

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

FELIPE BISCAGLIA LACERDA

**DETERMINAÇÃO DE PONTOS CRÍTICOS DE AVARIAS EM
UNIDADES TRANSFORMADORAS ATRAVÉS DE DESCOBERTA
DE CONHECIMENTO EM TEXTO**

Porto Alegre

2014

FELIPE BISCAGLIA LACERDA

**DETERMINAÇÃO DE PONTOS CRÍTICOS DE AVARIAS EM
UNIDADES TRANSFORMADORAS ATRAVÉS DE DESCOBERTA
DE CONHECIMENTO EM TEXTO**

Projeto de Diplomação apresentado ao Departamento de Engenharia Elétrica da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para Graduação em Engenharia Elétrica.

Orientador: Prof. Dr. Flávio Antônio Becon Lemos

Porto Alegre

2014

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

FELIPE BISCAGLIA LACERDA

**DETERMINAÇÃO DE PONTOS CRÍTICOS DE AVARIAS EM
UNIDADES TRANSFORMADORAS ATRAVÉS DE DESCOBERTA
DE CONHECIMENTO EM TEXTO**

Este trabalho de conclusão foi analisado e julgado adequado para a obtenção da Graduação em Engenharia Elétrica e aprovado em sua forma final pelo Orientador e pela Banca Examinadora da Universidade Federal do Rio Grande do Sul.

**Prof. Dr. Flávio Antônio Becon
Lemos**
Orientador

Prof. Dr. Altamiro Amadeu Susin
Coordenador

Apresentação em: 03/07/2014 às 16h.

BANCA EXAMINADORA

Flávio Antônio Becon Lemos (Doutor) - UFRGS _____

Carlos Oliva Pretto (Doutor) - UFPEL _____

Sérgio Luís Haffner (Doutor) - UFRGS _____

*Dedico este trabalho à minha filha
Bibiana, razão da minha vida.*

AGRADECIMENTOS

À minha mãe Gladis e minha irmã Flávia pelo apoio durante esta etapa.

Aos profissionais do Departamento de Proteção e Qualidade da Distribuição da Companhia Estadual de Distribuição de Energia Elétrica pelo profissionalismo.

Ao Prof. Dr. Flávio Antônio Becon Lemos pela orientação neste trabalho.

RESUMO

A interrupção no fornecimento da distribuição de energia elétrica é um grande transtorno tanto para os clientes quanto para as concessionárias responsáveis. Vários desses eventos são acarretados por avarias em transformadores de média tensão, fatos que requerem ações de manutenção demoradas e dispendiosas. À vista disso, o registro de tais acontecimentos é imprescindível para o emprego de futuras medidas preventivas. No entanto, esses apontamentos tornam-se incapazes de claras interpretações a partir do momento que transfiguram à grandes volumes de dados armazenados. Este trabalho propõe a análise de inconsistências nos registros de banco de dados de ocorrências encerradas de avarias em transformadores e especificação de níveis de criticidade (normal, crítico e muito crítico) inerentes à localização física das instalações, através da análise de reincidências. Com base em Descoberta de Conhecimento em Textos (*Knowledge-Discovery in Text* ou KDT) e algoritmo de aprendizagem por Redes Bayesianas, elaborou-se um meio de extrair informações presentes nas observações das ocorrências encerradas de maneira a reconhecer casos de avarias. A partir disso, a utilização de uma árvore de decisão a fim de especificar o quão crítico são os locais de instalação desses equipamentos. Os treinamentos dos algoritmos são apresentados através de métricas de qualidade (matrizes de confusão, curva ROC, coeficiente Kappa) e os resultados disponibilizados em gráficos e mapas dos pontos classificados representando o conhecimento extraído.

Palavras-chave: transformador. avaria. mineração de texto.

ABSTRACT

The outage in the electric power supply is unsuitable for customers and distribution utilities. Several forced outages are due to medium voltage transformers failures, which requires maintenance and costly actions. . Therefore, the recording of such events is essential to future preventive measures. However, these records become unable to support correct interpretations when large volumes of data are stored. This work proposes to examine inconsistencies in event records of database containing information about transformer damages and specify severity levels (normal, critical and very critical) inherent to the installation location of transformers in the network by analyzing recurrences. Based on Knowledge Discovery in Text (KDT) and learning algorithm by Bayesian networks, the methodology implements a method of extracting information contained in the occurrence in order to recognize instances of damage. After that, the methodology specifies how critical it's the location of equipment by a decision tree. The training algorithms are defined by quality metrics (confusion matrices, ROC curve, kappa coefficient) and the results shown in graphs and maps with points ranked representing the extracted knowledge.

Keywords: transformer. damage. text mining.

LISTA DE ILUSTRAÇÕES

Figura 1 – Mapa das UTs da concessionária CEEE-D	23
Figura 2 – Principais causas não programadas de avarias em transformadores na CEEE-D	25
Figura 3 – Estrutura para TR de linha aérea	29
Figura 4 – Componentes de um equipamento transformador trifásico .	31
Figura 5 – Chave fusível de distribuição tipo C	33
Figura 6 – Elo fusível de distribuição tipo botão	33
Figura 7 – Elo fusível tipo H	34
Figura 8 – Para-raios de MT	34
Figura 9 – Detalhes do cadastro de uma ocorrência encerrada	37
Figura 10 – Planilha de controle de avarias da oficina	38
Figura 11 – Custo por gigabyte de disco rígido nos últimos 30 anos . .	43
Figura 12 – A mineração de dados como uma etapa no processo de descoberta de conhecimento	45
Figura 13 – As etapas de (a) aprendizado e (b) teste para do processo de classificação	48
Figura 14 – Um modelo de classificação representado em várias formas, tais como: (a) regras SE-ENTÃO, (b) árvore de decisão, ou (c) rede neural	49
Figura 15 – Exemplo genérico diagnóstico-causa	50
Figura 16 – Diagnóstico genérico intercausal de efeito	50
Figura 17 – Diagnóstico genérico intercausal de causa	50
Figura 18 – Ilustração da validação cruzada para 10 partições	52
Figura 19 – Exemplo de Curva ROC	54
Figura 20 – Árvore de decisão para avaliação de pontos críticos	56
Figura 21 – Tela inicial da ferramenta WEKA	57
Figura 22 – Guia de pré-processamento de dados da aplicação <i>Weka Explorer</i>	58
Figura 23 – Análise de concordância de palavras via software AntConc	59
Figura 24 – Curva ROC para o primeiro treino de classificação	62
Figura 25 – Árvore de decisão para o classificador J48	65
Figura 26 – Verificação qualitativa das relações entre variáveis de texto	65
Figura 27 – Comparativo dos registros de avarias com código 304	68
Figura 28 – Comparativo dos falsos registros de avarias dados pelo código 304	69

Figura 29 –Quantidades de (a) UTs reincidentes e (b) UTs rolutadas
normais, críticas e muito críticas 70

Figura 30 –Classificações normais (verdes), críticas (amarelas) e muito
críticas (vermelhas) 71

LISTA DE TABELAS

Tabela 1 – Padrão de TRs da CEEE-D	30
Tabela 2 – Exemplo de matriz de confusão	52
Tabela 3 – Interpretação dos coeficientes Kappa	55
Tabela 4 – Lista de palavras equivalentes	60
Tabela 5 – Etapas de limpeza e integração dos dados	60
Tabela 6 – Amostra da tupla para a ocorrência 2014-2421	61
Tabela 7 – Comparativo entre classificações da oficina (OF), Rede Bayesiana (BN) e assistida (AS)	63
Tabela 8 – Matriz de confusão entre registros da oficina e classificados por BN para janeiro de 2014	67
Tabela 9 – Matriz de confusão entre registros com serviço 304 e classificados por BN para janeiro de 2014	67
Tabela 10 – Matriz de confusão entre registros com serviço 304 e classificados por BN para janeiro de 2013	68
Tabela 11 – Matriz de confusão entre registros com serviço 304 e classificados por BN para março de 2014	68
Tabela 12 – Matriz de confusão entre registros com serviço 304 e classificados por BN para abril de 2014	68
Tabela 13 – Comparativo entre verdadeiros positivos (VP), falsos positivos (FP), falsos negativos (FN) e verdadeiros negativos (VN)	69
Tabela 14 – Casos de avarias ocorridas em curtos períodos de tempo	70
Tabela 15 – Exemplos de observações das ocorrências encerradas	81
Tabela 16 – Códigos de causas para ocorrências encerradas	85

LISTA DE CÓDIGOS

Código 1 – Amostra de tokens resultantes	61
Código 2 – Métricas de qualidade do primeiro treino de classificação .	61
Código 3 – Métricas de qualidade do segundo treino de classificação .	63
Código 4 – Comparativo entre algoritmos classificadores: (1) BayesNet padrão, (2) BayesNet com máximo de dois pais, (3) DMNBtext, (4) J48 e (5) PART	64
Código 5 – Métricas de qualidade do terceiro treino de validação cruzada	66
Código 6 – Métricas de qualidade do primeiro treino de classificação (resultado completo)	79
Código 7 – Métricas de qualidade do segundo treino de classificação (resultado completo)	79
Código 8 – Métricas de qualidade do terceiro treino de classificação (resultado completo)	80
Código 9 – Comparativo entre algoritmos classificadores (resultado com- pleto)	80

NOMENCLATURA

ANEEL	Agência Nacional de Energia Elétrica
BN	<i>Bayesian Network</i>
BT	Baixa Tensão
CA	Corrente Alternada
CEEE-D	Companhia Estadual de Distribuição de Energia Elétrica
CM	<i>Confusion Matrix</i>
CV	<i>Cross-Validation</i>
GRM	Gerência Regional Metropolitana
KDD	<i>Knowledge Discovery in Databases</i>
KDT	<i>Knowledge Discovery in Texts</i>
MT	Média tensão
OE	Ocorrências Encerradas
OS	Ordem de Serviço
OSD	Operador de Sistema de Distribuição
PDA	<i>Personal Digital Assistant</i>
PRAT	Programa de Redução de Avarias em Transformadores
ROC	<i>Receiver Operator Characteristics</i>
SGD	Sistema de Gestão da Distribuição
SGS	Sistema de Gestão de Serviços
TR	Transformador
UT	Unidade Transformadora
VHF	<i>Very High Frequency</i>
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

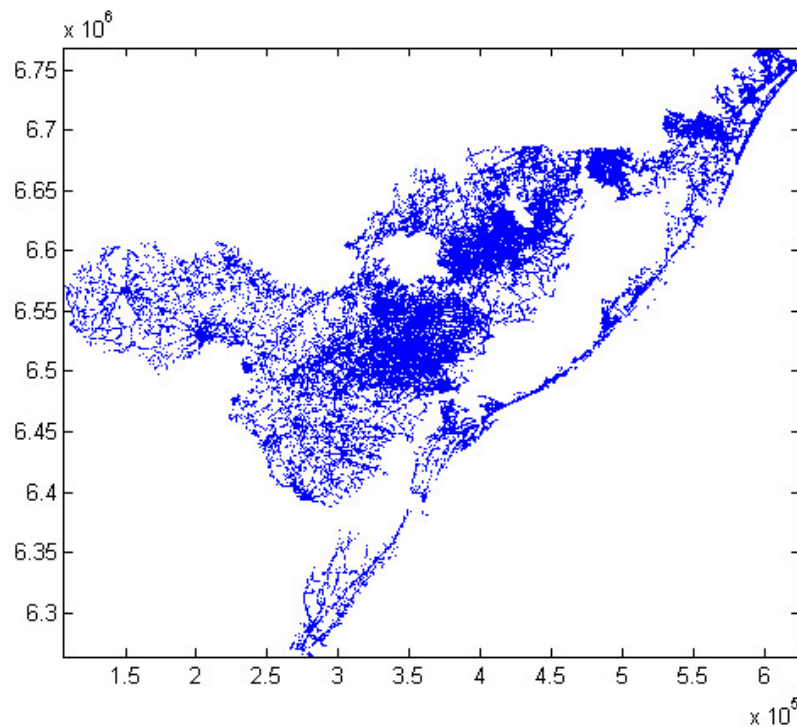
SUMÁRIO

1	INTRODUÇÃO	23
1.1	CARACTERIZAÇÃO DO PROBLEMA	24
1.2	OBJETIVO	26
1.3	CONCESSIONÁRIA CEEE-D	27
2	UNIDADE TRANSFORMADORA	29
2.1	EQUIPAMENTO TRANSFORMADOR	30
2.2	PROTEÇÕES	32
3	REGISTRO DE AVARIAS	35
3.1	OCORRÊNCIAS ENCERRADAS	35
3.2	PLANILHAS DAS OFICINAS	38
3.3	PRINCIPAIS CAUSAS	39
4	DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS	43
4.1	MINERAÇÃO DE DADOS	45
4.2	MINERAÇÃO DE TEXTOS	46
4.3	CLASSIFICAÇÃO	47
4.4	ALGORITMOS DE APRENDIZAGEM	48
4.5	AVALIAÇÃO DO PODER DE PREDIÇÃO	51
4.6	MATRIZ DE CONFUSÃO E MÉTRICAS DE QUALIDADE	52
4.6.1	Curva ROC	54
4.6.2	Coefficiente Kappa	55
4.7	AVALIAÇÃO DE PONTOS CRÍTICOS	55
4.8	FERRAMENTA WEKA	56
5	ESTUDO DE CASO: GRM	59
5.1	TRATAMENTO DOS DADOS	59
5.2	CLASSIFICAÇÃO	61
5.3	APLICAÇÃO DO MODELO	66
5.4	DETERMINAÇÃO DOS PONTOS CRÍTICOS	69
6	CONCLUSÕES	73
	REFERÊNCIAS	75
	APÊNDICE A RESULTADOS COMPLETOS DOS TREINOS	79
	APÊNDICE B OBSERVAÇÕES TEXTUAIS	81
	ANEXO A CÓDIGOS DE CAUSAS	85

1 INTRODUÇÃO

Utilizados aos milhares, o transformador é um equipamento elementar na distribuição de energia elétrica. A Figura 1 representa a região de concessão da distribuidora Companhia Estadual de Distribuição de Energia Elétrica (CEEE-D) com 73.308 pontos de Unidades Transformadoras (UT) em coordenadas UTM (*Universal Transversa de Mercator*). Entende-se por UT o conjunto de equipamentos físicos (transformador, chaves, poste, para-raios, etc.) e informações (coordenadas geográficas, localidade, zona, alimentador, etc.) dos locais de instalação dos transformadores de distribuição.

Figura 1 – Mapa das UTs da concessionária CEEE-D



Materiais de grande robustez são utilizados na fabricação desses equipamentos, com isso, torna-se maior a resistência às deteriorações causadas por elementos da natureza e fadigas do sistema. Manutenções preventivas regulares aliadas a um bom controle das cargas atendidas fazem com que o equipamento prolongue sua vida útil em dezenas de anos. Nos casos de avarias, o custo direto da troca do equipamento transformador é apenas uma fração no gravame da concessionária. Quanto a mobilização de recursos humanos, tem-se o deslocamento de equipes emergenciais, as quais possuem maior ônus em relação as programadas, além da sobrecarga nos canais de atendimento ao consumidor, seja esses em lojas físicas seja em canais de atendimento.

A inesperada falta de energia elétrica dada pelo imprevisto é inoportuna, deixa os clientes insatisfeitos e reflete no Índice Aneel de Satisfação do Consumidor (IASC) além de afetar os indicadores de qualidade do serviço definidos pela ANEEL – Agência Nacional de Energia Elétrica (2010, p. 39):

- DEC Duração Equivalente de Interrupção por Unidade Consumidora – indica o número de horas que, em média, as unidades consumidoras de determinado conjunto ficaram sem energia elétrica durante um determinado período: mensal, trimestral ou anual;
- FEC Frequência Equivalente de Interrupção por Unidade Consumidora – indica quantas vezes, em média, as unidades consumidoras de determinado conjunto sofreram interrupção;
- DIC Duração de Interrupção Individual por Unidade Consumidora – quantidade de horas que o consumidor ficou sem energia elétrica;
- FIC Frequência de Interrupção Individual por Unidade Consumidora – quantidade de interrupções que o consumidor experimentou no período de apuração (mensal, trimestral ou anual);
- DMIC Duração Máxima de Interrupção Contínua por Unidade Consumidora – indica o número de horas da maior interrupção experimentada pelo consumidor no período de apuração.

Caso a frequência das faltas torne-se relevante, o desgaste da imagem da empresa pode ser irreparável a ponto de, por força do hábito, ser erroneamente implicada à autoria de quaisquer outros problemas de energia dos quais não possui relação direta. Evitar esses acontecimentos é mais do que esquivar-se de simples transtornos ou economizar custos, trata-se de garantir os três princípios básicos do Código de Defesa do Consumidor no que se refere a prestação de serviços essenciais: qualidade, continuidade e universalidade. A identificação e análise de causas de avarias em equipamentos transformadores é mais um artifício no auxílio do controle da operação perante essas medidas. Através da investigação dos problemas sucedidos, a análise dos dados dos registros de ocorrências são relacionados de forma a diagnosticar os casos mais recorrentes e assim embasar futuros projetos e ou manutenções preventivas.

1.1 CARACTERIZAÇÃO DO PROBLEMA

Segundo o Programa de Redução de Avarias em Transformadores (PRAT) da CEEE-D (2009), para levantamentos de dados e análise das anormalidades e/ou ocorrências nos equipamentos transformadores, o procedimento elementar é a elaboração de

um relatório de consulta às ocorrências encerradas que possuam o serviço 304. Por esse código, define-se

304 - BT - POSTO DE TRANSFORMADOR: TRANSFORMADOR AVARIADO

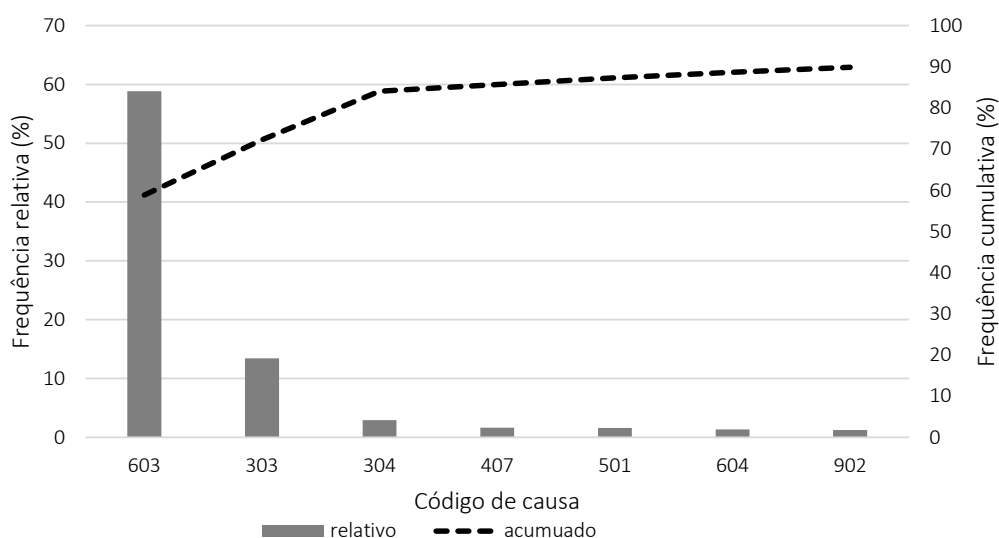
a troca do equipamento transformador em estado de avaria¹. Essa nomenclatura clara e intuitiva é vigente desde abril de 2014 (durante a realização deste trabalho). Anteriormente, registrada como

304 - BT - POSTO DE TRANSFORMADOR: TRANSFORMADOR

era condição de múltiplas interpretações, assinalada desde serviços de medições de tensão à trocas de elos fusíveis. Conseqüentemente, a resolução definida pelo PRAT é comprometida em função dessa inconsistência. Dada a dificuldade, muitos dos registros são pesquisados e analisados individualmente, com consideração aos campos de observações. Esses citam fatos ocorridos durante o atendimento, do mesmo modo que são utilizados para gravar os códigos dos equipamentos substituídos. É evidente o dispêndio de tempo e de recursos humanos, visto que acontecimentos tão importantes não tem (ou não tinham) sólidos itens de preenchimento. Esse caracteriza o primeiro problema abordado neste trabalho.

Em pesquisa ao banco de dados da CEEE-D, nos registros de ocorrências encerradas (OE), que possuem o serviço de troca de transformador (serviço 304) de modo não programado, cerca de 60% das causas são atribuídas aos próprios equipamentos (causa 603), seguidos de 14% descargas elétricas (causa 303), 4% vegetais (causa 304), 2% sobrecarga (causa 501) entre outros (códigos disponíveis no Apêndice A, pag. 85). A Figura 2 mostra o histograma da frequência das causas não programadas com os respectivos códigos. Os dados abrangem toda região da distribuidora nos de 2009 a 2013.

