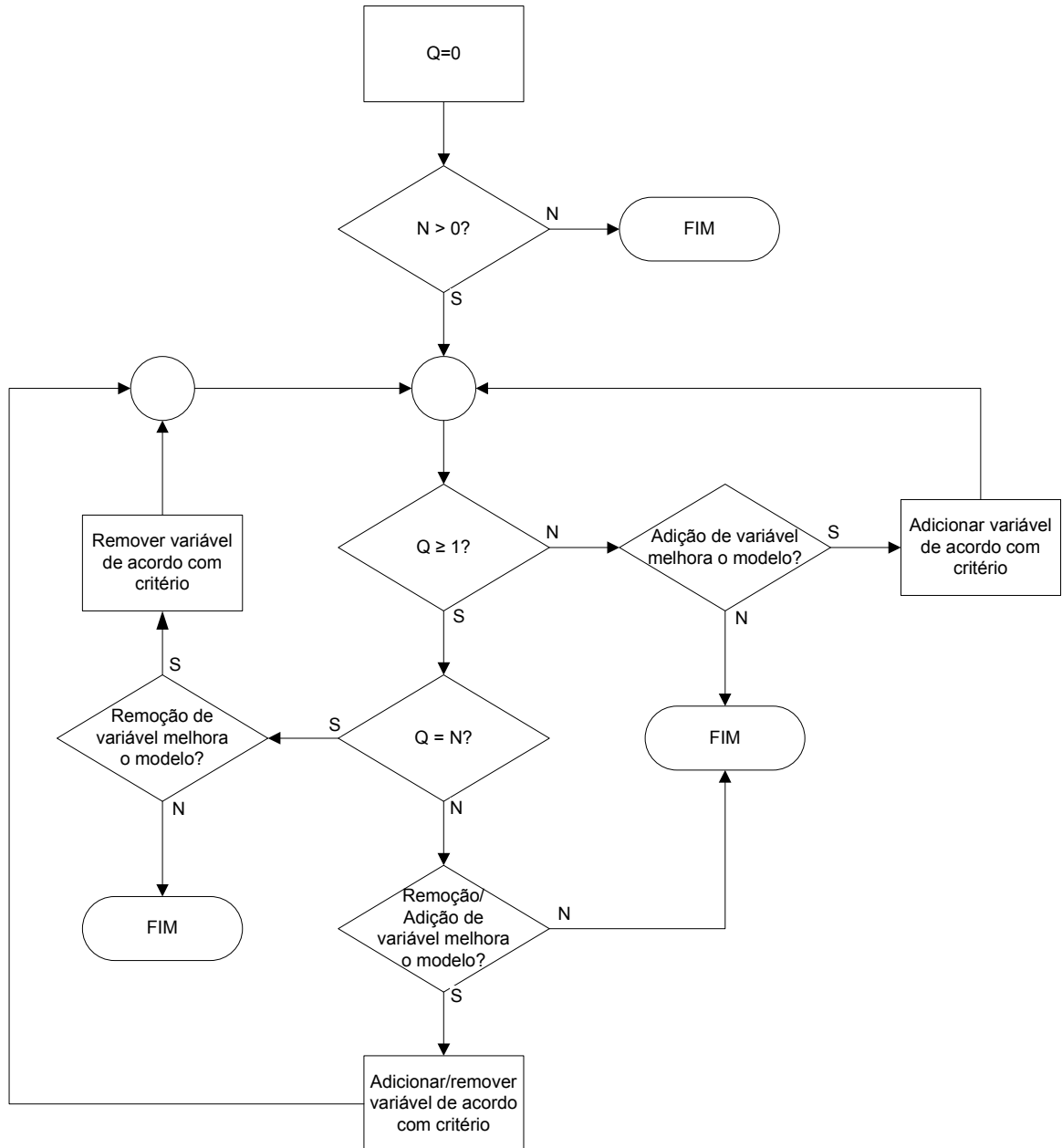


Figura 2.15 – Fluxograma simplificado de *forward stepwise*.

A regressão *stepwise* escolhe as variáveis que mais se adaptam ao modelo gerando menor erro de generalização da saída por uma função linear.

Mesmo que os dados não estejam correlacionados linearmente, o algoritmo pode apresentar fortes indícios das variáveis mais significativas na determinação da variável de saída.

Para aplicação nos dados da concessionária de energia, as variáveis categóricas foram transformadas via *flattening*.

O algoritmo de regressão *stepwise* utilizado é fornecido pelo programa computacional MATLAB. Os valores de  $p$  para inserção de variáveis foi  $p < 0,05$  e para retirada  $p > 0,1$ . O

algoritmo foi executado duas vezes, uma iniciando sem nenhuma variável no modelo (*forward stepwise*) e outra com todas as variáveis inclusas no modelo (*backward stepwise*).

### Forward stepwise

A figura 2.16 apresenta o histórico do RMSE durante a execução do algoritmo de *forward stepwise* [56]. Como a função de saída assume somente dois valores, "0" ou "1", e supondo que o valor de resposta não seja superior ao valor de saída, o maior valor de RMSE que se poderá obter será "1".

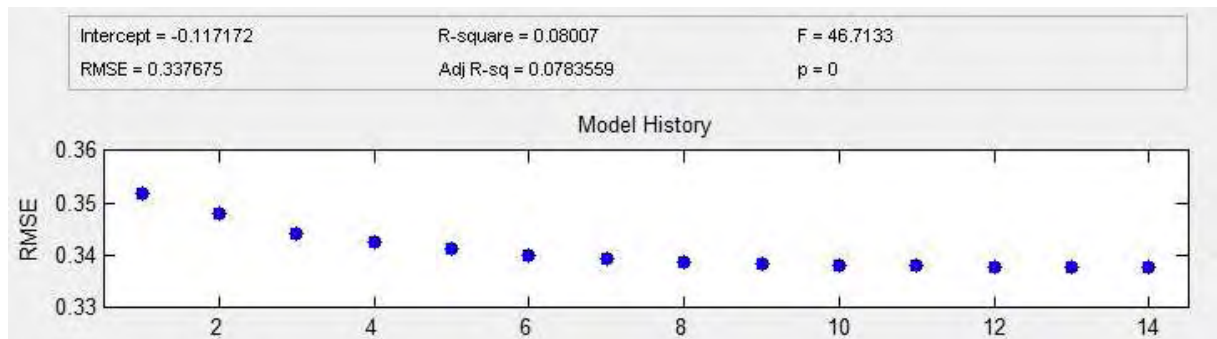


Figura 2.16 – Valores de RMSE para todas as etapas do algoritmo *forward stepwise*.

O RMSE inicial, para uma linha com valores de saída constantes traçado em defeito=0,144615, é igual a 0,35737 e com as variáveis selecionadas pelo algoritmo seu valor se tornou 0,33767. O algoritmo apresentou uma redução de 5% do RMSE, indicando que neste caso um valor constante está representando a saída de maneira similar ao uso das 13 variáveis selecionadas. As variáveis que compõem o modelo de resposta são apresentadas na tabela 2.10.

Tabela 2.10 – Variáveis selecionadas pelo algoritmo *forward stepwise*.

Variáveis
subtipo_equip="D58"
subtipo_equip="D58C"
subtipo_equip="D8L"
subtipo_equip="FV201"
subtipo_equip="M8L"
subtipo_equip="MY202"
contabil=4

instalacoes
dias_de_vida
cod_tipo_local
cod_localidade
cod_classe_principal=2
cod_classe_principal=4

## Backward stepwise

O histórico do RMSE durante a execução do algoritmo *backward stepwise* [56] é apresentado na figura 2.17.

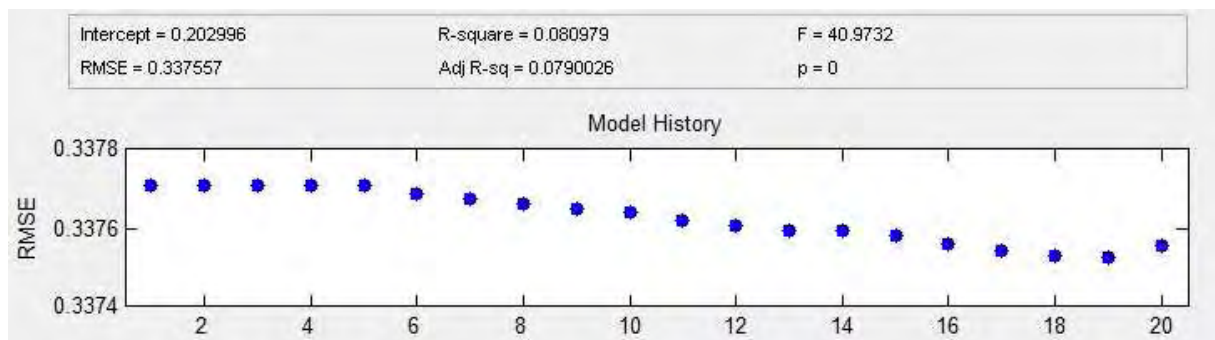


Figura 2.17 – Valores de RMSE para todas as etapas do algoritmo *backward stepwise*.

A utilização de todas variáveis (condição inicial) resulta em um RMSE igual a 0,337708. O RMSE obtido ao final do modelo foi 0,337557. A redução do RMSE foi menor do que 1% para este caso,

As variáveis selecionadas pelo modelo são apresentadas na tabela 2.11.

Tabela 2.11 – Variáveis selecionadas pelo algoritmo *backward stepwise*.

Variáveis
subtipo_equip="D58"
subtipo_equip="D8L"
subtipo_equip="FV201"
subtipo_equip="M1A"
subtipo_equip="M8L"
subtipo_equip="MV202"
subtipo_equip="MY202"
subtipo_equip="PN5DG"
contabil=6

instalacoes
dias_de_vida
cod_tipo_local
cod_localidade
cod_classe_principal=1
cod_classe_principal=3

### Variáveis finais selecionadas pela regressão *stepwise*

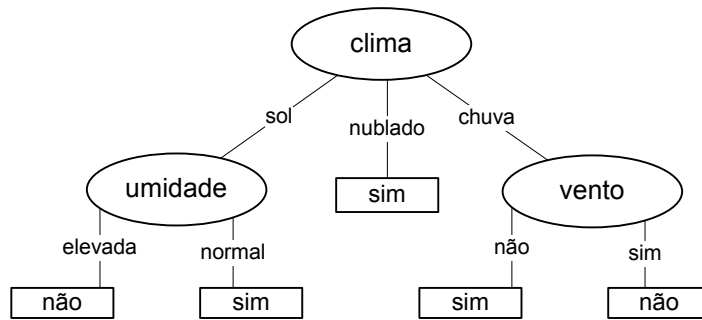
As tabelas 2.10 e 2.11 apresentam dados semelhantes. Faz-se a consideração de que se pelo menos um dos valores, de uma variável que recebeu *flattening*, for escolhido então a variável antes de flattening será escolhida. Assim, as variáveis selecionadas pela regressão são: `subtipo equip`, `contabil`, `instalacoes`, `dias_de_vida`, `cod_tipo_local`, `cod_localidade` e `cod_classe_principal`.

### 2.9.2 Árvores de decisão

A árvore de decisão é uma técnica de MD utilizada para classificação ou regressão. Ela possui este nome por se assemelhar a uma árvore vista de ponta-cabeça. A figura 2.18 apresenta uma árvore de decisão.

A árvore de decisão é constituída de duas partes:

- **Nó:** representa um teste de atributo. O nó que não possui ramo de entrada é denominado nó central e se encontra no início da árvore. Os nós que possuem ramos de entrada e saída são denominados nós internos. Qualquer outro tipo de nó é denominado de “folha”. Os nós “folhas” não representam testes, somente valores da variável de saída ou a própria variável.
- **Ramo:** dois nós são ligados por um ramo. Cada ramo representa um valor distinto existente na variável do nó superior para o subconjunto em estudo.



**Figura 2.18 – Exemplo de árvore de decisão.**

A utilidade das árvores de decisão para MD é perceptível quando são envolvidos algoritmos de indução. Algoritmos de indução são descritos como algoritmos que, a partir de um conjunto de dados de treinamento, generalizam de forma automática o relacionamento das variáveis de entrada com a variável de saída [57]. O objetivo básico é encontrar a árvore que proporcione o menor erro de generalização, isto é, que classifique corretamente a maior quantidade possível de dados. Entre os objetivos secundários destaca-se o desejo de obter a menor árvore possível. O tamanho da árvore influencia na complexidade em se entender o problema.

Determinar a árvore que apresente o menor erro de classificação é considerado um problema NP-difícil, portanto os algoritmos de indução devem ser heurísticas ou metaheurísticas que procuram pela melhor solução do problema. Esses algoritmos podem ser classificados em dois grupos de construção da árvore: cima-baixo e baixo-cima. Nos algoritmos “de cima para baixo” a árvore é montada a partir do nó central em direção as folhas. Esta metodologia é a mais utilizada na literatura.

Existem vários algoritmos de indução para árvores de decisão “de cima para baixo”: ID3, C4.5, CART [33]. Em geral, estes são heurísticas de partição que utilizam do algoritmo guloso para escolha do melhor nó segundo algum critério. Uma vez que um nó é escolhido, o problema é particionado criando subconjuntos para cada ramo do nó. Para cada subconjunto é escolhido o melhor nó e assim sucessivamente até que um critério de parada tenha sido satisfeito.

Os critérios de parada comuns para árvores de decisão são [33 e 58]:

- Todas as instâncias dos dados de treinamento contêm o mesmo valor de saída;
- A profundidade da árvore atingiu um valor máximo predefinido;
- O número de casos em uma folha é menor do que um valor mínimo;
- Se um nó for dividido e a quantidade de ramos for menor do que um valor mínimo;

- O critério de escolha do nó não satisfaz uma premissa (e.g. ser 1,5 vezes mais adequado do que os demais nós).

### ***Inductive Decision Tree (ID3)***

O algoritmo básico de aprendizagem de árvores de decisão é exemplificado pelo algoritmo ID3. Como dito anteriormente, o ID3 é um algoritmo de construção “de cima para baixo”, que assume como dados somente variáveis discretas. O uso de variáveis numéricas é possível desde que sejam discretizadas. Algoritmos mais recentes, como C4.5 e CART encontram automaticamente os pontos de discretização antes de iniciar o algoritmo. O algoritmo ID3 foi formulado por Ross Quinlan em 1986. [58 e 59]

Para a escolha do melhor atributo, ID3 utiliza o conceito de ganho de informação. O ganho de informação define quão bem um atributo classifica os dados da saída.

O ganho de informação é medido utilizando o conceito de entropia [36 e 45]. O conceito de entropia foi formulado por Claude Shannon em 1940 e é considerado um dos mais importantes trabalhos do século 20 [34].

A entropia mede a quantidade mínima média de bits por símbolo necessária para processar uma informação [60]. Dado um conjunto de dados  $S$  com valores discretos, a quantidade mínima média de bits necessários por símbolo para representar os valores de  $S$  é:

$$Entropia(S) = -\sum p_i \cdot \log_2 p_i \quad (2.9)$$

Onde  $p_i$  é a probabilidade do valor  $i$  no conjunto  $S$ .

Considere o exemplo da tabela 2.12. Ela contém informações sobre o tempo em uma manhã de sábado. A variável de saída indica se uma pessoa jogará tênis.

**Tabela 2.12 – Dados sobre tempo para se jogar tênis.**

<b>clima</b>	<b>temperatura</b>	<b>umidade</b>	<b>vento</b>	<b>jogar?</b>
sol	Quente	elevada	não	não
sol	Quente	elevada	sim	não
nublado	Quente	elevada	não	sim
chuva	ameno	elevada	não	sim
chuva	frio	normal	não	sim
chuva	frio	normal	sim	não
nublado	frio	normal	sim	sim

sol	ameno	elevada	não	não
sol	frio	normal	não	sim
chuva	ameno	normal	não	sim
sol	ameno	normal	sim	sim
nublado	ameno	elevada	sim	sim
nublado	frio	normal	não	sim
chuva	ameno	elevada	sim	não

A entropia da variável `clima` é dada por:

$$Entropia(clima) = -p_{sol} \cdot \log_2 p_{sol} - p_{nublado} \cdot \log_2 p_{nublado} - p_{chuva} \cdot \log_2 p_{chuva} \quad (2.10)$$

$$Entropia(clima) = -2 \cdot \left(\frac{5}{14}\right) \cdot \log_2 \left(\frac{5}{14}\right) - \left(\frac{4}{14}\right) \cdot \log_2 \left(\frac{4}{14}\right) = 1,577bits \quad (2.11)$$

Para as demais variáveis as entropias são:

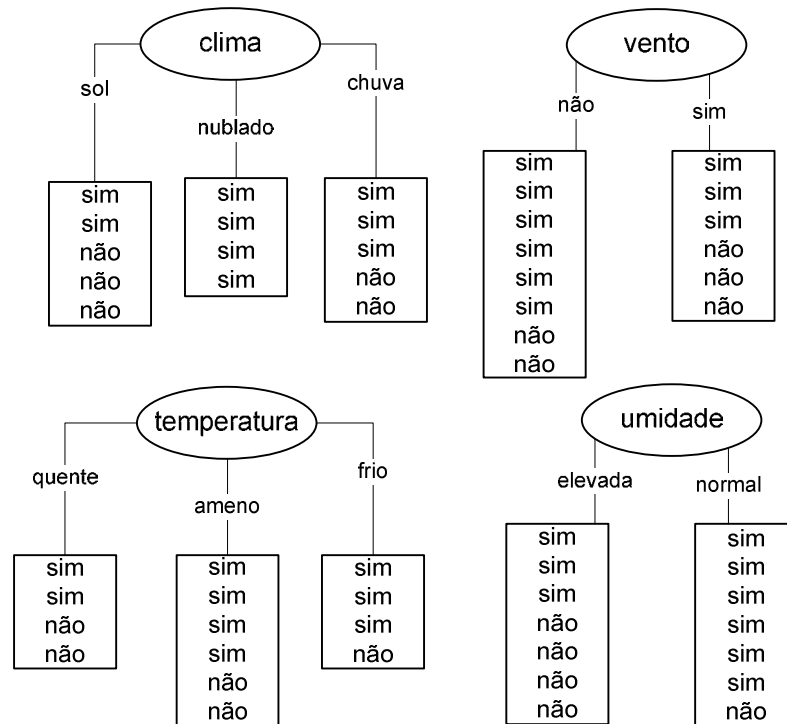
$$Entropia(temperatura) = 1,531bits \quad (2.12)$$

$$Entropia(umidade) = 1bit \quad (2.13)$$

$$Entropia(vento) = 0,985bit \quad (2.14)$$

$$Entropia(jogar ?) = 0,94bit \quad (2.15)$$

O ganho de informação é a diferença da quantidade de bits necessária sem a classificação e a quantidade de bits necessária após a classificação. Em outras palavras, representa a redução de bits quando se supõe que uma informação de entrada é conhecida. Para melhor compreensão será utilizado o exemplo anterior. A figura 2.19 apresenta os valores da resposta classificados para cada valor das variáveis.



**Figura 2.19 – Representação dos valores da variável de saída classificados de acordo com as variáveis de entrada.**

Quando a entropia de uma variável é encontrada utilizando outra variável, diz-se que a entropia encontrada é a entropia condicional, denotada por entropia( $S|T$ ), onde  $S$  é a variável de saída e  $T$  é a variável condicional.

Quando a entropia é encontrada para um valor específico da variável condicional, denomina-se entropia condicional específica. A entropia condicional é escrita em função das entropias condicionais:

$$Entropia(S | T) = \sum p(T = i) \cdot Entropia(S | T = i) \quad (2.16)$$

A entropia condicional específica para a variável de saída utilizando o valor `nublado` da variável `clima` é:

$$Entropia(jogar? | clima = nublado) = -\frac{4}{4} \cdot \log_2 \frac{4}{4} - \frac{0}{4} \cdot \log_2 \frac{0}{4} = 0bits \quad (2.17)$$

Observe que  $\log_2(0)$  não existe, mas sempre que este caso ocorrer ele será tratado como 0.

Para os demais valores de `clima` tem-se:

$$Entropia(jogar? | clima = sol) = Entropia_c(jogar? | clima = chuva) = 0,971bit \quad (2.18)$$

A entropia condicional da variável de saída em função da variável `clima` é:

$$Entropia(jogar? | clima) = \frac{5}{14} \cdot 0,971 + \frac{4}{14} \cdot 0 + \frac{5}{14} \cdot 0,971 = 0,693bit \quad (2.19)$$

A entropia condicional da variável de saída em função das demais variáveis de entrada são:

$$Entropia(jogar? | temperatura) = 0,911bit \quad (2.20)$$

$$Entropia(jogar? | umidade) = 0,79bit \quad (2.21)$$

$$Entropia(jogar? | vento) = 0,892bit \quad (2.22)$$

O ganho de informação indica quantos bits serão “economizados” para determinação da variável `S` se utilize o conhecimento da variável `T`.

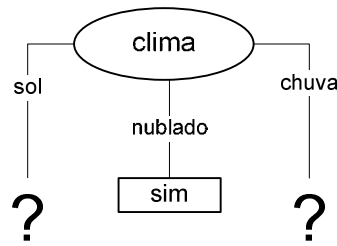
$$Ganho(S | T) = Entropia(S) - Entropia(S | T) \quad (2.23)$$

Os ganhos de informação da variável de saída do exemplo anterior são apresentados na tabela 2.13.

**Tabela 2.13 – Ganho de informação da variável de saída utilizando uma variável de entrada.**

Variável	Ganho
<code>clima</code>	0,247 bit
<code>temperatura</code>	0,029 bit
<code>umidade</code>	0,152 bit
<code>vento</code>	0,048 bit

A variável que apresenta maior ganho é `clima`, portanto `clima` é selecionada como o nó de solução para este subconjunto. A estrutura inicial da árvore é apresentada na figura 2.20.



**Figura 2.20 – Estrutura inicial da árvore de decisão.**

Para `clima` igual a `nublado` a variável de saída assume valor `sim`, para outro valor de `clima` a variável de saída apresenta mais de um tipo valor. O problema é particionado em dois grupos. No primeiro grupo entram os valores da tabela 2.12 que possuem valor de `clima` igual a `sol` e no segundo os valores da tabela que possuem valor de `clima` igual a `chuva`. Para cada grupo é realizado novo teste de ganho de informação. O processo se repete até que o critério de parada tenha sido atingido. A árvore completa é apresentada na figura 2.18.

Variáveis com muitos valores podem apresentar problemas para o algoritmo de ganho de informação. Uma vez que uma diversidade de valores tende a gerar resultados puros, variáveis com muitos valores tendem a ser classificadas como as melhores. Como alternativa é utilizada a taxa de ganho que é o ganho de informação dividido pela entropia da variável condicional. A taxa de ganho para os dados do exemplo são apresentados na tabela 2.14. A taxa de ganho é uma técnica utilizada no algoritmo C4.5.

**Tabela 2.14 – Taxa de ganho para as variáveis do problema sobre jogo de tênis.**

Variável	Ganho	Entropia da variável	Taxa de Ganho
<code>clima</code>	0,247 bit	1,577 bits	0,157
<code>temperatura</code>	0,029 bit	1,531 bits	0,019
<code>umidade</code>	0,152 bit	1 bit	0,152
<code>vento</code>	0,048 bit	0,985 bit	0,049

### **Aplicação de árvores de decisão no banco de dados**

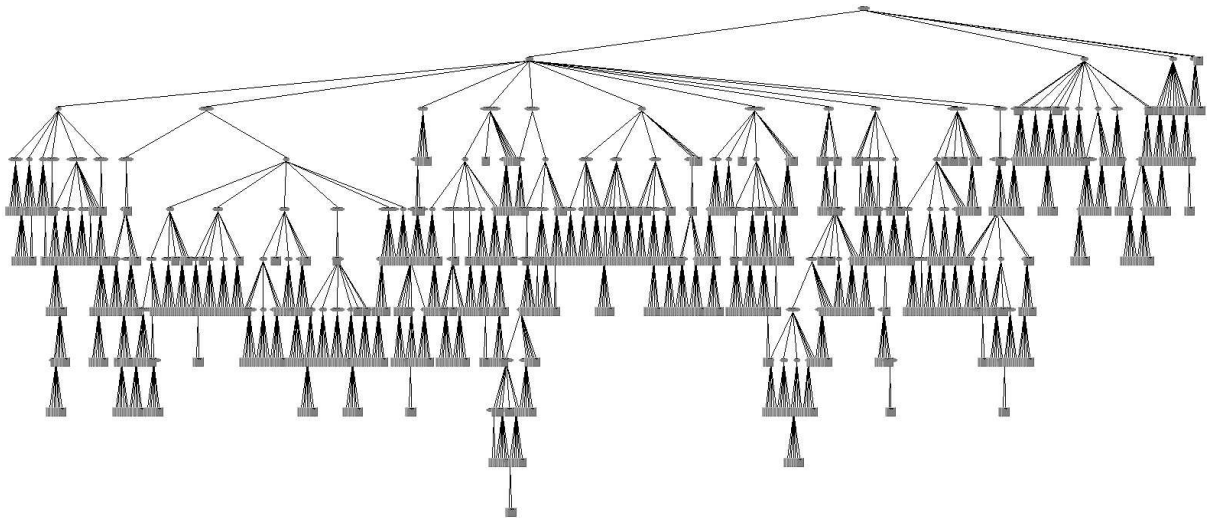
A árvore de decisão é uma técnica que classifica os elementos de acordo com um critério de saída. No problema de falha de medidores, a árvore gera regras que determinam os medidores com falha.

A árvore de decisão não encontra a correlação existente entre os dados de entrada, mas fornece um mecanismo de classificação da saída em função das entradas. As variáveis que se

encontram mais próximas ao nó central podem ser consideradas mais significativas, por serem, de acordo com a teoria, as que mais reduzem a informação.

Para construção de árvores de decisão foi utilizado o programa WEKA (*Waikato Environment for Knowledge Analysis*). O algoritmo de indução utilizado é o J48, uma versão em Java do algoritmo C4.5 de R. Quinlan.

A árvore obtida é muito grande tornando sua visualização difícil. Na figura 2.21 é mostrado um esboço da árvore completa. As regras desta árvore estão disponíveis no anexo B.



**Figura 2.21 – Árvore de decisão obtida com os dados da concessionária.**

A matriz de confusão é uma matriz quadrada que possui ordem igual a quantidade de valores possíveis que a variável de saída analisada assume, neste trabalho a matriz de confusão possui ordem 2. As linhas da matriz representam a resposta dos dados de observação (ou aprendizagem), enquanto as colunas representam a resposta de uma metodologia (*e.g.*, árvores de decisão). A tabela 2.16 apresenta a matriz de confusão gerada com os dados da árvore de decisão.

Cada valor da matriz de confusão representa a quantidade de dados com saída observada apresentada na linha como a saída classificada na coluna. Por exemplo, o elemento na linha 1 e coluna 1 representa a quantidade de registros do banco de dados que possuem saída observada "N" foram classificados como "N". Analogamente, o elemento da linha 1 e coluna 2 apresenta quantos registros que possuem saída observada "N" foram classificados como "S".

Tabela 2.15 – Matriz de confusão para a árvore de decisão.

