



**ANDRÉ LUIZ MEDEIROS**

**ALOCAÇÃO DE EQUIPES DE CAMPO PARA  
AVALIAÇÃO DE PERDAS NÃO TÉCNICAS DE  
ENERGIA ELÉTRICA: DESENVOLVIMENTO DE  
UM SISTEMA DE APOIO À DECISÃO**

**2013**

**ANDRÉ LUIZ MEDEIROS**

**ALOCAÇÃO DE EQUIPES DE CAMPO PARA AVALIAÇÃO DE  
PERDAS NÃO TÉCNICAS DE ENERGIA ELÉTRICA:  
DESENVOLVIMENTO DE UM SISTEMA DE APOIO À DECISÃO**

Tese apresentada à Universidade Federal de Lavras,  
como parte das exigências do Programa de Pós-  
Graduação em Administração, área de concentração  
em Gestão de Negócios, Economia e Mercado, para  
obtenção do título de Doutor.

Orientador

Dr. Ricardo Pereira Reis – UFLA

Coorientador

Dr. José Arnaldo Barra Montevechi – UNIFEI

**LAVRAS – MG  
2013**

**Ficha Catalográfica Elaborada pela Divisão de Processos Técnicos da  
Biblioteca da UFLA**

Medeiros, André Luiz.

Alocação de equipes de campo para avaliação de perdas não técnicas de energia elétrica : desenvolvimento de um sistema de apoio à decisão / André Luiz Medeiros. – Lavras : UFLA, 2013.

186 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2013.

Orientador: Ricardo Pereira Reis.

Bibliografia.

1. Setor elétrico. 2. Perdas não técnicas. 3. Pesquisa-ação. 4. Inspeção do consumo fraudulento. 5. Sistemas de informação. I. Universidade Federal de Lavras. II. Título.

CDD – 658.403

**ANDRÉ LUIZ MEDEIROS**

**ALOCAÇÃO DE EQUIPES DE CAMPO PARA AVALIAÇÃO DE  
PERDAS NÃO TÉCNICAS DE ENERGIA ELÉTRICA:  
DESENVOLVIMENTO DE UM SISTEMA DE APOIO À DECISÃO**

Tese apresentada à Universidade Federal de Lavras,  
como parte das exigências do Programa de Pós-  
Graduação em Administração, área de concentração  
em Gestão de Negócios, Economia e Mercado, para  
obtenção do título de Doutor.

APROVADA em 28 de fevereiro de 2013

Dr. Luiz Marcelo Antoniali	UFLA
Dr. Antônio Carlos dos Santos	UFLA
Dr. Alexandre Ferreira de Pinho	UNIFEI
Dr. Fabio Favaretto	UNIFEI
Dr. José Arnaldo Barra Montevechi	UNIFEI

Dr. Ricardo Pereira Reis  
Orientador

**LAVRAS – MG  
2013**

*Aos meus avôs, Carlos e Delziro (in memorian), que serviram de exemplo e de  
inspiração.*

*As minhas avós, Maria e Elvira, que cada uma ao seu modo, me inspiram na  
caminhada da vida.*

*Ao Luiz e à Arlete Medeiros, meu pai e minha mãe, pelo sublime dom da vida,  
pelo exemplo de ser humano, pelo apoio incondicional e pelo amor  
incontestável.*

*Aos meus irmãos, Frederico e Marcelo, que mesmo separados pela distância  
física, estão próximos pelo amor.*

*Ao Gleiber e à Izabel, meus segundos pais, pelo acolhimento, dedicação e  
carinho.*

*A minha filha Amanda, que além de me mostrar uma nova forma de ver o  
mundo, dedica amor incondicional.*

*Ao meu filho que logo chegará (Zequinha). Mesmo que ainda não esteja  
presente entre nós, já me proporciona momentos mágicos.*

*Em especial, à Glenia Junqueira Machado Medeiros, minha esposa,  
companheira, confidente, cúmplice e guerreira. Sem o seu apoio, compreensão e  
amor tenho certeza que tudo teria sido muito mais difícil.*

**DEDICO**

## AGRADECIMENTOS

À Universidade Federal de Lavras (UFLA) e ao Departamento de Administração e Economia (DAE), pela oportunidade concedida para realização do doutorado.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig) pela bolsa de estudo concedida por meio do Programa Mineiro de Capacitação Docente (PMCD) e pelo apoio técnico e científico.

Aos professores do DAE da UFLA, pelos ensinamentos transmitidos e harmoniosa convivência.

Ao mestre e amigo Dr. Ricardo Pereira Reis pela confiança, paciência, orientação, sincera amizade e seus ensinamentos que foram de grande relevância para a realização deste trabalho e meu crescimento pessoal e profissional.

Ao amigo e coorientador Dr. José Arnaldo Barra Montevechi pela confiança, compreensão e pelos conhecimentos compartilhados ao longo do desenvolvimento da tese.

Aos colaboradores da distribuidora de energia elétrica que foi objeto de pesquisa deste trabalho. Sem a “cumplicidade” deles, não teria sido possível.

Aos professores Dr. Luiz Marcelo Antoniali, Dr. Antônio Carlos dos Santos, Dr. Alexandre Ferreira de Pinho e Dr. Fabio Favaretto, pelos conhecimentos repassados e pela disposição em participarem como membros da banca examinadora.

Aos amigos e colegas mestrandos e doutorandos Adriano, Luiz Fernando (*in memorian*), Diego, Fernanda, João, Miguel, Viviane, Rodrigo, Elizete e todos os demais que contribuíram com a preciosa ajuda no desenvolvimento das atividades exigidas pelas disciplinas.

E não menos importante, a Deus por ter me dado sabedoria para pensar e a força para agir, elementos essenciais para construir um trabalho belo.

*“... algo simples, depende de uma estrutura extremamente complexa. Por isso as coisas simples são extraordinárias...”*

## RESUMO

As perdas de energia elétrica, principalmente a não técnica ou comercial, têm sido assunto prioritário das empresas concessionárias de distribuição de energia elétrica e órgãos reguladores. Isso porque elas têm gerado prejuízos estimados em R\$7 bilhões ao ano no Brasil. Uma das soluções para esse problema seria auditar ou inspecionar as instalações com suspeita de irregularidade no consumo. Entretanto, propostas isoladas como essa podem ser técnica e economicamente inviável para as distribuidoras. Considerando as discussões sobre esse tema, o objetivo deste trabalho foi, em conjunto com uma distribuidora de energia elétrica, construir, reflexivamente, ações que auxiliassem a alocação de equipes de campo, para viabilizar, técnica e economicamente, a inspeção de consumidores suspeitos de fraudar o consumo de energia. Especificamente, levantou-se, interativamente com a distribuidora, dificuldades existentes na alocação das equipes de campo e, por meio de reflexões à luz da teoria e das experiências da distribuidora, desenvolveu-se um sistema de apoio à decisão para auxiliar na solução do problema. Considerando o objetivo proposto este trabalho teve como objeto de estudo uma das 64 distribuidoras de energia que atuam no Brasil, que foi selecionada pela sua representatividade nacional e disponibilidade em participar do estudo. Para suportar a construção reflexiva e interativa das ações propostas, adotou-se a pesquisa-ação como procedimento técnico de pesquisa. Como resultado, pode-se destacar que foram necessários três ciclos de interação de pesquisa-ação para que, em um quarto ciclo, fosse desenvolvido o *software* de apoio à decisão, denominado de Perdas DSS. Ao usar a pesquisa-ação como procedimento técnico, foi possível, pouco a pouco, interativa e reflexivamente, usar a experiência dos colaboradores da distribuidora para construir um conjunto de ações que resultaram no *software* Perdas DSS, que poderá auxiliar na alocação das equipes de campo, viabilizando técnica e economicamente, a inspeção de consumidores suspeitos de fraudar o consumo de energia. Os resultados obtidos na análise da distribuidora de energia elétrica poderão ser comparados com o de outras distribuidoras que vivem situações semelhantes. Promovendo, desse modo, o progresso científico, que surge por meio da construção de teorias que evoluem, paulatinamente, do particular para o geral.

Palavras-chave: Perdas não técnicas. Energia elétrica. Sistema de apoio à decisão. Pesquisa-ação.

## ABSTRACT

The loss of electricity mainly non-technical or commercial, have been a matter of priority the concession of electricity distribution and regulators. That's because they have generated losses estimated at R\$7 billion a year in Brazil. One solution to this problem would be to audit or inspect the premises with suspected irregularity in consumption. However, isolated proposals like this can be technically and economically feasible for the distributors. Considering the discussions on this issue, the objective of this work was, together with an electricity distributor, build, reflexively, actions that would aid the allocation of field teams, to enable, technically and economically, to inspect suspected consumers to defraud the energy consumption. Specifically, stood up interactively with the distributor, difficulties in the allocation of field teams and through reflections about the light of the theory and the distributor experiences developed a decision support system to assist in problem solving. Considering the objective proposed this work had as object of study one of the 64 energy distributors that operate in Brazil, which was selected for its national representativeness and availability to participate of the study. To support the construction interactive and reflexive actions proposed, adopted the action-research as technical procedure research. As a result, it may be noted that it took three cycles of action-research interaction for in a fourth cycle was developed decision support software, called Losses DSS. By using action-research as a technical procedure, it was possible, little by little, interactive and reflexively use the experience of the collaborators from distributor to build a set of actions that resulted in Losses DSS software that may assist in the allocation of field teams , enabling technically and economically, to inspect suspected consumers to defraud the energy consumption. The results obtained in the analysis of the electricity distributor may be compared with those other distributors which remain similar situations. Promoting thereby, scientific progress which stood up through the construction of theories that evolve gradually of the particular to general.

Keywords: Non-technical losses. Electricity. Decision support system. Action-research.

## LISTA DE QUADROS

Quadro 1	Formas de fraude no consumo de energia, encontrado nas distribuidoras de energia elétrica.....	25
Quadro 2	Dimensões dos sistemas de informação computacional e suas características .....	40
Quadro 3	Tipos de sistemas de informação computacional e suas características .....	42
Quadro 4	Ciclos de pesquisa-ação com suas respectivas descrições, divididos em subitens deste trabalho.....	60
Quadro 5	Característica e sigla das variáveis selecionadas para serem analisadas no Ciclo PA-3 .....	99
Quadro 6	Características dos Métodos de Busca associados ao algoritmo <i>CfsSubsetEval</i> , usados na análise dos dados da distribuidora em estudo.....	107
Quadro 7	Conjunto de variáveis selecionadas pelos métodos de busca associados ao algoritmo <i>CfsSubsetEval</i> , usados na análise dos dados da distribuidora em estudo .....	108
Quadro 8	Características dos Métodos de Busca associados ao algoritmo <i>Ranker</i> , usados na análise dos dados da distribuidora em estudo .....	108
Quadro 9	Métodos de busca e classificação individual das variáveis, de acordo com o algoritmo <i>Ranker</i> , usados na análise dos dados da distribuidora em estudo .....	109
Quadro 10	Classificação das variáveis (M ou E), de acordo com os métodos de busca associados ao algoritmo <i>Ranker</i> , usados na análise dos dados da distribuidora em estudo.....	110

## LISTA DE FIGURAS

Figura 1	Ciclo de vida utilizado para o desenvolvimento de um sistema de apoio à decisão .....	49
Figura 2	Formas de classificações de pesquisas científicas.....	52
Figura 3	Passos para implantação da pesquisa-ação, como procedimento técnico .....	56
Figura 4	Fluxograma do Procedimento Operacional Padrão (POP) usado na atividade de seleção e liberação de inspeção, na distribuidora em estudo .....	69
Figura 5	Fluxograma usado por um colaborador na seleção e liberação de inspeção, em uma unidade regional da distribuidora em estudo .....	72
Figura 6	Fluxograma usado por um colaborador para adicionar filtro de alvos, em uma unidade regional da distribuidora em estudo.....	73
Figura 7	Fluxograma usado por um colaborador para ordenar dados filtrados, em uma unidade regional da distribuidora em estudo....	73
Figura 8	Fluxograma usado por um colaborador para selecionar as instalações a serem inspecionadas, em uma unidade regional da distribuidora em estudo .....	74
Figura 9	Fluxograma usado por um colaborador na seleção e liberação de inspeção, em uma unidade regional da distribuidora em estudo .....	76
Figura 10	Fluxograma usado por um colaborador para adicionar filtros de alvo, em uma unidade regional da distribuidora em estudo .....	77
Figura 11	Fluxograma usado por um colaborador para selecionar as instalações a serem inspecionadas, em uma unidade regional da distribuidora em estudo .....	78

Figura 12	Quantidade total de consumidores por unidade regional da distribuidora em estudo, no período de 12 meses.....	83
Figura 13	Distribuição percentual das classes de consumo de energia, por unidade regional da distribuidora em estudo, no período de 12 meses .....	84
Figura 14	Quantidade total de motivos existentes, por unidade regional da distribuidora em estudo, no período de 12 meses.....	84
Figura 15	Principais tipos de motivo registrado na unidade regional DC da distribuidora em estudo, no período de 12 meses.....	86
Figura 16	Principais tipos de motivo registrado na unidade regional DL da distribuidora em estudo, no período de 12 meses.....	87
Figura 17	Principais tipos de motivo registrado na unidade regional DN da distribuidora em estudo, no período de 12 meses.....	87
Figura 18	Principais tipos de motivo registrado na unidade regional DS da distribuidora em estudo, no período de 12 meses.....	88
Figura 19	Principais tipos de motivo registrado na unidade regional DMda distribuidora em estudo, no período de 12 meses .....	88
Figura 20	Principais tipos de motivo registrado na unidade regional DT da distribuidora em estudo, no período de 12 meses.....	89
Figura 21	Principais tipos de motivo registrado na unidade regional DO da distribuidora em estudo, no período de 12 meses.....	89
Figura 22	Quantidade total de inspeções realizadas, por unidade regional da distribuidora em estudo, no período de 12 meses.....	91
Figura 23	Alocação das equipes de campo de acordo com retorno e a probabilidade das inspeções serem procedentes, na distribuidora em estudo .....	94

Figura 24	Alocação das equipes de campo da distribuidora em estudo, usando como referência o resultado do algoritmo <i>SimpleKMeans</i> .....	119
Figura 25	Alocação das equipes de campo da distribuidora em estudo, de acordo com <i>SimpleKMeans</i> , usando como referência uma nova base de dados para validação da seleção das variáveis .....	123
Figura 26	Modelo da matriz de confusão usado na validação das variáveis da distribuidora em estudo .....	125
Figura 27	Matriz de confusão e estatísticas dos dados analisados da distribuidora em estudo .....	126
Figura 28	Procedimento Operacional Padrão (POP) da atividade de seleção e liberação de inspeção, incorporando a aplicação Perdas DSS, na distribuidora em estudo .....	135
Figura 29	Arquitetura da base de dados da aplicação Perdas DSS, desenvolvido para a distribuidora em estudo .....	137
Figura 30	Demonstração simplificada do algoritmo <i>k-means</i> , usado na aplicação desenvolvida para a distribuidora em estudo .....	143

## LISTA DE TABELAS

Tabela 1	Resultado das inspeções liberadas por unidade regional da distribuidora em estudo, no período de 12 meses.....	92
Tabela 2	Percentual de classificação correta e precisão média ponderada dos algoritmos de classificação, usado no teste e validação das variáveis, da distribuidora em estudo .....	115
Tabela 3	Consumo médio mensal por classe de consumidores e por número de fases, no período de 12 meses, em kWh, da distribuidora em estudo .....	121
Tabela 4	Resultado da única inspeção de campo realizada usando como referência o resultado gerado pela aplicação Perdas DSS, desenvolvido para a distribuidora em estudo .....	147

## SUMÁRIO

1	INTRODUÇÃO .....	16
2	AS PERDAS DO SETOR ELÉTRICO .....	20
2.1	Perdas globais e perdas técnicas .....	20
2.2	Perdas não técnicas ou comerciais .....	22
2.2.1	Perdas não técnicas: cenário internacional .....	25
2.2.2	Perdas não técnicas: cenário nacional .....	29
3	SISTEMA DE INFORMAÇÃO .....	39
3.1	Sistemas de informação: conceito, dimensões e tipos .....	39
3.3	<i>Decision support system</i> ou Sistema de apoio à decisão (DSS) .....	43
3.3.1	Introdução ao DSS .....	43
3.3.2	Conceito de DSS .....	44
3.3.3	Tipos ou subgrupos de DSS .....	45
3.3.4	Características de DSS .....	47
3.3.5	Etapas do desenvolvimento de um DSS .....	48
4	MATERIAL E MÉTODOS .....	50
4.1	Objeto de estudo .....	50
4.2	Classificação e método científico .....	50
4.3	Pesquisa-ação .....	53
5	RESULTADOS E DISCUSSÃO .....	60
5.1	Contexto e propósito .....	61
5.1.1	Razão para a ação .....	61
5.1.2	Razão para a pesquisa .....	63
5.2	Ciclo PA-1: POP e mapeamento das liberações .....	66
5.2.1	Obtenção de dados do Ciclo PA-1 .....	67
5.2.2	<i>Feedback</i> de dados do Ciclo PA-1 .....	67
5.2.3	Análise dos dados do Ciclo PA-1 .....	79

5.2.4	Plano de ação do Ciclo PA-1.....	80
5.3	Ciclo PA-2: histórico das inspeções realizadas .....	81
5.3.1	Obtenção de dados do Ciclo PA-2.....	81
5.3.2	<i>Feedback</i> de dados do Ciclo PA-2 .....	81
5.3.3	Análise dos dados do Ciclo PA-2.....	82
5.3.4	Plano de ação do Ciclo PA-2.....	93
5.4	Ciclo PA-3: análise das variáveis das inspeções realizadas .....	95
5.4.1	Obtenção de dados do Ciclo PA-3.....	95
5.4.2	<i>Feedback</i> de dados do Ciclo PA-3 .....	95
5.4.3	Análise dos dados do Ciclo PA-3.....	100
5.4.4	Plano de ação do Ciclo PA-3.....	113
5.4.5	Implementação do Ciclo PA-3.....	114
5.4.5.1	Teste e validação das variáveis, usando a mesma base de dados usada na seleção .....	115
5.4.5.2	Teste e validação das variáveis, usando a uma base de dados diferente da usada na seleção .....	120
5.4.6	Avaliação do Ciclo PA-3 .....	128
5.4.6.1	Ciclo PA-4: desenvolvimento da aplicação perdas DSS.....	131
5.4.7	Interesses e motivações – Ciclo PA-4.....	131
5.4.8	Objetivos e recursos a usados – Ciclo PA-4 .....	132
5.4.9	Processo decisório e Decisões-chave – Ciclo PA-4.....	133
5.4.10	Adequação das soluções propostas – Ciclo PA-4 .....	134
5.4.11	Projeto de desenvolvimento e de implantação – Ciclo PA-4.....	134
5.4.12	Avaliação do Ciclo PA-4 .....	149
6	CONCLUSÃO .....	151
	REFERÊNCIAS .....	154
	APÊNDICE .....	163
	ANEXO .....	196

## 1 INTRODUÇÃO

A distribuição de energia elétrica apresenta, historicamente, perdas que são caracterizadas pela diferença entre o volume da energia, disponibilizado às concessionárias distribuidoras (total comprado) e o volume de energia faturado (total vendido). A diferença entre o total comprado e vendido resulta na perda global de energia que, de acordo com a origem, pode ser classificada em perda técnica, quando vinculada aos materiais e equipamentos usados na distribuição e em perda não técnica ou comercial, quando relacionadas, principalmente, ao furto e à fraude no consumo de energia.

As perdas não técnicas ou comerciais no Brasil têm despertado a atenção de órgãos reguladores e de concessionárias distribuidoras de eletricidade. A preocupação dessas organizações se justifica porque a atual legislação do setor elétrico brasileiro estipula um valor máximo de repasse às tarifas aplicadas aos consumidores, referente a esse tipo de perda (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA – ANEEL, 2011). Dessa forma, reduzir as perdas comerciais possibilitaria, de modo geral, um aumento de receita das concessionárias e uma redução da tarifa de energia elétrica. Além do benefício econômico, outros poderiam ser obtidos como, por exemplo, a diminuição dos riscos de acidentes para os usuários e a melhoria na qualidade da energia e do serviço prestado pelas distribuidoras (FERREIRA et al., 2012).

As perdas não técnicas, considerando os fatores exógenos ao sistema elétrico, podem ser divididas em inadimplência, furto de energia e fraude no consumo de eletricidade. As perdas comerciais, em sua grande maioria estão focadas na fraude do consumo de energia que é caracterizada, principalmente, pela violação ou adulteração de medidores, com a intenção de redução ou eliminação da conta de eletricidade. É com base nesse conceito que este trabalho foi fundamentado.

As distribuidoras de energia elétrica a partir de bancos de dados, como o de clientes e o de registros de padrões de consumo, conseguem obter listagens de consumidores (unidades consumidoras ou instalações) suspeitos de fraudar o consumo de eletricidade (LEÓN et al., 2011). Entretanto, saber apenas quem e quais são essas unidades suspeitas pode não resolver o problema das perdas não técnicas. Para isso, uma das soluções é auditar ou inspecionar as instalações suspeitas, para a efetiva constatação ou não do problema.

A auditoria ou a inspeção de todos os suspeitos de consumo irregular poder ser técnica e economicamente inviável para as distribuidoras (NIZAR et al., 2006; QUEIROGA, 2005). Essa afirmação se fundamenta em diversos motivos, dos quais se destacam: 1) a quantidade de suspeitos de consumo irregular de energia, no Brasil, atinge a casa dos milhões (ANEEL, 2011); 2) seria necessário muito tempo para inspecionar todos os suspeitos; 3) inspecionar todos os suspeitos em um curto espaço de tempo demandaria elevado investimento em um grande número de equipes de inspeção (NIZAR et al., 2006); e 4) grande parte das instalações é apenas suspeita de consumo irregular (CALILI, 2005), sendo que, de fato, o consumo de energia pode estar regular. Pelo exposto, a possibilidade de ganho das distribuidoras na recuperação e/ou no incremento do faturamento de energia elétrica, pode ser menor do que o próprio custo de inspecionar todos os suspeitos.

Diante do exposto, a questão a ser investigada é: como alocar equipes de campo para viabilizar, técnica e economicamente, a inspeção de milhares de consumidores suspeitos de fraudar o consumo de energia elétrica? Além disso, é possível desenvolver algum instrumento que gere informações que auxilie na alocação de equipes?

Diversos autores abordaram o tema, inclusive com questões semelhantes (BERNARDES, 2010; CALILI, 2005; DICK, 1995; LEÓN et al., 2011; ONAT, 2010; ORTEGA, 2008; PENIN, 2008; QUEIROGA, 2005; REIS, 2005;

SMITH, 2004; STEADMAN, 2010; VIEIRALVES, 2005) entre outros. Analisando esses trabalhos, pode-se afirmar que, na maioria deles, foram utilizados recursos de tecnologia da informação, aliada às técnicas de descoberta de conhecimento em banco de dados, para analisar os dados disponíveis nas distribuidoras e melhorar a qualidade da informação sobre os suspeitos de fraudarem o consumo de energia.

Considerando as discussões sobre o tema e a questão de pesquisa levantada, o objetivo deste trabalho é, em conjunto com uma distribuidora de energia elétrica, construir, reflexivamente, ações que auxiliem na alocação de equipes de campo, para viabilizar, técnica e economicamente, a inspeção de consumidores suspeitos de fraudar o consumo de energia. Especificamente, pretende-se:

- a) levantar, interativamente com a distribuidora, as dificuldades existentes na alocação das equipes de campo; e
- b) desenvolver, por meio da reflexão à luz da teoria e da experiência da distribuidora, um sistema de apoio à decisão que auxilie na solução do problema.

Considerando o objetivo proposto, este trabalho teve como objeto de estudo uma das 64 distribuidoras de energia que atuam no Brasil. A principal justificativa para a escolha dessa distribuidora está no fato de que, além de atuar nacional e internacionalmente, ela se dispôs, mediante contrato de confidencialidade, a disponibilizar recursos humanos, tecnológicos e organizacionais para o desenvolvimento do trabalho.

Para suportar a construção reflexiva e interativa das ações propostas no objetivo, adotou-se a pesquisa-ação como procedimento técnico de pesquisa. A escolha desse procedimento se fundamentou na necessidade de gerar

conhecimento para solucionar um problema organizacional, a partir do desdobramento de uma série de ações ao longo do tempo.

Este trabalho está estruturado em cinco itens além desta Introdução. O item dois define as perdas técnicas e não técnicas, apresentando a importância do combate às perdas comerciais. Além disso, aborda-se, de forma mais abrangente, as perdas não técnicas e os principais trabalhos, internacionais e nacionais, já realizados sobre o tema.

No item três, faz-se uma abordagem sobre o conceito, as dimensões e os tipos de sistemas de informação aplicados a organizações e, na sequência, apresenta-se as características e o tipo do sistema de informação que melhor pode auxiliar na solução do problema das perdas não técnicas.

O quarto item é utilizado para explicitar o objeto de estudo, a classificação da pesquisa e o procedimento técnico utilizado no trabalho.

O item cinco apresenta e analisa os resultados encontrados no trabalho. Esse item está dividido em cinco subitens que seguem os passos de implantação de uma pesquisa-ação, facilitando a análise e a discussão dos resultados encontrados.

Por fim, no último item, apresentam-se as conclusões do trabalho. Na sequência, seguem as referências, os apêndices e os anexos utilizados.

## **2 AS PERDAS DO SETOR ELÉTRICO**

De acordo com levantamentos da Agência Nacional de Energia Elétrica (ANEEL), o prejuízo com a perda não técnica de energia elétrica no Brasil chega a quase R\$7 bilhões ao ano, o que representa em torno de 13% da energia consumida (SIMÃO, 2012). Além de constituir-se em um grave problema social, há também reflexos significativos sobre o valor da tarifa de energia e, conseqüentemente, sobre a eficiência econômica do país.

Considerando o cenário apresentado, o objetivo desse item é definir as perdas técnicas e não técnicas e apresentar a importância do combate às perdas comerciais. Para introduzir o assunto, faz-se uma abordagem sucinta sobre as perdas globais e as perdas técnicas no setor elétrico. Na sequência, aborda-se, de forma mais abrangente, as perdas não técnicas e, por fim, apresenta-se os principais trabalhos, internacionais e nacionais, já realizados sobre o tema.

### **2.1 Perdas globais e perdas técnicas**

A produção de riqueza de forma geral, seja na indústria, comércio, serviços ou atividades agroindustriais, está diretamente relacionada ao uso da energia elétrica. Apesar disso, a quantidade de energia utilizada nos processos de transformação ou até mesmo no uso residencial é diferente do volume total de energia produzido. Uma das justificativas para essa afirmação é que a quantidade de energia que uma companhia distribuidora fornece a uma rede elétrica é diferente da quantidade de energia entregue, na mesma rede.

Para Ortega (2008) a diferença entre a energia fornecida a uma rede e a energia entregue (faturada) é o que se chama, tecnicamente, de perda global de energia elétrica. Essa perda, ainda segundo o autor, pode ser mensurada pela

diferença entre a energia requerida por uma rede, ou a que a distribuidora precisa para abastecer seus clientes, e a energia efetivamente fatura daquela rede.

Diminuir as perdas do sistema elétrico tem se mostrado extremamente relevante. Para se ter uma ideia da magnitude desse problema, a Associação Brasileira das Distribuidoras de Energia Elétrica (Abradee) indica que a perda global de energia no Brasil se aproxima a 17% de toda a energia consumida, o que corresponde a 84.019 *gigawatts*-hora (PERDA..., 2012).

A perda global de energia pode ser classificada, quanto à origem, em perda técnica e perda não técnica ou comercial (INSTITUTO ACENDE BRASIL, 2007; LIMA, 2005; PENIN, 2008; QUEIROGA, 2005; VIEIRALVES, 2005).

As perdas técnicas são, de certo modo, “toleráveis” nos sistemas elétricos por serem causadas, de acordo com Calili (2005), Penin (2008) e Vieiralves (2005), por ações internas nos materiais, inerentes aos processos de transporte e dissipação de energia, nos diversos componentes dos sistemas elétricos. Tecnicamente, esse fenômeno é conhecido como efeito Joule e ocorre em cabos, condutores, transformadores, medidores e equipamentos.

Para Vieiralves (2005), as perdas técnicas podem receber, ainda, outras classificações quanto: à natureza, em perda de demanda e perda de energia; à localização, relacionada à perda global, a de transmissão e a de distribuição; e ao segmento, podendo ser no sistema de alta tensão, na subestação, na rede primária, em transformadores, na rede secundária, no ramal de serviço, em medidores, entre outros.

Para reduzir esse tipo de perda, Ortega (2008) afirma que inúmeras estratégias podem ser adotadas, como a tecnologia utilizada, a qualidade dos serviços de manutenção, a adequação do sistema elétrico ao mercado consumidor, o modo de operação dos sistemas, dentre outras. Por esses motivos,

Penin (2008) complementa que as perdas técnicas podem ser estimadas a partir de modelos matemáticos previamente estabelecidos.

Já as perdas não técnicas ou comerciais são causadas, de modo geral, pelo uso irregular da energia elétrica, ocasionado, principalmente, pela ação de terceiros ou por equipamentos defeituosos (ORTEGA, 2008). Ou seja, para Penin (2008) e Vieiralves (2005), esse tipo de perda está vinculado à falta de faturamento de parte da energia distribuída.

Por serem parte da perda global, juntamente com a perda técnica, os valores referentes às perdas não técnicas podem ser estimados pela subtração do valor das perdas técnicas das perdas globais de energia. Entretanto, Penin (2008) comenta que, dificilmente, seria possível diferenciar esses valores diretamente das perdas globais.

## **2.2 Perdas não técnicas ou comerciais**

De acordo com Brasil (2008), até pouco tempo, estudos sobre as perdas comerciais não eram comuns, sendo que as pesquisas sobre perdas, quase em sua totalidade, eram dedicadas às formas de mensurar as perdas técnicas. Uma das justificativas para tal situação é que os órgãos reguladores das distribuidoras de energia não forneciam incentivos adequados para o efetivo combate às perdas comerciais (INSTITUTO ACENDE BRASIL, 2007).

Atualmente, os órgãos reguladores e as distribuidoras vivem outra realidade. Mesmo com o aumento de investimentos e da fiscalização das distribuidoras, a Agência Nacional de Energia Elétrica (ANEEL) afirma que a perda não técnica de energia elétrica no Brasil chega a quase R\$7 bilhões ao ano, o que representa em torno de 13% da energia consumida (SIMÃO, 2012).

Ao contrário das perdas técnicas, as comerciais não deveriam ocorrer, sendo, portanto, consideradas “inaceitáveis” nos sistemas elétricos. Isso porque,

de acordo com Vieiralves (2005), dentre outros fatores, as perdas comerciais têm expressiva participação de diversos fatores exógenos ao setor elétrico, como: a questão social; nível de emprego e renda; composição da balança comercial regional e nacional.

Considerando os fatores exógenos, as perdas não técnicas podem ser divididas em três grupos distintos: a inadimplência, o furto de energia e a fraude no consumo de eletricidade.

A inadimplência, de modo geral, é definida como a recusa ou o não pagamento das faturas de energia elétrica na data estipulada pelos consumidores das distribuidoras de energia. Segundo o Instituto Acende Brasil (2007), estima-se que as contas vencidas e irrecuperáveis do setor elétrico correspondam a 1,2% do faturamento das distribuidoras, o que totalizaria cerca de R\$1 bilhão ao ano.

Entre as distribuidoras concessionárias, a inadimplência varia expressivamente sendo que, em algumas, o nível de inadimplência era da ordem de 5,6% em 12 meses (RESULTADO, 2008). A justificativa para tal variação é que esse fato está associado às características do ambiente institucional, jurídico, econômico, social e educacional da área de concessão das distribuidoras. Outro dado que merece destaque é que o setor público, relativamente, é a classe de consumo que mais contribui para o aumento do nível de inadimplência das distribuidoras (INSTITUTO ACENDE BRASIL, 2007; ORTEGA, 2008; PREFEITURAS..., 2012; VEIRALVES, 2005).

As perdas não técnicas por inadimplência são exaustivamente discutidas e avaliadas em vários trabalhos que, dentre eles se destacam: Araújo (2007), Calili (2005), Ortega (2008) e Penin (2008), além de relatórios técnicos como Instituto Acende Brasil (2007) e Resultado (2008).

Instituto Acende Brasil (2007) afirma ainda que as perdas por inadimplência e por furto e fraude são complementares. Pois, segundo o trabalho, um maior combate às fraudes e ao furto resulta na incorporação de

clientes com maior propensão à inadimplência. Da mesma forma, ao intensificar o combate à inadimplência, por meio da suspensão do fornecimento, há um aumento de furto e fraude no consumo.

O furto de energia é outro grupo de perdas comerciais. Esse é caracterizado principalmente por desvios no ramal de entrada (antes do medidor de energia) ou ligações clandestinas ao sistema energético da concessionária. Por ser difícil mensurar o volume de energia que é furtado em uma rede, geralmente o furto e a fraude no consumo de energia são tratados em conjunto.

O último grupo, fraude no consumo de energia, é caracterizado, principalmente, pela manipulação do total de energia consumido por um cliente. O

Quadro 1, representa uma relação das formas de fraude no consumo de energia e suas características.

De acordo com Penin (2008), quaisquer perdas que não possam ser enquadradas como técnicas são, comumente, consideradas como perdas não técnicas. Segundo o autor, isso é admissível devido à própria natureza dos cálculos elétricos de perdas, que só podem ser feitos sobre cargas conhecidas. Assim, como as cargas que geram as perdas não técnicas não são conhecidas é possível uma melhora no desempenho dos cálculos das perdas técnicas. Além disso, o autor afirma que o percentual de perdas técnicas adicionado às perdas não técnicas é quase irrelevante, não comprometendo, assim, o índice de perdas não técnicas.

Quadro 1 Formas de fraude no consumo de energia, encontrado nas distribuidoras de energia elétrica

FORMAS	CARACTERÍSTICA
Fraude nos medidores de energia	a) violação ou adulteração do medidor, com a intenção de redução ou eliminação de consumo da unidade consumidora – ponteiros do medidor deslocados, ligações do medidor invertidas, bobina de potencial interrompida, engrenagem do medidor substituída, sequência de fases invertida, curto-circuito na entrada ou saída do medidor; b) outro tipo de fraude ocorre quando uma unidade consumidora regular é cortada (por falta de pagamento) e o consumidor faz a religação direta à rede, por conta própria.
Irregularidades técnicas no equipamento	a) problema técnico nas instalações elétricas do consumidor que não foi provocado artificialmente ou por terceiros (mau funcionamento ou não funcionamento do medidor); b) geralmente causadas pela deterioração dos equipamentos. Entretanto, não podem ser estimadas pelas metodologias de cálculo de perdas técnicas conhecidas.
Administrativas (erros de leitura, processamento, consumo próprio e outros)	a) problemas nos processos administrativos das distribuidoras gerados, principalmente por erros de leitura do medidor, erro no processamento e faturamento.

Fonte: Organizado pelo autor a partir de Ortega (2008), Penin (2008), Queiroga (2005) e Reis (2005)

### 2.2.1 Perdas não técnicas: cenário internacional

Apesar de Brasil (2008) afirmar que existiam poucos estudos sobre perdas comerciais, internacionalmente, há muito esse assunto vem sendo discutido e estudado.

O trabalho de Dick (1995), por exemplo, apresenta como as distribuidoras no Reino Unido deveriam combater o problema, que na época, causavam prejuízos da ordem de 50 milhões de libras por ano. No trabalho, o autor apresentou a natureza e a escala do problema, com comentários sobre a incidência relativa de cada tipo de perda não técnica ou interferência no sistema elétrico. Além disso, o autor também abordou a estrutura legal em que o problema se enquadrava. Sua conclusão mostrou que as empresas deveriam usar as seguintes estratégias para vencer o problema: 1) combater as perdas, usando ações efetivamente coordenadas e estruturadas; 2) descobrir os furtos e as fraudes, por meio de denúncia de consumidores e denúncia de empregados não envolvidos com a detecção de fraudes; 3) realizar inspeções por empregados especializados; 4) recuperar as perdas, usando ações policiais e processos judiciais; e 5) prevenir novas ações de furto ou fraude, adotando recursos tecnológicos novos para dificultar e coibir tais ações.

Mais abrangente do que Dick (1995), o trabalho de Smith (2004) avaliou a extensão das perdas comerciais de energia em uma amostra de 102 países considerando o período de 1980 e 2000. O autor comentou que o sistema elétrico nunca estará 100% seguro contra ações de furto ou fraudes. E, apesar do percentual de ações como essas, em muitos sistemas, girarem em torno de 1% a 2% da energia gerada, o prejuízo financeiro era elevado em função da quantidade de energia distribuída.

O autor afirma ainda que os níveis de perda não técnica e os prejuízos financeiros, em muitos países, aumentaram de tal maneira que as distribuidoras estavam passando por crises financeiras. Além disso, o autor destacou que as perdas comerciais em alguns sistemas estavam estreitamente ligadas ao governo

e aos ambientes social, econômico e político. Segundo o autor, a corrupção social, econômica e política e o furto e a fraude de energia prospera fora de si. Com isso, as perdas poderiam ser reduzidas a níveis moderados por métodos, engenharia técnica e tecnologia. Entretanto, isso seria difícil de ser conseguido considerando a extensão da corrupção nos diversos níveis sociais.

Mais recentemente, outros autores estudaram o problema das perdas não técnicas sob a perspectiva socioeconômica: Steadman (2010) na Jamaica e Onat (2010) na Turquia.

O objetivo do trabalho de Steadman (2010) foi avaliar o padrão das perdas comerciais de energia na Jamaica a partir do ano de 1993 e tentar determinar quais os principais agentes responsáveis pela condução desses padrões. Em sua análise, o autor comenta que esse problema não é exclusivo da Jamaica, tornando-se mais frequente em países do Sul da Ásia, América do Sul, além de alguns incidentes isolados nos Estados Unidos. Entretanto, segundo o autor, no caso da Jamaica, o problema é alarmante, pois as perdas chegaram a 13% da energia líquida gerada e a 18% da energia total distribuída. Ao comparar as perdas por furto ou fraude de energia com o índice de preço ao consumidor da Jamaica, Steadman (2010) concluiu que nos períodos em que o custo de vida diminuiu as perdas comerciais também diminuíram e o volume de energia faturada foi maior, sugerindo a substituição das ações de furto e de fraude pela compra da energia. Além disso, essas ações ocorriam em todos os grupos analisados (residências, indústrias, pequenos comércios, grandes comércios) e em todas as classes sociais. Sendo mais expressivos em residências e pequenos comércios.

O trabalho de Onat (2010) avaliou o uso ilegal de energia na Turquia. Assim como Smith (2004), o autor afirmou que é praticamente impossível evitar as perdas no sistema de energia. Mas, em países desenvolvidos, há um controle dessas perdas que, geralmente, estão entre 1% e 2%, ao passo que na Turquia,

esse valor chegou a 14,4% do total da energia produzida em 2008. A conclusão do autor foi que seria necessário adotar medidas para barrar as perdas comerciais, como por exemplo, acelerar as políticas de privatização, adotar políticas de regionalização de preços, conscientizar a população das dimensões econômica e social do uso ilegal de energia e incentivar funcionários públicos e até mesmo os cidadãos a denunciarem o uso ilegal de energia elétrica.

Para finalizar a abordagem internacional, além dos trabalhos com perspectiva socioeconômica sobre o assunto, há também aqueles que propuseram abordagens aplicadas para combater o problema, usando como fundamento a tecnologia da informação associada a ferramentas estatísticas de análise de dados: Nagi et al. (2008), realizado na Malásia e outro é o de León et al. (2011), realizado na Espanha.

No trabalho de Nagi et al. (2008), os autores usaram como fundamento a mineração de dados (*data mining*), especificamente *Support Vector Machine* (SVM) para reduzir as perdas não técnicas no setor de distribuição. O modelo proposto pré-selecionava os clientes a serem inspecionados com base em irregularidades e comportamento de consumo anormal. Como resultado, o trabalho gerava classes para classificação que eram utilizadas para selecionar os suspeitos. Os autores concluíram afirmando que a simulação realizada provou que o método proposto seria mais eficaz do que ações praticadas pela distribuidora analisada.

