

## O PROCESSO *KDD* APLICADO À CRIAÇÃO DE ETIQUETA DE QUALIDADE DE ENERGIA ELÉTRICA RELACIONADA AOS AFUNDAMENTOS DE TENSÃO

### Anderson Roges Teixeira Góes

UFPR: Programa de Pós-Graduação em Métodos Numéricos em Engenharia (PPGMNE); e  
Departamento de Expressão Gráfica; CP: 19081; CEP: 81531-990; Curitiba, PR  
[artgoes@ufpr.br](mailto:artgoes@ufpr.br)

### Maria Teresinha Arns Steiner

UFPR: Programas de Pós-Graduação em Métodos Numéricos em Engenharia (PPGMNE) e de  
Engenharia de Produção (PPGEP); e  
PUCPR: Programa de Pós-Graduação em Engenharia de Produção e Sistemas (PPGEPS);  
CP: 19081; CEP: 81531-990; Curitiba, PR  
[tere@ufpr.br](mailto:tere@ufpr.br); [maria.steiner@pucpr.br](mailto:maria.steiner@pucpr.br)

### Rodrigo Antônio Peniche

COPEL – Companhia Paranaense de Energia  
[rodrigo.peniche@copel.com](mailto:rodrigo.peniche@copel.com)

### RESUMO

A busca pela qualidade em serviços prestados é muito discutida na atualidade, mas os trabalhos apresentados na literatura parecem não indicar técnicas que explorem bancos de dados com a finalidade de obter sua classificação. Diante deste contexto, neste artigo é proposta uma metodologia para a classificação da qualidade, de forma comparativa, utilizando o processo *KDD*. Como estudo de caso, a metodologia é aplicada à Qualidade de Energia Elétrica verificando a qualidade dos alimentadores de uma determinada subestação, levando em consideração a magnitude do afundamento, sua duração e sua frequência (quantidade de ocorrências no decorrer de certo período). Na busca da classificação dos alimentadores é utilizada a técnica de Redes Neurais Artificiais, com o algoritmo de aprendizado *back-propagation* e topologia com três neurônios para a camada de entrada, camada oculta (com número de neurônios variando de “1” a “20”) e um neurônio na camada de saída. Os resultados se apresentaram bastante promissores.

**PALAVRAS-CHAVES:** Processo *KDD*; Redes Neurais Artificiais; Qualidade da Energia Elétrica.

### ABSTRACT

The search for quality in services is much discussed today, but the papers in the literature seems don't indicate techniques which explore data bases in order to discover its classification. In this context, in this paper it is proposed a methodology for the quality classification, in a comparative way, using the *KDD* process. As a case study, the methodology is applied to Electrical Energy Quality verifying the feeders' quality of a substation, taking into account the sag magnitude, its duration and its frequency (quantity of occurrences during a certain period). Looking for the feeders' classification it is used the Artificial Neural Networks, with the back-propagation learning algorithm and topology with three neurons in the input layer, hidden layer (with number of neurons varying from “1” to “20”) and one neuron in the output layer. The results were much promised.

**KEY WORDS:** *KDD* Process; Artificial Neural Networks; Electrical Energy Quality.

## 1. Introdução

A qualidade nos serviços prestados pelas diversas empresas existentes na atualidade é crucial para um relacionamento satisfatório com seus consumidores. No setor energético, particularmente, tem ocorrido o crescimento no uso de equipamentos computadorizados de alta sensibilidade visando o atendimento de excelência aos seus clientes, através do fornecimento de Qualidade de Energia Elétrica (QEE).

Diante deste contexto, neste artigo é proposta uma metodologia para a classificação da QEE dos alimentadores de uma determinada subestação, de forma comparativa, levando em consideração a magnitude do afundamento, sua duração e sua frequência (quantidade de ocorrências no decorrer de certo período). Desta forma, com a obtenção desta classificação tem-se, como consequência, a criação de uma etiqueta de QEE.

A justificativa para a elaboração de tal etiqueta está no fato de ser uma alternativa a exigência da Agência Nacional de Energia Elétrica (ANEEL) que não define padrões de desempenho em relação aos afundamentos de tensão, mas solicita que “as distribuidoras devem acompanhar e disponibilizar, em bases anuais, o desempenho das barras de distribuição monitoradas” (ANEEL, 2008), informações que podem servir como referência para as unidades consumidoras atendidas pelo Sistema de Distribuição de Alta Tensão ou pelo Sistema de Distribuição de Média Tensão com cargas sensíveis a variações de tensão de curta duração.

Apesar da ANNEEL (2008) sugerir o monitoramento do desempenho das barras de distribuição, optamos por classificar alimentadores por dois motivos: i) a base de dados fornecidos pela concessionária de energia elétrica não dispunha de dados suficientes para aplicação em subestações; ii) temos a pretensão de utilização de dados reais e não fictícios/simulados

Nesta metodologia é utilizado o processo *KDD* (*Knowledge Discovery in Databases* ou Descoberta de Conhecimento em Bases de Dados) que, por sua vez, é composto de cinco fases, aplicadas ao problema aqui analisado.

Este artigo está organizado em cinco seções, incluindo esta introdução. Na segunda seção é escrito o problema. A revisão de literatura, indicando trabalhos correlatos e a inspiração para a criação desta metodologia é apresentada na terceira seção. A quarta seção apresenta o estudo de caso, sua aplicação e seus resultados. Por fim, a quinta seção realiza as conclusões e indica sugestões de continuidade e aplicação da metodologia.