		Classificação	
		N	S
Dados	N	2228	70
	S	215	382

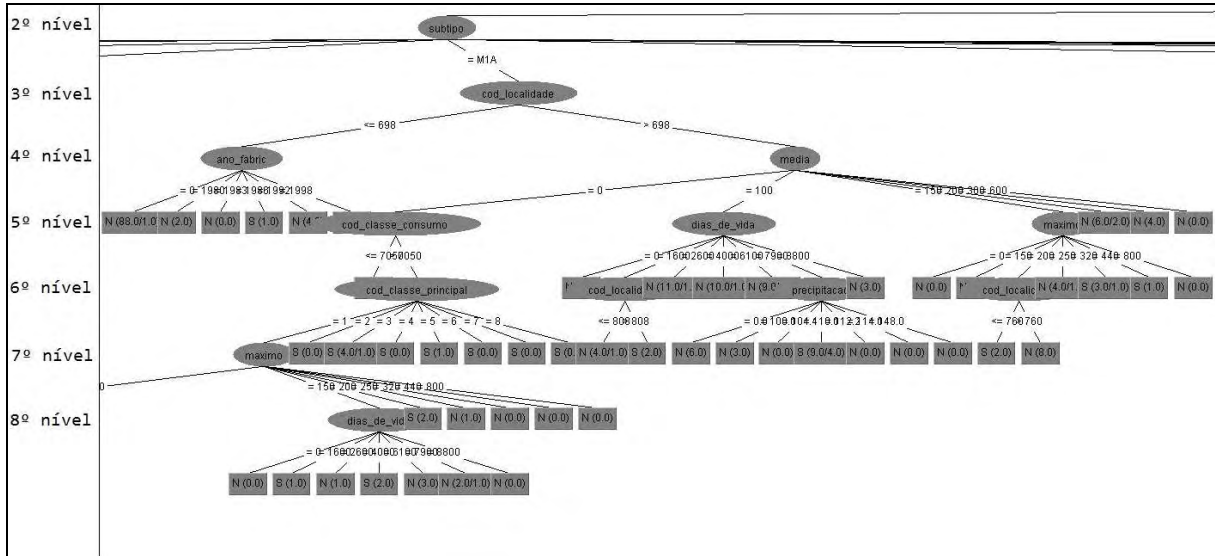
A matriz de confusão é uma matriz que possibilita a obtenção de várias medidas dos resultados: *accuracy*, *TP Rate*, *FP Rate*, *precision*, *recall*, *F-Measure*, etc. Se o leitor desejar obter informações sobre essas medidas, sugere-se a consulta de [61].

Neste trabalho, um dado relevante é a quantidade de equipamentos que apresentam defeito foram corretamente classificados como defeito. O elemento da linha 2 e coluna 2 da matriz de confusão apresenta a quantidade de registros que representam casos de defeito foram classificados como defeito enquanto o elemento desta mesma linha e coluna 1 apresenta a quantidade de registros que representam casos defeitos mas não foram classificados como tal. Assim, a porcentagem de registros que apresentam falha classificados corretamente pela árvore de decisão é 64%.

A árvore obtida na figura 2.21 possui 10 níveis de profundidade e 1115 folhas, uma árvore grande. Podemos fazer uma redução da árvore de decisão utilizando níveis: o primeiro nível é o nó central, o segundo nível são os nós descendentes do nó central e assim por diante.

O primeiro nível contém a variável `instalacoes`. Para os valores de 1, 2 e 3 de `instalacoes` o segundo nível contém a variável `subtipo`. Para o valor 4 de `instalacoes` o segundo nível contém a variável `ano_fabric`, e para os demais valores de `instalacoes` (5 e 6), não existe registros de equipamentos com falha.

A figura 2.22 apresenta um detalhe da árvore no segundo nível, quando `instalacoes=1` e `subtipo="M1A"`. Observa-se pela figura que a partir do quarto nível, esse segmento da árvore se torna muito complexo.



**Figura 2.22 – Detalhe da árvore para instalacoes="1" e subtipo="M1A".**

A utilização do quarto nível ou níveis mais profundos pode tornar o conhecimento da árvore muito específico para o conjunto de dados utilizado, podendo resultar em uma árvore que possui baixa eficiência para novos conjuntos de dados. Considerando somente os níveis 1, 2 e 3, as variáveis que apareceram na árvore foram: instalacoes, subtipo equip, ano\_fabric, media, cod\_localidade, cod\_classe\_principal, maximo, temperatura, cod\_tipo\_local, dias\_de\_vida, precipitacao.

Construindo uma árvore de decisão somente estas variáveis se obtém a matriz de confusão da tabela 2.16.

**Tabela 2.16 – Matriz de confusão para a árvore de decisão após poda.**

		Classificação	
		N	S
Dados	N	2226	72
	S	224	373

O percentual de casos de falha classificados corretamente pela nova árvore é 62,5%, representando uma redução de 2,4% (9 casos) na classificação. Assim, a remoção das variáveis cod\_tipo\_fase, contabil, cod\_classe\_consumo, grupo\_tensao\_fornecimento e altitude não influenciam no resultado de. A árvore gerada sem essas variáveis é muito

semelhante com a árvore da figura 2.21, sendo que nesta existem 9 níveis e 1114 folhas, uma a menos do que a árvore original.

## **2.10 Pós-processamento**

Na etapa de pós-processamento se faz a validação do modelo, eliminação de regras ou mesmo reavaliação das etapas anteriores com outras técnicas de MD. Esta etapa depende do conhecimento do especialista.

A validação do modelo pode ser feita através de diversas metodologias [62], porém em todas se utiliza erro de generalização. Antes de explicar erro de generalização deve-se definir alguns tipos de dados.

Os dados utilizados para a criação do modelo são chamados de dados de treinamento. Os dados utilizados para determinar o erro de generalização são chamados de dados de validação. Ambos os dados devem apresentar os valores das entradas e o valor da saída, *i.e.*, estes dados devem ser históricos e testados por metodologia válida, que no caso dos medidores de energia significa os dados da tabela RESULTADOS, que foram testados com através da metodologia de ensaio de medidores de energia elétrica.

O método de validação *hold-out* é o método mais simples existente na literatura. Neste método são utilizados os dados de treinamento e validação. O modelo é construído com os dados de treinamento e depois as entradas dos dados de validação são aplicadas ao modelo. A diferença entre a saída fornecida pelo modelo e o valor esperado da saída (fornecida pela resposta dos dados de validação) é o erro de generalização. Para casos em que a resposta é discreta, o erro de generalização é a quantidade de valores em que a saída do modelo diferenciou da saída esperada, dividida pela quantidade total de registros dos dados de validação, em forma percentual.

O erro de generalização é o erro resultante da utilização do modelo em dados que não foram utilizados para sua criação. Assim, o erro é uma medida de quão generalizado é o modelo.

### **2.10.1 Modelo final**

As técnicas de mineração de dados utilizadas possuem objetivos distintos. A regressão *stepwise* determina as variáveis de entrada mais significativas na definição de uma equação

matemática que representa a saída. A árvore de decisão mantém em seus nós superiores as variáveis que possibilitam a maior redução da quantidade de bits necessários para representar a saída, ou seja, as variáveis mais significativas para classificar a saída.

Pensando que as variáveis mais significativas para representação da saída são aquelas que possuem maior significância para a representação da saída através de uma equação e que possuem maior capacidade classificativa, através da redução da quantidade de bits para a classificação da saída. Estas variáveis são as que foram selecionadas por ambas as técnicas de mineração de dados, sendo consideradas significativas tanto para problemas de classificação quanto para problemas de regressão. As variáveis selecionadas por ambas as técnicas são: `subtipo equip`, `instalacoes`, `dias_de_vida`, `cod_tipo_local`, `cod_localidade` e `cod_classe_principal`. Um modelo de classificação (árvores de decisão) utilizando somente essas variáveis consegue classificar corretamente apenas 17,1% dos dados de falha.

Como o problema está mais para um problema de classificação do que um problema de regressão pode-se acrescentar variáveis que foram selecionadas somente pela técnica de classificação.

Acrescentando a variável `media`, que representa a quantidade de energia a que o medidor está submetido, a classificação correta de medidores com falha aumentou para 43,9%. O mesmo estudo utilizando a variável `maximo` apresenta uma classificação correta de 43,7%.

A utilização de ambas as variáveis apresenta 56,4% de classificações corretas, o que significa que as falhas nos medidores são melhores classificadas ora pela média do consumo mensal e ora pelo máximo valor de consumo mensal. Sendo assim, a inclusão de ambas variáveis é mais indicada para a construção do modelo.

A inserção da variável `temperatura` não altera a quantidade de casos de falha classificados corretamente, portanto esta variável pode ser descartada.

Inserindo a variável `precipitacao` ao conjunto de dados a quantidade de classificações corretas de equipamentos com falha aumentou para 61,1%. Já o uso da variável `ano_fabric` alterou de maneira pouco significativa a classificação. Com o uso de ambas as variáveis se obtém o mesmo subconjunto obtido pela árvore de decisão.

As variáveis selecionadas para compor o modelo final são: `subtipo equip`, `instalacoes`, `media`, `maximo`, `dias_de_vida`, `cod_tipo_local`, `cod_localidade`, `cod_classe_principal` e `precipitacao`.

## 2.11 Considerações finais

O processo de DCBD possibilita a descoberta de conhecimento previamente oculto nas bases de dados da concessionária. É uma tarefa dispendiosa e subjetiva, estando às custas do conhecimento do especialista e sua prática com várias técnicas para garantir a qualidade da saída obtida. Cabe ao especialista decidir quais variáveis devem ser utilizadas no processo, as técnicas utilizadas no pré-processamento, objetivos e técnicas utilizadas na mineração de dados, além de decidir como o resultado será pós-processado.

O algoritmo de regressão *stepwise* apresentou baixa redução de RMSE, indicando a baixa eficiência em se utilizar uma regressão linear para estimar uma variável de saída discreta. Mesmo assim, a análise da variância apontou como variáveis significativas aquelas selecionadas pelos três primeiros níveis da árvore de decisão, com exceção de `contabil`, indicando que a metodologia específica para seleção de principais variáveis de regressão linear consegue determinar algumas das principais variáveis para problemas onde a saída não é uma função linear das entradas. Como sugestão pode-se utilizar uma regressão *stepwise* onde os dados são ajustados através de uma regressão logística, técnica de regressão utilizada para saídas discretas.

A árvore de decisão obtida ainda é muito complexa para ser utilizada como metodologia de classificação dos medidores. A análise da árvore apresenta características interessantes. No nó central, existe uma grande distinção entre os medidores com apenas uma instalação e os demais, quase podendo-se afirmar que existe uma árvore própria para o caso de primeira instalação. Isto se relaciona à distribuição dos dados nos valores de `instalacoes`, para primeira instalação existem mais de 2000 casos com 30% destes de falha, para a segunda instalação existem 380 casos com apenas 5% destes sendo de medidores com falhas. Para terceira instalação a quantidade de casos é menor, e assim por diante.

A metodologia utilizada apresentou variáveis que classificaram, por meio de uma árvore de decisão, mais de 50% dos casos. Cada variável escolhida pode ser relacionada com a saída como segue:

- O modelo do equipamento (`subtipo Equip`) apresenta dados sobre a característica construtiva do equipamento. Os equipamentos agrupados são do mesmo fabricante e possuem características construtivas semelhantes. As

principais diferenças entre eles está no nível de tensão para o qual foram projetados.

- A quantidade de instalações do equipamento (*instalacoes*) indica que este passou por diversos padrões de consumo, e provavelmente teve um período sem registros relacionado a troca de localidades. Esta variável não aparenta ser a mais significativa, pois não indica como eram os padrões de consumo pelos quais o medidor passou, somente apontando que ele foi utilizado em padrões diferentes;
- O consumo médio (*media*) e máximo (*maximo*) na última UC apresenta informações sobre o padrão de consumo ao qual o equipamento está submetido.
- O tempo de permanência no último padrão é informado pela variável *dias\_de\_vida*.
- A variável *cod\_tipo\_local* informa se o local da última instalação é urbano ou rural. A probabilidade das instalações elétricas em zonas rurais serem atingidas por raio é maior, podendo danificar os equipamentos elétricos, inclusive medidores de energia. Outras características típicas de instalações em zona rural é a maior susceptibilidade à poeira e variações de temperatura
- A localidade (*cod\_localidade*) em que o medidor se encontra não apresenta uma relação específica com as falhas. Os dados de temperatura e altitude são em relação à localidade do equipamento, porém essas variáveis não apresentaram influência em nenhuma técnica. A localidade pode conter uma relação com a qualidade de energia (*e.g.*, harmônicos), probabilidade de raios, ou mesmo fraude.
- A variável *cod\_classe\_principal* apresenta uma relação direta com a quantidade de energia aferida pelo equipamento. Em unidades residenciais o consumo é relativamente baixo, comparado com os consumidores industriais.
- Por último, a quantidade média de chuva (*precipitacao*) sobre a cabine de medição pode ser um indicador de falhas nos equipamentos. Cabines de medição sujeitas a uma quantidade maior de chuvas se degradam mais rapidamente, dependendo da intensidade da chuva as vedações não são suficientes, permitindo a entrada de água na cabine ou mesmo no interior do medidor. Estas são considerações hipotéticas.

### 3 METODOLOGIA DE CÁLCULO DA PROBABILIDADE DE FALHA DOS EQUIPAMENTOS

---

---

As variáveis para a determinação de falhas nos medidores eletromecânicos de energia elétrica foram selecionadas no capítulo 2. Através delas é possível construir um modelo que determine os estados dos medidores instalados no parque, apresentando a probabilidade de falha associada a cada equipamento.

Um modelo estatístico é capaz de informar as condições dos equipamentos associando a cada um deles a probabilidade de falha do equipamento, *i.e.*, qual a probabilidade dele estar funcionando com erros de medição fora da faixa permitida. A partir de um modelo estatístico que contém as informações da probabilidade de falha de cada equipamento pode-se construir um programa computacional que auxilie na tomada de decisões informando quais equipamentos devem ser removidos do parque, selecionando os equipamentos com maior probabilidade de falha como indicados à remoção.

Para a construção do modelo de probabilidade de falha os equipamentos são agrupados em classes de acordo com a probabilidade de falha. Equipamentos com probabilidade de falhas próximas são agrupados. O agrupamento possibilita generalizar o comportamento de falha do equipamento, associando a este a probabilidade de falha da classe. Como resultado tem-se um modelo com baixa susceptibilidade a ruídos (novos casos com probabilidade de falhas divergente da classe) e de fácil visualização comportamental do parque.

A probabilidade de falha da classe é obtida através da divisão da quantidade de equipamentos que possuem na saída a indicação de falha pela quantidade total de equipamentos na classe. Por exemplo: em uma classe existem 87 equipamentos, 39 apresentam falha. A probabilidade de falha desta classe é definida como:

$$\frac{39}{87} \times 100 = 44,82\% \quad (3.1)$$

O modelo é dividido em duas etapas:

- Na primeira etapa as variáveis de entrada são analisadas, determinando-se pesos para cada faixa de valores delas;

- Na segunda etapa, os pesos de cada variável de entrada do medidor são somados e arredondados inferiormente para um número múltiplo de dez, formando grupos de valores (classes). A probabilidade de falha de cada uma dessas classes determinará a probabilidade de falha final dos medidores pertencentes a elas.

### 3.1 Determinação dos pesos das variáveis de entrada

De acordo com a metodologia de discretização do capítulo 2, as variáveis de entrada contínuas tiveram seus valores discretizados em faixas. Para a primeira etapa de construção do modelo, a quantidade de equipamentos em cada faixa de valores é analisada, comparando a quantidade de equipamentos que apresentam saídas que indiquem falha dos equipamentos com a quantidade total de equipamentos na mesma faixa. Assim, para cada faixa de valor existente nas oito variáveis de entrada, é associada uma probabilidade de falha. Esta probabilidade será denominada *peso*. Cada variável possui tantos valores de *peso* quantos forem a quantidade de faixa de valores que ela possuir. Nos casos em que as variáveis de entrada são discretas, os *pesos* são calculados para cada valor da variável.

As tabelas de 3.1 a 3.9 apresentam a probabilidade de falha dos equipamentos para cada faixa de valores das variáveis *subtipo equip*, *instalacoes*, *media*, *maximo*, *dias\_de\_vida*, *cod\_tipo\_local*, *cod\_localidade*, *cod\_classe\_principal* e *precipitacao*, respectivamente.

**Tabela 3.1 – Classificação da probabilidade de falha para a variável *subtipo equip*.**

<b>subtipo equip</b>	<b>Nº de casos</b>	<b>peso</b>
D58	378	35,98
D58C	1548	13,44
D8L	798	4,26
FV201	429	25,17
M1A	570	15,79
M8L	652	28,99
MFT120G	291	15,81
MV202	619	11,95
MY202	1005	3,68
PN5DG	427	11,71
T8L	274	14,23

Tabela 3.2 – Classificação da probabilidade de falha para a variável *instalacoes*.

<b>Instalacoes</b>	<b>Nº de casos</b>	<b>peso</b>
1	6139	15,54
2	670	6,12
3	152	8,55
4	22	13,64
5	6	0,00
6	2	0,00

Tabela 3.3 – Classificação da probabilidade de falha para a variável *media*.

<b>media</b>	<b>Nº de casos</b>	<b>peso</b>
0 <= valor < 100	1535	14,92
100 <= valor < 150	1487	14,86
150 <= valor < 200	1203	15,05
200 <= valor < 300	1311	12,89
300 <= valor < 600	989	13,75
valor >= 600	466	16,09

Tabela 3.4 – Classificação da probabilidade de falha para a variável *maximo*.

<b>maximo</b>	<b>Nº de casos</b>	<b>peso</b>
0 <= valor < 150	1076	16,36
150 <= valor < 200	918	16,67
200 <= valor < 250	1056	14,68
250 <= valor < 320	1063	12,42
320 <= valor < 440	1059	13,69
440 <= valor < 800	1046	13,10
valor >= 800	773	14,62

Tabela 3.5 – Classificação da probabilidade de falha para a variável *dias\_de\_vida*.

<b>dias_de_vida</b>	<b>Nº de casos</b>	<b>peso</b>
0 <= valor < 1600	1064	8,83
1600 <= valor < 2600	1068	9,64
2600 <= valor < 4000	977	14,64
4000 <= valor < 6100	1037	13,02
6100 <= valor < 7900	1040	22,12
7900 <= valor < 8800	922	22,23
8800 <= valor	883	11,44

Tabela 3.6 – Classificação da probabilidade de falha para a variável `cod_tipo_local`.

<code>cod_tipo_local</code>	Nº de casos	peso
RR	841	18,43
UB	6150	13,92

Tabela 3.7 – Classificação da probabilidade de falha para a variável `cod_localidade`.

<code>cod_localidade</code>	Nº de casos	peso
1	123	14,63
8	50	16,00
11	32	25,00
18	57	12,28
31	532	19,55
39	45	24,44
45	18	27,78
49	18	16,67
53	41	17,07
54	70	20,00
55	9	0,00
57	206	14,08
58	122	11,48
60	51	21,57
61	133	24,06
62	95	25,26
63	11	27,27
69	357	24,65
71	35	5,71
75	452	15,71
79	42	14,29
81	177	23,73
85	19	10,53
89	32	15,63
90	6	0,00
92	4	0,00
93	204	10,78
121	35	8,57
123	255	7,84
124	81	12,35
126	89	11,24
127	75	5,33
128	51	7,84
129	104	10,58
130	35	0,00
132	20	15,00
133	44	9,09
135	14	7,14
136	137	10,22
138	28	0,00
139	3	0,00
142	12	0,00
144	22	4,55
145	135	8,89
146	1	0,00

147	1	0,00
148	1	0,00
150	5	0,00
151	3	0,00
156	1	0,00
157	66	0,00
158	6	0,00
159	3	0,00
160	2	0,00
161	37	0,00
162	3	0,00
163	3	0,00
173	100	0,00
181	87	19,54
182	2	50,00
185	8	50,00
190	30	20,00
194	49	18,37
199	86	15,12
200	225	14,67
201	5	0,00
206	6	33,33
207	47	21,28
210	34	32,35
212	7	14,29
215	125	12,80
217	1	0,00
219	1	0,00
220	57	7,02
221	3	0,00
222	8	25,00
223	3	0,00
224	6	0,00
226	3	33,33
228	114	17,54
230	32	18,75
231	21	19,05
232	1	0,00
233	47	8,51
234	15	13,33
235	22	18,18
236	72	13,89
237	27	7,41
238	1	0,00
239	8	0,00
240	3	0,00
241	9	22,22
242	1	0,00
243	2	0,00
244	22	4,55
245	2	0,00
246	4	25,00
256	13	7,69
259	7	0,00
265	10	0,00
266	12	8,33
280	40	7,50
290	76	1,32

295	45	6,67
297	28	3,57
307	2	0,00
310	2	50,00
317	9	0,00
350	11	0,00
355	10	0,00
360	10	0,00
365	54	5,56
375	6	0,00
380	9	0,00
551	33	0,00
554	10	50,00
555	18	0,00
556	27	22,22
557	15	13,33
561	4	0,00
566	6	16,67
576	13	0,00
578	4	0,00
581	6	0,00
589	7	0,00
590	6	0,00
592	6	16,67
626	18	0,00
630	27	7,41
632	13	23,08
634	7	14,29
639	16	12,50
642	11	9,09
645	9	11,11
648	12	16,67
652	32	6,25
654	2	50,00
659	27	14,81
665	21	23,81
669	5	20,00
672	36	11,11
676	6	16,67
682	12	58,33
684	2	50,00
686	9	55,56
689	20	20,00
690	7	28,57
692	31	6,45
695	176	2,84
698	10	20,00
699	9	33,33
700	55	9,09
702	19	15,79
704	13	15,38
706	62	8,06
710	15	6,67
712	7	28,57
746	40	32,50
752	1	0,00
756	4	50,00
758	8	0,00

760	1	0,00
761	3	33,33
763	8	37,50
769	4	25,00
771	6	16,67
774	1	0,00
776	4	25,00
779	1	0,00
781	1	0,00
783	2	0,00
785	3	0,00
787	4	50,00
789	1	0,00
792	6	33,33
794	1	100,00
798	3	0,00
799	2	50,00
801	2	0,00
804	1	0,00
808	69	37,68
811	5	20,00
813	2	0,00
817	2	50,00
820	5	20,00
822	2	0,00
823	1	100,00
828	2	0,00
831	8	25,00
836	3	33,33
844	11	27,27
849	18	33,33
854	3	0,00
857	8	12,50
861	4	25,00
864	9	11,11
870	2	0,00
872	2	50,00
876	2	0,00
884	1	0,00
889	2	50,00
893	9	22,22
902	1	0,00
904	1	0,00
906	2	100,00
912	1	0,00
914	1	0,00
922	3	33,33
924	1	100,00
932	7	42,86

Tabela 3.8 – Classificação da probabilidade de falha para a variável `cod_classe_principal`.

<code>cod_classe_principal</code>	Nº de casos	peso
1	5474	14,01
2	113	21,24
3	735	11,43
4	558	20,79
5	85	18,82
6	10	20,00
7	14	7,14
8	2	50,00

Tabela 3.9 – Classificação da probabilidade de falha para a variável `precipitacao`.

Precipitação	Nº de casos	peso
0 <= valor < 100	1240	17,58
100 <= valor < 104,4	1104	15,13
104,4 <= valor < 110	1054	12,05
110 <= valor < 112,2	1157	22,13
112,2 <= valor < 114	1031	12,71
114 <= valor < 148	895	10,17
valor >= 148	510	4,12

### 3.2 Criação das classes finais

A segunda etapa do modelo consiste em determinar as classes finais aos quais os medidores devem pertencer.

Essas classes são determinados analisando os pesos que os medidores receberam em cada variável de entrada. Uma vez determinado o peso de todas as variáveis, eles devem ser somados e o valor final aproximado para o número inteiro inferior mais próximo múltiplo de dez. A tabela 3.10 apresenta a determinação do grupo de risco de um medidor exemplo.

Tabela 3.10 – Exemplo de determinação do grupo de risco de um medidor.

Variável	Faixa	Peso
<code>subtipo equip</code>	PN5DG	11,71
<code>instalacoes</code>	1	15,54
<code>media</code>	0 <= valor < 100	14,92
<code>maximo</code>	0 <= valor < 150	16,36
<code>dias de vida</code>	1600 <= valor < 2600	9,64
<code>cod_tipo_local</code>	UB	13,92
<code>cod_localidade</code>	215	12,80
<code>cod_classe_principal</code>	1	14,01
<code>precipitacao</code>	0 <= valor < 100	17,58
<b>TOTAL</b>		<b>126,48</b>
<b>APROXIMAÇÃO</b>		<b>120</b>

Com todos os medidores devidamente agrupados, calcula-se a probabilidade de falha de cada classe, através da divisão da quantidade de equipamentos que apresentam uma saída indicando falha pela quantidade total de equipamentos da classe. Esta também será a probabilidade de falha de cada equipamento pertencente à classe. A tabela 3.11 apresenta as classes finais, a probabilidade de falha dos mesmos e a quantidade de equipamentos em cada classe.

**Tabela 3.11 – Modelo final para determinação de probabilidade de falha nos medidores.**

<b>Classe</b>	<b>Nº de casos</b>	<b>Probabilidade de falha</b>
70	10	0,00%
80	86	1,16%
90	375	1,87%
100	780	4,23%
110	934	5,35%
120	1453	8,33%
130	1323	14,74%
140	747	19,14 %
150	581	32,01%
160	425	35,53%
170	224	40,63%
180	40	55,00%
190	8	75,00%
210	1	100,00%
220	4	100,00%

### **3.3 Minimização de erros**

Com o passar do tempo, a probabilidade de falha real dos medidores será diferente da probabilidade apresentada pelo modelo. Isto ocorre porque, devido à complexidade do problema, a modelação não é ideal.

O modelo não utiliza todas as variáveis que contribuem para os erros de medição, somente as que mais influenciam no processo de determinação dos erros. Além disso, variáveis que não apresentam influencia direta sobre os erros de medição, como `cod_localidade` e `subtipo equip`, não possuem uma probabilidade de falha constante com o tempo. Uma localidade não é a causadora direta de falha na medição, estando a falha associada a um comportamento específico existente na localidade, ainda não determinado. Por exemplo: considerando que o método de construção dos painéis/caixas/cubículos de medição seja uma característica de influencia nos erros de medição: em uma localidade, 40% dos

painéis/caixas/cubículos que estavam fora das padronizações foram reformados para um modelo de acordo com as normas atuais, como consequência, espera-se que nesta localidade a probabilidade de falha se torne menor.

Uma solução é periodicamente adicionar novos dados de ensaios e ajustar a probabilidade de falha das classes em função destes. Não é desejável que ocorra uma evolução do modelo, *i.e.*, não se deseja que os pesos das variáveis de entrada sejam recalculados, o que resultaria em classes de medidores com mesmo valor das classes atuais porém compostas por equipamentos com características divergentes das atuais, ou seja, resultando em um novo modelo. A metodologia de evolução do modelo parece ser a mais indicada para a adequação aos erros dos parâmetros, mas a alteração das classes de medidores impede a análise da variação da probabilidade de falha da saída do modelo com a inserção dos novos dados.

A inserção de novos dados deve ser utilizada para ajustar os valores das probabilidades de falhas das classes de medidores. Para cada conjunto de novos dados de ensaio, eles devem ser agrupados de acordo com a metodologia proposta anteriormente, utilizando os valores apresentados nas tabelas de 3.1 a 3.9. Uma vez agrupados, a probabilidade de falha deve ser obtida analisando somente os novos dados. O modelo deve ser reajustado utilizando esses novos dados de probabilidade de falha e os dados existentes na tabela 3.10. Os novos valores de probabilidade de falha são obtidos através da regressão via kernel.

### 3.3.1 Regressão via kernel

Para explicar a regressão via kernel, será realizada uma revisão sobre regressão.

Utilizando a convenção adotada no capítulo 2, uma regressão pode ser expressa da seguinte forma:

$$Y = g(X) + \varepsilon \quad (3.2)$$

Isto significa que a saída é uma função da entrada, acrescida de um erro  $\varepsilon$ .

As regressões podem ser classificadas em dois tipos, paramétricas e não-paramétricas. Em [63] é apresentada a diferença entre ambas: Entende-se que um modelo paramétrico assume que  $g(X)$  é uma função desconhecida num número finito de parâmetros, e nosso trabalho é estimar os parâmetros desconhecidos, por exemplo por mínimos quadrados. Em um

modelo não-paramétrico a relação funcional entre as duas variáveis vive num espaço de funções muito mais amplo: assumimos apenas que  $g(X)$  está num espaço de funções seguindo algumas restrições convenientes e buscamos uma combinação linear de funções desse espaço que aproximem bem de  $g(X)$ .