No trabalho de León et al. (2011), os autores propuseram o sistema especialista integrado (*Integrated Expert System – IES*) que usava dados amostrais reais, extraídos diretamente da base de dados de uma distribuidora de eletricidade na Espanha, para analisar e classificar os consumidores sob uma ou mais categorias diferentes, auxiliando na identificação e categorização das perdas comerciais. O sistema estava em fase de teste e era usado como um sistema de apoio à decisão (SAD ou *Decision Support System – DSS*).

De acordo com os autores, esse sistema usa como fundamento: módulo de mineração de texto, utilizado para coletar as informações dos agentes que realizam as inspeções de campo; módulo de mineração de dados, que define as regras de consumo dos clientes, analisando se está normal ou não; módulo de sistema especialista baseado em regras (*Rule Based Expert System* – RBES), que utiliza as informações geradas pelos outros dois módulos; e, por fim, outro módulo, formado por um conjunto de parâmetros estabelecidos por especialistas da distribuidora, utilizado para classificar os clientes com base nos problemas que apresentam.

Avaliando os resultados segundo os autores, percebe-se que um dos pontos mais complexos do trabalho foi processar a enorme quantidade e a grande variedade de dados dos clientes. E, apesar disso, o IES classificou as perdas não técnicas por meio da análise automática desses dados. Além disso, com o uso do IES foi possível gerar novas informações úteis sobre os clientes, aumentar a eficiência das inspeções dos clientes (inspeção de clientes que estão fraudando o consumo de energia), detectar novos casos e padrões de perdas não técnicas.

### **2.2.2 Perdas não técnicas: cenário nacional**

No contexto nacional, dois trabalhos abrangentes se destacam: o de Araújo (2007) e o de Penin (2008).

No primeiro, o objetivo principal foi aprofundar a discussão dos determinantes para o problema de perdas e inadimplência, englobando não somente o ponto de vista sociocultural, mas também variáveis de categorias como setorial, grau de desenvolvimento, renda e complexidade social. Para fechar o trabalho, o autor ainda analisou a questão das perdas sob o ponto de vista tarifário, técnico e comercial conferido pela Agência Nacional de Energia

Elétrica (ANEEL). Para explicar o problema a partir das variáveis propostas o autor utilizou análises estatísticas que possibilitaram a construção de modelos explicativos para o problema. Dentre as conclusões de Araújo (2007), as que mais se destacaram foram:

- a) as variáveis das categorias mencionadas explicavam 73% da perda de energia e 50% da inadimplência. Com isso, o autor afirmou que:
  - i) quanto maior a tarifa, maiores serão as perdas, o que contribui para o incremento da tarifa (criação de um círculo vicioso);
  - ii) os consumidores da classe residencial são os principais responsáveis pelas perdas;
  - iii) quanto maior o grau de capacitação das pessoas, melhores são as condições de vida e por conseguinte maior a noção dos direitos e deveres dos indivíduos (por isso, menores são as perdas);
  - e iv) a inadimplência está correlacionada positivamente com a perda da capacidade de pagamento da população e negativamente com a urbanização, ou seja, quanto mais urbanizado menores as taxas de inadimplência;
- b) há uma estreita ligação entre as perdas de energia e a inadimplência;
- c) para os Estados que compõem as regiões Sul e Sudeste as variáveis que melhor explicaram as perdas foram: i) tarifa média da concessionária; ii) posse de ar condicionado; iii) índice de posse de bens; e iv) taxa de mortos por agressão. Já nos Estados das regiões Norte, Nordeste e Centro-Oeste, as variáveis foram: i) percentual de consumo residencial; ii) taxa de urbanização; e iii) taxa de mortos por agressão;
- d) por fim, o autor sugeriu que a ANEEL crie incentivos adequados para prover a atualização da regulamentação necessária para que as distribuidoras combatam as perdas e a inadimplência

adequadamente. E recomenda às distribuidoras que desenvolvam novas tecnologias e procedimentos para dificultar o furto de energia e mecanismos mais eficazes de cobrança de faturas em atraso.

Com o objetivo distinto do trabalho de Araújo (2007), mas com algumas conclusões semelhantes, o trabalho de Penin (2008) analisou as perdas não técnicas, adotando uma abordagem multidisciplinar, considerando cinco aspectos distintos que são: a) usou uma abordagem multidisciplinar das perdas não técnicas e pormenorizou os cálculos das perdas técnicas; b) traçou um panorama (nacional e internacional) em relação à prevenção e combate ao furto e fraude de energia; c) examinou o histórico da utilização de ferramentas estatísticas e de mineração de dados (*data mining*); d) desenvolveu uma metodologia e um sistema computacional para subsidiar o combate às fraudes; 5) analisou os aspectos envolvidos na regularização e na recuperação de receitas, propondo uma metodologia para cálculo de consumo; e e) abordou o problema das perdas sob o ponto de vista econômico, dimensionando quanto deveria ser investido financeiramente na diminuição das perdas comerciais.

Pela abrangência da proposta de Penin (2008), várias conclusões foram apresentadas, das quais se destacam:

- a) o furto e a fraude de energia estão ligados fortemente à diversidade socioeconômica do país, o que aprofunda o problema. Seria necessário uma reavaliação global do processo de aplicação de recursos pelas distribuidoras/concessionárias para a redução das perdas de energia;
- b) divisão das perdas não técnicas em dois conjuntos: 1) PNT1 – que abrange especificamente furtos e fraudes; e 2) PNT2 – que engloba

erros em medidores, consumidores não medidos, iluminação pública e outros;

- c) investimento em campanhas de prevenção, caracterizando a perspectiva criminosa do furto e da fraude e programas socioeducativos em comunidades com elevados índices de perdas. Além de investir em tecnologia (como por exemplo, *Smart Grid* ou *IntelliGrid* medição remota), em outras formas de disponibilização da energia ao consumidor (como medidores pré-pagos); e qualificação das equipes de campo para atuarem na identificação, inspeção, combate e prevenção dos furtos e fraudes de energia;
- d) a mineração de dados (*data mining*) representou um notável avanço na busca de instalações com irregularidades no consumo de energia. Entretanto, para utilizar esses modelos, é muito importante a qualidade dos dados históricos das unidades consumidoras;
- e) o autor propôs a estimativa de consumo de energia de clientes fraudadores a partir da potência instalada por três métodos diferentes: regressão, tabela com intervalo fixo e tabela com frequência fixa;
- f) por fim, sob o ponto de vista econômico, melhorias deveriam ser adotadas nos modelos desenvolvidos, mesmo que, na opinião de especialistas, os valores encontrados sejam realistas.

Mesmo não sendo tão abrangentes quanto os trabalhos anteriores, outros trabalhos nacionais se destacam no combate e prevenção das perdas não técnicas. Os trabalhos de Cabral et al. (2006) e Cabral e Gontijo (2004) utilizaram a tecnologia da informação e a teoria *Rough Sets* (PAWLAK, 1982) no combate às perdas comerciais. No primeiro, a teoria proposta por Pawlak (1982) foi utilizada para reduzir o número de atributos condicionais e, a partir

desse pequeno conjunto de atributos gerar regras que identificassem e classificassem o comportamento fraudulento do consumidor. No segundo, a teoria *Rough Sets* foi usada em conjunto à descoberta de conhecimento em banco de dados (*Knowledge Discovery in Databases* – KDD) para detectar fraude no consumo de energia. A metodologia proposta por Cabral et al. (2006) fez uma detalhada avaliação entre clientes com consumo normal e com consumo fraudulento, a partir de dados históricos de distribuidoras, identificando comportamento fraudulento de consumo. Isso permitiu detectar em bancos de dados das distribuidoras clientes que apresentavam características fraudulentas, o que melhorou o resultado das equipes de campo, responsáveis pela auditoria dos consumidores suspeitos.

O trabalho de Calili (2005) tinha como objetivo a detecção de fraude na rede elétrica por meio de uma Pesquisa de Posses e Hábitos de Consumo (PPH), utilizando duas técnicas computacionais para análise: a Rede Neural (especificamente o algoritmo de Mapa Auto-organizável de Kohonen) e a Lógica Fuzzy. Metodologicamente, o autor desenvolveu seu trabalho da seguinte forma: 1) uso de redes neurais para formação de *clusters*, a partir de um banco de dados de cadastro disponibilizado pela distribuidora; 2) aplicação da PPH em cada um desses grupos identificados pela rede neural; e 3) aplicação da lógica Fuzzy para analisar os consumidores dos clusters, classificando-os em adimplentes, inadimplentes e fraudulentos. O modelo proposto foi testado na cidade Guarujá/SP, área de concessão da distribuidora ELEKTRO.

Calili (2005) fecha o trabalho com uma série de conclusões, dentre as quais se destacam: a) muitas vezes as inspeções são direcionadas a unidades de consumo que não possuem fraude; b) nos testes de validação, o modelo proposto conseguiu um percentual de acerto de 70,38% para os clientes com fraude e de 47,51% para clientes inadimplentes; c) na comparação com os resultados de campo, o modelo conseguiu um percentual de acerto de 68,92% para os clientes

com fraude e de 42,98% para clientes inadimplentes; e d) mesmo não testando o modelo em buscas diretas em campo (inspeções), para o autor, os resultados apresentados mostram que o modelo era robusto.

Outro trabalho que usou ferramentas estatísticas para analisar os consumidores de energia foi o de Queiroga (2005). Esse tinha como objetivo melhorar a identificação de perdas comerciais por meio do uso de *data mining*, evidenciando padrões que indicassem a possibilidade de fraude. Segundo o autor, simplesmente aumentar o número de equipes de inspeção não seria economicamente viável e poderia não prover melhorias significativas nos resultados das perdas. Metodologicamente, o trabalho seguiu o ciclo de mineração de dados proposto por Fayyad, Piatetsky-Shapiro e Smyth (1996). Para testar a aplicação, o autor usou a base de dados da empresa ESCELSA.

Após o desenvolvimento de vários ciclos de *data mining*, o autor chegou às seguintes conclusões: a) a dificuldade de acesso às bases de dados e a pouca qualidade dos dados armazenados na *Data Warehouse* (DWH) da empresa foram os principais limitantes para o desenvolvimento do trabalho; b) as séries de consumo com suas respectivas características foram os maiores geradores de padrões com poder discriminatório; e c) pelas limitações do trabalho, não foi possível comprovar, em campo, os resultados encontrados nos ciclos de mineração de dados. Entretanto, os resultados comparativos mostraram inspeções com acertos da ordem de 25% a 45%, enquanto que pelo método adotado pela empresa os acertos das inspeções de campo estavam entre 7% e 25%.

Enquanto Queiroga (2005) analisou o uso de classificadores a partir de algoritmos de aprendizado de máquina, Dias (2006) usou o conhecimento de especialistas, empregando essas regras para a pesquisa de suspeitos de fraude. O objetivo deste trabalho foi realizar um estudo sobre o conhecimento de especialistas em relação à fraude no consumo de energia e propor uma

arquitetura de um *software*, chamado de Sistema de Auxílio à Investigação de Perdas Elétricas (SAUIPE), que apoiasse o processo de seleção de suspeitos de fraude. Assim como no de Queiroga (2005), este trabalho usou dados da empresa ESCELSA, com cerca de 40.600 clientes inspecionados no ano de 2005.

Após o desenvolvimento do *software*, inclusão das regras obtidas com os especialistas, treinamento e análise dos suspeitos de fraude no consumo, o autor pode concluir: 1) o SAUIPE obteve um percentual de acerto, após testes de campo, de 27,17%; ao passo que, a forma tradicional de seleção da empresa tinha índice de 12%; e 2) foi possível realizar a documentação e formalização do conhecimento dos especialistas em uma base de conhecimento disponível e utilizável por outros profissionais e sistemas.

Outro trabalho que usou *data mining* foi o de Reis Filho (2006), que tinha objetivo de desenvolver um sistema de auxílio à detecção de fraudes em unidades consumidoras e identificação de medidores de energia com problemas.

Metodologicamente, Reis Filho (2006) também seguiu o ciclo de mineração de dados proposto por Fayyad, Piatetsky-Shapiro e Smyth (1996). Para testar a aplicação, o autor usou a base de dados de uma distribuidora, com cerca de 620 mil consumidores. Após desenvolver ciclos de *data mining*, usando Árvore de Decisão como uma técnica de inteligência artificial, o autor chegou a várias conclusões: a) a qualidade dos dados utilizados e a dificuldade de acessá-los limitaram o desenvolvimento do trabalho, impossibilitando a descoberta de novos e úteis conhecimentos; b) a técnica gerou bons resultados, sendo que os índices de acerto, nos testes de validação, ficaram em torno de 40%, ao passo que a distribuidora era em torno de 5%; e c) a técnica proposta não foi validada pela equipe de campo (inspeções), limitando a abrangência do modelo proposto.

Assim como Calili (2005) e Ortega (2008) também usou Redes Neurais como base para agrupar e classificar suspeitos de furto e fraude, com

comportamentos semelhantes. A partir da arquitetura *Multilayer Perceptron* (MPL) e do algoritmo *Levenberg-Marquardt* (LM), o objetivo do trabalho foi desenvolver um *software* de apoio à decisão, chamado de Sistema Inteligente para Identificação de Perdas Comerciais (SIIPERCOM). Dentre as funcionalidades da aplicação, ela deveria ser capaz de: a) fornecer informações para auxiliar no processo de seleção de clientes suspeitos de furto e fraude a serem inspecionados; e b) aumentar a produtividade da recuperação de energia por parte das distribuidoras. Metodologicamente, o trabalho seguiu as seguintes etapas de modelagem e desenvolvimento do sistema de apoio à decisão, obtenção da base de dados com histórico de inspeções, análise e pré-processamento da base de dados e avaliação do desempenho do sistema. Os dados utilizados no trabalho de Ortega (2008) foram cedidos pela empresa *LIGHT*, que, segundo o autor, era a terceira empresa com maior incidência de furto e fraude de energia do Brasil.

Após o desenvolvimento metodológico do trabalho, Ortega (2008) chegou a várias conclusões, destacando que, apesar de não ter realizado nenhuma inspeção de campo, nos testes de validação do *software* proposto, gerou os índices de acerto entre 44,2% e 62,6%, ao passo que, os índices de acerto normais da *Light* estavam entre 29,4% e 46,3%.

Apesar de usar como tema as perdas comerciais, o trabalho de Bernardes (2010) não fez uso de aplicações computadorizadas e nem de procedimentos estatísticos em suas análises. O objetivo do trabalho foi propor um plano estruturado de ação para atenuação de perdas comerciais, por meio de um método que integre as abordagens do Pensamento Sistêmico e Planejamento por Cenários (PSPC) e do Processo de Pensamento da Teoria das Restrições (TOC). Na condução do trabalho, o autor adotou como método de pesquisa a Pesquisa-Ação, sendo que o trabalho foi desenvolvido a partir de informações da empresa CEEE-D.

Após a condução metodológica do trabalho e avaliação dos resultados obtidos, o autor chegou às seguintes conclusões: a) o trabalho, além de ter propiciado a criação de um plano estruturado de ação, permitiu que a equipe visualizasse a questão das perdas de forma diferente da adotada até então; b) a riqueza de detalhes com que as perdas não técnicas foram discutidas pelo método PSPC-TOC proposto, gerou um conhecimento aprofundado sobre o tema, possibilitando maior sucesso nas ações de combate e prevenção do furto e fraude; c) permitiu a análise das ações estratégico/táticas estabelecidas ao contexto das perdas não técnicas, com a verificação da repercussão sistêmica de cada ação no mapa representativo da situação; e d) o plano de ação proposto incorporou, além de novas práticas, as já conhecidas e propostas em planos anteriores.

Um dos trabalhos mais recentes que trata sobre perdas comerciais é o de Ferreira et al. (2012). O objetivo do trabalho foi apresentar a proposta de um *software* de Georreferenciamento de Perdas Comerciais (GeoPerdas), que está sendo desenvolvido para disponibilizar informações georreferenciadas das perdas comerciais no Estado de Minas Gerais. Esse *software* auxiliará na prevenção e no combate ao furto e/ou fraude de energia elétrica por, segundo os autores, auxiliar as equipes de campo da distribuidora na rápida constatação e combate às irregularidades, diminuindo as ações de fraude por parte dos clientes.

O Item 2 apresentou trabalhos, nacionais e internacionais, com propostas que objetivavam solucionar o problema das perdas comerciais. De modo geral, pode-se afirmar que, dentre os trabalhos, não há unanimidade entre o método, o instrumento ou a técnica utilizada para combater as perdas comerciais. Mas, por outro lado, é possível afirmar que, a grande maioria, considera o uso de sistemas de informação computadorizado como elemento principal para gerar informações relevantes para tomada de decisão. Ou seja, a geração de

informação sobre as perdas, de um modo geral, foi um dos principais recursos utilizados pelos autores no combate às perdas não técnicas.

Considerando que a informação gerada a partir de sistemas de informação apropriados, realmente pode ser um dos elementos-chave na seleção e inspeção dos suspeitos de fraudar o consumo de eletricidade, o objetivo deste item é apresentar os tipos e as características de sistemas de informação capazes de auxiliar no combate às perdas comerciais.

### 3 SISTEMA DE INFORMAÇÃO

O objetivo desse item é definir sistema de informação e apresentar o tipo de sistema que será usado neste trabalho. No subitem a seguir, faz-se uma abordagem sobre o conceito, as dimensões e os tipos de sistemas de informação aplicados a organizações e, na sequência, apresenta as características e o tipo do sistema de informação que melhor pode auxiliar na solução do problema das perdas não técnicas.

#### 3.1 Sistemas de informação: conceito, dimensões e tipos

Sistema de informação baseado em computador, ou simplesmente sistema de informação (SI), pode ser definido como um conjunto de recursos (humanos, computacionais e organizacionais) que se inter-relacionam para coletar, armazenar, recuperar, processar e distribuir informações com o objetivo de apoiar processos gerenciais (planejamento, controle, comunicação e tomada de decisão) em organizações (CHECKLAND; HOLWELL, 1998; LAUDON; LAUDON, 2010; TURBAN; McLEAN; WETHERBE, 2004).

Para Laudon e Laudon (2010), o uso eficiente dos sistemas de informação depende do entendimento das três dimensões que os compõem, que são: a organizacional, a humana e a tecnológica, como representadas no Quadro 2.

Considerando as dimensões representadas no Quadro 2, um SI dificilmente será composto por apenas um computador pessoal e um *software*, como sugerido por Turban, McLean e Wetherbe (2004). Pois, a falta de um gerenciamento efetivo dos componentes dessas dimensões, poderá resultar em um desequilíbrio entre elas, o que pode comprometer o objetivo principal para o qual o SI foi desenvolvido, provocando falhas (LYYTINEN, 1988).

Quadro 2 Dimensões dos sistemas de informação computacional e suas características

FORMAS	CARACTERÍSTICA
Organizacional	a) são compostas por processos de negócios, que são as regras (instruções) que orientam as pessoas nos diversos procedimentos; b) integram os SI, ou seja, os sistemas de informação fazem parte das organizações; c) influenciam o uso de SI a partir da estrutura, história e cultura da organização.
Humana	a) são os indivíduos que trabalham com SI ou utilizam os resultados gerados por eles em uma organização; b) devem ser qualificadas para desenvolver e manter SI; c) para usar as informações e atingir os objetivos organizacionais, as pessoas devem ser treinadas.
Tecnológica	a) um dos muitos recursos usados para enfrentar as mudanças organizacionais; b) plataforma (estrutura) sobre a qual a organização pode montar seus SI; c) <i>hardware</i> : equipamentos físicos usados nas atividades de entrada, processamento e saída de um SI (computadores, impressora, processadores, monitores, teclados, dispositivos de leitura externo, etc.); d) <i>software</i> : conjunto de programas (instruções detalhadas e pré-programadas) que controlam e coordenam os componentes do <i>hardware</i> ; e) banco de dados: <i>softwares</i> que comandam a organização de dados em meios físicos de armazenamento; f) redes: dispositivos físicos e <i>softwares</i> que interligam diversos equipamentos de computação e transfere dados de uma localização física para outra.

Fonte: Organizado pelo autor a partir de Laudon e Laudon (2010) e O'Brien (2011)

Para Laudon e Laudon (2010), a dimensão organizacional relaciona-se à autoridade e à responsabilidade da estrutura hierárquica de uma organização, ou seja, como a organização executa e coordena o trabalho por meio da hierarquia e de seus processos de negócios. Ao passo que a humana está relacionada principalmente ao trabalho criativo impulsionado por novos conhecimentos e informações. Por fim, a dimensão tecnológica relaciona-se ao armazenamento de

dados e à comunicação de redes, formada essencialmente pela interação entre *hardware* e *software*.

Ainda segundo os autores, as três dimensões apresentadas constituem a infraestrutura de tecnologia da informação, fornecendo um conjunto de serviços tecnológicos necessários para montar seus sistemas de informação específicos que, de acordo com os diferentes objetivos, interesses, especializações e níveis dentro de uma organização, podem culminar em diferentes tipos de SI, como representado no Quadro 3.

Considerando os tipos representados no Quadro 3, pode-se afirmar que, apesar de diferentes, os três tipos de SI representados se complementam. Em uma organização, dificilmente será utilizado um DSS sem que um IRS esteja em pleno funcionamento. Da mesma forma, o uso de um GSS somente será efetivo se as informações a serem trocadas e compartilhadas forem relevantes para o conhecimento do grupo.

Outro ponto que vale destacar é que, apesar de serem complementares, as dimensões dos SI são mais ou menos relevantes em cada tipo de sistema de informação. Apesar de a dimensão tecnológica ser comum aos três tipos de SI, essa se mostra mais relevante nos GSS pelas próprias características das aplicações que são enquadradas nesse tipo de SI. Mesmo considerando que os três tipos de SI são utilizados dentro de uma organização, a dimensão organizacional apresenta-se com maior relevância nos IRS, principalmente pelo estreito vínculo que possui com as atividades operacionais e repetitivas do dia a dia da empresa. Ou seja, funções que se forem padronizadas (em processos organizacionais) e automatizadas (usando elementos da dimensão tecnológica), a dimensão humana poderia quase ser dispensada desse tipo. Situação que dificilmente ocorreria com os DSS, pois esse tipo de sistema depende da interação das pessoas com a tecnologia para, de forma criativa gerar informações que poderão auxiliar na solução de problemas pouco estruturados.

Quadro 3 Tipos de sistemas de informação computacional e suas características

TIPOS	CARACTERÍSTICAS
<p style="text-align: center;"><i>Information Reporting System (IRS)</i> ou Sistemas de Processamento de Transações (STPs)</p>	<ul style="list-style-type: none"> <li>a) monitora e controla transações e atividades operacionais internas da organização (vendas, recebimentos, entradas de dinheiro, folhas de pagamento, decisões de crédito etc.);</li> <li>b) o objetivo principal é gerar, com regularidade, relatórios pré-formatados com informações (dados) sintetizadas, facilitando o controle operacional e as decisões rotineiras em uma organização;</li> <li>c) utilizado em nível operacional, onde tarefas, recursos e metas são predefinidos e altamente estruturados;</li> <li>d) importante fonte de informação para outros tipos de sistema;</li> <li>e) alta dependência das dimensões organizacional e tecnológica;</li> <li>f) como exemplos, pode-se citar a módulos e relatórios gerados por <i>Enterprise Resource Planning (ERP)</i> e <i>Customer Relationship Management (CRM)</i>.</li> </ul>
<p style="text-align: center;"><i>Decision Support System (DSS)</i> ou Sistemas de Apoio à Decisão (SADs)</p>	<ul style="list-style-type: none"> <li>a) usa os dados coletados a partir de IRS (ou STPs) e até mesmo dados externos à organização para gerar informações que auxiliarão em decisões da gerência média e sênior da empresa;</li> <li>b) o objetivo principal é gerar informações que facilitem decisões pouco estruturadas, indo além do uso “reativo” da informação;</li> <li>c) possibilita que os usuários façam análises exploratórias de forma criativa e interativa;</li> <li>d) alta dependência das dimensões humana e tecnológica;</li> <li>e) como exemplos, pode-se citar a sistemas baseados em modelos (<i>model-based systems</i>) que usam desde simples planilhas eletrônicas até avançadas técnicas de modelagem e sistemas baseados em dados (usando <i>data warehouse</i> e ferramentas como <i>Online Analytical Processing – OLAP –</i> e <i>data maning</i>) e sistema de inteligência de negócio.</li> </ul>
<p style="text-align: center;"><i>Group Support System (GSS)</i> ou Sistema de colaboração e trabalho em equipe</p>	<ul style="list-style-type: none"> <li>a) usa tecnologia para facilitar a comunicação e colaboração entre membros de um grupo, por meio do compartilhamento e troca de informação e conhecimento;</li> <li>b) o objetivo principal é fornecer meios pelos quais funcionários de uma organização possam se comunicar e trabalhar com outros funcionários em equipes virtuais (que podem estar a quilômetros de distância);</li> <li>c) pode ser utilizado em diversos níveis da organização, gerando como benefícios produtividade, qualidade, inovação, atendimento aos clientes e desempenho financeiro;</li> <li>d) alta dependência das dimensões tecnológica e humana;</li> </ul>

“continua”

Quadro 3 “conclusão”

TIPOS	CARACTERÍSTICAS
<i>Group Support System (GSS)</i> ou Sistema de colaboração e trabalho em equipe	e) como exemplos, pode-se citar a e-mail e mensagens instantâneas, <i>wikis</i> (tipo de site para armazenamento e compartilhamento de conhecimento), redes sociais, repositórios de conhecimento, conferência por áudio e vídeo, entre outros.

Fonte: Organizado pelo autor a partir de Laudon e Laudon (2010) e Sun e Teng (2012)<sup>1</sup>

Considerando a interação das dimensões representadas no Quadro 2, com os tipos de SI representados no Quadro 3, pode-se afirmar que o tipo de SI mais recomendado para gerar informações que possam auxiliar na solução das perdas não técnicas é o DSS, cuja descrição mais detalhada é apresentada no subitem 3.3, *Decision support system* ou Sistema de apoio à decisão (DSS).

### 3.3 *Decision support system* ou Sistema de apoio à decisão (DSS)

#### 3.3.1 Introdução ao DSS

De acordo com Arnott e Pervan (2008), há mais de 35 anos que pesquisadores têm trabalhado com os sistemas de apoio à decisão, relatando e provando sua utilidade no auxílio à solução de problemas semiestruturados ou não estruturados. Segundo Fleury, Agostini e Lima (1984), os primeiros conceitos relacionados aos Sistemas de Apoio à Decisão (DSS) surgiram no início da década de 70 com o trabalho de Scott Morton (1971) e uma das

---

<sup>1</sup> Nos dos tipos de SI apresentados, os autores se fundamentaram em trabalhos como o de Anthony (1965), Galbraith (1973, 1995), McAfee (2006), McNurlin, Sprague e Bui (2009) e Teng e Calhoun (1996)

primeiras definições do modelo teórico de um DSS foi publicada por Keen e Scott Morton (1978).

### 3.3.2 Conceito de DSS

De acordo com Er (1988), apesar do uso aleatório do termo *Decision Support System* (DSS), as palavras que o formam devem ter a seguinte análise:

- a) *Decision* (decisão): implica na solução de problema, pois muitas vezes para se solucionar problemas é necessário tomar decisões. Assim, a solução de problema implica no uso de conhecimento, que muitas vezes estão em modelos incorporados nos DSS;
- b) *Support* (suporte ou apoio): implica no uso de tecnologia (*hardware* e *software*) que auxiliarão o processo de tomada de decisão;
- c) *System* (sistema): implica em um modelo de interação entre homem e máquina, desde o desenvolvimento até a implementação.

Considerando a análise do termo proposta por Er (1988), outros autores também seguiram essa mesma lógica para conceituar o DSS. Finlay (1994), por exemplo, faz parte de um grupo de pesquisadores que definem o DSS como um sistema computacional que auxilia o processo de tomada de decisão. As definições propostas por Dong e Srinivasan (2012) e Turban (1995) são mais abrangentes. Para os autores, os DSS são sistemas de informação interativos, flexíveis e adaptáveis, que auxiliam os tomadores de decisão por meio do uso de dados, modelos, experiências anteriores, visualizações e interface com usuário, a resolver problemas, pouco ou não estruturados, e a aperfeiçoar a decisão tomada.

Arnott e Pervan (2008) comentam que a ideia de que os computadores podem ser usados para apoiar, e não substituir os seres humanos em processos

de tomada de decisão é tão importante hoje, quanto era na década de 1970. E é a partir desse conceito que os autores afirmam que o objetivo principal de um DSS é fornecer subsídios para melhorar as decisões que os usuários desses tipos de sistema vão tomar, resultando em decisões “mais bem informadas”.

Nesse mesmo sentido, Fleury, Agostini e Lima (1984) argumentam que as aplicações DSS procuram combinar aspectos formais das decisões, passíveis de representação em um programa de computador, com os processos heurísticos e intuitivos da mente do tomador de decisão. Ou seja, nesses tipos de aplicação o computador pode desempenhar um relevante papel na tomada de decisões semiestruturadas (que não podem ser totalmente automatizadas), requerendo uma combinação de algoritmos e modelos formais, com a intuição e visão subjetiva do tomador de decisão. Com isso, para os autores, a preocupação fundamental dos DSS deve ser no aumento a eficácia dos processos decisórios (melhoraria da qualidade das decisões) e não simplesmente melhorar a eficiência dos procedimentos operacionais da empresa.

Keen e Gambino (1980) vão mais além. Para os autores, um DSS pode ser considerado como uma estratégia de desenvolvimento adaptativa, aplicada à situação onde o sistema “final” não pode ser pré-definido, mas deve evoluir por meio da interação entre o usuário, o sistema e o desenvolvedor do DSS.

### **3.3.3 Tipos ou subgrupos de DSS**

A necessidade constante de interação usuário, sistema e desenvolvedor faz com que as aplicações DSS se diversifiquem. Motivo pelo qual, ao longo dos mais de 35 anos de história dessa disciplina, surgiram vários subgrupos de pesquisa e prática, dos quais se destacam (ARNOTT; PERVAN, 2005, 2008):

- a) Sistema de apoio à decisão pessoal (*Personal Decision Support Systems – PDSS*): caracterizam-se por serem, geralmente, sistemas de pequena escala, desenvolvido por um gestor ou um pequeno número de gestores independentes, para apoiar uma tarefa de decisão.
- b) Sistema de apoio a grupos (*Group Support Systems – GSS*): usam uma combinação de tecnologias DSS e comunicação para facilitar o trabalho efetivo de grupos.
- c) Sistema de apoio à negociação (*Negotiation Support Systems – NSS*): é um DSS onde o foco principal do grupo de trabalho é a negociação entre partes opostas.
- d) Sistemas Inteligentes de Apoio à Decisão (*Intelligent Decision Support Systems – IDSS*): baseiam-se na aplicação de técnicas de inteligência artificial para apoiar a decisão.
- e) Sistemas de apoio à decisão baseados na gestão do conhecimento (*Knowledge Management-Based DSS – KMDSS*): sistemas que auxiliam a tomada de decisão ao armazenar, recuperar, transferir e aplicar conhecimentos a partir da memória individual e organizacional de um grupo, permitindo que membros do grupo tenham acesso a esse conhecimento.
- f) *Data Warehousing (DW)*: sistemas que fornecem uma ampla escala de dados para apoiarem a decisão.
- g) *Enterprise Reporting and Analysis Systems*: sistemas de apoio à decisão que tem como foco corporações, incluindo sistema executivos (*Executive Information Systems – EIS*), *Business Intelligence (BI)*, e mais recentemente, *Corporate Performance Management Systems (CPM)*.

### 3.3.4 Características de DSS

Apesar da diversidade dos subgrupos de subgrupos de pesquisa e prática, os DSS devem possuir características elementares, tais como (FLEURY; AGOSTINI; LIMA, 1984; LAUDON; LAUDON, 2010):

- a) auxílio na tomada de a decisão em si, normalmente enfocada nos níveis organizacionais mais altos;
- b) foco em problemas únicos, que se alteram com rapidez e que não possuem procedimentos de resolução totalmente predefinido;
- c) fácil interação, ou seja, são projetados de modo que os usuários consigam trabalhar diretamente com eles;
- d) podem recorrer a fontes de informações externas;
- e) uso de dados obtidos em outros banco de dados, de linguagem de alto nível, de processamento gráfico e de modelos matemáticos;
- f) usam uma série de modelos para analisar dados ou condensar grande quantidades de dados em formatos que possam ser analisados pelos tomadores de decisão; e
- g) enfatiza a flexibilidade, adaptabilidade e resposta rápida, sendo os sistemas iniciados e controlados pelo próprio usuário, adequando-se ao estilo pessoal de tomada de decisão dos mesmos.

Para que um DSS apresente as características apontadas acima, o desenvolvimento desse tipo de sistemas de informação deve passar pelas etapas descritas no subitem a seguir.

### 3.3.5 Etapas do desenvolvimento de um DSS

Não há um consenso sobre a melhor forma de desenvolvimento de um sistema de apoio à decisão. Avaliando vários trabalhos publicados, desde Keen e Scott Morton (1978), passando por Finlay (1994) até Dong e Srinivasan (2012), pode-se afirmar que diferentes propostas de desenvolvimento de DSS são encontradas.

Entretanto, pode-se afirmar que fases como a de análise, projeto, desenvolvimento e implantação, como apresentado por Fleury, Agostini e Lima (1984), são comuns às maiorias das propostas de desenvolvimento. Para os autores, considerando as características elementares de um DSS, a abordagem tradicional do desenvolvimento de sistemas torna-se pouco adequada para esse tipo de SI. A justificativa, segundo os autores, está no fato de que essas fases, não raramente, são combinadas em um único passo, o que acaba sendo repetido interativamente. Assim, para facilitar o desenvolvimento de um DSS por Fleury, Agostini e Lima (1984) apresentam o ciclo de vida do DSS, que é demonstrado na Figura 1.

De acordo com Fleury, Agostini e Lima (1984), a ideia é desenvolver o mais cedo possível, na fase de projeto, um modelo simplificado do DSS e torná-lo disponível ao usuário para que ele possa experimentá-lo, possibilitando a obtenção de resultados imediatos.

Após certo período de tempo, o DSS deve ser avaliado, modificado e expandido, enfatizando a interação usuário-sistema-desenvolvedor do DSS. Esse ciclo deve ser repetido, efetivando o processo adaptativo e evolutivo do desenvolvimento do DSS. Dessa forma, o DSS ajuda o tomador de decisão na conceituação do problema, não esperando que o mesmo seja capaz de descrever o processo de decisão antes que o sistema esteja concluído (FLEURY; AGOSTINI; LIMA, 1984).

É importante destacar que, em decorrência dessa abordagem, o desenvolvimento do DSS torna-se dependente do usuário. E, para Fleury, Agostini e Lima (1984), apenas um usuário imaginativo, e envolvido no processo, permite e efetiva evolução do DSS.

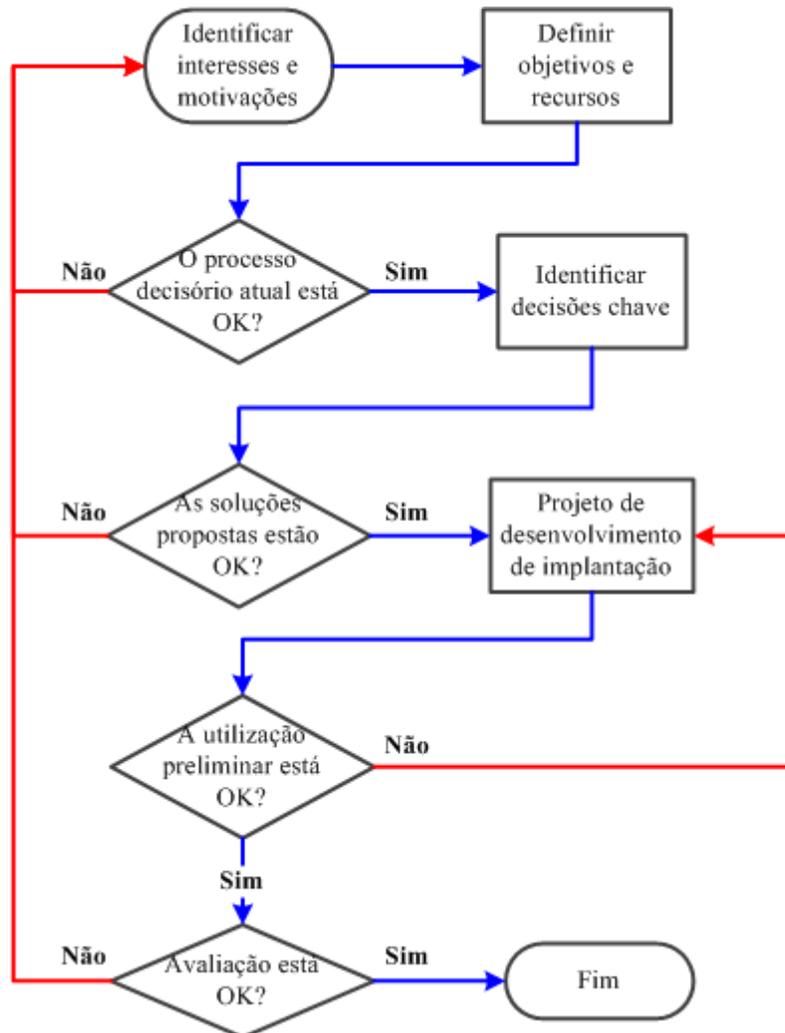


Figura 1 Ciclo de vida utilizado para o desenvolvimento de um sistema de apoio à decisão

Fonte: Adaptado de Fleury, Agostini e Lima (1984)

## **4 MATERIAL E MÉTODOS**

### **4.1 Objeto de estudo**

Este trabalho tem como objeto de estudo uma das 64 distribuidoras de energia que atuam no Brasil. Essa empresa faz parte de um grupo empresarial que é reconhecido pela sua dimensão e competência técnica, sendo considerado o maior grupo integrado do setor de energia elétrica (geração, transmissão e distribuição) do Brasil, constituído por 58 empresas e 10 consórcios. Para simplificar a referência à empresa, neste trabalho ela será tratada como Distribuidora de Energia Elétrica (DEE).

A principal justificativa para a escolha da DEE está no fato de que, além de atuar internacionalmente, em 19 Estados brasileiros e no Distrito Federal, ela se dispôs, mediante contrato de confidencialidade de dados e de pesquisa, a disponibilizar as informações necessárias para o desenvolvimento do trabalho. A DEE solicitou o sigilo por considerar que as informações geradas pelo trabalho são estratégicas para o combate às perdas não técnicas de energia.

Apesar da ampla atuação na maioria dos Estados brasileiros, apenas um Estado foi escolhido como amostra. De acordo com a DEE, a escolha dessa amostra para desenvolver o trabalho se fundamenta na amplitude territorial do Estado, na diversidade e quantidade de ocorrências de perdas não técnicas e na significativa atuação da DEE, que responde por 96% da área de concessão, com cerca de 6,8 milhões de consumidores regulares, em 774 municípios.

### **4.2 Classificação e método científico**

A pesquisa científica usa o processo formal, pragmático e sistemático do método científico para investigar e descobrir respostas para problemas (GIL,

1999; SILVA; MENEZES, 2005). Ou seja, o método científico é a base indispensável para se realizar pesquisas científicas.

Em organizações, a maioria dos problemas é representada por processos operacionais que podem ser complexos e de difícil representação científica, pois, em muitos casos, o desempenho de um processo operacional (geralmente mensurado pela qualidade, eficiência, custo, flexibilidade, entre outros) pode ser afetado por diferentes elementos, componentes e agentes (BERTRAND; FRASOO, 2002).

Considerando que este trabalho tem como objetivo principal auxiliar na solução de um problema que está diretamente relacionado a processos operacionais de uma organização, faz-se necessário classificá-lo metodologicamente, para que possa ser considerada uma pesquisa científica. A Figura 2 representa as diferentes formas de classificação de uma pesquisa científica.

A partir das classificações representadas na Figura 2, este estudo pode ser considerado como uma pesquisa científica, com as seguintes classificações:

- a) pesquisa aplicada, quanto à natureza – gera conhecimentos para aplicação prática, direcionados à solução de problemas específicos;
- b) pesquisa normativa e descritiva, quanto aos objetivos – pois visa desenvolver políticas, estratégias e ações para melhorar os resultados disponíveis em uma situação existente, encontrar uma solução ótima para um problema definido, ou comparar várias estratégias direcionadas a um problema específico. Além disso, uma das propostas do trabalho é analisar uma situação, que leve à compreensão e explicação de suas características;
- c) pesquisa qualitativa, quanto à forma de abordar o problema – por considerar que há uma relação dinâmica entre o mundo real e o

sujeito, onde a interpretação dos fenômenos e a atribuição dos significados são inerentes à pesquisa. Além disso, o trabalho fará uma abordagem quantitativa na análise dos dados a serem utilizados; e

- d) pesquisa-ação, quanto aos procedimentos técnicos – o procedimento aplica-se em casos em que a pesquisa é concebida e realizada em estreita associação a uma ação, ou a resolução de um problema coletivo ou específico.