## 2. Descrição do Problema

A COPEL (Companhia Paranaense de Energia Elétrica do Paraná), que supre as 399 cidades e 216 localidades situadas no estado do Paraná, Brasil, possuía, por ocasião da pesquisa, 378 subestações (SE) para atender cerca de 2.800.000 consumidores (residências, indústrias e outros). Em sua capital, Curitiba e região metropolitana, especificamente, existiam 30 subestações, com cerca de 300 alimentadores (aproximadamente 10 alimentadores por SE), atendendo cerca de 170.000 consumidores. Em 14 destas 378 subestações está instalado um aparelho, denominado *Power Quality Meter* (*PQM*; Medidor da Qualidade de Energia), que realiza, dentre outras medições, as vinculadas aos afundamentos e elevações de tensão.

Dos 14 *PQM*, seis deles estão instalados em seis das 30 SE da capital e região metropolitana: na SE da Cidade Industrial de Curitiba (CIC); Fazenda Rio Grande (também chamada de Fazenda Iguazu); Araucária; Piraquara; Tatuquara e Guatupê. No presente trabalho, a metodologia aqui desenvolvida é ilustrada fazendo-se a sua aplicação a apenas uma SE, a da CIC, que possuía 12 alimentadores, salientando-se de que poderá ser aplicada a qualquer SE, desde que possua o *PQM*.

Os eventos que ocorrem no sistema elétrico, devido aos mais diversos motivos, influenciam diretamente na QEE. Estes eventos podem ser de natureza accidental (queda de galhos de árvores; descargas atmosféricas) ou, então, programadas (manutenção preventiva), gerando assim, fenômenos de ordem comumente associados à QEE que são registrados pelo *PQM*.

Os registros (ou eventos) necessários ao desenvolvimento da proposta deste trabalho são armazenados em dois bancos de dados pela concessionária. Os registros do primeiro banco de dados, aqui denominado de BD01, são os captados pelo *PQM*. Cada um destes registros contém 17 informações (atributos), dentre os quais, a “identificação da oscilografia”; “data e hora de início do evento”; “data e hora de término do evento”; “tipo do evento” que pode ser, dentre outros, *Sag* (afundamento de tensão) ou *swell* (elevação de tensão); “Circuito”; “Status do evento”; “tensão remanescente ou *root mean square (RMS)*”, ou seja, a tensão que ainda “resta” com a ocorrência do evento; distorção harmônica total ou *total harmonic distortion (THD)*”.

Os registros do segundo banco de dados, aqui denominado de BD02, são captados pelo Sistema de Operações de Distribuição (SOD) da concessionária, e estão relacionados às interrupções. Estes registros, obtidos através de um *software* da própria concessionária, fornecem 29 atributos, dentre os quais se tem “identificação do alimentador”; “data e hora de início do evento”; “causa da interrupção”; “condição climática”; “quantidade de consumidores interrompidos”; “duração” e “componente afetado”.

Os dados utilizados no estudo são provenientes de um período de quatro meses, de fevereiro a maio de 2008, sendo que o BD01 ficou constituído de 352 registros e o BD02, de 422 registros. Faz-se necessário, então, a utilização de algum procedimento que analise e explore estas informações, transformando-as em conhecimento e, para isto, foi aqui utilizado o processo *KDD* visando a exploração dos dados obtendo-se, como consequência, a etiqueta da QEE.

Vale salientar que a metodologia aqui desenvolvida pode ser aplicada a qualquer subestação, assim como aos mais diversos problemas das mais variadas áreas.

### 3. Revisão da Literatura

A área elétrica concentra inúmeros trabalhos que tratam da QEE onde os autores utilizam de técnicas de Pesquisa Operacional, principalmente RNA, na busca da identificação, localização, classificação e/ou previsão de distúrbios: afundamentos de tensão, sobretensão, *Total Harmonic Distortion (THD)*, frequência, desequilíbrio do circuito, entre outros.

Mas estas pesquisas não tratam diretamente de classificar a QEE e sim, como já dito, os distúrbios que afetam tal qualidade (Trindade, 2005; Oleskovicz *et al.*, 2006; Silva *et al.*, 2007; Adepoju, Ogunjuyigbe e Alawode, 2007; e Caciotta, Giarnetti e Leccese, 2009).

No entanto, na literatura pesquisada foi possível encontrar dois artigos que tratam da classificação da QEE e propõe a criação de uma etiqueta. Um destes trabalhos é de autoria de Casteren *et al.* (2005) e o outro é de Cobben e Casteren (2006). E foi com a leitura destes, mais especificamente no trabalho de Casteren *et al.* (2005), que surgiu a inspiração para a proposta desta metodologia.

O trabalho desenvolvido por Cobben e Casteren (2006) apresenta métodos para a classificação da QEE, onde descrevem os níveis de qualidade para a tensão de energia, principais responsáveis pelas queixas de clientes e custos associados, com base em: pequenas variações de tensão, oscilação de tensão e quedas de tensão. Estes três métodos de classificação combinam a transparência com a simplicidade, uma vez que utilizam um sistema de classificação por meio de EQ proveniente de Casteren *et al.* (2005).

Já o trabalho Casteren *et al.* (2005) busca classificar os afundamentos de tensão de tal forma a indicar a responsabilidade pela causa do evento e mitigações (consumidor, fabricante do equipamento ou concessionária), examinando a duração e o valor remanescente de tais afundamentos.

De posse destes dados, os autores criam uma Etiqueta de Qualidade (EQ), de acordo com a frequência (número de ocorrências) ocorrida em cada uma das responsabilidades. Para tal classificam os afundamentos em uma tabela dividida em nove faixas, agrupados em três regiões, onde cada região representa uma área de responsabilidade.