A regressão via kernel é uma técnica não paramétrica onde as estimações são realizadas utilizando os dados de treinamento. Ela basicamente realiza uma média ponderada dos valores de saída, utilizando uma função kernel para determinar o valor dos pesos da ponderação. Em uma média sem ponderação, o valor de saída é obtido por:

$$\hat{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (3.3)$$

Em uma média ponderada, para cada valor de saída dos dados de treinamento é adicionado um peso  $p_i$ , ajustando suas contribuições para determinação do ponto de saída desejado. A média ponderada é apresentada na equação 3.4.

$$\hat{y} = \frac{\sum_{i=1}^n y_i \cdot p_i}{\sum_{i=1}^n p_i} \quad (3.4)$$

A distância ponderada é uma técnica de regressão baseada na média ponderada. Para determinação dos pesos a técnica utiliza uma modificação da medida de distância entre o ponto de entrada do dado que se deseja determinar,  $q$ , e os pontos de entrada dos dados de treinamento. Existem várias fórmulas de distância, mas geralmente utiliza-se a distância Euclidiana, apresentada na equação 3.3. Para a regressão ponderada utiliza-se o inverso da distância Euclidiana.

$$d(x_i, q) = \sqrt{(x_i - q)^2} \quad (3.3)$$

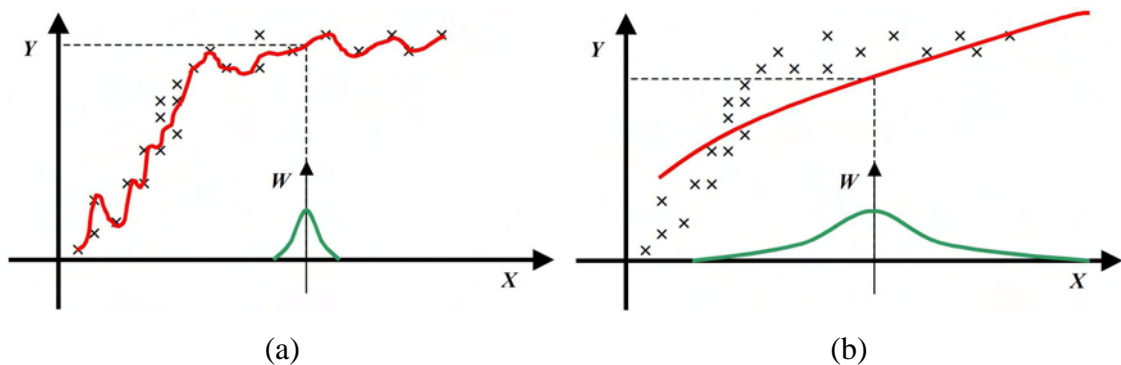
A regressão kernel é semelhante a regressão de distância ponderada, mas ao invés de utilizar o inverso da distância, aplica-se uma função kernel sobre a medida de distância. Existem vários tipos de função kernel, mas neste trabalho utilizaremos a função kernel normal, também conhecida como gaussiana. A equação de uma função kernel gaussiana

genérica é apresentada na equação 3.4. Os pontos que estiverem mais próximos de  $q$  receberão um peso maior enquanto que os pontos mais distantes receberão um peso menor, influenciando menos na determinação do valor de saída.

$$K(d(x_i, q)) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{d(x_i, q)^2}{2\sigma^2}\right) \quad (3.4)$$

A função kernel gaussiana genérica apresenta um parâmetro adicional,  $\sigma$ , denominado largura de banda. Este parâmetro altera a variância da função kernel. Em termos práticos, isto significa alterar o tamanho do espaço de pontos significativos na determinação da resposta local: um valor elevado de  $\sigma$  aumenta o espaço de pontos que influenciam na resposta de forma significativa, fazendo com que o resultado da regressão seja suave em todo o espectro; um valor muito pequeno de  $\sigma$  diminui o espaço de pontos que influenciam na resposta de maneira significativa, adequando-se a regressão muito bem aos dados de treinamento, porém, também aos ruídos.

No trabalho de Galotto [24] são apresentados os efeitos do uso de valores de  $\sigma$  elevado e pequeno. Aqui são apresentadas duas imagens do trabalho, uma para o caso de largura de banda com valor pequeno e outro com valor elevado. Ambas as imagens são apresentadas na figura 3.1. Os dados de treinamento estão indicados por  $\times$ , a curva vermelha representa a resposta da regressão kernel e a curva em verde é a representação da função kernel gaussiana genérica.



**Figura 3.1 – Efeito de  $\sigma$  sobre a regressão kernel. a) Resultado para  $\sigma$  pequeno. b) Resultado para  $\sigma$  elevado. Fonte: [24]**

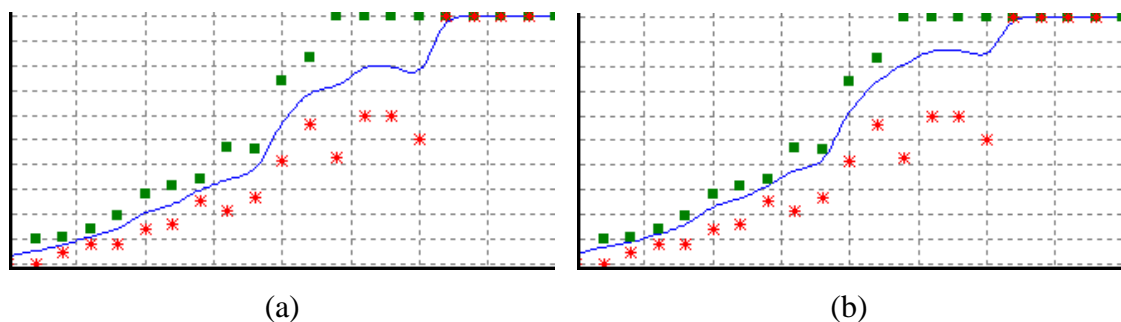
Pela figura se observa como  $\sigma$  altera a função kernel e, conseqüentemente, a resposta da regressão. O ideal é que o valor de  $\sigma$  seja suficientemente elevado para não estar sujeito a

ruídos, porém, seja suficientemente pequeno ao ponto de manter uma forma de onda próxima a real.

### 3.3.2 Peso inicial

Após a inserção de novos dados ao modelo, estes apresentarão novas probabilidades de falha para as classes, inclusive podendo apresentar probabilidades de falhas para classes que não estavam presentes no modelo original. A regressão kernel é utilizada para ajustar novos valores de probabilidade de falha, sendo que os novos pontos possuem influência tanto do modelo original quanto dos novos dados inseridos.

Para melhorar a resposta da regressão kernel um novo parâmetro, denominado `peso inicial`, foi criado. O `peso inicial` define a quantidade de vezes que os dados do modelo original serão repetidos no conjunto de dados de treinamento. Isto permite definir qual a importância do modelo original sobre os novos dados: um valor de `peso inicial` nulo representa que os dados originais possuem o mesmo peso que os novos dados na determinação da regressão. Um valor unitário indica que os dados originais possuem o dobro de importância em relação aos dados novos na determinação da resposta da regressão. A figura 3.2 apresenta um exemplo da influência do `peso inicial` na determinação de um novo modelo. Os pontos verdes representam os dados iniciais, os pontos vermelhos os novos dados e a curva em azul é a resposta da regressão kernel.



**Figura 3.2 – Efeito do `peso inicial` sobre a resposta da regressão kernel. a) Com `peso inicial` = 0. b) Com `peso inicial` = 1.**

Quando o `peso inicial` é nulo a regressão se localiza aproximadamente na metade da distância entre os pontos do modelo original e os novos dados, quando o valor de `peso inicial` é unitário a resposta da regressão kernel se aproxima dos dados originais.

### **3.4 Considerações finais**

Neste capítulo, as variáveis obtidas como mais significativas para determinação de erros nos medidores eletromecânicos de energia ativa foram utilizadas na construção de um modelo de determinação de probabilidade de falhas. O modelo permite determinar grupos de medidores com características de falhas semelhantes, e associar a cada grupo uma probabilidade de ocorrência de falha, podendo ser utilizado para estimar a situação de uma diversidade de equipamentos de medição instalados no parque de medição.

A técnica utilizada permite que a classificação seja realizada por meio de tabelas, onde para cada resposta o equipamento recebe um peso. A classe do equipamento é dada em função da soma e arredondamento desses pesos, o que torna o tempo de agrupamento um valor fixo, ao contrário de uma árvore de decisão que pode classificar um equipamento utilizando apenas uma variável e outro utilizando todas as variáveis.

Pela metodologia, a menor classe que se pode obter é 60 e a maior 300. Considerando que todas as classes nesse intervalo existam, a metodologia pode classificar um equipamento em uma das 25 classes existentes.

Como as classes não podem se alterar com o tempo, mesmo que a probabilidade de falha associada a cada variável de entrada (*e.g.*, localidade) se altere, a utilização de uma metodologia de minimização de erro da resposta da probabilidade de falhas das classes é necessária. A regressão kernel se mostra adequada ao problema apresentado, traçando a curva que melhor se ajusta aos pontos para a largura de banda selecionada.

## 4 APLICATIVO EM PLATAFORMA AMIGÁVEL

---

---

Este capítulo apresenta a interface desenvolvida para análise da probabilidade de falha dos medidores de energia.

O aplicativo foi desenvolvido utilizando a linguagem de programação Delphi. Para o armazenamento dos dados foi utilizado o banco de dados MySQL.

O aplicativo foi desenvolvido de forma que o banco de dados é iniciado juntamente com a aplicação e finalizado com a mesma, não sobrecarregando o sistema.

Serão apresentadas todas as funcionalidades, aqui descritas:

- Importação de dados com informações dos medidores instalados no parque;
- Importação de dados de ensaio de medidores;
- Criar novos modelos através dos dados de ensaios.
- Visualização da probabilidade de falha dos medidores instalados no parque;
- Sistema especialista para amostragens de medidores com maior probabilidade de falha;
- Impressão de relatórios.

### 4.1 *Tela inicial*

O aplicativo recebeu o nome WhFailure. A tela inicial do programa é apresentada na figura 4.1.



**Figura 4.1 – Tela inicial do aplicativo.**

Todas as interfaces do programa são confinadas dentro deste espaço.

## **4.2 Importação**

O aplicativo WhFailure possui uma interface específica para a realização de importações de dados relacionados com a constituição física e cronológica do parque de medidores de energia elétrica instalados em todas as áreas da concessionária, bem como os dados de consumo registrados pelos medidores durante um período de tempo estipulado (5 anos). Além disso, o aplicativo também possibilita a importação de dados relacionados com resultados experimentais realizados em uma população de unidades medidoras para avaliação de falhas nas mesmas.

O processo de importação é auxiliado por dois assistentes de importação: um responsável pelo auxílio na importação de dados relacionados com o parque e o consumo das unidades medidoras, e outro referente à importação de dados relacionados com ensaios de teste de falha de operação de medidores. A figura 4.2 apresenta o menu de importação.

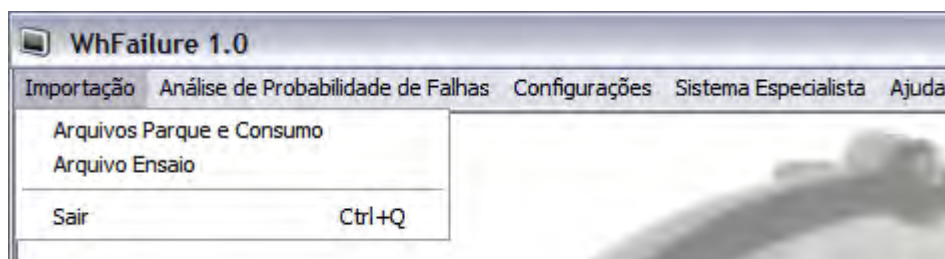


Figura 4.2 – Menu de importação.

### 4.2.1 Importação de dados do parque e de consumo dos medidores

O WhFailure, a partir do item `Arquivos Parque e Consumo` do menu `Importação`, possibilita a importação de dados relacionados com a constituição física e cronológica do parque de medidores de energia elétrica tipo indução instalados em toda a área de abrangência das regionais, bem como os dados de consumo registrados pelos mesmos durante um período de tempo de 5 anos. Os dados são importados simultaneamente a partir de dois arquivos no formato `.csv`, denominados de `uc_medidor.csv` (dados do parque de medidores) e `med_cons.csv` (dados de consumo), e devem respeitar uma padronização de formatação.

O formato *comma-separated values* (csv) define que os campos sejam separados por vírgula e que cada valor seja encapsulado por aspas.

O arquivo `uc_medidor.csv` deve ser gerado pela concessionária, contendo os registros com informações dos equipamentos instalados no parque de medição a serem importados pelo WhFailure, obedecendo a formatação e a ordem apresentada na Tabela 4.1.

Tabela 4.1 - Formatação padrão para geração do arquivo "uc\_medidor.csv".

Variável	Tipo	Tamanho máximo
uc	inteiro	4294967295
cod_tipo_fase	texto	2
cod_situacao	texto	2
cod_tipo_local	texto	2
cod_localidade	inteiro	65535
tipo equip	texto	2
num equip	texto	10
cod_tipo_espec	texto	3
data_inst equip	data	—
data_retirada equip	data	—
subtipo equip	texto	10
contabil	texto	3
cod_classe_principal	inteiro	255

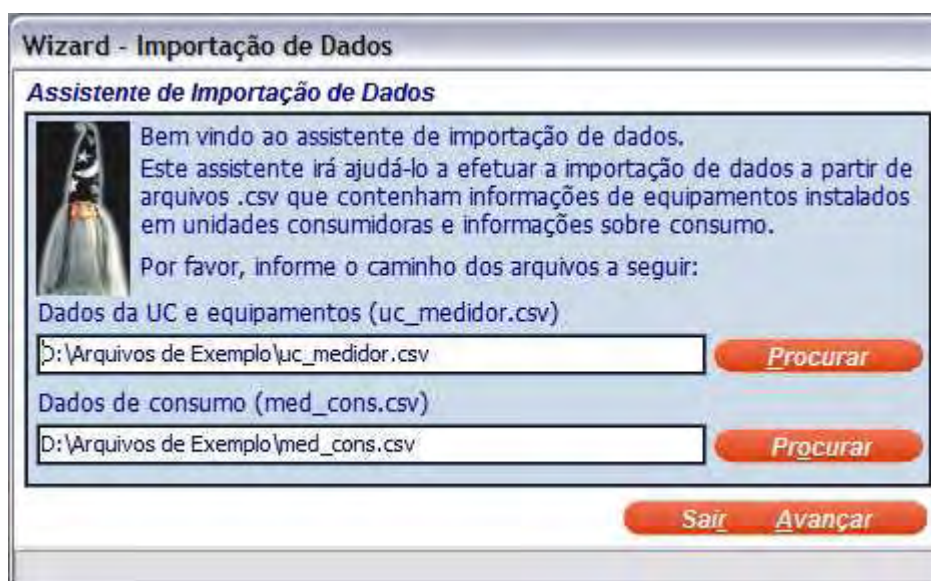
Uma variável adicional do banco de dados (*cod\_situacao*) é utilizada pelo aplicativo. A função da variável é indicar se a unidade consumidora está energizada, *i.e.*, se está ocorrendo a medição e faturamento. O uso desta garante que após o processo de importação, somente as unidades consumidoras com medidores em uso serão analisadas.

O arquivo *med\_cons.csv* deve conter os registros de consumo das unidades consumidoras a serem importados pelo WhFailure, obedecendo a formatação e ordem mostrada na tabela 4.2.

**Tabela 4.2 - Formatação padrão para geração do arquivo "med\_cons.csv".**

Variável	Tipo	Tamanho máximo
conta	inteiro	4294967295
leitura_1	real	9999999,9
leitura_2	real	9999999,9
leitura_3	real	9999999,9
...	...	...
leitura_60	real	9999999,9
data_ultima_leitura	data	—

A variável *data\_ultima\_leitura* indica a data em que ocorreu o último registro de leitura da UC, *leitura\_60*. O arquivo deve conter as informações de consumo dos últimos cinco anos da UC, terminando no mês apresentado pela data de última leitura. A figura 4.3 apresenta a primeira tela do assistente de importação dos dados de parque e consumo.



**Figura 4.3 – Assistente de importação de dados de parque e consumo.**

O programa permite importar vários dados de parque e consumo distintos. Com essa característica é possível caminhar historicamente pelas várias situações de parque de medidores instalados no aplicativo. Este originalmente contém a situação do parque de medidores em julho de 2008, período em que foram recebidos os dados.

Para que o usuário final tenha um controle sobre cada importação realizada, uma tela semelhante a da figura 4.4 é apresentada ao usuário antes de iniciar a importação dos dados. O preenchimento dos campos permite identificar qual o nome do usuário que realizou a importação, a data de importação, o período de referência do parque de medidores e pode-se adicionar um nome a este conjunto de dados.

**Wizard - Importação de Dados**

**Assistente de Importação de Dados**

Informações adicionais.  
Antes de realizar a importação é necessário preencher algumas informações. Por favor, informe o nome da pessoa responsável pela importação, o nome dado à importação e o período (mês/ano) a que os dados se referem.

Nome do responsável

Nome da atualização

Data de referência

Mês:

Ano:

Sair Voltar Avançar

Figura 4.4 – Tela de informações adicionais sobre a importação.

A última tela do assistente contém uma barra de progresso da importação informando em qual etapa da importação o aplicativo está. Durante a importação são realizadas as filtragens dos dados e obtenção dos demais campos necessários. O processo de obtenção desses dados é semelhante ao apresentado na seção de integração de dados do capítulo 2, em especial a figura 2.5. O último estágio do assistente de importação dos dados com informações do parque e consumo é apresentado na figura 4.5.

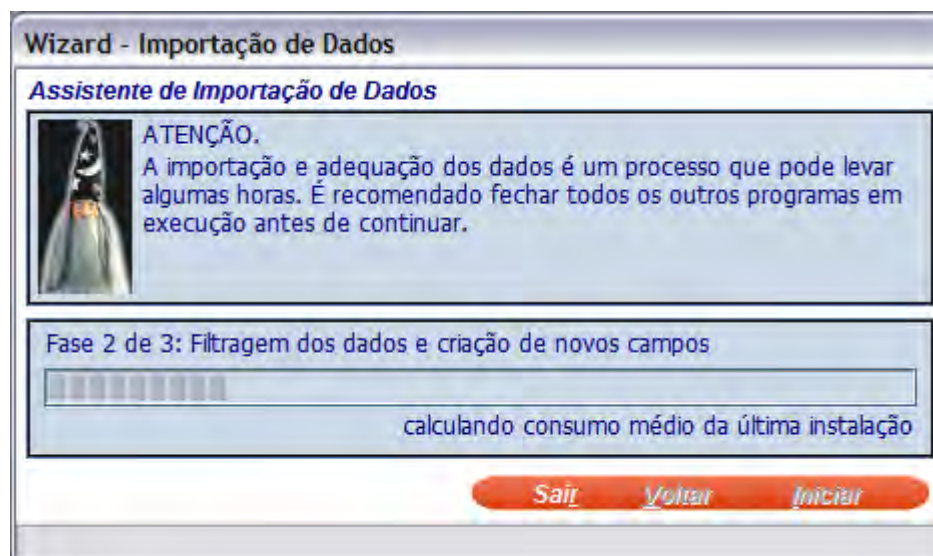


Figura 4.5 – Último estágio do assistente de importação dos registros com dados de parque e consumo.

## 4.2.2 Importação dos dados de ensaio

A partir do item *Arquivo Ensaio* do menu *Importação* o aplicativo possibilita a importação de dados relacionados com ensaios de teste de falha de operação de medidores.

Os dados são importados a partir de um arquivo no formato de planilha Excel (.xls). O arquivo deve obedecer a uma padronização de formatação específica, conforme exemplo apresentado na tabela 4.3: a primeira coluna deve conter o número de identificação do equipamento (*num Equip*); na segunda coluna deve ser apresentado o resultado do ensaio dielétrico; as terceira, quarta e quinta colunas devem apresentar o erro percentual obtido para os ensaios nominal, carga indutiva e carga pequena, respectivamente; a sexta coluna deve conter a resposta de falha do medidor, valor "0" indica que não existe falha no equipamento e valor "1" indica que existe falha.

Tabela 4.3 – Exemplo de planilha para importação de dados de ensaio.

Número do equipamento	Teste dielétrico	Carga nominal	Carga indutiva	Carga pequena	Falha
M02139380	APROVADO	0,844	0,497	0,053	0
M0223662	APROVADO	-1,888	-7,857	-125,900	1
M0221813	APROVADO	1,555	0,488	4,386	1
M0238445	REPROVADO				1

O preenchimento manual da sexta coluna permite ao usuário modificar a forma como é definida a falha dos medidores. Para este trabalho, o medidor recebe a notação de falha quando:

- O equipamento foi reprovado no teste dielétrico;
- O módulo do erro percentual de algum ensaio com carga foi superior a 3%.

O processo de importação armazena algumas informações resumidas sobre os medidores, como subtipo de equipamento importado e quantidade de registros. Estas informações não alteram o funcionamento da metodologia, tendo característica apenas descritiva. Para que o aplicativo registre corretamente as descrições dos dados por subtipo, deve-se utilizar o recurso de abas do Excel. Cada aba deve conter registros referentes a um subtipo, cujo nome deve estar presente na aba correspondente, conforme apresentado na figura 4.6. A figura 4.7 apresenta a primeira tela do assistente de importação de dados de ensaio.

345	3T1117467	APROVADO	-1,155	0,400	0,484	0
346	3T1096265	APROVADO	0,182	-0,697	-5,219	1
347	3T110651	APROVADO	0,588	1,322	1,233	0
348	3T1123335	APROVADO	1,864	2,402	0,715	0
349	3T1139282	APROVADO	1,620	2,091	0,928	0
350	3T1092790	APROVADO	-0,017	1,300	1,608	0
351	3T1124854	APROVADO	-47,180	-45,920	-53,790	1
352	3T1138804	APROVADO	2,124	2,644	1,448	0
353	3T112029	APROVADO	3,911	3,853	2,960	1
354	3T112027	APROVADO	0,284	0,924	1,833	0
355	3T109815	APROVADO	-0,464	1,408	1,757	0
356	3T110530	APROVADO	-28,640	-43,050	-89,290	1
357	3T1120948	APROVADO	0,697	1,948	0,588	0
358	3T113617	APROVADO	-0,282	-0,117	2,615	0
359	3T112635	APROVADO	2,424	3,651	3,415	1
360	3T112104	APROVADO	0,628	1,155	-0,224	0

Figura 4.6 – Exemplo de arquivo de ensaio com preenchimento de abas.

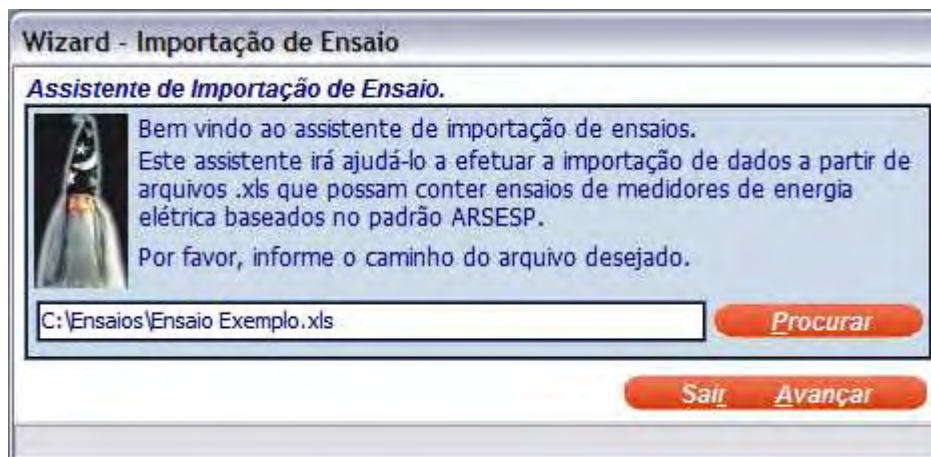


Figura 4.7 – Assistente de importação de dados de ensaio.

O programa armazena todas as importações de dados de ensaio em disco, permitindo que vários modelos de probabilidade de falha sejam realizados rapidamente. Para que haja um controle por parte do usuário sobre os ensaios, os mesmos campos da figura 4.4 são apresentados pelo assistente de importação de ensaios.

Para a utilização dos dados de ensaio para ajuste da probabilidade de falha dos grupos de medidores são necessárias as informações das demais categorias da metodologia do capítulo 3. Essas informações não estão presentes nos arquivos de ensaio sendo necessário realizar a interligação dos registros de ensaio com os demais dados da metodologia, *i.e.*, a interligação com os registros de situação de instalação e de consumo presentes na base de dados de parque e consumo. Para que este processo de importação possua uma melhor eficiência, estas duas bases de dados devem possuir a mesma referência de tempo ou a mais próxima possível. Durante a importação dos dados de ensaio é questionado ao usuário qual a base de dados de parque e consumo existente no aplicativo que deve ser utilizada para obtenção dos demais campos necessários. O nome do banco de dados de parque e consumo que acompanha o aplicativo é "Base Inicial". A figura 4.8 apresenta a tela de escolha da base de parque e consumo durante a importação dos dados de ensaio.

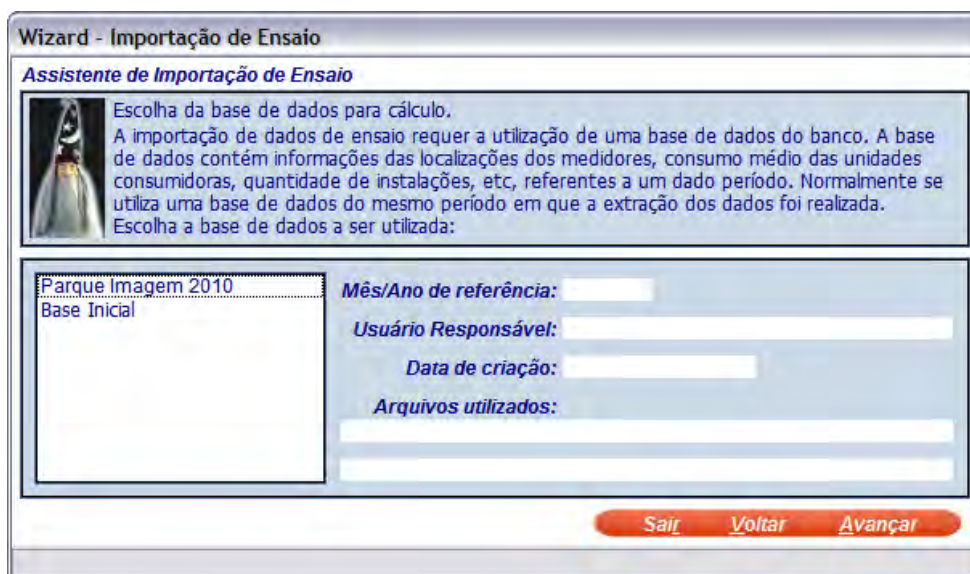


Figura 4.8 – Escolha da base de dados de parque e consumo durante a importação de dados de ensaio.

Por fim, os dados importados nos arquivos de ensaios poderão ser utilizados para ajustar os valores da probabilidade de falha, utilizando o assistente de Inteligência Artificial através do item Criar novo modelo utilizando IA do menu Análise de Probabilidade de Falhas. A figura 4.9 apresenta o último estágio da importação dos dados de ensaios.

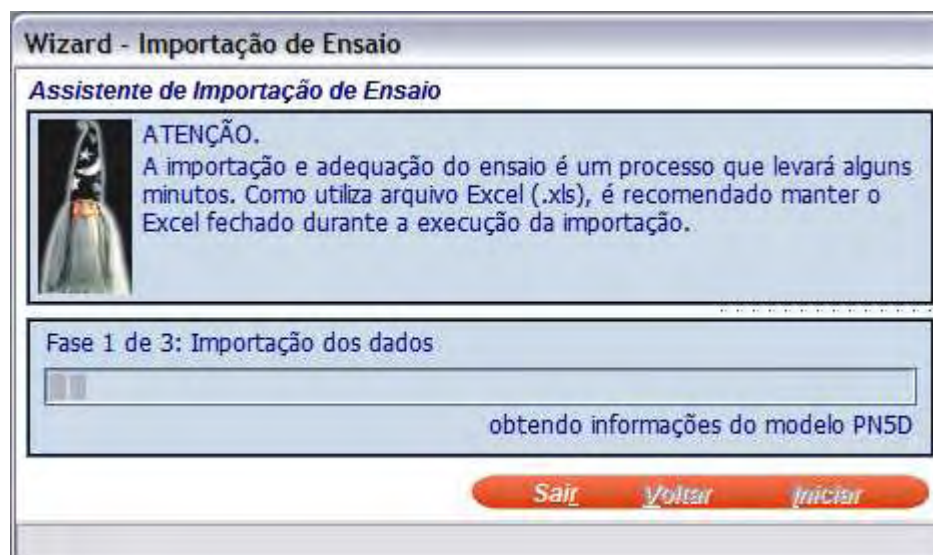
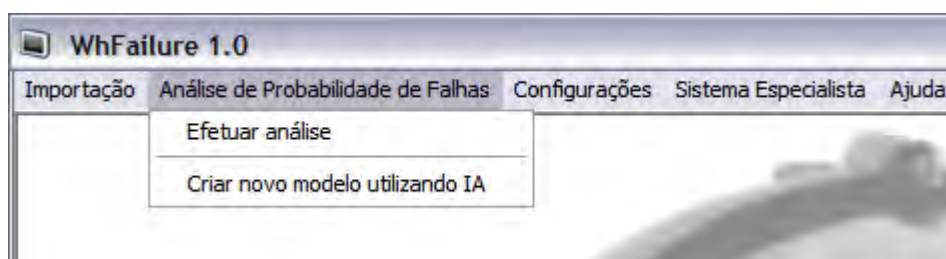


Figura 4.9 – Último estágio da importação de dados de ensaio.