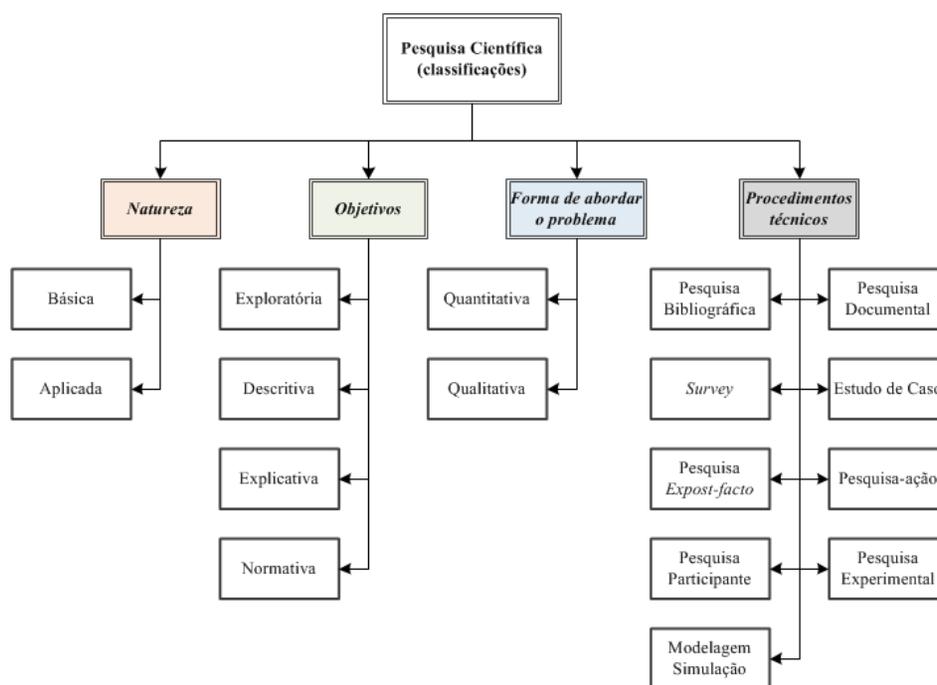


Figura 2 Formas de classificações de pesquisas científicas

Fonte: Organizado pelo autor a partir de Bertrand e Frasoo (2002) e Silva e Menezes (2005)

Considerando especificamente o procedimento técnico usado neste trabalho, a pesquisa-ação é a mais adequada ao seu desenvolvimento, pela necessidade de interação (FLEURY; AGOSTINI; LIMA, 1984) entre o

pesquisador e os agentes do objeto de pesquisa, para gerar conhecimento e auxiliar na solução do problema das perdas não técnicas. O subitem a seguir apresenta uma abordagem mais detalhada do procedimento usado neste trabalho.

### **4.3 Pesquisa-ação**

A pesquisa-ação é um procedimento de pesquisa com base empírica que é concebida e realizada em estreita associação com uma ação ou com a resolução de um problema coletivo, onde pesquisadores e participantes, que representam a situação ou o problema, atuam de forma cooperativa e participativa, gerando conhecimento (BERTRAND; FRASOO, 2002; COUGHLAN; COUGHLAN, 2002; THIOLENT, 2005).

A justificava para escolha desse procedimento de pesquisa se fundamenta na necessidade de gerar conhecimento para solucionar um problema organizacional, a partir do desdobramento de uma série de ações ao longo do tempo, que podem mudar ou melhorar o funcionamento de um sistema (COUGHLAN; BRANNICK, 2009). E, pela característica do problema, isso dificilmente seria possível, considerando os procedimentos positivistas tradicionais de pesquisa (COUGHLAN; COUGHLAN, 2002). Além disso, Westbrook (1995) argumenta que a pesquisa-ação se diferencia de outros procedimentos de pesquisa pelos seguintes motivos:

- a) apresenta grande relevância e aplicabilidade para as pessoas que vivenciam o problema que, muitas vezes não é estruturado;
- b) pode contribuir com a teoria; e
- c) garante, por meio de seu fundamento iterativo e intervencionista, proximidade ao contexto em que o trabalho está sendo desenvolvido, o que possibilita a “utilização” um grande número de

variáveis que não são disponibilizadas (ou se apresentam) de uma única vez.

Segundo Macke (2006), a pesquisa-ação pode ser definida como um procedimento técnico que visa à solução coletiva para de determinada situação ou problema, por meio de um processo de mudança planejada, contemplando, ao mesmo tempo, o processo de pesquisa e de intervenção (ação), os quais são resultantes da participação efetiva entre pesquisadores e participantes envolvidos.

Para Thiollent (2005), o princípio desse método de pesquisa consiste na intervenção, por meio do trabalho conjunto entre pesquisadores e membros da organização, para: a) definição do problema; b) busca de soluções; e c) aprofundamento do conhecimento científico disponível. Motivos pelos quais, para o autor, esse procedimento deve agregar várias técnicas de pesquisa social, para estabelecer uma estrutura coletiva, participativa e ativa ao nível de captação de informação e solução de problemas.

De forma complementar, Coughlan e Coughlan (2002) comentam que a pesquisa-ação é, ao mesmo tempo, uma sequência de eventos e uma abordagem para resolução de problemas. Segundo os autores, como sequência de eventos esse procedimento compreende ciclos iterativos de coleta de dados, gerando *feedback* aos participantes, análise de dados, planejamento de ações e avaliação das ações, que levam a um novo ciclo de coleta de dados, análise, planejamento e avaliação, e assim por diante. Como abordagem de solução de problema, os autores afirmam que o procedimento é uma aplicação do método científico para descobrir fatos e experimentação para solução de problemas práticos, sendo necessárias ações que envolvem a colaboração e cooperação para se obter soluções. Concluindo, os autores argumentam que as soluções encontradas constituem-se em um dos resultados desejados, pois, tão importante quanto elas

é o aprendizado com os resultados intencionais e não intencionais obtidos, a contribuição para a teoria e o conhecimento científico.

Como se pode perceber, para usar a pesquisa-ação como procedimento técnico de pesquisa é necessário que se tenha um problema real de pesquisa, que seja relevante gerencialmente, que apresente um resultado incerto para a organização e que possa ser submetido a uma rigorosa investigação, contemplando a análise e implementação de ações (COUGHLAN; COUGHLAN, 2002). A partir das condições de uso apresentadas, Coughlan e Coughlan (2002) afirmam que a implementação de uma pesquisa-ação compreende três tipos de passo: a) um pré-passo, para entender a contexto e propósito da pesquisa; b) seis passos principais, compreendendo a obtenção, *feedback*, e análise de dados, planejamento, implantação e avaliação de ações; e c) um meta-passo, que corresponde ao monitoramento dos passos anteriores. A Figura 3 ilustra os passos da implantação da pesquisa-ação.

Avaliando os passos ilustrados na Figura 3, Coughlan e Coughlan (2002) argumentam que, para entender o contexto e propósito da pesquisa proposto no pré-passo, é importante responder dois questionamentos.

- a) Qual é a razão para a ação? Ou seja, é necessário saber, por exemplo, por que o projeto é necessário e quais são as forças econômicas, políticas, sociais e técnicas que levam a ação.
- b) Qual é a razão para a pesquisa? Ou seja, é importante saber, por exemplo, por que vale a pena estudar o problema, se a pesquisa-ação realmente é uma metodologia adequada e qual a contribuição gerada para o conhecimento.

Os passos principais do projeto estão relacionados com os dados e consequentemente com a ação. Cada passo deve atender aos seguintes propósitos (COUGHLAN; COUGHLAN, 2002):

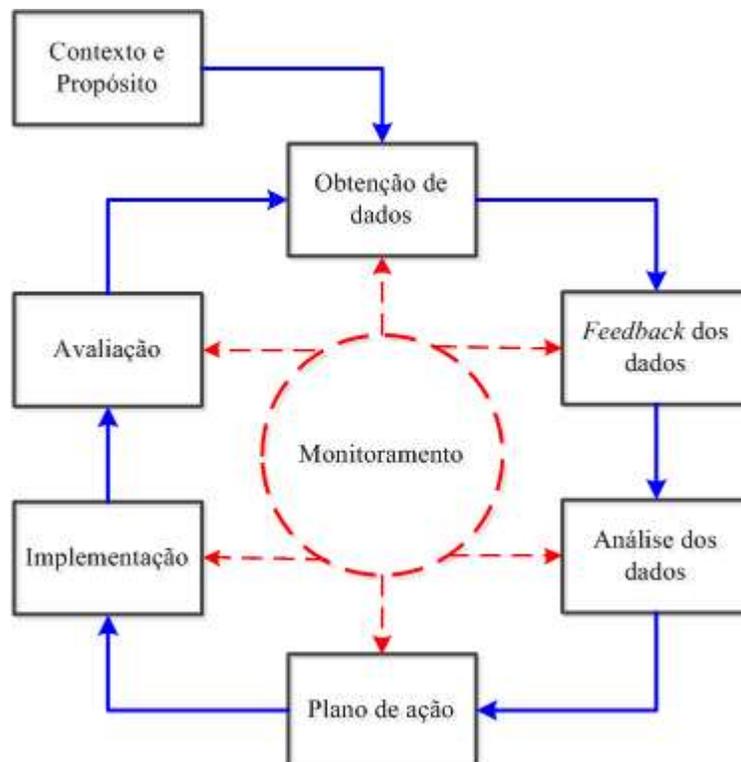


Figura 3 Passos para implantação da pesquisa-ação, como procedimento técnico  
 Fonte: Organizado pelo autor a partir de Coughlan e Coughlan (2002)

- a) Obtenção de dados: relaciona-se à forma de obtenção de dados, como por exemplo, a partir de relatórios operacionais, financeiros ou de *marketing*. Também, podem ser obtidos a partir de observações, discussões ou entrevistas. Ou seja, os dados obtidos nesse passo devem ser o resultado do envolvimento e participação do pesquisador no dia a dia da organização em estudo.

- b) *Feedback* dos dados: os dados obtidos devem ser apresentados e compartilhados entre os pesquisadores e os membros da organização que também participam da pesquisa, minimizando, por exemplo, problemas por falta de dados.
- c) Análise dos dados: é um passo crítico nesse procedimento, pois a análise de dados deve ser feita de forma colaborativa, ou seja, em conjunto entre os pesquisadores e os membros da empresa, partindo do pressuposto que eles conhecem bem a organização, conhecem o trabalho e serão os atores que farão a implantação das ações. Os critérios e ferramentas para análise dos dados devem ser discutidos e orientados para o objetivo da pesquisa e das intervenções.
- d) Plano de ação: com os dados analisados, é necessário criar, também em conjunto (pesquisadores e membros da empresa) um plano para implementar a ação. Nesse passo, alguns questionamentos podem ser feitos, tais como: o que é necessário mudar? Em quais partes da organização? Quais tipos de mudança são necessários? Qual a estrutura é necessária? Qual a resistência deve ser gerenciada?
- e) Implementação: na maioria das vezes, a ação é implementada pelos membros da organização, seguindo os planos estabelecidos no passo anterior.
- f) Avaliação: é a chave para a aprendizagem e envolve reflexões sobre os resultados intencionais e não intencionais da ação e sobre o processo como um todo. Por meio desse passo, os consequentes ciclos de planejamento e ação beneficiarão da experiência do processo com um todo. Sem esse passo, as chances de ineficácia e frustração da equipe podem levar ao fracasso da pesquisa-ação.

Por fim, de acordo com Coughlan e Coughlan (2002), o meta-passo monitoramento destaca-se como o foco da pesquisa acadêmica. Para os autores o monitoramento deve ser realizado em todos os passos e também nos ciclos de planejamento, implantação e ação que forem necessários para solucionar o problema e gerar conhecimento necessário. Enquanto os membros da organização devem se preocupar com os resultados práticos, o pesquisador deve-se preocupar com o andamento do projeto e com o monitoramento do processo de aprendizagem.

Considerando o exposto, pode-se afirmar que a pesquisa-ação deve apresentar as seguintes particularidades (GUMMESSON, 1999):

- a) o pesquisador não é um simples observador. Ele trabalha ativamente com o problema a ser solucionado;
- b) objetiva resolver um problema e, ao mesmo tempo, contribuir com a ciência;
- c) consiste num procedimento interativo, que requer a cooperação entre o pesquisador e os membros da organização em que o trabalho está sendo desenvolvido, necessitando de contínuos ajustes face aos novos eventos e/ou informações;
- d) busca desenvolver uma compreensão holística durante o projeto, reconhecendo a sua complexidade;
- e) relaciona-se, fundamentalmente com a mudança;
- f) requer entendimento do enquadramento ético, dos valores e das normas que são usadas em contexto particular (principalmente entre os pesquisadores e os *stakeholders* da organização);
- g) pode utilizar todos os métodos de coleta de dados (ferramentas quantitativas, qualitativas – como entrevistas e *surveys*);

- h) requer um conhecimento prévio do ambiente corporativo, das condições de negócio, da estrutura e da operacional da organização, além das bases teóricas do problema a ser resolvido;
- i) realiza-se em tempo real e promovendo reflexão e aprendizagem na organização; e
- j) usa critérios de avaliação diferente dos adotados pela ciência positivista. A pesquisa-ação reflete a cooperação entre os pesquisadores e os membros da organização? A pesquisa-ação é guiada para os resultados práticos a partir de uma constante reflexão do processo de melhoria ou mudança organizacional? A pesquisa-ação considera uma pluralidade de saber que garante a integridade teórico-conceitual, por meio do uso adequado da metodologia? O projeto da pesquisa-ação é significativo e trará resultados duradouros para a organização?

Fazendo um paralelo entre o ciclo de vida de um sistema de apoio à decisão, representado na Figura 1 e os passos de implantação de uma pesquisa-ação, representado na Figura 2, pode-se afirmar que, apesar de estarem relacionados a conceitos distintos, as duas estruturas usadas neste trabalho são complementares, além de apresentarem etapas ou passos que possuem o mesmo objetivo e/ou finalidade.

## 5 RESULTADOS E DISCUSSÃO

Este tópico tem o objetivo de apresentar e analisar os resultados encontrados neste trabalho. Para facilitar a análise e a discussão dos resultados encontrados, este item será subdividido em subitens e tópicos, que seguirão os passos de implantação de uma pesquisa-ação, representado na Figura 3. Os resultados deste trabalho são apresentados em quatro ciclos de pesquisa-ação. O Quadro 5 representa os ciclos de pesquisa-ação e uma descrição resumida de cada um deles, que são divididos em subitens.

Quadro 4 Ciclos de pesquisa-ação com suas respectivas descrições, divididos em subitens deste trabalho

<b>Subitem</b>	<b>Nome</b>	<b>Descrição resumida</b>
5.1	Contexto e propósito	Pré-passo da pesquisa-ação que apresenta resposta para duas perguntas: qual é a razão para a ação? E qual é a razão para a pesquisa?
5.2	Ciclo PA-1: POP e mapeamento das liberações	Primeiro ciclo interativo da pesquisa-ação que avaliou o uso do Procedimento Operacional Padrão (POP) de seleção e liberação de inspeções aos consumidores com suspeita de irregularidade nas Unidades Regionais DC e DS da Distribuidora de Energia Elétrica em estudo (DEE).
5.3	Ciclo PA-2: Histórico das inspeções realizadas	Segundo ciclo da pesquisa-ação que avaliou, entre outros dados, o histórico das inspeções realizadas pela DEE.
5.4	Ciclo PA-3: Análise das variáveis das inspeções realizadas	Terceiro ciclo da pesquisa-ação que analisou, com profundidade, as variáveis do histórico das inspeções já realizadas pela DEE.
5.5	Ciclo PA-4: Desenvolvimento da aplicação Perdas DSS	Quarto e último ciclo da pesquisa-ação. Nesse ciclo, desenvolveu-se o sistema de apoio à decisão para alocar de equipes de campo, chamado de Perdas DSS.

Fonte: Organizado pelo autor

## **5.1 Contexto e propósito**

Neste pré-passo do trabalho, é fundamental obter resposta para duas perguntas: qual é a razão para a ação? E qual é a razão para a pesquisa?

Para obter as respostas e cumprir esse passo, foi agendada, junto à Distribuidora de Energia Elétrica (DEE), uma reunião entre o pesquisador e os funcionários selecionados pela DEE para compor a equipe de trabalho.

Nessa reunião, realizada na sede da DEE, o pesquisador apresentou a metodologia (pesquisa-ação) que seria usada no desenvolvimento do trabalho, destacando a importância da interação e da cooperação entre a equipe de trabalho (colaboradores da DEE e o pesquisador).

Para facilitar o desenvolvimento da pesquisa-ação, definiu-se que três funcionários da DEE estariam envolvidos diretamente com a pesquisa-ação, sendo que um seria o coordenador do projeto na distribuidora (coordenador/DEE) e outros dois colaboradores (colaborador A/DEE e colaborador B/DEE) que auxiliariam o pesquisador no que fosse necessário para o desenvolvimento do trabalho. Na sequência, o pesquisador realizou uma entrevista não estruturada com o coordenador/DEE e com os colaboradores A/DEE e B/DEE, que são descritas nos subitens a seguir.

### **5.1.1 Razão para a ação**

Na parte inicial da entrevista, o coordenador/DEE fez uma ampla explanação sobre o problema das perdas não técnicas. De acordo com o ele, as perdas comerciais de energia elétrica geravam prejuízos da ordem de R\$6 bilhões por ano no Brasil. O volume de energia referente a esse prejuízo seria suficiente para abastecer, ininterruptamente, o consumo de média e baixa potência do Estado de Minas Gerais por um ano. O coordenador informou ainda

que, o valor referente às perdas comerciais poderia ser facilmente calculado, subtraindo do volume total de energia destinado à distribuição a soma entre o volume total de energia faturado (vendido aos consumidores) e o volume total de perdas técnicas (que pode ser mensurado matematicamente). Essa diferença, de acordo com o coordenador/DEE, representava prejuízo, tanto para a distribuidora quanto para os consumidores de energia elétrica, pois parte dessas perdas eram revertidas em prejuízo financeiro para distribuidora e outra parte era transferida para o preço da energia elétrica, que é pago pelo consumidor.

O coordenador/DEE comentou que, apesar da perda comercial média da DEE ser menor do que a perda média encontrada em outras distribuidoras nacionais, a estimava era que cerca de 5% dos quase 6,5 milhões de consumidores regulares (com medidor ligado) da DEE fraudavam o consumo de energia. Apesar dessa estimativa, o banco de dados com suspeitos de fraudarem o consumo de energia da DEE se aproximava de 25% do total de consumidores regulares.

Para o coordenador/DEE, esse era um dos principais problemas do departamento de proteção da receita da DEE, pois não era possível inspecionar todos os suspeitos de irregularidade no consumo de energia, em função da disponibilidade de equipe de eletricitas.

Segundo o colaborador A/DEE, as equipes de eletricitas disponíveis na distribuidora eram insuficientes para executar todas as inspeções, pois, além dessa atividade, os eletricitas também respondiam por outros serviços prestados pela DEE (como os serviços de urgência, manutenções preventivas e corretivas, entre outros). O coordenador/DEE complementou dizendo que, mesmo se a DEE conseguisse auditar todos os consumidores suspeitos e, na grande maioria houvesse fraude no consumo de energia, a empresa poderia ter outros problemas internos, em consequência dos trâmites legais de cobrança e recuperação do

volume de energia fraudada (resultado do pequeno efetivo de pessoal na área gerencial).

Diante do exposto pelos participantes da pesquisa-ação, nessa primeira parte da entrevista, foi possível identificar que os principais motivadores para a ação do trabalho foram: a) o prejuízo econômico e financeiro que as perdas não técnicas geram para a distribuidora e para os consumidores de energia em geral; e b) a dificuldade que a distribuidora possui em combater as perdas não técnicas, considerando ampla quantidade de suspeitos de consumo irregular de energia e a limitação de pessoal, técnico e gerencial da DEE.

### **5.1.2 Razão para a pesquisa**

Na tentativa de compreender melhor o problema, o pesquisador questionou aos colaboradores da DEE quem eram as pessoas responsáveis por indicar ou selecionar, quais seriam os suspeitos de irregularidade que passariam por auditoria no consumo de energia.

De acordo com o coordenador/DEE, a distribuidora possuía uma estrutura matricial e, no caso da recuperação de perdas não técnicas, dois setores estavam diretamente envolvidos. O primeiro, com apelo funcional e centralizado na sede da empresa era o responsável por decisões estratégicas e financeiras. O segundo, operacional e descentralizado, ficava responsável pela execução das estratégias e dos serviços a campo.

A estrutura operacional, de acordo com o coordenador/DEE, era dividida em sete unidades distribuídas por regiões específicas do Estado em que a distribuidora atua (objeto de estudo). Essas unidades receberam as seguintes denominações: a) Unidade Regional DC; b) Unidade Regional DL; c) Unidade Regional DM; d) Unidade Regional DS; e) Unidade Regional DN; f) Unidade Regional DO; e g) Unidade Regional DT.

Ainda segundo o coordenador, cada unidade era responsável pelo cumprimento das metas pactuadas com a unidade funcional, pela seleção das unidades consumidoras que deveriam ser inspecionadas, por estabelecer como as inspeções deveriam ser executadas, além de executar os demais serviços das unidades regionais.

De acordo com o colaborador A/DEE, as inspeções dos suspeitos de irregularidade, nesse contexto, eram prejudicadas por dois motivos principais: a) o número insuficiente de equipes de eletricitas disponíveis para realizar as inspeções; e b) os outros serviços de campo, executados pelas mesmas equipes de eletricitas, tinham mais prioridade do que a inspeção dos suspeitos de irregularidade (ou seja, na lista de prioridades, o serviço de inspeção de suspeitos de que era o tinha a menor prioridade).

Para o coordenador/DEE, considerando os motivos apresentados pelo colaborador A/DEE, e por outros mais, havia uma proposta, ainda não implantada, de que da unidade funcional também fosse responsável pelo processo de inspeção dos suspeitos de irregularidade. Com a proposta, a unidade funcional deveria centralizar a estratégia de inspeção e também de distribuição das ordens de serviço de inspeção às equipes de eletricitas.

Após a explanação dos colaboradores da DEE, o pesquisador perguntou como os responsáveis pelas unidades regionais selecionavam os suspeitos que deveriam ser inspecionados pelas equipes de eletricista (equipes de campo), ou seja, quais eram as regras usadas para determinar quais unidades consumidoras deveriam ser inspecionadas?

Segundo os colaboradores A/DEE e B/DEE, as unidades regionais deveriam usar um Procedimento Operacional Padrão (POP) que orientava a atividade de seleção das unidades consumidoras e liberação das ordens de inspeção (ordens de serviço de campo) das que deveriam ser inspecionadas. Assim, de acordo com os colaboradores, a seleção deveria seguir, dentre outras,

três características principais: a) pelos motivos que geraram a suspeita de irregularidade; b) pelo potencial de incremento e recuperação de faturamento; e c) pelo ramo de atividade do suspeito.

De acordo com o coordenador/DEE, os motivos são as razões pelas quais os consumidores passaram a fazer parte do banco de suspeitos. Continuando, ele afirmou que os motivos poderiam ser gerados internamente, na distribuidora, ou terem origens externas. Uma queda maior do que 20% no consumo médio de energia, por mais de três meses, por exemplo, seria um motivo interno que torna um consumidor suspeito de irregularidade. Um exemplo de motivo externo seria a denúncia de fraude feita, por exemplo, por um vizinho do consumidor. Assim, dentre os vários motivos registrados, os responsáveis pelas unidades regionais devem priorizar os que possuem maior chance efetiva de irregularidade, ou seja, na denúncia de um vizinho pode haver mais chance de se encontrar fraude em uma unidade suspeita do que simplesmente a queda no consumo de energia.

Para o colaborador B/DEE, o potencial de incremento e recuperação do faturamento era importante porque, ao se detectar a fraude no consumo, duas situações poderiam ocorrer: a) o consumidor que fraudava o consumo (consumo menor do que o real), após a inspeção passaria a ter um consumo maior, com a eliminação da fraude (incremento no faturamento); b) constatada a fraude, a DEE faria uma cobrança retroativa, referente ao volume de energia consumido indevidamente (recuperação do faturamento).

Por fim, o ramo de atividade do consumidor também era importante na detecção de fraude porque, segundo o colaborador A/DEE, os consumidores da distribuidora eram classificados de acordo com a finalidade do uso de energia. Assim, consumidores comerciais, por exemplo, principalmente aqueles com utilização intensamente de energia (como açougues e padarias), apresentariam

maior potencial de recuperação e incremento no faturamento, do que consumidores residenciais.

A partir das explicações apresentadas, o pesquisador solicitou o POP usado pelas unidades regionais na atividade de seleção e liberação das inspeções a serem realizadas. Além disso, também foi agendada uma visita à Unidade Regional DC e à Unidade Regional DS, para verificar se o POP estava sendo seguido. Nessa visita, também deveria ser levantado quais eram os critérios que as duas unidades regionais estavam usando para selecionar e liberar as inspeções dos consumidores com suspeita de irregularidade.

Diante do exposto pelos participantes da pesquisa-ação, na segunda parte da entrevista, foi possível identificar que os principais motivadores para a pesquisa foram: a) a necessidade de estudar um problema com repercussão financeira e social; b) a adequação do método de pesquisa adotado, pois a cada interação entre os membros participantes da pesquisa, novos resultados poderiam ser gerados; e c) a possibilidade de contribuir com a geração de conhecimento, a partir da solução de um problema que é comum a todas as 64 distribuidoras de energia que atuam no Brasil.

As visitas agendadas deram início ao primeiro ciclo interativo da pesquisa-ação (coleta de dados, *feedback* aos participantes, análise de dados, planejamento de ações e avaliação das ações), que será descrita no subitem a seguir.

## **5.2 Ciclo PA-1: POP e mapeamento das liberações**

O Ciclo PA-1 corresponde ao primeiro ciclo interativo completo que analisou, por meio de coleta de dados e entrevista, o uso do POP de seleção e liberação de inspeções aos consumidores com suspeita de irregularidade nas Unidades Regionais DC e DS da DEE. A descrição desse ciclo, que teve a

duração total de 3 meses, seguiu os passos da implantação da pesquisa-ação representado na Figura 3. Os subitens a seguir apresentam os resultados das entrevistas.

### **5.2.1 Obtenção de dados do Ciclo PA-1**

Os dados usados no Ciclo PA-1 foram obtidos a partir das seguintes fontes:

- a) POP usado na atividade de seleção e liberação de inspeção dos consumidores suspeitos de irregularidade: obtido diretamente na seda da DEE com o auxílio do colaborador B/DEE. O POP era um documento operacional de uso interno na distribuidora;
- b) fluxo da atividade de seleção e liberação de inspeção dos consumidores suspeitos de irregularidade nas Unidades Regionais DC e DS: o fluxo do processo foi obtido mediante entrevista não estruturada com os colaboradores da DEE, responsáveis pela seleção e liberação de inspeção em cada unidade visitada.

### **5.2.2 Feedback de dados do Ciclo PA-1**

Antes de realizar a visita nas unidades regionais o pesquisador obteve juntamente com o colaborador B/DEE o POP usado na atividade de seleção e liberação de inspeção dos consumidores suspeitos de irregularidade (ANEXO A).

Analisando o POP, o pesquisador encontrou dificuldade de interpretar a sequência de atividades que deveriam ser executadas. Conversando com o colaborador B/DEE, eles optaram por criar um fluxograma que fosse capaz de

representar o documento. Para o pesquisador, o uso do fluxograma poderia facilitar a análise do processo nas unidades regionais a serem visitadas. O fluxograma que é representado na Figura 4 foi validado com o colaborador B/DEE e faz um resumo geral da atividade descrita pelo POP (ANEXO A).

De posse do fluxograma, foi realizada a visita à Unidade Regional DC, com a participação do pesquisador e do colaborador B/DEE.

Após ser apresentado aos demais colaboradores da DEE, o pesquisador abordou rapidamente qual seria o objetivo do projeto e também o objetivo daquela visita à unidade. Após as devidas apresentações, o colaborador B/DEE colocou o pesquisador em contato com colaborador C/DEE, que era o responsável pela atividade de seleção e liberação de inspeção naquela unidade.

O colaborador C/DEE mencionou que tinha achado o projeto muito interessante e se prontificou a ajudar no que fosse necessário. Segundo ele, esse tipo de projeto era justamente o que os colaboradores que trabalhavam com as perdas não técnicas estavam buscando, “adicionar inteligência ao processo de inspeção de suspeitos de irregularidade”.

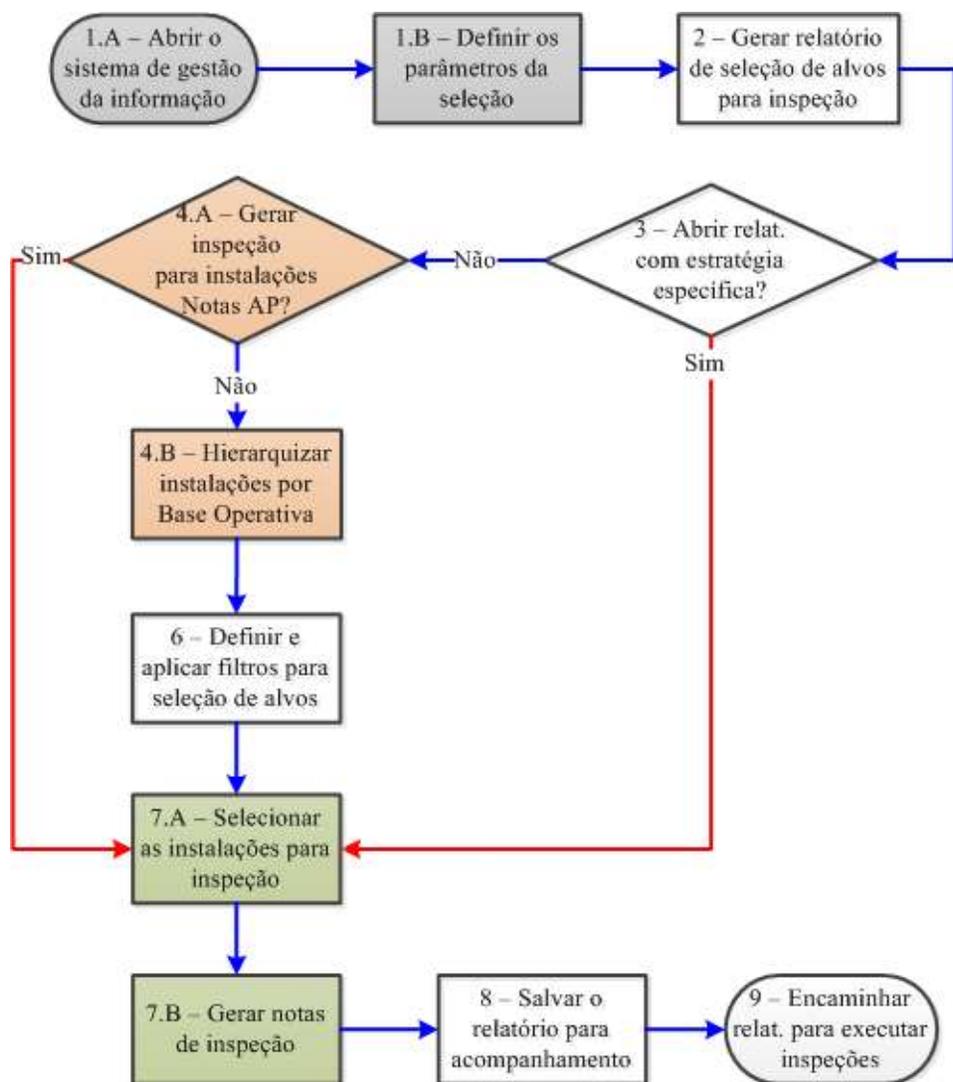


Figura 4 Fluxograma do Procedimento Operacional Padrão (POP) usado na atividade de seleção e liberação de inspeção, na distribuidora em estudo

Fonte: Organizado pelo autor a partir do ANEXO A

Considerando que o colaborador C/DEE era o responsável pela atividade, o pesquisador solicitou, antes de mostrar o fluxograma do POP, que

ele descrevesse qual era, exatamente, a sua rotina de trabalho. O colaborador C/DEE mencionou que, exceto quando ele estava de férias, ele era a única pessoa responsável pela liberação das ordens de serviço de inspeção para cada uma das bases operativas da Unidade Regional DC. Comentou ainda que a DC possuía características distintas das demais unidades regionais. A justificativa para essa afirmação, segundo o colaborador C/DEE, era que a DC é a única unidade que: a) possui uma pessoa dedicada à atividade de análise e liberação das inspeções (no caso o colaborador C/DEE); b) trabalhava com mais de uma base operativa (ou seja, possuía diversas equipes de eletricitas distribuídas por diferentes sub-regiões); e c) possuía equipes de eletricitas dedicadas exclusivamente às inspeções.

Segundo os colaboradores B/DEE e C/DEE, nas demais unidades regionais, na maioria dos casos, o mesmo colaborador que era responsável por essa atividade (seleção e liberação) também era responsável por outras atividades como, a cobrança de clientes e outros serviços internos. Além disso, nas demais unidades, a disponibilidade de equipe de pessoal própria também era menor. E, para os colaboradores, esses fatos poderiam explicar as significativas diferenças nos resultados de inspeção entre a DC e as demais unidades regionais.

Antes de iniciar a descrição de sua rotina, o colaborador C/DEE mencionou que o objetivo do departamento de proteção da receita da DEE, em específico o setor de perdas, era “realizar o mínimo de inspeções possível, com o maior potencial de recuperação de receita (*kWh*)”. Isso porque, internamente, a DEE também possuía capacidade limitada de processamento das inspeções realizadas.

De acordo com os colaboradores, caso no número de inspeções procedentes (consumidores que efetivamente fraudam o consumo de energia) aumentasse muito, seria gerada uma série de outros processos internos que poderiam causar problemas ainda maiores para a DEE e por um tempo

relativamente longo, pois envolveria outros setores da DEE como a movimentação de arquivos, cobrança, recurso na ouvidoria, recurso na ouvidoria da Agência Nacional de Energia Elétrica (ANEEL), entre outros, que poderiam chegar até o judiciário (quando houvesse processos civis). Por isso, a DEE precisaria (de acordo com os colaboradores B/DEE e C/DEE) melhorar a técnica de seleção de alvos (suspeitos de irregularidade no consumo de energia) para aumentar o Índice de Acerto das Inspeções (número de inspeções procedentes, ou seja, com fraude comprovada), para recuperar o faturamento do consumo (o volume de energia que não foi cobrado até a constatação) e, conseqüentemente, incrementar o faturamento (ou seja, cobrar o volume real de energia consumido, considerando um período de 12 meses a partir da regularização). Assim, os colaboradores concluíram que o foco era realizar o menor número de inspeções possível, com o maior potencial de recuperação e incremento de faturamento. Na seqüência, o colaborador C/DEE descreveu qual o processo adotado por ele na atividade de seleção e liberação das inspeções.

Descrevendo a atividade, ele comentou que as ordens de inspeção eram geradas semanalmente, mas ele fazia um acompanhamento diário do envio das ordens para as equipes de eletricitas. Isso porque havia uma tentativa de manter uma produtividade mínima semanal por base operativa da DC e evitar uma quantidade excessiva de inspeções pendentes. A atividade de seleção e liberação das ordens de inspeção realizada pelo colaborador C/DEE está representada no fluxograma da Figura 5.

Com base na descrição feita pelo colaborador C/DEE, o fluxograma representado na Figura 5 deveria contar ainda com três outros fluxogramas, que são representados nas Figura 6, 7 e 8.

Ao validar o fluxograma representado na Figura 5 com o colaborador C/DEE, ele afirmou que o fluxo, apresentado pelo pesquisador, realmente correspondia à atividade que ele desenvolvia. Mas, segundo ele, dependendo da

base operativa da unidade DC, ele adotava diferentes procedimentos. Isso porque havia diferentes especificidades para cada uma das bases e de nada adiantava usar critérios que não representassem tais especificidades.