### 4. Metodologia e Obtenção dos Resultados

Como já comentado, fez-se uso do processo *KDD* para compor a metodologia para a criação da etiqueta de QEE que ficou composta, basicamente, pelas seguintes etapas: pré-

processamento dos dados (limpeza e transformação dos dados); associação dos dados dos bancos de dados (BD01 e BD02) e, finalmente, a criação da etiqueta propriamente dita. Nesta última etapa fez-se uso de uma técnica de *Data Mining* (ou Mineração de Dados; aplicação de técnicas de Reconhecimento de Padrões para classificação), as Redes Neurais Artificiais.

#### 4.1. Pré-processamento dos Dados

No BD01 foi realizada a limpeza ou pré-processamento das informações obtidas. Nesta etapa, vários dos atributos constantes em BD01, foram retirados pelo fato de serem desnecessários para o objetivo aqui proposto que visa à criação da etiqueta de QEE de acordo com os afundamentos de tensão. Assim, o número de atributos foi reduzido de 17 para 9. Com relação ao pré-processamento do BD02, o número de atributos foi reduzido de 29 para apenas 6 pelo mesmo motivo.

Em reunião com engenheiros elétricos da concessionária, estes informaram que deveriam ser considerados apenas os registros em que o atributo “Tipo” fosse “Acidental”. Desta forma, houve nova filtragem no BD02 e permaneceram apenas cinco atributos a serem considerados. Como conseqüência desta nova “limpeza”, tem-se que o número de registros em BD02 foi alterado de 422 para 181.

Assim, ao final desta etapa tem-se que dos 352 registros, cada um deles com 17 atributos originais do BD01, apenas nove serão utilizados nas próximas fases. Para o BD02, dos 422 registros, apenas 181 serão considerados, sendo que dos 29 atributos originais, apenas cinco são considerados.

Ao serem analisados os registros no BD01, pôde-se verificar que para o atributo “Identificação da Oscilografia” são enumerados vários registros com mesmo valor em ordem cronológica. Esses registros se referem a uma determinada configuração do *PQM* para detectar o início de eventos relacionados à QEE, bem como aos eventos que estão ocorrendo naquele momento.

A partir desta análise, para cada grupo de registros do tipo “Sag”, com mesma “Identificação da Oscilografia”, é transformado em um único registro, diminuindo não apenas a quantidade de registros do BD01, mas também a quantidade de atributos, visto que “Tipo do evento” e “Status do Evento” não são mais necessários. Quanto ao atributo “Data e hora final” é considerado o registro que possui o horário “mais distante” do registrado em “Data do início”, gerando assim, um novo atributo denominado “Duração”, em milissegundos.

O atributo “Data e hora do início”, contendo data e hora do registro, foi dividido em outros dois, denominados “Data Início” e “Hora Início”. O mesmo ocorreu com “Data e hora final”, que foi dividido em “Data Final” e “Hora Final”.

Outra transformação nos dados do BD01 se refere aos atributos que indicam a tensão remanescente em cada uma das fases. A metodologia aplicada não leva em consideração em qual fase ocorreu o afundamento de tensão, mas sim se tal evento ocorreu ou não; caso positivo, dentre os valores registrados, o de menor magnitude registrada (ou seja, de pior valor) é o registrado. Desta forma, para cada grupo de registros com a mesma “Identificação de Oscilografia” foi verificada a menor tensão remanescente no grupo, independentemente da fase. Assim, esse valor foi registrado no novo atributo “Tensão remanescente”, e os atributos “Tensão RMS fase A”, “Tensão RMS fase B” e “Tensão RMS fase C” foram excluídos do BD01. A título de ilustração, alguns registros são apresentados no quadro 01 a seguir.

**Quadro 01** - Alguns dos registros do BD01 após a transformação dos dados

<b>Id. Osc.</b>	<b>Data Início</b>	<b>Hora Início</b>	<b>Data Final</b>	<b>Hora Final</b>	<b>Duração</b>	<b>Circuito</b>	<b>RMS</b>
9	2008-02-06	07:28:35.034	2008-02-06	07:28:35.252	218	0	60,1
10	2008-02-06	20:04:14.805	2008-02-06	20:04:14.990	185	1	35,9
...	...	...	...	...	...	...	...

Como já mencionado, a metodologia para a criação da etiqueta de QEE de um alimentador leva em consideração três atributos: tensão remanescente, duração e quantidade de ocorrências. Os dois primeiros atributos estão no BD01 (quadro 02), o terceiro é o resultado de uma simples contagem de ocorrências. No entanto, o BD01 não indica o alimentador que foi afetado pelo evento, pois os dados referentes aos alimentadores constam do BD02. Assim sendo, faz-se necessário associar os registros do BD01 com os registros do BD02, apresentada na próxima seção.

#### 4.2 Associação dos Dados (BD01 e BD02)

Para a associação dos dados contidos nos bancos de dados BD01 e BD02 foram utilizados os atributos vinculados a tempo. Mais especificamente foram utilizados os atributos “Data do Início” e “Hora do Início” do BD01 e “Data de Início” e “Hora de Início” do BD02.