### 4.3 Análise de probabilidade de falhas

O aplicativo apresenta uma interface para a realização de consultas de probabilidade de falhas dos equipamentos. A consulta é realizada através de uma configuração de caso através da seleção de um conjunto de modelos de medidores, regionais, consumo e escala de probabilidade de falha. A consulta é acessada pelo item *Efetuar análise* do menu *Análise de Probabilidade de Falhas*, conforme apresentado na figura 4.10.



**Figura 4.10 – Menu de análise de probabilidade de falhas.**

Outra funcionalidade acessada por meio deste menu é a criação de novos modelos de probabilidade de falhas, utilizando a metodologia de correção da probabilidade de falhas apresentada no capítulo anterior.

#### 4.3.1 Efetuar análise

Nesta interface é possível observar a estratificação da probabilidade de falhas do resultado da análise através de histogramas, tabelas e, utilizando recursos gráficos de plotagem, a situação dos medidores de energia agrupados por localidade em um mapa. Além disso, é possível a emissão de relatórios das análises efetuadas no formato de planilhas eletrônicas do Excel.

A Figura 4.11 apresenta a visão geral da interface de análise de probabilidade de falhas.

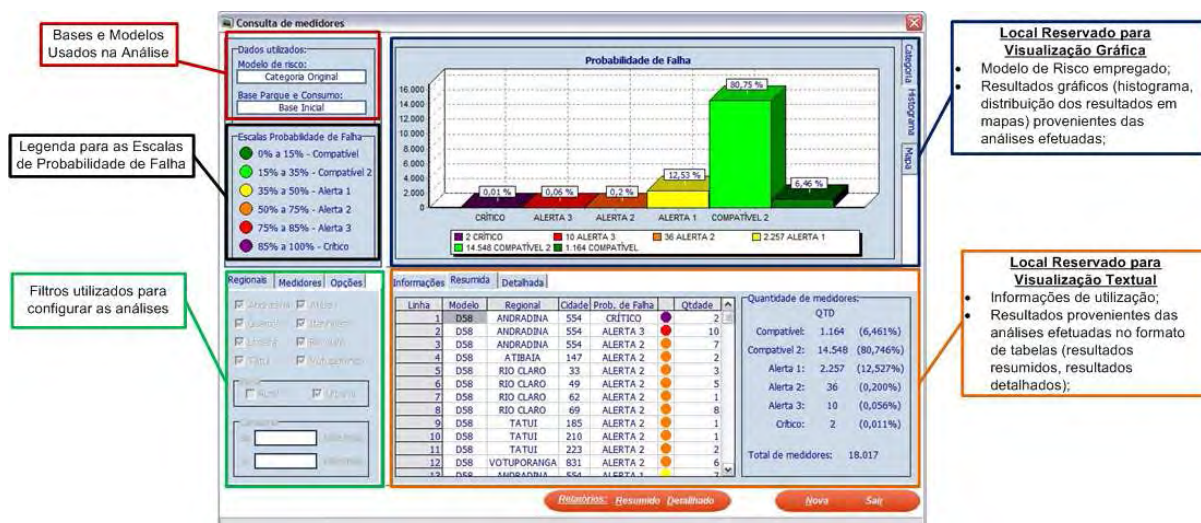


Figura 4.11 – Interface de análise de probabilidade de falhas.

A interface é dividida basicamente em cinco regiões:

- Bases e modelos utilizados na análise: apresenta o nome da importação de dados de parque e consumo utilizado para a análise, bem como o nome da atualização do modelo que descreve o comportamento da probabilidade de falha dos equipamentos;
- Legendas para as escalas de probabilidade de falhas;
- Filtros: devem ser preenchidos antes da consulta. Apresentam os critérios de consulta de equipamentos. Pode-se utilizar como critérios as regionais da concessionária de energia, o modelo agrupado do equipamento, definir faixas de consumo médio mensal, limitar a busca por zona rural ou urbana, limitar a busca por número de fases dos equipamentos e limitar os resultados por escalas de probabilidade de falhas;
- Visualização gráfica: apresenta o resultado da consulta dos equipamentos utilizando histogramas e mapas;
- Visualização textual: apresenta o resultado da consulta dos equipamentos em forma de tabelas;

A tabela 4.4 apresenta as escalas de probabilidade de falhas utilizadas no aplicativo.

Tabela 4.4 – Escalas de probabilidade de falhas utilizada no aplicativo.

Faixa de probabilidade de falha	Nome	Cor
0% a 15%	Compatível	Verde escuro
15% a 35%	Compatível 2	Verde claro
35% a 50%	Alerta 1	Amarelo
50% a 75%	Alerta 2	Laranja
75% a 85%	Alerta 3	Vermelho
85% a 100%	Crítico	Roxo

A seção de filtros é definida como a área ativa da tela, pois o usuário é obrigado a interagir com ela para realizar a consulta. O usuário deve escolher, pelo menos, quais modelos de medidores devem ser avaliados. A figura 4.12 apresenta uma vista explodida da seção de filtros.

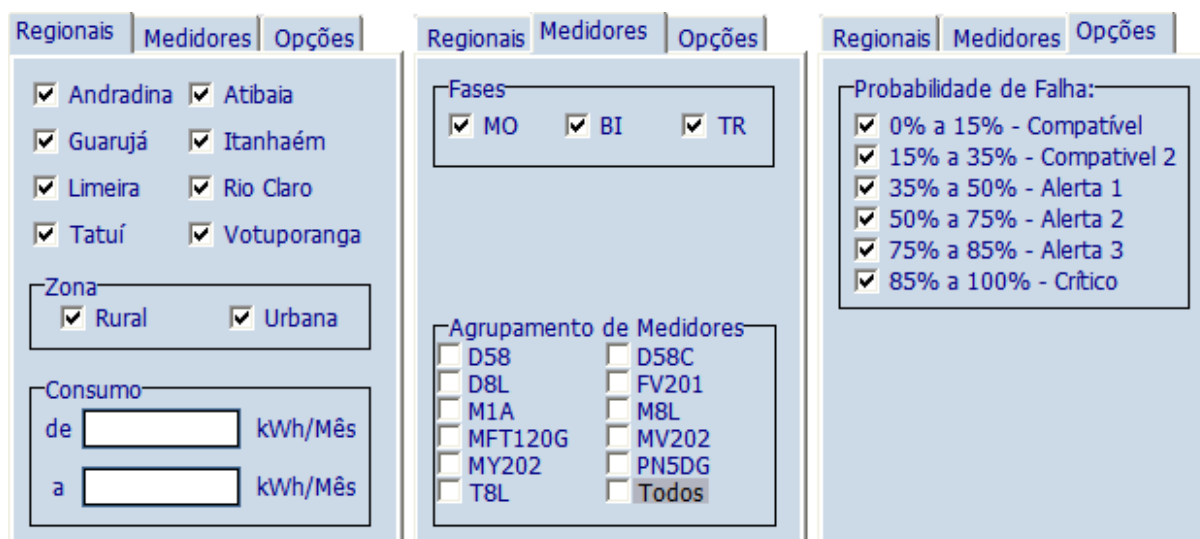


Figura 4.12 – Abas da seção de filtros da interface de análise de probabilidade de falhas.

A informação de consumo só deve ser preenchida caso se deseje limitar uma quantidade mínima e/ou máxima de consumo médio mensal.

Uma vez realizada a consulta, a área gráfica apresentará um histograma da distribuição dos medidores nas escalas de probabilidade de falhas e um mapa do parque da concessionária de energia, com informações sobre a quantidade de medidores em cada escala de probabilidade de falhas por cada localidade. As figuras 4.13 e 4.14 apresentam, respectivamente, o histograma e o mapa da interface de análise de probabilidade de falhas.

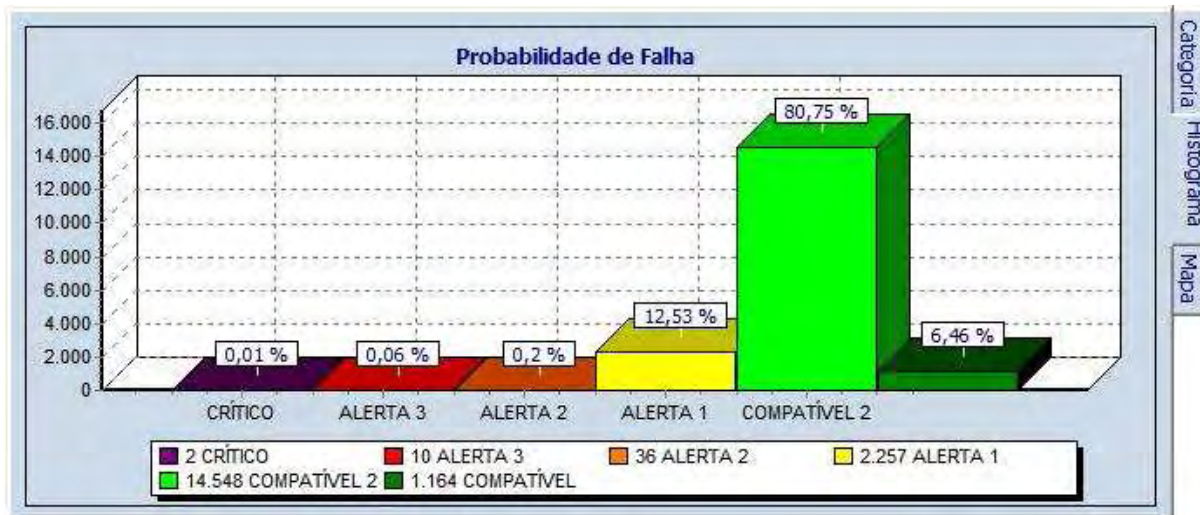


Figura 4.13 – Histograma da interface de análise de probabilidade de falhas.

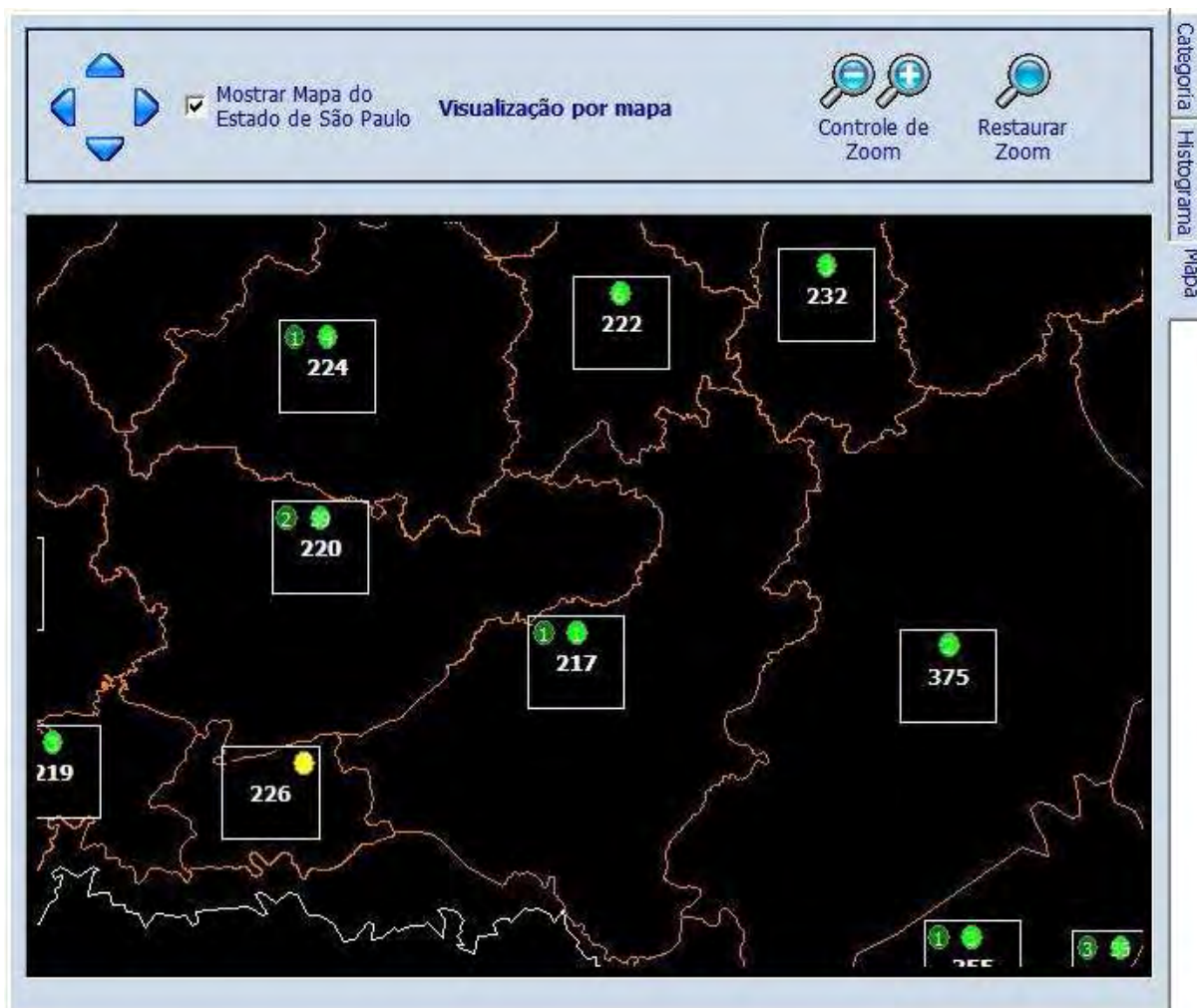


Figura 4.14 – Mapa de distribuição da probabilidade de falhas em cada escala por localidade.

A área textual apresentará dois tipos de tabelas, uma com dados resumidos e outra com dados detalhados. A tabela com dados resumidos apresentará um agrupamento de

informações, apresentando o modelo agrupado, a localidade, a escala de probabilidade e como última informação a quantidade de equipamentos. A figura 4.15 apresenta uma imagem da tabela resumida. Ao lado desta tabela existe um informativo sobre a distribuição dos equipamentos nas escalas de probabilidade de falhas.

A tabela detalhada apresenta a informação por equipamento. A tabela contém a identificação do equipamento, modelo agrupado, quantidade de fases, tipo de local, localidade, quantidade de instalações do equipamento, consumo médio mensal e escala de probabilidade de falha. A figura 4.16 apresenta uma imagem desta tabela. Como a quantidade de resultados pode ser muito elevada, limitou-se a quantidade de registros para 1000, adicionando-se uma barra de navegação superior.

Linha	Modelo	Regional	Cidade	Prob. de Falha	Qtidade
1	D58	ANDRADINA	554	CRÍTICO	2
2	D58	ANDRADINA	554	ALERTA 3	10
3	D58	ANDRADINA	554	ALERTA 2	7
4	D58	ATIBAIA	147	ALERTA 2	2
5	D58	RIO CLARO	33	ALERTA 2	3
6	D58	RIO CLARO	49	ALERTA 2	5
7	D58	RIO CLARO	62	ALERTA 2	1
8	D58	RIO CLARO	69	ALERTA 2	8
9	D58	TATUI	185	ALERTA 2	1
10	D58	TATUI	210	ALERTA 2	1
11	D58	TATUI	223	ALERTA 2	2
12	D58	VOTUPORANGA	831	ALERTA 2	6
13	D58	ANDRADINA	554	ALERTA 1	7

Quantidade de medidores: QTD

Compatível: 1.164 (6,461%)

Compatível 2: 14.548 (80,746%)

Alerta 1: 2.257 (12,527%)

Alerta 2: 36 (0,200%)

Alerta 3: 10 (0,056%)

Crítico: 2 (0,011%)

Total de medidores: 18.017

Figura 4.15 – Tabela resumida para análise da probabilidade de falhas.

Linha	Número	Modelo	Fases	Zona	Regional	Cidade	Instalacoes	Média kWh	Prob. de Falha
1	3T065765	D58	3	UB	ANDRADINA	554	1	705,6	CRÍTICO
2	3T078133	D58	3	UB	ANDRADINA	554	1	760,9	CRÍTICO
3	3T064280	D58	3	UB	ANDRADINA	554	1	502,3	ALERTA 3
4	3T064284	D58	3	UB	ANDRADINA	554	1	539,4	ALERTA 3
5	3T065714	D58	3	UB	ANDRADINA	554	1	347,4	ALERTA 3
6	3T065743	D58	3	UB	ANDRADINA	554	1	3168,1	ALERTA 3
7	3T065746	D58	3	UB	ANDRADINA	554	1	1244,5	ALERTA 3
8	3T065799	D58	3	UB	ANDRADINA	554	1	1894,7	ALERTA 3
9	3T066337	D58	3	UB	ANDRADINA	554	1	211,0	ALERTA 3
10	3T066339	D58	3	UB	ANDRADINA	554	1	448,3	ALERTA 3
11	3T066340	D58	3	UB	ANDRADINA	554	1	297,8	ALERTA 3
12	3T070989	D58	3	UB	ANDRADINA	554	1	956,0	ALERTA 3
13	3T000422	D58	2	UB	RIO CLARO	60	1	648,2	ALERTA 2

Figura 4.16 – Tabela detalhada para análise da probabilidade de falhas.

A interface de análise de probabilidade de falhas possibilita a emissão de dois tipos de relatórios após a execução da consulta, um relatório resumido e um relatório detalhado. Basicamente, esses relatórios são as tabelas de mesmo nome apresentadas na interface. Eles são gerados no programa Excel, e contêm um cabeçalho com os critérios de pesquisa para a obtenção das análises (configuração dos filtros). O relatório resumido apresenta exatamente as informações da tabela resumida, já o relatório detalhado apresenta todas as informações da tabela detalhada acrescida da informação das datas da primeira instalação e da última instalação do equipamento. Na figura 4.17 se pode observar o cabeçalho de um relatório resumido.


 <b>Relatório de Pesquisa de Medidores</b>					
Critérios de Pesquisa					
Base de Dados: Base Inicial			Modelo de Risco: Categoria Original		
Fases:	Monofásico, Bifásico, Trifásico		Consumo:	de	—
Zonas:	Rural e Urbana			até	—
Regionais: Andradina, Atibaia, Guarujá, Itanhaém, Limeira, Rio Claro, Tatuí, Votuporanga.					
Modelos: TBL.					
Probabilidade de Falha: Compatível, Compatível 2, Alerta 1, Alerta 2, Alerta 3, Crítico.					
Resultados da Pesquisa					
Modelo	Regional	Cidade	Risco	Escala de Risco	Quantidade
TBL	ANDRADINA	654	CRÍTICO		2
TBL	ANDRADINA	654	ALERTA 2		4
TBL	VOTUPORANGA	867	ALERTA 2		2
TBL	ANDRADINA	554	ALERTA 1		16
TBL	ANDRADINA	665	ALERTA 1		1

Figura 4.17 – Cabeçalho de um relatório resumido.

### 4.3.2 Criar novo modelo utilizando IA

O WhFailure permite a criação de novos modelos de probabilidade de falhas atualizando a probabilidade de falha de modelos anteriores com os dados obtidos de algum arquivo de ensaio. A interface de criação de novos modelos permite salvar o novo modelo

gerado, dando-lhe um nome específico, ou apenas visualizar a alteração da probabilidade de falha com os novos dados.

O processo de criação de um novo modelo é efetuado através de um assistente acessado pelo item Criar novo modelo utilizando IA do menu Análise de probabilidade de falhas.

A Figura 4.18 apresenta a tela inicial do assistente. Neste primeiro passo de geração de um novo modelo de probabilidade de falhas, também denominado modelo de risco, deve-se selecionar dentro das bases de dados disponíveis o modelo de risco a partir do qual será realizado o ajuste das probabilidades de falhas, e o conjunto de dados de ensaios a ser utilizado para a regressão. O modelo original do aplicativo é denominado "Categoria Original". Na tela são apresentadas várias informações sobre a criação dos dados, como o nome do usuário, a data de criação, o período de referência dos dados e outras informações específicas: para os modelos de risco as informações dos modelos anteriores e ensaios utilizados; para os dados de ensaio as informações sobre a localização do arquivo de importação e a quantidade de dados importados ordenados por modelo.

**Wizard - Inteligência Artificial**

**Assistente de Geração de Inteligência Artificial**

Bem-vindo ao assistente de Inteligência Artificial.  
 Este assistente permite atualizar a probabilidade de falha de um modelo de risco de medidores utilizando-se um banco de dados de ensaios no padrão ARSESP, com auxílio de técnicas de inteligência artificial.  
 O assistente também permite simular a evolução do modelo de risco.  
 Para iniciar, escolha um modelo de risco e um banco de dados de ensaios.

**Modelos de risco disponíveis:**

Categoria Original

Mês/Ano de referência: 07/2008

Usuário Responsável: \_\_\_\_\_

Data de criação: \_\_\_\_\_

Base de dados: Dados internos do programa

Ensaio utilizado: Dados internos do programa

**Banco de dados de ensaios disponíveis:**

Ensaio Teste

Mês/Ano de referência: 03/2010

Usuário Responsável: Engenheiro F.

Data de criação: 26/04/2010 17:56:27

Arquivos utilizados:  
 D:\Arquivos de Exemplo\Ensaio Exemplo.xls  
 D58C, 382 | MFB-120G, 431 | PN5D, 372 | FX221, 420 | D8L, 415 |

Sair Avançar

Figura 4.18 – Assistente de geração de modelo de probabilidade de falhas.

Após a seleção das bases de dados desejadas, o assistente apresenta uma interface de ajuste de parâmetros da regressão e um gráfico com a resposta do modelo, conforme figura 4.19. No gráfico, os valores em verde representam a resposta do modelo de risco utilizado, os valores em vermelho representam as probabilidades de falhas obtidas a partir dos dados de ensaio e a curva em azul representa a curva que descreve as novas probabilidades de falhas.

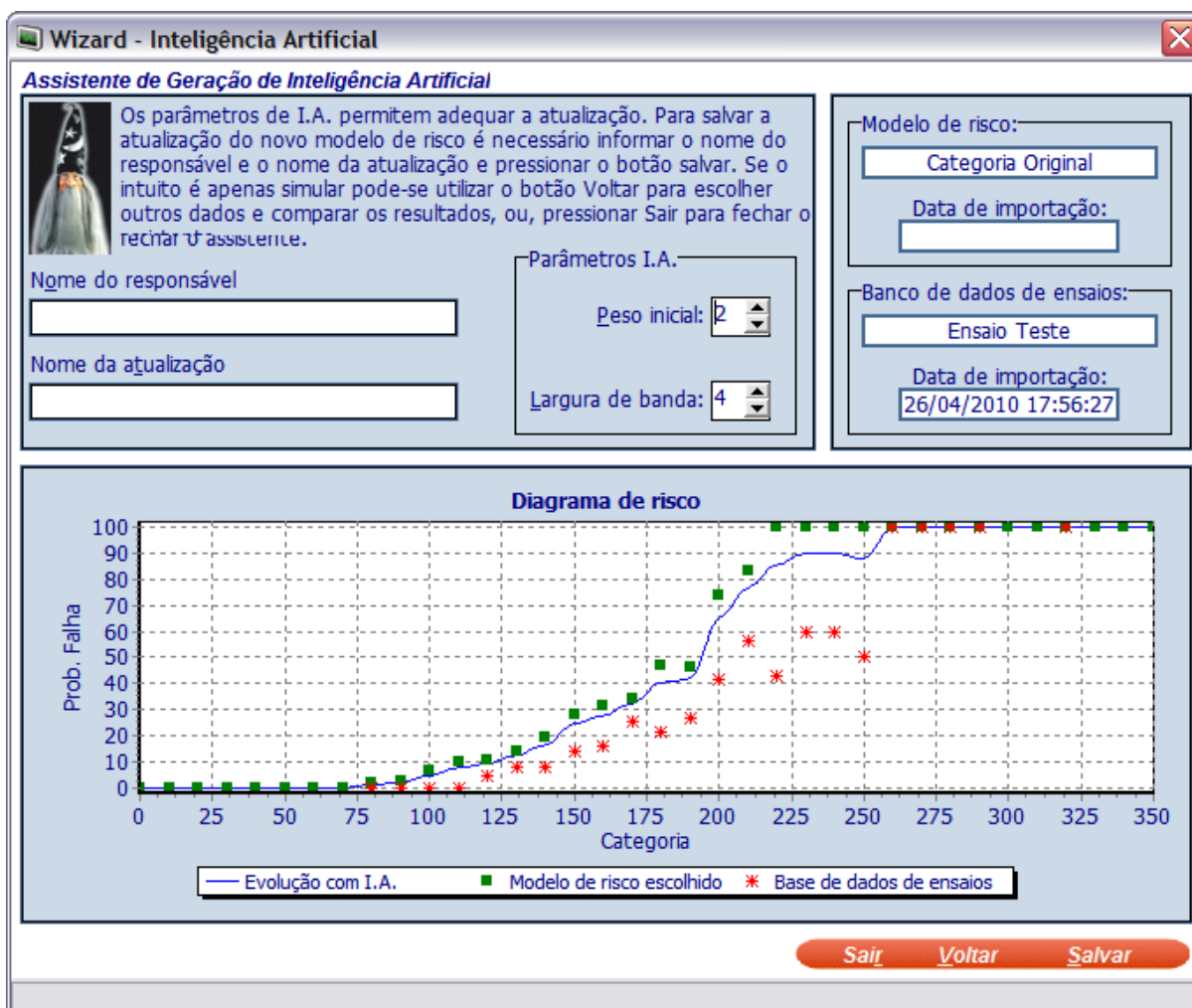
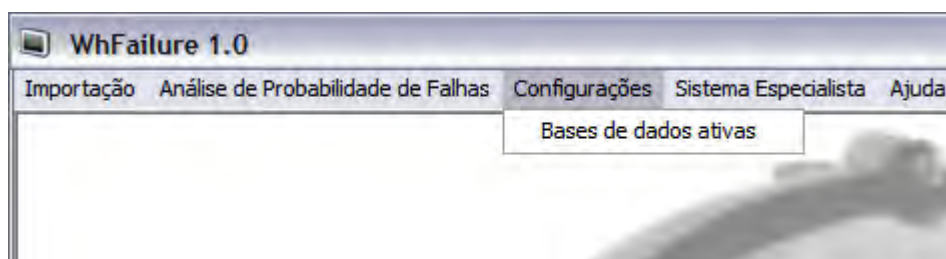


Figura 4.19 – Visualização do ajuste das probabilidades de falhas para os novos dados de ensaio.

Além de possibilitar a simulação da curva de probabilidade de falhas para diversos valores de peso inicial e largura de banda é possível salvar a resposta da regressão como um novo modelo, preenchendo-se as informações de usuário responsável pela criação e acrescentando um nome para o modelo.

## 4.4 Configurações

A interface de configurações possibilita selecionar qual a base de dados obtida pela importação dos arquivos de parque e consumo será utilizada para análises e amostragens, além da seleção de um modelo de probabilidade de falha dentre o modelo inicial e os novos modelos criados através do assistente de criação de novos modelos. A interface é acessada através do item `Bases de dados ativas` do menu `Configurações`, conforme figura 4.20.