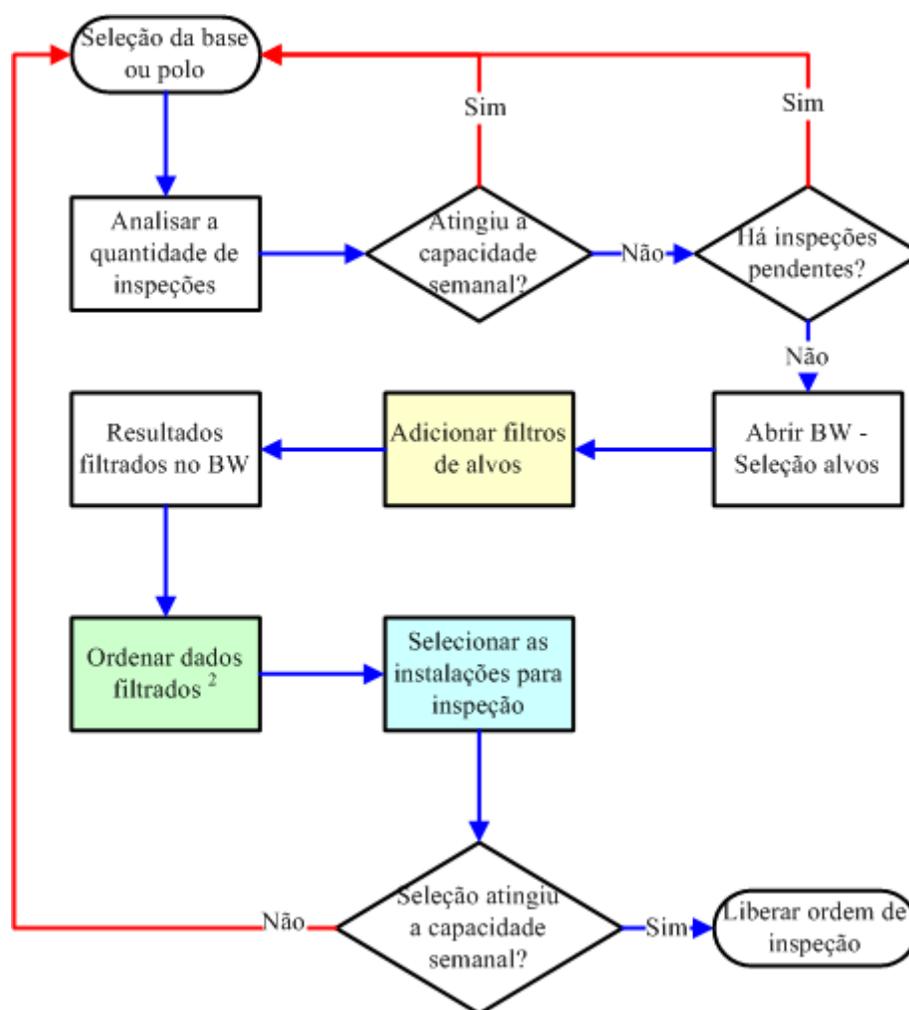


Figura 5 Fluxograma usado por um colaborador na seleção e liberação de inspeção, em uma unidade regional da distribuidora em estudo  
Fonte: Organizado pelo autor a partir de entrevista realizada com o colaborador C/DEE

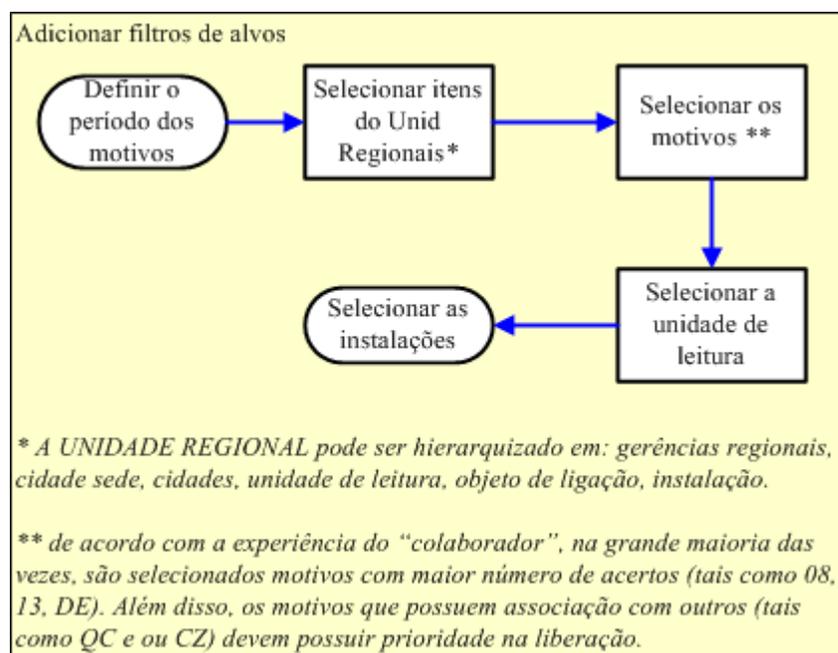


Figura 6 Fluxograma usado por um colaborador para adicionar filtro de alvos, em uma unidade regional da distribuidora em estudo

Fonte: Organizado pelo autor a partir de entrevista realizada com o colaborador C/DEE

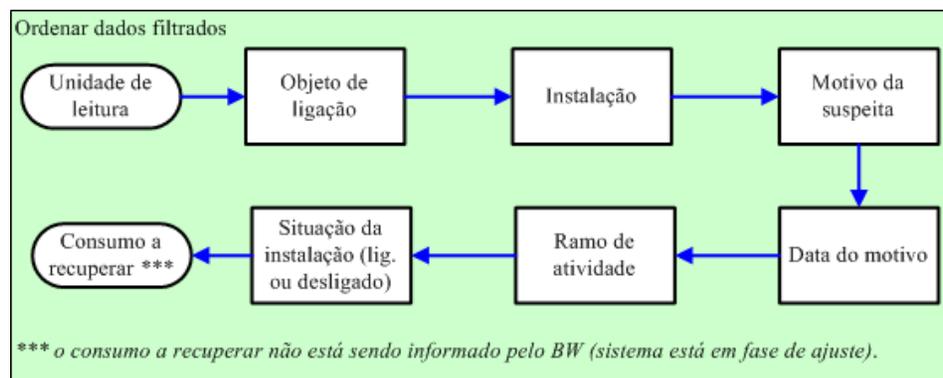


Figura 7 Fluxograma usado por um colaborador para ordenar dados filtrados, em uma unidade regional da distribuidora em estudo

Fonte: Organizado pelo autor a partir de entrevista realizada com o colaborador C/DEE

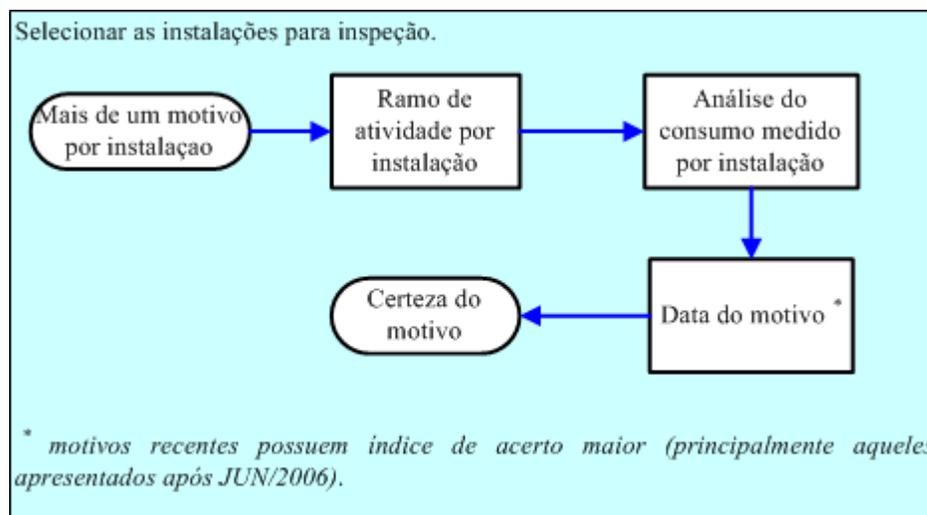


Figura 8 Fluxograma usado por um colaborador para selecionar as instalações a serem inspecionadas, em uma unidade regional da distribuidora em estudo

Fonte: Organizado pelo autor a partir de entrevista realizada com o colaborador C/DEE

Fazendo uma comparação preliminar entre o fluxograma representado na Figura 4 com o descrito pelo colaborador C/DEE (Figura 5), pôde-se perceber que o POP da atividade de seleção e liberação de inspeção, estabelecido pela DEE, não estava sendo seguido pelo colaborador C/DEE, pois quase nenhuma ação do fluxograma da Figura 4 estava sendo seguido pelo colaborador.

Após a visita à Unidade Regional DC, o pesquisador visitou a Unidade Regional DS, cujo responsável pela seleção e liberação de inspeção na unidade era o colaborador D/DEE. Após ser apresentado, o pesquisador novamente apresentou a motivação do projeto e também o objetivo daquela visita à unidade.

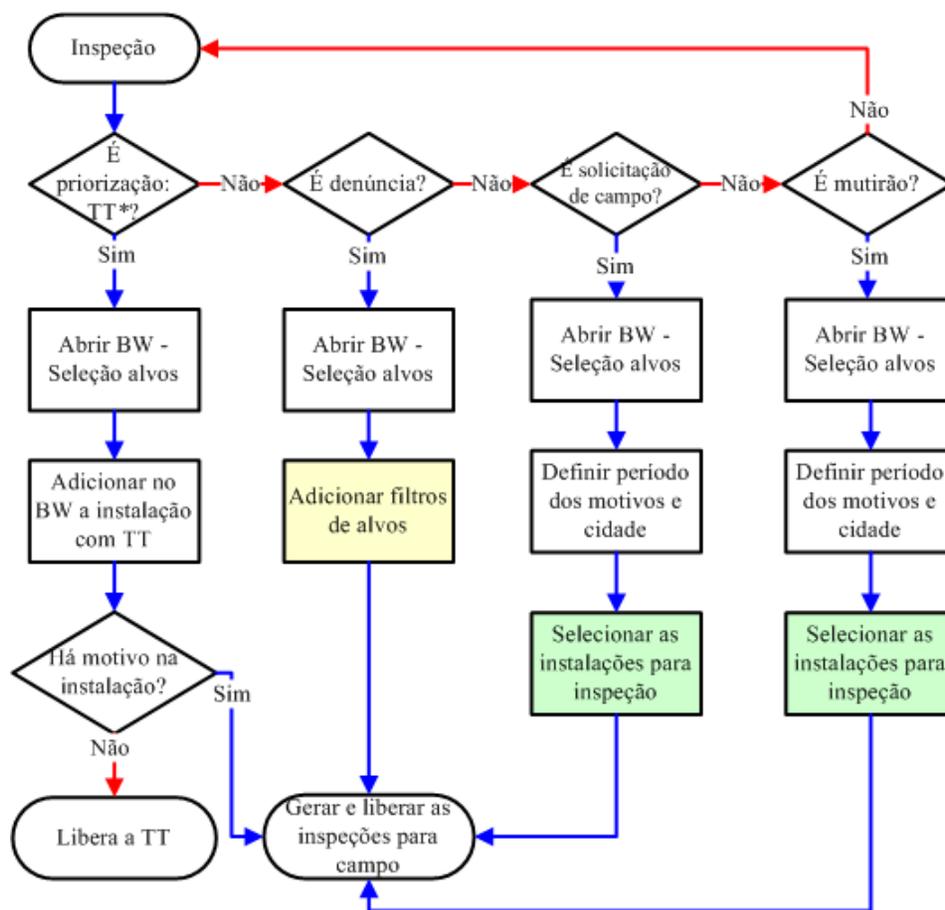
O colaborador D/DEE mencionou que ele era o supervisor responsável pela unidade DS. Por esse motivo, além de atuar na seleção e liberação inspeção, ele também coordenava os demais setores ligados ao processo de recuperação de perdas, tais como: cálculo de consumo a recuperar, documentação, processo, entre outros.

Segundo relatos do colaborador, a atividade de seleção e liberação das inspeções na unidade seguia as recomendações da unidade funcional. Por isso, para ele, seria importante salientar que as liberações das inspeções passavam por duas situações principais, uma relacionada com a priorização de inspeção em trocas de titularidade (orientada pela DEE) e outra relacionada com o tratamento de denúncias. Fora essas duas situações, as suspeitas de fraude eram inspecionadas via mutirões realizados em localidades pré-estabelecidas ou a partir da solicitação dos eletricitistas (ou seja, solicitação de campo para uma dada localidade). Por fim, apenas em casos onde ele sabia que a equipe de campo estava com a meta de serviço abaixo do pactuado é que ele selecionava e liberava as ordens de inspeção.

Apesar dessas peculiaridades, o colaborador descreveu a atividade de seleção e liberação das ordens de inspeção, está representada no fluxograma da Figura 9.

Assim como aconteceu no fluxograma do colaborador C/DEE (Figura 5), o fluxograma do colaborador D/DEE (Figura 9), também contou com dois outros fluxogramas, que estão representados nas Figura 10 e 11.

Para facilitar o entendimento do fluxograma da atividade, o colaborador D/DEE o descreveu de forma simplificada. Segundo ele, a recomendação da DEE era para que, no momento da solicitação de troca de titularidade (TT), fosse gerado um procedimento interno para verificar se não havia alguma suspeita de irregularidade no consumo do imóvel. Se houvesse, os responsáveis pela seleção e liberação das ordens de inspeção deveriam liberar, prioritariamente, uma ordem de inspeção para os consumidores que estivessem naquela situação. O colaborador comentou ainda que aquela recomendação deveria evitar que o novo titular do imóvel continuasse fraudando o consumo (caso ela fosse confirmada), mesmo sem saber.



\* TT é Troca de Titularidade

Figura 9 Fluxograma usado por um colaborador na seleção e liberação de inspeção, em uma unidade regional da distribuidora em estudo

Fonte: Organizado pelo autor a partir de entrevista realizada com o colaborador D/DEE

De acordo com o colaborador D/DEE, por um lado essa priorização era interessante, pois poderia evitar indisposição posterior com um cliente que, possivelmente, não tinha nada a ver com a fraude. Mas, por outro lado, para a unidade regional, aquela média era pouco benéfica por dois motivos: a) a quantidade de solicitações de troca de titularidade era muito elevada e, como a

quantidade de suspeitas também era alta, quase toda a capacidade de inspeção das equipes acabava sendo destinada para aquele tipo de inspeção; e b) a suspeita, de acordo com o colaborador, muitas das vezes, não era confirmada, o que resultava em um baixo Índice de Acerto da unidade que comprometia, conseqüentemente, a motivação das equipes de campo na realização das inspeções.

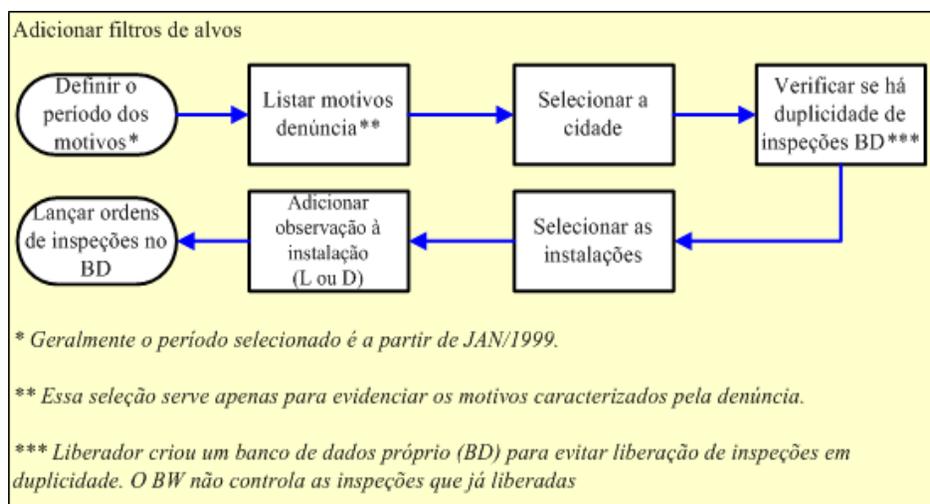


Figura 10 Fluxograma usado por um colaborador para adicionar filtros de alvo, em uma unidade regional da distribuidora em estudo

Fonte: Organizado pelo autor a partir de entrevista realizada com o colaborador D/DEE

Outro item destacado no fluxograma eram as denúncias de fraude. Segundo ele, aquelas inspeções eram liberadas o quanto antes porque, geralmente, as denúncias partiram dos leituristas e, em muitos casos, de vizinhos. Com isso, o Índice de Acerto das inspeções realizadas, naqueles casos era elevado, aumentando a recuperação das perdas e dinamizando o processamento interno das liberações.

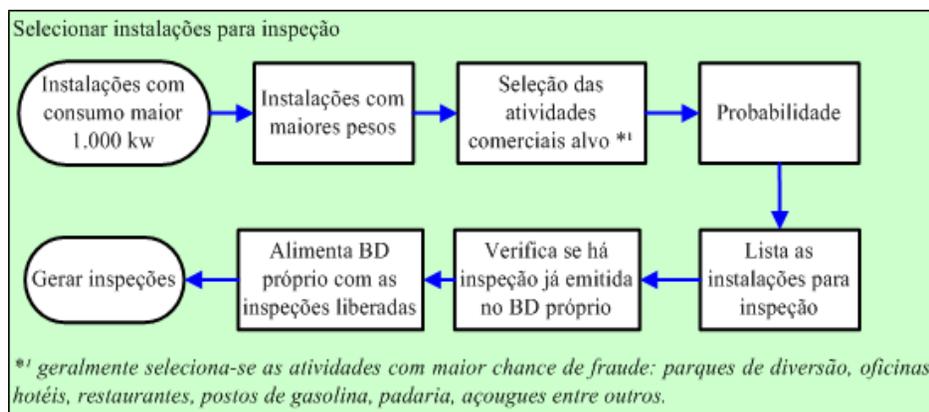


Figura 11 Fluxograma usado por um colaborador para selecionar as instalações a serem inspecionadas, em uma unidade regional da distribuidora em estudo

Fonte: Organizado pelo autor a partir de entrevista realizada com o colaborador D/DEE

Em outro caso, quando as equipes de eletricitas solicitavam suspeitos para inspeção, o colaborador D/DEE comentou que a seleção e liberação eram feitas diretamente no sistema de informação da DEE (BW), dando prioridade aos motivos que possuíam as maiores chances de acerto (Figura 11).

A última situação tratada pelo colaborador D/DEE foi sobre as inspeções realizadas em mutirão. De acordo com ele, o mutirão era muito eficiente porque, além de servir como forma de integração das equipes de campo, estimulava os leituristas a fazerem a denúncia. Além disso, o mutirão mostrava à população consumidora que a distribuidora estava trabalhando para combater as fraudes de consumo, inibindo, conseqüentemente, novas fraudes. Ressaltou ainda o colaborador, que o mutirão além de gerar custos para a unidade regional, com deslocamento de equipe de campo, alimentação e hospedagem, também aumentava o registro de ordens de serviço na unidade regional onde o mutirão está sendo realizado. Por esse motivo, a maioria dos gerentes das regionais não costumava comprometer seus orçamentos com ações que poderiam gerar um

baixo Índice de Acerto e com risco de não cumprir a meta acordada com a Unidade Funcional.

Ao apresentar o fluxograma da Figura 9 ao colaborador D/DEE, ele afirmou que o fluxo correspondia à atividade que ele desenvolvia no dia a dia de trabalho e que o mesmo era válido para aquela unidade.

Assim, como foi feito com o fluxograma descrito pelo colaborador C/DEE (Figura 5), ao comparar, preliminarmente, o fluxograma representado na Figura 4 com o descrito pelo colaborador D/DEE (Figura 9), percebe-se que o POP de seleção e liberação de inspeção, estabelecido pela DEE, também não estava sendo seguido pelo colaborador D/DEE.

Após a compilação de todos os dados coletados nas unidades regionais da DEE, realizou-se uma reunião com o colaborador A/DEE para analisar os dados coletados. A análise detalhada dos dados levantados é apresentada no tópico a seguir.

### **5.2.3 Análise dos dados do Ciclo PA-1**

Para o pesquisador, apesar dos colaboradores C/DEE e D/DEE trabalharem na mesma empresa e serem orientados a seguirem o mesmo POP para a atividade de seleção e liberação de inspeção, como também receberem o mesmo treinamento e usarem o mesmo sistema de gestão, o processo de liberação das ordens de inspeção, em cada unidade regional, possuía características totalmente distintas.

Para o pesquisador, essa situação poderia ser explicada justamente por estarem em unidades totalmente distintas, com características e metas diferentes. E, a partir desse resultado, o pesquisador sugeriu que a atividade de seleção e liberação de inspeções fosse centralizada em uma única unidade, com já havia

sido comentado pelo coordenador/DEE, e que o processo fosse revisto e padronizado.

Segundo o colaborador A/DEE a distribuidora já estava viabilizando uma nova estrutura física e novos colaboradores para reorganizar o setor da empresa que ficaria responsável por essas atividades. Mesmo assim, o resultado encontrado nesse ciclo tinha sido muito importante para confirmar a suspeita que ele e o coordenador/DEE já tinham sobre o processo. Com relação à padronização do processo, o colaborador comentou que, mesmo com a centralização, ainda continuará existindo alguma diferenciação no processo de seleção, pois as unidades regionais possuem características distintas.

Diante do exposto pelo colaborador A/DEE, o pesquisador argumentou que os dados coletados no Ciclo PA-1 foram importantes, mas teriam limitada contribuição para a solução do problema. Principalmente se considerar que a distribuidora já estava agindo para minimizar as discrepâncias na atividade de seleção e liberação de inspeções encontradas nas unidades regionais.

Para que a pesquisa-ação não fugisse do foco, o colaborador A/DEE sugeriu que fosse montado um plano de ação, que será descrito no tópico a seguir.

#### **5.2.4 Plano de ação do Ciclo PA-1**

O colaborador A/DEE sugeriu que, pela dificuldade de uma ação mais direta na implantação e na avaliação da ação proposta para o Ciclo PA-1, a pesquisa-ação deveria retomar a premissa do problema que era melhorar o processo de alocação das equipes de campo (eletricistas) para realizar o menor número de inspeções possível com o maior Índice de Acerto.

A partir dessa proposta, o pesquisador sugeriu então que o Ciclo PA-1 fosse encerrado e um novo ciclo interativo da pesquisa-ação (coleta de dados,

*feedback* aos participantes, análise de dados, planejamento de ações e avaliação das ações) fosse iniciado, tendo como fundamentado na análise do histórico das inspeções já executadas pela distribuidora. Para o pesquisador, a análise histórica poderia ser usada para se obter novas informações sobre as inspeções.

O colaborador A/DEE concordou com a proposta e colocou o colaborador B/DEE à disposição para auxiliar na coleta de dados do no ciclo de pesquisa-ação.

### **5.3 Ciclo PA-2: histórico das inspeções realizadas**

O Ciclo PA-2 corresponde ao ciclo interativo que avaliou, entre outros dados, o histórico das inspeções realizadas pela DEE. A descrição e análise dos dados serão feitas nos subitens a seguir e usará como referência os passos da implantação da pesquisa-ação, como representado na Figura 3. Esse ciclo completo teve a duração de 6 meses.

#### **5.3.1 Obtenção de dados do Ciclo PA-2**

Os dados usados no Ciclo PA-2 foram obtidos diretamente na sede da DEE com o auxílio dos colaboradores A/DEE, B/DEE e C/DEE. Esses dados, em sua grande maioria, foram extraídos do Sistema Integrado de Gestão Empresarial (*Enterprise Resource Planning – ERP*) da distribuidora e por meio de entrevistas não estruturadas com os colaboradores da DEE.

#### **5.3.2 *Feedback* de dados do Ciclo PA-2**

Para entender as características das perdas não técnicas na distribuidora, o pesquisador solicitou ao colaborador A/DEE que dados gerais sobre as perdas

não técnicas fossem disponibilizados, por unidade regional. O pesquisador justificou a solicitação afirmando que era importante fazer um paralelo entre as unidades regionais da DEE e as perdas comerciais, como forma de direcionar as próximas etapas da pesquisa-ação. O colaborador concordou com a solicitação e com a justificativa do pesquisador, disponibilizando os dados.

Após processar os dados, referente a um período de 12 meses, de inspeções realizadas, o pesquisador, o coordenador/DEE e os colaboradores DEE se reuniram para analisá-los. A análise detalhada dos dados levantados é apresentada no tópico a seguir.

### **5.3.3 Análise dos dados do Ciclo PA-2**

No período de 12 meses analisado, a DEE contava com mais de 6,5 milhões de consumidores ativos nas sete unidades regionais. Como pode ser visto na Figura 12, a unidade regional DC é a que registrava o maior número de consumidores na DEE (27,8% do total). Além disso, pode-se afirmar que a unidade regional com o menor número de consumidores registrados era na DO, com apenas 8,3% do total.

Apesar da diferença no número de consumidores de cada unidade regional, ao serem segmentados por classe de consumo, percebe-se que distribuição percentual dessas classes era relativamente semelhante entre as unidades, como pode ser visto na Figura 13.

A unidade DC se destacava por possuir as maiores classes residencial (87,2%) e comercial/serviços (9,9%) e também a menor classe rural (1,2%). Por outro lado, a unidade DN apresentava as menores classes residencial (74,0%) e industrial (0,6%) e a maior classe rural (16,4%). A unidade regional DO possuía, percentualmente, a maior classe industrial (1,7%). A menor classe comercial

era da DN com 7,4%. A classe classificada como outros, englobava os consumidores relacionados ao poder público, iluminação pública, serviços públicos em geral e consumo próprio da DEE. Nessa classe, a unidade DN apresentava o maior percentual (1,5%) e a DC o menor (0,5%).

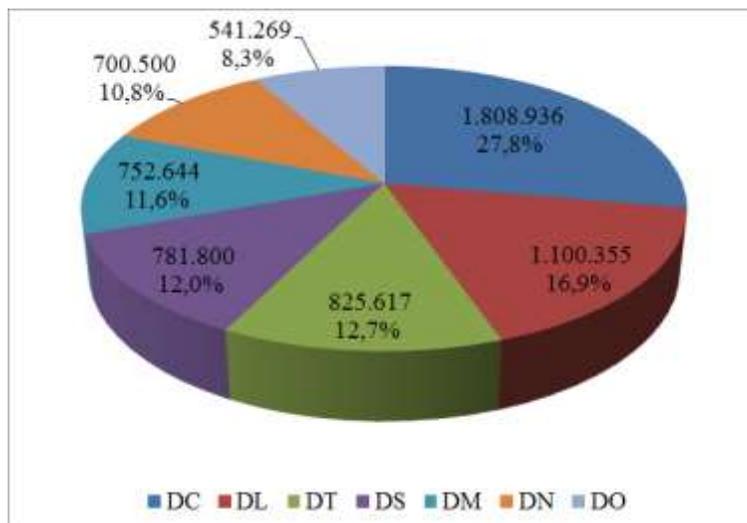


Figura 12 Quantidade total de consumidores por unidade regional da distribuidora em estudo, no período de 12 meses

Fonte: Organizado pelo autor a partir de dados obtidos em relatórios da DEE

De posse da distribuição dos consumidores por unidade regional e por classe de consumo de energia, foram analisadas as suspeitas de irregularidade no consumo de energia por unidade regional. A Figura 14 representa a quantidade de motivos existente por unidade regional.

No período analisado, a DEE contava com mais de 1,9 milhão de motivos registrados. Como pode ser visto na Figura 14, a unidade regional DC é a que possuía o maior número de motivos registrados (38,4% do total). Além disso, pode-se afirmar que a unidade regional com o menor número motivos no período analisado foi a DO, com apenas 7,8% do total.

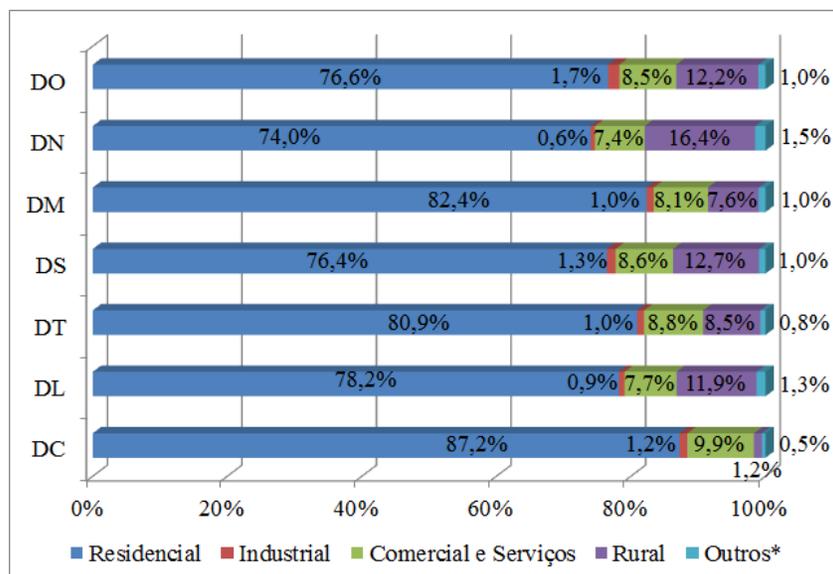


Figura 13 Distribuição percentual das classes de consumo de energia, por unidade regional da distribuidora em estudo, no período de 12 meses

Fonte: Organizado pelo autor a partir de dados obtidos em relatórios da DEE

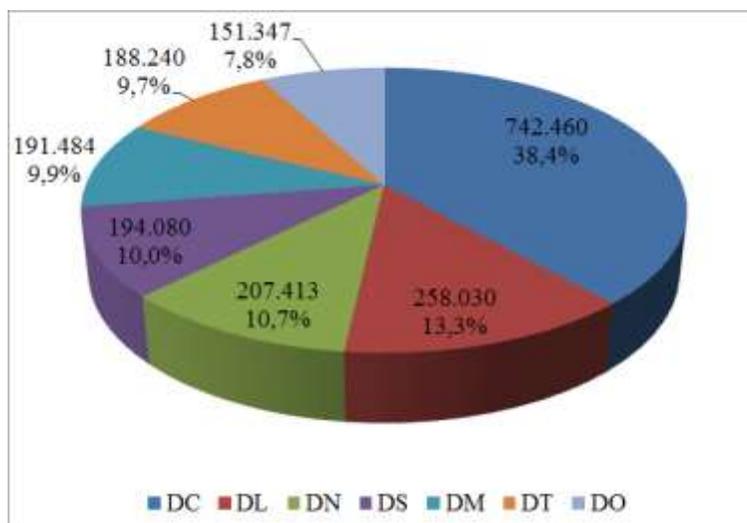


Figura 14 Quantidade total de motivos existentes, por unidade regional da distribuidora em estudo, no período de 12 meses

Fonte: Organizado pelo autor a partir de dados obtidos em relatórios da DEE

Ao comparar os valores da Figura 12 com os da Figura 14, pode-se afirmar que a unidade DC, além de apresentar o maior número de consumidores, também apresentava a maior quantidade de motivos registrados no período. De acordo com os relatórios da DEE, o total de 742.460 motivos estava distribuídos em 575.719 instalações. Ou seja, cada instalação da DC possuía, em média, 1,3 motivos registrados. De forma oposta, a unidade DO além de apresentar a menor número de consumidores também apresentava a menor quantidade de motivos registrados. Apesar disso, a relação motivo por instalação na unidade DO era semelhante à da DC, com 1,2 motivo por instalação.

Analisando o resultado, o colaborador A/DEE, disse que esse resultado já era esperado. Para ele, as regiões com o maior número de consumidores tendiam a ter mais irregularidades no consumo de energia. Assim, a quantidade de motivo registrado por região evidenciava, de acordo com o colaborador, a região que deveria apresentar o maior número de consumidores fraudando o consumo de energia. Apesar da aparente superficialidade dessa informação, ela foi importante para direcionar ações e investigações futuras.

Outra análise realizada foi em relação aos tipos de motivo registrados. De acordo com o colaborador B/DEE, os consumidores com suspeita de consumo irregular eram incluídos no banco de suspeitos a partir do motivo que gerou a suspeita. Na DEE vários motivos poderiam gerar suspeita de consumo irregular, entretanto os mais comuns eram:

- a) QC: Queda de consumo ou demanda;
- b) CZ: Registro de consumo zero;
- c) 08: Suspeita de irregularidade;
- d) 13: Suspeita medidor defeituoso/manipulado;
- e) PB: Inspeção periódica em baixa tensão;
- f) PN: Parceiro de negócio com irregularidade em outra instalação;

- g) 01: Caixa de medição c/ selos rompidos (padrão externo); e
- h) PM: Inspeção periódica em média tensão.

Os tipos de motivo mais recorrentes nas unidades regionais da DEE, no período de 12 meses, são representados a partir da Figura 15 até a Figura 21.

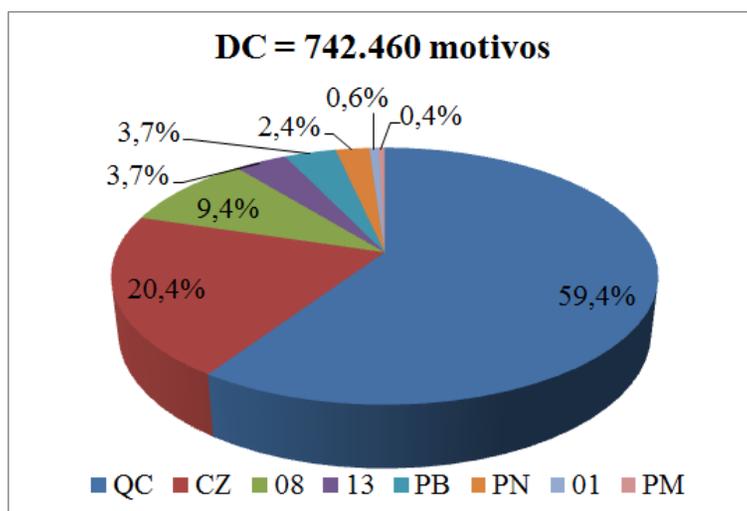


Figura 15 Principais tipos de motivo registrado na unidade regional DC da distribuidora em estudo, no período de 12 meses

Fonte: Organizado pelo autor a partir de dados obtidos em relatórios da DEE

Analisando os tipos de motivos, os que apresentaram maior quantidade de registros, em todas as unidades regionais, foi o QC (com variação de 59,4% a 68,9% dos registros), seguido por CZ (variando de 16,03% a 24,87%) e pelo 01 (variando de 0,6% a 17,3%).

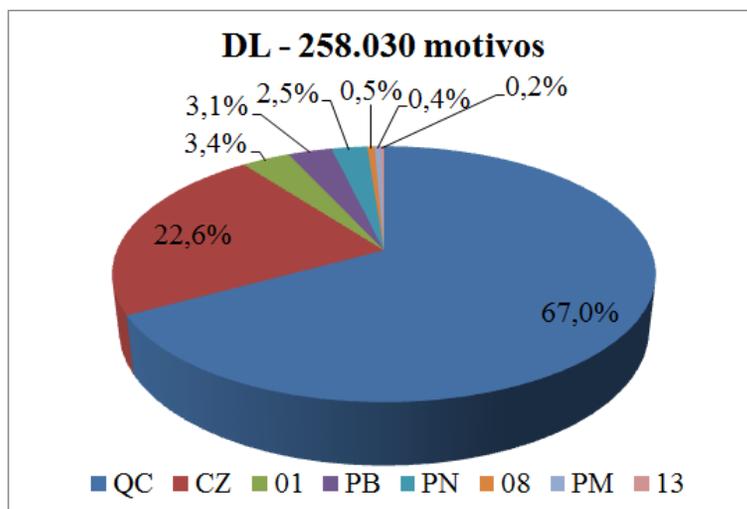


Figura 16 Principais tipos de motivo registrado na unidade regional DL da distribuidora em estudo, no período de 12 meses

Fonte: Organizado pelo autor a partir de dados obtidos em relatórios da DEE

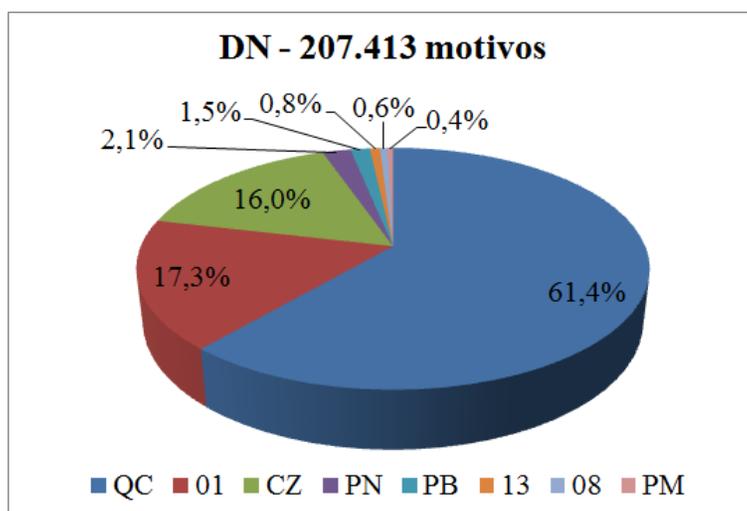


Figura 17 Principais tipos de motivo registrado na unidade regional DN da distribuidora em estudo, no período de 12 meses

Fonte: Organizado pelo autor a partir de dados obtidos em relatórios da DEE

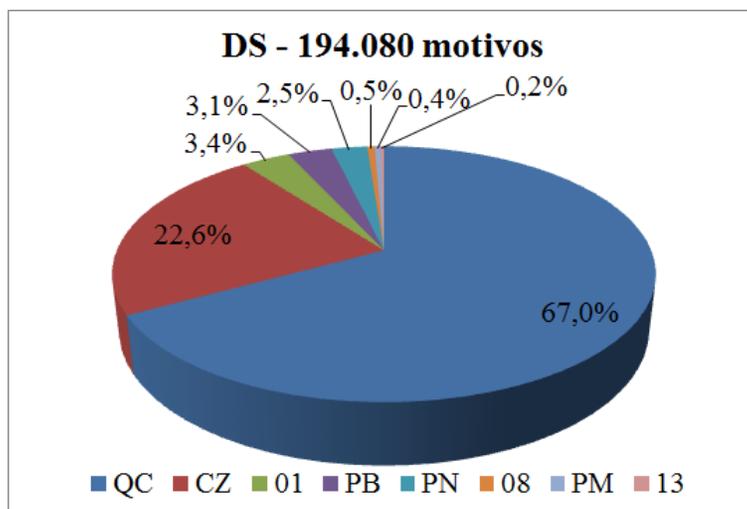


Figura 18 Principais tipos de motivo registrado na unidade regional DS da distribuidora em estudo, no período de 12 meses

Fonte: Organizado pelo autor a partir de dados obtidos em relatórios da DEE

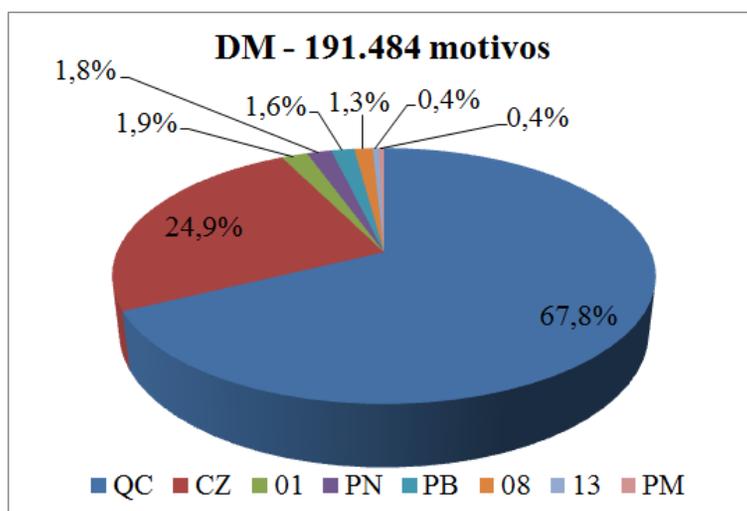


Figura 19 Principais tipos de motivo registrado na unidade regional DM da distribuidora em estudo, no período de 12 meses

Fonte: Organizado pelo autor a partir de dados obtidos em relatórios da DEE

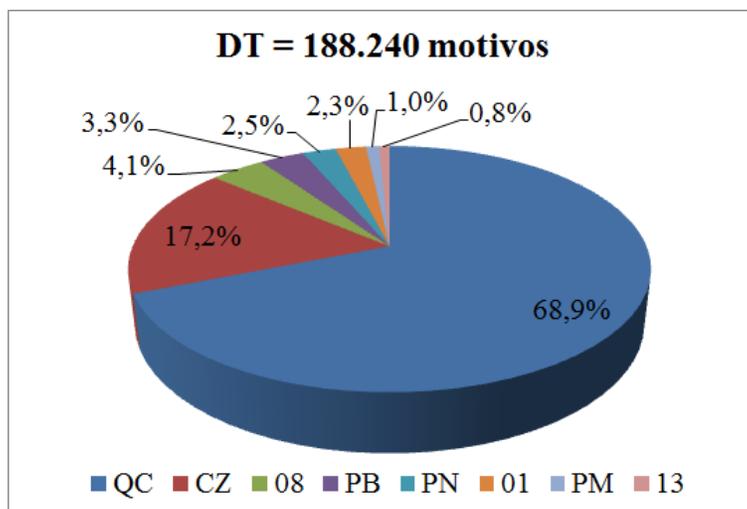


Figura 20 Principais tipos de motivo registrado na unidade regional DT da distribuidora em estudo, no período de 12 meses

Fonte: Organizado pelo autor a partir de dados obtidos em relatórios da DEE

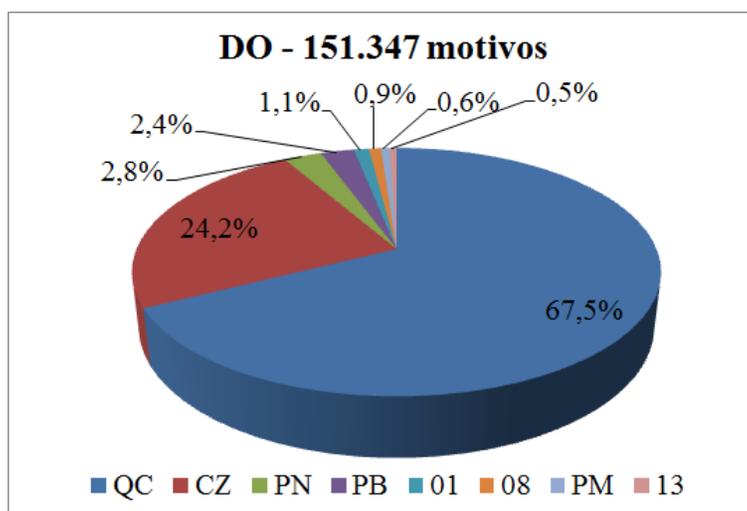


Figura 21 Principais tipos de motivo registrado na unidade regional DO da distribuidora em estudo, no período de 12 meses

Fonte: Organizado pelo autor a partir de dados obtidos em relatórios da DEE

Um fato que chamou a atenção foi que os motivos do tipo QC e CZ apresentavam a maior quantidade de registro em quase todas as unidades regionais. Na média, a soma dos dois tipos de motivos respondia por mais de 80% de todos os motivos registrados na DEE. A única unidade regional em que QC e CZ não eram os principais motivos era na DN, mas mesmo assim os dois motivos juntos respondiam por mais de 77% de todos os motivos registrados. A unidade que apresentava o menor percentual dos dois motivos juntos era a DC (68,8%) e a que apresentava o maior percentual era a DM (92,6%).

Para o colaborador C/DEE, a justificativa para essa situação era porque o ERP da distribuidora registrava, automaticamente, os consumidores que apresentavam os motivos QC e CZ. Ou seja, foi criado um conjunto interno de regras que quando uma instalação as atendesse, o sistema de gestão automaticamente registrava o consumidor no banco de suspeitos. Segundo o colaborador, essas regras justificavam os quase 25% dos consumidores regulares estarem no banco de suspeitos, ao passo que menos de 5% dos consumidores efetivamente fraudavam o consumo. Por essa razão, era tão importante descobrir entre os suspeitos quais os consumidores que realmente estavam cometendo fraude.