Figura 2 – Principais causas não programadas de avarias em transformadores na CEEE-D



¹ Avariado: Sem condições de operação (CEEE-D, 2006, p. 3)

Dentre as causas de avarias em transformadores, pode-se citar algumas que são evidentes pela saliência de suas consequências, logo dispensam análises mais aprofundadas: colisão de veículos, animais, incêndio, objetos na rede. Para outras, leva-se em consideração itens subjetivos como a experiência e/ou empenho do técnico, fatos não observáveis na ocasião (vento, descarga atmosférica) e históricos de atendimento (casos de reincidências).

A Figura 2 constata uma disparidade entre as atribuições, principalmente em relação à causa 603 - atribuição ao equipamento transformador. Pela análise de históricos, averigua-se múltiplos casos de reincidências desse código, no mesmo local, em curtos períodos de tempo. Fica evidente que, nesses casos, têm-se um fator inerente à UT, ou seja, às condições as quais os equipamentos são submetidos no local.

Quanto à determinação de causas, constata-se nitidamente a pouca confiabilidade em muitos registros, seja por desatender o histórico de ocorrências, seja por falta de instrução no preenchimento. Esse define o segundo problema tratado neste trabalho.

1.2 OBJETIVO

Devido às grandes quantidades de informações armazenadas, o uso da tecnologia se torna indispensável. O aparecimento de técnicas de inteligência artificial, através algoritmos de aprendizado, surge em um passado recente. O trabalho de Pretto et al. (2006), dá-se a proposta de coleta de dados pré-determinados no local do evento utilizando computação móvel e, através de classificadores de arvores de decisão e matriz pertinência, são determinadas as causas de desligamentos nas redes de distribuição. Com mesmo intuito, mas devido a indisponibilidade de sistemas automatizados de coleta, Tronchoni et al. (2010) estabelece análises através de Redes Bayesianas, em um banco de dados contendo históricos de desligamentos. Ambos extraem informações úteis através do processo de Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases* ou KDD).

O propósito deste trabalho é 1) quantificar as incertezas inerentes à quantidade de transformadores avariados na concessionária de distribuição CEEE-D e, a partir disso, 2) reiterar as causas de avarias em unidades transformadoras, especialmente às falhas conferidas aos próprios equipamentos transformadores (causa 603). Sem grandes pormenores, o que se propõe não é estabelecer uma nova classificação a ponto de gerar/utilizar um código de causa pré-definida com características de avarias, e sim qualitativa no sentido de apontar os locais vulneráveis à essas ocorrências, assim torná-los suscetíveis de assistência continuada. Para esse fim, considera-se as informações já disponíveis no banco de dados da empresa, relativos aos registros de OE, acrescidos de planilhas fornecidas por suas oficinas. Isso é realizado utilizando 1) o processo de Descoberta de Conhecimento

em Textos (*Knowledge Discovery in Texts* ou KDT) a fim de encontrar padrões nos preenchimentos, com algoritmo de aprendizado de Redes Bayesianas, nos conteúdos textuais presentes nos campos de observações. Concluída essa etapa, a 2) reiteração das causas é concebida através da imposição de uma árvore de decisão empírica, a qual determina o grau de criticidade do ponto, considerado seu histórico de avarias.

1.3 CONCESSIONÁRIA CEEE-D

A Companhia Estadual de Distribuição de Energia Elétrica – CEEE-D é uma empresa de economia mista pertencente ao Grupo CEEE, concessionária dos serviços de distribuição de energia elétrica na região sul-sudeste do Estado do Rio Grande do Sul. Com área de concessão que compreende a região Metropolitana, Sul, Litoral e Campanha gaúcha, a CEEE Distribuição atende a 72 municípios, abrangendo 73.627 km², o que corresponde aproximadamente a 34% do mercado consumidor do Rio Grande do Sul (CEEE-D, 2014a).

A CEEE Distribuição atendeu, em 2013, um total de 1,6 milhão de unidades consumidoras, o que equivale a cerca de 4,8 milhões de pessoas ou um terço da população gaúcha, distribuindo diretamente 9.976 GWh.

Devida a grande extensão de sua concessão, a CEEE-D está dividida em Gerências Regionais, que são unidades administrativas situadas em cidades estratégicas as quais abrangem os municípios circunvizinhos. São elas:

GRM	Gerência Regional Metropolitana
GRLN	Gerência Regional Litoral Norte
GRS	Gerência Regional Sul
CRCS	Centro Regional Centro Sul
CRC	Centro Regional da Campanha
CRLS	Centro Regional Litoral Sul

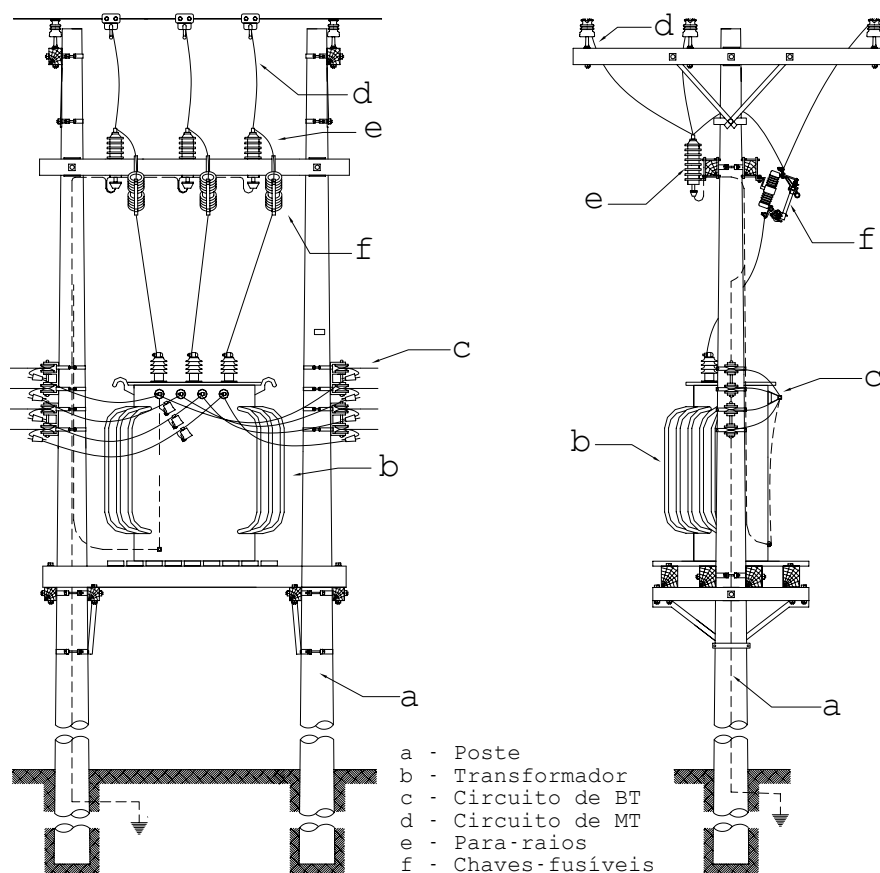
Dos grandes centros urbanos às pequenas aglomerações citadinas tanto quanto localidades rurais, sua vastidão alonga-se a:

- Extensão de rede primária (Urbana e Rural): 36.658 km;
- Extensão de rede secundária (Urbana e Rural): 17.973 km;
- Postes: 760.425;
- Transformadores de distribuição: 59.295.

2 UNIDADE TRANSFORMADORA

Para compreender a composição da unidade transformadora, a Figura 3 mostra um exemplo de padrão de estruturas para equipamentos instalados em plataforma. Essa configuração é comumente utilizada, trata-se de uma UT para distribuição de linha aérea. Essencialmente suportada por postes de madeira, concreto ou ferro, possui no centro o essencial equipamento transformador (TR), o qual fica próximo do circuito de baixa tensão de distribuição¹ (BT) (ou secundário). Entre o circuito de média tensão de distribuição² (MT) e o equipamento principal, tem-se as proteções de para-raios e chaves fusíveis. As demais partes móveis são formadas por ferragens, grampos e tábuas com um total final de 36 itens.

Figura 3 – Estrutura para TR de linha aérea



Fonte (CEEE-D, 2008, s. 2-2)

¹ Tensão entre fases cujo valor eficaz é igual ou inferior a 1 kV (ANEEL, 2012, p.19).

² Tensão entre fases cujo valor eficaz é superior a 1 kV e inferior a 69 kV (ANEEL, 2012, p.23).

2.1 EQUIPAMENTO TRANSFORMADOR

Constituído basicamente por bobinas de fios enroladas em torno de um núcleo ferromagnético comum, onde a única interação é dada por meio da ação de um campo magnético. Consiste em um mecanismo que converte a energia de corrente elétrica alternada (CA) com certa frequência, para energia elétrica CA, na mesma frequência porém com diferente nível de tensão. O rebaixamento ou elevação dessa tensão aumenta ou reduz a corrente do circuito de modo que não se altere a potência desse. Os transformadores de potência recebem diversos nomes, dependendo de como são utilizados nos sistemas de potência. Para o transformador abordado, tem-se ligado no enrolamento primário ou enrolamento de entrada a rede de MT, e no enrolamento secundário ou enrolamento de saída uma rede de BT, logo é denominado de transformador de distribuição.

No sistema de distribuição, assim como na transmissão, as perdas são proporcionais ao quadrado da corrente que percorre os condutores. Desse modo, com a utilização de transformadores, a elevação da tensão distribuída por um fator de 10 permitirá reduzir as perdas elétricas em 100 vezes em razão da diminuição das correntes pelo mesmo fator (CHAPMAN; LASCHUK, 2013, p. 66). Com isso há uma melhora significativa na queda de tensão ao longo do circuito, eliminando restrições de alcance e de capacidade do sistema, por consequência uma distribuição de melhor qualidade a longas distâncias. Assim, a energia elétrica CA pode ser concentrada em subestações, em seguida distribuída com baixas perdas, e finalmente sua tensão é rebaixada novamente no destino final. Para concessionária CEEE-D, tem-se em uso os transformadores com as especificações fundamentais mostradas na Tabela 1.

Tabela 1 – Padrão de TRs da CEEE-D

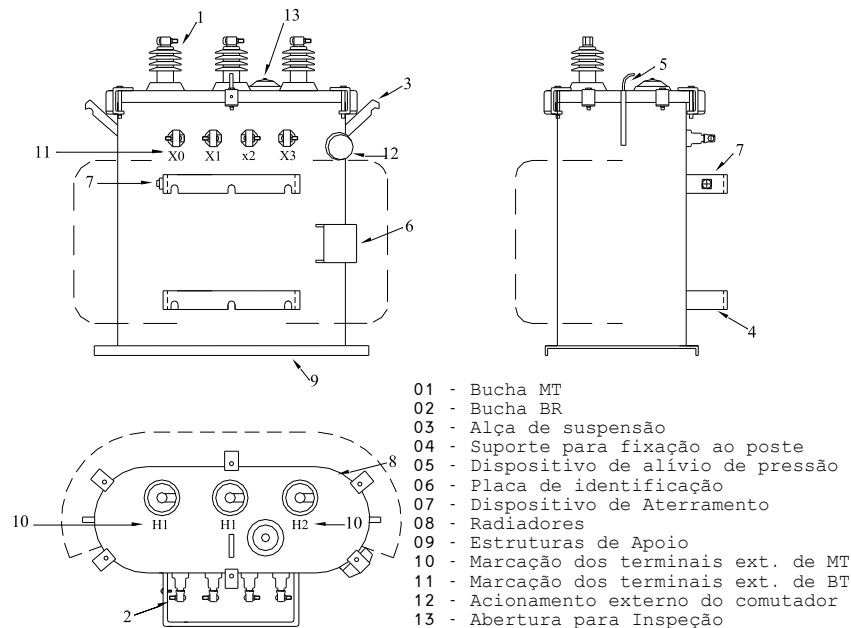
Trifásicos Potência (kVA)	Monofásicos Potência (kVA)	Média Tensão Nominal (V)	Baixa Tensão Nominal (V)	Enrolamento	Isolamento
15, 30, 45, 75, 112,5, 150, 225, 300 e 500	10, 15 e 25	13,8 ou 23,1	220/127 ou 380/220	Cobre	Óleo isolante

Fonte: (CEEE-D, 2013a)

COMPONENTES

Segundo a norma que estabelece o padrão dos materiais para redes aéreas de distribuição (CEEE-D, 2013c, s. 10-16d), tem-se as partes fundamentais de um TR trifásico conforme a Figura 4.

Figura 4 – Componentes de um equipamento transformador trifásico



Fonte: (CEEE-D, 2013c, s. 10-16d)

Os principais componentes são definidos como:

Bucha Componente de porcelana destinado ao isolamento elétrico, que assegura a passagem isolada de um condutor através de uma parede não isolante e de evitar a entrada de elementos contaminadores.

Comutador Dispositivo que permite alterar a relação de transformação do transformador que pode ser interno ou externo ao equipamento.

Enrolamento Conjunto das espiras que constituem um circuito elétrico de um transformador.

Junta Elastômero resistente a ação do óleo aquecido, à ação da umidade e dos raios solares, que serve como vedante.

Radiador Dispositivo que aumenta a superfície de irradiação, para facilitar a dissipação do calor. Normalmente constituído de tubos ou chapas metálicas dispostos na lateral do equipamento.

Tanque Recipiente destinado a conter a parte ativa e o meio isolante do transformador.

Terminal Parte condutora de transformador destinado a sua ligação elétrica a um circuito externo.

Essas peças, muitas vezes relatadas nas observações dos atendimentos, são relevantes na determinação do cenário da avaria, especialmente onde a causa é o próprio

transformador. Entre os defeitos recorrentes destacam-se: vazamento de óleo, quebra da bucha e falha na comutação.

2.2 PROTEÇÕES

A corrente de elevada intensidade que ocorre durante um curto-circuito pode trazer efeitos mecânicos e térmicos aos equipamentos ligados ao circuito sob falha. Os efeitos mecânicos, cujas forças são proporcionais ao quadrado da corrente instantânea, podem deformar condutores e romper materiais isolantes. Já os efeitos térmicos estão ligados ao tempo de permanência da falha e ao valor eficaz da corrente e podem produzir um aquecimento excessivo dos materiais condutores e isolantes, com sua consequente deterioração. Para minimizar os efeitos produzidos pelas correntes de falhas, é comum o uso dos seguintes equipamentos:

- Chaves fusíveis
- Para-raios
- Aterramento

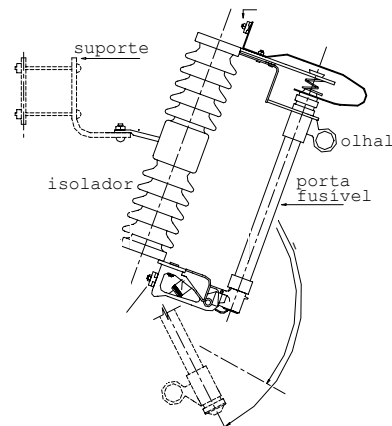
Chaves fusíveis

Um fusível em um circuito elétrico vigia e protege o circuito e equipamentos nele inseridos de danos que podem ser causados pelas sobrecorrentes dentro dos limites nominais. A forma como o fusível executa suas funções depende não apenas da boa fabricação, mas também de sua correta aplicação e da atenção que recebe depois de instalado (ABNT, 2011, p. 31). A Figura 5 mostra as partes componentes fundamentais de uma chave fusível de distribuição tipo expulsão de abertura automática (Tipo C) .

Por ocasião da circulação de sobrecorrente em uma chave fusível, devido ao efeito térmico, o elemento fusível se funde, interrompendo o circuito. A alta temperatura do arco provoca a queima e a decomposição parcial do revestimento interno do cartucho, gerando gases que interrompem o arco no instante de corrente nula (ABNT, 2011, p. 60). A pressão dentro do cartucho aumenta em função dos incrementos de temperatura e a geração dos gases cria condições dentro do tubo que ajudam a desionizar o caminho do arco. A pressão exercida também ajuda a manter a condição de circuito aberto, uma vez que as partículas ionizadas forçam a abertura das extremidades do cartucho, sendo expelidas em seguida.

Os transformadores de distribuição devem ser protegidos através de chaves fusíveis de distribuição com base do tipo C, com porta-fusível de 100A e capacidade de interrupção assimétrica mínima de 10kA para tensão de 15kV e 6,3kA para a tensão de 23kV (CEEE-D, 2013b, p. 21).

Figura 5 – Chave fusível de distribuição tipo C

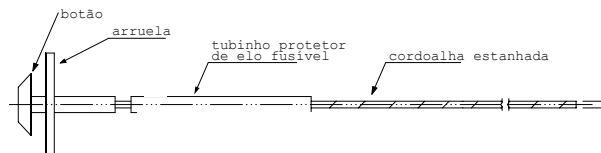


Fonte:(CEEE-D, 2013c, s. 10-13)

Elo fusível

O elo fusível é composto pelo elemento fusível, tubinho, cordoalha, botão, tubo protetor e arruela (CEEE-D, 2008, p. 6), conforme Figura 6.

Figura 6 – Elo fusível de distribuição tipo botão



Fonte:(CEEE-D, 2013c, s. 10-10)

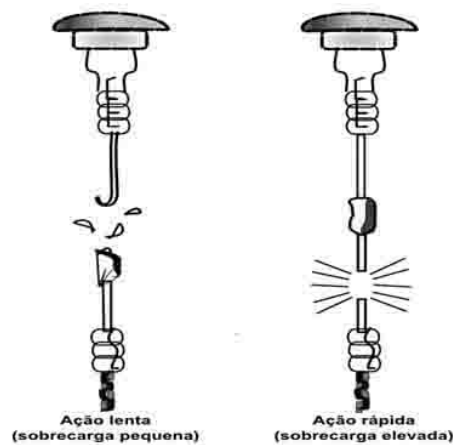
Os fusíveis tipo H são ditos de alto surto. São utilizados na proteção de transformadores, pois suportam as correntes de partida de motores e da energização de transformadores. Esses fusíveis são constituídos de dois elementos fusíveis em série: um de ação rápida e outro de ação lenta. Os elos tipo H possuem as seguintes bitolas: 1H, 2H, 3H e 5H. A Figura 7 ilustra a diferença entre a ocorrência de uma ação lenta e uma ação rápida.

É aconselhável substituir todos os três elos fusíveis quando o elo fusível de uma ou duas fases de um circuito trifásico operou, a menos que haja certeza de que nenhuma sobrecorrente passou pelo elo fusível não fundido (ABNT, 2011, p. 34).

Para-raios e Aterramento

Dispositivo elétrico que protege equipamento elétricos contra sobretensões transitórias elevadas e limita a duração e frequentemente a intensidade da corrente subsequente

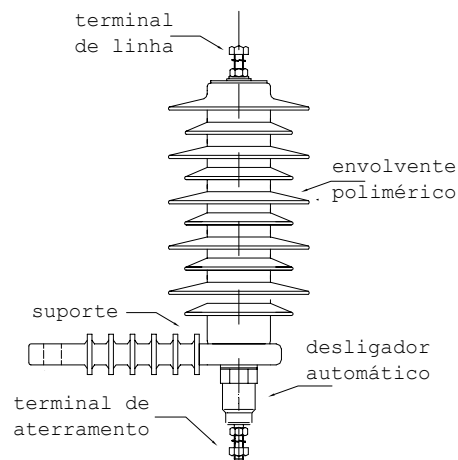
Figura 7 – Elo fusível tipo H



Fonte:(CEEE-D, 2008, p. 7)

o valor da corrente subsequente. De uso obrigatório em toda unidade transformadora, é mostrado na Figura 8.

Figura 8 – Para-raios de MT



Fonte:(CEEE-D, 2013c, seção 10-13)

Devem ser aterrados todos os para-raios e carcaças de equipamentos (transformadores, religadores, etc.) e, havendo condutor neutro secundário, a ligação a terra deve ser em comum aos para-raios, ao neutro e às carcaças. Não deve haver ponto de rede secundária afastado mais de 100 metros de um aterramento, assim como todo fim dessa deve ter neutro aterrado(CEEE-D, 2013b, p. 10).

3 REGISTRO DE AVARIAS

Para gestão de ocorrências, a CEEE-D conta com o Sistema de Gestão da Distribuição (SGD). Nesse, os dados são registrados e disponibilizados para consulta em todas gerências regionais. Em alguns setores, além desse banco de dados centralizado, pode-se encontrar outras formas peculiares de gerência. No caso de avarias em transformadores, tem-se na oficina o monitoramento realizado através de planilhas de controle.