**Figura 4.20: Menu de configurações.**

O aplicativo WhFailure permite a existência de diferentes configurações de dados registradas para as análises, entretanto apenas um conjunto de dados de parque e consumo e um modelo de probabilidade de falha são considerados ativos para efetuar as análises e amostragens. Este fato possibilita um grau de flexibilidade elevado, permitindo a utilização de diversas probabilidades de falhas para analisar um mesmo conjunto de dados, ou vice-versa. A figura 4.21 apresenta a interface de configuração do aplicativo.

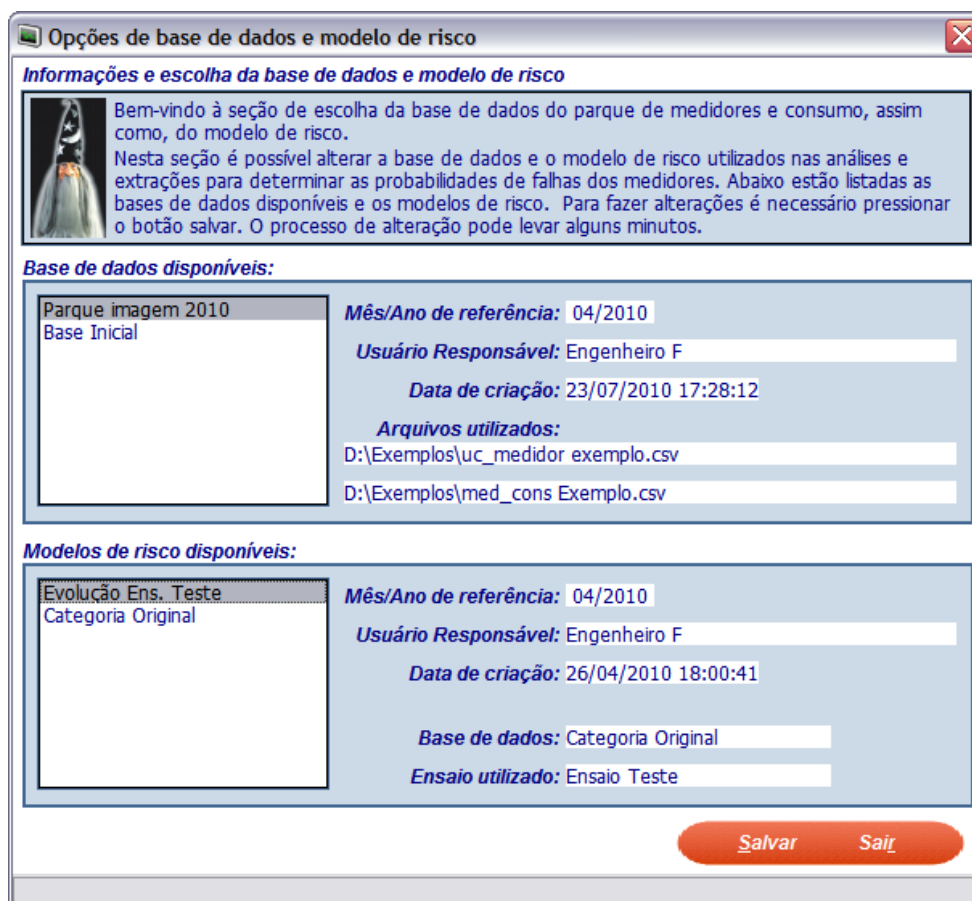


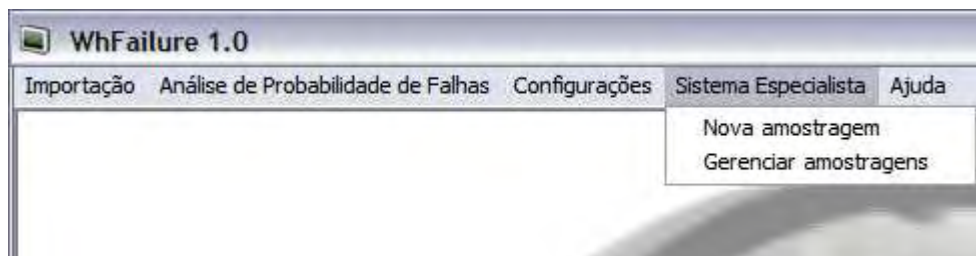
Figura 4.21 – Interface de configuração da imagem do parque e modelo de probabilidade de falhas selecionados.

A interface apresenta as informações sobre as bases de dados e modelos, como usuário responsável, data de criação ou importação, etc. Destes, talvez o campo mais importante é o período de referência, que deve ser igual ou próximo para ambas as seleções, de modo a aumentar a eficiência do algoritmo.

#### 4.5 Sistema especialista para amostragens

O aplicativo apresenta uma interface para a realização de amostragens de medidores, selecionando aqueles que apresentam maior probabilidade de falha. A ferramenta emite relatórios em formato de planilha eletrônica do Excel que apresentam informações sobre a localização dos equipamentos selecionados. A interface para escolha da quantidade de equipamentos a serem amostrados e criação dos resultados é acessada pelo item *Nova amostragem* do menu *Sistema Especialista*. As amostragens previamente realizadas são

salvas no aplicativo para futuras reimpressões e são disponibilizadas através do item Gerenciar amostragens no mesmo menu, conforme figura 4.22.



**Figura 4.22 – Menu do sistema especialista para amostragens.**

A interface para realização de amostragens de medidores de energia elétrica é apresentada na figura 4.23. A amostragem é realizada sobre o banco de dados de parque e consumo selecionado ativo na interface de configuração utilizando o modelo de probabilidade de falha também selecionado como ativo na mesma interface.

O lado direito da interface apresenta duas tabelas. Elas possuem três colunas contendo, respectivamente, os grupos finais em utilização, a quantidade de equipamentos pertencente ao grupo, e o valor da probabilidade de falha associado ao grupo. A tabela no extremo direito apresenta os dados para o modelo selecionado na interface de configuração e a tabela à esquerda desta apresenta os dados para o modelo que acompanha o aplicativo. Assim, é possível visualizar a diferença entre a probabilidade de falha atual e a probabilidade de falha original. A parte superior esquerda apresenta campos cujo preenchimento são obrigatórios para geração das amostragens (nome da amostragem e o usuário responsável) além de apresentar informações sobre os dados e modelo utilizados.

**Sistema Especialista**

**Sistema Especialista para amostragem de medidores**

Bem-vindo ao sistema especialista de amostragem. Nesta seção é possível amostrar um conjunto de medidores em função de suas probabilidades de falhas, simular e calcular a acurácia esperada para a amostragem e gerar relatórios com os dados dos equipamentos amostrados pelo sistema especialista.

Nome do responsável:

Nome da amostragem:

Base de dados:

Distribuição da Amostragem

Total a amostrar:

Amostra dirigida:

Amostra aleatória:


Simulação da acurácia esperada

Na Amostra:

Parte Dirigida:

Parte Aleatória:

Dirigida (Modelo Original):



Dados com Modelo Original			Dados com Modelo Atualizado		
Cat.	Qtd	Risco	Cat.	Qtd	Risco
50	2	0	50	2	0
60	1445	0	60	1401	0
70	24749	0	70	23540	0,1
80	135364	1,9	80	131079	1,4
90	269932	2,8	90	268966	2,2
100	403634	6,4	100	400909	4,8
110	367411	10,4	110	371623	7,7
120	279243	10,9	120	279940	9,3
130	197612	14	130	199294	12,6
140	115833	19,2	140	117455	16,7
150	57043	28,4	150	57497	24,5
160	24352	31,8	160	24775	27,9
170	8750	34,5	170	8816	32,4
180	3949	47,2	180	4004	40,4
190	1482	46,2	190	1473	42,4
200	720	73,7	200	740	65,2
210	725	83,3	210	724	76,5
220	713	100	220	712	85,5

**Figura 4.23 – Interface de amostragem de equipamentos.**

Na parte inferior esquerda da interface existem três campos que determinam a quantidade de equipamentos que devem compor o relatório. O primeiro campo, *total a amostrar*, define o tamanho da amostra, o segundo campo, *amostra dirigida*, define quantos equipamentos serão selecionados considerando a probabilidade de falhas associada, o terceiro campo, *amostra aleatória*, determina quantos equipamentos devem ser selecionados aleatoriamente. A seleção aleatória é importante para a metodologia, pois esses dados devem ser inseridos novamente no aplicativo após ensaio, atualizando as probabilidades de falhas. A quantidade de equipamentos de amostra dirigida deve ser superior a 50% e inferior a 80% da quantidade total de equipamentos, garantindo que sempre exista uma parcela de dados de ensaio obtidos de forma aleatória para realimentação do aplicativo e melhoria da probabilidade de falha do modelo.

Uma funcionalidade importante da ferramenta é a possibilidade de simular a acurácia esperada de uma amostragem, *i.e.*, a quantidade percentual de equipamentos que devem apresentar falha. Na figura 4.24 é apresentado um exemplo de simulação da acurácia esperada. A acurácia é apresentada em três partes: a acurácia da amostra, a acurácia da parte dirigida e a acurácia da parte aleatória. Por último é apresentada a acurácia esperada da parte dirigida utilizando os valores originais de probabilidade de falhas.

Distribuição da Amostragem	Simulação da acurvidade esperada
Total a amostrar: <input type="text" value="5000"/>	Na Amostra: 57,87%
Amostra dirigida: <input type="text" value="4000"/>	Parte Dirigida: 70,32%
Amostra aleatória: <input type="text" value="1000"/>	Parte Aleatória: 8,08%
	Dirigida (Modelo Original): 78,80%




Figura 4.24 – Exemplo de simulação da acurvidade esperada.

Uma vez preenchidos os campos obrigatórios para a amostragem, e ao término do processo de amostragem o relatório estará disponível para impressão no item Gerenciar amostragens do menu Sistema Especialista. Também será questionado ao usuário, após o processo de cálculo, se ele deseja emitir o relatório.

A interface de gerenciamento de amostragens apresentada na figura 4.25 contém uma lista com os nomes de todas as amostragens já realizadas. Uma vez selecionada a amostragem é possível visualizar informações sobre o usuário responsável, data de criação e informações sobre o tamanho da amostra e acurvidade esperada.



**Gerenciamento de amostragens**

Bem-vindo a seção de gerenciamento de amostragens.  
Nesta janela é possível verificar todas as amostragens realizadas. Ao lado são apresentadas informações sobre o usuário que realizou a amostra, data em que foi realizada, o banco de dados utilizado e o modelo de risco. Também são apresentadas informações do tamanho da amostra e da acurvidade esperada.  
Se desejar gerar novamente algum relatório, escolha a amostra e pressione o botão Relatório.

**Amostragens realizadas:**

Amostragem MAR-2010	<b>Usuário Responsável:</b> Engenheiro A
	<b>Data de criação:</b> 27/04/2010 08:21:16
	<b>Parque Utilizado:</b> Base Inicial
	<b>Modelo de risco:</b> Categoria Original
	<b>Estratificação:</b>
	<b>Acurvidade esperada:</b>
<b>Total:</b> 5.000	<b>Total:</b> 64,96%
<b>Dirigida:</b> 4.000	<b>Dirigida:</b> 78,78%
<b>Aleatória:</b> 1.000	<b>Aleatória:</b> 9,69%

**Relatório** **Sair**

Figura 4.25 – Interface de gerenciamento de amostragens.

O relatório é emitido utilizando uma planilha eletrônica do Excel. No cabeçalho são apresentadas todas as informações sobre a amostragem. Os campos do relatório apresentam o número de identificação do equipamento, o modelo agrupado, a quantidade de fases, o tipo de local, o município e a unidade consumidora em que está instalado. A figura 4.26 apresenta um exemplo de relatório de amostragem.


 <b>Relatório de Amostragem de Medidores</b>						
<b>Base de dados:</b> Base Inicial			<b>Modelo de Risco:</b> Categoria Original			
Informações da amostra						
<b>Tamanho da extração:</b> 5.000			<b>Amostra dirigida:</b>		4.000	
			<b>Amostra aleatória:</b>		1.000	
<b>Acertividade esperada da extração:</b> 64,96%			<b>Somente parte dirigida:</b>		78,78%	
			<b>Somente parte aleatória:</b>		9,69%	
Resultados da Pesquisa						
Número Equipamento	Modelo	Fases	Localização	Regional	Cidade	Unidade Consumidora
3T0113182	D58	3	RR	ANDRADINA	654	8061904
3T012119	D58	3	RR	ANDRADINA	654	8062234
3T018213	D58	3	RR	ANDRADINA	654	8064230

Figura 4.26 – Exemplo de relatório de amostragem de medidores.

## 4.6 Considerações finais

O aplicativo apresentado neste capítulo permite aplicar facilmente a metodologia do capítulo 3 em um conjunto de medidores. As diversas ferramentas, como importação de dados de parque e consumo de medidores, visualização de resultados, emissão de relatórios e sistema especialista, permitem uma facilidade para aplicação da metodologia sobre qualquer conjunto de dados desejado, gerando relatórios para reuniões com os equipamentos mais indicados para aferição, de acordo com a probabilidade de falha.

A interface de análise de probabilidade de falha permite a visualização dos estados dos equipamentos, permitindo analisar um tipo específico de equipamento, uma gama de equipamentos que obedecem a determinados critérios de seleção ou mesmo todo o parque de medidores. Com a interface de atualização da probabilidade de falhas, novos dados de ensaio são facilmente incorporados ao modelo, ajustando a probabilidade de falhas.

O aplicativo está voltado para utilização pela concessionária de energia, a qual é a portadora das informações de parque e consumo.

## 5 ESTUDO DE CASO

---

Neste capítulo é realizado um estudo de caso do parque da concessionária utilizando a metodologia desenvolvida. É analisada a probabilidade de falha dos medidores instalados em unidades consumidoras monofásicas. A justificativa da escolha desse tipo de medidor é apresentada a seguir.

### 5.1 Introdução

Os medidores da concessionária de energia elétrica destinados à medição de consumo dos clientes podem ser divididos em quatro grupos, de acordo com o tipo de medidor e tipo de fases: eletrônicos, eletromecânicos monofásicos, eletromecânicos bifásicos e eletromecânicos trifásicos.

As unidades consumidoras monofásicas representam grande parte dos clientes da concessionária de energia elétrica deste estudo de caso. A tabela 5.1 apresenta a distribuição da quantidade de equipamentos do parque por tipo de equipamento de medição e a quantidade percentual de energia aferida por cada grupo no ano de 2005.

**Tabela 5.1 – Distribuição de medidores no parque da concessionária de acordo com o tipo de medidor e número de fases, ano 2005.**

<b>Tipo de medidor</b>	<b>Quantidade no parque</b>	<b>Percentual de consumo total</b>
Eletrônico	8.132 (0,43%)	55,39%
Indução monofásico	849.046 (45,07%)	11,27%
Indução bifásico	790.211 (41,95%)	14,20%
Indução trifásico	227.488 (12,08%)	19,14%

A quantidade elevada de consumidores monofásicos torna mais difícil a determinação do estado operacional destes medidores instalados nestas unidades via metodologia convencional. Estas unidades apresentam um consumo relativo baixo, representando a menor renda por unidade consumidora da concessionária. Entretanto, a concessionária se obriga a mantê-los aferidos em sua área de concessão, em conformidade com os limites de erros de medição estabelecidos pela legislação pertinente. De acordo com a legislação vigente o tipo

de fase depende da carga instalada na UC, portanto pode-se justificar que as UCs monofásicas geralmente apresentarão um consumo médio inferior do que as demais.

Admitindo-se apenas os medidores eletromecânicos, a tabela 5.2 apresenta os resultados da análise da quantidade de equipamentos instalados no parque da concessionária de acordo com o tipo de medidor para o ano de 2008, obviamente excluindo-se as informações sobre os medidores eletrônicos.

**Tabela 5.2 – Distribuição de medidores no parque da concessionária de acordo com o tipo de medidor eletromecânico e número de fases, ano 2008.**

<b>Tipo de medidor</b>	<b>Quantidade no parque</b>	<b>Consumo total em kWh</b>
Indução monofásico	857.756 (43,48%)	107.601.562 (28,28%)
Indução bifásico	880.928 (44,66%)	149.439.075 (39,28%)
Indução trifásico	233.853 (11,86%)	123.410.147 (32,44%)
<b>TOTAL</b>	<b>1.972.537</b>	<b>380.450.784</b>

Pela tabela 5.2 observa-se que a quantidade de UCs bifásicas se tornou maior que a quantidade de UCs monofásicas, caracterizando expansão de consumo nesta concessionária exemplo de caso. Contudo, o consumo médio por equipamento é maior nas UC bifásicas (169kWh/mês) do que nas UCs monofásicas (125kWh/mês), justificando-se a escolha da análise.

## **5.2 Aplicação da metodologia sobre os equipamentos**

Uma vez justificada a análise dos medidores eletromecânicos monofásicos é necessário preparar os dados para aplicação da metodologia, ou seja, a etapa de pré-processamento de descoberta de conhecimento em base de dados deve ser aplicada ao conjunto. Esta etapa engloba a aplicação dos filtros para seleção apenas de medidores eletromecânicos monofásicos destinados a utilização em unidades consumidoras.

A quantidade de equipamentos monofásicos obtida após a aplicação de todo o processo de filtragem e cruzamento de dados foi 764.046.

A aplicação da metodologia de análise de probabilidade de falha não compreende todo o parque de medição, sendo que existem modelos agrupados de equipamentos e localidades que não existem na metodologia. Os casos que não podem ser avaliados pela metodologia devem ser descartados.

Na análise existem 689.439 casos de modelos que não são englobados pela metodologia, 3.977 casos de equipamentos sem informação de consumo e 910 casos de localidades que não são englobadas pela metodologia. Um total de 690.473 equipamentos não puderam ser avaliados pela metodologia. Isto equivale a 90% dos equipamentos.

Dos 73.573 equipamentos agrupados, 16 equipamentos apareceram no grupo 200, que não existe no modelo original. A regressão kernel foi utilizada para estimar a probabilidade de falhas deste grupo em 87,5%.

Os grupos com a quantidade de equipamentos associado a cada um deles e a probabilidade de falhas são apresentados na tabela 5.3.

**Tabela 5.3 – Grupos de medidores e quantidade de equipamentos para o estudo de caso.**

<b>Grupo</b>	<b>Nº de casos</b>	<b>Probabilidade de falha</b>
80	4	1,16%
90	278	1,87%
100	3312	4,23%
110	10452	5,35%
120	16519	8,33%
130	16791	14,74%
140	13496	19,14%
150	6659	32,01%
160	3250	35,53%
170	1772	40,63%
180	617	55,00%
190	131	75,00%
200	16	87,50%
210	53	100,00%
220	62	100,00%
230	95	100,00%
240	65	100,00%
250	1	100,00%

Existem equipamentos monofásicos distribuídos em todos os grupos do modelo final. Pode-se determinar a probabilidade de falha média dos medidores monofásicos fazendo o somatório do resultado da multiplicação da quantidade de casos de cada grupo pela probabilidade de falha associada ao grupo, e dividir tudo isso pelo total de casos analisados, conforme a equação 5.1.

$$\bar{Pr} = \frac{\sum (Pr \times N^{\circ} \text{ de casos})}{\sum N^{\circ} \text{ de casos}} \quad (5.1)$$

Sendo  $Pr$  a probabilidade de falha do grupo e  $\bar{Pr}$  a probabilidade de falha média.

A probabilidade de falha média associada aos medidores monofásicos é de 16,14%. Isto quer dizer que dos 73.573 equipamentos analisados, 11.875 devem apresentar falhas.

### **5.3 Considerações finais**

Este capítulo apresentou um estudo de caso real sobre os medidores de energia elétrica monofásicos instalados no parque da concessionária no ano de 2008. Para determinação do tempo de vida dos medidores, considerou-se a data final de 25 de Julho de 2010, aproximadamente dois anos após a imagem do parque. A situação apresentada se torna hipotética pois se estabelece que o consumo médio e máximo não se alteraram e que não houve remoção de equipamentos ou instalação de equipamentos novos. O que se deve ressaltar é a importância de uma metodologia para estimar a probabilidade de falhas nos equipamentos submetidos a diversas condições.

Somente uma pequena quantidade dos medidores monofásicos puderam ser avaliada através da metodologia. A metodologia pôde ser utilizada em 10% do parque de medidores monofásicos. Assim, existe uma dificuldade em afirmar se a quantidade de medidores estimados com falha é grande ou pequena.

Como alternativa a análise da probabilidade de falha, utiliza-se a acurácia. No estudo de caso, se for feita a aferição de 73.573 equipamentos, estima-se que 16,14% tenham falhas. Se for feita a aferição de apenas 1.000 equipamentos, a metodologia indica que 706 equipamentos terão problemas, isto é uma acurácia de 70,6%. A concessionária de energia afirmou que, considerando-se a metodologia atual utilizada, a acurácia média deles está entre 7% até 10%.

Finalmente, observa-se que a grande limitação da metodologia desenvolvida está na falta de dados de ensaios para modelos distintos de medidores, e, ainda um histórico limitado de informações de aferição e falhas, considerando-se que muitos modelos possuíam poucos ou nenhum caso de ensaio registrado.

## 6 CONCLUSÃO E CONTINUIDADE DA PESQUISA

---

---

A construção de modelos representativos da probabilidade de falhas de medidores eletromecânicos de energia elétrica é uma nova ferramenta de auxílio na criação de planos de inspeção e troca de equipamentos. O aplicativo desenvolvido permite monitorar a situação de todo o parque de medição, com possibilidade de emissão de relatórios sobre as condições estimadas dos equipamentos em todas as áreas do parque, de acordo com critérios de busca. Além disso, engloba uma ferramenta que auxilia no processo de retirada de equipamentos para ensaio laboratorial, apontando os equipamentos com maiores probabilidades de falhas. A interface amigável, com o uso de gráficos e mapas, permite que usuários leigos consigam utilizar facilmente o aplicativo.

A metodologia proposta neste trabalho impede a classificação exata da situação de um único equipamento, mas permite generalizar o comportamento do equipamento ao associá-lo a um grupo de equipamentos com características semelhantes. O resultado é uma visualização mais genérica do estado dos equipamentos no parque de medição, apresentando a probabilidade de falha do equipamento.

A técnica de regressão kernel permite que os valores das probabilidades de falhas dos grupos de medidores sejam atualizados, assemelhando-se ao comportamento real observado. A regressão kernel é uma técnica que estima a resposta com base em uma memória, *i.e.*, um conjunto de pontos de entradas e saídas previamente existentes. A principal dificuldade desta técnica é trabalhar com respostas que estão fora dos limites da memória, como pontos inferiores ao menor ponto existente na memória. Nesses casos, o valor será igual ao valor estimado para o ponto limite da memória. Inicialmente esta característica da regressão kernel não influencia no algoritmo desenvolvido, pois os grupos de limite inferior e superior possuem, respectivamente, os valores 0% e 100.

O processo de descoberta de conhecimento em bases de dados possibilita descobrir padrões e relacionamentos em grandes bases de dados que a priori estão ocultos. É uma ferramenta abrangente, utilizada em diversas áreas do conhecimento. Sua utilização foi essencial para determinar as variáveis mais significativas na classificação de falhas em medidores eletromecânicos de energia elétrica ativa.

A etapa de pré-processamento, ressaltando o tratamento de dados discrepantes, possibilitou a construção de uma árvore de decisão de qualidade. A aplicação do algoritmo

J48 sem a remoção dos dados discrepantes de modelos de medidores resulta em uma árvore de decisão muito grande e espalhada, sendo o modelo do equipamento a primeira variável selecionada. A escolha é justificada pela grande quantidade de modelos com apenas um ou dois casos com mesma classificação, assim a árvore consegue classificar 100% desses modelos, sendo a melhor variável para classificação. Após o devido tratamento dessa variável se obteve uma árvore menos espalhada, suficiente para se obter as principais variáveis do modelo.

A regressão *stepwise* é uma técnica muito interessante, cuja função é determinar as principais variáveis de entrada que definem uma variável de saída através de regressão linear. A aplicação da técnica ao conjunto de dados da concessionária apresentou pouca redução do RMSE, indicando que uma função linear não representa adequadamente a saída binária. Independente disto, a análise da variância é uma grande auxiliadora para determinação das variáveis mais significativas para classificar os casos de falha em medidores eletromecânicos. Os algoritmos *forward* e *backward stepwise* apresentaram como resultados variáveis próximas. Se for feita a consideração que ao se escolher um elemento de uma variável transformada via *flattening* se está escolhendo a variável como um todo, então os algoritmos escolheram as mesmas variáveis.

Com a mistura de ambas as técnicas foi possível selecionar o conjunto de variáveis mais indicado para construir um modelo de determinação de falhas. A quantidade de variáveis escolhidas foi a metade da quantidade de variáveis que se acreditava ter relação com as falhas.

Durante o estudo de caso se observou que o modelo ainda não está adequado para analisar o comportamento de todo o parque de medição, sendo que somente 10% dos equipamentos foram analisados. A variável que causou a maior restrição foi o modelo do equipamento. Nos dados do parque da concessionária foram observados 74 modelos agrupados de equipamentos instalados em UCs monofásicas, porém somente 2 destes fazem parte da metodologia proposta.

A variável `localidade` não estabelece um efeito claro sobre as falhas nos equipamentos. Neste trabalho foi estudado os efeitos dos dados climáticos, que estão correlacionados com a localidade, e somente a precipitação apresentou influência significativa sobre o conjunto de dados. Outro ponto a se considerar sobre esta variável são os dados discrepantes. Neste trabalho, os dados discrepantes desta variável não foram tratados, mesmo com a variável se encontrando em condições semelhantes a `subtipo Equip`.

Por último, este trabalho finaliza sem a validação em campo do modelo desenvolvido, considerando-se que a concessionária admitida para estudos prevê a utilização da ferramenta

para o ano de 2011. Como as variáveis do modelo final não possuem uma distribuição abundante em todos os valores que podem assumir, decidiu-se por um modelo mais abrangente, não separando dados atuais existentes para validação.

Com base nessas conclusões, faz-se as seguintes sugestões para a continuidade da pesquisa:

- Obter novos dados de ensaios, englobando modelos de medidores que não fazem parte da metodologia proposta, amostrados de forma aleatória e representativa do parque atual da concessionária;
- Realizar o pré-processamento da variável `localidade`, que apresenta vários campos com apenas um ou dois casos;
- Utilizar outras técnicas de mineração de dados, tanto para obtenção das variáveis mais significativas como para a metodologia de agrupamento de equipamentos;
- Inserir novas variáveis no processo de DCBD, tentando obter modelos que utilizam variáveis mais diretamente relacionadas com a saída e/ou que assumam um leque menor de valores. Isto inclui determinar a existência de variáveis relacionadas com a localidade que influenciam sobre o medidor;
- Experimentar outras técnicas de discretização dos dados, verificando as novas variáveis selecionadas;
- Na técnica de regressão *stepwise*, utilizar uma regressão logística ao invés de uma regressão linear para determinar a saída. Acredita-se que uma regressão logística possa reduzir significativamente o RMSE para valor de saída discreto, obtendo melhor qualidade na seleção das variáveis;
- Validar o modelo proposto, utilizando-se novos conjuntos de dados de ensaios.