Para entender melhor o perfil dos consumidores fraudadores, analisou-se também o resultado das inspeções que já haviam sido executadas. A Figura 22 representa a quantidade total de inspeções realizadas, por unidade regional da DEE, no período de 12 meses.

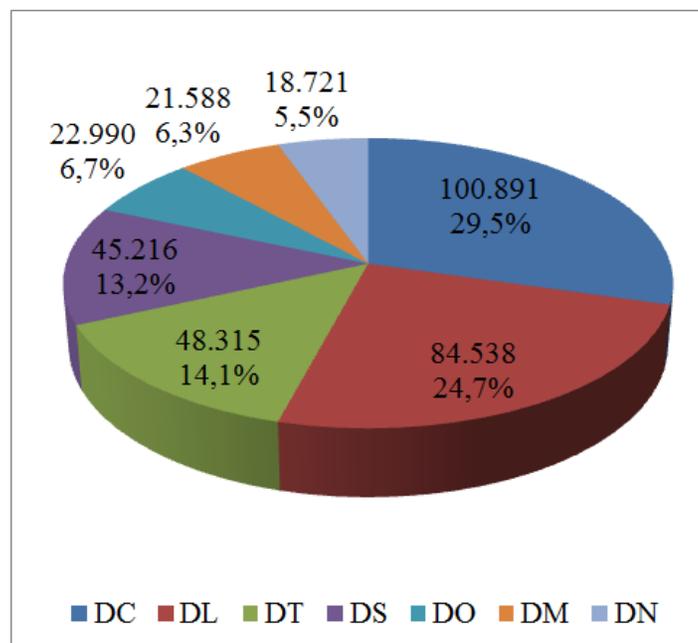


Figura 22 Quantidade total de inspeções realizadas, por unidade regional da distribuidora em estudo, no período de 12 meses

Fonte: Organizado pelo autor a partir de dados obtidos em relatórios da DEE

Analisando a quantidade de inspeções realizadas (Figura 22), a unidade regional DC foi a que realizou o maior número de inspeções no período (100.981), seguido pela unidade DL, com 84.538 inspeções realizadas. Vale destacar que essas duas unidades regionais também registravam o maior número de consumidores (Figura 12) e de motivos (Figura 14).

No período em análise, a DEE liberou um total de 342.259 notas de inspeções. Desse total, 26,0% das inspeções realizadas foram procedentes, ou seja, ao realizar a inspeção foi constatado que havia fraude no consumo de energia. Por outro lado, em 48,8% das instalações não foi encontrado nenhuma irregularidade no consumo, ou seja, o resultado da inspeção foi improcedente. No restante das instalações (mais de 25%), as equipes de campo foram impedidas de realizar a inspeção nas instalações, não sendo possível comprovar

a existência ou não de fraude. Com isso, o resultado da inspeção era impedido. No geral, o Índice de Acerto (IA) médio da DEE foi de 34,8%. O IA considera apenas as instalações inspecionadas, assim esse índice é o resultado da razão entre o total de instalações com resultado procedente pela diferença entre o total de notas de inspeção e a quantidade de inspeções com resultado impedido.

O resultado das inspeções que já haviam sido executadas, por unidade regional da DEE, no período de 12 meses é representado na Tabela 1.

A partir das informações da Tabela 1, deve-se destacar o resultado da unidade DC. Essa foi a região com o maior número de notas de inspeção liberadas, com o maior índice de acerto, com o maior percentual de inspeções impedidas e com o menor percentual de inspeções improcedentes.

Outro resultado que se destacou foi o da unidade DL. Com o segundo maior número de inspeções liberadas, essa unidade foi a que teve o maior percentual de inspeções improcedentes, o menor percentual de inspeções procedentes e, conseqüentemente, o menor índice de acerto de todas as regionais.

Tabela 1 Resultado das inspeções liberadas por unidade regional da distribuidora em estudo, no período de 12 meses

UNIDADE REGIONAL	NOTAS	INSPEÇÕES			ÍNDICE
	LIBERADAS Total	Procedentes %	Improcedentes %	Impedidas %	DE ACERTO %
DC	100.891	33,5	21,1	45,4	61,4
DN	18.721	34,4	57,5	8,1	37,4
DO	22.990	30,8	53,9	15,3	36,4
DM	21.588	25,7	55,2	19,1	31,8
DS	45.216	23,9	53,3	22,8	31,0
DT	48.315	24,8	56,2	19,1	30,6
DL	84.538	15,8	70,5	13,7	18,3

Fonte: Organizado pelo autor a partir de dados obtidos em relatórios da DEE

De acordo com o colaborador C/DEE, os dois destaques descritos refletiam bem a situação relatada por ele anteriormente. Como a DC era a única unidade regional que contava com um colaborador específico para selecionar e liberar inspeções e que tinha as equipes de campo dedicadas a realizar inspeções, o resultado da regional destacou-se positivamente. Como essa situação não ocorria na DL, o resultado obtido pela regional destacou-se negativamente.

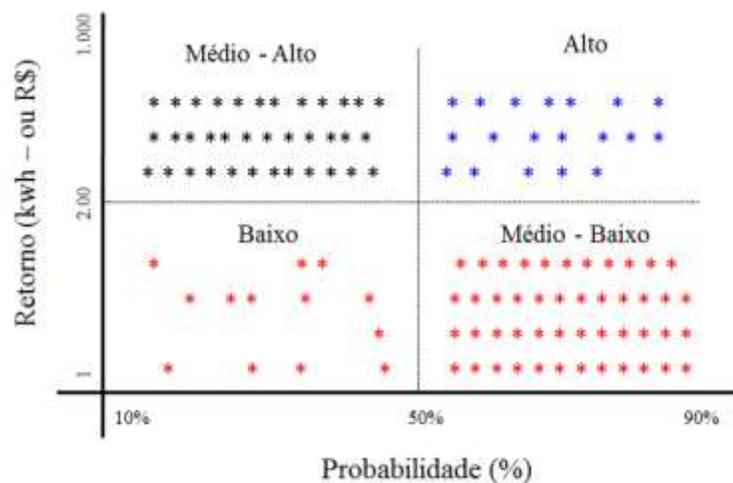
Diante dos resultados encontrados, o pesquisador argumentou que análise histórica das inspeções realizadas pela DEE foi extremamente importante para compreender melhor o problema e visualizar, de forma mais abrangente, a situação de cada unidade regional. Ainda segundo o pesquisador, o plano de ação a ser desenvolvido deveria ser pautado nos resultados obtidos pela unidade regional DC. O tópico a seguir apresenta o plano de ação sugerido pela equipe da pesquisa-ação.

#### **5.3.4 Plano de ação do Ciclo PA-2**

Como ação do Ciclo PA-2, o pesquisador sugeriu que fosse realizada uma análise mais detalhada das inspeções realizadas pelas unidades regionais da DEE. Mas que, para isso, seria necessário obter outros dados e com um número maior de variáveis. Para o pesquisador, essa nova análise deveria ser pautada, sobretudo, nos resultados positivos obtidos pela unidade DC.

O colaborador A/DEE concordou com a sugestão do pesquisador. Segundo ele, pelas análises e reuniões já realizadas, “o projeto já tinha ganhado maturidade suficiente para iniciar análises mais aprofundadas sobre o problema”. E o interessante seria que, como resultado desse plano de ação fosse possível alocar as equipes de campo, conforme o potencial de retorno das instalações suspeitas (recuperação/incremento de faturamento) e a probabilidade da inspeção ser procedente, para cada unidade regional. Ou seja, para o

colaborador, o resultado deveria se aproximar do gráfico representado na Figura 23.



\* Instalações suspeitas de irregularidade no consumo de energia

Figura 23 Alocação das equipes de campo de acordo com retorno e a probabilidade das inspeções serem procedentes, na distribuidora em estudo

Fonte: Organizado pelo autor a partir de entrevista com o colaborador A/DEE

Segundo o colaborador A/DEE, com resultados semelhantes ao representado na Figura 23, parte dos problemas com a alocação das equipes poderia ser resolvido. Porque, de acordo com ele os colaboradores, responsáveis pela seleção e liberação das inspeções, poderiam direcionar as ações das equipes de campos para as instalações com alto potencial de retorno e com alta probabilidade de ter irregularidade. Evitando, conseqüentemente, as instalações que com baixo potencial de retorno e baixa probabilidade de irregularidade.

O pesquisador afirmou que era possível de se conseguir o resultado da Figura 23. Entretanto, para isso, dados mais bem detalhados deveriam ser usados.

Considerando que uma nova base de dados deveria ser obtida, o pesquisador, juntamente com o coordenador/DEE e os colaboradores julgaram que seria mais prudente encerrar o Ciclo PA-2 e iniciar um novo ciclo de pesquisa-ação.

#### **5.4 Ciclo PA-3: análise das variáveis das inspeções realizadas**

O Ciclo PA-3 corresponde ao ciclo interativo que analisou, com profundidade, as variáveis do histórico das inspeções já realizadas pela DEE. A descrição desse ciclo, que teve a duração de 10 meses será feita nos subitens a seguir e usará como referência os passos da implantação da pesquisa-ação, como representado na Figura 3.

##### **5.4.1 Obtenção de dados do Ciclo PA-3**

Os dados usados no Ciclo PA-3 foram obtidos diretamente na sede da DEE com o auxílio dos colaboradores A/DEE, B/DEE e C/DEE. Os dados, em sua grande maioria, foram extraídos do Sistema Integrado de Gestão Empresarial *Enterprise Resource Planning* (ERP) da distribuidora. Além disso, outros dados foram obtidos por meio de entrevistas não estruturadas com os colaboradores da DEE, realizadas principalmente com o responsável pela seleção e liberação de inspeção da Unidade Regional DC.

##### **5.4.2 Feedback de dados do Ciclo PA-3**

Antes de iniciar a extração dos dados no ERP, o colaborador A/DEE e o pesquisador utilizaram a técnica de *brainstorming* para elencar algumas variáveis que poderiam ser utilizadas no Ciclo PA-3 e como elas poderiam

auxiliar na solução do problema levantado. A partir da utilização dessa técnica, foi possível elencar, entre outras variáveis, as seguintes:

- a) data dos motivos (data em que foi registrado o motivo que gerou a suspeita em uma dada instalação);
- b) quantidade de motivos por instalação (motivos registrados em cada instalação);
- c) tempo médio dos motivos (quantidade de dias que cada motivo aparece na instalação);
- d) tempo médio da irregularidade (quantidade de dias que a instalação ficou com irregularidade – essa variável se aplica apenas às instalações inspecionadas e que a fraude no consumo foi comprovada);
- e) data da geração da nota de inspeção (dia em que os colaboradores responsáveis pela seleção e liberação da inspeção geraram a ordem de serviço para que as equipes de campo realizassem a inspeção);
- f) data efetiva da inspeção (dia em que as equipes de campo efetivamente realizaram a inspeção na instalação suspeita);
- g) taxa de crescimento de motivos (evolução da quantidade de suspeitos de irregularidade no banco de dados da DEE);
- h) índice de acerto/recuperação de energia (é a razão entre a quantidade de inspeções realizadas e a quantidade de irregularidades encontradas nas inspeções);
- i) grupo de suspeitos (caracteriza e agrupa as instalações com suspeitas de irregularidade);
- j) equipe de inspeção (identificação das equipes de campo ou equipes de eletricitistas responsáveis pela inspeção das instalações);

- k) número de consumidores (quantidade de consumidores suspeitos de irregularidade por unidade regional);
- l) capacidade de inspeção por unidade regional (número máximo de inspeções que cada unidade regional é capaz de realizar em um período); e
- m) região/local da inspeção (cidade ou bairro em que a inspeção deve ser ou foi realizada).

Após esse levantamento foi solicitado ao colaborador B/DEE que ele extraísse os dados a partir do ERP da DEE. O colaborador B/DEE informou que o relatório de Análise de Inspeções gerado pelo próprio ERP, apresentava um número significativo de variáveis incluindo inclusive algumas já estabelecidas no *brainstorming*. O relatório de Análise de Inspeção contemplava as seguintes variáveis:

- a) instalação (número do relógio que mede o consumo de energia);
- b) objeto de ligação (imóvel em que a instalação foi feita. Por exemplo, em um prédio residencial com vários apartamentos, o prédio é o objeto de ligação e os relógios de consumo dos apartamentos as várias instalações);
- c) unidade de leitura (sequência numérica que agrupa três dados: razão, local e rota),
- d) código SVC (código do serviço de campo – registra o código do resultado da inspeção realizada);
- e) retorno do SVC (retorno do serviço de campo – registra o resultado que a equipe de eletricitista obteve: Procedente, Improcedente e Impedido);

- f) ramo de atividade (atividade fim para qual a energia é utilizada na instalação);
- g) número de fases (quantidade de fases elétricas disponíveis na instalação – monofásico, bifásico ou trifásico);
- h) tensão (condições de segurança do sistema de distribuição e utilização de energia elétrica – alta, média e baixa tensão); e
- i) dias entre a geração da nota de inspeção e a inspeção (diferença de período entre a liberação da nota de inspeção e a inspeção realizada pela equipe de campo).

Considerando a quantidade de variáveis e o número de inspeções realizadas, o colaborador B/DEE sugeriu que na primeira análise, fosse usado como amostra o resultado das inspeções realizadas em apenas um município da unidade regional DC, adotando como referência um período de 6 meses. Segundo o colaborador, a proposta feita se fundamentava no fato de que a regional DC era a que apresentava o maior IA de todas as unidades e o município selecionado era o que possuía o maior número de suspeitos e inspeções realizadas. Não vendo motivos contrários à sugestão, o pesquisador e o colaborador A/DEE concordaram com a proposta apresentada.

Após a extração dos dados do ERP da DEE, o pesquisador e o colaborador A/DEE fizeram uma avaliação geral dos dados e concluíram que algumas das variáveis anteriormente elencadas seriam de difícil operacionalização nesse primeiro levantamento. As variáveis que fizeram parte dessa análise, com suas respectivas características, são representadas no Quadro 5.

Quadro 5 Característica e sigla das variáveis selecionadas para serem analisadas no Ciclo PA-3

<b>VARIÁVEIS</b>	<b>CARACTERÍSTICA</b>	<b>SIGLA</b>
1. Unidade Regional	Local em análise	HIERARQUIA_ POLITICA
2. Tensão	Tensão instalada na unidade consumidora	TENSAO
3. Dia de leitura	Dado obtido a partir da unidade de leitura	DIA_LEITURA
4. Local de leitura	Dado obtido a partir da unidade de leitura	LOCAL_LEITURA
5. Rota de leitura	Dado obtido a partir da unidade de leitura	ROTA_LEITURA
6. Classe	Classe em que a instalação de enquadra	CLASSE
7. Ramo de atividade	Ramo de Atividade em que a instalação de enquadra	RAMO_ ATIVIDADE
8. Número de fases	Monofásico, Bifásico ou Trifásico	NUMERO_FASES
9. Motivo mais antigo	1º motivo que gerou a suspeita da instalação	MOT_ANTIGO
10. Data do motivo mais antigo menos a data de emissão da nota de inspeção	Dado obtido a partir da diferença entre a data do motivo mais antigo e a data de emissão da nota de inspeção (em dias)	DATAMOTIVO_ DATAENOTA_ DIAS
11. Data de emissão da nota de inspeção menos a data de inspeção	Dado obtido a partir da diferença entre a data de emissão da nota de inspeção e a data efetiva da inspeção (em dias)	DATAENOTA_ DATAINSPECAO_ DIAS
12. Grupo responsável pela execução da inspeção	Número da equipe de campo que realizou a inspeção	GR_EXEC_ INSPECAO
13. Colaborador que executou a inspeção	Número de matrícula do colaborador que realizou a inspeção	EXECUTOR_ INSPECAO
14. Retorno do serviço de campo	Procedente, Improcedente, Impedido	RETORNO_SVC
15. Cód. serviço de campo	Código do serviço de campo que indica a situação encontrada na instalação suspeita	COD_SVC
16. Número de irregularidades encontradas	Quantidade de irregularidade encontrada na instalação suspeita	NIE

Fonte: Organizado pelo autor a partir de informações da DEE

Considerando as variáveis selecionadas, o pesquisador iniciou a análise dos dados, que é descrita no tópico a seguir.

### 5.4.3 Análise dos dados do Ciclo PA-3

A análise dos dados levantados usou como fundamento o processo de mineração de dados (*data mining*) e a descoberta de conhecimento a partir de banco de dados. A justificativa para a escolha desse processo se fundamenta nos seguintes fatores:

- a) avaliando trabalhos que também trataram sobre o tema, a grande maioria usava o processo de mineração de dados com fundamento para encontrar soluções para o problema (CABRAL; GONTIJO, 2004; LEÓN et al., 2011; NAGI et al., 2008; ORTEGA, 2008; PENIN, 2008; QUEIROGA, 2005; REIS, 2005; REIS FILHO (2006); VIEIRALVES, 2005) entre outros);
- b) mesmo usando a amostra selecionada pelo colaborador B/DEE, o volume de dados extraído era muito elevado para usar estatísticas convencionais; e
- c) usar o processo de *datamining* possibilitaria encontrar padrões e gerar informações que, até então, eram imperceptíveis aos colaboradores da DEE.

Para realizar um processo de mineração de dados, métricas que vão desde análises estatísticas simples até complexos modelos de redes neurais podem ser utilizados. Além disso, utilizar qualquer uma dessas métricas, considerando o volume de dados a serem processados, sem o auxílio de um *software* específico é muito difícil. Diante do exposto, foi utilizado o *software*

*Weka 3: Data Mining Software in Java*, no processo de mineração dos dados. O *Weka* é um *software* de fonte aberta (*open source software*) que reúne uma coleção de algoritmos de máquina de aprendizagem (*machine learning*) que podem ser aplicados diretamente a um conjunto de dados ou pode-se usar seu próprio código (*Weka Java code*). Os algoritmos contidos no *Weka* podem ser usados para pré-processamento, classificação, regressão, *clustering*, associação e visualização (WITTEN; FRANK, 2005).

Witten e Frank (2005) comentam que o primeiro passo em um processo de mineração de dados é a preparação dos dados de entrada (limpeza e pré-processamento). Segundo os autores esse passo geralmente consome grande parte dos recursos a serem investidos em um processo de mineração de dados porque, de forma geral, é necessário seguir padrões determinados para que os dados sejam considerados válidos.

Assim, após a definição do *software*, os dados levantados foram preparados seguindo as etapas propostas por Witten e Frank (2005), que são:

- a) **resumo dos dados:** a primeira etapa consiste em agrupar os dados das variáveis em um conjunto de exemplos (*instances*). Os exemplos são as informações para se transformar em o aprendizado (conhecimento), de acordo com o conjunto de atributos (características) que eles possuem. Nessa etapa é necessário, por exemplo, limpar os dados das variáveis ou mesmo fundir e ou separar variáveis;
- b) **dados escassos:** a segunda etapa consiste em analisar os dados das variáveis que, muitas vezes, são compostos de valores iguais a zero ou a um conjunto muito amplo de texto. Dados com essas características dificultam o processo de mineração de dados (os dados são lidos e operacionalizados em forma de matriz);

- c) tipo de atributo: a terceira etapa consiste em avaliar qual o tipo de atributo mais indicado para cada variável. O *Weka* processa arquivos do tipo *Attribute-Relation File Format* (ARFF) que aceitam dados do tipo nominal e numérico. Assim, é necessário identificar qual o tipo de dado que cada variável deve assumir;
- d) valores desconhecidos: na maior parte dos arquivos de dados há valores que são desconhecidos. Nesses casos, os valores desconhecidos devem ser substituídos por um ponto de interrogação (?);
- e) valores imprecisos: é importante verificar os arquivos de mineração de dados cuidadosamente para evitar atributos e valores não desejáveis;
- f) padronização ARFF: a última etapa consiste em padronizar os exemplos em um formato cujo conjunto de dados se tornem independentes, desordenados e sem qualquer relação entre os exemplos.

Como abordado por Witten e Frank (2005), esse passo realmente foi o mais difícil, desgastante e que mais consumiu tempo do pesquisador na análise dos dados. No processo de limpeza da base de dados da DEE os seguintes procedimentos foram realizados:

- a) limpeza e padronização dos dados (eliminar caracteres especiais, como “ ”, ç, ã, á, à, í, õ, entre outros);
- b) transformação de todas as letras da base de dados em maiúscula;
- c) substituição de células em branco, Não atribuído e com # por ?;
- d) substituição de (,) por ( \_ ); de ( . ) por ( ); e de espaço por ( \_ ).

No pré-processamento dos dados, o pesquisador adotou os seguintes procedimentos:

- a) desmembramento da variável unidade de leitura em três variáveis: dia de leitura, local e rota (para isso foi aplicado um algoritmo com a seguinte estrutura: {Dia\_Leitura =MUDAR(D2;3;6;"\_D")},{Rota\_Leitura =MUDAR(D2;1;6;"R\_")},{Local\_Leitura =MUDAR(EXT.TEXTO(D3;3;4);1;0;"L\_")};
- b) exclusão da variável unidade de leitura (após separar dia, local e rota);
- c) cálculo da quantidade de retorno do SVC para instalações com resultado Procedente, Improcedente e Impedido;
- d) exclusão das células: % Proced; % Imped; Qtde. Notas; Notas c/ Calc; Notas s/ Calc.;
- e) exclusão das instalações em que o campo RETORNO DO SVC = CANCELADO;
- f) criação de uma coluna LINHAS A SEREM EXCLUÍDAS para excluir as instalações com mais de uma irregularidade encontrada;
- g) definição das LINHAS A SEREM EXCLUÍDAS.
- h) exclusão das linhas com o texto "EXCLUIR LINHA";
- i) exclusão da coluna LINHAS A SEREM EXCLUÍDAS;
- j) transformação das células Cod\_SVC {=MUDAR(T2;1;0;"Cod\_")};
- k) criação da coluna DataMotivo\_DataENota\_dias {Data\_Emissao\_Nota - Data\_Motivo};
- l) criação da coluna DataMotivo\_DataInspecao\_dias {Data\_Inspecao - Data\_Motivo};

- m) criação da coluna DataENota\_DataInspecao\_dias {Data\_Inspecao - Data\_ Emissao\_Nota };
- n) exclusão da coluna número da nota;
- o) exclusão dos campos: Data do Motivo, Data Emissao Nota, Data Inspeção, Objeto de Ligação, Instalação, Motivos;
- p) criação da coluna NIE (número de irregularidades encontradas – vinculado ao Retorno do SVC).

Outro fator que causou dificuldade na preparação foi a disponibilidade dos dados pela DEE em arquivo digital no formato de planilha eletrônica (arquivo XLS – Excel). Após a limpeza e o pré-processamento dos dados, o arquivo eletrônico teve que ser convertido para o formato arquivo separado por vírgula (CSV) e, por fim, padronizado no formato ARFF.

Um fato relevante considerado antes do processamento dos dados foi que as máquinas de aprendizagem de *data mining* são desenvolvidas para aprender com os exemplos (*instances*) da base de dados em análise, considerando as variáveis mais relevantes para a tomada de decisão. Assim, bases de dados com variáveis pouco relevantes em relação ao objeto de aprendizagem podem comprometer o desempenho dos algoritmos, gerando um efeito negativo na análise (resultado pouco significativo).

Como o objetivo do processamento usando a mineração de dados era gerar resultados significativos e relevantes, o passo seguinte à preparação dos dados foi selecionar as variáveis mais relevantes do banco de dados.

De acordo com Fayyad, Piatetsky-Shapiro e Smyth (1996) e Witten e Frank (2005), a melhor forma de selecionar variáveis relevantes na solução de um problema é fazê-lo manualmente, baseando-se no conhecimento do problema a ser solucionado. Entretanto, isso exige um profundo conhecimento do problema, o que muitas vezes é subjetivo e não pode ser validado

estatisticamente. Além dessa forma, os autores destacam que a seleção poderia ser feita por meio de métodos automáticos, utilizando métricas estatísticas ou algoritmos “especialistas”.

Na análise dos dados coletados na DEE, as duas formas de selecionar variáveis foram utilizadas, ou seja, tanto a manual quanto a automática. Na manual, os colaboradores A/DEE e B/DEE julgaram que três variáveis poderiam ser retiradas da análise. As variáveis retiradas manualmente com suas respectivas justificativas são:

- a) GR\_EXEC\_INSPECAO: apenas as informações do serviço da própria DEE seriam processadas (não seriam analisados serviços prestados por terceiros);
- b) EXECUTOR\_INSPECAO: há um número muito grande de executores das ordens de inspeção, o que poderia dificultar a análise dos dados; e
- c) RETORNO\_SVC: desconsiderar as instalações inspecionadas em que essa variável presente como retorno IMPEDIDO.

Após o filtro manual, utilizaram-se algoritmos de seleção automática presentes no *software Weka* para selecionar as variáveis. O objetivo da seleção automática era determinar quais seriam as variáveis mais relevantes. Para isso, dois conjuntos distintos de algoritmos foram usados: a) avaliou um subconjunto de variáveis; e b) avaliou individualmente as variáveis.

Na avaliação das variáveis por subconjunto, foi usado o algoritmo *CfsSubsetEval*, como proposto por Witten e Frank (2005). Esse algoritmo mensura a habilidade preditiva de cada variável individualmente e o grau de redundância entre elas, priorizando aquelas que são altamente correlacionadas

com a variável resposta (no caso, foi usado como resposta a variável RESULTADO\_SVC), mas com baixa correlação com as outras variáveis.

O algoritmo *CfsSubsetEval* no Weka está associado a diferentes métodos de busca, que, de acordo com Witten e Frank (2005), examinam o "espaço" da variável de resposta para encontrar um bom subconjunto de variáveis "explicativas". Nesse caso, a qualidade da resposta é mensurada pelo subconjunto do atributo que se está avaliando (RETORNO\_SVC). Os métodos de busca associados ao algoritmo *CfsSubsetEval* usados na seleção das variáveis são representados no Quadro 6.

Todos os seis métodos de busca associados ao algoritmo *CfsSubsetEval* foram utilizados na identificação do subconjunto de variáveis que melhor definisse a variável resposta (RETORNO\_SVC). As do resumo do conjunto de variáveis selecionadas pelos métodos de busca são representadas no Quadro 7. O APÊNDICE A apresenta o resultado estatístico e detalhados dos métodos de busca.

Avaliando os subconjuntos apresentados pelos métodos de busca (Quadro 7), pode-se afirmar que apenas o método *RandomSearch* (HALL; SMITH, 1999) apresentou resultado diferente dos demais métodos. Enquanto os outros cinco métodos selecionaram como subconjunto de variáveis, COD\_SVC e NIE, o *RandomSearch* considerou como relevante DIA\_LEITURA, COD\_SVC e NIE.

Para avaliar as variáveis individualmente, foi testado o algoritmo *Ranker*, como proposto por McGregor et al. (2004). De acordo com os autores, a avaliação simples de variáveis é uma forma de ordená-las individualmente e selecionar as variáveis que geram os melhores resultados para variável resposta (RETORNO\_SVC). Assim, como na avaliação por subconjunto, na avaliação simples foram utilizados diferentes métodos de busca associados ao algoritmo

*Ranker*. O Quadro 8 representa os métodos de busca relacionados a esse algoritmo e suas respectivas características.

Quadro 6 Características dos Métodos de Busca associados ao algoritmo *CfsSubsetEval*, usados na análise dos dados da distribuidora em estudo

MÉTODOS DE BUSCA	CARACTERÍSTICAS
<i>BestFirst</i>	Pesquisa o "espaço" de subconjunto de atributos por " <i>greedy hillclimbing augmented</i> " com facilidade de recuo. <i>Best first</i> pode começar com o conjunto vazio de atributos e realizar buscas para frente ou começar com todo o conjunto de atributos e pesquisar para trás, ou iniciar em qualquer ponto e pesquisar em ambas as direções (considera todas as adições e subtrações de atributos simples em um determinado ponto).
<i>ExhaustiveSearch</i>	Executa uma pesquisa exaustiva através do "espaço" do subconjunto de atributos a partir do conjunto vazio de atributos. Apresenta o melhor subconjunto encontrado.
<i>GeneticSearch</i>	Executa uma busca utilizando o algoritmo genético simples.
<i>GreedyStepwise</i>	Executa uma pesquisa <i>greedy</i> para frente ou para trás no espaço de subconjunto do atributo. Pode começar com todos ou nenhum dos atributos de um ponto arbitrário no espaço, sendo que a adição/eliminação de todos os atributos restantes resulta em uma diminuição na avaliação. Também pode produzir uma lista ordenada de atributos por atravessar o espaço de um lado para o outro e elencar a ordem que os atributos são selecionados.
<i>RandomSearch</i>	Executa uma busca aleatória no subconjunto do espaço de atributos. Se não for definido o conjunto inicial, a pesquisa começa a partir de um ponto aleatório, fornecendo o melhor subconjunto encontrado. Se o conjunto de início for informado, a pesquisa é aleatória nos subconjuntos que são tão bons ou melhores do que o ponto de início.
<i>RankSearch</i>	Utiliza um atributo/subconjunto avaliador para classificar todos os atributos. Se um subconjunto avaliador for especificado, então é utilizada uma seleção para frente gerando uma lista ordenada. A partir da lista ordenada de atributos, subconjuntos de tamanho crescente são avaliados (por exemplo: o melhor atributo, o melhor atributo mais o próximo melhor atributo, e assim por diante, até que o melhor atributo conjunto seja relatado).

Fonte: Organizado pelo autor a partir Goldberg (1989), Liu e Setiono (1996) e Witten e Frank (2005)

Quadro 7 Conjunto de variáveis selecionadas pelos métodos de busca associados ao algoritmo *CfsSubsetEval*, usados na análise dos dados da distribuidora em estudo

MÉTODOS DE BUSCA	VARIÁVEIS SELECIONADAS
<i>BestFirst</i>	COD_SVC; NIE
<i>ExhaustiveSearch</i>	COD_SVC; NIE
<i>GeneticSearch</i>	COD_SVC; NIE
<i>GreedyStepwise</i>	COD_SVC; NIE
<i>RandomSearch</i>	DIA_LEITURA; COD_SVC; NIE
<i>RankSearch</i>	COD_SVC; NIE

Fonte: Resultado da pesquisa

Quadro 8 Características dos Métodos de Busca associados ao algoritmo *Ranker*, usados na análise dos dados da distribuidora em estudo

MÉTODOS DE BUSCA	CARACTERÍSTICA
<i>ReliefAttributeEval</i> <b>A1</b>	Avalia o valor de um atributo por amostragem repetitiva em um exemplo e considera o valor do dado para o atributo mais próximo do exemplo da mesma classe e de classes diferentes. Funciona tanto em classe de dados discretos quanto contínuos.
<i>InfoGainAttributeEval</i> <b>A2</b>	Avalia o valor de um atributo pela medição do ganho da informação com relação à classe.
<i>ChiSquareAttributeEval</i> <b>A3</b>	Avalia o valor de um atributo calculando o valor da estatística qui-quadrado em relação à classe.
<i>GainRatioAttributeEval</i> <b>A4</b>	Avalia o valor de um atributo pela medição da taxa de ganho em relação à classe.
<i>SymmetricalUncertAttributeEval</i> <b>A5</b>	Avalia o valor de um atributo através da medição da incerteza simétrica em relação à classe.

Fonte: Organizado pelo autor a partir Hall e Holmes (2003), Kira e Rendell (1992), Kononenko (1994) e Robnik-Sikonja e Kononenko (1997)

Os cinco métodos de busca associados ao algoritmo *Ranker* foram utilizados na avaliação simples das variáveis. O Quadro 9 representa o resultado de cada método de busca e a ordem das variáveis que geram os melhores resultados para a variável resposta (RETORNO\_SVC).

Quadro 9 Métodos de busca e classificação individual das variáveis, de acordo com o algoritmo *Ranker*, usados na análise dos dados da distribuidora em estudo

<b>MÉTODOS DE BUSCA</b>	<b>CLASSIFICAÇÃO DAS VARIÁVEIS</b>
<i>ReliefFAttributeEval</i> <b>A1</b>	a) COD_SVC b) LOCAL_LEITURA c) ROTA_LEITURA d) DIA_LEITURA e) TENSAO
<i>InfoGainAttributeEval</i> <b>A2</b>	a) COD_SVC b) NIE c) LOCAL_LEITURA d) DIA_LEITURA e) ROTA_LEITURA
<i>ChiSquareAttributeEval</i> <b>A3</b>	a) COD_SVC b) NIE c) LOCAL_LEITURA d) DIA_LEITURA e) ROTA_LEITURA
<i>GainRatioAttributeEval</i> <b>A4</b>	a) COD_SVC b) TENSAO c) NIE d) CLASSE e) RAMO_ATIVIDADE
<i>SymmetricalUncertAttributeEval</i> <b>A5</b>	a) COD_SVC b) NIE c) TENSAO d) LOCAL_LEITURA e) RAMO_ATIVIDADE

Fonte: Resultado da pesquisa

Analisando os resultados representados no Quadro 9 pode-se afirmar que cada método de busca ranqueou de modo distinto as variáveis que geravam os melhores resultados para RETORNO\_SVC. Apesar dessa diferença, algumas variáveis assumiram a mesma ordem em todos os métodos testados, como é o caso de: a) COD\_SVC – que assume sempre a primeira posição na classificação; b) NIE – que assume a segunda posição na classificação de três algoritmos; e c) RAMO\_ATIVIDADE – com a quinta posição na classificação de dois algoritmos.

Comparando os resultados representados no Quadro 7 com os do Quadro 9, pode-se afirmar que, mesmo usando algoritmos de seleção com características distintas, as variáveis que melhor definiriam o RETORNO\_SVC eram praticamente as mesmas.

Mas, apesar do resultado encontrado, os colaboradores A/DEE, B/DEE e C/DEE acreditavam que usar apenas aquelas variáveis ranqueadas como base para a seleção e liberação de inspeções, poderia gerar resultados práticos pouco significativos. Essa afirmação, segundo os colaboradores, poderia ser justificada pela experiência prática que eles já haviam acumulado sobre o problema.

Considerando o exposto e a diversidade de ordenação de cada um dos métodos apresentados, o pesquisador e os colaboradores criaram uma forma de classificação das variáveis (M: manter na análise e E: excluir da análise), que é representada no Quadro 10.

Quadro 10 Classificação das variáveis (M ou E), de acordo com os métodos de busca associados ao algoritmo *Ranker*, usados na análise dos dados da distribuidora em estudo

VARIÁVEIS	A 1	A 2	A 3	A 4	A 5	TOT. M	TOT. E	RESULTADO
TENSAO	M	M	M	M	M	5	0	Manter na análise
DIA_LEITURA	M	M	M	M	M	5	0	Manter na análise
LOCAL_LEITURA	M	M	M	M	M	5	0	Manter na análise
ROTA_LEITURA	M	M	M	E	E	3	2	Manter na análise
CLASSE	M	M	M	M	M	5	0	Manter na análise
RAMO_ATIVIDADE	E	M	M	M	M	4	1	Manter na análise
NUMERO_FASES	E	E	E	M	E	1	4	Excluir da análise
MOT_ANTIGO	E	E	E	E	E	0	5	Excluir da análise
DATAMOTIVO_ DATAENOTA_ DIAS	E	E	E	E	E	0	5	Excluir da análise
DATAENOTA_ DATAINSECAO_ DIAS	E	E	E	E	E	0	5	Excluir da análise
COD_SVC	M	M	M	M	M	5	0	Manter na análise
NIE	M	E	E	E	M	2	3	Excluir da análise

Fonte: Resultado da pesquisa

A classificação das variáveis representadas no Quadro 10, além de considerar todas as variáveis, usou como referência os cinco métodos de busca (de A1 a A5) associados ao algoritmo *Ranker* para gerar dois resultados possíveis para cada variável: a) “Manter na análise” (M) e b) “Excluir da análise” (E). As variáveis classificadas com a letra “M” foram aquelas que assumiram classificações de 1 a 8, de acordo com cada método de busca. Já as classificadas com a letra “E”, foram aquelas que obtiveram classificações de 9 a 12. O critério para classificar as variáveis em “M” ou “E” considerou a prática dos colaboradores da DEE e ponderações sobre os resultados apresentados pelos algoritmos.

Após a classificação, apurou-se o total de “M” (TOT. M) e o total de “E” (TOT. E) atribuído a cada variável. Para estabelecer quais variáveis deveriam ser mantidas ou excluídas da análise, adotou-se o mesmo critério da classificação anterior, as oito variáveis com o maior número de letra “M” foram mantidas na análise e as demais foram excluídas da análise.

Por fim, avaliando o resultado do Quadro 10, tanto os colaboradores da DEE quanto o pesquisador, julgaram que as variáveis que receberam a indicação de serem mantidas na análise, realmente eram condizentes com a realidade do problema. Da mesma forma, as variáveis que receberam indicação para serem excluídas faziam pouco sentido na atividade prática dos colaboradores responsáveis pela seleção e liberação de inspeção.

Antes de tomar a decisão final em relação ao resultado do Quadro 10, os colaboradores B/DEE e C/DEE ainda fizeram as seguintes ponderações:

- a) a variável NUMERO\_FASES, apesar ter recebido indicação de ser excluída da análise, deveria ser mantida por representar a complexidade das inspeções a serem realizadas e determinar, por

exemplo, qual o tipo de equipe de campo deveria ser enviado para realizar a inspeção;

- b) a variável NIE (número de irregularidades encontradas) deveria ser excluída da análise porque, mesmo tendo se mostrado relevante para os métodos de busca, o dado dessa variável é calculado somente após as inspeções serem efetivamente realizadas. Ou seja, não há como saber o número de irregularidades antes das inspeções serem executadas;
- c) outra variável que deveria ser excluída é a COD\_SVC. A justificativa é a mesma usada para a variável NIE. O dado que alimenta essa variável só será conhecido após a realização da inspeção de campo.

Considerando os argumentos dos colaboradores e ponderando os benefícios das variáveis, usou-se o recurso da seleção manual e a variável NUMERO\_FASES foi mantida na análise, ao passo que as variáveis COD\_SVC e NIE foram excluídas da análise.

Após essa última rodada de análise, os colaboradores DEE e o pesquisador concluíram que apenas sete das dezesseis variáveis deveriam ser mantidas na análise.

A partir dos resultados encontrados, os colaboradores da DEE e pesquisador concluíram que a análise das variáveis foi extremamente relevante porque possibilitou determinar, experimentalmente, quais variáveis realmente são significativas para a seleção e liberação de inspeções. Para os colaboradores da DEE, essa informação foi importante porque, apesar de trabalhar todos os dias com o problema eles não tinham a noção da importância das variáveis. Sendo que em alguns casos, variáveis que foram excluídas da análise eram usadas como referência para a decisão de selecionar e liberar inspeções.

O pesquisador argumentou que, mesmo considerando a relevância do conhecimento descoberto, ainda era necessário testar se as variáveis mantidas na análise seriam suficientes para gerar bons resultados na recuperação das perdas comerciais. O tópico a seguir apresenta o plano de ação seguido pela equipe da pesquisa-ação.

#### **5.4.4 Plano de ação do Ciclo PA-3**

Considerando os resultados obtidos com a análise dos dados, o objetivo principal do plano de ação do Ciclo PA-3 era validar as variáveis selecionadas.

Para atingir o objetivo proposto, a equipe estabeleceu que as seguintes ações deveriam ser realizadas:

- a) considerando a mesma base de dados usada para selecionar as variáveis, a equipe deveria testar e validar o resultado gerado pelas variáveis;
- b) a equipe deveria testar e validar o resultado gerado pelas variáveis selecionadas, considerando uma segunda base de dados, com dados diferentes da usada anteriormente.

Nas duas ações propostas, as variáveis selecionadas seriam consideradas válidas se:

- a) o percentual de classificação correta dos dados utilizados fosse maior do que o percentual médio de resultado PROCEDENTE da distribuidora DEE; e

- b) a partir dos resultados encontrados, fosse possível gerar um gráfico de alocação das equipes de campo, semelhante ao representado na Figura 23.

Após estabelecer as ações a serem desenvolvidas e o critério de avaliação das ações, a equipe iniciou a implantação, que é descrita no tópico a seguir.