3.1 OCORRÊNCIAS ENCERRADAS

Iniciado o processo de reclamação, a primeira etapa é realizada por um atendente da Central de Atendimento ao Consumidor (CAC), responsável pelo recebimento das informações dos consumidores no que se refere à problemas no fornecimento de energia. É feita uma Nota de Reclamação, documento no qual o atendente registra as informações dos consumidores para tomada de providências por parte do Centro de Operação de Distribuição (COD). Esse órgão é destinado a supervisionar e coordenar as atividades operativas do sistema de distribuição, e manter contato com os operadores de subestação e/ou Centro de Operação do Sistema (COS). Na tela do COD, o software de gerenciamento analisa as informações e as agrupa conforme entidades em comum. É determinado o grau de urgência do incidente e a equipe que atenderá o evento (CEEE-D, 2006). Os procedimentos da norma de Operação da Distribuição NT-84.001 segundo CEEE-D (2014b) são:

Envio de Equipes a Campo

O operador de sistema de distribuição (OSD) realiza o despacho de serviços alocando as ocorrências para equipes que estão logadas no SGD/SGS¹ (Sistema de Gestão de Serviços), são essas dos tipos:

- a) Equipes *Off Line*: equipes de atendimento de ocorrências no SGD/SGS que atuarão sem o auxílio do PDA², utilizando comunicação via rádio VHF ou celular, sendo necessário a intervenção direta do OSD para consulta e encerramento de ocorrências;
- b) Equipes *On-line*: equipes de atendimento de ocorrências no SGD/SGS que atuarão com o auxílio do PDA. Essa ferramenta possibilita o atendimento de

¹ Sistema de Gestão de Serviços: Software utilizado para o tratamento das Ordens de Serviço. Esse sistema é integrado ao SGD no momento do login para acesso a todas as suas funções.

² Personal Digital Assistant: Equipamento de comunicação com o OSD, ou com o SGS, utilizado pelos eletricitistas para recebimento das OS referentes ao atendimento da ocorrências na rede de distribuição.

ocorrências sem a intervenção direta do OSD, salvo em situações específicas definidas pelos CODs. (ex.: manobras).

Despacho de Ocorrências a Equipes *On-line*

O SGS envia a respectiva Ordem de Serviço³ (OS) para o PDA, contendo todos os dados da ocorrência gerada no SGD.

Identificação do Defeito/Falha

Procedimento realizado em campo no atendimento das ocorrências, no qual a equipe localiza e identifica o defeito, informando ao OSD o dispositivo da rede que atuou (ex.: chave, transformador entre outros), bem como, se foi defeito isolado.

Restabelecimento do Fornecimento

A equipe providencia a recomposição do sistema elétrico no trecho atingido pela ocorrência, de acordo com os procedimentos necessários.

Encerramento da Ocorrência

O encerramento da ocorrência é feito através do preenchimento no SGD/SGS dos dados informados pela equipe que efetuou o atendimento. Depois do encerramento, todos os dados ficam disponíveis no SGD para os devidos tratamentos. São efetuados:

- a) Pelo COD do atendimento por equipes *off line*, onde essas devem informar ao OSD os dados do atendimento verificados em campo:
 - A causa da interrupção;
 - Serviço executado e materiais utilizados;
 - Horários referentes ao atendimento da interrupção; chegada e conclusão.
- b) Pelo eletricista do atendimento por equipes *on line*: onde essas devem preencher diretamente no PDA os mesmos dados descritos no item a) para que sejam enviados ao SGS, e encerrada a ocorrência no SGD. Com o propósito de auxiliar na comprovação da causa, sempre que houver condições, recomenda-se que as equipes *on line* utilizem o recurso de “anexar imagem” no PDA antes do encerramento da OS.

A Figura 9 mostra a tela de consulta ao banco de dados para uma OE com causa 303 – Descarga Elétrica, onde verifica-se a seleção da guia “Serviços”. Destaque para (1)

³ Ordem de Serviço: Instrumento lógico do SGS para encaminhamento do atendimento necessário na rede de energia elétrica originado por uma ocorrência, para ser atendido por equipes médias de emergência através do PDA.

o registro do serviço 304 (ainda na antiga nomenclatura) e (2) o texto das observações sobre os serviços executados.

Figura 9 – Detalhes do cadastro de uma ocorrência encerrada

Detalhes da Ocorrência Encerrada

Ocorrência Abrangência Defeito *Instalação Transformadora* Defeito/Falha *045008356*
 Número *2014-1728* Equipamentos *1 / 045008356*

Auditoria
 Controle *JOB* Abrangência Original *Cliente Isolado* Usuário de Criação
 Defeito/Falha Original *734782*

Localização
 Alimentador *CGU 8 - AL8 - UPRT* OS *3490179* Proprietário UT *COMPANHIA*
 Localidade *CANGUCU* Localização *Rural* Tipo *Não Programada* Alterações
 Endereço do P.D.F. : *EST ARVORITO 8356 -A*
 Início da Ocorrência *02/01/2014 08:52:24* Final da Ocorrência *03/01/2014 17:40:00* Operador *F338583*

Detalhes
 Clientes Afetados *2* Clientes Especiais Afetados *0* Clientes VIPs Afetados *0* Informações Importantes *Não* Comunicações *2*

Observações | Causa | Serviços | Horários | Eletricistas | Odômetros | Dec/Eec | Materiais Utilizados | Int. Sit. de Emergência

Observações sobre os Serviços Executados
 B1TMD5 DE 5KVA, WEG, N:35682, CONT:20964-J
 ACESSO BOM LEVAR PARA RAID.
 FOI SUBSTITUIDO TR E PARA-ARID. **2)**

Serviços Executados na Ocorrência
 BT - POSTO DE TRANSFORMADOR
 PARA-RAIOS
 BT - POSTO DE TRANSFORMADOR **1)**
 TRANSFORMADOR AVARIADO

Notas de Serviço | Manobras | Detalhar PD | Clientes Atingidos | Mensagens | Sair

O acesso ao banco de dados das OE se dá através da ferramenta administrativa SGD/SIQUAL, disponibilizado e inter-relacionado em sete tabelas através do ano e número da ocorrência. São estas:

Ocorrência localização, datas, horas, clientes atingidos, etc.;

Comunicados conta, endereços, telefones e etc. dos reclamantes;

Eletricistas eletricitas encarregados, motoristas;

Equipes veículos, equipes e localização;

Observações observações sobre a ocorrência, serviços, causas;

Serviços serviços efetuados na ocorrência, por códigos;

Manobras tipo de manobra, equipamento e horário autorizado.

A OE constitui o evento registrado no software SGD, relativo ao atendimento realizado por equipes de campo, identificando sua localização geográfica, causa, tipo de

equipamento, operação a ser executada, horários, etc. A comunicação é em tempo real e feita por rádio ou telefone (CEEE-D, 2011).

Dentre as tabelas acima citadas, têm-se como as mais relevantes: Ocorrência, Serviços e Observações. A primeira por incluir a causa da ocorrência, local do defeito/falha, e datas de atendimento. A segunda, por conter de forma estruturada os serviços executados. Através de observações textuais, a última engloba fatos relevantes e muitas vezes detalhados dos procedimentos empregados. As demais tabelas agregam informações não fundamentais a este trabalho.

Sobre esse método de controle, tem-se as seguintes vantagens:

- Prós: quantidade de informações, dados centralizados, dinâmico, acessível;
- Contras: incertezas associadas ao preenchimento, alguns módulos pouco utilizados (movimentação de transformadores).

3.2 PLANILHAS DAS OFICINAS

A distribuidora dispõe de duas Turmas de Suprimentos e Equipamentos para aviação e posterior encaminhamento dos transformadores avariados para as oficinas terceirizadas que executam os consertos. Após a avaria do equipamento o transporte até as Turmas de Suprimentos e Equipamentos, feito por caminhões, está propenso à espera da disponibilidade mínima de equipamentos nas regionais e, evidentemente, dos veículos. Em consequência, se espera que essas planilhas não possuam dados mensalmente completos.

Sendo esse setor responsável pelo conserto dos transformadores, espera-se que esta seja uma fonte precisa de informação, isto é, dado que há o registro na planilha, certamente está avariado. Apesar disso, erros de preenchimento são recorrentes, inviabilizando muitas vezes a identificação do evento. Ademais, por tratar-se de um arquivo, dispõe de pouca acessibilidade. A Tabela 10 mostra parte de uma planilha de controle de avarias para a GRM.

Figura 10 – Planilha de controle de avarias da oficina

Circuito (3)		Município (4)		TRANSFORMADOR RETIRADO (5)			TRANSFORMADOR INSTALADO (6)		Data Avaria (7)	Ocorrência (8)
				kVA	Contrafo	N° Série (*)	kVA	Contrafo		
4133	Porto Alegre	112,5	32783C		112,5	300025H	22/01/2014	52335		
2855	Porto Alegre	75	15677J		75	85226B	22/01/2014	52686		
13192	Porto Alegre	75	87442G		112,5	16150B	23/01/2014	54381		
3874	Porto Alegre	112,5	7620/1		112,5	14077F	22/01/2014	51443		

Sobre esse método de controle, tem-se as seguintes vantagens:

- Prós: registro irrefutável de avaria;
- Contras: descentralizado, particularizado, dados restritos, preenchimento errôneo, pode conter equipamentos em manutenção preventiva.

3.3 PRINCIPAIS CAUSAS

A determinação da causa em uma ocorrência aponta a circunstância formadora da descontinuidade do fornecimento de energia. Assim como a inspeção visual, os demais procedimentos são listados na Tarefa Técnica 44 - Testar e Isolar Transformador de Distribuição do Manual do Eletricista de Rede Aérea de Distribuição da CEEE-D (2011). As causas são divididas em grupos, alusivos à autoria do evento: Serviços programados (100), Meio Ambiente (300), Meio de Terceiros (400), Consumidor (500), Falha de componentes (600), Falha humana (700), Outros códigos (900), Próprios do sistema (500) (lista completa dos códigos disponíveis no Apêndice A, pag. 85). Os itens a seguir caracterizam algumas das principais causas de avarias em transformadores suscetíveis a reincidências de modo a demandar especial atenção, isto é, pontos críticos.

MEIO AMBIENTE

Interrupção ocasionada por intervenção do meio ambiente, objetos presos na rede elétrica, animais, pássaros, chuvas, granizo, tempestades, ventos, temperaturas extremas, ou outras condições climáticas adversas que possam influenciar num desligamento (TRONCHONI, 2008, p. 83).

Descarga elétrica (303) O desligamento de uma linha de distribuição pode ocorrer devido ao impacto direto da descarga sobre a linha, como também devido à tensão induzida na linha por uma descarga que ocorra próxima ela. Como decorrência, arcos podem ser gerados tanto na linha como nos transformadores (INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS, 2014). É relacionada diretamente às condições climáticas e localidades de elevado índice cerâmico.

Vegetal (304) Ocorre em regiões onde árvores são plantadas de maneira desordenada ou em regiões rurais, de modo que galhos encostam nos cabos de BT e/ou MT provocando curto circuito.

MEIO DE TERCEIROS

Quando ocorrem desligamentos ocasionados por vandalismo, furto ou outra interferência humana é atribuída essa causa. Muitas causas desse grupo estão relacionadas à áreas de maior complexidade urbana, com ocupações desordenadas e altos índices de violência.

Vandalismo (402) Avarias ocasionadas pela ação humana de caráter intencional e destrutivo.

Ligações à revelia (407) Além do aspecto financeiro, o furto de energia prejudica a integridade física da rede elétrica, pois causa sobrecargas e desligamentos indevidos no sistema de distribuição. Isso prejudica a qualidade e aumenta, ainda mais, os custos para a adequada prestação do serviço à medida que os investimentos necessários em distribuição, transmissão e geração de energia são remunerados por quem paga corretamente sua conta de luz (VIDINICH; NERY, 2009, p. 83). A falta de controle sobre as cargas, e principalmente o desequilíbrio das ligações no circuito geralmente resultam em avarias nos transformadores.

Colisão de veículo (403) Causadas por colisão de veículos, que afetam o patrimônio da CEEE-D, a qual consiste de avaria de postes, cruzetas, condutores e outros materiais da rede de distribuição de energia elétrica, inclusive transformadores e demais equipamentos de proteção e manobra. Nesses casos os registros das ocorrências com essa causa promovem processos judiciais por parte da distribuidora, com intento de cobrar indenização por danos materiais. Por conseguinte, conclui-se que registros com essas causas são pouco suscetíveis a erros.

Furto de rede (409) Interrupções de energia decorrentes de furto de componentes da rede elétrica ou equipamentos.

PRÓPRIO SISTEMA

Caracteriza eventos ocorridos no próprio sistema elétrico, contudo não tendo como origem falha de seus componentes.

Sobrecarga (501) A sobrecarga ocorre em circuitos sem defeitos, onde são proporcionadas sobrecorrentes devido ao excesso de carga. Esse fato tem relação direta com temperaturas ou muito elevadas ou muito baixas em horários de ponta, como também ao aumento de carga à revelia associada à troca descomedida do disjuntor de BT pelo consumidor. Associar esse código à avarias em transformadores é incomum (ou deveria ser), pois os indícios são revelados por queimas sucessivas no elo fusível (atuação da proteção).

FALHA DE COMPONENTES

Interrupção de fornecimento causada por falha do componente devido aos mais diversos motivos.

Transformador avariado (603) Para falhas no equipamento transformador, parte principal da unidade transformadora, tem-se as seguintes danos: Quebra da bucha de porcelana, falha interna dada pela comutação (troca de TAP), vazamento de óleo ou no tanque ou radiador ou junta, curto-circuito nas espiras, entre outros. Considera-se esses fatores devidos ou ao final de vida útil ou defeitos de fabricação. Em caso de falta de manutenção preventiva ou agentes externos, outras causas são designadas. Essa causa é muito empregada para eventos que possuem ausência de evidências, logo é a mais utilizada e a que compreende maiores erros.

Condutor (604) e Conexão (610) Origina-se pelo rompimento do condutor e quebra de conectores instalados na rede (cunha, parafusos, grampo de linha-viva), causando curto-circuito.

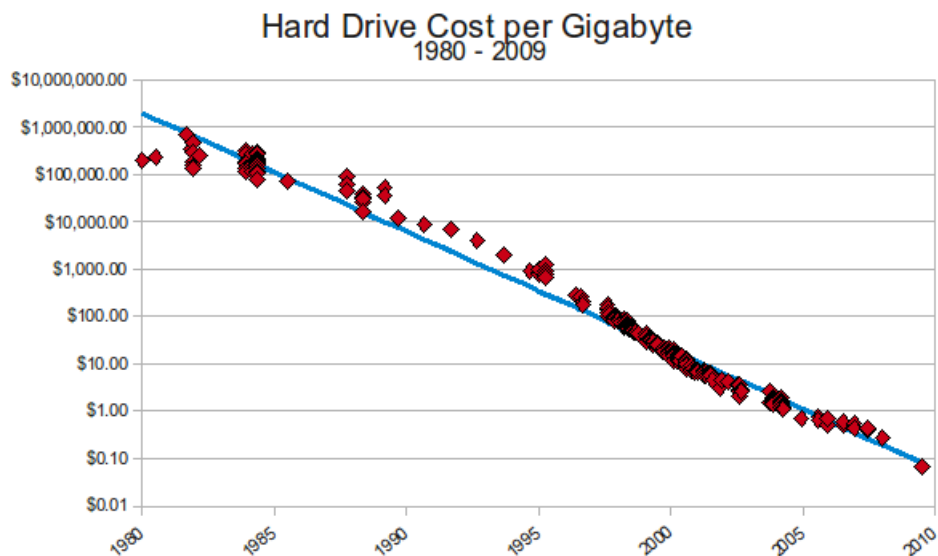
Elo fusível queimado (619) Falha do próprio componente – elo fusível – devendo-se constatar no local que não há nenhum fator externo que tenha ocasionado a interrupção, e o elo não apresentar nenhum indício de fusão no material. Ainda pode apresentar falha de componente no caso de constatação de que o elo encontrado não se encontra de acordo com os padrões especificados (principalmente com relação ao tubo). Resume-se na não atuação da proteção do transformador.

4 DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS

Ao longo dos últimos 30 anos, o custo do espaço de armazenagem de dados tem caído pela metade a cada a cada 14 meses (diminuição de uma ordem de magnitude a cada 48 meses). A Figura 11 ilustra esses pontos em conjunto com a equação de regressão linear simples (KOMOROWSKI, 2009) (coeficiente de correlação $r = 0.9916$) que modela esse comportamento:

$$custo = 10^{-0,2502(ano-1980)+6,304}$$

Figura 11 – Custo por gigabyte de disco rígido nos últimos 30 anos



Fonte:(KOMOROWSKI, 2009)

A boa combinação de custo por gigabyte e o desempenho faz com que a unidade de disco rígido tradicional seja o principal meio de armazenamento para os provedores de nuvem, que os compram aos milhares a cada ano. O decrescimento dos preços ao longo dos anos é surpreendente: o custo médio por gigabyte caiu de US\$437.500,00 em 1980 para US\$0,05 em 2013 (MCLELLAN, 2014). Isso evidencia que, nos mais diversos âmbitos, dados estão sendo coletados e acumulados em um ritmo vertiginoso. A Descoberta de Conhecimento em Banco de Dados (*Knowledge Discovery in Databases* ou KDD) é um conjunto de métodos e técnicas aplicáveis no auxílio da extração de informações úteis (conhecimento) dessas grandes bases. No cerne do processo está a Mineração de Dados (*Data Mining* ou DM), parte do processo que consiste na aplicação de algoritmos especí-

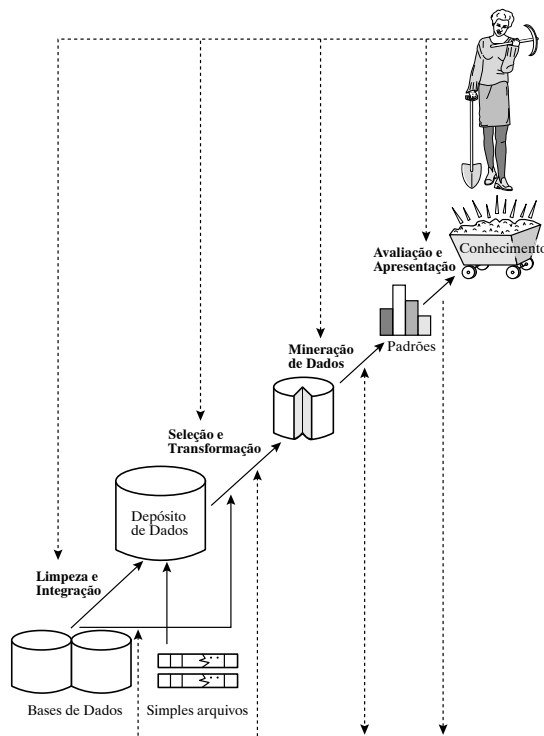
ficos para detecção de padrões de dados. (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Como um processo, a descoberta de conhecimento está representada na Figura 12. Fundamenta-se em uma sequência iterativa das etapas (HAN, 2006, p. 7):

1. **Limpeza de dados:** remoção de ruídos e dados inconsistentes;
2. **Integração de dados:** combinação de múltiplas fontes de dados;
3. **Seleção de dados:** dados relevantes para a tarefa de análise são recuperados do banco de dados;
4. **Transformação de dados:** transformação ou consolidação em formatos adequados para a mineração através da realização de operações de resumo ou de agregação;
5. **Mineração de dados:** processo essencial onde os métodos inteligentes são aplicados a fim de extrair padrões de dados;
6. **Avaliação do padrão:** identificação dos padrões verdadeiramente interessantes que representam conhecimento baseado em algumas medidas de interesse;
7. **Apresentação do conhecimento:** utilização de técnicas de visualização e representação do conhecimento para apresentar o conhecimento extraído para o utilizador.

Os passos de 1 a 4 apresentam diferentes formas de pré-processamento de dados, onde esses são preparados para a mineração. A etapa de mineração de dados pode interagir com o usuário ou uma base de conhecimento. Os padrões interessantes são apresentados para o usuário e podem ser armazenados como novo conhecimento na base de conhecimento.

Figura 12 – A mineração de dados como uma etapa no processo de descoberta de conhecimento



Fonte: Traduzido de Han (2006, p. 6)

4.1 MINERAÇÃO DE DADOS

Traduz-se na aplicação de análise de dados e algoritmos de descoberta que, de acordo com as limitações de eficiência computacional aceitáveis, produzem uma enumeração particular de padrões (ou modelos) sobre os dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996, p. 6). Isto é, explora e analisa grandes quantidades de dados com objetivo de encontrar padrões e regras significativas para algum fim (BERRY; LINOFF, 2000, p. 7).

Muitas técnicas, algoritmos e tipos de modelos são englobados pelo conceito de mineração de dados, sempre com o objetivo de buscar informação útil em meio a dados. Dado isso, a DM é formulada em termos das seis tarefas seguintes:

- Classificação;
- Estimação;
- Predição;
- Agrupamento por afinidade ou regras de associação;
- Agregação (*Clustering*);
- Descrição e visualização.