## REFERÊNCIAS

- [1] WIKIPÉDIA. **History of electrical engineering**. [S.l: s.n.], 2010. Disponível em: <[http://en.wikipedia.org/wiki/History\\_of\\_electrical\\_engineering](http://en.wikipedia.org/wiki/History_of_electrical_engineering)>. Acesso em: 30 abr. 2010.
- [2] MUNSON, R. **From Edison to Enrol: the business of power and what it means for the future of electricity**. Westport: Praeger, 2005. 206 p.
- [3] DAHLE, D. **Historical timeline of electric meters**. [S.l: s.n.], 2010. Disponível em: <<http://watthourmeters.com/history.html>>. Acesso em: 30 abr. 2010.
- [4] BUSHBELL, O. J.; McCOY, W. E.; VAUGHN, F. A. **Electrical metermen's handbook**. Washington: Edison Electric Institute, 1912. 1070 p.
- [5] SKRABEC, Q. R. Jr. **George Westinghouse: gentle genius**. [S.l. s.n.]. Disponível em: <<http://books.google.com>>. Acesso em: 01 maio 2010.
- [6] DOTTO, G. H. **Estudo do comportamento de medidores eletromecânicos de energia elétrica frente às distorções harmônicas de corrente**. 2006. 92 f. Trabalho de Conclusão de Curso (Graduação em Engenharia Elétrica) – Universidade Regional do Noroeste do Estado do Rio Grande do Sul, Ijuí, 2006.
- [7] BRASIL. Agência Nacional de Energia Elétrica. **Consulta Pública nº 015/2009**. Realizada em 30 março de 2009 para obter subsídios e informações para implantação da medição eletrônica em baixa tensão. Brasília, 2009. Disponível em: <[www.aneel.gov.br](http://www.aneel.gov.br)>. Acesso em: 02 maio 2010.
- [8] LANDIS+GYR. **Manual do medidor M12: medidor monofásico de energia ativa**. Curitiba: Landis + Gyr Equipamentos de Medição, 2007. Disponível em: <[http://www.landisgyr.com.br/upload/opcoes/M12\\_por\\_6p.pdf](http://www.landisgyr.com.br/upload/opcoes/M12_por_6p.pdf)>. Acesso em: 02 maio 2010.
- [9] BRASIL. Agência Nacional de Energia Elétrica. Estudo de vida útil econômica e taxa de depreciação. In: **Anexos da Audiência Pública 012/2006**. Realizada entre 08 set. 2006 a 04 out. 2006 para estabelecer e equalizar taxas anuais de depreciação das unidades consumidoras. Brasília, 2006. Disponível em: <<http://www.aneel.gov.br>>. Acesso em: 02 maio 2010.

- [10] BARBIERI, R. Medidores eletrônicos. In: SEMINÁRIO INTERNACIONAL SOBRE MEDIÇÃO ELETRÔNICA, 1., 2008, Brasília. **Apresentações...** Brasília: ANEEL, 2008. Disponível em: <<http://www.aneel.gov.br/area.cfm?idArea=654&idPerfil=2>>. Acesso em: 02 maio 2010.
- [11] LAMIN, H. **Medição eletrônica em baixa tensão:** aspectos regulatórios e recomendações para implantação. 2009. 184 f. Dissertação (Mestrado em Engenharia Elétrica) – Faculdade de Tecnologia, Universidade de Brasília – UnB, Brasília, 2009.
- [12] BRASIL. Agência Nacional de Energia Elétrica. **Nota Técnica nº 00132009-SRD/ANEEL.** Brasília, 2009. Disponível em: <<http://www.aneel.gov.br>>. Acesso em: 02 maio 2010.
- [13] SEMINÁRIO INTERNACIONAL SOBRE MEDIÇÃO ELETRÔNICA, 1., 2008, Brasília. **Apresentações....** Brasília: ANEEL, 2008. Disponível em: <<http://www.aneel.gov.br/area.cfm?idArea=654&idPerfil=2>>. Acesso em: 02 maio 2010.
- [14] KRUG, S. R. **Aplicação do método de design macroergonômico no projeto de postos de trabalho:** estudo de caso de posto de pré-calibração de medidores de energia monofásicos. 2000. 194 f. Dissertação (Mestrado profissional em Engenharia da Produção) – Escola de Engenharia, Universidade Federal do Rio Grande do Sul – UFRS, Porto Alegre, 2000.
- [15] EL-SADEK, M. Z.; IBRAHIM, E. N. A.; KAMEL, R. M. Watthour meters readings errors in presence of harmonics and active filters. In: INTERNATIONAL CONFERENCE ON ELECTRICAL, ELECTRONIC AND COMPUTER ENGINEERING, 2004, Cairo. **Proceedings...** New York: IEEE, 2004. p. 759-762.
- [16] DENG, Z.; XU, B.; MEI, G. CHEN, Q.; ZHANG, Z. Influence of harmonic powers on metering error of induction watthour meter. **Proceedings of the Chinese Society for Electrical Engineering (CSEE)**, Beijing, v. 22, n. 4, p. 138-143, abr. 2002.
- [17] DORNIJAN, A. Jr.; EMBRIZ-SANTANDER, E.; GILANI, A.; LAMER, G. Watthour meter accuracy under controlled unbalanced harmonic voltage and current conditions. **IEEE Transactions on Power Delivery**, Piscataway, v. 11, n. 1, p. 64-70, jan. 1996.
- [18] IGNEA, A.; CHIVU, M. Influenta componentei armonice de ordinul III asupra contoarelor electrice de inductie monofazate si aparat pentru verificarea acestei influente. **Metrologia Aplicata**, Bucuresti, v. 35, n. 1, p. 33-34, 1988.

- [19] SIQUEIRA, L.; CARNEIRO, A. M.; DE OLIVEIRA, M. Assessment of harmonic distortions on power and energy measurements. In: INTERNATIONAL HARMONICS AND QUALITY OF POWER, 9., 2000, Orlando. **Proceedings...** Orlando: IEEE, 2000.
- [20] CANESIN, C. A.; GONÇALVES, F. A. S.; ORIGA, L. C.; ROSSI, J. C.; GALOTTO, L. Jr.; GODOY, R. B.; SCIAMANA, M.; SILVA, L. C. E. Sistema de medição e modelação de erros em medidores de energia elétrica ativa. **Eletrônica de Potência**, Campinas, v. 13, n. 1, p. 33-43, fev. 2008.
- [21] BRASIL. **Agência Nacional de Energia Elétrica**. Apresenta informações sobre o setor elétrico. Disponível em: <[www.aneel.gov.br](http://www.aneel.gov.br)>. Acesso em: 13 maio 2010.
- [22] AGÊNCIA Reguladora de Saneamento e Energia do Estado de São Paulo. Disponível em: <[www.arsesp.sp.gov.br](http://www.arsesp.sp.gov.br)>. Acesso em: 13 maio 2010.
- [23] BRASIL. **Instituto Nacional de Metrologia, Normalização e Qualidade Industrial**. Disponível em: <[www.inmetro.gov.br](http://www.inmetro.gov.br)>. Acesso em: 13 maio 2010.
- [24] GALOTTO, L., Jr. **Análise de compensação de falta em sensores aplicada em controle de motores**. 2006. 162 f. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal do Mato Grosso do Sul – UFMS, Campo Grande, 2006.
- [25] BASSETO, I. F. F. **Estudo de confiabilidade de compressores alternativos semi-herméticos de sistemas de refrigeração**. 2007. 157 f. Dissertação (Mestrado em Engenharia Mecânica) – Escola Politécnica, Universidade de São Paulo – USP, São Paulo, 2007.
- [26] WIKIPÉDIA. **Failure rate**. [S.l. s.n.], 2010. Disponível em: <[http://en.wikipedia.org/wiki/Failure\\_rate](http://en.wikipedia.org/wiki/Failure_rate)>. Acesso em: 15 maio 2010.
- [27] DIAS, A. Projeto para a confiabilidade aplicado ao processo de implantação de uma rede de gás. **Revista Brasileira de Gestão e Desenvolvimento do Produto**, Florianópolis, v. 2, mar. 2002. Disponível em: <<http://www.moscoso.org/pub/docs/books/fmc2.graco.unb.br/eletronorte2005/fmea/fmea-projetoproduto.pdf>>. Acesso em: 17 maio 2010.
- [28] SHUMWAY, T. Forecasting bankrupt more accurately: a simple hazard model. **Journal of Business**, Chicago, v. 74, n. 1, p. 101-124, 2001.

- [29] ALKAIM, J. L. **Metodologia para incorporar conhecimento intensivo às tarefas de manutenção centrada na confiabilidade aplicada em ativos de sistemas elétricos**. 2003. 239 f. Dissertação (Mestrado em Engenharia da Produção) – Universidade Federal de Santa Catarina – UFSC, Florianópolis, 2003.
- [30] LUZ, A. F. **Uma metodologia baseada em algoritmo de otimização por enxame de partículas para manutenção preventiva focada em confiabilidade e custo**. 2009. 66 f. Dissertação (Mestrado profissional em Engenharia Nuclear) – Instituto de Engenharia Nuclear, Comissão Nacional de Energia Elétrica – CNEN, Rio de Janeiro, 2009.
- [31] GRAZIANO, N. **Análise de confiabilidade e melhoria da taxa de falhas para cubículos classe 15 kV**. 2006. 92 f. Dissertação (Mestrado em Energia) – Programa Interunidades de Pós-Graduação em Energia – PIPGE, Universidade de São Paulo – USP, São Paulo, 2006.
- [32] VANA, C. D.; MELO, C. F.; NICOLAT, J. M.; PORTUGAL, M. A.; IKEDA, M. Desenvolvimento de software para estimativa e acompanhamento da vida útil econômica de medidores de energia. In: CONGRESSO DE INOVAÇÃO TECNOLÓGICA EM ENERGIA ELÉTRICA - CITENEL, 2., 2003. Salvador. **Anais...** [S.l.]: ANEEL, 2003.
- [33] MAIMON, O.; ROKACH, L. **The data mining and knowledge discovery handbook**. New York: Springer, 2005. 1383 p.
- [34] PIATETSKY-SHAPIRO, G.; PARKER, G. **Data mining course**. [S.l.: KDnuggets, 20--?]. Disponível em: <[http://www.kdnuggets.com/data\\_mining\\_course/](http://www.kdnuggets.com/data_mining_course/)>. Acesso em: 16 abr. 2010.
- [35] FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, La Canada, v. 17, p. 37-54, 1996.
- [36] ADRIAANS, P.; ZANTINGE, D. **Data Mining**. Boston: Addison-Wesley, 1997.
- [37] HAN, J.; KAMBER, M. **Data mining: concepts and techniques**. San Francisco: Elsevier, 2006.
- [38] FAYYAD, U.; PIATETSKY-SHAPIRO, G.; PADHRAIC, S.; RAMASAMY, U. **Advances in Knowledge Discovery and Data Mining**. Menlo Park: AAAI, 1996.

- [39] FAYYAD, U.; HAUSSLER, D.; STOLORZ, P. KDD for science data analysis: issues and examples. In: KNOWLEDGE DISCOVERY AND DATA MINING CONFERENCE – KDD, 2., 1996. **Proceedings...** Menlo Park: AAAI, 1996. p. 82-87.
- [40] HOLSHEIMER, M.; KERSTEN, M.; MANNILA, H.; TOIVONEN, H. A perspective on databases and data mining. In: KNOWLEDGE DISCOVERY AND DATA MINING CONFERENCE – KDD, 1., 1995. **Proceedings...** Menlo Park: AAAI, 1995. p. 150-155.
- [41] CARVALHO, D. R.; BUENO, M.; NETO, W. A.; LOPES, L. R. Ferramenta de pré e pós-processamento para data mining. In: SEMINÁRIO DE COMPUTAÇÃO - SEMINCO, 12., 2003, Blumenau. **Anais...** [S.l; s.n], 2003. p. 131-139.
- [42] LENZERINI, M. Data integration: a theoretical perspective. In: SYMPOSIUM OF PRINCIPLES OF DATABASE SYSTEMS, 21., 2002, Madison, Wisconsin. **Proceedings...** New York: ACM, 2002. p. 223-246.
- [43] SARAWAGI, S.; RAISINGHANI, V. T. (Orgs.) **Cleaning methods in data warehouse**. [S.l;s.n], 1999. Disponível em: <<http://www.it.iitb.ac.in/~rvijay/seminar/index.html>>. Acesso em: 13 abr. 2010.
- [44] BOSCARIOLI, C. Pré-processamento de dados para descoberta de conhecimento em banco de dados: uma visão geral. In: CONGRESSO DE TECNOLOGIAS PARA GESTÃO DE DADOS E METADADOS DO CONE SUL – CONGED, 3., 2005. **Anais...** Guarapuava: Unicentro, 2005. p. 101-120.
- [45] BORGES, A. **Aram**: uma linguagem de programação para o ambiente Virtuosi. 10 f. Projeto final (Bacharelado em Ciência da Computação) – Pontifícia Universidade Católica do Paraná, Curitiba, 2004. Disponível em: <[http://www.biblioteca.pucpr.br/tede/tde\\_arquivos/14/TDE-2006-02-10T095302Z-276/Publico/ANEXO\\_B.pdf](http://www.biblioteca.pucpr.br/tede/tde_arquivos/14/TDE-2006-02-10T095302Z-276/Publico/ANEXO_B.pdf)>. Acesso em: 06 maio 2010.
- [46] WIKIPÉDIA. **Caractere de escape**. [S.l;s.n], 2009. Disponível em: <[http://pt.wikipedia.org/wiki/Caractere\\_de\\_escape](http://pt.wikipedia.org/wiki/Caractere_de_escape)>. Acesso em: 06 maio 2010.
- [47] HE, Z.; XU, X.; DENG, S. A fast greedy algorithm for outlier mining. **ArXiv Computer Science E-prints**. Ithaca: Cornell University Library, jul. 2005. Disponível em: <<http://arxiv.org/abs/cs/0507065>>. Acesso em: 15 abr. 2010.
- [48] FAYYAD, U. M.; IRANI, K. B. On the handling of continuous-valued attributes in decision tree generation. **Machine Learning**, Boston, v. 8, n. 1, p. 87-102, Jan. 2005.

- [49] KOTSIANTIS, S.; KANELLOPOULOS, D. Discretization techniques: a recent survey. **GESTS International Transactions on Computer Science and Engineering**, Seoul, v. 32, n. 1, p. 47-58, Jul. 2006.
- [50] PATRICIO, C. M. M. M. **Detecção de fraude ou erro de medição em grandes consumidores de energia elétrica utilizando rough sets baseado em dados históricos e em dados em tempo real**. 2005. 138 f. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal do Mato Grosso do Sul – UFMS, Campo Grande, 2005.
- [51] ANTON, H. A.; BUSBY, R. C. **Álgebra linear contemporânea**. São Paulo: Bookman, 2003. 615 p.
- [52] WIKIPÉDIA. **Moore-Penrose pseudoinverse**. [S.l.: s.n.], 2010. Disponível em: <[http://en.wikipedia.org/wiki/Moore-Penrose\\_pseudoinverse](http://en.wikipedia.org/wiki/Moore-Penrose_pseudoinverse)>. Acesso em: 21 maio 2010.
- [53] GRECHANOVSKY, E. Stepwise regression procedures: overview, problems, results, and suggestions. **Annals of the New York Academy of Sciences**, New York, v. 491, n. 1, p. 197-232, Dez. 2006.
- [54] HOCKING, R. R. The analysis and selection of variables in linear regression. **Biometrics**, Washington, v. 32, n. 1, p. 1-49, Mar. 1976.
- [55] RIBEIRO, C. C.; ARAGÃO, M. V. S. P. de. **Metaheurísticas**. [S.l.: s.n.], 1998. Disponível em: <<http://www.inf.puc-rio.br/~poggi>>. Acesso em: 05 jul. 2010.
- [56] AWASTHI, S. **General stepwise regression**. [S.l.: s.n.], 2000. Disponível em: <<http://www.uta.edu/faculty/sawasthi/Statistics/stgsr.html>>. Acesso em: 22 maio 2010.
- [57] ROKACH, L.; MAIMON, O. **Data mining with decision trees: theory and applications**. Singapore: World Scientific, 2008. 262 p.
- [58] BOUCKAERT, R. R.; FRANK, E.; HALL, M.; KIRKBY, R.; REUTEMANN, P.; SEEWALD, A.; SCUSE, D. **WEKA manual for version 3-6-2**. Hamilton: University of Waikato, 2010. Disponível em: <<http://prdownloads.sourceforge.net/weka/WekaManual-3-6-2.pdf?download>>. Acesso em: 24 abr. 2010.

- [59] MITCHELL, T. M. **Machine learning**. New York: McGraw-Hill, 1997. 432 p.
- [60] MOORE, A. W. **Information gain**. [S.l.: s.n.], 2003. Disponível em: <<http://www.autonlab.org/tutorials/infogain11.pdf>>. Acesso em: 23 maio 2010.
- [61] CUNNINGHAM, P. Evaluation in machine learning. In: EUROPEAN CONFERENCE ON MACHINE LEARNING AND PRINCIPLES AND PRACTICE OF KNOWLEDGE DISCOVERY IN DATABASES – ECML PKDD, 2009, Bled. **Tutorials...** Ljubljana: ECML PKDD, 2009. Disponível em: <<http://www.ecmlpkdd2009.net/program/tutorials/evaluation-in-machine-learning/>>. Acesso em: 12 maio 2010.
- [62] BLUM, A.; KALAI, A.; LANGFORD, J. Beating the hold-out: bounds for k-fold and progressive cross-validation. In: ANNUAL WORKSHOP ON COMPUTATIONAL LEARNING THEORY, 20., 1999, Santa Cruz. **Proceedings...** New York: ACM, 1999, p. 203-208.
- [63] ZUBEN, V. **Regressão paramétrica e não-paramétrica**. Material da disciplina IA353 – Redes neurais. Campinas: Universidade Estadual de Campinas, 2000. Disponível em: <[ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia353\\_00/topico10\\_00.pdf](ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia353_00/topico10_00.pdf)>. Acesso em: 04 ago. 2010.
- [64] **Figura 1.1 – WIKIPÉDIA. Edison bulb**. [S.l.: s.n.], 2006. Disponível em: <[http://en.wikipedia.org/wiki/File:Edison\\_bulb.jpg](http://en.wikipedia.org/wiki/File:Edison_bulb.jpg)>. Acesso em: 02 maio 2010.
- [65] **Figura 1.3 – WIKIPEDIA. Shallenberger-meter**. [S.l.: s.n.], 2008. Disponível em: <<http://de.wikipedia.org/w/index.php?title=Datei:Shallenberger-meter.jpg>>. Acesso em: 02 maio 2010.

## APÊNDICE A – Agrupamento de modelos semelhantes

Este anexo apresenta o agrupamento de medidores de acordo com critérios informados pela concessionária de energia.

A tabela A apresenta os modelos agrupados existentes na tabela RESULTADOS.

**Tabela A – Modelos agrupados utilizados na tabela RESULTADOS.**

<b>Modelo Original</b>	<b>Modelo Agrupado</b>	
B4C3V	B4C3V	
B4C3V12015		
C1X412010	C1X4	
C1X412015		
C1X42405		
C1X424010		
C1X424015		
D58-C/S	D58	
D5804AC		
D58AC3R		
D5804AL		
D5812015JC		
D5812015RC		
D5812030JF		
D5812030RF		
D5812050		
D5824015JJ		
D5824015RJ		
D5824030JM		
D5824030RM		
D5824050		
D58C12015		D58C
D58C120MC		
D58C120NC		
D58C120OC		
D58C120XC		
D58C2401MJ		
D58C240NJ		
D58C240XJ		

D58C240XJ5	
D8L	D8L
D8L12015	
D8L12015KC	
D8L24015	
D8LM	
D8LR	
F7212015	
F7212015DC	
F7224015DJ	
F7224015IJ	
F72G12015	F72G
FV-201-120	FV201
FV201	
FV20112015	
FV20124015	
FV20212015	FV202
FX-221-120	FX221
FX-221-240	
FY201120	FY201
FY20112015	
FY20112030	
FY201120VC	
FY201120VF	
FY201240	
FY20124015	
M1A12005AC	
M1A12015	
M1A24015	
M1A-G12015	M1AG
M1A-G24015	
M1AG-P-TNI	
M1A-T	M1AT
M1A-T12015	
M8C12015	M8C
M8C24015	
M8L12015	M8L
M8L24015	
MF79G120	MF79G
MF79G12015	

MFB120G	MFB120G
MFB120G120	
MFT120-G	MFT120G
MFT120G	
MFT120G120	
MY-202120	MY202
MY-202240	
MY202	
PN5D-G-TNI	PN5DG
PN5DG	
PN5DG12015	
PN5DG24015	
PN5T-G	PN5TG
PN5T-G-TNI	
PN5T-G120	
PN5T-G240	
SL-1621	SL1621
SL-1621120	
SL-1621240	
SL1621	
T4C3V	T4C3V
T4C3V1205	
T4C3V12015	
T4C3V12030	
T4C3V120VC	
T4C3V24015	
T8L C/SENS	T8L
T8L07IF	
T8L07IO	
T8L12015SC	
T8L12015UC	
T8L12015WC	
T8L24015SJ	
T8L24015UJ	
T8L37IF	
T8L37IO	
T8LC/S07LO	
T8LC/S37L0	
T8LM07KC	
T8LM07KO	

## APÊNDICE B – Árvore de decisão completa

Este anexo apresenta a árvore de decisão completa obtida no capítulo 4 em formato texto.

```

instalacoes = 1
|  subtipo = D58
|  |  media = 0
|  |  |  dias_de_vida = 0: N (2.0/1.0)
|  |  |  dias_de_vida = 1600: N (3.0/1.0)
|  |  |  dias_de_vida = 2600: N (7.0/1.0)
|  |  |  dias_de_vida = 4000: N (0.0)
|  |  |  dias_de_vida = 6100: S (4.0/1.0)
|  |  |  dias_de_vida = 7900: S (2.0)
|  |  |  dias_de_vida = 8800
|  |  |  |  altitude = 0: N (0.0)
|  |  |  |  altitude = 400: N (0.0)
|  |  |  |  altitude = 500: N (4.0)
|  |  |  |  altitude = 600: S (3.0/1.0)
|  |  |  |  altitude = 615: S (2.0)
|  |  |  |  altitude = 680: S (1.0)
|  |  |  |  altitude = 800: N (0.0)
|  |  |  media = 100
|  |  |  |  altitude = 0: N (0.0)
|  |  |  |  altitude = 400: S (1.0)
|  |  |  |  altitude = 500: N (7.0)
|  |  |  |  altitude = 600: S (5.0/1.0)
|  |  |  |  altitude = 615
|  |  |  |  |  cod_localidade <= 61: N (3.0/1.0)
|  |  |  |  |  cod_localidade > 61: S (2.0)
|  |  |  |  altitude = 680: S (1.0)
|  |  |  |  altitude = 800: N (0.0)
|  |  |  media = 150
|  |  |  |  maximo = 0: N (0.0)
|  |  |  |  maximo = 150: S (1.0)
|  |  |  |  maximo = 200: N (3.0)
|  |  |  |  maximo = 250: N (6.0/1.0)
|  |  |  |  maximo = 320
|  |  |  |  |  cod_localidade <= 57: S (2.0)
|  |  |  |  |  cod_localidade > 57: N (3.0/1.0)

```

```

| | | maximo = 440: N (3.0/1.0)
| | | maximo = 800: N (1.0)
| | media = 200
| | | cod_tipo_local = RR: S (3.0/1.0)
| | | cod_tipo_local = UB: N (25.0/3.0)
| | media = 300
| | | cod_classe_principal = 1
| | | | temperatura = 0.0: S (0.0)
| | | | temperatura = 20.6: S (0.0)
| | | | temperatura = 21.5: S (1.0)
| | | | temperatura = 21.6
| | | | | maximo = 0: S (0.0)
| | | | | maximo = 150: S (0.0)
| | | | | maximo = 200: S (0.0)
| | | | | maximo = 250: S (0.0)
| | | | | maximo = 320: S (4.0/1.0)
| | | | | maximo = 440
| | | | | precipitacao = 0.0: N (0.0)
| | | | | precipitacao = 100.0
| | | | | | dias_de_vida = 0: N (0.0)
| | | | | | dias_de_vida = 1600: N (2.0/1.0)
| | | | | | dias_de_vida = 2600: S (1.0)
| | | | | | dias_de_vida = 4000: N (2.0/1.0)
| | | | | | dias_de_vida = 6100: N (2.0/1.0)
| | | | | | dias_de_vida = 7900: N (1.0)
| | | | | | dias_de_vida = 8800: N (0.0)
| | | | | precipitacao = 104.4: N (0.0)
| | | | | precipitacao = 110.0: N (3.0)
| | | | | precipitacao = 112.2: S (1.0)
| | | | | precipitacao = 114.0: N (0.0)
| | | | | precipitacao = 148.0: N (0.0)
| | | | | maximo = 800: S (5.0/1.0)
| | | | temperatura = 21.7: N (3.0)
| | | | temperatura = 23.1: S (0.0)
| | | cod_classe_principal = 2
| | | | dias_de_vida = 0: S (0.0)
| | | | dias_de_vida = 1600: S (0.0)
| | | | dias_de_vida = 2600: N (1.0)
| | | | dias_de_vida = 4000: S (0.0)
| | | | dias_de_vida = 6100: S (2.0)
| | | | dias_de_vida = 7900: S (0.0)

```

```

|   |   |   |   dias_de_vida = 8800: S (3.0/1.0)
|   |   |   |   cod_classe_principal = 3
|   |   |   |   maximo = 0: N (0.0)
|   |   |   |   maximo = 150: N (0.0)
|   |   |   |   maximo = 200: N (0.0)
|   |   |   |   maximo = 250: N (0.0)
|   |   |   |   maximo = 320: N (0.0)
|   |   |   |   maximo = 440: S (3.0/1.0)
|   |   |   |   maximo = 800: N (7.0/1.0)
|   |   |   |   cod_classe_principal = 4
|   |   |   |   cod_localidade <= 39: N (2.0)
|   |   |   |   cod_localidade > 39: S (10.0/1.0)
|   |   |   |   cod_classe_principal = 5: N (1.0)
|   |   |   |   cod_classe_principal = 6: S (0.0)
|   |   |   |   cod_classe_principal = 7: S (0.0)
|   |   |   |   cod_classe_principal = 8: S (0.0)
|   |   |   |   media = 600
|   |   |   |   cod_localidade <= 63
|   |   |   |   temperatura = 0.0: N (0.0)
|   |   |   |   temperatura = 20.6: N (0.0)
|   |   |   |   temperatura = 21.5: S (2.0)
|   |   |   |   temperatura = 21.6
|   |   |   |   |   cod_classe_principal = 1: N (2.0/1.0)
|   |   |   |   |   cod_classe_principal = 2: S (3.0/1.0)
|   |   |   |   |   cod_classe_principal = 3
|   |   |   |   |   precipitacao = 0.0: N (0.0)
|   |   |   |   |   precipitacao = 100.0: N (6.0)
|   |   |   |   |   precipitacao = 104.4: N (0.0)
|   |   |   |   |   precipitacao = 110.0: S (1.0)
|   |   |   |   |   precipitacao = 112.2: N (6.0/1.0)
|   |   |   |   |   precipitacao = 114.0: N (0.0)
|   |   |   |   |   precipitacao = 148.0: N (0.0)
|   |   |   |   |   cod_classe_principal = 4: S (3.0/1.0)
|   |   |   |   |   cod_classe_principal = 5: N (6.0/2.0)
|   |   |   |   |   cod_classe_principal = 6: N (0.0)
|   |   |   |   |   cod_classe_principal = 7: N (0.0)
|   |   |   |   |   cod_classe_principal = 8: N (0.0)
|   |   |   |   |   temperatura = 21.7: N (6.0/2.0)
|   |   |   |   |   temperatura = 23.1: N (0.0)
|   |   |   |   |   cod_localidade > 63: S (8.0)
|   |   |   |   subtipo = D58C

```

```

| |   cod_localidade <= 69
| |   |   cod_localidade <= 18
| |   |   |   temperatura = 0.0: N (0.0)
| |   |   |   temperatura = 20.6
| |   |   |   |   maximo = 0: N (2.0/1.0)
| |   |   |   |   maximo = 150: S (0.0)
| |   |   |   |   maximo = 200: S (0.0)
| |   |   |   |   maximo = 250: S (1.0)
| |   |   |   |   maximo = 320: N (1.0)
| |   |   |   |   maximo = 440: S (2.0)
| |   |   |   |   maximo = 800: S (0.0)
| |   |   |   temperatura = 21.5
| |   |   |   |   cod_classe_principal = 1
| |   |   |   |   |   maximo = 0: N (2.0)
| |   |   |   |   |   maximo = 150: N (3.0)
| |   |   |   |   |   maximo = 200
| |   |   |   |   |   |   dias_de_vida = 0: N (0.0)
| |   |   |   |   |   |   dias_de_vida = 1600: N (1.0)
| |   |   |   |   |   |   dias_de_vida = 2600: N (0.0)
| |   |   |   |   |   |   dias_de_vida = 4000: N (1.0)
| |   |   |   |   |   |   dias_de_vida = 6100: S (3.0/1.0)
| |   |   |   |   |   |   dias_de_vida = 7900: N (4.0/2.0)
| |   |   |   |   |   |   dias_de_vida = 8800: N (0.0)
| |   |   |   |   |   maximo = 250: N (0.0)
| |   |   |   |   |   maximo = 320: N (3.0/1.0)
| |   |   |   |   |   maximo = 440: N (2.0/1.0)
| |   |   |   |   |   maximo = 800: N (0.0)
| |   |   |   |   cod_classe_principal = 2: N (0.0)
| |   |   |   |   cod_classe_principal = 3: N (5.0)
| |   |   |   |   cod_classe_principal = 4: N (0.0)
| |   |   |   |   cod_classe_principal = 5: N (0.0)
| |   |   |   |   cod_classe_principal = 6: N (0.0)
| |   |   |   |   cod_classe_principal = 7: N (0.0)
| |   |   |   |   cod_classe_principal = 8: N (0.0)
| |   |   |   temperatura = 21.6: N (0.0)
| |   |   |   temperatura = 21.7: N (0.0)
| |   |   |   temperatura = 23.1: N (0.0)
| |   |   cod_localidade > 18: N (67.0)
| |   cod_localidade > 69
| |   |   media = 0
| |   |   |   maximo = 0