#### **5.4.5 Implementação do Ciclo PA-3**

A validação das variáveis selecionadas poderia ser realizada de diversas formas. Entretanto, considerando o volume de dados envolvido nessa análise e a facilidade de testar diversos algoritmos, também se optou por utilizar a mineração de dados para validar as variáveis selecionadas.

Análises por *data mining* são processos em que a complexidade aumenta de acordo com a necessidade de resultados mais detalhados e precisos. Mas, nem sempre é necessário usar algoritmos complexos para obter resultados precisos. Por esse motivo, a validação das variáveis começou usando os algoritmos simples e, de acordo com a necessidade de resultados mais detalhados, algoritmos mais complexos seriam utilizados.

Três diferentes técnicas de mineração de dados foram utilizadas. De acordo com Fayyad, Piatetsky-Shapiro e Smyth (1996), Han, Kamber e Pei (2011) e Witten e Frank (2005) as técnicas utilizadas possuem, de modo geral, as seguintes características:

- a) classificação: representa de um conjunto de modelos que descrevem e diferenciam classes ou conceitos, com o propósito de utilizar o

modelo para prever uma classe de objetos que ainda não foram classificados;

- b) associação: representa padrões descritivos ou conjunto de valores que podem ser úteis para estabelecer relações entre itens;
- c) agrupamento ou *clustering*: identifica aglomerações que descrevem os dados a partir de padrões de similaridade entre objetos.

#### 5.4.5.1 Teste e validação das variáveis, usando a mesma base de dados usada na seleção

Para testar e validar as variáveis, a classificação foi a primeira técnica de mineração de dados utilizada. Os algoritmos usados e suas respectivas características são apresentados no APÊNDICE B. O resumo com as estatísticas de todos os algoritmos de classificação são apresentadas no APÊNDICE C.

Analisando os resultados estatísticos dos algoritmos de classificação, pode-se afirmar que as variáveis selecionadas conseguiram descrever e diferenciar os resultados da variável de resposta (RETORNO\_SVC). Dentre os onze algoritmos de classificação que foram utilizados na primeira técnica, o percentual de classificação correta das mais de 8.300 instalações da base de dados, foi maior do que o percentual médio de resultado PROCEDENTE da distribuidora DEE, como pode ser visto na Tabela 2.

Tabela 2 Percentual de classificação correta e precisão média ponderada dos algoritmos de classificação, usado no teste e validação das variáveis, da distribuidora em estudo

ALGORITMO	CLASSIFICAÇÃO	PRECISÃO MÉDIA
	CORRETA	PONDERADA
	%	%
Decorate	79,9	80,1
AODE	77,1	77,3
PART	76,3	76,3

“continua”

Tabela 2 “conclusão”

ALGORITMO	CLASSIFICAÇÃO	PRECISÃO MÉDIA
	CORRETA	PONDERADA
	%	%
REPTree	75,6	76,1
J48	73,8	74,3
Logistic	73,3	73,5
SMO	71,4	71,8
<i>BayesNet</i>	69,1	70,9
LogitBoost	68,7	68,8
OneR	65,4	65,9
MultlayerPerceptron	51,6	75,0

Fonte: Resultados da pesquisa

Apesar de todos os algoritmos terem apresentado resultados maiores do que a média de resultado da DEE, alguns algoritmos tiveram resultado mais significativo do que outros, como demonstrado na Tabela 2. Além disso, pode-se afirmar que o algoritmo o *Decorate* foi o que apresentou o maior percentual de classificação correta (79,9%), ou seja, em termos absolutos significa dizer que foi possível classificar corretamente 6.649 inspeções de um total de 8.315. Além disso, esse foi o algoritmo que apresentou a maior precisão média ponderada (80,1%)

Por outro lado, o algoritmo com o menor percentual de classificação correta foi o *Multlayer Perceptron* (51,6%) e o que apresentou a menor precisão média ponderada foi o *OneR* (65,9%).

A segunda técnica de mineração de dados utilizada no teste e validação das variáveis foi a de associação. Os algoritmos usados e suas respectivas características são apresentados no APÊNDICE D. O resumo com as estatísticas dos três os algoritmos de associação são apresentadas no APÊNDICE E.

Diferentemente dos algoritmos de classificação, que predisseram quais inspeções foram classificadas de forma correta ou não, os algoritmos de associação descreveram quais foram os padrões que melhor representaram um dado conjunto de associação.

De modo geral, pode-se afirmar que os algoritmos *Apriori* e *FilteredAssociator* apresentaram resultados muito semelhantes para as dez principais regras de associação apresentadas. Dentre as regras apresentadas, as duas abaixo serão usadas como exemplo:

- a) Retorno\_SVC = PROCEDENTE: 4.286 inspeções ==> Tensao = BT: 4.274 inspeções; conf:(1); e
- b) Tensao = BT; Ramo\_Atividade = RA\_380: 5.672 inspeções ==> Classe=RESIDENCIAL: 5.633 inspeções; conf:(0.99).

Como se pode perceber, a análise do resultado dos algoritmos acima destacados é um pouco diferente. Analisando a regra (a), pode-se afirmar que dos 4.286 resultados procedentes do RETORNO\_SVC, 4.274 inspeções foram realizadas em baixa tensão (BT), com confiança de 100% na regra. Já sobre a regra (b), pode-se afirmar que das 5.672 inspeções realizadas em instalações de baixa tensão (BT) com ramo de atividade RA\_380, 5.633 inspeções pertenciam à classe residencial, com confiança de 99% na regra.

Apesar de apresentar regras diferentes das apresentadas pelos algoritmos *Apriori* e *FilteredAssociator*, o algoritmo *PredictiveApriori* segue os mesmos critérios para apresentar as regras que melhor representaram um dado conjunto de associação, como pode ser visto na lista:

- a) Local\_Leitura = L\_0530; Ramo\_Atividade = RA\_380; Retorno\_SVC = PROCEDENTE: 293 inspeções ==> Tensao = BT: 293 inspeções, sendo acc: (0.99499);
- b) Dia\_Leitura = 11\_D; Retorno\_SVC = PROCEDENTE: 288 inspeções ==> Tensao = BT: 288 inspeções, sendo acc:(0.99499);

- c) Rota\_Leitura = R\_30; Ramo\_Atividade = RA\_380; Retorno\_SVC = PROCEDENTE: 203 inspeções ==> Tensao = BT: 203 inspeções, sendo acc: (0.99497); e
- d) Dia\_Leitura = 13\_D; Local\_Leitura = L\_8015; Retorno\_SVC = PROCEDENTE: 190 inspeções ==> Tensao = BT: 190 inspeções, sendo acc:(0.99496).

Analizando a regra apresentada no item (a), por exemplo, pode-se afirmar que das 293 inspeções realizadas no local de leitura L\_0503, com ramo de atividade RA\_380 e RETORNO\_SVC procedente, todas as 293 eram de baixa tensão, com índice de precisão de 99,5%.

Apesar da relevância dos algoritmos de associação, os colaboradores da DEE acharam que essa forma de organização dos resultados, além de ser difícil de ser interpretada, demandaria uma ampla e complexa análise para determinar qual deveria ser a regra mais completa e que melhor representasse, por exemplo, as inspeções com RETORNO\_SVC procedente.

A terceira e última técnica de mineração de dados utilizada no teste e validação das variáveis foi a de agrupamento. Os algoritmos usados e suas respectivas características são apresentados no APÊNDICE F. O resumo com as estatísticas dos três algoritmos de associação são apresentadas no APÊNDICE G.

De modo geral, pode-se afirmar que, apesar de não predizerem quais inspeções foram classificadas de forma correta, como ocorreu na técnica de classificação, os algoritmos de agrupamento identificaram quais os resultados das variáveis que apresentava maior similaridade com a variável resposta RETORNO\_SVC.

Analizando o algoritmo *SimpleKMeans*, por exemplo, pode-se afirmar que 66% dos casos em que o RETORNO\_SVC foi procedente, as instalações

inspeccionadas tinham as seguintes características: Tensao = BT; Dia\_Leitura = 13\_D; Local\_Leitura = L\_8015; Rota\_Leitura = R\_30; Classe = RESIDENCIAL; Ramo\_Atividade = RA\_380.

Já de acordo com o algoritmo *FarthestFirst*, pode-se afirmar que 80% das inspeções em que o RETORNO\_SVC foi improcedente, as instalações tinham as seguinte características: Tensão = BT; Dia\_Leitura = 03\_D; Local\_Leitura = L\_0631; Rota\_Leitura = R\_12; Classe = RESIDENCIAL; Ramo\_Atividade = RA\_380.

De acordo com o colaborador A/DEE, a técnica de agrupamento gerou resultados mais interessantes para a seleção de inspeções do que as outras técnicas. E segundo ele, com as características dos *clusters*, seria bem mais fácil selecionar quais as instalações teriam aquelas características em específico.

Adotando o resultado do algoritmo *SimpleKMeans* como referência, seria possível gerar um gráfico de alocação das equipes de campo, como representado na Figura 24.

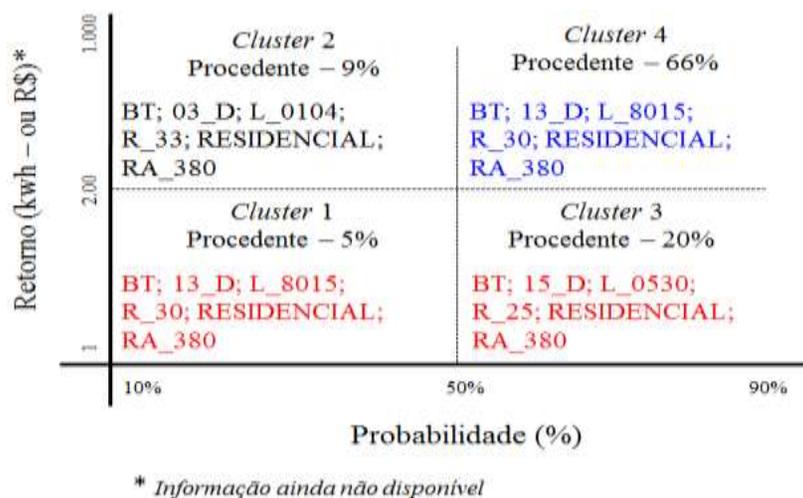


Figura 24 Alocação das equipes de campo da distribuidora em estudo, usando como referência o resultado do algoritmo *SimpleKMeans*

Fonte: Resultado da pesquisa

O resultado gerado pela técnica de agrupamento (Figura 24) possibilitaria que a DEE alocasse as equipes de inspeção de acordo com as características das instalações suspeitas. Ou seja, pelos resultados da Figura 24, os colaboradores responsáveis pela seleção de instalações iriam priorizar as que vivessem as características do *Cluster 4* (Figura 24).

Entretanto, de acordo com o pesquisador, a Figura 24 ainda não estava completa porque informações sobre o retorno ainda não haviam sido processadas. Por esse motivo, não seria possível diferenciar, por exemplo, qual o retorno esperado do *cluster 1* ou do *cluster 4*, sabendo que eles possuem as mesmas característica.

Assim, antes de realizar o teste e a validação das variáveis usando uma nova base de dados, o pesquisador sugeriu aos colaboradores da distribuidora DEE que dados como probabilidades e consumo médio mensal por classe e por número de fase fossem adicionados à nova base de dados.

#### **5.4.5.2 Teste e validação das variáveis, usando a uma base de dados diferente da usada na seleção**

Motivados com os resultados do teste anterior, os colaboradores da DEE levantaram os dados solicitados pelo pesquisador. A Tabela 3 representa o consumo médio mensal por classe de consumidores e por número de fases, no período de 12 meses, em kWh.

De acordo com o colaborador C/DEE, os valores referentes aos consumos médios (Tabela 3) eram apenas uma aproximação dos valores médios obtidos em um período de 12 meses. Desse modo, em novas aplicações, esses valores deveriam ser revisados e atualizados com frequência.

Tabela 3 Consumo médio mensal por classe de consumidores e por número de fases, no período de 12 meses, em kWh, da distribuidora em estudo

CLASSES	NÚMERO DE FASES	CONSUMO MÉDIO
	%	kWh
Residencial	Monofásico	103,3
	Bifásico	170,5
	Trifásico	347,9
Industrial	Monofásico	115,9
	Bifásico	244,0
	Trifásico	1.112,6
Comércio/Serviço	Monofásico	129,4
	Bifásico	290,6
	Trifásico	1.420,9
Rural	Monofásico	200,0
	Bifásico	401,8
	Trifásico	1.207,5
Poder público	Monofásico	139,4
	Bifásico	442,5
	Trifásico	2.406,5
Serviços públicos	Monofásico	255,5
	Bifásico	975,4
	Trifásico	5.228,2

Fonte: Organizado pelo autor a partir de informações da DEE

De posse das médias de consumo, estavam faltando apenas os dados relativos às probabilidades para completar o gráfico de alocação das equipes de campo (Figura 23). Questionado sobre esses, o colaborador C/DEE informou que o ERP da DEE ainda não estava preparado para gerar probabilidades referentes às inspeções realizadas. E, segundo ele, para validar a análise das variáveis, as probabilidades poderiam ser geradas, fora do ERP da empresa, a partir de uma amostra.

O pesquisador, diante dessa situação, concordou com o colaborador e, para evitar qualquer pré-julgamento dos dados, sugeriu que a amostra fosse selecionada pelos colaboradores da DEE e que todas as probabilidades fossem geradas por eles. Nesse caso, o pesquisador ficaria responsável pelo teste e

validação das variáveis. Os colaboradores concordaram e calcularam as probabilidades que foram incorporadas aos dados.

Considerando que a técnica de mineração de dados por agrupamento foi a que gerou resultados mais satisfatórios, segundo os colaboradores da DEE, o pesquisador sugeriu que nesse teste fosse utilizado apenas o algoritmo de agrupamento *SimpleKMeans*. A justificativa para a proposta, de acordo com o pesquisador, foi de que nem sempre é necessário usar algoritmos complexos para obter resultados precisos. Ou seja, apesar da simplicidade do algoritmo de agrupamento sugerido, esperava-se resultados precisos. Além disso, caso o algoritmo não gerasse resultados satisfatórios, outros algoritmos e outras técnicas poderiam ser usados. Os colaboradores concordaram com a proposta e comentaram que também tinham preferência no uso do algoritmo, pelo resultado gerado.

A amostra selecionada pelos colaboradores da DEE contava com 2.499 instalações, de um município da unidade regional DC. Apesar de já terem sido inspecionadas pelas equipes de campo da DEE, apenas os colaboradores da DEE sabiam quais eram os resultados das inspeções realizadas.

Assim como ocorreu na base de dados anterior, o algoritmo *SimpleKMeans* foi usado para identificar os padrões que apresentavam maior similaridade com a variável resposta RETORNO\_SVC. Considerando os dados da amostra, os seguintes resultados foram encontrados:

- a) 45,5% apresentaram características como: BT; 03\_D; L\_0631; R\_12; RESIDENCIAL; RA\_380; IMPROCEDENTE. Consumo médio mensal (CM): 103,3 kWh. Probabilidade (Prob.): 21%;
- b) 37,5% apresentaram características como: BT; 13\_D; L\_8015; R\_30; RESIDENCIAL; RA\_380; PROCEDENTE. CM: 103,3 kWh. Prob.: 34,5%;

- c) 11,4% apresentaram características como: MT; 44\_D; L\_0109; R\_02; OUTROS\_SERVICOS\_E\_OUTRAS\_ATIV; RA\_999; PROCEDENTE. CM: 1.420 kWh. Prob.: -x-%;
- d) 5,6% dos casos possuíam variáveis com características como: BT; 15\_D; L\_0530; R\_25; RESIDENCIAL; RA\_380; PROCEDENTE. CM: 103,3 kWh. Prob.: 34,3%.

A partir dos resultados supracitados, foi possível gerar o gráfico de alocação das equipes de campo, como representado na Figura 25.

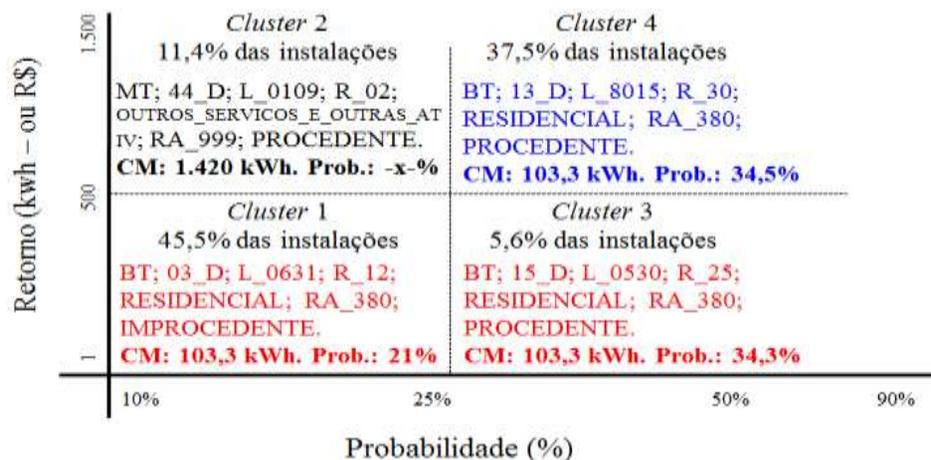


Figura 25 Alocação das equipes de campo da distribuidora em estudo, de acordo com *SimpleKMeans*, usando como referência uma nova base de dados para validação da seleção das variáveis

Fonte: Resultado da pesquisa

Avaliando o resultado da Figura 25, os colaboradores da DEE e o pesquisador concluíram que as características nos agrupamentos se aproximavam da realidade das inspeções realizadas. Entretanto, um fato que chamou a atenção foi que os percentuais das instalações dos agrupamentos se diferiam dos realmente obtidos pela DEE. O pesquisador respondeu que, uma

das possíveis justificativas para essa diferença, é que as instalações foram agrupadas por semelhança. Assim, para o algoritmo, uma dada instalação se agrupa, por semelhança, às instalações do *Cluster* 4, por exemplo. Porém na prática, ela pertence ao *Cluster* 3. Ou seja, mesmo com o delineamento dos grupos, há uma área de intercessão entre eles que pode resultar na diferença entre o resultado apresentado pelo algoritmo e o resultado das inspeções efetivamente realizadas.

Dando sequência na análise, das 2.499 inspeções analisadas, o algoritmo classificou como PROCEDENTE 1.234 instalações (49,37%) e como IMPROCEDENTE 1.265 instalações (50,62%). Entretanto, comparando com o resultado encontrado pelas equipes de campo, foram encontradas 1.420 (56,8%) instalações com resultado PROCEDENTE e 1.079 (43,2%) instalações com resultado IMPROCEDENTE.

Os colaboradores da DEE questionaram a divergência entre o resultado do algoritmo e o que as equipes de campo efetivamente encontraram. Para dirimir qualquer dúvida, o pesquisador sugeriu que fosse utilizado uma matriz de confusão, seguindo as propostas de Kohavi e Provost (1998), Santos e Azevedo (2005) e Weiss e Kulikowsky (1991).

De acordo com os autores, a matriz de confusão é usada para comparar o número de classificações corretas com o número de previsões efetuadas para cada caso, considerando um conjunto de exemplos. A Figura 26 representa o modelo da matriz de confusão e as equações de 1 a 6 apresentam as estatísticas relacionadas à matriz.

		PREVISÃO	
		Positivo	Negativo
REAL	Positivo	TP	FN
	Negativo	FP	TN

Figura 26 Modelo da matriz de confusão usado na validação das variáveis da distribuidora em estudo

Fonte: Organizado pelo autor a partir de Kohavi e Provost (1998), Santos e Azevedo (2005) e Weiss e Kulikowsky (1991)

$$\text{Acuracidade (AC):} \quad = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (1)$$

$$\text{Verdadeiro positivo (TP):} \quad = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

$$\text{Falso positivo (FP):} \quad = \frac{FP}{FP + TN} \times 100\% \quad (3)$$

$$\text{Verdadeiro negativo (TN):} \quad = \frac{TN}{FP + TN} \times 100\% \quad (4)$$

$$\text{Falso negativo (FN):} \quad = \frac{FN}{TP + FN} \times 100\% \quad (5)$$

$$\text{Precisão (P):} \quad = \frac{TP}{TP + FP} \times 100\% \quad (6)$$

Considerando o modelo da matriz representado na Figura 26 e as equações de 1 a 6, a Figura 27 representa a matriz de confusão e suas estatísticas para os dados em análise.

		ALGORITMO	
		Procedente	Improcedente
INSPEÇÕES	Procedente	736	684
	Improcedente	498	581

ESTATÍSTICAS	
1. Acuracidade (AC):	52,7%
2. Verdadeiro positivo (TP):	51,8%
3. Falso positivo (FP):	46,2%
4. Verdadeiro negativo (TN):	53,8%
5. Falso negativo (FN):	48,2%
6. Precisão (P):	59,6%

Figura 27 Matriz de confusão e estatísticas dos dados analisados da distribuidora em estudo

Fonte: Resultados da pesquisa

Comparando o resultado real das inspeções com a previsão do algoritmo, pode-se afirmar que, pela similaridade entre as instalações, o algoritmo gerou um conjunto intermediário de resultados que pode ser descrito em duas situações:

- a) situação 1 – Falso Negativo: nessa situação, o algoritmo julgou que 684 instalações teriam resultado IMPROCEDENTE ao passo que, pelas inspeções, o resultado foi PROCEDENTE. A situação apresentada (FN: 48,2%) poderia ser prejudicial à DEE, pois ela deixaria de inspecionar instalações que estão fraudando o consumo de energia;
- b) situação 2 – Falso Positivo: nessa situação o algoritmo julgou que 498 inspeções teriam resultado PROCEDENTES, ao passo que, na verdade, o resultado real foi IMPROCEDENTE. Apesar de menos grave do que a situação 1, essa pode gerar desperdício de recursos,

pois seria o mesmo que inspecionar instalações em que o resultado será IMPROCENTE (FP: 46,2%).

Apesar das situações descritas, pode-se afirmar que o algoritmo tem uma precisão de 59,6% (P), ou seja, dentre as inspeções realizadas ele agrupou corretamente 736 instalações com resultado PROCEDENTE e 581 instalações com resultado IMPROCEDENTE (Figura 27).

De modo geral, o pesquisador considerou que o resultado gerado pelo algoritmo era significativo, pois o índice de acerto encontrado pela equipe de campo foi 56,8%. Os colaboradores da DEE, por outro lado, consideraram que o resultado do algoritmo ficou aquém do imaginado, pois se fosse considerado apenas os resultados PROCEDENTES indicados pelo algoritmo, as equipes de inspeção teriam tido um índice de acerto de 29,5%.

O pesquisador contra-argumentou dizendo que essa análise não era condizente com a realidade do problema. Segundo o pesquisador, se os colaboradores da DEE fossem avaliar o índice de acerto do algoritmo, eles deveriam usar como referência a precisão do algoritmo que foi de 59,4%. Além disso, caso o algoritmo tivesse sido utilizado para selecionar as instalações, as equipes de campo teriam inspecionado apenas 1.234 instalações e não as 2.499. Essa diferença representaria uma redução de quase 50% no número de inspeções a serem realizadas e, apesar disso, o índice de acerto da equipe de campo teria aumentado quase 5%, saltado dos atuais 56,8% para 59,4%.

Após a implementação das ações do Ciclo PA-3, o pesquisador, o coordenador/DEE e os colaboradores DEE envolvidos na pesquisa-ação se reuniram para avaliar o resultado das ações implantadas. A avaliação das ações é apresentada no tópico a seguir.

### 5.4.6 Avaliação do Ciclo PA-3

A avaliação do Ciclo PA-3 considerou os seguintes aspectos:

- a) transformação de dados em informações;
- b) uso de novos instrumentos computacionais para analisar dados;
- c) relevância das variáveis selecionadas no processo de seleção e liberação de inspeções;
- d) as dificuldades de utilizar a mineração de dados pelos colaboradores no dia a dia da DEE;
- e) novas etapas da pesquisa-ação; e
- f) as mudanças ocorridas na DEE.

Houve um consenso entre os participantes da pesquisa-ação sobre a importância e relevância do Ciclo PA-3. Segundo os colaboradores da DEE, o ciclo serviu para reforçar a necessidade de adicionar ao processo de combate às perdas não técnicas mais informação. As ações, até então implantadas na DEE, não tinham considerado o uso dos dados históricos das inspeções como fonte de informação. O pesquisador complementou afirmando que a informação gerada durante esse ciclo será fundamental para os próximos ciclos da pesquisa-ação.

De acordo com o coordenador/DEE, a disponibilização de novas tecnologias da informação como, por exemplo, o *software* utilizado, também foi muito relevante. Segundo ele, dificilmente os colaboradores da DEE conseguiriam processar tantos dados, de uma única vez, sem o auxílio computacional.

Além da agilidade no processamento dos dados, o *software* assumiu papel preponderante na seleção das variáveis que foram consideradas essenciais na seleção e liberação de inspeções. Segundo o colaborador C/DEE, dificilmente

ele conseguiria convencer o seu superior de que a combinação de apenas sete variáveis fosse suficiente para auxiliar na seleção das instalações com suspeita de irregularidade.

O pesquisador destacou também a importância do algoritmo de agrupamento. O algoritmo conseguiu identificar as características das variáveis que descreviam a similaridade entre o resultado das inspeções realizadas (PROCEDENTE e IMPROCEDENTE). E, a partir dos padrões descritos, foi possível gerar o gráfico de alocação das equipes de campo, como havia sido solicitado.

O pesquisador ressaltou ainda que, a princípio, o uso do algoritmo pode gerar uma aparente piora no índice de acerto da DEE. Entretanto, no médio prazo, o índice poderá ser compensado pela queda no número de inspeções a serem realizadas e pela precisão das equipes de campo. Atendendo, conseqüentemente, os anseios do setor de proteção da receita da DEE, que é realizar menos inspeções com maior índice de acerto. Além disso, ressaltou-se também que esse ciclo, ainda mais do que nos ciclos anteriores, foi vital para os resultados alcançados a interação entre os colaboradores e os dados processados.

Diante das ponderações feitas, os membros da pesquisa-ação julgaram que o Ciclo PA-3 foi cumprido com sucesso, sendo que todas as ações e resultados gerados foram validados.

O colaborador C/DEE chamou a atenção para a dificuldade de usar a mineração de dados no dia a dia da empresa. Pois, além de ser responsável pela atividade de seleção e liberação de inspeções, ele também desempenha outras rotinas de trabalho, o que pode inviabilizar o seu uso. O ideal, segundo ele, é que fosse criado um procedimento automatizado que gerasse as informações para os colaboradores responsáveis pela seleção e liberação das inspeções.

O pesquisador argumentou que também já havia pensado sobre isso e que gostaria de fazer a proposta do desenvolvimento de sistema de apoio à

decisão (SAD ou DSS) que fosse utilizado como suporte para alocar de equipes de campo (Perdas DSS). A justificativa para o desenvolvimento desse DSS, segundo o pesquisador, estava pautada em outros trabalhos, como por exemplo o de León et al. (2011).

Entretanto, segundo o pesquisador, para o desenvolvimento desse aplicativo, além de considerar os resultados do Ciclo PA-3, deveria ser realizado em um novo ciclo interativo da pesquisa-ação (coleta de dados, *feedback* aos participantes, análise de dados, planejamento de ações e avaliação das ações). Os colaboradores gostaram da ideia e aceitaram o desafio do desenvolvimento do *software* Perdas DSS.

Para finalizar a reunião de avaliação, o coordenador/DEE informou que iriam ocorrer mudanças na DEE e também na condução da pesquisa-ação.

Segundo ele, a DEE já havia terminado a estruturação da unidade que ficaria responsável por controlar as perdas não técnicas. Assim, a atividade de seleção e liberação de inspeções seria centralizada na unidade funcional da DEE e cada unidade regional passaria a contar com um colaborador específico para esse fim. Segundo ele, essa ação evitaria as disfunções apresentadas no Ciclo PA-1.

O coordenador informou também que ele tinha conseguido uma promoção interna na DEE e por isso além de estar mudando de setor também estava deixando o grupo de trabalho da pesquisa-ação. Com isso, quem assumiu a coordenação da pesquisa-ação na DEE foi o colaborador A/DEE. Para evitar confusões, a partir de agora o coordenador da pesquisa-ação na DEE será tratado como coordenador A/DEE.

Com a alteração na estrutura da equipe da pesquisa-ação, o colaborador C/DEE deveria ter contato direto com o pesquisador, para auxiliá-lo no processo de desenvolvimento do aplicativo Perdas DSS. O colaborador B/DEE, apesar de continuar compondo a equipe, ficaria dedicado responsável pela atividade de

seleção e liberação de inspeções e por outras funções administrativas no setor de perdas da DEE. Além disso, a equipe passou a contar também com o colaborador D/DEE.

#### **5.4.6.1 Ciclo PA-4: desenvolvimento da aplicação perdas DSS**

O Ciclo PA-4 corresponde ao ciclo interativo que desenvolveu o sistema de apoio à decisão para alocar de equipes de campo, chamado de Perdas DSS. O desenvolvimento dessa aplicação seguirá o ciclo de vida do DSS sugerido por Fleury, Agostini e Lima (1984) e demonstrado na Figura 1. Apesar de considerar passos diferentes dos usados na implantação da pesquisa-ação (Figura 3), a utilização o ciclo de vida do DSS não comprometeu o trabalho porque tanto o ciclo de vida do DSS quanto à pesquisa ação usam como elemento principal a interação entre usuário-sistema-desenvolvedor.

A descrição desse ciclo, que teve duração de 12 meses, é apresentada nos tópicos a seguir.

#### **5.4.7 Interesses e motivações – Ciclo PA-4**

Os interesses e as motivações para o desenvolvimento da aplicação Perdas DSS surgiu, naturalmente, com a evolução do trabalho de pesquisa-ação. O Ciclo PA-3 destaca com detalhes as principais motivações.

Para o coordenador A/DEE, a aplicação Perdas DSS foi mais um instrumento que passou a integrar o processo de recuperação de perdas não técnicas. De acordo com o coordenador, ele deverá ser utilizado para processar os dados históricos das inspeções realizadas e comparar com os dados dos suspeitos de irregularidade, aumentando a efetividade do combate das perdas não técnicas.

A motivação para o desenvolvimento da aplicação Perdas DSS, segundo o pesquisador, se fundamentou na necessidade de manipular, processar e analisar grandes volumes de dados, gerando informações para facilitar a alocar equipes de campo, responsáveis pela inspeção de consumidores suspeitos de fraudar o consumo de energia.

O uso do aplicativo Perdas DSS para auxiliar na tomada de decisões operacionais também foi um grande motivador. Pois, grande parte dos teóricos sobre DSS afirma que esse tipo de sistema de informação é recomendado e muito utilizado no apoio à decisões estratégicas, muitas vezes ligadas ao comando de uma organização.

#### **5.4.8 Objetivos e recursos a usados – Ciclo PA-4**

O objetivo principal do aplicativo Perdas DSS é gerar informações para que os colaboradores da DEE (responsáveis pela atividade de selecionar e liberar as inspeções) possam decidir quais consumidores suspeitos de fraudar o consumo de energia devem ser inspecionados. Ou seja, o *software* não deverá substituir os colaboradores na atividade seleção e liberação de inspeções. Ele será um instrumento facilitador da decisão (apoio).

Para garantir o objetivo da aplicação, os recursos usados foram agrupados nas dimensões humana, tecnológica e organizacional, como representado no Quadro 2.

Os recursos da dimensão humana foram compostos basicamente pelos colaboradores da DEE que iriam trabalhar com o Perdas DSS, o pesquisador e os desenvolvedores da aplicação. O desenvolvimento da aplicação Perdas DSS ficou sob a responsabilidade de uma empresa especializada em desenvolvimento de *software*. O pesquisador, que ficou responsável pelo acompanhamento do desenvolvimento da aplicação, comentou que a contratação da empresa se

justificava pelo ganho no tempo de desenvolvimento da aplicação, pela qualidade final do *software* desenvolvido e pela garantia do serviço. Além de acompanhar o desenvolvimento, o pesquisador também ficou responsável por treinar os colaboradores da DEE.

Os recursos da dimensão tecnológica tiveram origens internas e externas à DEE. Durante o desenvolvimento da aplicação, a empresa contratada disponibilizou a plataforma de desenvolvimento, *hardwares*, *softwares* e banco de dados. Após o desenvolvimento, a DEE teria que manter, com recursos internos, os *hardwares* necessários para o correto funcionamento da aplicação, *software* e banco de dados para fornecer e armazenar dados, além das redes para transferência de dados.

Por fim, com relação aos recursos da dimensão organizacional, além de uma nova estrutura física e de gestão, a DEE implantou uma nova rotina para a atividade de seleção e liberação de inspeções. E foi considerando essa nova rotina que a aplicação Perdas DSS foi desenvolvida.

#### **5.4.9 Processo decisório e Decisões-chave – Ciclo PA-4**

O processo decisório analisado está relacionado à atividade de seleção e liberação de inspeções, que é executada no setor de combate às perdas comerciais e proteção da receita da DEE. Mesmo com a mudança estrutural da unidade, o POP usado na seleção e liberação de inspeção dos consumidores suspeitos de irregularidade no consumo de energia, representado na Figura 4, foi mantido.

Segundo o colaborador C/DEE, após as mudanças estruturais, o POP passou a ter mais sentido. Por esse motivo, o processo decisório tinha sido mantido e, até então, estava vigente.

Analisando o fluxograma do POP pode-se afirmar que a decisão-chave da atividade é, e continuará sendo, determinar quais as instalações serão selecionadas para inspeção.

#### **5.4.10 Adequação das soluções propostas – Ciclo PA-4**

Sabendo que a decisão-chave da atividade não tinha sido mudada e, considerando o desenvolvimento do *software* Perdas DSS, o pesquisador sugeriu que o Procedimento Operacional Padrão (POP) da atividade de seleção e liberação de inspeções fosse atualizado. Ao incorporar o *software*, a atividade passaria a ter um POP semelhante ao representado na Figura 28.

#### **5.4.11 Projeto de desenvolvimento e de implantação – Ciclo PA-4**

O projeto de desenvolvimento do *software* Perdas DSS se fundamentou no processo operacional executado na DEE. Ou seja, seu desenvolvimento tinha como principal objetivo suportar a atividade de seleção e liberação de inspeções.

Antes de iniciar o seu desenvolvimento, o pesquisador, juntamente com os desenvolvedores, tentaram caracterizar o Perdas DSS. De acordo com a classificação proposta por Arnott e Pervan (2005, 2008), o *software* seria um *mix* entre os sistemas de apoio à decisão baseados na gestão do conhecimento e o *Data Warehousing* e *Data Mining*. Essa classificação foi importante para direcionar o desenvolvimento do *software*.

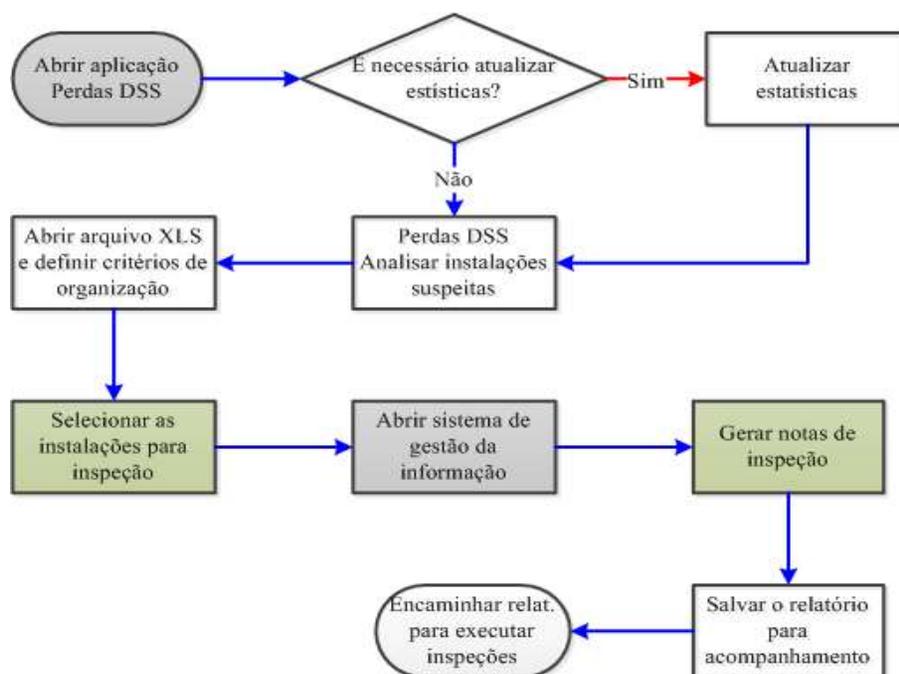


Figura 28 Procedimento Operacional Padrão (POP) da atividade de seleção e liberação de inspeção, incorporando a aplicação Perdas DSS, na distribuidora em estudo

Fonte: Organizado pelo autor a partir de entrevista realizada com colaboradores da DEE

O primeiro passo do projeto de desenvolvimento foi definir a arquitetura da base de dados. Considerando o volume de dados a ser processado, o pesquisador, juntamente com a empresa contratada para o desenvolvimento, concluíram que o *software* deveria ter um banco de dados próprio, que deveria ser alimentado com dados que são registrados diretamente no ERP da DEE.

De acordo com a empresa contratada, essa escolha seria a mais acertada e se justificava pelos seguintes motivos:

- a) volume de dados a serem processados – considerando o número de inspeções realizadas anualmente e o número de suspeitos que não

para de aumentar, com um banco de dados próprio a capacidade de processamento de dados seria exponencialmente maior do que se o *software* fizesse apenas leitura de dados a partir do ERP da DEE;

- b) segurança da informação – com o banco de dados próprio, evitaria possível perda de dados pela falta de comunicação da aplicação com o ERP da DEE, por exemplo;
- c) controle do tráfego de dados – com o banco de dados próprio, as consultas ao ERP da DEE poderiam ser programadas para serem executadas em momentos que a rede estivesse ociosas, evitando, portanto, congestionamento na rede e lentidão na transferência de dados.

Após a aprovação da decisão por parte dos colaboradores da DEE, iniciou-se o desenho da arquitetura da base de dados. Esquemáticamente, a base de dados foi projetada com arquitetura relativamente simples, com um pequeno número de tabelas interligadas e com variáveis bem definidas. A Figura 29 ilustra a arquitetura da base de dados da aplicação Perdas DSS. A arquitetura representada na Figura 29 ilustra a relação entre as tabelas que serão implementadas na base dados.

Definida a arquitetura da base de dados o passo seguinte foi definir o Sistema Gerenciador de Banco de Dados (SGBD). De acordo com os desenvolvedores da empresa contratada, apesar da arquitetura “simples” da base de dados, o SGBD deve ser robusto o bastante para suportar o armazenamento, de forma segura e confiável, de um grande volume de dados.



Dentre as várias opções de SGBD disponíveis no mercado, os desenvolvedores e o pesquisador optaram pelo *PostgreSQL*. De acordo com *The PostgreSQL Global Development Group*<sup>2</sup>, o *PostgreSQL* é um avançado sistema gerenciador de banco de dados objeto relacional (SGBDOR), de código aberto, com mais de 15 anos de desenvolvimento mundial e ativo.

Dentre as justificativas que levaram a equipe a escolher o *PostgreSQL* como SGBDOR pode-se destacar:

- a) possui arquitetura que, além de seguir padrões internacionais, é confiável e capaz de manter a integridade dos dados;
- b) possui código aberto, oferecendo o mais baixo custo total de propriedade (TCO), o que reduz de forma significativa os custos de administração, suporte e licenciamento;
- c) suporta tamanho ilimitado de registro: não impõe limites no tamanho de armazenamento dos tipos de dados (*The Oversized-Attribute Storage Technique* – TOAST – os limites de armazenamento de dados são impostos pela configuração do *hardware*);
- d) roda em todos os sistemas operacionais (GNU/Linux, Unix – AIX, BSD, HP-UX, SGI IRIX, Mac OS X, *Solaris*, Tru64 – e MS Windows);
- e) inclui a maior parte dos tipos de dados do ISO SQL:1999, incluindo *INTEGER*, *NUMERIC*, *BOOLEAN*, *CHAR*, *VARCHAR*, *DATE*, *INTERVAL*, e *TIMESTAMP*;

---

<sup>2</sup> The PostgreSQL Global Development Group. **About**. Disponível em: <<http://www.postgresql.org/about/>>. Acesso em: 10 jan. 2013.