As três primeiras tarefas (classificação, estimação e predição) consistem em exemplos de mineração de dados direta, ou seja, tem o objetivo de encontrar o valor de uma variável-alvo específica. As outras três são exemplos de atividades da mineração de dados indireta, onde o objetivo é descobrir a estrutura em dados sem relação a uma variável-alvo específica.

4.2 MINERAÇÃO DE TEXTOS

Normalmente, os estudos em aprendizado de máquina trabalham com dados estruturados. Entretanto, uma grande quantidade de informações é armazenada em textos, que são dados semiestruturados. Por exemplo, um documento pode conter alguns campos estruturados, como título, autor, data de publicação, categoria, e assim por diante, mas também conter alguns componentes de texto em grande parte não estruturados, como resumo e conteúdo (HAN, 2006, p. 614).

A Mineração de Textos (*Text Mining* ou TM) tem como objetivo tratar essa informação semiestruturada, com origem na Mineração de Dados, sendo, por isso, chamada também de *Text Data Mining*, etapa do processo de Descoberta de Conhecimento em Textos (*Knowledge Discovery in Texts* ou KDT). Analogamente à mineração de dados, a mineração de texto foi definida por Tan (1999, p. 1) como “processo de extração de padrões ou conhecimentos interessantes e não triviais de documentos de texto”.

Para o processamento, a ideia básica é determinar individualmente a influência de cada palavra em um texto. Logo, os documentos são representados como um conjunto de palavras soltas chamado “saco de palavras” (*bag-of-words*), isto é, uma matriz dispersa onde cada linha L_i corresponde à um documento e cada coluna equivale à um termo T_j . O valor do campo (L_i, T_j) é dado pela frequência do termo T_j no documento L_i . Apesar de essa representação descartar informações sequenciais dadas pela ordem das palavras, ela está por trás de muitas técnicas de recuperação de documentos notavelmente eficazes e populares (WITTEN, 2005, p. 5).

Ao contrário do método do “saco de palavras”, textos técnicos necessitam de informações sequenciais fornecidas pelo arranjo das palavras. Por exemplo, as palavras soltas: queimado, capacitor e transformador, não expressam qual equipamento de fato está avariado. Assim, Cavnar (1995, p. 46) aborda o método de representação por conjuntos de palavras circunvizinhas denominados N-Grams.

Nas etapas de KDD da Figura 12, a TM corresponde à mineração de dados, como previsto. Para as demais etapas, as mudanças são insignificantes, afora a limpeza do texto. O trabalho de Miner, Elder e Hill (2012, p. 46) cita sete procedimentos básicos característicos:

1. **Seleção:** Escolha do âmbito do texto a ser processado (documentos, parágrafos, frases, etc.)
2. **Tokenização:** fraciona o texto em palavras (*bag-of-words*), ou conjunto de palavras (N-Grams), discretas chamadas *tokens*;
3. **Stopwords:** remoção de palavras frequentes e com pouca relevância (tais como preposições, artigos, conjunções);
4. **Stemming:** remoção de prefixos e sufixos a fim de restringir a palavra à seu radical (por exemplo: trocar, trocado e trocou é reduzido para “troc”).
5. **Normalização da ortografia:** unificar erros ortográficos e outras variações de ortografia em um único *token* (por exemplo: fenômeno (Brasil) e fenómeno (Portugal), trocar e trocra para erro ortográfico);
6. **Limites da frase:** delimitação do início e fim dos trechos;
7. **Minuscularizar:** tornar todo o texto minúsculo (ou maiúsculo).

4.3 CLASSIFICAÇÃO

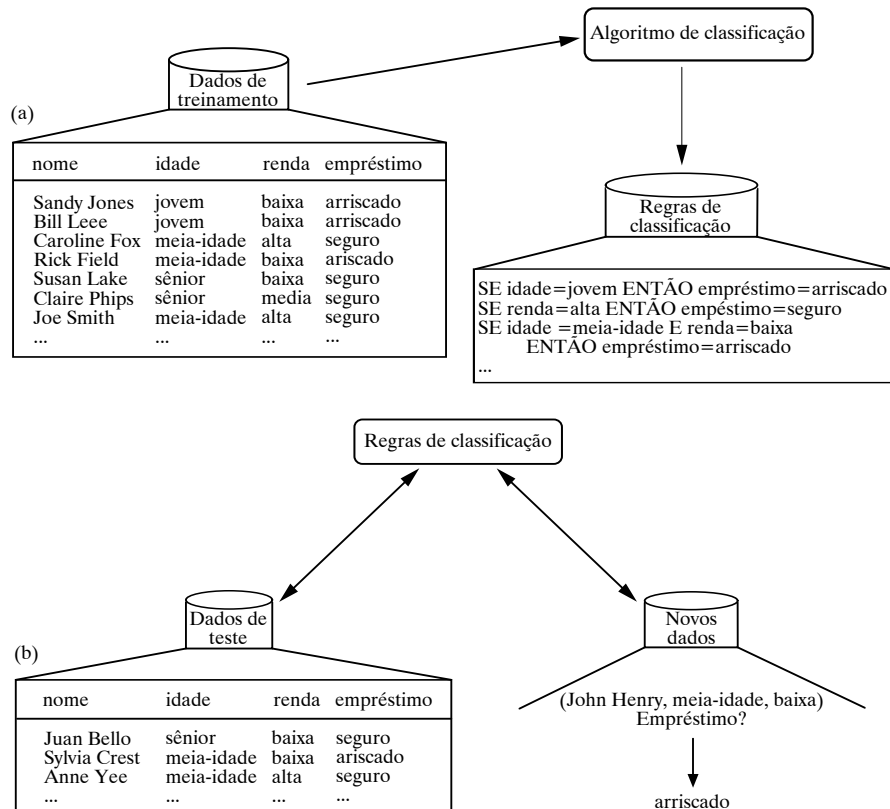
A classificação é uma das atividades possíveis na mineração de dados. Define-se como o processo de encontrar um modelo (ou função) que descreve e distingue classes de dados ou conceitos, a fim de ser capaz de usar esse modelo para prever a classe de objetos cujo rótulo de classe é desconhecido (HAN, 2006, p. 24).

No primeiro passo, um classificador é construído descrevendo um conjunto pré-determinado de classes de dados (ou conceitos). Essa é a etapa de aprendizagem (ou fase de treinamento), onde um algoritmo de classificação constrói o modelo classificador através da análise (aprendizado) de um conjunto de treinamento composto de tuplas¹. Uma tupla \mathbf{X} é representada por um vetor de atributos n -dimensional, $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$, que retoma n medidas (Sandy Jones, jovem, baixa e arriscado), registradas para os n atributos (nome, idade, renda e empréstimo) do banco de dados. Assume-se que cada tupla é pertencente a uma classe pré-definida, determinada por um atributo do banco de dados chamado atributo de rótulo da classe (empréstimo). Como o rótulo da classe de cada tupla de treinamento é fornecido, essa etapa também é conhecido como aprendizado supervisionado (HAN, 2006, p. 285).

Essa primeira etapa do processo de classificação pode ser vista como mapeamento de uma função $y = f(\mathbf{X})$, que é capaz de prever a classe associada ao rótulo y de uma determinada tupla \mathbf{X} . No exemplo da Figura 13 (a), o mapeamento é representado como

¹ Tupla: linha de um banco de dados, formada por um conjunto de colunas representando um registro.

Figura 13 – As etapas de (a) aprendizado e (b) teste para do processo de classificação



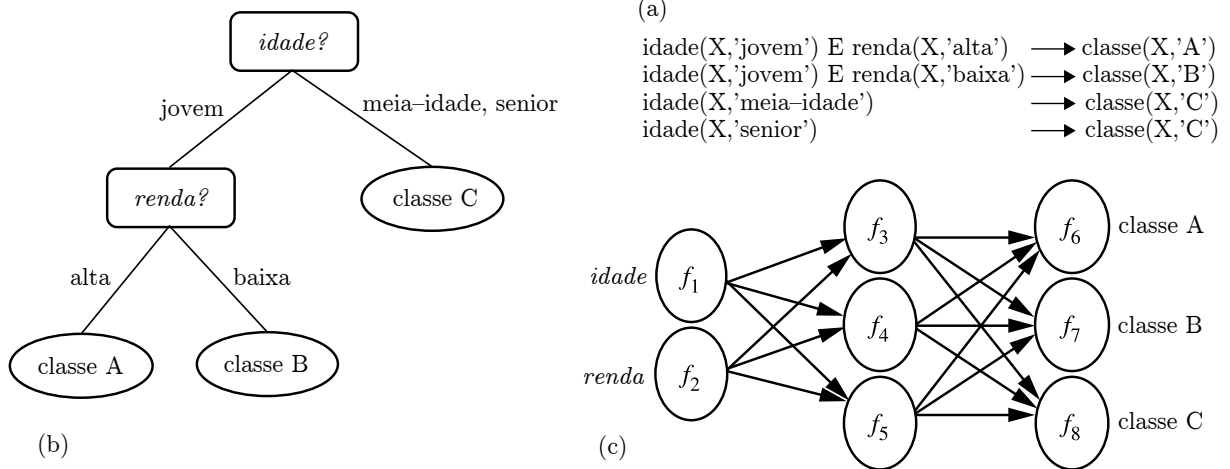
Fonte: Traduzido de Han (2006, p. 287)

regras de classificação que identificam pedidos de empréstimo como sendo ou seguro ou arriscado. Essas regras fornecem uma compreensão mais profunda do conteúdo do banco de dados, bem como podem ser utilizadas para categorizar tuplas de dados futuros, de modo a, por exemplo, aprovar ou rejeitar novos candidatos à empréstimos (Figura 13 (b))

4.4 ALGORITMOS DE APRENDIZAGEM

A escolha do algoritmo está diretamente ligada aos padrões a serem encontrados assim como o nível de interpretação exposto pelo modelo final. Podem-se utilizar algoritmos indutores de árvores de decisão ou regras de classificação, por exemplo, se o objetivo é uma obter uma classificação de fácil regimento. Entre os tipos mais frequentes de representação de padrões destacam-se: árvores de decisão, regras SE-ENTÃO, modelos lineares, modelos não-lineares (Redes Neurais Artificiais), modelos baseados em exemplos (*K-Nearest Neighbor* ou KNN) e modelos de dependência probabilística (Redes Bayesianas do inglês *Bayesian Network* – BN) (REZENDE, 2003, p. 319). A Figura 14 apresenta os formatos de regras de classificação, árvores de decisão e rede neural.

Figura 14 – Um modelo de classificação representado em várias formas, tais como: (a) regras SE-ENTÃO, (b) árvore de decisão, ou (c) rede neural



Fonte: Traduzido de Han (2006, p. 25)

REDES BAYESIANAS

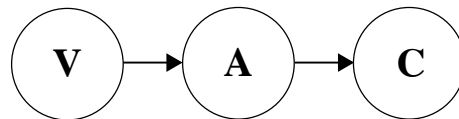
Uma Rede Bayesiana é uma estrutura de dados que representa as dependências entre variáveis e fornece uma especificação concisa de qualquer distribuição de probabilidade conjunta total. É representada por um grafo orientado em que cada nó é identificado com informações de probabilidade quantitativa (RUSSELL; NORVIG, 2004, p. 480). A especificação completa é dada a seguir:

1. Um conjunto de variáveis aleatórias constitui os nós da rede. As variáveis podem ser discretas ou contínuas;
2. Um conjunto de vínculos orientados ou seta conecta pares de nós. Se houver uma seta do nó X até o nó Y , X será denominada “pai” de Y ;
3. Cada nó X_i tem uma distribuição de probabilidade condicional $P(X_i|Pais(X_i))$ que quantifica o efeito dos pais sobre o nó;
4. O grafo não tem nenhum ciclo orientado (e conseqüentemente é um grafo acíclico orientado).

A topologia da rede – o conjunto de nós e vínculos – especifica os relacionamentos de independência condicional que são válidos no domínio, de um modo que se tornará claro em breve. O significado intuitivo de uma seta em uma rede construída corretamente em geral é que X tem uma influência direta sobre Y . Normalmente é fácil para um especialista em domínios descobrir quais são as influências existentes no domínio – na verdade, é muito mais fácil do que realmente especificar as probabilidades em si.

O diagrama diagnóstico-causa mostrado na Figura 15 descreve a influência dos elementos pais sobre os filhos de forma exclusiva. Cada filho possui apenas um pai, assim A é filho de V e C é filho de A. Um exemplo dessa relação é o (V) ato de chover, (A) descargas atmosféricas (C) causa de avaria: descarga elétrica (Adaptado de Silva (2010, p. 33)).

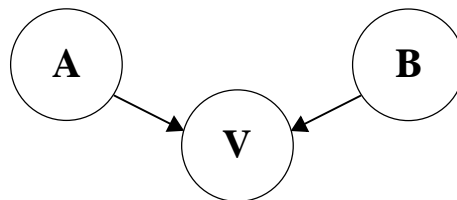
Figura 15 – Exemplo genérico diagnóstico-causa



Fonte: (SILVA, 2010, p. 33)

No diagrama intercausal de causa (Figura 16), o filho possui dois ou mais pais. Cada filho possui V causas A e B. É uma situação de causas diversas que pode ser ilustrada por (A) vazamento de óleo, (B) troca de TR e (V) ocorrência de avaria.

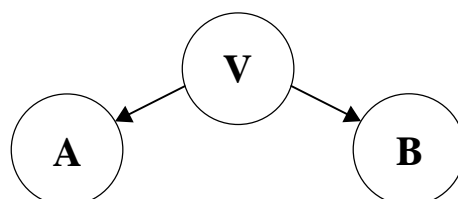
Figura 16 – Diagnóstico genérico intercausal de efeito



Fonte: (SILVA, 2010, p. 34)

Finalmente, invertendo-se o sentido das interligações (Figura 17) encontra-se o diagrama intercausal de efeito onde um pai (V) pode possuir vários filhos (A e B). As consequências múltiplas descrevem, por exemplo, uma (V) colisão de veículo sobre o estado final do (A) poste e (B) transformador.

Figura 17 – Diagnóstico genérico intercausal de causa



Fonte: (SILVA, 2010, p. 34)

A aprendizagem bayesiana simplesmente calcula a probabilidade de cada hipótese, considerando-se os dados, e faz previsões de acordo com ela. Isto é, as previsões são feitas com uso de todas as hipóteses, ponderadas por suas probabilidades, em vez de utilizar apenas uma única "melhor" hipótese. Desse modo, a aprendizagem é reduzida à inferência probabilística (RUSSELL; NORVIG, 2004, p. 691)

4.5 AVALIAÇÃO DO PODER DE PREDIÇÃO

Um dos critérios mais importantes para comparar o desempenho de algoritmos de classificação é o seu poder de predição. Para isso, pode-se empregar os próprios dados utilizados em sua elaboração. Porém, esses resultados podem ocasionar estimativas otimistas devido a superespecialização do algoritmo de aprendizado.

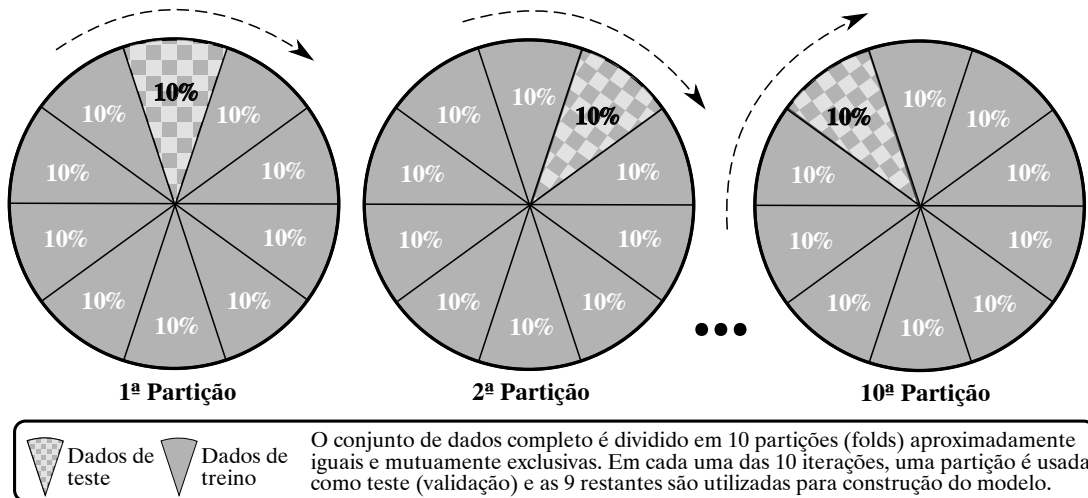
O poder de predição é a acurácia da previsão do modelo esperado quando ele é aplicado a novos casos, os quais não foram incluídos no conjunto de dados de treino. Essa só pode ser medida assintoticamente em um grande número de novos casos. Infelizmente, na maioria dos problemas do mundo real a quantidade de dados disponíveis de mesma população é bastante limitado, conseqüentemente é necessário o uso de estimadores (MAIMON; LAST, 2001, p. 55). Holdout, subamostragem aleatória, Validação Cruzada e Bootstrap são técnicas comuns para avaliação da acurácia com base em partições de amostras aleatórias dos dados fornecidos.

VALIDAÇÃO CRUZADA

A estimação por Validação Cruzada em r -Fold (*Cross-Validation* ou CV) baseia-se na divisão aleatória dos n exemplos em r partições mutuamente exclusivas (*folds*) de tamanho aproximadamente igual a n/r exemplos. Os exemplos nas $(r - 1)$ partições são usadas para treinamento e a hipótese induzida é testada na partição remanescente. Esse processo é repetido r vezes, cada vez considerando uma partição diferente para teste (REZENDE, 2003). A Figura 18 ilustra 3 etapas da validação cruzada para 10 partições.

Nos estudos realizados por Kohavi (1995, p. 7), conclui-se que o número de 10 partições é ótimo para se estimar a acurácia do classificador, mesmo que o poder computacional permita usar mais.

Figura 18 – Ilustração da validação cruzada para 10 partições



Fonte: Traduzido de (OLSON; DELEN, 2008, p. 142)

4.6 MATRIZ DE CONFUSÃO E MÉTRICAS DE QUALIDADE

Uma matriz de confusão (*Confusion Matrix* ou CM) é uma matriz $L \times L$, onde L é o número de diferentes valores de rótulos. A Tabela 2 é um exemplo de matriz de confusão (SMOLINSKI; MILANOVA; HASSANIEN, 2008, p. 56).

Tabela 2 – Exemplo de matriz de confusão

	Predição Negativa	Predição Positiva
Atual Negativo	a	b
Atual Positivo	c	d

Fonte: (SMOLINSKI; MILANOVA; HASSANIEN, 2008, p. 56)

Nessa, o quadrante a indica aqueles tuplas que foram preditas como negativas e na verdade eram negativas. Quadrante b é composto por aquelas tuplas que foram preditas como positivas, mas na verdade eram negativas, também conhecidos como falsos positivos. O quadrante c apresenta as tuplas que foram classificados como negativos, mas na verdade eram positivas, também conhecidos como falsos negativos. E finalmente, o quadrante d contém as tuplas que foram classificados como positivas e realmente eram positivas. Assim, usando a matriz de confusão, é mais fácil definir algumas métricas:

Acurácia (*accuracy*) é o percentual de tuplas que estão corretamente classificadas, entre todas as tuplas que são dadas uma classificação.

$$\text{acurácia} = (a + d) / (a + b + c + d)$$

Sensibilidade (ou Taxa de Verdadeiros Positivos) é o percentual de classificação positiva. Para um classificador muito sensível, raramente se deixará de encontrar

o verdadeiro positivo. Ideal para ser utilizado quando o custo ocasionado por se deixar de encontrar é alto (tais como, de forma incorreta prever que um paciente canceroso não possui câncer).

$$\text{sensibilidade} = d/(c + d)$$

Especificidade (ou Taxa de Verdadeiros Negativos) indica a porcentagem de classificação negativa. Um classificador com alta especificidade raramente rotula um falso positivo. Ao contrário da sensibilidade, é ótimo para ser utilizado quando o custo por encontrar é alto (como rotular um paciente não canceroso como canceroso).

$$\text{especificidade} = a/(a + b)$$

Taxa de Falsos Positivos representa o percentual de tuplas que são classificadas como positivas, mas na realidade são negativas.

$$\text{tfn} = b/(a + b) = 1 - \text{especificidade}$$

Taxa de Falsos Negativos é a porcentagem de tuplas que são classificadas como negativas, mas na realidade são positivas.

$$\text{tfn} = c/(c + d) = 1 - \text{sensibilidade}$$

Precisão (*precision*) indica o número de exceções a uma regra. Por exemplo, uma precisão de 4/5, indica que há uma exceção à regra.

$$\text{precisão} = d/(b + d)$$

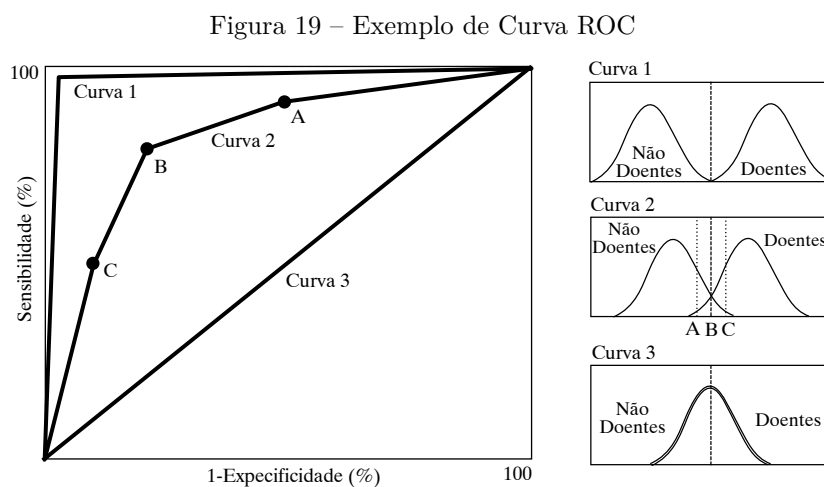
Cobertura (*recall*) é o valor da cobertura de casos, muito influenciada pela sensibilidade e pouco pela especificidade. É calculada por número de casos cobertos pelo número total de casos aplicáveis.