Os engenheiros da concessionária definiram os intervalos de associação destes bancos de dados da seguinte forma: se no BD02 um registro possui para o atributo “Componente Afetado” a informação “RA” (Religamento Automático da chave) deve-se levar em consideração um intervalo de tempo de no máximo 5 minutos em relação a um registro do BD01, já que quando se tem uma chave do tipo “RA”, os bancos de dados registram o evento simultaneamente. Caso a informação não seja “RA”, mas sim outros tipos de chaves, o intervalo a ser considerado é de até 2h. Esta associação gerou um novo banco de dados, denominado BD03 contendo 169 registros. Ou seja, dos 352 registros de BD01 e dos 181 registros de BD02, existem 169 registros associados conforme o critério acima definido.

O quadro 02, composto de 10 colunas, apresenta alguns exemplos/registros desta associação. As informações das colunas de 1 a 5 são dados provenientes do BD01 e as das colunas 6 a 10, são as suas respectivas associações contidas no BD02. Além disto, como forma de identificar os 12 alimentadores desta subestação, estes serão denominados de AA, AB, AC, ..., AK, e AL.

**Quadro 02 - Exemplos dos registros BD03 (associação de BD01 com BD02)**

Id. Osc.	Data do Início	Hora do Início	Duração	RMS	Alimentador	Componente Afetado	Data Início	Hora Início	Duração
117	28/04/2008	14:28:43	185	46,3	AC	Cruzam. aéreo	28/04/2008	14:30	135
117	28/04/2008	14:28:43	185	46,3	AF	Atuação do RA	28/04/2008	14:29	1
117	28/04/2008	14:28:43	185	46,3	AF	Atuação elo fusível	28/04/2008	14:42	41
121	28/04/2008	18:07:50	202	42,6	AC	Condutor - AT	28/04/2008	18:32	344
121	28/04/2008	18:07:50	202	42,6	AI	Condutor - BT	28/04/2008	18:16	403
136	02/05/2008	7:13:10	705	28,7	AF	Poste	02/05/2008	07:15	36
139	08/05/2008	11:13:04	168	42,4	AI	Atuação do RA	08/05/2008	11:14	0

Neste quadro observa-se que um mesmo registro do BD01, pode ter mais de uma associação com o BD02, como no caso das três primeiras linhas deste quadro, onde o “Identificação de Oscilografia” é 117. Isto indica que o evento “captado” na subestação, também foi “captado”, ou se originou, em dois alimentadores “AC” e “AF”, sendo dois registros para o “AF” com componentes afetados diferentes: “Cruzamento aéreo”, “Atuação do RA”, ou simplesmente “RA” e “Atuação elo fusível”.

#### 4.3 Criação da etiqueta de QEE dos alimentadores de uma subestação, de forma comparativa

Desta forma, foi iniciada a classificação de cada registro do BD03 (quadro 03) construindo um quadro classificatório adaptado em relação ao proposto por Casteren *et al.* (2005), tendo como atributos: a tensão remanescente, a duração e a quantidade de ocorrências. A divisão proposta ao quadro ficou da seguinte forma, baseado em Casteren *et al.* (2005):

- foram consideradas duas faixas para a duração do evento:  $\leq 500$  e  $> 500$  milissegundos; e
- cinco intervalos de tensão remanescente: 10% a 19%, 20% a 39%, 40% a 59%, 60% a 79% e 80% a 90%.

O vínculo entre a duração e a tensão remanescente pode ser melhor compreendida observando-se o quadro 03, onde são apresentadas as 10 possíveis classes, aqui denominadas de C1, C2,... a C10. É evidente que quanto menor for a duração e maior for a tensão remanescente do evento, menos pior será a QEE daquele evento. Tem-se, assim, que a QEE dos eventos fica hierarquizada da seguinte forma:  $C1 \geq C2 \geq \dots \geq C10$ .

Para exemplificar a referida classificação, os registros hachurados do quadro 02 estão devidamente classificados, conforme o quadro 03 no quadro 04 a seguir.

**Quadro 03** - Classificação considerando a duração e a tensão remanescente dos registros

RMS (%)	Duração	
	$\leq 500$ milissegundos	$> 500$ milissegundos
80 a 90%	C1	C2
60 a 79%	C3	C4
40 a 59%	C5	C6
20 a 39%	C7	C8
10 a 19%	C9	C10

**Quadro 04** - Classificação dos registros do quadro 02 conforme quadro 03

Duração (milissegundos)	RMS (%)	Alimentador	Classificação do registro
185	46,3	AC	C5
185	46,3	AF	C5
185	46,3	AF	C5
202	42,6	AC	C5
202	42,6	AI	C5
705	28,7	AF	C8
168	42,4	AI	C5

Ao realizar esta classificação com os 169 registros do BD03 têm-se que as quantidades de registros dos alimentadores AA, por exemplo, é apresentada no quadro 05. Da mesma forma, tem-se a classificação dos registros para os demais alimentadores da subestação da CIC. O quadro 05 mostra que o alimentador "AA" possui dois eventos do tipo C5; um do tipo C7 e um do tipo C8.

**Quadro 05** - Classificação dos afundamentos de tensão do alimentador "AA"

RMS (%)	Duração	
	$\leq 500$ milissegundos	$> 500$ milissegundos
80 a 90%	0	0
60 a 79%	0	0
40 a 59%	2	0
20 a 39%	1	1
10 a 19%	0	0

Considerando todos os 169 registros de todos os 12 alimentadores da subestação da CIC, tem-se o quadro 06, onde se pode observar que apenas três destas faixas possuem registros: C5, C7 e C8.