- f) suporta o armazenamento de objetos binários, incluindo figuras, sons ou vídeos;
- g) possui interfaces nativas de programação para C/C++, Java, .Net, Perl, Python, Ruby, Tcl, ODBC; e
- h) além de vários recursos de programação disponíveis (consultas complexas; chaves estrangeiras; integridade transacional; suporte ao modelo híbrido objeto relacional; gatilhos e visões; linguagem procedural para procedimentos de armazenagem; indexação por texto; estrutura para guardar dados Georeferenciados *Post* (GIS).

Após a escolha do SGBDOR, os desenvolvedores fizeram a transcrição da arquitetura da base de dados para o PostgreSQL, que após pequenas adequações, foi validada pelo pesquisador e pelos colaboradores da DEE.

O terceiro passo do projeto de desenvolvimento foi a escolha da linguagem de programação.

Há várias definições para o termo, mas de forma bem simples pode-se dizer que a linguagem de programação é um conjunto de regras sintáticas e semânticas usadas para definir um programa de computador. Ou seja, é um método padronizado para expressar um conjunto de instruções a um computador, especificando, precisamente, sobre quais dados em quais ele vai atuar, como os dados serão armazenados ou transmitidos e quais ações devem ser tomadas sob várias circunstâncias (AHO et al., 2008; DERSHEM; JIPPING, 1995; FISCHER; GRODZINSKY, 1993; MELO; SILVA, 2003).

De acordo com os desenvolvedores da empresa contratada, uma característica muito desejada nas linguagens de programação é que ela seja orientada a objetos. A Programação Orientada a Objetos (POO –*Object-Oriented Programming*) é um paradigma de análise, projeto e programação de sistemas de

*software* baseado na composição e interação entre diversas unidades de *software* chamadas de objetos (SICA, 2006).

Tal como ocorre com os SGBD, há várias linguagens de programação orientada a objetos no mercado. Dentre as várias opções, a opção feita foi pela linguagem JAVA. JAVA é uma linguagem de programação desenvolvida na década de 90, pela empresa *Sun Microsystems*. Além de ser orientada a objetos, é considerada uma linguagem de alto nível, com sintaxe similar à da linguagem C++. Diferentemente das linguagens convencionais, que são compiladas para código nativo, a linguagem Java é compilada para um "bytecode" que é executado por uma máquina virtual (DEITEL, 2005).

De acordo com a equipe, a linguagem JAVA foi escolhida por várias razões, das quais se destacam:

- a) possui código aberto de programação; o que reduz de forma significativa os custos de administração, suporte e licenciamento;
- b) possui interface nativa com o SGBDOR PostgreSQL;
- c) simples de programar;
- d) conta com recursos de rede (extensa biblioteca de rotinas que facilitam a cooperação com protocolos TCP/IP, como HTTP e FTP);
- e) facilidades de internacionalização (suporta nativamente caracteres Unicode);
- f) distribuída com um vasto conjunto de bibliotecas (ou APIs); e
- g) conta com Carga Dinâmica de Código (programas em Java são formados por uma coleção de classes armazenadas independentemente e que podem ser carregadas no momento de utilização, minimizando o uso de memória).

Após a definição da linguagem de programação, os desenvolvedores da empresa contratada iniciaram programação do *software* Perdas DSS. Durante a programação, foram implementados vários algoritmos (descrição de como um computador executa uma operação simples e específica) para que a aplicação executasse quatro conjuntos de funções: segurança, importação, planejamento e operacional.

Os algoritmos da função de segurança, são responsáveis pelo cadastro, consulta, edição e validação dos usuários da aplicação Perdas DSS. As telas com a página inicial do *software* e as que apresentam a função segurança implementada podem ser vistas nos APÊNDICES H e I.

Os da função importação são responsáveis pelo processo de importação de dados do ERP da DEE para o *software*. Nesse processo, três grandes conjuntos de dados são importados: i) as inspeções realizadas pela DEE; ii) as instalações com suspeitas de irregularidade no consumo; e iii) as classes de atividade das instalações (são importantes para gerar o consumo médio mensal por classe). As telas que apresentam a função importação implementada podem ser vistas nos APÊNDICES J e K.

As duas últimas funções são as mais importantes da aplicação Perdas DSS. A função de planejamento agrupa os algoritmos responsáveis pela configuração dos locais ou unidade regionais e também por apresentar o gráfico da alocação de equipes de campo de acordo com retorno e a probabilidade das inspeções serem procedentes, como proposto na Figura 23. As telas que apresentam a função planejamento que foram implementadas, podem ser vistas nos APÊNDICES Q e R.

A função operacional agrupa os algoritmos mais importantes da aplicação. Para executar essa função com perfeição, quatro conjuntos de algoritmos foram usados. Um conjunto é responsável por calcular as estatísticas

das inspeções realizadas, gerando, por exemplo, as probabilidades de cada variável.

Um segundo conjunto permite que os usuários analisem as estatísticas geradas para cada variável. Isso permite que os colaboradores da DEE obtenham informações sobre as inspeções realizadas.

O terceiro conjunto de algoritmos é responsável por gerar os gráficos das instalações a serem inspecionadas. Esses algoritmos são responsáveis por gerar um gráfico de dispersão e um com o número de inspeções realizadas.

Por fim, o quarto conjunto de algoritmos é responsável por analisar as instalações suspeitas e gerar um arquivo em formato de planilha eletrônica (.xls), que pode ser analisado pelos colaboradores da DEE. Dentre os algoritmos desse conjunto, o principal algoritmo implantado foi o *K-Means* (LLOYD, 1982).

Como apresentado no tópico 5.4.5, do Ciclo PA-3, esse algoritmo de agrupamento foi e está sendo usado para identificar quais os resultados das variáveis que apresentam maior similaridade com a variável resposta RETORNO\_SVC. Existem algumas variações do algoritmo, entretanto na aplicação Perdas DSS foi implantado a estrutura-padrão, como proposto por Hamerly e Elkan (2002). A Figura 30 demonstra, de maneira simplificada, o funcionamento do algoritmo *k-means*.

As telas que apresentam a função operacional que foi implementada, podem ser vistas nos APÊNDICES de L, M, N, O e P.

O último passo do desenvolvimento foi a compilação da aplicação, com o tratamento prévio dos eventuais erros que surgiram.

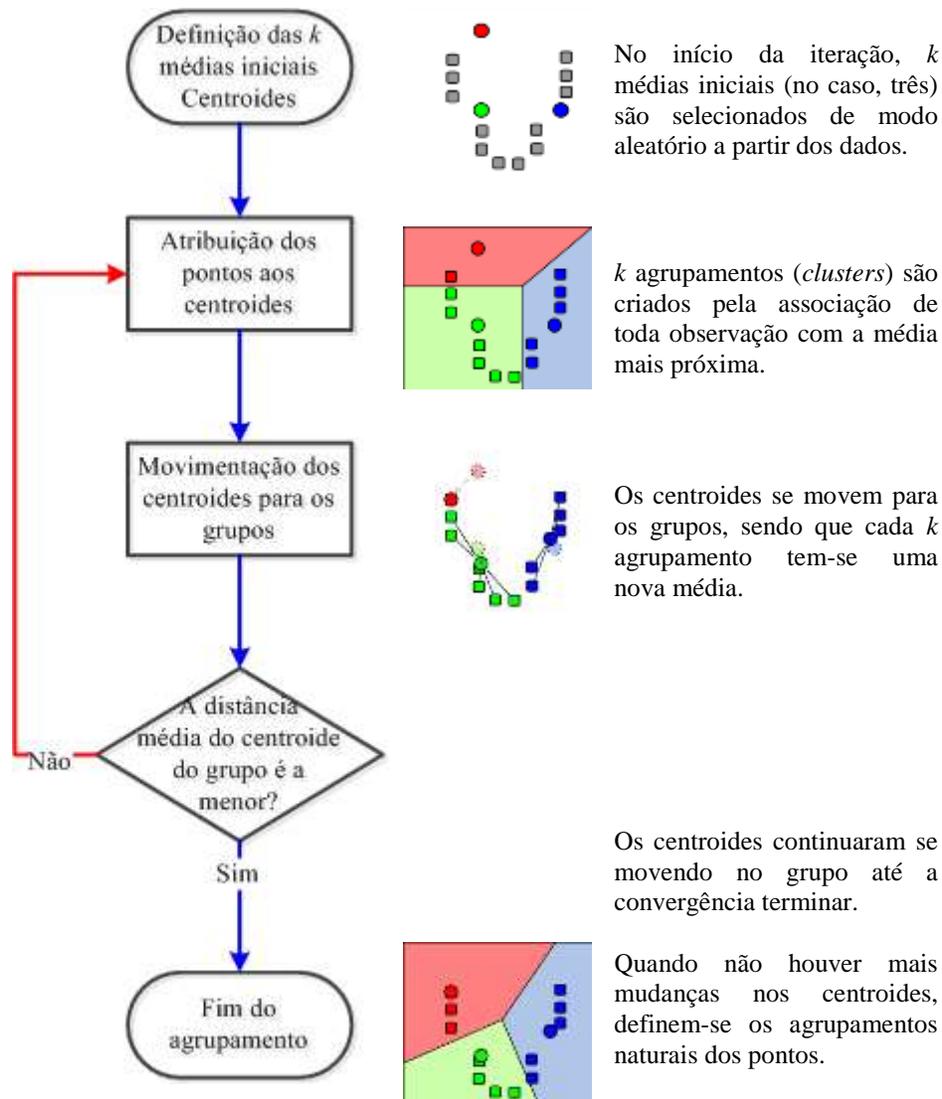


Figura 30 Demonstração simplificada do algoritmo *k-means*, usado na aplicação desenvolvida para a distribuidora em estudo

Fonte: Organizado pelo autor a partir de Hamerly e Elkan (2002) e Lloyd (1982)

Terminada a fase de desenvolvimento, realizou-se uma reunião com a participação dos desenvolvedores, do pesquisador e dos colaboradores da DEE.

O objetivo principal da reunião foi demonstrar o *software* Perdas DSS e submetê-lo a uma primeira avaliação, para posterior implantação.

Na demonstração, o pesquisador e os desenvolvedores apresentaram a interface gráfica e realizaram algumas aplicações práticas. Os colaboradores da DEE gostaram da parte gráfica e acharam que, pelo menos a princípio, a aplicação era fácil de ser utilizada.

Avaliando as funcionalidades da aplicação, o coordenador A/DEE sugeriu que as seguintes adequações fossem implantadas:

- a) inclusão de alguns filtros para facilitar a consulta de dados;
- b) criação de um gráfico resumo da região que estiver sendo analisada;
- c) incorporação ao gráfico da quantidade de equipes de eletricitista necessárias para a região que estiver sendo analisada.

Acertadas as adequações a serem feitas na aplicação, ficou agendado uma nova reunião na DEE, para apresentar a versão final da aplicação e estabelecer um cronograma de implantação do *software* na unidade funcional da distribuidora.

O pesquisador sugeriu então, que até a data na nova reunião, a DEE enviasse alguns arquivos com dados das inspeções realizadas e com dados das instalações suspeitas de fraudar o consumo de energia, para que testes fossem realizados com a aplicação. A sugestão foi aceita pelos colaboradores da DEE.

Ao final da reunião, o coordenador A/DEE informou que o colaborador C/DEE tinha assumido outras funções na equipe de trabalho e que, a partir de então, ele não mais acompanharia a evolução da aplicação Perdas DSS, deixando essa função a cargo do colaborador E/DEE.

No prazo estabelecido, os desenvolvedores fizeram as adequações solicitadas e apesar do atraso, a equipe DEE enviou os dados que foram

importados para a base de dados da aplicação. Apesar dos dados terem sido processados pela aplicação Perdas DSS, não foi possível testar o resultado gerado com os colaboradores responsáveis pela atividade de seleção e liberação das inspeções.

Na nova reunião, estavam presentes os desenvolvedores da empresa contratada, o pesquisador e os colaboradores da DEE. O pesquisador iniciou a reunião apresentando as adequações que haviam sido solicitadas e o resultado que elas geraram na aplicação.

Após a apresentação, os colaboradores da DEE avaliaram os resultados da aplicação. Segundo o colaborador E/DEE, a aplicação era exatamente o que eles estavam precisando para “fortalecer o combate às perdas não técnicas”. Entretanto, para ele, alguns ajustes precisavam ser realizados.

De acordo com o colaborador havia uma série de pontos em que a aplicação precisava ser ajustada. O pesquisador argumentou que muitos dos ajustes solicitados não foram contemplados nem nos Ciclos PA anteriores e nem no processo de desenvolvimento da aplicação. Assim, após uma ampla discussão técnica entre os colaboradores DEE, o pesquisador e os desenvolvedores, apenas 8 ajustes deveriam ser implementados.

Antes de terminar a reunião, o pesquisador informou que, em função dos ajustes na aplicação, não tinha como estabelecer um cronograma de implantação do *software* na distribuidora. Além disso, ele solicitou um novo conjunto de dados para alimentar a aplicação e tentar avaliá-la com resultados de campo, ou seja, disponibilizar resultados para que os colaboradores os analisem, selecionasse e liberasse inspeções para serem executadas pela equipe de campo. Para terminar a reunião, o colaborador E/DEE solicitou à equipe de desenvolvedores a configuração de um servidor para que o *software* Perdas DSS fosse implantado. Solicitação que foi atendida prontamente pela equipe.

Após a implantação dos ajustes solicitados pelo colaborador E/DEE e no intervalo entre a última reunião realizada e a que havia sido agendada, a equipe DEE enviou ao pesquisador vários arquivos com dados das inspeções realizadas e de suspeitos de consumo irregular de energia. Esses dados foram importados e analisados pela aplicação Perdas DSS.

De posse do resultado da análise, o pesquisador entrou em contato com o colaborador E/DEE e disse que havia processado os dados e que gostaria que eles fizessem um teste de campo. O colaborador concordou prontamente, e solicitou que fosse enviado um relatório contendo suspeitos de três municípios distintos da unidade regional DS. Estando de acordo com a solicitação do colaborador E/DEE, o pesquisador enviou o relatório com as informações. Junto com o relatório, o pesquisador informou que os colaboradores que fossem selecionar as inspeções deveriam seguir o fluxograma do POP da atividade de seleção e liberação de inspeção, como representado na Figura 28.

Após várias semanas e antes da reunião, o colaborador E/DEE entrou em contato com o pesquisador e apresentou o resultado do único teste de campo realizado, desde o desenvolvimento da aplicação. O resultado do teste de campo é representado na Tabela 4.

Analisando a Tabela 4, o colaborador da DEE argumentou que, para o grupo de municípios da regional DS, o índice de acerto médio tinha sido menor do que 7%. Segundo ele, em um período de 3 meses, a DEE havia liberado 8.333 notas de inspeção na regional DS, obtendo, sem o uso da aplicação, um índice médio de acerto de quase 13%.

Tabela 4 Resultado da única inspeção de campo realizada usando como referência o resultado gerado pela aplicação Perdas DSS, desenvolvido para a distribuidora em estudo

RESULTADO	MUNICÍPIOS DA UNIDADE REGIONAL			TOTAL
	DS			
	A	B	C	
Total de suspeitos	357	52	249	658
Inspeções liberadas	8	20	19	47
Resultado procedente	1	1	1	3
Resultado improcedente	6	17	8	31
Resultado impedido	1	2	6	9
Inspeções canceladas	0	0	4	4
<b>Índice de Acerto</b>	<b>12,5%</b>	<b>5%</b>	<b>5,3%</b>	<b>6,3%</b>

Fonte: Resultados da pesquisa

Considerando a colocação do colaborador, o pesquisador apresentou os seguintes contra-argumentos:

- a) se comparado ao índice de acerto médio da DEE na regional DS, 6,3% realmente aparentar estar muito abaixo da média. Entretanto, a DEE conseguiu uma média de quase 13% realizando mais de 8 mil inspeções. Ao passo que, no teste realizado usando o relatório da aplicação, foram liberadas apenas 47 inspeções;
- b) deve-se considerar também que, dentre as 47 inspeções liberadas quase 20% das instalações não foram inspecionadas porque os consumidores, por algum motivo, impediram a inspeção. Sem contar que no município C, quatro das 19 inspeções liberadas foram canceladas. No total, quase 28% das inspeções liberadas não foram realizadas.

O colaborador disse que entendia as colocações do pesquisador, mas julgava que o índice de acerto do *software* tinha ficado muito abaixo do que ele esperava. O pesquisador sugeriu então, que novos testes de campo fossem

realizados, usando como referência os municípios da unidade regional DC, por exemplo. O colaborador concordou com a sugestão e solicitou que novos dados fossem enviados.

Antes de encerrar o contato, o pesquisador questionou o colaborador sobre a reunião e a programação de implantação da aplicação. O colaborador informou que o servidor (*hardware*) solicitado para abrigar a aplicação ainda não havia sido adquirido e que, por problemas operacionais, ainda não tinha o retorno de quando ele estaria disponível. A partir dessa informação, o pesquisador disse que poderia liberar, por 30 dias, o acesso remoto da aplicação, para que os colaboradores da DEE se familiarizassem com a aplicação e, assim que estrutura de *hardware* fosse disponibilizada, a implantação seria realizada.

Com a anuência do colaborador da DEE, o pesquisador disponibilizou o acesso remoto da aplicação e fez um rápido treinamento com o colaborador E/DEE, que ficou responsável por multiplicá-lo entre os demais colaboradores da DEE.

Durante o período em que a aplicação ficou disponível remotamente para os colaboradores da DEE, o colaborador E/DEE entrou em contato várias vezes com o pesquisador, para solucionar algumas dúvidas que ele tinha. Entretanto, as dúvidas estavam mais relacionadas com o desenvolvimento da aplicação do que com o resultado que a aplicação gerou para a DEE. Em relação aos resultados, os colaboradores da DEE realizaram apenas mais um teste de campo que, segundo o colaborador, não foram muito satisfatórios.

Ao ser questionado se ele havia seguido o fluxograma da atividade de seleção e liberação de inspeção (Figura 28), ele disse que não. Segundo o colaborador, ele simplesmente solicitou a análise das instalações suspeitas, gerou o relatório com as instalações e as enviou para as equipes de campo. Não realizando, portanto, a função dele que era de analisar, selecionar e liberar as inspeções.

Diante do apresentado, pode-se afirmar que até a presente data a distribuidora não havia disponibilizado a infraestrutura necessária para a implantação do *software* Perdas DSS. E, como o prazo de acesso remoto havia terminado e a DEE não estava usando o *software* para a atividade de alocação das equipes de campo.

#### **5.4.12 Avaliação do Ciclo PA-4**

Avaliando o Ciclo PA-4, pode-se afirmar que os dois primeiros passos (5.5.1 Interesse e motivação e 5.5.2 Objetivos e recursos usados) que antecederam o desenvolvimento da aplicação Perdas DSS estavam bem fundamentados, fruto evolução dos ciclos anteriores da pesquisa-ação e da interação entre os participantes.

Durante o processo de desenvolvimento da aplicação, pode-se afirmar que, de certo modo, houve distanciamento entre o pesquisador e os colaboradores da DEE, que pode ser justificado pelo próprio processo de desenvolvimento da aplicação, que ficou a cargo de uma empresa contratada. Entretanto, o aplicativo foi desenvolvido e todas as solicitações feitas pelos colaboradores da DEE foram atendidas.

Outro ponto que merece ser destacado é que, mesmo com a centralização dos colaboradores responsáveis pela seleção e liberação de inspeções e com a sugestão de um novo POP para a atividade, durante o Ciclo PA-4, não houve indícios que essas mudanças geraram melhorias nos resultados.

Apesar de não ter sido implantada na DEE, a aplicação está pronta para o uso, como pode ser comprovado pelas imagens do *software* que vão do APÊNDICE H ao APÊNDICE R. Entretanto, também não foi possível avaliar qual a real contribuição da aplicação na atividade de alocação das equipes de campo e nem no combate às perdas não técnicas.

Como mencionado por Fleury, Agostini e Lima (1984), a interação usuário-sistema-desenvolvedor é essencial para o sucesso do DSS. Tomando essa afirmação como referência, pode-se afirmar que um dos fatores que pode ter prejudicado o término do projeto foi a sucessiva mudança de colaboradores que ocorreu no final do projeto. Como as pessoas que passaram a participar da pesquisa-ação não conseguiam visualizar o objetivo da aplicação e não tinham participado dos ciclos anteriores, depreendeu-se que o *software* Perdas DSS ficou para segundo plano.

Com esse resultado, deu-se por finalizado o Ciclo PA-4 e todo o ciclo de pesquisa-ação na distribuidora DEE.

## 6 CONCLUSÃO

As perdas não técnicas de energia elétrica são e serão tema recorrente entre as distribuidoras de eletricidade e órgãos reguladores, principalmente em épocas como as de hoje, nas quais as mudanças climáticas minam a capacidade de geração de energia no Brasil. Portanto, o combate ao consumo irregular de energia, além de ser encarado como uma forma de proteção da receita e recuperação financeira, também deve ser vista como a solução de um problema social, que expõe, diariamente, milhares de consumidores ao risco de acidente e à baixa qualidade da energia e dos serviços prestados pelas distribuidoras.

Apesar de vários trabalhos nacionais e internacionais tratarem sobre o tema não há um consenso quanto à melhor forma de solucionar o problema. Alguns estudos apontam para a adoção de novas políticas e práticas para o setor, outros para o uso de recursos tecnológicos como forma de coibir e dificultar as fraudes e a maioria aponta para o uso da tecnologia da informação aliada a complexos algoritmos e cálculos computacionais como a solução para o problema. Entretanto, nenhuma das soluções propostas será efetiva na redução das perdas comerciais, enquanto for identificado apenas o consumidor suspeito. Para cessar a fraude é preciso ter certeza do ato. É necessário auditar, inspecionar cada um daqueles consumidores que foram previamente identificados. O desafio está na maneira de se fazer isso e definir o que é prioritário, quando há milhares de suspeitos e poucas dezenas equipes de campo.

Com base nesses questionamentos pode-se afirmar que este trabalho atingiu seu objetivo. Por meio da pesquisa-ação como procedimento técnico, foi possível, pouco a pouco, interativa e reflexivamente, usar a experiência dos colaboradores da distribuidora para construir um conjunto de ações que resultaram no *software* Perdas DSS. Esse *software* poderá auxiliar na alocação

das equipes de campo, viabilizando técnica e economicamente, a inspeção de consumidores suspeitos de fraudar o consumo de energia.

Apesar de não haver sido implantado na distribuidora, até a apresentação final deste trabalho, não há como negar que o *software* de apoio à decisão desenvolvido, chamado de Perdas DSS, pode ser considerado como um dos principais resultados deste trabalho. Entretanto, tão importante quanto o produto desenvolvido, foi a geração de uma teoria emergente, que se desenvolveu a partir de uma síntese, originária de dados e que evoluiu a partir da utilização prática de um corpo teórico bem estabelecido e consolidado.

Outro ponto que deve ser ressaltado é o comprometimento dos participantes da pesquisa. Enquanto houver pessoas comprometidas com um projeto, melhores serão os resultados. Assim, de nada adiantou o desenvolvimento de um instrumento para auxiliar no processamento e análise de dados se as pessoas não estiverem comprometidas com o resultado a ser produzido.

Este trabalho, adotando a pesquisa-ação como procedimento técnico, não tinha o objetivo de criar um conhecimento universal e amplamente aceito. Porém, os resultados obtidos na análise da distribuidora de energia elétrica poderão ser comparados com o de outras distribuidoras que vivem situações semelhantes. Promovendo, desse modo, o progresso científico, que surge por meio da construção de teorias que evoluem, paulatinamente, do particular para o geral.

Para trabalhos futuros, sugere-se que outras distribuidoras sejam pesquisadas, na tentativa de replicar as ações deste trabalho e acrescentar novos aspectos teóricos aos já desenvolvidos. Neles, poderia ser adicionado à aplicação Perdas DSS, elementos de geoposicionamento, que poderiam ser usados para a otimização de rotas de inspeção das equipes de campo. Como resultado,

poder-se-ia aumentar o número de inspeções realizadas por dia, em uma mesma localidade, ao minimizar, por exemplo, o caminho a ser percorrido pela equipe.

Para a administração este trabalho demonstra sua importância por vários fatores, dos quais se destacam:

- a) desenvolvimento de um sistema de apoio à decisão para gerar informação, que é insumo preponderante para que gestores, de modo geral, tomem de decisões em ambiente de incerteza;
- b) possibilidade de melhoria nos resultados das inspeções, contribuindo para a economia de recursos da distribuidora, viabilizando portanto, técnica e economicamente, a inspeção de consumidores suspeitos de fraudar o consumo de energia; e
- c) a importância da efetiva participação das pessoas na melhoria dos processos gerenciais das empresas e, especificamente no caso deste trabalho, no comprometimento com os resultados da pesquisa.

## REFERÊNCIAS

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. **Consumo irregular de energia gera prejuízo de R\$ 8,1 bilhões ao ano**. 2011. Disponível em: <[http://www.ANEEL.gov.br/aplicacoes/noticias/Output\\_Noticias.cfm?Identidade=4160&id\\_area=90](http://www.ANEEL.gov.br/aplicacoes/noticias/Output_Noticias.cfm?Identidade=4160&id_area=90)>. Acesso em: 10 jul. 2012.

AHO, A. V. et al. **Compiladores: princípios, técnicas e ferramentas**. São Paulo: Addison-Wesley, 2008. 634 p.

ANTHONY, R. N. **Planning and control systems: framework for analysis**. Boston: Harvard University, 1965.

ARAÚJO, A. C. M. **Perdas e inadimplência na atividade de distribuição de energia elétrica no Brasil**. 2007. 125 f. Tese (Doutorado em Planejamento Energético) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2007.

ARNOTT, D.; PERVAN, G. A critical analysis of decision support systems research. **Journal of Information Technology**, Germantown, v. 20, n. 2, p. 67-87, 2005.

ARNOTT, D.; PERVAN, G. Eight key issues for the decision support systems discipline. **Decision Support Systems**, Amsterdam, v. 44, n. 3, p. 657-672. 2008.

BERNARDES, M. L. **Proposta de um plano de ação para atenuação de perdas não técnicas de distribuição de energia elétrica em uma empresa do Rio Grande do Sul**. 2010. 222 p. Dissertação (Mestrado em Engenharia de Produção e Sistemas) – Universidade do Vale dos Sinos, São Leopoldo, 2010.

BERTRAND, J. W. M.; FRASOO, J. C. Modelling and simulation. Operations management research methodologies using quantitative modeling. **International Journal of Operations & Production Management**, Bradford, v. 22, n. 2, p. 241-264, 2002.

BRASIL. Agência Nacional de Energia Elétrica. Superintendência de Regulação dos Serviços de Distribuição. **Nota técnica nº 342/2008**. Brasília, 2008. Disponível em: <<http://www.ANEEL.gov.br/arquivos/PDF/Nota20TC3A9cnica2034220-20Perdas20NC3A3o20TC3A9cnicas.pdf>>. Acesso em 30 jun. 2011.

CABRAL, J. E. et al. **Methodology for fraud detection using rough sets**. 2006. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1635791&isnumber=34297>>. Acesso em: 14 ago. 2012.

CABRAL, J. E.; GONTIJO, E. M. **Fraud detection in electrical energy consumers using rough sets**. 2004. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1400905&isnumber=30424>>. Acesso em: 14 ago. 2012.

CALILI, R. F. **Desenvolvimento de sistema para detecção de perdas comerciais em redes de distribuição de energia elétrica**. 2005. Dissertação (Mestrado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2005.

CHECKLAND, P.; HOLWELL, S. **Information, systems and information systems: making sense of the field**. Chichester: J. Wiley, 1998. 278 p.

COUGHLAN, D.; BRANNICK, T. **Doing action research in your organization**. 3rd ed. London: SAGE, 2009. 184 p.

COUGHLAN, P.; COUGHLAN, D. Action research for operations management. **International Journal of Operations & Production Management**, Bradford, v. 22, n. 2, p. 220-240, 2002.

DEITEL, H. M. **Java: como programar**. 6th ed. São Paulo: Pearson education do Brasil, 2005. 1097 p.

DERSHEM, H. L.; JIPPING, M. J. **Programming Languages: structures and models**. 2nd ed. Boston: PWS, 1995. 432 p.

DIAS, H. B. P. **Uma Abordagem baseada em conhecimento para apoio ao combate às Perdas Comerciais na Distribuição de Energia Elétrica**. 2006. 96 p. Dissertação (Mestrado em Informática) – Universidade Federal do Espírito Santo, Vitória, 2006. Disponível em: <[http://www.btdt.ufes.br/tde\\_busca/arquivo.php?codArquivo=353](http://www.btdt.ufes.br/tde_busca/arquivo.php?codArquivo=353)> Acesso em: 28 jun. 2011.

DICK, A. J. Theft of electricity. How UK electricity companies detect and deter. **European Convention**, p. 90-95, May 1995. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=491546&isnumber=10615>>. Acesso em: 1 jul. 2011.

DONG, C. S. J.; SRINIVASAN, A. Agent-enabled service-oriented decision support systems. **Decision Support System**, Amsterdam, v. -, n. -, p. --, 2012. In Press, Corrected Proof. Disponível em: <<http://dx.doi.org/10.1016/j.dss.2012.05.047>>. Acesso em: 4 abr. 2011.

ER, M. C. Decision support systems: a summary, problems, and future trends. **Decision Support Systems**, Amsterdam, v. 4, n. 3, p. 355-363, 1988.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, Palo Alto, v. 17, n. 3, p. 37-54, 1996. Disponível em: <<http://www.aaai.org/ojs/index.php/aimagazine/article/view/1230/1131>>. Acesso em: 4 abr. 2011.

FERREIRA, A. C. S. et al. Implantação do projeto GeoPerdas: georreferenciamento de perdas comerciais na CEMIG. In: ENCONTRO NACIONAL DE GEOPROCESSAMENTO DO SETOR ELÉTRICO, 2., 2012, Recife. **Anais eletrônicos...** Recife: ENGEIO, 2012. Disponível em: <<http://www.engeio2012.com.br/index.php/download-das-trabalhos-tecnicos?download=46:trabalhos-tecnicos>>. Acesso em: 15 jul. 2012.

FINLAY, P. N. **Introducing decision support systems**. 2nd ed. Oxford : NCC Blackwell, 1994. 274 p.

FISCHER, A. E.; GRODZINSKY, F. **The anatomy of programming languages**. Englewood Cliffs: Prentice Hall, 1993. 557 p.

FLEURY, A. C. C.; AGOSTINI, C. G.; LIMA, R. R. Sistemas de apoio a decisão em projetos de engenharia. **Revista de Administração**, Santa Maria, v. 19, n. 3, p. 28-33, 1984.

GALBRAITH, J. R. **Designing organizations: an executive briefing on strategy, structure and process**. San Francisco: Jossey-Bass, 1995.

GALBRAITH, J. R. **Organization design**. Reading: Addison-Wesley, 1973.

GIL, A. C. **Métodos e técnicas de pesquisa social**. São Paulo: Atlas, 1999.

GOLDBERG, D. E. **Genetic algorithms in search, optimization and machine learning**. New York: Addison-Wesley, 1989. 432 p.

GUMMESSON, E. **Qualitative methods in management research**. 2nd ed. Thousand Oaks: Sage, 1999. 264 p.

HALL, M. A.; HOLMES, G. Benchmarking attribute selection techniques for discrete class data mining. **IEEE Transactions on Knowledge and Data Engineering**, Los Alamitos, v. 15, n. 6, 1374-1388, Nov./Dec. 2003.

HALL, M. A.; SMITH, L. A. Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper. In: INTERNATIONAL FLORIDA ARTIFICIAL INTELLIGENCE RESEARCH, 12., 1999, New York. **Proceedings...** New York: AAI, 1999. p. 235-239.

HAMERLY, G.; ELKAN, C. Alternatives to the k-means algorithm that find better clusterings. In: INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 11., 2002, McLean. **Proceedings...** McLean: ACM, 2002. 1 CD ROM.

HAN, J.; KAMBER, M.; PEI, J. **Data mining: concepts and techniques**. 3. ed. Waltham: M. Kaufmann, 2011. 744 p.

INSTITUTO ACENDE BRASIL. **Perdas e inadimplência no setor elétrico**. São Paulo, 2007. (Cadernos de Política Tarifária). Disponível em: <[http://www.acendebrasil.com.br/archives/files/estudos/Caderno\\_05\\_Perdas\\_e\\_Inadimplencia.pdf](http://www.acendebrasil.com.br/archives/files/estudos/Caderno_05_Perdas_e_Inadimplencia.pdf)>. Acesso em: 29 jun. 2011.

KEEN, P. G. W.; GAMBINO, T. J. **Building a decision support system: the mythical man-month revisited**. 1980. Disponível em: <[http://ia700304.us.archive.org/5/items/buildingdecision00keen/building decision00keen.pdf](http://ia700304.us.archive.org/5/items/buildingdecision00keen/building%20decision00keen.pdf)>. Acesso em: 4 abr. 2012.

KEEN, P. G. W.; SCOTT MORTON, M. S. **Decision support systems: an organizational perspective**. Reading: Addison-Wesley, 1978.

KIRA, K.; RENDELL, L. A. A practical approach to feature selection. In: INTERNATIONAL WORKSHOP ON MACHINE LEARNING, 9., San Francisco, 1992. **Proceedings...** São Francisco: M. Kaufmann, 1992. p. 249-256.

KOHAVI, R.; PROVOST, F. Glossary of terms. **Machine Learning**, Heidelberg, v. 30, n. 2/3, p. 271-274, 1998.

KONONENKO, I. Estimating attributes: analysis and extensions of RELIEF. In: EUROPEAN CONFERENCE ON MACHINE LEARNING ON MACHINE LEARNING, 1., 1994, New York. **Proceedings...** New York: Springer-Verlag, 1994. p. 171-182. Disponível em: <<http://dl.acm.org/citation.cfm?id=188427>>. Acesso em: 22 jun. 2012.

LAUDON, K.; LAUDON, J. **Sistemas de informações gerenciais**. 9. ed. São Paulo: Pearson Prentice Hall, 2010. 428 p.

LEÓN, C. et al. Integrated expert system applied to the analysis of non-technical losses in power utilities. **Expert Systems with Applications**, Adelphi, v. 38, n. 8, p. 10274-10285, Aug. 2011.

LIMA, D. A. **Convergência tarifária**: remédio regulatório para o livre acesso. Brasília: ANEEL, 2005. 16 p. Disponível em: <[http://www.ANEEL.gov.br/arquivos/PDF/Convergencia\\_Tarifaria.pdf](http://www.ANEEL.gov.br/arquivos/PDF/Convergencia_Tarifaria.pdf)>. Acesso em: 4 abr. 2011.

LIU, H.; SETIONO, R. A probabilistic approach to feature selection: a filter solution. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 13., 1996, San Francisco. **Proceedings...** San Francisco: M. Kaufmann, 1996. p. 319-327.

LLOYD, S. P. Least squares quantization in PCM. **IEEE Transactions on Information Theory**, Zurich, n. 28, v. 2, p. 129-137, 1982.

LYYTINEN, K. Expectation failure concept and systems analysts' view of information system failures: results of an exploratory study. **Information & Management**, Amsterdam, v. 14, p. 45-56, 1988.

MACKE, J. A pesquisa-ação como estratégia de pesquisa qualitativa. In: GODOL, C.; BANDEIRA DE MELO, R.; SILVA, A. **Pesquisa qualitativa em estudos organizacionais**: paradigmas, estratégias e métodos. São Paulo: Saraiva, 2006. p. 207-239.

McAFEE, A. Mastering the three worlds of information technology. **Harvard Business Review**, Cambridge, v. 84, n. 11, p. 141-149, 2006.

McGREGOR, A. et al. **Flow clustering using machine learning techniques**. 2004. Disponível em: <[http://security.riit.tsinghua.edu.cn/~bhyang/paper\\_read/PAM2004\\_Flow%20Clustering%20Using%20Machine%20Learning%20Techniques.pdf](http://security.riit.tsinghua.edu.cn/~bhyang/paper_read/PAM2004_Flow%20Clustering%20Using%20Machine%20Learning%20Techniques.pdf)>. Acesso em: 22 jan. 2012.

McNURLIN, B. C.; SPRAGUE, R. H.; BUI, T. **Information systems management in practice**. 8th ed. Upper Saddle River: Prentice Hall, 2009.

MELO, A. C. V.; SILVA, F. S. C. **Princípios de linguagens de programação**. São Paulo: E. Blücher, 2003. 211 p.

NAGI, J. et al. Non-Technical loss analysis for detection of electricity theft using support vector machines. In: POWER AND ENERGY CONFERENCE, 2., 2008, Malaysia. **Anais eletrônicos...** Malaysia: PECon, 2008. p. 907-912. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4762604&isnumber=4762428>>. Acesso em: 4 jul. 2011.

NIZAR, A. H. et al. Load profiling method in detecting non-technical loss activities in a power utility. In: INTERNATIONAL POWER AND ENERGY CONFERENCE, 1., Malaysia, 2006. **Anais eletrônicos...** Malaysia: PECon 2006. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=04154468>>. Acesso em: 4 jul. 2011.

O'BRIEN, J. A. **Sistemas de informação e as decisões gerenciais na era da Internet**. 3. ed. São Paulo: Saraiva, 2011. 492 p.

ONAT, N. **Transmission and distribution losses of Turkey's power system**. Istanbul: Marmara University, 2010. Disponível em: <<http://www.wseas.us/e-library/conferences/2010/Tunisia/EPERES/EPERES-28.pdf>>. Acesso em: 4 jul. 2011.

ORTEGA, G. V. C. **Redes neurais na identificação de perdas comerciais do setor elétrico**. 2008. 184p. Dissertação (Mestrado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

PAWLAK, Z. Rough sets. **International Journal of Parallel Programming**, Heidelberg, v. 11, n. 5, p. 341-356, 1982. Disponível em: <<http://dx.doi.org/10.1007/BF01001956>>. Acesso em: 14 ago. 2012.

PENIN, C. A. S. **Combate, prevenção e otimização das perdas comerciais de energia elétrica**. 2008. 214 p. Tese (Doutorado em Engenharia) – Escola Politécnica da Universidade de São Paulo, São Paulo, 2008. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/3/3143/tde-14082008-092248/pt-br.php>>. Acesso em: 11 jan. 2010.

PERDA brasileira de energia é quase o dobro da média mundial. **O Estado de São Paulo**, Brasília, 12 fev. 2010. Disponível em: <<http://www.estadao.com.br/noticias/impresso,perda-brasileira-de-energia-e-quase-o-dobro-da-media-mundial-,834727,0.htm>>. Acesso em: 1 jun. 2012.

PREFEITURAS inadimplentes religam energia por conta própria, diz Eletrobras. **Gazetaweb**, Maceió, 6 ago. 2012. Disponível em: <<http://gazetaweb.globo.com/noticia.php?c=319690&e=6>>. Acesso em: 14 ago. 2012.

QUEIROGA, R. M. **Uso de técnicas de data mining para detecção de fraudes em energia elétrica**. 2005. – 147 p. Dissertação (Mestrado em Informática) – Universidade Federal do Espírito Santo, Vitória, 2005.

REIS, C. Z. **Eficácia de solução tecnológica para redução de furtos de energia em empresas distribuidoras: estudo de caso**. 2005. – 81 p. Dissertação (Mestrado em Engenharia de Produção) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2005.

REIS FILHO, J. **Sistema inteligente baseado em árvore de decisão, para apoio ao combate às perdas comerciais na distribuição de energia elétrica**. 2006. 174 f. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Uberlândia, Uberlândia, 2006.

RESULTADO consolidado da inadimplência do setor de distribuição no 3T08. Brasília: ABRADÉE, 2008. Disponível em: <[http://www.abradee01.org/download/relatorios/Inadimplencia\\_2005-abr.pdf](http://www.abradee01.org/download/relatorios/Inadimplencia_2005-abr.pdf)>. Acesso em: 14 ago. 2012.

ROBNIK-SIKONJA, M.; KONONENKO, I. An adaptation of RELIEF for attribute estimation in regression. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 14., 1997. **Proceedings...** San Francisco: M. Kaufmann, 1997. p. 296-304.

SANTOS, M. F.; AZEVEDO, C. S. **Data mining: descoberta de conhecimento em base de dados**. Lisboa: FCA, 2005. 214 p.

SCOTT MORTON, M. S. **Management decision systems: computer-based support for decision making**. Boston: Harvard University, 1971.