Medida-F (*F-measure*) é uma métrica de avaliação popular para problemas de desequilíbrio. Representa uma média harmônica entre cobertura (*recall*) e precisão. A média harmônica de dois números tende a estar mais perto do menor dos dois. Assim, um valor alto da medida-F garante que tanto recall e precisão são razoavelmente alta (GU; CAI; ZHU, 2009)

$$\text{medida} - F = (2 * \text{precisão} * \text{cobertura})/(\text{precisão} + \text{cobertura})$$

4.6.1 Curva ROC

A Curva ROC (*Receiver Operator Characteristics*) constitui uma forma de integrar as informações obtidas pela sensibilidade e especificidade para um determinado teste (PASQUALOTTO et al., 2012, p. 3). Ela é constituída desenhando-se a taxa de verdadeiro-positivos (sensibilidade) contra a taxa de falso-positivos (1-especificidade) ao longo de uma faixa de pontos de corte (Figura 19 (a)). Os testes com melhores sensibilidade e especificidade encontram-se no canto superior esquerdo da curva, onde um aumento da sensibilidade é obtido à custa de pouca ou nenhuma perda da especificidade.



Fonte: (PEREIRA et al., 2014)

Quando se tem uma variável contínua, resultado da aplicação de uma classificação quantitativa, e se pretende transformá-la numa variável dicotômica, do tipo doente/não doente, temos que utilizar um determinado valor na escala contínua que discrimine entre essas duas classes. A esse valor dá-se o nome de *cutoff point* (PEREIRA et al., 2014).

O valor escolhido como *cutoff point* vai influenciar as características da classificação, como exemplificado na curva 2 da Figura 19. Quanto maior é o *cutoff point* maior é a especificidade do teste mas menor é a sensibilidade (ponto C da curva 2); e quanto menor o *cutoff point* maior é a sensibilidade mas menor é a especificidade (ponto A da curva 2).

Assim, a intenção com que se utilizará a classificação vai influenciar a escolha do *cutoff point*, logo, das características do teste. No exemplo da curva 2 da Figura 19, se pretendemos um classificador muito sensível e menos específico, escolhe-se um *cutoff point* menor (ponto A), obtendo-se uma menor proporção de falsos negativos e uma maior proporção de falsos positivos; se pretendemos um teste muito específico e menos sensível, escolhe-se um *cutoff point* maior (ponto C), obtendo-se uma menor proporção de falsos positivos e uma maior proporção de falsos negativos.

Classificadores de menor poder discriminatório têm curvas mais próximas da diagonal, que vai do canto inferior esquerdo ao canto superior direito. Também é útil para se comparar o desempenho de diferentes testes de classificação, pois a acurácia global do classificador corresponde à área sob a Curva ROC. Um modelo com precisão perfeita terá uma área de 1,0.

4.6.2 Coeficiente Kappa

A medida Kappa (*Kappa statistic*) descreve a intensidade da concordância entre dois ou mais métodos de classificação (BALTAR; OKANO, 201-?). Consiste no número de respostas concordantes, ou seja, no número de casos cujo resultado é o mesmo entre os classificadores. Assim, tem-se o grau de concordância além do que seria esperado tão somente pelo acaso. O valor máximo 1 representa total concordância e os valores próximos, e até abaixo de 0, indicam nenhuma concordância, ou a concordância foi exatamente a esperada pelo acaso. O trabalho de Landis, Koch et al. (1977) preconiza a interpretação vista na Tabela 3.

Tabela 3 – Interpretação dos coeficientes Kappa

Valores de Kappa	Interpretação
<0	Sem acordo
0 – 0,19	Baixa concordância
0,20 – 0,39	Acordo justo
0,40 – 0,59	Concordância moderada
0,60 – 0,79	Acordo substancial
0,80 – 1,00	Concordância quase perfeita

Fonte: (LANDIS; KOCH et al., 1977)

Essa avaliação de concordância através do Kappa é utilizada quando as escalas são categóricas e sempre quando compara-se dois ou mais classificadores.

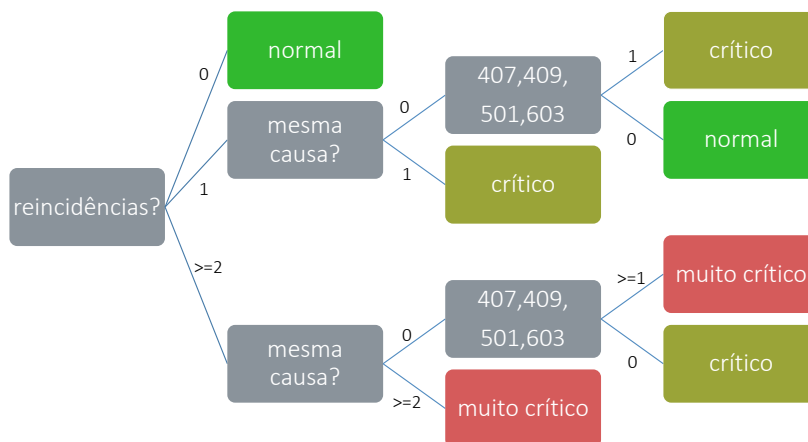
4.7 AVALIAÇÃO DE PONTOS CRÍTICOS

Para determinar o nível de criticidade de uma UT, deve-se analisar seu histórico de ocorrências. A reincidência de atendimentos referentes à causas não programadas é sinal da exposição do equipamento a uma condição peculiar de sua localização. Como visto anteriormente, circunstâncias onde têm-se registros sucessivos de causas atribuídas ao próprio equipamento (causa 603), não atentam para essa observância. Assim como, por exemplo, pontos com frequentes avarias devido à vegetais, também devem ser assistidos conforme sua recorrência.

Não só causas iguais constituem uma relação de reincidência. Para o caso analítico 603, os antecedentes de ligações à revelia (407), furtos de rede (409) e sobrecarga (501) indicam condições vinculadas ao local. Nessas situações, classifica-se esses pontos de modo a condicioná-los a patamares de criticidade.

A Figura 20 mostra uma árvore de decisão, a qual representa nada mais que uma disposição gráfica de um conjunto de regras SE-ENTÃO aplicáveis à uma UT. As determinações utilizadas são definidas de forma empírica e abrangem apenas registros contendo TR avariados para um período de 5 anos.

Figura 20 – Árvore de decisão para avaliação de pontos críticos



Pelas regras, verifica-se que em um atendimento de emergência onde a causa é o transformador avariado, não constatadas reincidências, considera-se o ponto como normal. Oposto à isso, a constatação de duas reincidências ou mais, ora de mesma causa ora comportando as causas 407, 409, 501 e 603, atribui um nível alto de criticidade ao ponto, ou seja, tem-se a necessidade de um acompanhamento especial para esse local.

O intuito dessa avaliação, como descrito inicialmente no objetivo deste trabalho, é reiterar as causas de avarias em UT, especialmente às falhas conferidas aos próprios equipamentos transformadores. É expectável da árvore de decisão vista na Figura 20 o reconhecimento de locais problemáticos das mais diversas causas, sobretudo assentamentos urbanos com baixos índices socioeconômicos e elevadas taxas de furto de energia.

4.8 FERRAMENTA WEKA

Formada por uma coleção de algoritmos de aprendizado de máquina, a ferramenta computacional WEKA (*Waikato Environment for Knowledge Analysis*) é um instrumento de fácil utilização para principiantes, pois seus processos podem ser aplicados diretamente a um conjunto de dados. Não obstante, disponibiliza configurações de parâmetros avançados e interface para linhas de comando, atendendo também os usuários mais experientes.

Desenvolvido para a plataforma Java, tem sua execução assegurada em diversos sistemas operacionais. Contém ferramentas para pré-processamento de dados, classificação, regressão, agregação, regras de associação e visualização (HALL et al., 2009). WEKA é um software de código aberto sob a GNU (*General Public License*), desenvolvido pelo Machine Learning Group at the University of Waikato (2014).

A tela inicial (Figura 21), chamada de *Weka GUI Chooser*, serve de ponto de partida para os módulos de aplicações e ferramentas de apoio. São disponibilizados quatro botões, dados os quatro principais aplicativos. São eles:

- *Explorer*: ambiente para explorar dados;
- *Experimenter*: ambiente para a realização de experimentos e realização de testes estatísticos entre sistemas de aprendizagem.
- *KnowledgeFlow*: ambiente com as mesmas funções que o *Explorer*, mas com uma interface arrastar e soltar;
- *SimpleCLI*: interface de linha de comando simples que permite a execução direta de comandos WEKA para sistemas operacionais que não fornecem sua própria interface de linha de comando.

Figura 21 – Tela inicial da ferramenta WEKA

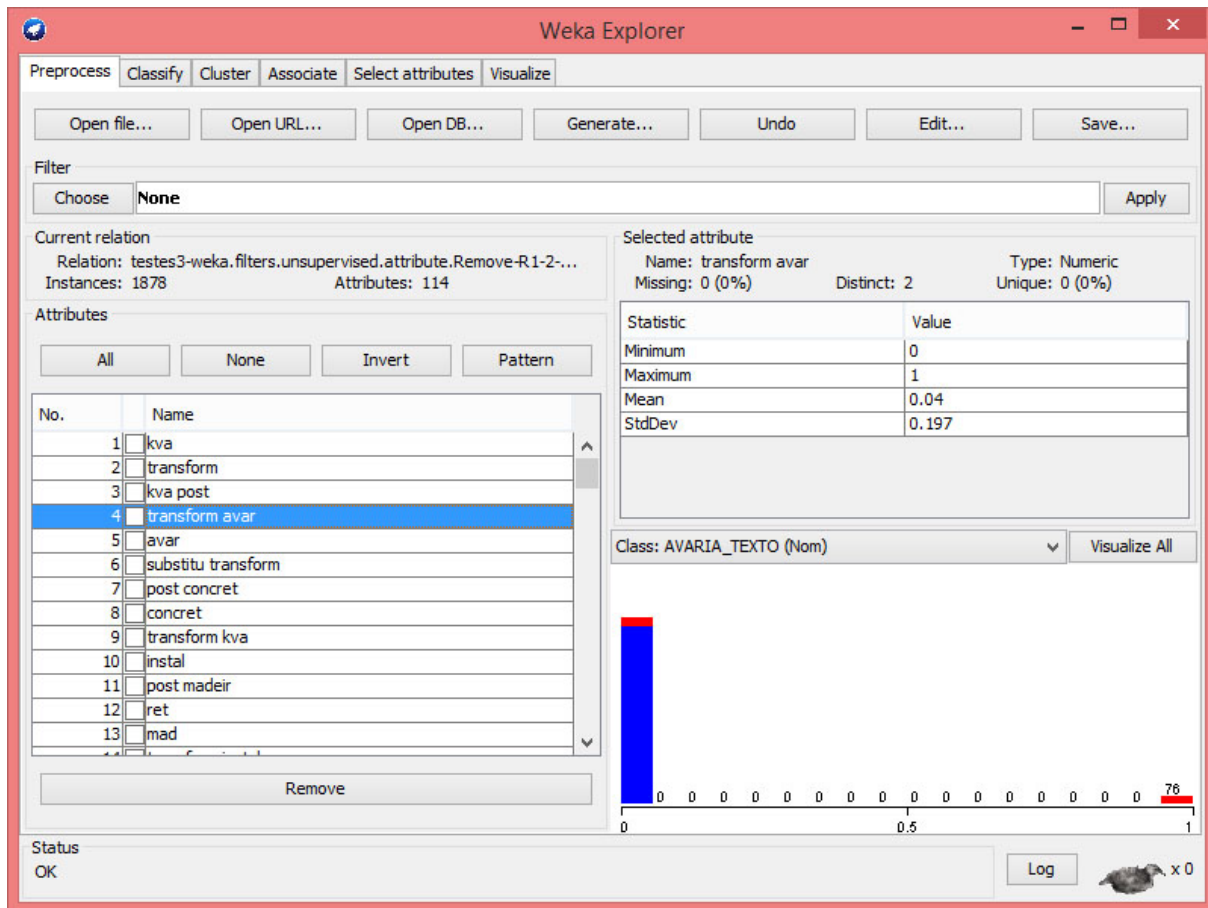


A Figura 22 mostra o *Weka Explorer*, ambiente que permite a execução manual das etapas individuais de mineração de dados. Tem-se no topo da janela, abaixo da barra de título, a guia de obtenção de dados para pré-processamento – *Preprocess*, visualização em 2d – *Visualizer*, e algoritmos de mineração de dados separados por funções:

- *Classify*: treino e testes de classificação e regressão;
- *Cluster*: classificação por agregação (agrupamento);
- *Associate*: descoberta de regras de associação;

- *Select attributes*: seleção de atributos mais relevantes.

Figura 22 – Guia de pré-processamento de dados da aplicação *Weka Explorer*



Os dados podem ser extraídos diretamente de páginas no formato html, pastas com arquivos em texto ou bancos de dados. Para realização de treinamentos simples, faz-se a análise de dados a partir de arquivos com valores separados por vírgulas (*Comma-separated values* ou CSV). Basicamente, o arquivo CSV é importado para dentro do aplicativo através de um conversor chamado *CSVLoader*, o qual o transforma no formato ARFF (*Attribute-Relation File Format*). Esse é definido como um arquivo de texto ASCII que descreve a lista de instâncias que compartilham um conjunto de atributos.

Outro ambiente é o *Weka Experiment Environment*, o qual permite ao usuário criar, executar, modificar e analisar experiências de uma forma mais conveniente do que é possível quando o processamento dos sistemas individualmente. Pode-se criar experiências que executam vários sistemas contra uma série de conjuntos de dados e, em seguida, analisar os resultados para determinar qual deles representam (estatisticamente) o melhor desempenho (HALL et al., 2009, p. 61).

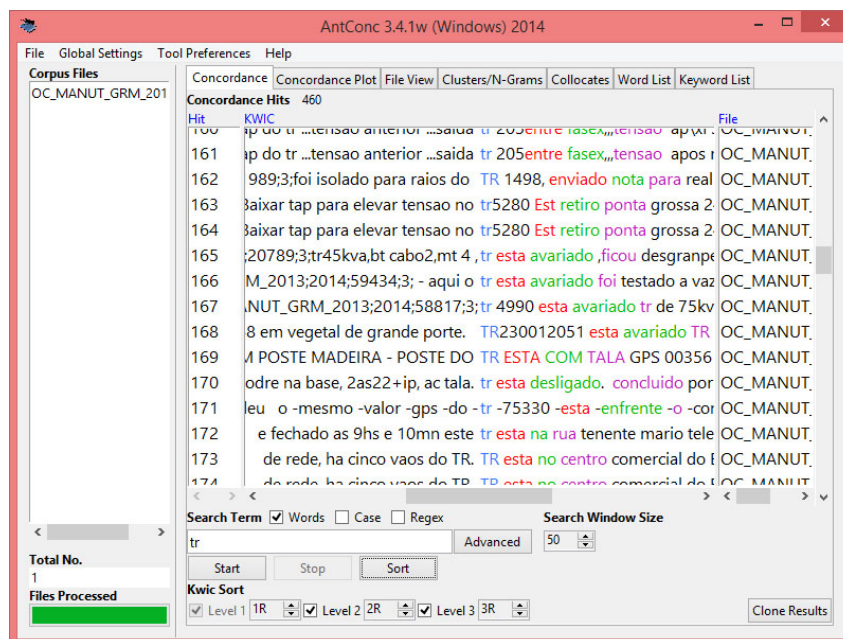
5 ESTUDO DE CASO: GRM

Seleciona-se as ocorrências encerradas, não programadas, da Gerência Regional Metropolitana para o mês de janeiro de 2014, em um total de 3.657 registros. A GRM é composta por 177 alimentadores em 19 municípios. As informações são registradas e/ou comunicadas por 395 eletricitistas integrantes de 134 equipes.

5.1 TRATAMENTO DOS DADOS

Para cada registro, os 3 campos de observações (serviços, causas e ocorrência) são concatenados. A análise textual com auxílio do software AntConc (Figura: 23) possibilita averiguar as mais frequentes palavras, gírias e erros ortográficos capazes de influenciar o resultado da análise. Em um primeiro momento, essa verificação possibilita escolher expressões as quais se deve atentar.

Figura 23 – Análise de concordância de palavras via software AntConc



Retiram-se as palavras pouco relevantes (*stopwords*) e as demais reduzem-se às primitivas (*stemming*). Uma lista de modelos de equipamentos a serem substituídos por palavras simbólicas é criada, onde por exemplo: o cabo da baixa tensão 4 CA é substituído por “txyzca”, tem-se então a informação de que um cabo foi especificado, independentemente de suas características. Uma atenção especial para as palavras circunvizinhas à transformador (referenciado também como TR, transf., trafo, contrafo, cc, TD) as quais são exemplificadas na Tabela 4.

Tabela 4 – Lista de palavras equivalentes

Palavra	Equivalente	Palavra	Equivalente
subst	substituir	tr	transformador
substitui??o	substituir	transf.	transformador
substitu?d?	substituir	trafo	transformador
trocad?	substituir	vazando	vazamento
troca	substituir	vasamento	vazamento
trocar	substituir	vazou	vazamento

A Tabela 5 mostra a execução das etapas referentes à ocorrência encerrada número 2014-2421.

Tabela 5 – Etapas de limpeza e integração dos dados

Etapa	Observação
texto original	rede desregulada. - por favor fazer ordem p instalac?o de espacadores no cc 34 !rede desregulada. TR COM VAZAMENTO DE OLEO, 3x45kVA, CC 34, suporte, poste de concreto. ficou aberto com elos trocados, desgrampeado. Regular rede e instalar espacadores. Trocar dois para-raios. BT 3#2(4)CA. Av Getulio Vargas, 166.
filtros, ortografia, minúsculização, substituições	rede desregulada fazer ordem instalacao espacadores cc rede desregulada transformador vazamento oleo kva cc suporte poste concreto aberto elo trocados desgrampeado regular rede instalar espacadores substituir dois para-raio bt txyzca getulio vargas
<i>stemming, stopwords</i>	red desregul faz ordem instalac espac cc red desregul transform vazament ole kva cc suport post concret abert elo troc desgramp regul red instal espac substitui pararai bt txyzca getuli varg

Os filtros de acentos, correção ortográfica e minúsculização são efetuados de forma assistida no software de planilha eletrônica Microsoft Excel, através da criação de macros de substituição em massa. As *stopwords* utilizadas são disponibilizadas por Porter (2001), já o processo de *stemming* é efetuado através de um software resultante do trabalho de Junior, Imamura e Rezende (2001). Desconsidera-se registros com observações duplicadas, em branco e até 20 caracteres.

Os textos são inseridos no *Weka Explorer* e convertidos para tuplas com no máximo 250 símbolos (*tokens*), esses formados por agrupamentos de até 2 palavras por meio do comando *NGramTokenizer*. Reduzidas pelo avaliador *InfoGainAttributeEval* a 93 atributos relevantes, o Código 1 mostra algumas das palavras símbolo resultantes.

Assim, a tupla final para a ocorrência encerrada da Tabela 5 tem alguns de seus parâmetros disponíveis na Tabela 6.