```

```

| | | | | temperatura = 0.0
| | | | | | cod_classe_principal = 1
| | | | | | | precipitacao = 0.0: N (0.0)
| | | | | | | precipitacao = 100.0: S (1.0)
| | | | | | | precipitacao = 104.4: N (3.0/1.0)
| | | | | | | precipitacao = 110.0: N (0.0)
| | | | | | | precipitacao = 112.2: N (3.0/1.0)
| | | | | | | precipitacao = 114.0: N (0.0)
| | | | | | | precipitacao = 148.0: N (0.0)
| | | | | | cod_classe_principal = 2: N (0.0)
| | | | | | cod_classe_principal = 3: N (4.0)
| | | | | | cod_classe_principal = 4: N (0.0)
| | | | | | cod_classe_principal = 5: N (0.0)
| | | | | | cod_classe_principal = 6: N (0.0)
| | | | | | cod_classe_principal = 7: N (0.0)
| | | | | | cod_classe_principal = 8: N (0.0)
| | | | | temperatura = 20.6
| | | | | | cod_tipo_local = RR: S (5.0/1.0)
| | | | | | cod_tipo_local = UB
| | | | | | | dias_de_vida = 0: N (3.0)
| | | | | | | dias_de_vida = 1600: S (2.0)
| | | | | | | dias_de_vida = 2600: N (1.0)
| | | | | | | dias_de_vida = 4000: N (1.0)
| | | | | | | dias_de_vida = 6100: N (2.0)
| | | | | | | dias_de_vida = 7900: N (2.0/1.0)
| | | | | | | dias_de_vida = 8800: N (0.0)
| | | | | temperatura = 21.5: N (4.0/1.0)
| | | | | temperatura = 21.6: S (6.0/1.0)
| | | | | temperatura = 21.7: N (4.0)
| | | | | temperatura = 23.1: N (0.0)
| | | | | maximo = 150
| | | | | | dias_de_vida = 0: N (3.0)
| | | | | | dias_de_vida = 1600: N (3.0/1.0)
| | | | | | dias_de_vida = 2600: S (1.0)
| | | | | | dias_de_vida = 4000: N (1.0)
| | | | | | dias_de_vida = 6100: N (4.0/1.0)
| | | | | | dias_de_vida = 7900: N (1.0)
| | | | | | dias_de_vida = 8800: N (0.0)
| | | | | maximo = 200: N (4.0)
| | | | | maximo = 250
| | | | | | dias_de_vida = 0: N (0.0)

```

```

| | | | | dias_de_vida = 1600: S (1.0)
| | | | | dias_de_vida = 2600: N (0.0)
| | | | | dias_de_vida = 4000: N (2.0)
| | | | | dias_de_vida = 6100: N (2.0)
| | | | | dias_de_vida = 7900: N (0.0)
| | | | | dias_de_vida = 8800: N (0.0)
| | | | | maximo = 320: N (3.0)
| | | | | maximo = 440: N (1.0)
| | | | | maximo = 800: N (0.0)
| | | | media = 100
| | | | | temperatura = 0.0
| | | | | dias_de_vida = 0: N (2.0)
| | | | | dias_de_vida = 1600: N (4.0)
| | | | | dias_de_vida = 2600: N (4.0/1.0)
| | | | | dias_de_vida = 4000: N (4.0/1.0)
| | | | | dias_de_vida = 6100
| | | | | | cod_localidade <= 130: S (2.0)
| | | | | | cod_localidade > 130: N (3.0/1.0)
| | | | | dias_de_vida = 7900: N (3.0)
| | | | | dias_de_vida = 8800: N (0.0)
| | | | | temperatura = 20.6: N (8.0)
| | | | | temperatura = 21.5
| | | | | maximo = 0: N (2.0)
| | | | | maximo = 150: N (0.0)
| | | | | maximo = 200: S (3.0/1.0)
| | | | | maximo = 250: N (1.0)
| | | | | maximo = 320: N (0.0)
| | | | | maximo = 440: S (1.0)
| | | | | maximo = 800: N (0.0)
| | | | | temperatura = 21.6
| | | | | maximo = 0: N (1.0)
| | | | | maximo = 150: S (3.0/1.0)
| | | | | maximo = 200: N (2.0)
| | | | | maximo = 250: N (1.0)
| | | | | maximo = 320: S (1.0)
| | | | | maximo = 440: N (0.0)
| | | | | maximo = 800: N (0.0)
| | | | | temperatura = 21.7
| | | | | maximo = 0: N (0.0)
| | | | | maximo = 150: N (1.0)
| | | | | maximo = 200: N (2.0)

```

```

| | | | | maximo = 250: S (1.0)
| | | | | maximo = 320: N (0.0)
| | | | | maximo = 440: N (2.0)
| | | | | maximo = 800: N (0.0)
| | | | | temperatura = 23.1: N (0.0)
| | | | |
| | | | | media = 150
| | | | |
| | | | | temperatura = 0.0
| | | | |
| | | | | maximo = 0: N (0.0)
| | | | | maximo = 150: N (0.0)
| | | | | maximo = 200
| | | | |
| | | | | | dias_de_vida = 0: N (0.0)
| | | | | | dias_de_vida = 1600: N (1.0)
| | | | | | dias_de_vida = 2600: S (1.0)
| | | | | | dias_de_vida = 4000: S (3.0/1.0)
| | | | | | dias_de_vida = 6100: N (2.0)
| | | | | | dias_de_vida = 7900: S (1.0)
| | | | | | dias_de_vida = 8800: N (0.0)
| | | | | |
| | | | | | maximo = 250
| | | | | |
| | | | | | | altitude = 0: S (0.0)
| | | | | | | altitude = 400: S (0.0)
| | | | | | | altitude = 500: S (0.0)
| | | | | | | altitude = 600: S (0.0)
| | | | | | | altitude = 615: N (1.0)
| | | | | | | altitude = 680: S (3.0)
| | | | | | | altitude = 800: N (2.0/1.0)
| | | | | |
| | | | | | maximo = 320
| | | | | | | dias_de_vida = 0: N (0.0)
| | | | | | | dias_de_vida = 1600: N (1.0)
| | | | | | | dias_de_vida = 2600: N (2.0/1.0)
| | | | | | | dias_de_vida = 4000: N (3.0)
| | | | | | | dias_de_vida = 6100: S (1.0)
| | | | | | | dias_de_vida = 7900: N (1.0)
| | | | | | | dias_de_vida = 8800: N (0.0)
| | | | | |
| | | | | | maximo = 440: N (3.0/1.0)
| | | | | | maximo = 800: N (0.0)
| | | | | |
| | | | | | temperatura = 20.6: N (18.0/1.0)
| | | | | |
| | | | | | temperatura = 21.5
| | | | | |
| | | | | | maximo = 0: N (0.0)
| | | | | | maximo = 150: S (1.0)
| | | | | | maximo = 200: N (3.0)
| | | | | |
| | | | | | maximo = 250: N (6.0/1.0)

```

```

| | | | | maximo = 320: N (1.0)
| | | | | maximo = 440: S (1.0)
| | | | | maximo = 800: N (0.0)
| | | | | temperatura = 21.6
| | | | | precipitacao = 0.0
| | | | | | cod_classe_principal = 1: S (2.0)
| | | | | | cod_classe_principal = 2: N (0.0)
| | | | | | cod_classe_principal = 3: N (2.0)
| | | | | | cod_classe_principal = 4: N (0.0)
| | | | | | cod_classe_principal = 5: N (0.0)
| | | | | | cod_classe_principal = 6: N (0.0)
| | | | | | cod_classe_principal = 7: N (0.0)
| | | | | | cod_classe_principal = 8: N (0.0)
| | | | | precipitacao = 100.0: N (2.0)
| | | | | precipitacao = 104.4: N (2.0/1.0)
| | | | | precipitacao = 110.0: N (0.0)
| | | | | precipitacao = 112.2: N (0.0)
| | | | | precipitacao = 114.0: N (1.0)
| | | | | precipitacao = 148.0: N (0.0)
| | | | | temperatura = 21.7: N (2.0/1.0)
| | | | | temperatura = 23.1: N (0.0)
| | | | media = 200
| | | | | cod_tipo_local = RR: N (5.0)
| | | | | cod_tipo_local = UB
| | | | | | precipitacao = 0.0
| | | | | | temperatura = 0.0: N (0.0)
| | | | | | temperatura = 20.6
| | | | | | | dias_de_vida = 0: N (0.0)
| | | | | | | dias_de_vida = 1600: N (2.0)
| | | | | | | dias_de_vida = 2600: S (1.0)
| | | | | | | dias_de_vida = 4000: N (1.0)
| | | | | | | dias_de_vida = 6100: N (2.0)
| | | | | | | dias_de_vida = 7900: N (2.0/1.0)
| | | | | | | dias_de_vida = 8800: N (0.0)
| | | | | | temperatura = 21.5: N (2.0/1.0)
| | | | | | temperatura = 21.6: N (4.0)
| | | | | | temperatura = 21.7: S (1.0)
| | | | | | temperatura = 23.1: N (0.0)
| | | | | | precipitacao = 100.0
| | | | | | maximo = 0: N (0.0)
| | | | | | maximo = 150: N (0.0)

```

```

| | | | | | maximo = 200: N (0.0)
| | | | | | maximo = 250: N (2.0)
| | | | | | maximo = 320: N (2.0)
| | | | | | maximo = 440: S (1.0)
| | | | | | maximo = 800: N (0.0)
| | | | | | precipitacao = 104.4
| | | | | | dias_de_vida = 0: N (1.0)
| | | | | | dias_de_vida = 1600: N (3.0/1.0)
| | | | | | dias_de_vida = 2600: S (1.0)
| | | | | | dias_de_vida = 4000: S (3.0/1.0)
| | | | | | dias_de_vida = 6100: N (2.0/1.0)
| | | | | | dias_de_vida = 7900: N (1.0)
| | | | | | dias_de_vida = 8800: N (0.0)
| | | | | | precipitacao = 110.0
| | | | | | maximo = 0: N (0.0)
| | | | | | maximo = 150: N (0.0)
| | | | | | maximo = 200: N (1.0)
| | | | | | maximo = 250: N (4.0/1.0)
| | | | | | maximo = 320
| | | | | | | dias_de_vida = 0: S (0.0)
| | | | | | | dias_de_vida = 1600: N (1.0)
| | | | | | | dias_de_vida = 2600: S (0.0)
| | | | | | | dias_de_vida = 4000: S (2.0)
| | | | | | | dias_de_vida = 6100: N (1.0)
| | | | | | | dias_de_vida = 7900: S (3.0)
| | | | | | | dias_de_vida = 8800: S (0.0)
| | | | | | maximo = 440: N (0.0)
| | | | | | maximo = 800: N (0.0)
| | | | | | precipitacao = 112.2: N (4.0/1.0)
| | | | | | precipitacao = 114.0
| | | | | | dias_de_vida = 0: N (0.0)
| | | | | | dias_de_vida = 1600: S (1.0)
| | | | | | dias_de_vida = 2600: N (3.0)
| | | | | | dias_de_vida = 4000: N (5.0/1.0)
| | | | | | dias_de_vida = 6100: N (7.0/2.0)
| | | | | | dias_de_vida = 7900: N (0.0)
| | | | | | dias_de_vida = 8800: N (0.0)
| | | | | | precipitacao = 148.0: S (3.0)
| | | | | media = 300
| | | | | cod_classe_principal = 1
| | | | | temperatura = 0.0: N (8.0/1.0)

```

```

| | | | | | temperatura = 20.6
| | | | | | | dias_de_vida = 0: N (0.0)
| | | | | | | dias_de_vida = 1600: N (2.0)
| | | | | | | dias_de_vida = 2600: N (2.0/1.0)
| | | | | | | dias_de_vida = 4000: N (3.0)
| | | | | | | dias_de_vida = 6100: N (1.0)
| | | | | | | dias_de_vida = 7900: S (3.0/1.0)
| | | | | | | dias_de_vida = 8800: N (0.0)
| | | | | | temperatura = 21.5: N (7.0/3.0)
| | | | | | temperatura = 21.6: N (4.0)
| | | | | | temperatura = 21.7: N (5.0)
| | | | | | temperatura = 23.1: N (0.0)
| | | | | | cod_classe_principal = 2: N (1.0)
| | | | | | cod_classe_principal = 3: N (13.0)
| | | | | | cod_classe_principal = 4: N (1.0)
| | | | | | cod_classe_principal = 5: S (1.0)
| | | | | | cod_classe_principal = 6: N (0.0)
| | | | | | cod_classe_principal = 7: N (0.0)
| | | | | | cod_classe_principal = 8: N (0.0)
| | | | | media = 600
| | | | | altitude = 0: N (0.0)
| | | | | altitude = 400: N (0.0)
| | | | | altitude = 500: N (0.0)
| | | | | altitude = 600: S (3.0/1.0)
| | | | | altitude = 615: S (4.0/1.0)
| | | | | altitude = 680: N (3.0)
| | | | | altitude = 800: N (0.0)
| | | | | subtipo = D8L
| | | | | ano_fabric = 0
| | | | | cod_tipo_fase = MO: S (1.0)
| | | | | cod_tipo_fase = BI
| | | | | | cod_classe_principal = 1
| | | | | | | dias_de_vida = 0: N (6.0)
| | | | | | | dias_de_vida = 1600
| | | | | | | media = 0: N (8.0)
| | | | | | | media = 100: N (2.0)
| | | | | | | media = 150: S (1.0)
| | | | | | | media = 200: N (1.0)
| | | | | | | media = 300: S (1.0)
| | | | | | | media = 600: N (0.0)
| | | | | | dias_de_vida = 2600: N (12.0/3.0)

```

```

| | | | | dias_de_vida = 4000: N (15.0/2.0)
| | | | | dias_de_vida = 6100: N (14.0)
| | | | | dias_de_vida = 7900
| | | | | | media = 0: S (1.0)
| | | | | | media = 100: N (3.0)
| | | | | | media = 150
| | | | | | | cod_localidade <= 130: N (2.0)
| | | | | | | cod_localidade > 130: S (3.0/1.0)
| | | | | | media = 200: N (4.0)
| | | | | | media = 300: N (1.0)
| | | | | | media = 600: N (0.0)
| | | | | dias_de_vida = 8800
| | | | | | maximo = 0: N (2.0)
| | | | | | maximo = 150: N (6.0/1.0)
| | | | | | maximo = 200: N (1.0)
| | | | | | maximo = 250: N (5.0/1.0)
| | | | | | maximo = 320: S (7.0/3.0)
| | | | | | maximo = 440: N (5.0/1.0)
| | | | | | maximo = 800: S (1.0)
| | | | | cod_classe_principal = 2: N (1.0)
| | | | | cod_classe_principal = 3: N (16.0)
| | | | | cod_classe_principal = 4: N (4.0)
| | | | | cod_classe_principal = 5: N (2.0)
| | | | | cod_classe_principal = 6: N (0.0)
| | | | | cod_classe_principal = 7: N (0.0)
| | | | | cod_classe_principal = 8: N (0.0)
| | | | | cod_tipo_fase = TR: N (2.0/1.0)
| | | ano_fabric = 1980: N (16.0)
| | | ano_fabric = 1983: N (69.0)
| | | ano_fabric = 1986: N (1.0)
| | | ano_fabric = 1992: N (0.0)
| | | ano_fabric = 1998: N (0.0)
| | subtipo = FV201
| | | cod_classe_principal = 1
| | | | media = 0
| | | | | altitude = 0: N (0.0)
| | | | | altitude = 400: N (0.0)
| | | | | altitude = 500: S (1.0)
| | | | | altitude = 600: N (6.0/1.0)
| | | | | altitude = 615
| | | | | | maximo = 0: N (4.0/1.0)

```

```

| | | | | maximo = 150: N (2.0)
| | | | | maximo = 200: S (1.0)
| | | | | maximo = 250: N (0.0)
| | | | | maximo = 320: N (1.0)
| | | | | maximo = 440: N (0.0)
| | | | | maximo = 800: N (0.0)
| | | | | altitude = 680: N (0.0)
| | | | | altitude = 800: N (0.0)
| | | | | media = 100
| | | | | cod_classe_consumo <= 9101
| | | | | precipitacao = 0.0: N (0.0)
| | | | | precipitacao = 100.0: N (2.0)
| | | | | precipitacao = 104.4
| | | | | | dias_de_vida = 0: S (0.0)
| | | | | | dias_de_vida = 1600: S (0.0)
| | | | | | dias_de_vida = 2600: S (0.0)
| | | | | | dias_de_vida = 4000: S (1.0)
| | | | | | dias_de_vida = 6100: N (2.0)
| | | | | | dias_de_vida = 7900: S (0.0)
| | | | | | dias_de_vida = 8800: S (2.0)
| | | | | precipitacao = 110.0: N (5.0)
| | | | | precipitacao = 112.2
| | | | | | maximo = 0: N (0.0)
| | | | | | maximo = 150: N (5.0/1.0)
| | | | | | maximo = 200: N (1.0)
| | | | | | maximo = 250: S (3.0/1.0)
| | | | | | maximo = 320: S (1.0)
| | | | | | maximo = 440: N (0.0)
| | | | | | maximo = 800: N (0.0)
| | | | | precipitacao = 114.0: N (2.0)
| | | | | precipitacao = 148.0: N (0.0)
| | | | | cod_classe_consumo > 9101: S (2.0)
| | | | | media = 150
| | | | | cod_localidade <= 39: S (4.0)
| | | | | cod_localidade > 39
| | | | | | temperatura = 0.0: N (0.0)
| | | | | | temperatura = 20.6: S (5.0/2.0)
| | | | | | temperatura = 21.5: N (7.0)
| | | | | | temperatura = 21.6: N (9.0/2.0)
| | | | | | temperatura = 21.7: N (0.0)
| | | | | | temperatura = 23.1: N (0.0)

```

```

| | | media = 200
| | | | maximo = 0: N (0.0)
| | | | maximo = 150: N (0.0)
| | | | maximo = 200: N (0.0)
| | | | maximo = 250: N (13.0/3.0)
| | | | maximo = 320
| | | | | precipitacao = 0.0: N (0.0)
| | | | | precipitacao = 100.0: N (0.0)
| | | | | precipitacao = 104.4: N (1.0)
| | | | | precipitacao = 110.0: N (3.0/1.0)
| | | | | precipitacao = 112.2
| | | | | | cod_localidade <= 55: N (3.0/1.0)
| | | | | | cod_localidade > 55: S (2.0)
| | | | | precipitacao = 114.0: S (1.0)
| | | | | precipitacao = 148.0: N (0.0)
| | | | maximo = 440: S (3.0/1.0)
| | | | maximo = 800: N (2.0)
| | | media = 300
| | | | maximo = 0: S (0.0)
| | | | maximo = 150: S (0.0)
| | | | maximo = 200: S (0.0)
| | | | maximo = 250: S (0.0)
| | | | maximo = 320: S (4.0)
| | | | maximo = 440
| | | | | altitude = 0: S (0.0)
| | | | | altitude = 400: S (0.0)
| | | | | altitude = 500: N (2.0)
| | | | | altitude = 600
| | | | | | ano_fabric = 0: N (0.0)
| | | | | | ano_fabric = 1980: N (3.0)
| | | | | | ano_fabric = 1983: S (2.0)
| | | | | | ano_fabric = 1986: N (0.0)
| | | | | | ano_fabric = 1992: N (0.0)
| | | | | | ano_fabric = 1998: N (0.0)
| | | | | altitude = 615: S (5.0/1.0)
| | | | | altitude = 680: S (1.0)
| | | | | altitude = 800: S (0.0)
| | | | maximo = 800: S (0.0)
| | | media = 600: N (1.0)
| | cod_classe_principal = 2: S (1.0)
| | cod_classe_principal = 3

```

```

| | | dias_de_vida = 0: N (3.0)
| | | dias_de_vida = 1600: N (4.0)
| | | dias_de_vida = 2600: S (1.0)
| | | dias_de_vida = 4000: N (2.0)
| | | dias_de_vida = 6100
| | | | media = 0: N (0.0)
| | | | media = 100: N (2.0/1.0)
| | | | media = 150: N (1.0)
| | | | media = 200: N (2.0)
| | | | media = 300: S (2.0)
| | | | media = 600: N (0.0)
| | | dias_de_vida = 7900: N (4.0/2.0)
| | | dias_de_vida = 8800: N (1.0)
| | cod_classe_principal = 4: N (0.0)
| | cod_classe_principal = 5: N (3.0/1.0)
| | cod_classe_principal = 6: S (1.0)
| | cod_classe_principal = 7: N (0.0)
| | cod_classe_principal = 8: N (0.0)
| subtipo = M1A
| | cod_localidade <= 698
| | | ano_fabric = 0: N (88.0/1.0)
| | | ano_fabric = 1980: N (2.0)
| | | ano_fabric = 1983: N (0.0)
| | | ano_fabric = 1986: S (1.0)
| | | ano_fabric = 1992: N (4.0)
| | | ano_fabric = 1998: N (0.0)
| | cod_localidade > 698
| | | media = 0
| | | | cod_classe_consumo <= 7050: N (7.0)
| | | | cod_classe_consumo > 7050
| | | | | cod_classe_principal = 1
| | | | | | maximo = 0
| | | | | | | dias_de_vida = 0
| | | | | | | | cod_localidade <= 760: N (2.0)
| | | | | | | | cod_localidade > 760: S (6.0/2.0)
| | | | | | | | dias_de_vida = 1600: N (5.0/1.0)
| | | | | | | | dias_de_vida = 2600: S (3.0)
| | | | | | | | dias_de_vida = 4000
| | | | | | | | | altitude = 0: S (1.0)
| | | | | | | | | altitude = 400: N (2.0/1.0)
| | | | | | | | | altitude = 500: N (2.0)

```

```

| | | | | | | | altitude = 600: N (0.0)
| | | | | | | | altitude = 615: N (0.0)
| | | | | | | | altitude = 680: N (0.0)
| | | | | | | | altitude = 800: N (0.0)
| | | | | | | | dias_de_vida = 6100: N (6.0/3.0)
| | | | | | | | dias_de_vida = 7900
| | | | | | | | precipitacao = 0.0
| | | | | | | | | cod_localidade <= 811: S (2.0)
| | | | | | | | | cod_localidade > 811: N (5.0/2.0)
| | | | | | | | precipitacao = 100.0: S (1.0)
| | | | | | | | precipitacao = 104.4: N (0.0)
| | | | | | | | precipitacao = 110.0: N (9.0/3.0)
| | | | | | | | precipitacao = 112.2: N (0.0)
| | | | | | | | precipitacao = 114.0: N (0.0)
| | | | | | | | precipitacao = 148.0: N (0.0)
| | | | | | | | dias_de_vida = 8800: N (0.0)
| | | | | | | | maximo = 150
| | | | | | | | dias_de_vida = 0: N (0.0)
| | | | | | | | dias_de_vida = 1600: S (1.0)
| | | | | | | | dias_de_vida = 2600: N (1.0)
| | | | | | | | dias_de_vida = 4000: S (2.0)
| | | | | | | | dias_de_vida = 6100: N (3.0)
| | | | | | | | dias_de_vida = 7900: N (2.0/1.0)
| | | | | | | | dias_de_vida = 8800: N (0.0)
| | | | | | | | maximo = 200: S (2.0)
| | | | | | | | maximo = 250: N (1.0)
| | | | | | | | maximo = 320: N (0.0)
| | | | | | | | maximo = 440: N (0.0)
| | | | | | | | maximo = 800: N (0.0)
| | | | | | | | cod_classe_principal = 2: S (0.0)
| | | | | | | | cod_classe_principal = 3: S (4.0/1.0)
| | | | | | | | cod_classe_principal = 4: S (0.0)
| | | | | | | | cod_classe_principal = 5: S (1.0)
| | | | | | | | cod_classe_principal = 6: S (0.0)
| | | | | | | | cod_classe_principal = 7: S (0.0)
| | | | | | | | cod_classe_principal = 8: S (0.0)
| | | | media = 100
| | | | dias_de_vida = 0: N (5.0)
| | | | dias_de_vida = 1600
| | | | cod_localidade <= 808: N (4.0/1.0)
| | | | cod_localidade > 808: S (2.0)

```

```

|   |   |   |   dias_de_vida = 2600: N (11.0/1.0)
|   |   |   |   dias_de_vida = 4000: N (10.0/1.0)
|   |   |   |   dias_de_vida = 6100: N (9.0/2.0)
|   |   |   |   dias_de_vida = 7900
|   |   |   |   |   precipitacao = 0.0: N (6.0)
|   |   |   |   |   precipitacao = 100.0: N (3.0)
|   |   |   |   |   precipitacao = 104.4: N (0.0)
|   |   |   |   |   precipitacao = 110.0: S (9.0/4.0)
|   |   |   |   |   precipitacao = 112.2: N (0.0)
|   |   |   |   |   precipitacao = 114.0: N (0.0)
|   |   |   |   |   precipitacao = 148.0: N (0.0)
|   |   |   |   dias_de_vida = 8800: N (3.0)
|   |   |   media = 150
|   |   |   |   maximo = 0: N (0.0)
|   |   |   |   maximo = 150: N (5.0)
|   |   |   |   maximo = 200
|   |   |   |   |   cod_localidade <= 760: S (2.0)
|   |   |   |   |   cod_localidade > 760: N (8.0)
|   |   |   |   maximo = 250: N (4.0/1.0)
|   |   |   |   maximo = 320: S (3.0/1.0)
|   |   |   |   maximo = 440: S (1.0)
|   |   |   |   maximo = 800: N (0.0)
|   |   |   media = 200: N (6.0/2.0)
|   |   |   media = 300: N (4.0)
|   |   |   media = 600: N (0.0)
|   subtipo = M8L
|   |   maximo = 0
|   |   |   dias_de_vida = 0
|   |   |   |   precipitacao = 0.0: N (0.0)
|   |   |   |   precipitacao = 100.0: N (2.0/1.0)
|   |   |   |   precipitacao = 104.4: N (0.0)
|   |   |   |   precipitacao = 110.0: N (4.0/1.0)
|   |   |   |   precipitacao = 112.2: N (2.0)
|   |   |   |   precipitacao = 114.0: S (1.0)
|   |   |   |   precipitacao = 148.0: N (0.0)
|   |   |   dias_de_vida = 1600
|   |   |   |   ano_fabric = 0: N (0.0)
|   |   |   |   ano_fabric = 1980: N (3.0/1.0)
|   |   |   |   ano_fabric = 1983: N (4.0/1.0)
|   |   |   |   ano_fabric = 1986: S (1.0)
|   |   |   |   ano_fabric = 1992: N (0.0)