**Quadro 06 - Classificação dos afundamentos de tensão da subestação analisada considerando todos os registros**

RMS (%)	Duração	
	≤ 500 milissegundos	> 500 milissegundos
80 a 90%	0	0
60 a 79%	0	0
40 a 59%	149	0
20 a 39%	15	5
10 a 19%	0	0

Para se obter a “qualidade média” da subestação em análise, pode-se dividir a quantidade de eventos do quadro 06 por 12 (total de alimentadores), obtendo-se os dados do quadro 07, já devidamente arredondados.

**Quadro 07 - Classificação média dos afundamentos de tensão na subestação analisada**

RMS (%)	Duração	
	≤ 500 milissegundos	> 500 milissegundos
80 a 90%	0	0
60 a 79%	0	0
40 a 59%	13	0
20 a 39%	2	1
10 a 19%	0	0

Assim, para criar a etiqueta de QEE foi estabelecido o valor de seis faixas, sendo a “Faixa A” a de melhor QEE e a “Faixa F” a de pior QEE. Para cada faixa foram multiplicados fatores, aleatórios, determinando assim os valores do limite superior de cada faixa. Desta forma, o quadro 07 representa a média dos alimentadores, ou seja, o limite superior da “Faixa C”.

O limite superior da “Faixa A” (quadro 08) pode ser obtido multiplicando os valores do quadro 07 por 0,25. O limite superior da “Faixa B” pode ser obtido multiplicando os valores do quadro 07 pelo fator 0,50. Ao multiplicar os valores do quadro 07 por 1,5 obtém o limite superior da “Faixa D”. O limite superior da “Faixa E” é obtido multiplicando pelo fator 2 os valores do quadro 07. Por fim, tem-se o limite superior da “Faixa F” é obtido verificando o maior valor apresentado para os alimentadores analisados.

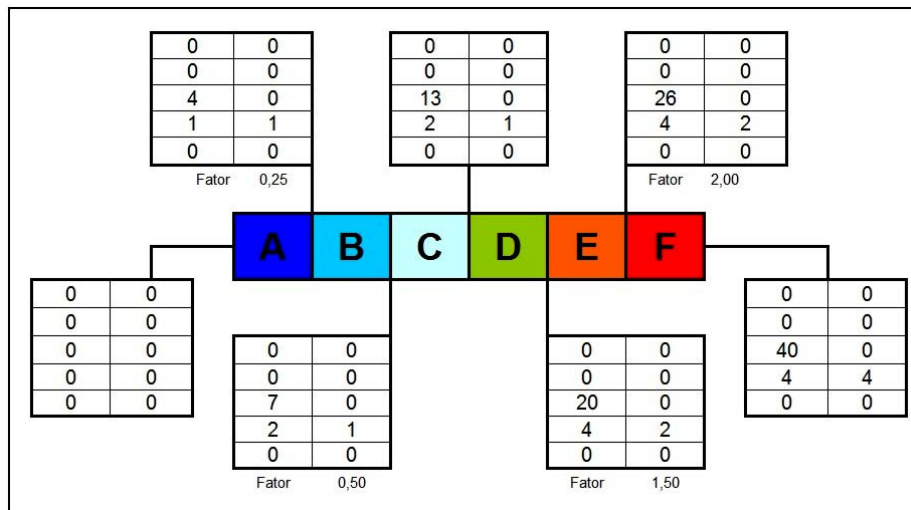
**Quadro 08 - Limite superior da “Faixa A” da etiqueta de classificação da QEE de um alimentador em relação à subestação**

RMS (%)	Duração	
	≤ 500 milissegundos	> 500 milissegundos
80 a 90%	0	0
60 a 79%	0	0
40 a 59%	4	0
20 a 39%	1	1
10 a 19%	0	0

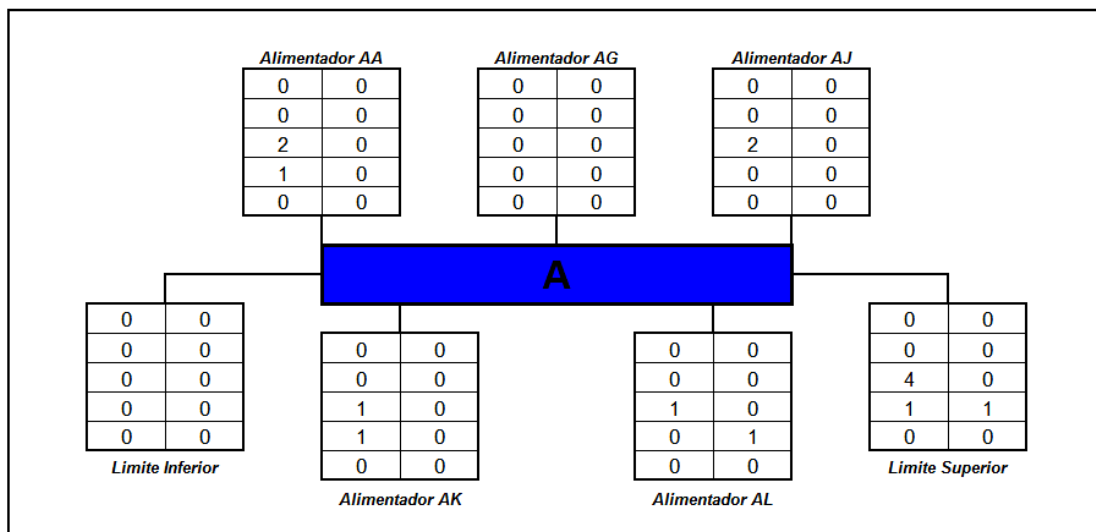
A figura 01, a seguir, apresenta a etiqueta de classificação da QEE dos alimentadores da subestação da CIC, de forma comparativa, com os limites de cada faixa definidos anteriormente.