Código 1 – Amostra de tokens resultantes

```
transform avar, avar, instal, substitu transform, ret, transform kva,
retir instal, instal retir, transform instal, vazament ole, vazament,
desgramp, transform retir, avar kva, condutor said, transform queim,
avar transform, retir transform, abert desgramp, test vazi
```

Tabela 6 – Amostra da tupla para a ocorrência 2014-2421

No.	classe	suport	transform	transform avar	instal	concret	...	vazament ole
95	1	1	1	0	1	1	...	1

5.2 CLASSIFICAÇÃO

Para classificar as ocorrências encerradas como possuindo troca de transformador (1) ou não (0), utiliza-se a classificação assistida ou uma fonte confiável de informação. Dada essas opções, uma planilha de transformadores avariados fornecida pela oficina teve seus dados cruzados com as observações de treino. A escolha deve-se à confiabilidade das informações positivas dessa fonte e, desse modo, não correr o risco de gerar um padrão incorreto.

Dessa planilha (vista na Figura 10 da página 38) tem-se o conhecimento de 114 avarias no mês de 01/2014. Dessas, 5 não se relacionam diretamente com as ocorrências do banco de dados seja por erro de preenchimento, seja por se tratar de ocorrências programadas. Enfim, são rotuladas 109 observações como contendo avaria.

Primeiro Treino

Para criação do modelo, usa-se a opção *Use Training Set* na guia *Classify*. O classificador é alterado para *BayesNet*, em configuração padrão. O resultado está disponível no Código 2 (resultado completo para esse e demais treinos no Apêndice A, pag. 79).

Código 2 – Métricas de qualidade do primeiro treino de classificação

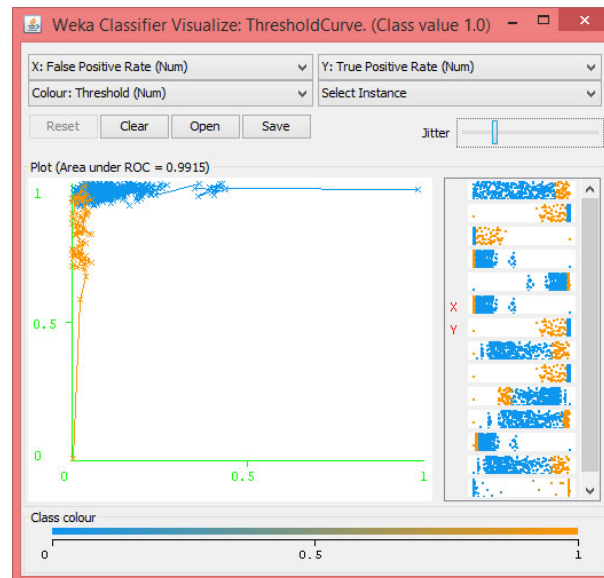
Correctly Classified Instances	1816	96.6986 %
Incorrectly Classified Instances	62	3.3014 %
TP Rate	0.967	
FP Rate	0.037	
Precision	0.998	
Recall	0.967	
F-Measure	0.982	
R0C Area	0.98	
Class	0.0	
Weighted Avg.	0.967	0.036
	0.977	0.967
	0.97	0.98
a	b	<-- classified as
1711	58	a = 0.0
4	105	b = 1.0

A taxa de classificação incorreta (*Incorrectly Classified Instances*) de 3,3%, induz a priori que o modelo tem uma boa precisão, porém, dado que nos dados de treino as subs-

tuições de transformadores são em torno de 5,8% dos registros, deve-se adotar outros valores como referência. O número absoluto de instâncias incorretamente classificadas (62), se comparadas às instâncias classificadas como positivas pela Oficina (109), ocasionam em um número alto, esse discriminado pela matriz de confusão (*Confusion Matrix*), onde o modelo encontrou 58 falsos positivos e 4 falsos negativos. A análise inicial dá o entendimento de que o modelo criado é um péssimo classificador. Têm-se a confiabilidade positiva (precisão) (0,644) baixa em relação à confiabilidade negativa (0,998), logo o modelo encontrado é muito melhor em encontrar negativos do que positivos, o que é esperado devido a discrepância entre as classes.

A área da curva ROC (Figura 24) associada às classes é de 0,98 (muito próxima a 1), o que significa uma grande sensibilidade e especificidade por parte do modelo, isto é, o classificador consegue distinguir muito bem as classes.

Figura 24 – Curva ROC para o primeiro treino de classificação



É sabido que os registros da planilha fornecida pela oficina não compreendem todas as avarias, então, em relação às métricas determinadas, considera-se um bom resultado. Isso deve-se à pequena dimensão dos falsos positivos e negativos e, principalmente, à área da curva ROC. Assim, nessa etapa, supõe-se 58 novos registros positivos e 4 negativos, além daqueles fornecidos pela oficina.

Classificação Assistida

Com o intuito de melhorar as métricas de qualidade do modelo, são averiguados no Microsoft Excel os 62 resultados falsos encontrados pelo classificador, ou seja, as divergências encontradas na matriz de confusão do primeiro treino (Código 2). A Tabela 7 mostra alguns exemplos das comparações, onde P[1] e P[0] são as probabilidades de positivo e negativo respectivamente dadas pelo modelo Bayesiano. Esses parâmetros são de

grande auxílio na tarefa rotulamento dos dados, pois a grande maioria de probabilidade zero é descartada de leitura.

Tabela 7 – Comparativo entre classificações da oficina (OF), Rede Bayesiana (BN) e assistida (AS)

Ano	Ocor.	OF	BN	P[0]	P[1]	AS	Observação
2014	1527	1	0	0.814	0.186	1	- trocar cc 5928 - suporte - concreto - 3 x 45 - vazando óleo - rua pedro a de freitas filho, 70 .
2014	16185	1	0	1	0	0	atuada fase c do circ 11011 - pne - elo 2h - normalizado
2014	56383	0	1	0.003	0.997	1	-substituir tr queimado 3x75 kva suporte poste de madeira, rua cel inacio de quadros numeral 93
2014	56821	0	1	0	1	1	tr avariado, 1x10kva, suporte poste de madeira, localizado frente ao reclamante 2 elo 1 h, 1 elo 40 k contrafo inst 28424

Para os 4 casos de falsos negativos, apenas em 1 há a descrição de avaria devido à vazamento de óleo (ocorrência 1527 da Tabela 7). Identificou-se 52 verdadeiros positivos nos 58 falsos positivos da matriz de confusão. Ademais, foram encontrados 2 registros que não tiveram classificação positiva tanto pela planilha quanto pelo classificador.

Segundo Treino

Coloca-se as novas classificações no Weka, com os mesmos parâmetros de treino anteriores. Constata-se um aumento no número de símbolos considerados pelo avaliador de 93 para 113 itens relevantes a fim de adequar os novos rótulos. O Código 3 mostra o resultado do novo modelo.

Código 3 – Métricas de qualidade do segundo treino de classificação

Correctly Classified Instances	1869	99.5208 %
Incorrectly Classified Instances	9	0.4792 %
TP Rate	0.995	
FP Rate	0	
Precision	1	
Recall	0.995	
F-Measure	0.997	
ROC Area	1	
Class	0.0	
Weighted Avg.	0.995	
a	b	<-- classified as
1709	9	a = 0.0
0	160	b = 1.0

Há expressiva melhora nos novos valores onde destaca-se a melhora no equilíbrio entre a cobertura e a precisão dada pela medida-F ($F_Measure$). Dos 9 casos de falsos positivos, nenhum de fato possui claramente descrita a avaria e troca do transformador,

porém a sua maioria contém informações sobre desligamentos, trocas de postes com transformador e trocas de TRs por aumento de carga, logo são de fato casos falsos. Tem-se então as limitações impostas pela avaliação do classificador e/ou seus parâmetros.

Otimização

Com a necessidade de diminuir o número de falsos positivos, treina-se novamente o conjunto de testes através da ferramenta *Weka Experiment Environment*, para outros parâmetros e/ou outros classificadores: **BayesNet**, **DMNBtext** (*Discriminative Multinomial Naive Bayes*), **J48** e **PART** (os dois últimos árvores de decisão). A configuração dos parâmetros dos algoritmos são feitas de forma a mantê-los os mais simples possíveis. Os treinos são realizados através de CV para 10 partições (*folds*). No caso do classificador BN, utiliza-se nova configuração com limite de nós pais aumentado para 2 (padrão 1), isto é, cada variável aleatória pode ser influenciada por outras duas. Os resultados são disponibilizados no Código 4.

Código 4 – Comparativo entre algoritmos classificadores: (1) BayesNet padrão, (2) BayesNet com máximo de dois pais, (3) DMNBtext, (4) J48 e (5) PART

	(1)	(2)	(3)	(4)	(5)
Percent_correct	99.50	99.69	99.55	99.11	99.01
Kappa_statistic	0.97	0.98	0.97	0.94	0.94 *
Area_under_ROC	1.00	1.00	1.00	0.97 *	0.97
F_measure	1.00	1.00	1.00	1.00	0.99
Num_false_positives	0.04	0.11	0.59 v	0.84 v	1.02 v
Num_false_negatives	0.90	0.48	0.25 *	0.83	0.83
Mean_absolute_error	0.01	0.00	0.05 v	0.01 v	0.01 v
False_positive_rate	0.00	0.01	0.04 v	0.05 v	0.06 v
False_negative_rate	0.01	0.00	0.00 *	0.00	0.00

Pelas métricas resultantes, verifica-se um melhor desempenho do algoritmo (2) – BN com pais limitados à 2 variáveis – em relação à medida Kappa (0,98), pequeno número de falsos positivos (0,11) e falsos negativos (0,48) além de possuir a maior porcentagem de acertos (99,69) no que se refere à CV. É interessante ressaltar que o algoritmo J48 possui uma árvore simples de classificação (podada) vista na Figura 25, a qual relaciona a palavra “suporte” com as observações de registros avariados. Segundo a árvore, consideram-se positivos os textos contendo essa palavra em decorrência da descrição do tipo de suporte utilizado para o TR substituído.

Pela guia *Visualize* do *Weka Explorer*, é feita a verificação qualitativa das relações entre as variáveis de texto. A Figura 26 mostra a ligação entre a palavra “suporte” (eixo X) e “transformador” (eixo Y), com valores que vão de 1 para verdadeiro e 0 para falso, referentes à presença dessas em uma tupla. A cor azul representa avaria positiva, oposto da vermelha.

Figura 25 – Árvore de decisão para o classificador J48

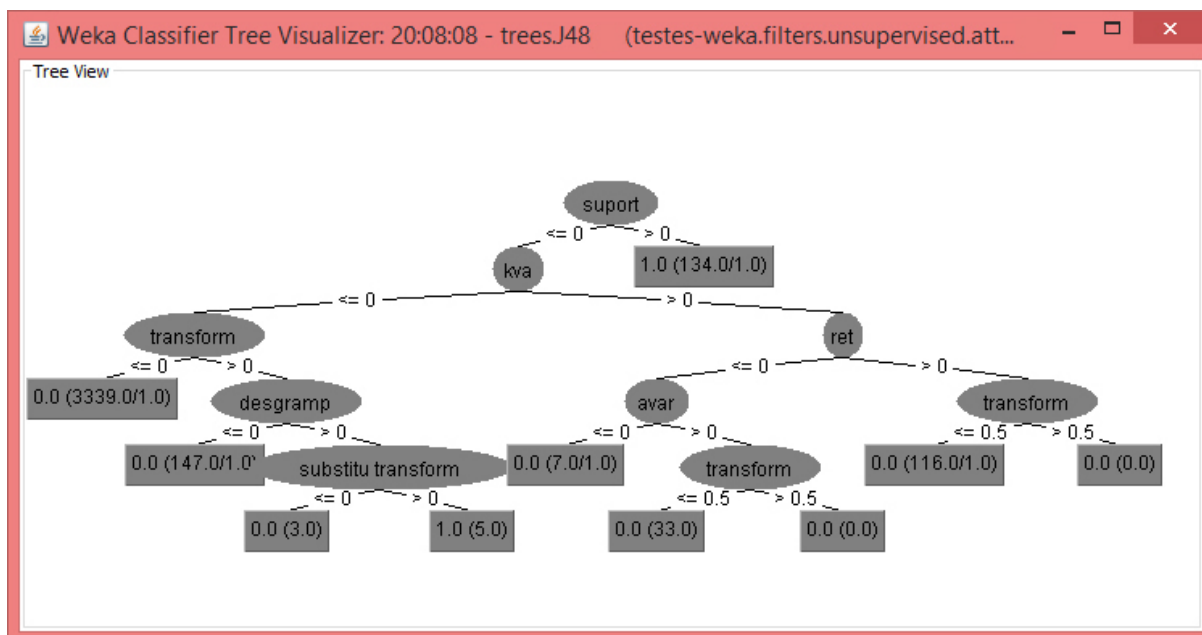
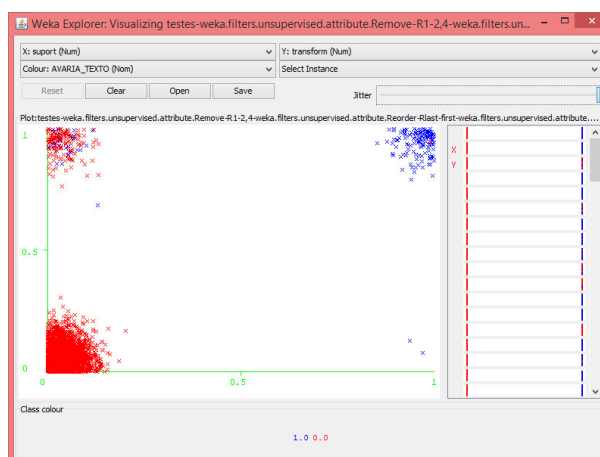


Figura 26 – Verificação qualitativa das relações entre variáveis de texto



No canto superior direito, os pontos acumulados em azul denotam que observações que contém ambas palavras são classificadas como positivas (em sua grande maioria). No canto superior esquerdo, a cor vermelha predominante demonstra que a palavra “transformador” desacompanhada de “suporte” não está presente na maioria dos casos positivos.

Modelo Final

Feitas as comparações, o algoritmo de aprendizagem escolhido – Rede Bayesiana para 2 pais – é analisado quanto aos dados de treino através de CV, onde o Código 5 dispõe o resultado.

A medida Kappa (0,98) mostra que os 10 modelos criados por CV apresentam uma “concordância quase perfeita”, bem como a área sob a curva ROC (1) mostra que

Código 5 – Métricas de qualidade do terceiro treino de validação cruzada

Correctly Classified Instances	1873									99.7338 %
Incorrectly Classified Instances	5									0.2662 %
Kappa statistic	0.9831									
Mean absolute error	0.0044									
Root mean squared error	0.0496									
Relative absolute error	2.7991									%
Root relative squared error	17.7571									%
Total Number of Instances	1878									
	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC	Area	Class		
	0.998	0.006	0.999	0.998	0.999	1	0.0			
	0.994	0.002	0.975	0.994	0.985	1	1.0			
Weighted Avg.	0.997	0.006	0.997	0.997	0.997	1				
a	b	<-- classified as								
1714	4	a = 0.0								
1	159	b = 1.0								

o modelo tem uma precisão “perfeita”. A cobertura (*recall*) próxima de 1 indica que praticamente todas as avarias corretas foram encontradas em relação a tudo que deveria ter sido encontrado. A precisão indica que a grande maioria dos registros foram rotulados corretamente em relação ao total detectado e a medida-F mostra que há equilíbrio entre a precisão e a cobertura. Através desses resultados verifica-se que o desbalanceamento entre as classes não representa um problema na determinação dos rótulos.

5.3 APLICAÇÃO DO MODELO

Encontrado o modelo final, esse está pronto para classificar outros registros. Para isso, além dos dados de treino referentes ao período de janeiro de 2014, utiliza-se outros meses: janeiro de 2013 e março e abril de 2014, com mesmas características: ocorrências não programadas, alimentadores e eletricitas. Registros sem observações são considerados negativos, visto que todos os registros da planilha da oficina dispunham de observações nas ocorrências. Os resultados pretendidos não abrangem todas as avarias posto que os dados utilizados são referentes à manutenção, logo não incluem causas relacionadas à meio de terceiros por exemplo.

Dados de Treino

Apenas 64,8% (105 de 162) dos casos de avarias reais foram registradas pela oficina para o período de janeiro de 2014, como mostra a MC na Tabela 8. Demonstra-se que 96,3% (105 de 109) do total de seus registros são confirmados como positivos pelo algoritmo, ou seja, retoma a confiabilidade tanto da planilha fornecida quanto do classificador. A diferença de 3,67% (4 casos) se dá ou pelo não apontamento da avaria nas observações ou por manutenções preventivas nos equipamentos. É importante ressaltar

que em alguns casos os textos são dúbios, suscetíveis a interpretação do usuário.

Tabela 8 – Matriz de confusão entre registros da oficina e classificados por BN para janeiro de 2014

Oficina		jan/2014		
0	1			
1712	4	0		BN
57	105	1		

xcvxc

No mesmo período a MC da Tabela 9 mostra que, coincidentemente, a quantidade de registros contendo o serviço 304 e rotulados positivos pelo classificador BN é idêntica (162). Porém, 26 desses registros são incoerentes com as descrições, um erro de 16% (26/162) na identificação de TR avariados. Em relação ao serviço 304, grande quantidade dos falsos positivos (13) é relacionada à troca de TAP e pequenos defeitos do equipamento TR, normalizados in loco. Já para os falsos negativos (13), 4 são observações em branco e os demais são de fato TRs avariados. Um ótimo desempenho, já esperado tratando-se dos dados de treino.

Tabela 9 – Matriz de confusão entre registros com serviço 304 e classificados por BN para janeiro de 2014

304		jan/2014		
0	1			
1703	13	0		BN
13	149	1		

Demais Registros

Para janeiro de 2013, período de características semelhantes aos dados de treino, 27 registros são considerados errados pelo classificador, como mostra a MC na Tabela 10. A partir do serviço 304, considera-se um erro de 29% (27/93) dos registros envolvendo avarias em transformadores. Sete dos 15 casos de falsos positivos devem-se à ausência de observações. Em março e abril de 2014 (Tabelas 11 e 12) o número de avarias diminui se comparados aos meses do verão, isso torna os erros mais significativos. Para esses meses são 8 e 3 respectivamente o número de ocorrências denominadas como avarias que não possuem observações.

Do serviço 304 em relação à classificação feita pelo algoritmo BN, considerando um erro absoluto relativo de de 2,8%, a Tabela 13 compara os verdadeiros e falsos, positivos e negativos para os meses analisados.

Tabela 10 – Matriz de confusão entre registros com serviço 304 e classificados por BN para janeiro de 2013

304		jan/2013		BN
0	1	0	1	
2543	15	0		
12	78	1		

Tabela 11 – Matriz de confusão entre registros com serviço 304 e classificados por BN para março de 2014

304		mar/2014		BN
0	1	0	1	
3109	12	0		
5	58	1		

Tabela 12 – Matriz de confusão entre registros com serviço 304 e classificados por BN para abril de 2014

304		abr/2014		BN
0	1	0	1	
2851	7	0		
7	42	1		

Verifica-se que independente do total de avarias, há uma constância na dimensão dos registros falsos até março, mês anterior à implementação do serviço 304 com inclusão do termo "AVARIADO" à nomenclatura. O histograma na Figura 27 ilustra melhor esse aspecto. Recordando o erro de aproximado de 3% na classificação do modelo BN, para o histograma na Figura 28, mais uma vez é explanado o contínuo registro de erros de preenchimento, tendendo à queda para o mês de abril. Assim, tem-se uma melhora tanto nos registros onde deixou-se de registrar a avaria quanto nos serviços onde o código foi marcado com intuito de representar outra atividade.