```

```

|   |   |   |   ano_fabric = 1998: N (0.0)
|   |   |   | dias_de_vida = 2600: N (4.0/1.0)
|   |   |   | dias_de_vida = 4000: N (9.0/1.0)
|   |   |   | dias_de_vida = 6100
|   |   |   |   precipitacao = 0.0: S (0.0)
|   |   |   |   precipitacao = 100.0: N (3.0/1.0)
|   |   |   |   precipitacao = 104.4: S (0.0)
|   |   |   |   precipitacao = 110.0: S (3.0)
|   |   |   |   precipitacao = 112.2: N (3.0/1.0)
|   |   |   |   precipitacao = 114.0: S (0.0)
|   |   |   |   precipitacao = 148.0: S (0.0)
|   |   |   | dias_de_vida = 7900: S (5.0/1.0)
|   |   |   | dias_de_vida = 8800: N (1.0)
|   |   |   | maximo = 150
|   |   |   |   precipitacao = 0.0: N (0.0)
|   |   |   |   precipitacao = 100.0
|   |   |   |   |   cod_localidade <= 49
|   |   |   |   |   |   dias_de_vida = 0: S (0.0)
|   |   |   |   |   |   dias_de_vida = 1600: N (1.0)
|   |   |   |   |   |   dias_de_vida = 2600: S (2.0)
|   |   |   |   |   |   dias_de_vida = 4000: S (0.0)
|   |   |   |   |   |   dias_de_vida = 6100: N (4.0/2.0)
|   |   |   |   |   |   dias_de_vida = 7900: S (3.0)
|   |   |   |   |   |   dias_de_vida = 8800: S (0.0)
|   |   |   |   |   |   cod_localidade > 49: N (2.0)
|   |   |   |   precipitacao = 104.4: N (0.0)
|   |   |   |   precipitacao = 110.0
|   |   |   |   |   dias_de_vida = 0: S (1.0)
|   |   |   |   |   dias_de_vida = 1600: N (3.0)
|   |   |   |   |   dias_de_vida = 2600: N (1.0)
|   |   |   |   |   dias_de_vida = 4000: N (4.0/1.0)
|   |   |   |   |   dias_de_vida = 6100: N (8.0/3.0)
|   |   |   |   |   dias_de_vida = 7900: N (3.0)
|   |   |   |   |   dias_de_vida = 8800: N (0.0)
|   |   |   |   precipitacao = 112.2: S (5.0/1.0)
|   |   |   |   precipitacao = 114.0
|   |   |   |   |   dias_de_vida = 0: N (1.0)
|   |   |   |   |   dias_de_vida = 1600: S (2.0)
|   |   |   |   |   dias_de_vida = 2600: S (1.0)
|   |   |   |   |   dias_de_vida = 4000: N (1.0)
|   |   |   |   |   dias_de_vida = 6100: N (4.0/1.0)

```

```

| | | | dias_de_vida = 7900: N (0.0)
| | | | dias_de_vida = 8800: N (0.0)
| | | | precipitacao = 148.0: N (0.0)
| | | | maximo = 200
| | | | precipitacao = 0.0: N (0.0)
| | | | precipitacao = 100.0
| | | | | media = 0: N (2.0/1.0)
| | | | | media = 100: S (5.0/2.0)
| | | | | media = 150: N (5.0)
| | | | | media = 200: N (0.0)
| | | | | media = 300: N (0.0)
| | | | | media = 600: N (0.0)
| | | | precipitacao = 104.4: N (0.0)
| | | | precipitacao = 110.0
| | | | | dias_de_vida = 0: S (2.0)
| | | | | dias_de_vida = 1600: N (2.0/1.0)
| | | | | dias_de_vida = 2600: N (3.0)
| | | | | dias_de_vida = 4000
| | | | | | media = 0: N (2.0/1.0)
| | | | | | media = 100: S (1.0)
| | | | | | media = 150: N (2.0)
| | | | | | media = 200: N (0.0)
| | | | | | media = 300: N (0.0)
| | | | | | media = 600: N (0.0)
| | | | | dias_de_vida = 6100: N (5.0)
| | | | | dias_de_vida = 7900: S (1.0)
| | | | | dias_de_vida = 8800: N (0.0)
| | | | | precipitacao = 112.2
| | | | | | dias_de_vida = 0: S (1.0)
| | | | | | dias_de_vida = 1600: N (2.0)
| | | | | | dias_de_vida = 2600: S (1.0)
| | | | | | dias_de_vida = 4000: S (1.0)
| | | | | | dias_de_vida = 6100
| | | | | | | ano_fabric = 0: S (0.0)
| | | | | | | ano_fabric = 1980: N (1.0)
| | | | | | | ano_fabric = 1983: S (5.0/1.0)
| | | | | | | ano_fabric = 1986: N (2.0/1.0)
| | | | | | | ano_fabric = 1992: S (0.0)
| | | | | | | ano_fabric = 1998: S (0.0)
| | | | | | dias_de_vida = 7900: N (1.0)
| | | | | | dias_de_vida = 8800: S (0.0)

```

```

| | | precipitacao = 114.0: S (4.0)
| | | precipitacao = 148.0: N (0.0)
| | maximo = 250
| | | cod_localidade <= 63
| | | | ano_fabric = 0: N (0.0)
| | | | ano_fabric = 1980
| | | | | precipitacao = 0.0: N (0.0)
| | | | | precipitacao = 100.0: S (1.0)
| | | | | precipitacao = 104.4: N (0.0)
| | | | | precipitacao = 110.0: N (1.0)
| | | | | precipitacao = 112.2: N (2.0/1.0)
| | | | | precipitacao = 114.0: N (2.0/1.0)
| | | | | precipitacao = 148.0: N (0.0)
| | | | ano_fabric = 1983
| | | | | altitude = 0: N (0.0)
| | | | | altitude = 400: N (0.0)
| | | | | altitude = 500: N (4.0/1.0)
| | | | | altitude = 600: N (4.0/1.0)
| | | | | altitude = 615: N (6.0)
| | | | | altitude = 680: S (3.0/1.0)
| | | | | altitude = 800: N (0.0)
| | | | ano_fabric = 1986: N (4.0)
| | | | ano_fabric = 1992: N (0.0)
| | | | ano_fabric = 1998: N (0.0)
| | | cod_localidade > 63: S (3.0)
| | maximo = 320: S (7.0/1.0)
| | maximo = 440: N (4.0/1.0)
| | maximo = 800: N (2.0)
| subtipo = MFT120G
| | cod_classe_principal = 1
| | | precipitacao = 0.0
| | | | maximo = 0: N (1.0)
| | | | maximo = 150: N (0.0)
| | | | maximo = 200: S (1.0)
| | | | maximo = 250: N (2.0)
| | | | maximo = 320: N (2.0/1.0)
| | | | maximo = 440
| | | | | altitude = 0: S (2.0)
| | | | | altitude = 400: N (3.0)
| | | | | altitude = 500: N (0.0)
| | | | | altitude = 600: N (0.0)

```

```

| | | | altitude = 615: N (0.0)
| | | | altitude = 680: N (0.0)
| | | | altitude = 800: N (0.0)
| | | | maximo = 800: N (5.0/2.0)
| | | precipitacao = 100.0: N (1.0)
| | | precipitacao = 104.4: N (3.0)
| | | precipitacao = 110.0: N (0.0)
| | | precipitacao = 112.2: N (0.0)
| | | precipitacao = 114.0: N (0.0)
| | | precipitacao = 148.0: N (0.0)
| | cod_classe_principal = 2
| | | media = 0: N (3.0)
| | | media = 100: N (0.0)
| | | media = 150: N (0.0)
| | | media = 200: N (1.0)
| | | media = 300: N (2.0)
| | | media = 600: S (1.0)
| | cod_classe_principal = 3: N (21.0)
| | cod_classe_principal = 4
| | | media = 0
| | | | cod_classe_consumo <= 4011: N (6.0)
| | | | cod_classe_consumo > 4011
| | | | dias_de_vida = 0: N (2.0/1.0)
| | | | dias_de_vida = 1600: S (0.0)
| | | | dias_de_vida = 2600: S (4.0)
| | | | dias_de_vida = 4000: N (2.0)
| | | | dias_de_vida = 6100: S (0.0)
| | | | dias_de_vida = 7900: S (0.0)
| | | | dias_de_vida = 8800: S (0.0)
| | | media = 100
| | | | dias_de_vida = 0: N (2.0)
| | | | dias_de_vida = 1600: S (3.0/1.0)
| | | | dias_de_vida = 2600
| | | | precipitacao = 0.0: S (3.0/1.0)
| | | | precipitacao = 100.0: N (2.0)
| | | | precipitacao = 104.4: S (1.0)
| | | | precipitacao = 110.0: N (0.0)
| | | | precipitacao = 112.2: N (0.0)
| | | | precipitacao = 114.0: N (0.0)
| | | | precipitacao = 148.0: N (0.0)
| | | dias_de_vida = 4000: S (1.0)

```

```

| | | | dias_de_vida = 6100: N (0.0)
| | | | dias_de_vida = 7900: N (0.0)
| | | | dias_de_vida = 8800: N (0.0)
| | | media = 150: N (6.0/1.0)
| | | media = 200
| | | | dias_de_vida = 0: S (1.0)
| | | | dias_de_vida = 1600: N (1.0)
| | | | dias_de_vida = 2600
| | | | | maximo = 0: N (0.0)
| | | | | maximo = 150: N (0.0)
| | | | | maximo = 200: N (0.0)
| | | | | maximo = 250: N (0.0)
| | | | | maximo = 320: N (4.0)
| | | | | maximo = 440
| | | | | | cod_localidade <= 692: S (2.0)
| | | | | | cod_localidade > 692: N (2.0)
| | | | | | maximo = 800: S (3.0/1.0)
| | | | dias_de_vida = 4000: N (2.0)
| | | | dias_de_vida = 6100: N (0.0)
| | | | dias_de_vida = 7900: N (0.0)
| | | | dias_de_vida = 8800: N (0.0)
| | | media = 300
| | | | cod_classe_consumo <= 4011: N (10.0/1.0)
| | | | cod_classe_consumo > 4011
| | | | | cod_localidade <= 684: S (4.0/1.0)
| | | | | cod_localidade > 684: N (5.0)
| | | media = 600: N (14.0)
| | cod_classe_principal = 5
| | | altitude = 0: N (4.0)
| | | altitude = 400
| | | | dias_de_vida = 0: S (3.0/1.0)
| | | | dias_de_vida = 1600: N (1.0)
| | | | dias_de_vida = 2600: N (3.0/1.0)
| | | | dias_de_vida = 4000: N (1.0)
| | | | dias_de_vida = 6100: N (0.0)
| | | | dias_de_vida = 7900: N (0.0)
| | | | dias_de_vida = 8800: N (0.0)
| | | altitude = 500: N (0.0)
| | | altitude = 600: N (0.0)
| | | altitude = 615: N (0.0)
| | | altitude = 680: N (0.0)

```

```

| | | altitude = 800: N (0.0)
| | cod_classe_principal = 6: N (2.0)
| | cod_classe_principal = 7: N (4.0)
| | cod_classe_principal = 8: N (0.0)
| subtipo = MV202
| | ano_fabric = 0: N (0.0)
| | ano_fabric = 1980: N (0.0)
| | ano_fabric = 1983: N (0.0)
| | ano_fabric = 1986
| | | temperatura = 0.0: N (0.0)
| | | temperatura = 20.6: N (13.0)
| | | temperatura = 21.5: S (1.0)
| | | temperatura = 21.6: N (0.0)
| | | temperatura = 21.7: N (6.0)
| | | temperatura = 23.1: N (4.0)
| | ano_fabric = 1992
| | | cod_tipo_fase = MO: N (6.0)
| | | cod_tipo_fase = BI
| | | | cod_classe_principal = 1
| | | | | precipitacao = 0.0
| | | | | dias_de_vida = 0
| | | | | | media = 0: N (1.0)
| | | | | | media = 100: N (3.0)
| | | | | | media = 150: N (2.0/1.0)
| | | | | | media = 200: S (1.0)
| | | | | | media = 300: N (0.0)
| | | | | | media = 600: N (0.0)
| | | | | dias_de_vida = 1600
| | | | | | maximo = 0: S (1.0)
| | | | | | maximo = 150: S (2.0)
| | | | | | maximo = 200: N (4.0/1.0)
| | | | | | maximo = 250: N (2.0)
| | | | | | maximo = 320: N (0.0)
| | | | | | maximo = 440: N (2.0/1.0)
| | | | | | maximo = 800: N (0.0)
| | | | | dias_de_vida = 2600
| | | | | | media = 0: S (6.0/1.0)
| | | | | | media = 100
| | | | | | | maximo = 0: N (0.0)
| | | | | | | maximo = 150: N (2.0)
| | | | | | | maximo = 200: N (3.0/1.0)

```

```

| | | | | | | | | maximo = 250: N (0.0)
| | | | | | | | | maximo = 320: S (1.0)
| | | | | | | | | maximo = 440: N (0.0)
| | | | | | | | | maximo = 800: N (0.0)
| | | | | | | | | media = 150: N (3.0/1.0)
| | | | | | | | | media = 200: N (7.0/1.0)
| | | | | | | | | media = 300: S (4.0/1.0)
| | | | | | | | | media = 600: N (1.0)
| | | | | | | | | dias_de_vida = 4000
| | | | | | | | | media = 0: N (6.0)
| | | | | | | | | media = 100: N (2.0)
| | | | | | | | | media = 150: N (3.0)
| | | | | | | | | media = 200: N (2.0/1.0)
| | | | | | | | | media = 300: S (1.0)
| | | | | | | | | media = 600: N (0.0)
| | | | | | | | | dias_de_vida = 6100: N (1.0)
| | | | | | | | | dias_de_vida = 7900: N (0.0)
| | | | | | | | | dias_de_vida = 8800: N (0.0)
| | | | | | | | | precipitacao = 100.0
| | | | | | | | | dias_de_vida = 0: N (0.0)
| | | | | | | | | dias_de_vida = 1600: S (1.0)
| | | | | | | | | dias_de_vida = 2600: N (8.0/1.0)
| | | | | | | | | dias_de_vida = 4000: N (5.0/2.0)
| | | | | | | | | dias_de_vida = 6100: N (0.0)
| | | | | | | | | dias_de_vida = 7900: N (0.0)
| | | | | | | | | dias_de_vida = 8800: N (0.0)
| | | | | | | | | precipitacao = 104.4: N (17.0/3.0)
| | | | | | | | | precipitacao = 110.0: N (0.0)
| | | | | | | | | precipitacao = 112.2: N (0.0)
| | | | | | | | | precipitacao = 114.0: S (1.0)
| | | | | | | | | precipitacao = 148.0: N (0.0)
| | | | | | | | | cod_classe_principal = 2: N (0.0)
| | | | | | | | | cod_classe_principal = 3: N (29.0/3.0)
| | | | | | | | | cod_classe_principal = 4
| | | | | | | | | media = 0: S (3.0/1.0)
| | | | | | | | | media = 100: S (2.0)
| | | | | | | | | media = 150: N (5.0)
| | | | | | | | | media = 200: S (2.0)
| | | | | | | | | media = 300: N (2.0)
| | | | | | | | | media = 600: N (0.0)
| | | | | | | | | cod_classe_principal = 5

```

```

| | | | | maximo = 0: N (2.0/1.0)
| | | | | maximo = 150: N (0.0)
| | | | | maximo = 200: N (0.0)
| | | | | maximo = 250: S (1.0)
| | | | | maximo = 320: N (0.0)
| | | | | maximo = 440: N (4.0)
| | | | | maximo = 800: N (0.0)
| | | | | cod_classe_principal = 6: N (2.0)
| | | | | cod_classe_principal = 7: N (0.0)
| | | | | cod_classe_principal = 8: S (1.0)
| | | | | cod_tipo_fase = TR: N (3.0/1.0)
| | | | | ano_fabric = 1998: N (0.0)
| | | | | subtipo = MY202
| | | | | temperatura = 0.0: N (89.0)
| | | | | temperatura = 20.6: N (17.0)
| | | | | temperatura = 21.5
| | | | | cod_localidade <= 49: S (5.0/2.0)
| | | | | cod_localidade > 49: N (2.0)
| | | | | temperatura = 21.6
| | | | | precipitacao = 0.0: N (0.0)
| | | | | precipitacao = 100.0
| | | | | dias_de_vida = 0: N (8.0/1.0)
| | | | | dias_de_vida = 1600: N (5.0/1.0)
| | | | | dias_de_vida = 2600: N (6.0)
| | | | | dias_de_vida = 4000
| | | | | media = 0: N (3.0)
| | | | | media = 100: N (1.0)
| | | | | media = 150: N (0.0)
| | | | | media = 200: S (7.0/3.0)
| | | | | media = 300: N (2.0/1.0)
| | | | | media = 600: N (1.0)
| | | | | dias_de_vida = 6100: N (1.0)
| | | | | dias_de_vida = 7900: N (1.0)
| | | | | dias_de_vida = 8800: N (1.0)
| | | | | precipitacao = 104.4: N (0.0)
| | | | | precipitacao = 110.0: N (31.0/6.0)
| | | | | precipitacao = 112.2: N (13.0)
| | | | | precipitacao = 114.0: N (0.0)
| | | | | precipitacao = 148.0: N (0.0)
| | | | | temperatura = 21.7
| | | | | cod_tipo_local = RR: N (18.0)

```



```

| | | | ano_fabric = 1992: S (4.0)
| | | | ano_fabric = 1998
| | | | | altitude = 0: N (1.0)
| | | | | altitude = 400: S (3.0/1.0)
| | | | | altitude = 500: N (3.0/1.0)
| | | | | altitude = 600: S (1.0)
| | | | | altitude = 615: N (4.0)
| | | | | altitude = 680: N (3.0)
| | | | | altitude = 800: N (0.0)
| | | | maximo = 150
| | | | | media = 0
| | | | | | cod_localidade <= 221: N (6.0)
| | | | | | cod_localidade > 221: S (3.0/1.0)
| | | | | media = 100
| | | | | | temperatura = 0.0: N (0.0)
| | | | | | temperatura = 20.6: N (2.0)
| | | | | | temperatura = 21.5: N (2.0/1.0)
| | | | | | temperatura = 21.6: N (0.0)
| | | | | | temperatura = 21.7: S (2.0)
| | | | | | temperatura = 23.1: N (0.0)
| | | | | media = 150: N (3.0)
| | | | | media = 200: N (0.0)
| | | | | media = 300: N (0.0)
| | | | | media = 600: N (0.0)
| | | | maximo = 200
| | | | | temperatura = 0.0: N (2.0)
| | | | | temperatura = 20.6: N (7.0)
| | | | | temperatura = 21.5
| | | | | | ano_fabric = 0: N (0.0)
| | | | | | ano_fabric = 1980: N (0.0)
| | | | | | ano_fabric = 1983: N (0.0)
| | | | | | ano_fabric = 1986: N (0.0)
| | | | | | ano_fabric = 1992: N (2.0)
| | | | | | ano_fabric = 1998: S (2.0)
| | | | | temperatura = 21.6: N (3.0)
| | | | | temperatura = 21.7: N (7.0)
| | | | | temperatura = 23.1: N (0.0)
| | | | maximo = 250
| | | | | dias_de_vida = 0: N (0.0)
| | | | | dias_de_vida = 1600
| | | | | | temperatura = 0.0: S (0.0)

```



```

| | | | | | maximo = 800: N (3.0)
| | | | | dias_de_vida = 4000
| | | | | | media = 0: N (2.0)
| | | | | | media = 100: N (0.0)
| | | | | | media = 150
| | | | | | | cod_classe_consumo <= 4011: N (2.0)
| | | | | | | cod_classe_consumo > 4011: S (2.0)
| | | | | | media = 200: N (3.0)
| | | | | | media = 300: N (3.0)
| | | | | | media = 600: N (4.0)
| | | | | dias_de_vida = 6100
| | | | | | media = 0: N (7.0/2.0)
| | | | | | media = 100: N (5.0/1.0)
| | | | | | media = 150: N (1.0)
| | | | | | media = 200
| | | | | | | maximo = 0: N (0.0)
| | | | | | | maximo = 150: N (0.0)
| | | | | | | maximo = 200: N (0.0)
| | | | | | | maximo = 250: N (0.0)
| | | | | | | maximo = 320: N (0.0)
| | | | | | | maximo = 440: S (5.0/2.0)
| | | | | | | maximo = 800: N (3.0/1.0)
| | | | | | media = 300: N (13.0/1.0)
| | | | | | media = 600
| | | | | | | precipitacao = 0.0
| | | | | | | | cod_localidade <= 700: N (5.0/2.0)
| | | | | | | | cod_localidade > 700: S (3.0)
| | | | | | | precipitacao = 100.0: S (0.0)
| | | | | | | precipitacao = 104.4: N (3.0/1.0)
| | | | | | | precipitacao = 110.0: S (0.0)
| | | | | | | precipitacao = 112.2: S (0.0)
| | | | | | | precipitacao = 114.0: S (0.0)
| | | | | | | precipitacao = 148.0: S (0.0)
| | | | | dias_de_vida = 7900
| | | | | | ano_fabric = 0: N (0.0)
| | | | | | ano_fabric = 1980
| | | | | | | maximo = 0: N (1.0)
| | | | | | | maximo = 150: N (0.0)
| | | | | | | maximo = 200: S (1.0)
| | | | | | | maximo = 250: N (0.0)
| | | | | | | maximo = 320: N (1.0)

```

```

| | | | | | maximo = 440: N (2.0/1.0)
| | | | | | maximo = 800: N (4.0)
| | | | | | ano_fabric = 1983: S (3.0)
| | | | | | ano_fabric = 1986: N (0.0)
| | | | | | ano_fabric = 1992: N (0.0)
| | | | | | ano_fabric = 1998: N (0.0)
| | | | | | dias_de_vida = 8800: N (13.0/1.0)
| | | | | | cod_classe_principal = 5: S (1.0)
| | | | | | cod_classe_principal = 6: N (0.0)
| | | | | | cod_classe_principal = 7: N (1.0)
| | | | | | cod_classe_principal = 8: N (0.0)
| | | | | | cod_tipo_local = UB: N (79.0/8.0)
instalacoes = 2
| | | | | | subtipo = D58: N (5.0/1.0)
| | | | | | subtipo = D58C
| | | | | | | | | | | | cod_classe_principal = 1
| | | | | | | | | | | | maximo = 0: N (17.0/1.0)
| | | | | | | | | | | | maximo = 150: N (8.0/1.0)
| | | | | | | | | | | | maximo = 200: N (12.0)
| | | | | | | | | | | | maximo = 250: N (11.0)
| | | | | | | | | | | | maximo = 320: N (3.0/1.0)
| | | | | | | | | | | | maximo = 440: N (6.0/1.0)
| | | | | | | | | | | | maximo = 800: S (1.0)
| | | | | | | | | | | | cod_classe_principal = 2: N (1.0)
| | | | | | | | | | | | cod_classe_principal = 3: N (9.0)
| | | | | | | | | | | | cod_classe_principal = 4: N (2.0/1.0)
| | | | | | | | | | | | cod_classe_principal = 5: N (3.0)
| | | | | | | | | | | | cod_classe_principal = 6: N (0.0)
| | | | | | | | | | | | cod_classe_principal = 7: N (0.0)
| | | | | | | | | | | | cod_classe_principal = 8: N (0.0)
| | | | | | | | | | | | subtipo = D8L
| | | | | | | | | | | | dias_de_vida = 0: N (30.0/2.0)
| | | | | | | | | | | | dias_de_vida = 1600: N (2.0)
| | | | | | | | | | | | dias_de_vida = 2600: S (1.0)
| | | | | | | | | | | | dias_de_vida = 4000: N (0.0)
| | | | | | | | | | | | dias_de_vida = 6100: N (0.0)
| | | | | | | | | | | | dias_de_vida = 7900: N (0.0)
| | | | | | | | | | | | dias_de_vida = 8800: N (0.0)
| | | | | | | | | | | | subtipo = FV201
| | | | | | | | | | | | cod_classe_principal = 1: N (18.0)
| | | | | | | | | | | | cod_classe_principal = 2: N (0.0)

```



```

|   |   |   cod_classe_principal = 3: N (6.0)
|   |   |   cod_classe_principal = 4: N (0.0)
|   |   |   cod_classe_principal = 5: N (0.0)
|   |   |   cod_classe_principal = 6: N (0.0)
|   |   |   cod_classe_principal = 7: N (0.0)
|   |   |   cod_classe_principal = 8: N (0.0)
|   |   media = 100
|   |   |   altitude = 0: N (0.0)
|   |   |   altitude = 400: N (2.0/1.0)
|   |   |   altitude = 500: S (1.0)
|   |   |   altitude = 600: S (1.0)
|   |   |   altitude = 615: N (2.0)
|   |   |   altitude = 680: N (0.0)
|   |   |   altitude = 800: N (0.0)
|   |   media = 150: N (6.0)
|   |   media = 200: N (1.0)
|   |   media = 300: N (6.0)
|   |   media = 600: N (2.0/1.0)
|   subtipo = MY202
|   |   precipitacao = 0.0: S (1.0)
|   |   precipitacao = 100.0: N (6.0)
|   |   precipitacao = 104.4: N (1.0)
|   |   precipitacao = 110.0: N (7.0)
|   |   precipitacao = 112.2: N (8.0)
|   |   precipitacao = 114.0: N (42.0)
|   |   precipitacao = 148.0
|   |   |   media = 0: N (7.0)
|   |   |   media = 100: N (2.0)
|   |   |   media = 150: S (1.0)
|   |   |   media = 200: N (4.0)
|   |   |   media = 300: N (3.0)
|   |   |   media = 600: N (1.0)
|   subtipo = PN5DG
|   |   media = 0
|   |   |   precipitacao = 0.0
|   |   |   |   altitude = 0: N (0.0)
|   |   |   |   altitude = 400: N (1.0)
|   |   |   |   altitude = 500: S (2.0)
|   |   |   |   altitude = 600: N (0.0)
|   |   |   |   altitude = 615: N (4.0/1.0)
|   |   |   |   altitude = 680: N (0.0)

```

```

| | | | altitude = 800: N (0.0)
| | | precipitacao = 100.0
| | | | maximo = 0: N (4.0)
| | | | maximo = 150: S (3.0/1.0)
| | | | maximo = 200: N (0.0)
| | | | maximo = 250: N (0.0)
| | | | maximo = 320: N (0.0)
| | | | maximo = 440: N (0.0)
| | | | maximo = 800: N (0.0)
| | | precipitacao = 104.4: N (6.0)
| | | precipitacao = 110.0: N (0.0)
| | | precipitacao = 112.2: N (0.0)
| | | precipitacao = 114.0: N (0.0)
| | | precipitacao = 148.0: N (0.0)
| | media = 100: N (11.0/1.0)
| | media = 150: N (3.0)
| | media = 200: N (7.0)
| | media = 300: N (2.0)
| | media = 600: N (1.0)
| subtipo = T8L: N (17.0)
instalacoes = 3
| subtipo = D58: N (1.0)
| subtipo = D58C
| | precipitacao = 0.0: N (1.0)
| | precipitacao = 100.0: N (0.0)
| | precipitacao = 104.4: N (3.0)
| | precipitacao = 110.0: N (6.0)
| | precipitacao = 112.2
| | | media = 0: S (3.0/1.0)
| | | media = 100: N (0.0)
| | | media = 150: N (0.0)
| | | media = 200: N (2.0)
| | | media = 300: N (0.0)
| | | media = 600: N (0.0)
| | precipitacao = 114.0: N (4.0/1.0)
| | precipitacao = 148.0: N (1.0)
| subtipo = D8L: N (10.0)
| subtipo = FV201: N (7.0)
| subtipo = M1A: N (4.0)
| subtipo = M8L
| | maximo = 0: N (4.0)

```

```

| | maximo = 150: S (1.0)
| | maximo = 200: N (2.0)
| | maximo = 250: N (0.0)
| | maximo = 320: N (0.0)
| | maximo = 440: N (0.0)
| | maximo = 800: N (0.0)
| subtipo = MFT120G: N (4.0)
| subtipo = MV202: N (14.0/2.0)
| subtipo = MY202: N (34.0/1.0)
| subtipo = PN5DG
| | temperatura = 0.0: N (2.0)
| | temperatura = 20.6: N (4.0)
| | temperatura = 21.5: S (1.0)
| | temperatura = 21.6: S (1.0)
| | temperatura = 21.7
| | | cod_localidade <= 220: N (2.0)
| | | cod_localidade > 220: S (2.0)
| | temperatura = 23.1: N (0.0)
| subtipo = T8L: N (7.0)
instalacoes = 4
| ano_fabric = 0: N (2.0/1.0)
| ano_fabric = 1980: N (4.0)
| ano_fabric = 1983: N (5.0)
| ano_fabric = 1986: N (3.0)
| ano_fabric = 1992: S (1.0)
| ano_fabric = 1998: N (5.0/1.0)
instalacoes = 5: N (6.0)
instalacoes = 6: N (2.0)

```

Number of Leaves : 1115

Size of the tree : 1335