Criada a etiqueta basta verificar em qual faixa cada quadro que representa o alimentador se enquadra. No entanto, esta tarefa não é tão simples, pois dos 12 alimentadores, apenas cinco se enquadram nestas faixas de valores, todos com classificação de qualidade “A”: AA, AG, AJ, AK e AL, conforme mostrado na figura 02, a seguir. Tem-se que os cinco alimentadores apresentam valores para C5 pertencentes ao intervalo discreto [0, 4] e para C7 e C8, no intervalo discreto [0,1].

Os demais alimentadores não podem ser classificados diretamente, uma vez que, por exemplo, para o alimentador AH, C5 é igual a 16, o que indica que sua classificação seria D. No entanto, C7 e C8 para este alimentador estão fora dos intervalos destas classes, já que C7 possui valor “1” e, por outro lado, C8 possui valor “0”. Assim, para realizar a classificação dos demais alimentadores, pode-se fazer uso de técnicas de *Data Mining*. Aqui neste trabalho, far-se-á uso das RNA capaz de realizar esta tarefa, apresentada na seção 4.4 a seguir.



**Figura 01** - Etiqueta de classificação da QEE dos alimentadores, de forma comparativa da subestação da CIC



**Figura 02** - Alimentadores classificados diretamente na etiqueta de QEE

#### 4.4. Técnica de *Data Mining*

A etapa de *DM* é a mais importante do processo *KDD*, uma vez que é neste momento que se aplicam as técnicas para o Reconhecimento de Padrões, seja através de procedimentos heurísticos ou metaheurísticos. Neste trabalho, como já comentado, tais procedimentos são aplicados visando à classificação da QEE de alimentadores de uma subestação.

A técnica aqui utilizada para classificar os alimentadores que não puderam ser classificados diretamente, como mostrado na seção 4.3 anterior são as RNAs. O algoritmo de



aprendizado utilizado foi o *back-propagation* (ou retro-propagação do erro) e foi implementado em Visual Basic 6.0.

As RNA treinadas possuíam três entradas ( $C_5$ ,  $C_7$  e  $C_8$ ) para a camada de entrada, camada oculta (com número de neurônios variando de “1” a “20”) e um neurônio na camada de saída. Os pesos iniciais, para cada simulação, foram definidos de forma aleatória, no intervalo de (-1, 1).

A forma de avaliação das RNAs, aqui utilizada, foi o *holdout* estratificado, ou seja, o conjunto de dados para o treinamento foi dividido em dois subconjuntos: 2/3 para o conjunto de treinamento e 1/3 para o conjunto de teste. Assim, como são 6 faixas de classificação (de A a F, sendo 30 registros por faixa) tem-se 180 registros, dos quais 120 são utilizados para o treinamento e 60 para testes.

O treinamento das RNAs ocorreu por seis vezes, um treinamento para cada classe sendo que, neste caso, a rede terá apenas uma saída: o registro pertence a uma determinada classe (ou não pertence). Assim, realiza-se o treinamento para a classe A, fazendo com que a rede “aprenda” o que é um registro da classe A (valor próximo a “0”) e o que não é A (valor próximo a “1”, ou seja, B, C, D, E e F). Em seguida, realiza-se outro treinamento para a classe B, retirando-se os dados do conjunto A, fazendo com que a rede “aprenda” o que é um registro da classe B e o que não é B (C, D, E e F) e assim por diante para as classes: C, D e E. Ao realizar o último treinamento (Classe E), quando um registro não é classificado como E, ele é automaticamente F (última classe).

Assim, ao se desejar classificar um novo registro à rede treinada, este deverá ser “apresentado” a todas as seis redes, sendo que a sua classificação será de acordo com a rede que apresentar o valor de saída mais próximo de “0”.

Para o treinamento, o conjunto de dados foi dividido em três subconjuntos, denominados de  $SC_1$ ,  $SC_2$  e  $SC_3$ , sendo que cada um contém 10 registros de cada faixa da etiqueta de classificação (de A a F), totalizando 60 registros em cada subconjunto, ou seja, 180 dados.

Com os subconjuntos citados são realizadas as simulações, onde denominados cada etapa de testes de: 1º *holdout*, quando os subconjuntos  $SC_1$  e  $SC_2$  formam o conjunto de treinamento e  $SC_3$  o conjunto de teste; 2º *holdout*, quando os subconjuntos  $SC_1$  e  $SC_3$  formando o conjunto de treinamento e  $SC_2$  o conjunto de teste; e 3º *holdout*, quando os subconjuntos  $SC_2$  e  $SC_3$  formando o conjunto de treinamento e  $SC_1$  o conjunto de testes.

Para cada uma destas etapas, a rede foi treinada cinco vezes, variando-se o conjunto de pesos iniciais. Assim, tem-se um total de 1500 testes (3 etapas x 5 conjuntos de pesos iniciais x 20 quantidades de neurônios na camada escondida x 5 faixas de classificação). A função de ativação utilizada foi a sigmoideal-logística.

O treinamento da rede era finalizado ao se atingir uma das seguintes três condições: 1.000 iterações; erro médio quadrático menor ou igual a  $10^{-4}$ ; ou número de registros classificados incorretamente igual a zero.

O treinamento teve início onde a RNA teve que reconhecer se o padrão pertence (ou não) ao conjunto A. A tabela 01 mostra que para o 1º *holdout*, 1º conjunto de pesos iniciais, teve o melhor resultado obtido com 20 neurônios na camada escondida, 431 iterações e erro quadrático médio igual a  $6,41 \times 10^{-4}$ . O melhor resultado do treinamento no 1º *holdout* para a “Faixa A” foi a 3ª simulação, uma vez que o erro médio quadrático é o menor.