Figura 27 – Comparativo dos registros de avarias com código 304

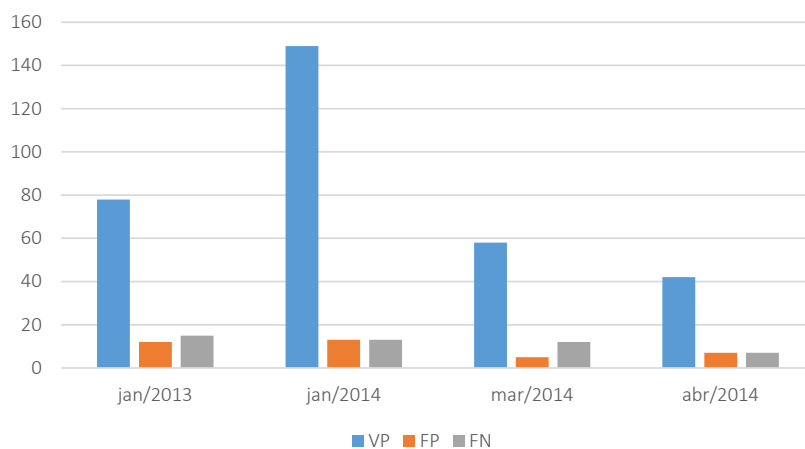
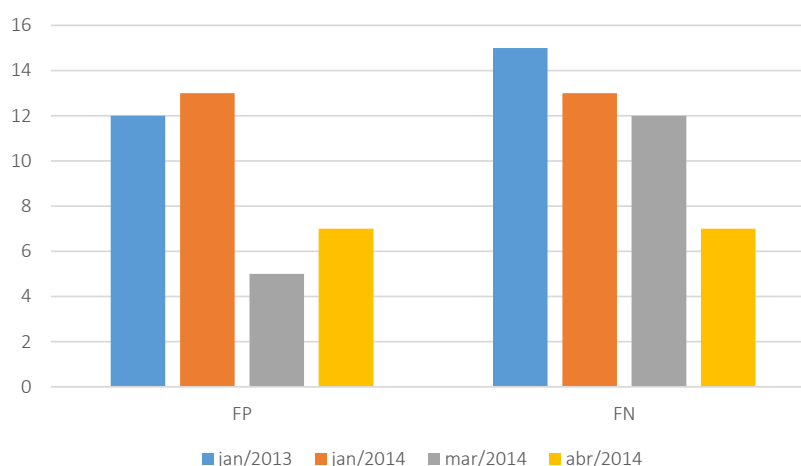


Tabela 13 – Comparativo entre verdadeiros positivos (VP), falsos positivos (FP), falsos negativos (FN) e verdadeiros negativos (VN)

Mês	VP	FP	FN	VN
jan/2013	78	12	15	2543
jan/2014	149	13	13	1703
mar/2014	58	5	12	3109
abr/2014	42	7	7	2851

Figura 28 – Comparativo dos falsos registros de avarias dados pelo código 304



5.4 DETERMINAÇÃO DOS PONTOS CRÍTICOS

Das ocorrências onde o modelo classificador BN rotulou como positivas para avaria nas observações no mês de janeiro de 2014, faz-se o levantamento dos códigos das UTs. Os dados históricos são organizados por mês, no período de janeiro de 2009 a janeiro de 2014, onde considera-se apenas um registro de avaria por mês.

O levantamento das reincidências aponta que apenas 58% dos casos de avarias ocorre pela primeira vez, sendo que 18% apresentaram 2 ou mais avarias nos últimos 5 anos, conforme demonstrado no gráfico da Figura 29 (a). A árvore de decisão (vista anteriormente na Figura 20 pag. 56) aplicada aos dados resulta em altos índices de criticidades anormais (42%) como mostra o gráfico da Figura 29 (b).

Para a mesma UT, há condições de avarias registradas em diferentes ocorrências. Ao todo são 12 casos, onde 2 desses são vistos na Tabela 14. A leitura das observações explicita a condição de dano ao equipamento (vazamento de óleo) onde determinou-se como causa o mesmo. Assim como em outros registros semelhantes a esses, desatende-se o histórico, mesmo que a curtíssimo prazo, evidenciado pela observação da ocorrência 46171 onde é claro o domínio dessa informação por parte da equipe designada.

Nesses exemplos é perceptível um fator inerente às condições as quais os equipamentos são submetidos no local. Segundo as regras de classificação utilizadas, determinou-

Figura 29 – Quantidades de (a) UTs reincidentes e (b) UTs rotuladas normais, críticas e muito críticas

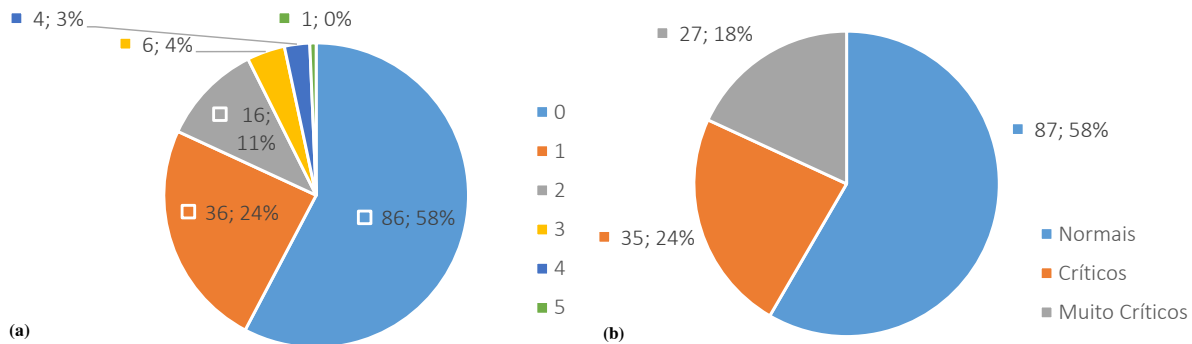


Tabela 14 – Casos de avarias ocorridas em curtos períodos de tempo

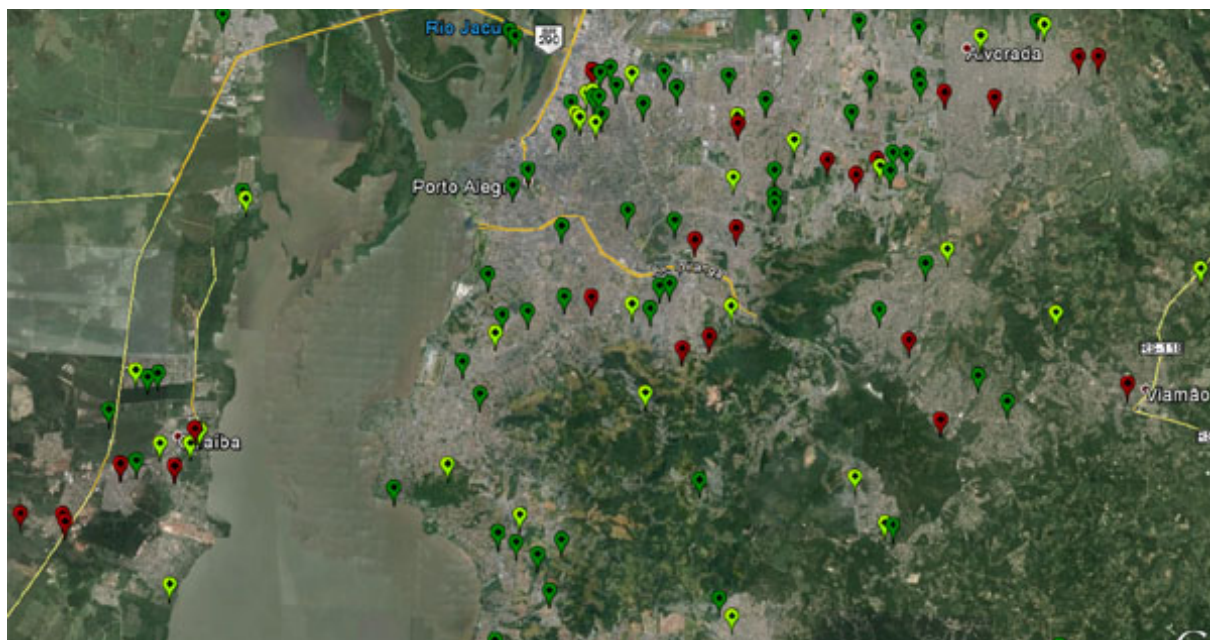
Ano Ocor.	UT	Causa	Observação	Data	Crit.
2014 42873	93016253	603	substituir tr 3x112,5 kva vazando óleo, ficou em linha, suporte poste de concreto esquina da rua sao jose com rua conego scherer ficou ligado porem vazando bastante óleo cc. 16253 contrafo retirado 35038-j contrafo instalado 115646-c	20/01/14 16:50	C
2014 46171	93016253	603	tr queimado com vazamento de óleo - rua sao jose, 417 - centro - 3x112,5kva - suporte poste de concreto - informado por pacheco trocado tr 3x112,5kva por 150kva cc. 16253 o mesmo havia queimado ontem contrafo retirado 115646- contrafo instalado 688635-f	21/01/14 16:30	C
2014 50965	149013192	603	tr avariado 3x75, suporte poste de madeira com tala, localizado na rua cel inacio quadros 50 instalado 9329-j retirado 53334-a	21/01/14 19:22	C
2014 54381	149013192	603	-substituir tr queimado 3x75 kva suporte poste de madeira, rua cel inacio de quadros numeral 93	23/01/14 14:40	C

se a UT 093016253 com condição crítica (C) devido a reincidência no mês de março de 2012 e a UT 149013192 em razão de uma avaria em janeiro de 2013. A imagem na Figura 30 mostra a disposição dos [pontos no mapa](#) da região metropolitana através do software *Google Earth*.

Realizando-se uma análise qualitativa dos resultados obtidos para o município de Porto Alegre, os locais muito críticos concentram-se na sua grande maioria em regiões periféricas urbanas com problemas socioeconômicos – bairros São José, Bom Jesus, etc. Para os rotulados normais, há predominância de regiões centro-periféricas de classe média. Já a área sem marcações no meio da área urbana de Porto Alegre compreende bairros nobres – bairros Rio Branco, Bela Vista e Mont Serrat. Ainda é possível constatar aglomerados de ocorrências de avarias na zona norte, região industrial leve – bairros Floresta,

São Geraldo e Navegantes.

Figura 30 – Classificações normais (verdes), críticas (amarelas) e muito críticas (vermelhas)



6 CONCLUSÕES

As empresas de distribuição de energia elétrica possuem uma grande quantidade de dados oriunda de procedimentos de operação e manutenção. Entretanto, a transformação desses dados em modelos de análises de informações ainda são pouco explorados frente as suas potencialidades para suporte a decisão, podendo ser considerados tardios não só na concessionária CEEE-D mas também na maioria das empresas. Em geral, muitos estudos e avaliações, mesmo que repetitivos, ainda são efetuados de forma manual, ocasionando dispêndios de tempo e recursos humanos. Procedimentos triviais como a simples composição de uma planilha de avarias podem conter pequenos desacertos, os quais com o tempo induzem imprecisões e podem ocasionar desperdício de recursos financeiros e impactar na margem de remuneração dos ativos da empresa.

A proposição da implementação de metodologias que utilizam descoberta de conhecimento no banco de dados de manutenção da concessionária CEEE-D, com intuito de reproduzir atividades já existentes e desempenhadas por técnicos de forma manual, mostrou-se eficiente. Os textos analisados, não submetidos a um padrão de preenchimento e apontados por técnicos diversificados, apresentaram-se em sua grande maioria homogêneos quanto as avarias em transformadores, justificáveis pelo bom desempenho do algoritmo classificador. Por tratar-se de um procedimento autônomo, sua execução agiliza, valida e acima de tudo dispensa especialistas para essas funções. Na otimização do modelo classificador, o aprendizado por árvore de decisão retornou a palavra “suporte” como importante para a avaliação das observações, fator que elucida os conceitos de extração do conhecimento (expressão não alusiva à avaria), visualização (composição gráfica da árvore acessível visualmente), e iteração (inicialmente considerada irrelevante, passa a integrar o enfoque antes somente concedido às palavras circunvizinhas à “transformador”). Nos resultados obtidos, verifica-se que os erros são independentes da quantidade total de casos de avarias nos períodos analisados, também mostra uma significativa melhora no preenchimento para o mês de abril (o qual teve a nomenclatura do serviço melhorada). Os poucos falsos resultados encontrados são, na sua maioria, passíveis de interpretação de especialistas, ou seja, possuem alguns traços de avarias mas não são claros o suficiente (como pode ser vistos os registros no Apêndice B).

De forma simples, a árvore de decisão imposta na determinação do grau de criticidade dos pontos expõe em seus resultados que as causas atribuídas às avarias em transformadores não se limitam apenas às características do evento, mas também às condições peculiares aos locais de instalação (unidades transformadoras). Assim, na condição atual, a frequência relativa de danos justificados com os próprios equipamentos é inapta de

interpretação, ou seja, muitos desses registros tratam de locais com iminente sobrecarga por motivos como roubo de energia e falhas de operação, sendo esse último ao considerar a troca de equipamentos desprezando os precedentes. A evidência dos lugares críticos altera a percepção de avarias antes compreendidas como “transformadores defeituosos em locais aleatórios” para “locais suscetíveis de defeitos em transformadores”, dessa forma mais palpável o problema a ser combatido.

Não é à toa que o software Weka se consolidou como a ferramenta de mineração de dados mais utilizada em ambiente acadêmico. Livre e de código aberto, seu ponto forte é a classificação disponibilizada por dezenas de algoritmos, os quais demandaram conhecimentos básicos para sua utilização. O módulo Weka Experimenter propiciou a execução de comparativos entre estes algoritmos de maneira a expor seus pontos fortes. Redes Bayesianas, o método probabilístico escolhido, mostrou-se bastante eficiente em identificar tanto textos extensos quanto incompletos.

Para os objetivos definidos, utilizou-se nas classificações a aplicação de algoritmos de aprendizagem (onde a máquina encontra o padrão) e a determinação de regras empíricas (padrão consolidado por um especialista). Em ambos, as aplicações se dão de forma praticável e módica, isto é, possuem condições propícias de aplicabilidade e dispensam o emprego de supercomputadores e/ou softwares privados. O conhecimento (informação útil) pode estar presente em toda e qualquer forma de dado armazenado – textos, dados ou inter-relacionados – e a CEEE-D possui um imenso potencial para o emprego desses recursos, os quais podem vir a ter grande auxílio na diminuição das ocorrências não programadas.

SUGESTÕES PARA TRABALHOS FUTUROS

Visto que para cada registro utilizado neste trabalho tem-se em média o cadastro de apenas 1,5 códigos de serviços realizados, os algoritmos utilizados na recuperação de informações de TRs avariados pode ser conduzida para reaver outros procedimentos, e assim completar as ocorrências encerradas.

Quanto à criticidade, um aumento suficiente no número de locais de pode definir regiões críticas, além dos pontos encontrados. Essas, organizadas de forma a mapear problemas conforme as causas: locais com problemas ambientais, sociais e com falta de manutenções por exemplo.

Para as planilhas fornecidas pela oficina, há possibilidade de correção dos códigos através de comparações com os cadastrados e posterior preenchimento do módulo de controle e com isso centralizar a informação em um banco de dados.

A identificação dos usuários que possuem mais registros errôneos é ideal para de forma focalizada alinhar as regras de preenchimento.

REFERÊNCIAS

- Agência Nacional de Energia Elétrica. *Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST: Qualidade da energia elétrica – módulo 8*. [S.l.], 2010. 70 p. 24
- Agência Nacional de Energia Elétrica. *Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST: Cartilha de acesso ao sistema de distribuição*. [S.l.], 2012. 26 p. 29
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 7282: Dispositivos fusíveis de alta tensão – dispositivos tipo expulsão – requisitos e métodos de ensaio*. 2. ed. Rio de Janeiro, 2011. 87 p. 32, 33
- BALTAR, V. T.; OKANO, V. *Análise de Concordância*. 201–? Laboratório de Epidemiologia e Estatística – Lee. Disponível em: <<http://www.lee.dante.br/pesquisa/kappa/#author>>. Acesso em: 26 de maio de 2014. 55
- BERRY, M. J. A.; LINOFF, G. S. *Mastering Data Mining: The art and science of customer relationship management*. 1. ed. New York, NY, USA: Wiley, 2000. 494 p. (Wiley Computer Publishing). 45
- CAVNAR, W. B. Using an n-gram-based document representation with a vector processing retrieval model. In: *Overview of the Third Text REtrieval Conference (TREC-3)*. Gaithersburg, MD, USA: [s.n.], 1995. p. 269–277. 46
- CHAPMAN, S.; LASCHUK, A. *Fundamentos de máquinas elétricas*. 5. ed. Porto Alegre: Mcgraw Hill - Artmed, 2013. 30
- COMPANHIA ESTADUAL DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA. *TTD-00.002: Termos relacionados com operação e manutenção de linhas e redes aéreas de distribuição*. Porto Alegre, 2006. 15 p. 25, 35
- COMPANHIA ESTADUAL DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA. *PTD-00.004: Padronização de distribuição: Estruturas para equipamentos*. Porto Alegre, 2008. 64 p. 29
- COMPANHIA ESTADUAL DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA. *PRAT: Programa de redução de avarias em transformadores*. Porto Alegre, 2009. 87 p. 24
- COMPANHIA ESTADUAL DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA. *Manual do Eletricista de Rede Aérea de Distribuição*. 1. ed. Porto Alegre, 2011. 333 p. 38, 39
- COMPANHIA ESTADUAL DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA. *ETD-00.001: Especificação de distribuição: Transformador de distribuição*. Porto Alegre, 2013. 18 p. 30
- COMPANHIA ESTADUAL DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA. *NTD-00.001: Elaboração de projetos de redes aéreas de distribuição urbanas*. Porto Alegre, 2013. 28 p. 32, 34

COMPANHIA ESTADUAL DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA.
PTD-00.001: Padronização de distribuição: Materiais para redes aéreas de distribuição - equipamentos e proteções. Porto Alegre, 2013. 171 p. 30, 31, 33, 34

COMPANHIA ESTADUAL DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA.
Institucional. [S.l.], 2014. Disponível em: <<http://www.cee.com.br/>>. Acesso em: 15 de abril de 2014. 27

COMPANHIA ESTADUAL DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA.
NT-84.001: Operação da distribuição. Porto Alegre, 2014. 35

COMPANHIA ESTADUAL DE ENERGIA ELÉTRICA. *CPSD*: Curso de proteções do sistema de distribuição. Porto Alegre, 2008. Volume 1, 13 p. 33, 34

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. Advances in knowledge discovery and data mining. In: FAYYAD, U. M. et al. (Ed.). Menlo Park, CA, USA: American Association for Artificial Intelligence, 1996. cap. From data mining to knowledge discovery: an overview, p. 1–34. 44, 45

GU, Q.; CAI, Z.; ZHU, L. Classification of imbalanced data sets by using the hybrid re-sampling algorithm based on isomap. In: CAI, Z. et al. (Ed.). *ISICA*. [S.l.]: Springer, 2009. (Lecture Notes in Computer Science, v. 5821), p. 287–296. 53

HALL, M. et al. The WEKA data mining software: An update. *SIGKDD Explorations*, ACM, New York, NY, USA, v. 11, n. 1, p. 10–18, Nov. 2009. ISSN 1931-0145. 57, 58

HAN, J. *Data mining: Concepts and techniques*. Amsterdam Boston San Francisco, CA: Elsevier Morgan Kaufmann, 2006. 770 p. 44, 45, 46, 47, 48, 49

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. *Grupo de Eletricidade Atmosférica – ELAT*. 2014. Disponível em: <<http://www.inpe.br/webelat/homepage/menu/infor/relampagos.e.efeitos/sistema.eletrico.php>>. Acesso em: 5 março 2014. 39

JUNIOR, J. C.; IMAMURA, C.; REZENDE, S. Avaliação de um algoritmo de stemming para a língua portuguesa. *Proceedings of the 2nd Congress of Logic Applied to Technology (LABTEC 2001)*, v. 2, p. 267–274, 2001. Disponível em: <<http://www.nilc.icmc.usp.br/nilc/tools/stemmer.html>>. 60

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995. (IJCAI'95), p. 1137–1143. 51

KOMOROWSKI, M. *A history of storage cost*. 2009. Disponível em: <<http://www.mkomo.com/cost-per-gigabyte/>>. Acesso em: 2 de maio de 2014. 43

LANDIS, J. R.; KOCH, G. G. et al. *The measurement of observer agreement for categorical data*. [S.l.]: International Biometric Society, 1977. 159–174 p. 55

Machine Learning Group at the University of Waikato. *Weka 3: Data mining software in java*. Hamilton, New Zealand: [s.n.], 2014. Disponível em: <<http://www.cs.waikato.ac.nz/~ml/weka/>>. Acesso em: 8 de março de 2014. 57

- MAIMON, O.; LAST, M. *Knowledge Discovery and Data Mining: The info-fuzzy network (ifn) methodology*. Dordrecht Boston: Springer, 2001. (Massive Computing). ISBN 9780792366478. 51
- MCLELLAN, C. *Storage in 2014: An overview*. 2014. Disponível em: <<http://www.zdnet.com/storage-in-2014-an-overview-7000024712/>>. Acesso em: 2 de maio de 2014. 43
- MINER, G.; ELDER, J.; HILL, T. *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*. [S.l.]: Academic Press, 2012. (Academic Press). 46
- OLSON, D. L.; DELEN, D. *Advanced Data Mining Techniques*. Berlin: Springer Berlin Heidelberg, 2008. 192 p. 52
- PASQUALOTTO, A. C. et al. *Métodos Diagnósticos: Consulta rápida*. 2. ed. Porto Alegre: Artmed Editora, 2012. 1167 p. 54
- PEREIRA, A. da C. et al. *Avaliação de Testes Diagnósticos*. 2014. Faculdade de Medicina da Universidade do Porto. Disponível em: <http://stat2.med.up.pt/cursop/print_script.php3?capitulo=tdiagnosticos&numero=4&titulo=Avaliacao%20de%20testes%20de%20dianostico>. Acesso em: 4 de maio de 2014. 54
- PORTER, M. F. *Snowball: A language for stemming algorithms*. 2001. Disponível em: <<http://snowball.tartarus.org/texts/introduction.html>>. Acesso em: 14 de março de 2014. 60
- PRETTO, C. O. et al. Utilização de computação móvel para qualificação de rotinas de operação e manutenção de redes de distribuição. *Sba: Controle & Automação Sociedade Brasileira de Automatica*, scielo, v. 17, p. 446 – 458, dez. 2006. ISSN 0103-1759. 26
- REZENDE, S. *Sistemas inteligentes: fundamentos e aplicações*. 1. ed. Barueri: Manole, 2003. 525 p. ISBN 9788520416839. 48, 51
- RUSSELL, S.; NORVIG, P. *Inteligência artificial*. Rio de Janeiro: Elsevier Campus, 2004. 1021 p. 49, 51
- SILVA, G. N. da. *Jogos de Licitações: O estado da arte no brasil*. 1. ed. São Paulo: Biblioteca 24 horas, 2010. 72 p. 50
- SMOLINSKI, T. G.; MILANOVA, M. G.; HASSANIEN, A.-E. *Computational Intelligence in Biomedicine and Bioinformatics: Current trends and applications*. Berlin: Springer, 2008. 432 p. (Studies in Computational Intelligence). 52
- TAN, A.-H. Text mining: The state of the art and the challenges. *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*, p. 65–70, 1999. 46
- TRONCHONI, A. B. *Identificação de causas de desligamentos não programados em redes de distribuição*. Dissertação (Mestrado) — Pontifícia Universidade Católica Do Rio Grande do Sul, Porto Alegre, mar. 2008. 39

TRONCHONI, A. B. et al. Descoberta de conhecimento em base de dados de eventos de desligamentos de empresas de distribuição. *Sba: Controle & Automação Sociedade Brasileira de Automatica*, scielo, v. 21, p. 185 – 200, abr. 2010. ISSN 0103-1759. 26

VIDINICH, R.; NERY, G. A. L. Pesquisa e desenvolvimento contra o furto de energia. *Revista Pesquisa e Desenvolvimento da ANEEL*, n. 3, p. 15, jul. 2009. ISSN 11981-9803. 40

WITTEN, I. H. Text mining. *Practical handbook of Internet computing*, Chapman and Hall/CRC Press, Boca Raton, Florida, p. 14–1, 2005. 46

APÊNDICE A – RESULTADOS COMPLETOS DOS TREINOS

Nos Códigos 6, 7 e 8 os resultados completos dos três treinos de classificação vistos anteriormente nas páginas 61, 63 e 66 respectivamente.