**Tabela 01** - Melhores resultados no treinamento do 1º *holdout* para a “Faixa A” da etiqueta de classificação utilizando as RNAs

Conjuntos de pesos iniciais	Quantidade de neurônios na camada escondida	Iterações	% de Acertos	Erro quadrático médio (aprox)
1	20	431	100%	$6,41 \times 10^{-4}$
2	18	561	100%	$5,17 \times 10^{-4}$
3	20	530	100%	$4,93 \times 10^{-4}$
4	17	387	100%	$6,80 \times 10^{-4}$

5	12	144	100%	$1,21 \times 10^{-03}$
---	----	-----	------	------------------------

Realizados os testes para a “Faixa B”, ainda no 1º *holdout*, o melhor resultado é determinado pela 5ª simulação (16 neurônios na camada escondida; 642 iterações; nenhum registro classificado errado e erro médio quadrático igual a  $3,68 \times 10^{-03}$ ). Da mesma forma, tem-se os resultados para as demais faixas: “Faixa C” (16 neurônios, 182 iterações e erro quadrático médio igual a  $1,75 \times 10^{-04}$ ); “Faixa D” (13 neurônios; 232 iterações e erro quadrático médio  $1,39 \times 10^{-02}$ ); Faixa “E” (12 neurônios; 376 iterações e erro médio quadrático igual a  $3,67 \times 10^{-04}$ ).

Assim, um resumo dos resultados do 1º *holdout* é apresentado na tabela 02, a seguir. Com estes treinamentos foram realizados os testes e nesta etapa a porcentagem de acerto foi total (100%). Cabe observar que quando um exemplo é apresentado à rede correspondente a “Faixa E”, se a rede não “conseguir” classificá-lo como sendo desta faixa, este será automaticamente classificado como “Faixa F” (já que esta é a última faixa). Da mesma forma foram realizados o 2º e o 3º *holdouts*.

**Tabela 02 - Melhores resultados do 1º *holdout* para a etiqueta de classificação - RNA**

Faixa	Sequência de simulação	TREINAMENTO			Erro quadrático médio (aprox)	TESTE % de acertos
		Quantidade de neurônios na camada escondida	Iterações	% de Acertos		
A	3	20	530	100%	$4,93 \times 10^{-04}$	100%
B	5	16	642	100%	$3,68 \times 10^{-03}$	100%
C	5	16	182	100%	$1,75 \times 10^{-04}$	100%
D	2	18	1000	99,16%	$1,39 \times 10^{-02}$	100%
E	1	12	376	100%	$3,67 \times 10^{-04}$	100%

Pode-se verificar que os resultados foram bastante satisfatórios, uma vez que em apenas duas de 15 classificações, as RNA não conseguiram aprender apenas um exemplo em cada destas duas classificações, gerando 98,33% de acerto para o conjunto de testes.

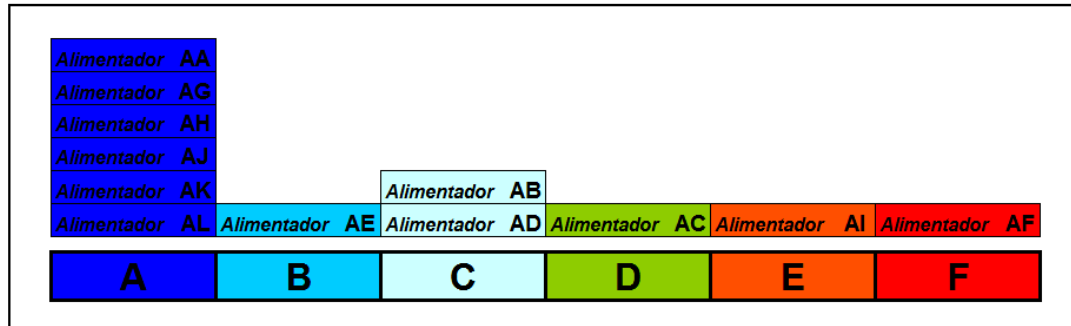
Em cada etapa do *holdout* foram apresentados às redes, além dos exemplos dos conjuntos de testes, também os valores referentes a cada alimentador, objetivo do treinamento. O resultado desta classificação está no quadro 09.

**Quadro 09 - Resultado da classificação dos alimentadores em cada etapa do holdout - RNA**

Alimentador	1º <i>holdout</i>	2º <i>holdout</i>	3º <i>holdout</i>	Classificação Média
AA	A	A	A	A
AB	C	C	C	C
AC	D	D	D	D
AD	C	C	B	C
AE	B	B	B	B
AF	F	F	F	F
AG	A	A	A	A
AH	A	A	A	A
AI	E	E	E	E
AJ	A	A	A	A
AK	A	A	A	A
AL	A	A	A	A

No quadro 09 anterior, a coluna “Classificação Média” (última coluna) indica a classificação de maior ocorrência nas colunas anteriores. Caso não haja uma classificação com maior ocorrência, está será definida como a pior situação.

Ainda referente ao quadro 09, apesar dos alimentadores AA, AG, AJ, AK e AL já terem as suas classificações, conforme descrito na seção 4.3, os mesmos também foram apresentados à rede e confirmaram as classificações já obtidas. Assim, têm-se seis alimentadores com classificação “A”, um alimentador com classificação “B”, dois alimentadores com classificação “C”, um alimentador com classificação “D”, um alimentador com classificação “E” e um alimentador com classificação “F”. (figura 03)



**Figura 03** - Etiqueta de qualidade com a classificação dos alimentadores

## 5. Conclusões

A busca pela qualidade em serviços prestados é muito discutida na atualidade. No entanto, os trabalhos apresentados na literatura não indicam técnicas que explorem bancos de dados com a finalidade de obter a classificação para esta qualidade.

Com a metodologia de criação de etiqueta de qualidade, de forma comparativa, no contexto *KDD* apresentada neste artigo, esta ausência na classificação de qualidade é solucionada.

Com os resultados obtidos na classificação dos alimentadores em relação aos afundamentos de tensão, a concessionária poderá analisar o comportamento destes, verificando os classificados como sendo de baixa qualidade e aplicar medidas de mitigação para que os mesmos tenham sua classificação “alavancada”. Para isto, tendo a etiqueta de qualidade gerada pode-se dar continuidade na busca da identificação do “por que” de tal classificação dos alimentados, ou seja, por que o alimentador AF tem qualidade “F”? Assim, sugerimos que para responder a esta pergunta, uma alternativa é aplicar novamente o processo *KDD*, mas agora com a finalidade de extrair regras dos registros deste alimentador, onde devem ser considerados outros atributos, como equipamentos, clima, manutenção e outros que devem ser discutidos com um especialista da área. Com isso, outros índices da qualidade, como DEC, FEC, DIC e FIC, serão influenciados por tal identificação.

Também como já comentado, vemos nesta metodologia, se aplicada a comparação de subestações, uma alternativa a exigência da ANEEL na solicitação de que “as distribuidoras devem acompanhar e disponibilizar, em bases anuais, o desempenho das barras de distribuição monitoradas”, que sirvam de referência para as unidades consumidoras atendidas pelo Sistema de Distribuição de Alta Tensão ou pelo Sistema de Distribuição de Média Tensão com cargas sensíveis a variações de tensão de curta duração.

Apesar do problema aqui apresentado estar relacionado à área elétrica, já temos aplicado tal metodologia na área educacional e estamos visualizando sua aplicação em outras áreas, como por exemplo, saúde. Com isto, fica comprovada a versatilidade de tal metodologia.

Quanto à técnica apresenta - RNA - esta se mostrou eficiente na busca de classificar os alimentadores que não tiveram classificação direta na etiqueta de qualidade e ainda classificou corretamente os alimentadores que já se sabia a classificação da qualidade.

Outras técnicas como, *Support Vector Machine*, Algoritmos Genéticos e heurísticas que consideram distância euclidiana e de mahalanobis, estão em testes para este caso, bem como, nas outras áreas citadas anteriormente. Este artigo apresentou apenas uma das técnicas que compõe

esta pesquisa, fruto de uma tese de doutorado do Programa de Pós-graduação em Métodos Numéricos em Engenharia da Universidade Federal do Paraná.

Os fatores multiplicadores (0,25; 0,50; 1,50 e 2) utilizados para a obtenção dos limites superior da faixas A, B, D e E da etiqueta de qualidade, foram determinados aleatoriamente. Com isso, sugerimos que em trabalhos futuros estes sejam explorados com a finalidade de defini-los.

Por fim, devido a versatilidade da metodologia esta tem muito ainda que ser explorada com diferentes técnicas e áreas.

## Referências

**Adepoju, G. A.; Ogunjuyigbe, S. O. A.; Alawode, K. O.** Application of Neural Network to Load Forecasting in Nigerian Electrical Power System. *The Pacific Journal of Science and Technology*. Spring. V. 8. 2007, p. 68-72.

**ANEEL Agência Nacional de Energia Elétrica.** Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST: Módulo 8 – Qualidade da Energia Elétrica. 2008.

**Caciotta, M.; Giarnetti, S.; Leccese, F.** Hybrid Neural Network System for Electric Load Forecasting of Telecommunication Station. *XIX IMEKO World Congress - Fundamental and Applied Metrology*. Lisboa, Portugal. 2009, p. 657-661

**Casteren, J. F. L. van.; Enslin, L. H. R.; Hulshorst, W. T. J.; Kilng, W.L.; Hamoen, M. D.; Cobben, J. F. G.** Acustomer oriented approach to the classification of voltage dips. In: *The 18th International Conference and exhibition on Electricity Distribuion - CIRED 2005*.

**Cobben, J. F. G.; Casteren, J. F. L.** Classification Methodologies for Power Quality. *Electrical Power Quality & Utilization Magazine*. V. 2. 2006.

**Oleskovicz, M.; Coury, D. V. ; Carneiro, A. A. F. M.; Arruda, E. F.; Delmont, O.; Souza, S. A.** Estudo comparativo de ferramentas modernas de análise aplicadas à qualidade da energia elétrica. *Revista Controle & Automação*. Vol. 17 no 3. Julho, agosto e setembro 2006.

**Silva, K. M.; Souza, B. A.; Brito, N. S. D.; Dantas, K. M. C.; Costa, F. B.; Silva, S. S. B.** Detecção e classificação de faltas a partir da análise de registros oscilográficos via redes neurais artificiais e Transformada Wavelet. *Revista Controle & Automação*. Vol. 18 no 2. Maio e junho 2007.

**Trindade, R. M.** Sistema Digital de Detecção e Classificação de Eventos de Qualidade de Energia. Juiz de Fora, 2005. *Dissertação (Mestrado em Engenharia Elétrica – Universidade Federal de Juiz de Fora)*.