Código 6 – Métricas de qualidade do primeiro treino de classificação (resultado completo)

```

=== Evaluation on training set ===
=== Summary ===
Correctly Classified Instances    1816                96.6986 %
Incorrectly Classified Instances    62                 3.3014 %
Kappa statistic                    0.755
Mean absolute error                 0.034
Root mean squared error             0.1816
Relative absolute error             30.9339 %
Root relative squared error        77.6564 %
Total Number of Instances          1878
=== Detailed Accuracy By Class ===
      TP Rate FP Rate Precision Recall  F-Measure  ROC Area  Class
      0.967  0.037   0.998  0.967    0.982    0.98   0.0
      0.963  0.033   0.644  0.963    0.772    0.98   1.0
Weighted Avg. 0.967  0.036   0.977  0.967    0.97     0.98
=== Confusion Matrix ===
  a    b  <-- classified as
1711   58 |    a = 0.0
  4   105 |    b = 1.0

```

Código 7 – Métricas de qualidade do segundo treino de classificação (resultado completo)

```

=== Evaluation on training set ===
=== Summary ===
Correctly Classified Instances    1869                99.5208 %
Incorrectly Classified Instances    9                  0.4792 %
Kappa statistic                    0.97
Mean absolute error                 0.0053
Root mean squared error             0.0683
Relative absolute error             3.3851 %
Root relative squared error        24.4549 %
Total Number of Instances          1878
=== Detailed Accuracy By Class ===
      TP Rate FP Rate Precision Recall  F-Measure  ROC Area  Class
      0.995  0         1      0.995    0.997    1       0.0
      1     0.005   0.947  1       0.973    1       1.0
Weighted Avg. 0.995  0         0.995  0.995    0.995    1
=== Confusion Matrix ===
  a    b  <-- classified as
1709   9 |    a = 0.0
  0  160 |    b = 1.0

```

Código 8 – Métricas de qualidade do terceiro treino de classificação (resultado completo)

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      1873                99.7338 %
Incorrectly Classified Instances      5                   0.2662 %
Kappa statistic                      0.9831
Mean absolute error                  0.0044
Root mean squared error              0.0496
Relative absolute error              2.7991 %
Root relative squared error          17.7571 %
Total Number of Instances            1878
=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  ROC  Area  Class
                0.998   0.006   0.999     0.998   0.999      1    0.0
                0.994   0.002   0.975     0.994   0.985      1    1.0
Weighted Avg.  0.997   0.006   0.997     0.997   0.997      1
=== Confusion Matrix ===
  a    b  <-- classified as
1714   4 |    a = 0.0
  1  159 |    b = 1.0

```

O resultado completo para comparativo entre algoritmos classificadores pode ser visto no Código 9, antes visto na página 64.

Código 9 – Comparativo entre algoritmos classificadores (resultado completo)

```

                (1)      (2)      (3)      (4)      (5)
Percent_correct  99.50 |  99.69  99.55   99.11  99.01
Kappa_statistic  0.97 |  0.98   0.97   0.94   0.94 *
Area_under_ROC   1.00 |  1.00   1.00   0.97 *  0.97
F_measure        1.00 |  1.00   1.00   1.00   0.99
Num_false_positives  0.04 |  0.11  0.59 v  0.84 v  1.02 v
Num_false_negatives  0.90 |  0.48  0.25 *  0.83   0.83
Mean_absolute_error  0.01 |  0.00  0.05 v  0.01 v  0.01 v
False_positive_rate  0.00 |  0.01  0.04 v  0.05 v  0.06 v
False_negative_rate  0.01 |  0.00  0.00 *  0.00   0.00
Key:
(1) bayes.BayesNet '-D -Q bayes.net.search.local.K2 -- -P 1 -S BAYES
-E bayes.net.estimate.SimpleEstimator -- -A 0.5' 746037443258775954
(2) bayes.BayesNet '-D -Q bayes.net.search.local.K2 -- -P 2 -S BAYES
-E bayes.net.estimate.SimpleEstimator -- -A 0.5' 746037443258775954
(3) bayes.DMNBtext '-I 1' 5932177450183457085
(4) trees.J48 '-C 0.25 -M 2' -217733168393644444
(5) rules.PART '-M 2 -C 0.25 -Q 1' 8121455039782598361

```

APÊNDICE B – OBSERVAÇÕES TEXTUAIS

A Tabela 15 mostra registros divergentes entre os treinos (BN1, BN2 e BN3), planilha da oficina (Of), serviço 304 (304) e classificação assistida (AS).

Tabela 15 – Exemplos de observações das ocorrências encerradas

Ano	Ocor.	304	Of	BN1	BN2	BN3	AS	Observação
2014	2421	0	0	1	1	1	1	rede desregulada. - por favor fazer ordem p instalac?o de espaçadores no cc 34 !rede desregulada. tr com vazamento de óleo, 3x45kva, cc 34, suporte, poste de concreto. ficou aberto com elos trocados, desgrampeado. regular rede e instalar espaçadores. trocar dois para-raios. bt 3#2(4)ca. av getulio vargas, 166.
2014	22026	0	0	1	1	1	1	- fase a, gasto 1 elo 1h trocar para-raios fase a tr avariado, cc 12226, 3x15kva, suporte poste de madeira. facil aceso caminhao. rua da conquista, 10 - assentmaneto dezenove de setembro. um grampo esta emperrado e nao foi aberto. foi trocado tr contrafo ret 160735 j 3x15kva inst 27294 j 3x30kva
2014	61215	0	0	1	1	1	1	poste m11 u4, cabo 4caa, ea, haste francklin tr 16083 1x10 kva, poste com tr queimou e quebrou e esta tombado no chao, ultimo poste da rede de mt, gps 181065, ficou aberta passagem de mt no gps 181078 que isola apenas este tr retirado 27266b instalado 57868d
2014	86340	0	0	1	1	1	1	[f358622 - 31/01/2014 18:46:09] - passado 4para valdirnei 99161047 caminhao [f164160 - 31/01/2014 21:08:27] - solicitada o apoio da equipe 9095 ao apoio da eq. ceee [f352306_opr - 01/02/2014 11:01:48] - ***** equipe do caminhao foi agredida no local, quebraram vidro do caminhao, estao na delegacia prestando queixa, nao irao retornar ao local durante a noite substituir poste m11 gps45991 quebrado na base. est b2 e c2 quebradas. rede de mt e bt cabo 2ca, ip e telefonia. substituir tr 3x45kva avariado. cc23324 - ch1661 ficou fechada e aberta conexao rigida na derivacao para cc23324.. rua joao inacio da silveira, 115 substituir ch e cruzeta do cc23329 quebrada. cc23329 ficou aberto. rua joao inacio da silveira, 1500 equipe do caminhao foi agredida no local, quebraram vidro do caminhao, estao na delegacia prestando queixa, nao irao retornar ao local durante a noite
2014	1527	1	1	0	1	1	1	- trocar cc 5928 - suporte - concreto - 3 x 45 - vazando óleo - rua pedro a de Freitas filho, 70 .
2014	52697	1	0	0	1	1	1	tr avariado 3x75, suporte poste de medeira com tala, localizado na rua cel inacio quadros 50 instlado 9329-j retirado 53334-a

Tabela 15 – Exemplos de observações das ocorrências encerradas (continuação)

Ano	Ocor.	304	Of	BN1	BN2	BN3	AS	Observação
2014	77120	1	0	0	1	1	1	[f350907 - 30/01/2014 04:42:09] - repassado para equipe de plantao trocar tr do cc 45 de 3 x 75 kva queimado suporte em c11 rua antonio fideles x rua belo amorin de abreu — atras da delegacia
2014	53609	0	0	1	1	1	0	refazer cruzamento da pernambuco com cairu junto ao poste do circuito. cabo bt 120mm multiplexado substituir tr 3x112,5kva poste concreto suporte cabo bt 120mm cabo de saida. 70mm avenida pernambuco em frente 1328
2014	70675	0	0	1	1	1	0	dois postes m9 quebrados na base nao aceitam tala gps 27973 e 27174 m9-hc-bt4ca passantes com acesso para caminhao aprumar m11 sem gps dois vaos apos a chave 3335 da est anizio jose goulart 19105 cc 4410 fciou aberto cc 230004410 contrafo inst 30190 g ret 598122
2014	77405	0	0	1	1	1	0	passagens se bateram pois chaves cairam do suporte. trocar com urgencia, poste do tr esta podre, estruturas tambem, ficou desligado e desgrampeado. m11-n3-c2-tr-3as11-ip-crt, cc 4180 pode colocar poste de concreto. trocar poste, estruturas completas, chaves e para-raios, tr bifascio, de suporte. bt1#4(4)ca mt 2#4caa rua hortencia, 580 - duas quadras antes do mercado oliveira entrar a esquerda. pda 96. foi colocado tala e refixadas chaves.
2014	40562	0	1	1	1	0	1	poste m9 ja possui tala necessita troca urgente. gps do poste 028532 cabo da bt 4ca 2 as 22 padronizar jogo de chaves na acriso prates, altura num 7914 equipe ao soltar estrutura do poste a ser substituido, poste m9 anterior tombou sobre o caminhao, levando em sequencia mais dois postes m11, incluido o poste do cc 4399, rompendo rede mt cabo 4ca. retirado 43045/5 instalado 24354a
2014	65575	0	0	1	1	0	0	tr circ 18 estava atuado 3 fases rede desregulada se bateu atuando tr agencia vai regular rede e colocar espacadores e uma tala provisoria no poste programar substituicao de poste m11-n1-c2-tr112,5kva circ 18 -4 as22-ip-tel - mt 1/0ca - bt 1/0ca - gps 235178 rua silvana narvaez esq rua marcal miranda
2014	81196	0	0	1	1	0	0	desarme do al, 50,51,a,b,c, 79 dois postes m9 quebrados, necessaria troca, pode colocar concreto 4kn, rede passante, cabo 2ca. emendar um va oda fase a da rede bt, mt nao arrebeitou. quebrado poste poste m11 com tr, cc 10440. colocar poste de concreto 6kn, eh em fim de rede. trocar chaves, para-raios, estruturas. m11-n3-c2-tr-ecy1, fim de rede mt. rua onze, num 15 ao 90. quebrou poste do clienet da casa num 75. xxxxxxxxxxxxxxxx esta com gaspar servico abaixo deve ser repassado gps 7982 7983 7984 7988 postes podres na base m9-2as22-bt(2ca) nao aceitam tala todos sao passantes aceitam duplo t poste 7985 m9-2as22-bt(2ca) fim de cc bt refazer cruzamento de bt rua onze esq com rua 13 jardim salome cabo bt 2ca mais 4 dt 9400 para ser trocado, mais cruzamento.
2014	72761	1	0	0	1	0	0	equipe constatou queda de tensao no cc, o mesmo havia sido substituido anteriormente, foi necessario desliga-lo para mudanca do tap do mesmo equipe realizou mundaca de tap do tr 127 localizado na av salgado filho 1091, mas tensoa no cliente quando com carga ligada esta com 198v, tr no local e de 45kva, verificar possibilidade de desdobrar cc ou substituicao de tr por 75kva

Tabela 15 – Exemplos de observações das ocorrências encerradas (continuação)

Ano	Ocor.	304	Of	BN1	BN2	BN3	AS	Observação
2014	86369	0	0	1	1	0	0	efetuado poda em vegetais de grande porte caídos sobre a mt em 3 pontos diferentes foi reposto 2 vaos 2 condutores do cc 4182 e substituído baixadas gps 28625 repor 3vaos 1 condutor cabo 4 caa condutores estão queimados substituir vaos +ou- 100 metros de cabo bec do pesqueiro 1785 - - ***** cheg 8h44 foram emendados e refeitos 3 vaos de mt, gasto 50m 4caa, 3 alca e 1k1 feito podas e regulado fase e neutro da rede bt do cc 4183, gasto 2k1 poste podre, programar troca com brevidade. tem acesso caminhão, fica dentro da propriedade. aceita tala. substituir chave e para-raios. tr 1x10kva. m11-ul-tr-2as11, bt e mt passantes, cabo 4caa. beco do pesqueiro, 1790. tr pode ser visto da porteira, tem caseiro no local, só buzinar. gps 26670, tem derivação monobucha com conexão rígida que atende este ramal de mt do cc 4183, instalar chave fusível.
2014	4601	0	0	0	1	0	0	devido poste concreto cônico instalado para passar rede futuramente, rede de mt arrebentou ao bater no poste, já que não foi colocado nenhum tipo de proteção feito 1 vao fase c na mt também refazer 1 vao de neutro, e barramento do tr 3315, na rua bezerra de menezes com leon deniz, cabo da bt 1/0
2014	82734	1	1	0	0	0	0	[f363031 - 31/01/2014 05:55:09] - tr aberto com elos trocados circuito desgrampeado e tirado da bt situação encontrada. - 3x75 suporte
2014	16185	0	1	0	0	0	0	atuada fase c do circ 11011 - pne - elo 2h - normalizado
2014	61357	0	1	0	0	0	0	==realizar poda em vegetal de grande porte caído sobre a rede de bt cabo 1/0 e mt cabo 2ca. necessário motosserra. recompor um vao a 4 condutores de bt cabo 1/0ca e um vao a 3 condutores de mt cabo 2ca. repor ramais de ligação 30m de ramal monofásico 10mm
2014	13328	1	0	0	0	0	0	ocorrência gerada pela alteração da ocorrência 2014-12950 passado para ari da subterrânea - cabo subterrâneo da fase c que alimenta o cc. 45813 queimado (esta com fita amarela), o mesmo ficou desconectado para instalar cabo reserva (não tem cor) e necessário equipamento de solda para soldar o mesmo no barramento do cc. a chave do cc. 45813 ficou aberta na rua butui número 105 síndico ciente serviço realizado, finalizado por marcelo subterrânea.
2014	21540	1	0	0	0	0	0	
2014	26331	1	0	0	0	0	0	
2014	39981	1	0	0	0	0	0	
2014	41816	1	0	0	0	0	0	- - regulado tap do tr ...tensão anterior ...saída tr 205entre fase,,tensão após regulado –tirado do tap 2**pro 3**ok - regulado tap do tr ...tensão anterior ...saída tr 205entre fase,,tensão após regulado –tirado do tap 2**pro 3**ok
2014	43371	1	0	0	0	0	0	ocorrência encerrada pelo pda ficou trancada na tela foi corrigida e encerrada pelo código erro no sistema já relatado a Indra
2014	48842	1	0	0	0	0	0	arrebentado 1 vao a 1 cond fase a. cabo 2ca poste do tr esta podre, já com pe de amigo podre m11, n1, c2, mt e bt cabo 2ca, ip, crt gps: 6084 rua Portugal, 56

Tabela 15 – Exemplos de observações das ocorrências encerradas (conclusão)

Ano	Ocor.	304	Of	BN1	BN2	BN3	AS	Observação
2014	49725	1	0	0	0	0	0	na av. dr. jose athanasio 149 susbtituir tr 112,5 cc 160 poste de concreto com suporte. tr ficou ligado descumprindo ordem do cod em deixa-lo desligado.
2014	63870	1	0	0	0	0	0	galho de coqueiro caiu sobre a rede, realizada poda fase a elo 3h
2014	77772	1	0	0	0	0	0	numero nota 10172011 - ocorrencia: 73728 baixar tap para elevar tensao no tr5280 est retiro ponta grossa 2450 numero nota 10172011 - ocorrencia: 73728 baixar tap para elevar tensao no tr5280 est retiro ponta grossa 2450 tr 5280, fechado as 10:50..lficou no tap 4.l....segue tensoes= a=128...b=127...l.c=128....lab}227,{bc}228....{ac} 22
2014	85936	1	0	0	0	0	0	
2014	75866	0	0	0	1	0	0	substituir poste m11 quebrado, com tr pendido seguro pela rede. risco iminente de queda. gps 15643 m11, n2, c2, tr 3x75kva , 1 afastador, 4as11, mt4ca, bt 1/0ca. na rua ariovaldo s jansen, 237, estalagem. levar jg de para raios. foi aberta a ch 909 por seguranca. nao aceita tala.

ANEXO A – CÓDIGOS DE CAUSAS

As possíveis causas referentes às ocorrências encerradas são dadas na Tabela 16.

Tabela 16 – Códigos de causas para ocorrências encerradas

	Cod.	Descrição		Cod.	Descrição	
MEIO DE TERCEIROS	401	pandorga	CLIENTE	551	defeito interno	
	402	vandalismo		552	sobrecarga (interna)	
	403	colisao de veiculo		RELG	religie de urgencia/normal	
	404	emp. serv. publico ou cont.	MEIO AMBIENTE	301	poluicao	
	405	jogo de bola		302	maresia	
	406	predio em demolicao		303	descarga eletrica	
	407	ligacoes a revelia		304	vegetal	
	408	objeto sobre a rede		305	animais	
	409	furto de rede		306	erosao	
	490	falta de seguranca publica		307	incendio	
499	outros	308		inundacao		
FALHA COMPONENTES	601	isolador		309	calamidade publica	
	602	para-raios		399	outros	
	603	transformador	PROGRAMADOS	99	subtransmissao	
	604	condutor		100	externo CEEE-D	
	605	capacitor		101	alteracao p/ melhoria	
	606	religador		102	alteracao p/ ampliacao	
	607	regulador		201	manutencao preventiva	
	608	seccionalizador		202	manutencao corretiva	
	609	chaves	HUMANA	701	operacao	
	610	conexao		702	construcao	
	611	poste pendido		703	manutencao	
	612	poste podre		704	de ligacao	
	613	poste quebrado	OUTROS	802	externo CEEE-D	
	614	poste queimado		803	desarme SE / LT subtransmissão	
	615	cruzeta queimada		902	desprendimento / desconhecida	
	616	cruzeta podre		903	adequacao de rede	
	617	mao francesa		904	serviços impropriedades	
	618	medicao		905	serviços nao emergenciais	
	619	elo fusivel queimado		SISTEMA	501	sobrecarga
	620	ramal de ligacao			502	manobra localizacao falha
621	TC	503	manobra por seguranca			
622	TP	504	manobra transf. de carga			
623	disjuntor	505	manobra previa			
624	passagem de at	506	ordem de manobra			
625	cruzeta quebrada	ERAC	alívio de carga solicitado pelo ons			
626	rede desregrada	PICK	pick			
627	vao muito longo					
699	outros					