SICA, C. **PHP orientado a objetos: fale a linguagem da internet**. Rio de Janeiro: Ciência Moderna, 2006. 216 p.

SILVA, E. L.; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. 4. ed. Florianópolis: UFSC, 2005. 138 p.

SIMÃO, E. ‘Gatos’ causam prejuízos de R\$ 7 bi e encarecem tarifas. **O Estado de São Paulo**, Brasília, 12 fev. 2012. Disponível em: <<http://www.estadao.com.br/noticias/impresso,gatos-causam-prejuizo-de-r-7-bi-e-encarecem-tarifas-,834723,0.htm>>. Acesso em: 1 jun. 2012.

SMITH, T. B. Electricity theft: a comparative analysis. **Energy Policy**, Guildford, v. 32, n. 18, p. 2067-2076, Dec. 2004.

STEADMAN, K. **Electricity theft in Jamaica**. New York: University of New York at Binghamton, 2010. v. 1. Disponível em: <<http://www2.binghamton.edu/economics/graduate/documents/prospectus-by-k-steadman.pdf>>. Acesso: 4 jul. 2011.

SUN, J.; TENG, J. T. C. Information systems use: construct conceptualization and scale development. **Computers in Human Behavior**, New York, v. 28, n. 5, p. 1564-1574, Sept. 2012.

TENG, J. T. C.; CALHOUN, K. J. Organizational computing as a facilitator of operational and managerial decision making: An exploratory study of managers' perceptions. **Decision Sciences**, Malden, v. 27, n. 4, p. 673-710, 1996.

THIOLLENT, M. **Metodologia da pesquisa-ação**. 14. ed. São Paulo: Cortez, 2005. 132 p.

TURBAN, E. **Decision support system and expert system: management Support Systems**. 4th ed. New York: Prentice Hall, 1995. 930 p.

TURBAN, E.; MCLEAN, E.; WETHERBE, J. **Tecnologia da informação para gestão: transformando os negócios na economia digital**. 3. ed. Porto Alegre: Bookman, 2004. 645 p.

VIEIRALVES, E. X. **Proposta de uma metodologia para avaliação das perdas comerciais dos sistemas elétricos: o caso Manaus**. 2005. – 180 p. Dissertação (Mestrado em Planejamento de Sistemas Energéticos) - Universidade Estadual de Campinas, Campinas, 2005.

WEISS, S. M.; KULIKOWSKY, C. A. **Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems**. San Francisco: M. Kaufmann, 1991. 223 p.

WESTBROOK, R. Action research, a new paradigm for research in production and operations management. **International Journal of Operations & Production Management**, Exeter, v. 15, n. 12, p. 6-20, 1995.

WITTEN, I. H.; FRANK, E. **Data mining**: practical machine learning tools and techniques. 2d ed. San Francisco: M. Kaufmann, 2005. 560 p.

**APÊNDICE**

## APÊNDICE A

### Resultado dos métodos de busca associado ao algoritmo *CfsSubsetEval*

```

=== Run information === for all Search Method

Evaluator: weka.attributeSelection.CfsSubsetEval
Search: weka.attributeSelection.BestFirst -D 1 -N 5
Relation: RetornoSVC-weka.filters.unsupervised.attribute.Remove-R1-
weka.filters.unsupervised.attribute.Remove-R12-13
Instances: 10171
Attributes: 14
    Tensao
    Dia_Leitura
    Local_Leitura
    Rota_Leitura
    Classe
    Ramo_Atividade
    Numero_Fases
    Mot_Antigo
    DataMotivo_DataENota_dias
    DataMotivo_DataInspecao_dias
    DataENota_DataInspecao_dias
    Retorno_SVC
    Cod_SVC

NIE
Evaluation mode: evaluate on all training data

```

```

1 = Attribute Selection on all input data =

Search Method:
    Best first.
    Start set: no attributes
    Search direction: forward
    Stale search after 5 node
    expansions
    Total number of subsets evaluated:
    65
    Merit of best subset found: 0.583

Attribute Subset Evaluator (supervised,
Class (nominal): 12 Retorno_SVC):
    CFS Subset Evaluator
    Including locally predictive
attributes

Selected attributes: 13, 14 : 2
    13 - Cod_SVC
    14 - NIE

```

```

2 = Attribute Selection on all input data =

Search Method:
    Exhaustive Search.
    Start set: no attributes
    Number of evaluations: 8192
    Merit of best subset found: 0.583

Attribute Subset Evaluator (supervised,
Class (nominal): 12 Retorno_SVC):
    CFS Subset Evaluator
    Including locally predictive
attributes

Selected attributes: 13, 14 : 2
    13 - Cod_SVC
    14 - NIE

```

Apêndice A – Continua

**3 = Attribute Selection on all input data =****Search Method:****Genetic search.**

Start set: no attributes  
 Search Method:  
 Genetic search.  
 Start set: no attributes  
 Population size: 20  
 Number of generations: 20  
 Probability of crossover: 0.6  
 Probability of mutation: 0.033  
 Report frequency: 20  
 Random number seed: 1

## Initial population

merit	scaled	subset
0.25467	0.22651	1 2 4 7 9 11 13
0.06967	0.09835	1 4 5 6 7 10 11
0.06638	0.09607	1 4 5 7 9
0.30509	0.26145	4 6 7 8 10 13 14
0.13702	0.14501	1 9 10 11 14
0.28426	0.24701	1 2 4 6 7 8 9 13 14
0.0747	0.10183	1 3 4 6 7 8 11
0.16232	0.16254	1 7 8 14
0.30259	0.25972	2 3 4 6 10 13 14
0.04881	0.0839	3 6 7 8 9 11
0.01761	0.06228	8
0.02278	0.06586	11
0.39834	0.32605	6 10 13 14
0.04691	0.08258	4 5 8 11
0.23363	0.21194	2 3 4 5 7 9 13
0.13455	0.1433	1 2 4 6 7 8 10 11 14
0.32703	0.27664	2 3 8 10 13 14
0.08624	0.10983	1
0.21102	0.19628	4 5 6 7 9 10 13

**3 = Attribute (cont.)**

Generation: 20

merit	scaled	subset
0.52577	0.6643	13 14
0.52577	0.6643	13 14
0.44629	0.21159	11 13 14
0.52577	0.6643	13 14
0.41101	0.01062	1 7 13 14
0.52577	0.6643	13 14
0.44417	0.19949	10 13 14
0.44683	0.21463	3 13 14
0.52577	0.6643	13 14
0.44683	0.21463	3 13 14
0.44136	0.18351	4 13 14
0.52577	0.6643	13 14
0.58339	0.99254	13
0.44772	0.2197	5 13 14
0.52577	0.6643	13 14
0.44349	0.19562	8 13 14
0.44136	0.18351	4 13 14
0.58339	0.99254	13
0.52577	0.6643	13 14
0.58339	0.99254	13

Attribute Subset Evaluator (supervised,  
 Class (nominal): 12 Retorno\_SVC):  
 CFS Subset Evaluator  
 Including locally predictive  
 attributes

Selected attributes: 13, 14 : 2  
 13 - Cod\_SVC  
 14 - NIE

**4 = Attribute Selection on all input data =****Search Method:****Greedy Stepwise (forwards).**

Start set: no attributes  
 Merit of best subset found: 0.583

Attribute Subset Evaluator (supervised,  
 Class (nominal): 12 Retorno\_SVC):  
 CFS Subset Evaluator  
 Including locally predictive  
 attributes

Selected attributes: 13, 14 : 2  
 13 - Cod\_SVC  
 14 - NIE

**5 = Attribute Selection on all input data =****Search Method:****Random search.**

Start set: no attributes  
 Number of iterations: 2048 (25.0%  
 of the search space)  
 Merit of best subset found: 0.445

Attribute Subset Evaluator (supervised,  
 Class (nominal): 12 Retorno\_SVC):  
 CFS Subset Evaluator  
 Including locally predictive attributes

Selected attributes: 2, 13, 14 : 3  
 2 - Dia\_Leitura  
 13 - Cod\_SVC  
 14 - NIE

## Apêndice A – Conclusão

```
6 = Attribute Selection on all input data =  
Search Method:  
  RankSearch  
  Attribute evaluator :  
weka.attributeSelection.GainRatioAttributeEval  
  Attribute ranking :  
  13 Cod_SVC  
  1 Tensao  
  14 NIE  
  5 Classe  
  6 Ramo_Atividade  
  3 Local_Leitura  
  2 Dia_Leitura  
  11 DataENota_DataInspecao_dias  
  8 Mot_Antigo  
  4 Rota_Leitura  
  10 DataMotivo_DataInspecao_dias  
  9 DataMotivo_DataENota_dias  
  7 Numero_Fases  
  Merit of best subset found : 0.583  
  
Attribute Subset Evaluator (supervised, Class  
(nominal): 12 Retorno_SVC):  
  CFS Subset Evaluator  
  Including locally predictive attributes  
  
Selected attributes: 13, 14 : 2  
  13 - Cod_SVC  
  14 - NIE
```

## APÊNDICE B

Quadro 1B *Data mining*: técnica de classificação – Algoritmos usados e suas características, usados no estudo

Tipo de Classificador	Algoritmo	Característica
<i>Bayes</i>	<i>AODE</i>	Preciso na classificação da média geral de um pequeno espaço de alternativa baseado em modelos <i>naïve-Bayes</i> que são frágeis (e, portanto, menos prejudicial). O algoritmo é computacionalmente eficiente, garantindo simultaneamente a sua grande precisão na classificação de muitas tarefas de aprendizagem.
	<i>BayesNet</i>	<i>Bayes Network Learning</i> usam vários algoritmos pesquisa e medidas de qualidade. Classe básica para a classificação <i>Bayes Network</i> . Fornece a estrutura dos dados (rede de estrutura, as distribuições de probabilidade condicional, etc.) e facilidades comuns aos a <i>Bayes Network</i> , como K2 e B. Para maiores informações: < <a href="http://www.cs.waikato.ac.nz/~remco/weka.pdf">http://www.cs.waikato.ac.nz/~remco/weka.pdf</a> >
<i>Trees</i>	<i>J48</i>	Classe para gerar árvores de decisão do tipo C4.5, podadas ou não. Para maiores informações, consultar: QUINLAN R. C4.5: Programs for Machine Learning. San Mateo: M. Kaufmann, 1993.
	<i>REPTree</i>	Constrói uma árvore decisão/regressão utilizando informações ganho/variância e a poda, utilizando o redutor de erro por poda (com <i>backfitting</i> ). Apenas sorteia valores para atributos numéricos uma vez. Valores em falta são tratadas através da divisão dos correspondentes casos em peças (isto é, como no C4.5).
<i>Rules</i>	<i>OneR</i>	Classe para a construção e utilização de um classificador 1R; em outras palavras, utiliza o erro mínimo do atributo para predição, discretizando atributos numéricos. Para maiores informações, consultar: HOLTE, R. C. Very simple classification rules perform well on most commonly used datasets. Machine Learning, Heidelberg, v. 11, p. 63-91, 1993.

“continua”

Quadro 1B “continuação”

Tipo de Classificador	Algoritmo	Característica
<i>Rules</i>	<i>PART</i>	<p>Classe para gerar uma lista de decisão PART. Utiliza <i>separate-and-conquer</i>. Constrói uma árvore de decisão parcial C4.5 em cada iteração e faz o "melhor" em uma regra folha.</p> <p>Para maiores informações, consultar: FRANK E.; WITTEN, I. H. Generating accurate rule sets without global optimization. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 15., 1998, San Francisco. Proceedings... San Francisco: M. Kaufmann, 1998. p. 144-151.</p>
<i>Functions</i>	<i>SMO</i>	<p>Implementa o algoritmo de otimização sequencial mínimo de John Platt para treinar um vetor suporte de classificação.</p> <p>Esta implementação global substitui todos os valores em falta e transforma os atributos requeridos em nominal. Também normaliza todos os atributos por padrão. (Nesse caso, os coeficientes de saída são baseados nos dados normalizados e não nos dados originais – isto é importante para interpretar o classificador.)</p> <p>Problemas <i>Multiclasse</i> são resolvidos usando classificação <i>pairwise</i> (1-vs-1 e se modelos logísticos são construídos de acordo com engate <i>pairwise</i> de Hastie e Tibshirani, 1998).</p> <p>Para maiores informações, consultar:</p> <p>PLATT, J. Machines using sequential minimal optimization. In: SCHOELKOPF, B.; BURGESS C.; SMOLA, A. Advances in Kernel methods: support vector learning. [S.: s. n.], 1998.</p> <p>KEERTHI, S. S. et al. Improvements to Platt's SMO algorithm for SVM classifier design. Neural Computation, Cambridge, v. 13, n. 3, p. 637-649, 2001.</p> <p>HASTIE, T.; TIBSHIRANI R. Classification by pairwise coupling. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 12., Cambridge. Proceedings... Cambridge: MIT, 1998.</p>

“continua”

Quadro 1B “conclusão”

Tipo de Classificador	Algoritmo	Característica
<i>Neural Networks</i>	<i>Multilayer Perceptron</i>	Classificador que usa <i>backpropagation</i> para classificar os casos. Esta rede pode ser construída manualmente, criadas por um algoritmo ou ambos. A rede também pode ser monitorada e modificada durante o período de formação. Os nós nesta rede são todos sigmóides (exceto quando a classe é numérica, caso em que a saída dos nós tornam-se unidades lineares <i>unthresholded</i> ).
<i>Metalearning</i>	<i>Decorate</i>	<i>Metalearning</i> para a construção de diversos conjuntos de classificadores, utilizando especialmente para treinamento de exemplos artificialmente construídos. Experimentos abrangentes têm demonstrado que esta técnica é consistentemente mais precisas do que a base <i>classifier, bagging and random forests</i> . Este classificador também é mais preciso do que <i>boosting</i> em pequenos conjuntos de treinamento e tem desempenho comparável a amplos conjuntos de treinamento.
	<i>LogitBoost</i>	Classe para realização de regressão logística aditiva. Esta classe realiza classificação utilizando um regime de regressão como a base de aprendizagem, sendo indicado na classificação de problemas multiclasse. Realiza <i>cross-validation</i> interna, mostrando ser eficiente na determinação do número apropriado de interações.

## APÊNDICE C

Tabela 1C *Data mining*: técnica de classificação – Resumo das estatísticas dos algoritmos, usados no estudo

Algoritmo	Resumo Estatístico																																								
AODE	<p>=== Summary ===</p> <p><b>Correctly Classified Instances</b>      <b>6.413</b>    <b>77,126%</b>  <b>Incorrectly Classified Instances</b>    <b>1.902</b>    <b>22,874%</b>  Kappa statistic                            0.5407  Mean absolute error                      0.2043  Root mean squared error                0.3232  Relative absolute error                 61,331%  Root relative squared error            79,194%  Total Number of Instances            8.315</p>																																								
	<p>=== Detailed Accuracy By Class ===</p> <table border="1"> <thead> <tr> <th></th> <th>TP Rate</th> <th>FP Rate</th> <th>Precision</th> <th>Recall</th> <th>F-Measure</th> <th>ROC Area</th> <th>Class</th> </tr> </thead> <tbody> <tr> <td></td> <td>0.826</td> <td>0.287</td> <td>0.754</td> <td>0.826</td> <td>0.788</td> <td>0.852</td> <td>PROCEDENTE</td> </tr> <tr> <td></td> <td>0.713</td> <td>0.174</td> <td>0.794</td> <td>0.713</td> <td>0.751</td> <td>0.852</td> <td>IMPROCEDENTE</td> </tr> <tr> <td></td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>IMPEDIDO</td> </tr> <tr> <td><b>Weighted Avg.</b></td> <td><b>0.771</b></td> <td><b>0.232</b></td> <td><b>0.773</b></td> <td><b>0.771</b></td> <td><b>0.77</b></td> <td><b>0.852</b></td> <td></td> </tr> </tbody> </table>		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class		0.826	0.287	0.754	0.826	0.788	0.852	PROCEDENTE		0.713	0.174	0.794	0.713	0.751	0.852	IMPROCEDENTE		0	0	0	0	0	0	IMPEDIDO	<b>Weighted Avg.</b>	<b>0.771</b>	<b>0.232</b>	<b>0.773</b>	<b>0.771</b>	<b>0.77</b>	<b>0.852</b>	
		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class																																	
	0.826	0.287	0.754	0.826	0.788	0.852	PROCEDENTE																																		
	0.713	0.174	0.794	0.713	0.751	0.852	IMPROCEDENTE																																		
	0	0	0	0	0	0	IMPEDIDO																																		
<b>Weighted Avg.</b>	<b>0.771</b>	<b>0.232</b>	<b>0.773</b>	<b>0.771</b>	<b>0.77</b>	<b>0.852</b>																																			
<p>=== Confusion Matrix ===</p> <table border="1"> <thead> <tr> <th>a</th> <th>b</th> <th>c</th> <th>&lt;- classified as</th> </tr> </thead> <tbody> <tr> <td>3.541</td> <td>745</td> <td>0</td> <td>a = PROCEDENTE</td> </tr> <tr> <td>1.157</td> <td>2.872</td> <td>0</td> <td>b = IMPROCEDENTE</td> </tr> <tr> <td>0</td> <td>0</td> <td>0</td> <td>c = IMPEDIDO</td> </tr> </tbody> </table>	a	b	c	<- classified as	3.541	745	0	a = PROCEDENTE	1.157	2.872	0	b = IMPROCEDENTE	0	0	0	c = IMPEDIDO																									
a	b	c	<- classified as																																						
3.541	745	0	a = PROCEDENTE																																						
1.157	2.872	0	b = IMPROCEDENTE																																						
0	0	0	c = IMPEDIDO																																						
BayesNet	<p>=== Summary ===</p> <p><b>Correctly Classified Instances</b>      <b>5.746</b>    <b>69,104%</b> %  <b>Incorrectly Classified Instances</b>    <b>2.569</b>    <b>30,896%</b> %  Kappa statistic                            0,3755  Mean absolute error                      0,2299  Root mean squared error                0,3694  Relative absolute error                 69,034%  Root relative squared error            90,522%  Total Number of Instances            8.315</p>																																								
	<p>=== Detailed Accuracy By Class ===</p> <table border="1"> <thead> <tr> <th></th> <th>TP Rate</th> <th>FP Rate</th> <th>Precision</th> <th>Recall</th> <th>F-Measure</th> <th>ROC Area</th> <th>Class</th> </tr> </thead> <tbody> <tr> <td></td> <td>0.853</td> <td>0.481</td> <td>0.654</td> <td>0.853</td> <td>0.74</td> <td>0.769</td> <td>PROCEDENTE</td> </tr> <tr> <td></td> <td>0.519</td> <td>0.147</td> <td>0.768</td> <td>0.519</td> <td>0.62</td> <td>0.769</td> <td>IMPROCEDENTE</td> </tr> <tr> <td></td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>IMPEDIDO</td> </tr> <tr> <td><b>Weighted Avg.</b></td> <td><b>0.691</b></td> <td><b>0.319</b></td> <td><b>0.709</b></td> <td><b>0.691</b></td> <td><b>0.682</b></td> <td><b>0.769</b></td> <td></td> </tr> </tbody> </table>		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class		0.853	0.481	0.654	0.853	0.74	0.769	PROCEDENTE		0.519	0.147	0.768	0.519	0.62	0.769	IMPROCEDENTE		0	0	0	0	0	0	IMPEDIDO	<b>Weighted Avg.</b>	<b>0.691</b>	<b>0.319</b>	<b>0.709</b>	<b>0.691</b>	<b>0.682</b>	<b>0.769</b>	
		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class																																	
	0.853	0.481	0.654	0.853	0.74	0.769	PROCEDENTE																																		
	0.519	0.147	0.768	0.519	0.62	0.769	IMPROCEDENTE																																		
	0	0	0	0	0	0	IMPEDIDO																																		
<b>Weighted Avg.</b>	<b>0.691</b>	<b>0.319</b>	<b>0.709</b>	<b>0.691</b>	<b>0.682</b>	<b>0.769</b>																																			
<p>=== Confusion Matrix ===</p> <table border="1"> <thead> <tr> <th>a</th> <th>b</th> <th>c</th> <th>&lt;- classified as</th> </tr> </thead> <tbody> <tr> <td>3.654</td> <td>632</td> <td>0</td> <td>a = PROCEDENTE</td> </tr> <tr> <td>1.936</td> <td>2.092</td> <td>1</td> <td>b = IMPROCEDENTE</td> </tr> <tr> <td>0</td> <td>0</td> <td>0</td> <td>c = IMPEDIDO</td> </tr> </tbody> </table>	a	b	c	<- classified as	3.654	632	0	a = PROCEDENTE	1.936	2.092	1	b = IMPROCEDENTE	0	0	0	c = IMPEDIDO																									
a	b	c	<- classified as																																						
3.654	632	0	a = PROCEDENTE																																						
1.936	2.092	1	b = IMPROCEDENTE																																						
0	0	0	c = IMPEDIDO																																						

“continua”

Tabela 1C “continuação”

J48	<p>=== Summary ===</p> <p><b>Correctly Classified Instances</b>      <b>6.137</b>    <b>73,806%</b> %  <b>Incorrectly Classified Instances</b>    <b>2.178</b>    <b>26,194%</b> %  Kappa statistic                            0.4731  Mean absolute error                      0.2391  Root mean squared error                0.3457  Relative absolute error                 71,775%  Root relative squared error            84,725%  Total Number of Instances            8.315</p>																																								
	<p>=== Detailed Accuracy By Class ===</p> <table border="1"> <thead> <tr> <th></th> <th>TP Rate</th> <th>FP Rate</th> <th>Precision</th> <th>Recall</th> <th>F-Measure</th> <th>ROC Area</th> <th>Class</th> </tr> </thead> <tbody> <tr> <td></td> <td>0.822</td> <td>0.351</td> <td>0.714</td> <td>0.822</td> <td>0.764</td> <td>0.796</td> <td>PROCEDENTE</td> </tr> <tr> <td></td> <td>0.649</td> <td>0.178</td> <td>0.774</td> <td>0.649</td> <td>0.706</td> <td>0.796</td> <td>IMPROCEDENTE</td> </tr> <tr> <td></td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>IMPEDIDO</td> </tr> <tr> <td><b>Weighted Avg.</b></td> <td><b>0.738</b></td> <td><b>0.267</b></td> <td><b>0.743</b></td> <td><b>0.738</b></td> <td><b>0.736</b></td> <td><b>0.796</b></td> <td></td> </tr> </tbody> </table>		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class		0.822	0.351	0.714	0.822	0.764	0.796	PROCEDENTE		0.649	0.178	0.774	0.649	0.706	0.796	IMPROCEDENTE		0	0	0	0	0	0	IMPEDIDO	<b>Weighted Avg.</b>	<b>0.738</b>	<b>0.267</b>	<b>0.743</b>	<b>0.738</b>	<b>0.736</b>	<b>0.796</b>	
		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class																																	
	0.822	0.351	0.714	0.822	0.764	0.796	PROCEDENTE																																		
	0.649	0.178	0.774	0.649	0.706	0.796	IMPROCEDENTE																																		
	0	0	0	0	0	0	IMPEDIDO																																		
<b>Weighted Avg.</b>	<b>0.738</b>	<b>0.267</b>	<b>0.743</b>	<b>0.738</b>	<b>0.736</b>	<b>0.796</b>																																			
<p>=== Confusion Matrix ===</p> <table border="1"> <thead> <tr> <th>a</th> <th>b</th> <th>c</th> <th>&lt;- classified as</th> </tr> </thead> <tbody> <tr> <td>3.521</td> <td>765</td> <td>0</td> <td>a = PROCEDENTE</td> </tr> <tr> <td>1.413</td> <td>2.616</td> <td>0</td> <td>b = IMPROCEDENTE</td> </tr> <tr> <td>0</td> <td>0</td> <td>0</td> <td>c = IMPEDIDO</td> </tr> </tbody> </table>	a	b	c	<- classified as	3.521	765	0	a = PROCEDENTE	1.413	2.616	0	b = IMPROCEDENTE	0	0	0	c = IMPEDIDO																									
a	b	c	<- classified as																																						
3.521	765	0	a = PROCEDENTE																																						
1.413	2.616	0	b = IMPROCEDENTE																																						
0	0	0	c = IMPEDIDO																																						
REPTree	<p>=== Summary ===</p> <p><b>Correctly Classified Instances</b>      <b>6.285</b>    <b>75,586%</b> %  <b>Incorrectly Classified Instances</b>    <b>2.030</b>    <b>24,414%</b> %  Kappa statistic                            0.5089  Mean absolute error                      0.217  Root mean squared error                0.3337  Relative absolute error                 65,165%  Root relative squared error            81,785%  Total Number of Instances            8.315</p>																																								
	<p>=== Detailed Accuracy By Class ===</p> <table border="1"> <thead> <tr> <th></th> <th>TP Rate</th> <th>FP Rate</th> <th>Precision</th> <th>Recall</th> <th>F-Measure</th> <th>ROC Area</th> <th>Class</th> </tr> </thead> <tbody> <tr> <td></td> <td>0.837</td> <td>0.331</td> <td>0.729</td> <td>0.837</td> <td>0.78</td> <td>0.83</td> <td>PROCEDENTE</td> </tr> <tr> <td></td> <td>0.669</td> <td>0.163</td> <td>0.795</td> <td>0.669</td> <td>0.726</td> <td>0.83</td> <td>IMPROCEDENTE</td> </tr> <tr> <td></td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>IMPEDIDO</td> </tr> <tr> <td><b>Weighted Avg.</b></td> <td><b>0.756</b></td> <td><b>0.249</b></td> <td><b>0.761</b></td> <td><b>0.756</b></td> <td><b>0.754</b></td> <td><b>0.83</b></td> <td></td> </tr> </tbody> </table>		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class		0.837	0.331	0.729	0.837	0.78	0.83	PROCEDENTE		0.669	0.163	0.795	0.669	0.726	0.83	IMPROCEDENTE		0	0	0	0	0	0	IMPEDIDO	<b>Weighted Avg.</b>	<b>0.756</b>	<b>0.249</b>	<b>0.761</b>	<b>0.756</b>	<b>0.754</b>	<b>0.83</b>	
		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class																																	
	0.837	0.331	0.729	0.837	0.78	0.83	PROCEDENTE																																		
	0.669	0.163	0.795	0.669	0.726	0.83	IMPROCEDENTE																																		
	0	0	0	0	0	0	IMPEDIDO																																		
<b>Weighted Avg.</b>	<b>0.756</b>	<b>0.249</b>	<b>0.761</b>	<b>0.756</b>	<b>0.754</b>	<b>0.83</b>																																			
<p>=== Confusion Matrix ===</p> <table border="1"> <thead> <tr> <th>a</th> <th>b</th> <th>c</th> <th>&lt;- classified as</th> </tr> </thead> <tbody> <tr> <td>3.589</td> <td>697</td> <td>0</td> <td>a = PROCEDENTE</td> </tr> <tr> <td>1.333</td> <td>2.696</td> <td>0</td> <td>b = IMPROCEDENTE</td> </tr> <tr> <td>0</td> <td>0</td> <td>0</td> <td>c = IMPEDIDO</td> </tr> </tbody> </table>	a	b	c	<- classified as	3.589	697	0	a = PROCEDENTE	1.333	2.696	0	b = IMPROCEDENTE	0	0	0	c = IMPEDIDO																									
a	b	c	<- classified as																																						
3.589	697	0	a = PROCEDENTE																																						
1.333	2.696	0	b = IMPROCEDENTE																																						
0	0	0	c = IMPEDIDO																																						

“continua”

Tabela 1C “continuação”

<i>OneR</i>	<p>=== Summary ===</p> <p><b>Correctly Classified Instances</b>      <b>5.438</b>    <b>65,400%</b> %  <b>Incorrectly Classified Instances</b>    <b>2.877</b>    <b>34,600%</b> %  Kappa statistic                            0.3027  Mean absolute error                    0.2307  Root mean squared error                0.4803  Relative absolute error                 69,258%  Root relative squared error            117,700%  Total Number of Instances            8.315</p>																																								
	<p>=== Detailed Accuracy By Class ===</p> <table border="1"> <thead> <tr> <th></th> <th><i>TP Rate</i></th> <th><i>FP Rate</i></th> <th><i>Precision</i></th> <th><i>Recall</i></th> <th><i>F-Measure</i></th> <th><i>ROC Area</i></th> <th><b>Class</b></th> </tr> </thead> <tbody> <tr> <td></td> <td>0.769</td> <td>0.468</td> <td>0.636</td> <td>0.769</td> <td>0.696</td> <td>0.65</td> <td><b>PROCEDENTE</b></td> </tr> <tr> <td></td> <td>0.532</td> <td>0.231</td> <td>0.684</td> <td>0.532</td> <td>0.598</td> <td>0.65</td> <td><b>IMPROCEDENTE</b></td> </tr> <tr> <td></td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td><b>IMPEDIDO</b></td> </tr> <tr> <td><b>Weighted Avg.</b></td> <td><b>0.654</b></td> <td><b>0.253</b></td> <td><b>0.659</b></td> <td><b>0.654</b></td> <td><b>0.649</b></td> <td><b>0.65</b></td> <td></td> </tr> </tbody> </table>		<i>TP Rate</i>	<i>FP Rate</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>	<i>ROC Area</i>	<b>Class</b>		0.769	0.468	0.636	0.769	0.696	0.65	<b>PROCEDENTE</b>		0.532	0.231	0.684	0.532	0.598	0.65	<b>IMPROCEDENTE</b>		0	0	0	0	0	0	<b>IMPEDIDO</b>	<b>Weighted Avg.</b>	<b>0.654</b>	<b>0.253</b>	<b>0.659</b>	<b>0.654</b>	<b>0.649</b>	<b>0.65</b>	
		<i>TP Rate</i>	<i>FP Rate</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>	<i>ROC Area</i>	<b>Class</b>																																	
	0.769	0.468	0.636	0.769	0.696	0.65	<b>PROCEDENTE</b>																																		
	0.532	0.231	0.684	0.532	0.598	0.65	<b>IMPROCEDENTE</b>																																		
	0	0	0	0	0	0	<b>IMPEDIDO</b>																																		
<b>Weighted Avg.</b>	<b>0.654</b>	<b>0.253</b>	<b>0.659</b>	<b>0.654</b>	<b>0.649</b>	<b>0.65</b>																																			
<p>=== Confusion Matrix ===</p> <table border="1"> <thead> <tr> <th><b>a</b></th> <th><b>b</b></th> <th><b>c</b></th> <th><b>&lt;- classified as</b></th> </tr> </thead> <tbody> <tr> <td>3.295</td> <td>991</td> <td>0</td> <td>a = PROCEDENTE</td> </tr> <tr> <td>1.686</td> <td>2.143</td> <td>0</td> <td>b = IMPROCEDENTE</td> </tr> <tr> <td>0</td> <td>0</td> <td>0</td> <td>c = IMPEDIDO</td> </tr> </tbody> </table>	<b>a</b>	<b>b</b>	<b>c</b>	<b>&lt;- classified as</b>	3.295	991	0	a = PROCEDENTE	1.686	2.143	0	b = IMPROCEDENTE	0	0	0	c = IMPEDIDO																									
<b>a</b>	<b>b</b>	<b>c</b>	<b>&lt;- classified as</b>																																						
3.295	991	0	a = PROCEDENTE																																						
1.686	2.143	0	b = IMPROCEDENTE																																						
0	0	0	c = IMPEDIDO																																						
<i>PART</i>	<p>=== Summary ===</p> <p><b>Correctly Classified Instances</b>      <b>6.340</b>    <b>76,248%</b> %  <b>Incorrectly Classified Instances</b>    <b>1.975</b>    <b>23,752%</b> %  Kappa statistic                            0.524  Mean absolute error                    0.219  Root mean squared error                0.3309  Relative absolute error                 65,759%    %  Root relative squared error            81,097%  Total Number of Instances            8.315</p>																																								
	<p>=== Detailed Accuracy By Class ===</p> <table border="1"> <thead> <tr> <th></th> <th><i>TP Rate</i></th> <th><i>FP Rate</i></th> <th><i>Precision</i></th> <th><i>Recall</i></th> <th><i>F-Measure</i></th> <th><i>ROC Area</i></th> <th><b>Class</b></th> </tr> </thead> <tbody> <tr> <td></td> <td>0.787</td> <td>0.263</td> <td>0.761</td> <td>0.787</td> <td>0.773</td> <td>0.834</td> <td><b>PROCEDENTE</b></td> </tr> <tr> <td></td> <td>0.737</td> <td>0.213</td> <td>0.765</td> <td>0.737</td> <td>0.75</td> <td>0.834</td> <td><b>IMPROCEDENTE</b></td> </tr> <tr> <td></td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td><b>IMPEDIDO</b></td> </tr> <tr> <td><b>Weighted Avg.</b></td> <td><b>0.762</b></td> <td><b>0.239</b></td> <td><b>0.763</b></td> <td><b>0.762</b></td> <td><b>0.762</b></td> <td><b>0.834</b></td> <td></td> </tr> </tbody> </table>		<i>TP Rate</i>	<i>FP Rate</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>	<i>ROC Area</i>	<b>Class</b>		0.787	0.263	0.761	0.787	0.773	0.834	<b>PROCEDENTE</b>		0.737	0.213	0.765	0.737	0.75	0.834	<b>IMPROCEDENTE</b>		0	0	0	0	0	0	<b>IMPEDIDO</b>	<b>Weighted Avg.</b>	<b>0.762</b>	<b>0.239</b>	<b>0.763</b>	<b>0.762</b>	<b>0.762</b>	<b>0.834</b>	
		<i>TP Rate</i>	<i>FP Rate</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>	<i>ROC Area</i>	<b>Class</b>																																	
	0.787	0.263	0.761	0.787	0.773	0.834	<b>PROCEDENTE</b>																																		
	0.737	0.213	0.765	0.737	0.75	0.834	<b>IMPROCEDENTE</b>																																		
	0	0	0	0	0	0	<b>IMPEDIDO</b>																																		
<b>Weighted Avg.</b>	<b>0.762</b>	<b>0.239</b>	<b>0.763</b>	<b>0.762</b>	<b>0.762</b>	<b>0.834</b>																																			
<p>=== Confusion Matrix ===</p> <table border="1"> <thead> <tr> <th><b>a</b></th> <th><b>b</b></th> <th><b>c</b></th> <th><b>&lt;- classified as</b></th> </tr> </thead> <tbody> <tr> <td>3.372</td> <td>914</td> <td>0</td> <td>a = PROCEDENTE</td> </tr> <tr> <td>1.061</td> <td>2.968</td> <td>0</td> <td>b = IMPROCEDENTE</td> </tr> <tr> <td>0</td> <td>0</td> <td>0</td> <td>c = IMPEDIDO</td> </tr> </tbody> </table>	<b>a</b>	<b>b</b>	<b>c</b>	<b>&lt;- classified as</b>	3.372	914	0	a = PROCEDENTE	1.061	2.968	0	b = IMPROCEDENTE	0	0	0	c = IMPEDIDO																									
<b>a</b>	<b>b</b>	<b>c</b>	<b>&lt;- classified as</b>																																						
3.372	914	0	a = PROCEDENTE																																						
1.061	2.968	0	b = IMPROCEDENTE																																						
0	0	0	c = IMPEDIDO																																						

“continua”

Tabela 1C “continuação”

<i>SMO</i>	<p>=== Summary ===</p> <p><b>Correctly Classified Instances</b>      <b>5.936</b>    <b>71,369%</b> %  <b>Incorrectly Classified Instances</b>    <b>2.379</b>    <b>28,611%</b> %  Kappa statistic                            0.4243  Mean absolute error                    0.2858  Root mean squared error                0.371  Relative absolute error                 85,812%    %  Root relative squared error            81,097%  Total Number of Instances            8.315</p>																																								
	<p>=== Detailed Accuracy By Class ===</p> <table border="1"> <thead> <tr> <th></th> <th>TP Rate</th> <th>FP Rate</th> <th>Precision</th> <th>Recall</th> <th>F-Measure</th> <th>ROC Area</th> <th>Class</th> </tr> </thead> <tbody> <tr> <td></td> <td>0.801</td> <td>0.379</td> <td>0.692</td> <td>0.801</td> <td>0.743</td> <td>0.711</td> <td>PROCEDENTE</td> </tr> <tr> <td></td> <td>0.621</td> <td>0.199</td> <td>0.746</td> <td>0.621</td> <td>0.678</td> <td>0.711</td> <td>IMPROCEDENTE</td> </tr> <tr> <td></td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>IMPEDIDO</td> </tr> <tr> <td><b>Weighted Avg.</b></td> <td><b>0.714</b></td> <td><b>0.292</b></td> <td><b>0.718</b></td> <td><b>0.714</b></td> <td><b>0.711</b></td> <td><b>0.711</b></td> <td></td> </tr> </tbody> </table>		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class		0.801	0.379	0.692	0.801	0.743	0.711	PROCEDENTE		0.621	0.199	0.746	0.621	0.678	0.711	IMPROCEDENTE		0	0	0	0	0	0	IMPEDIDO	<b>Weighted Avg.</b>	<b>0.714</b>	<b>0.292</b>	<b>0.718</b>	<b>0.714</b>	<b>0.711</b>	<b>0.711</b>	
		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class																																	
	0.801	0.379	0.692	0.801	0.743	0.711	PROCEDENTE																																		
	0.621	0.199	0.746	0.621	0.678	0.711	IMPROCEDENTE																																		
	0	0	0	0	0	0	IMPEDIDO																																		
<b>Weighted Avg.</b>	<b>0.714</b>	<b>0.292</b>	<b>0.718</b>	<b>0.714</b>	<b>0.711</b>	<b>0.711</b>																																			
<p>=== Confusion Matrix ===</p> <table border="1"> <thead> <tr> <th>a</th> <th>b</th> <th>c</th> <th>&lt;- classified as</th> </tr> </thead> <tbody> <tr> <td>3.434</td> <td>852</td> <td>0</td> <td>a = PROCEDENTE</td> </tr> <tr> <td>1.527</td> <td>2.502</td> <td>0</td> <td>b = IMPROCEDENTE</td> </tr> <tr> <td>0</td> <td>0</td> <td>0</td> <td>c = IMPEDIDO</td> </tr> </tbody> </table>	a	b	c	<- classified as	3.434	852	0	a = PROCEDENTE	1.527	2.502	0	b = IMPROCEDENTE	0	0	0	c = IMPEDIDO																									
a	b	c	<- classified as																																						
3.434	852	0	a = PROCEDENTE																																						
1.527	2.502	0	b = IMPROCEDENTE																																						
0	0	0	c = IMPEDIDO																																						
<i>Logistic</i>	<p>=== Summary ===</p> <p><b>Correctly Classified Instances</b>      <b>6.092</b>    <b>73,265%</b> %  <b>Incorrectly Classified Instances</b>    <b>2.223</b>    <b>26,735%</b> %  Kappa statistic                            0.4627  Mean absolute error                    0.2394  Root mean squared error                0.3458  Relative absolute error                 71,891%    %  Root relative squared error            84,751%  Total Number of Instances            8.315</p>																																								
	<p>=== Detailed Accuracy By Class ===</p> <table border="1"> <thead> <tr> <th></th> <th>TP Rate</th> <th>FP Rate</th> <th>Precision</th> <th>Recall</th> <th>F-Measure</th> <th>ROC Area</th> <th>Class</th> </tr> </thead> <tbody> <tr> <td></td> <td>0.802</td> <td>0.341</td> <td>0.714</td> <td>0.802</td> <td>0.756</td> <td>0.799</td> <td>PROCEDENTE</td> </tr> <tr> <td></td> <td>0.659</td> <td>0.198</td> <td>0.758</td> <td>0.659</td> <td>0.705</td> <td>0.799</td> <td>IMPROCEDENTE</td> </tr> <tr> <td></td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>IMPEDIDO</td> </tr> <tr> <td><b>Weighted Avg.</b></td> <td><b>0.733</b></td> <td><b>0.272</b></td> <td><b>0.735</b></td> <td><b>0.733</b></td> <td><b>0.731</b></td> <td><b>0.799</b></td> <td></td> </tr> </tbody> </table>		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class		0.802	0.341	0.714	0.802	0.756	0.799	PROCEDENTE		0.659	0.198	0.758	0.659	0.705	0.799	IMPROCEDENTE		0	0	0	0	0	0	IMPEDIDO	<b>Weighted Avg.</b>	<b>0.733</b>	<b>0.272</b>	<b>0.735</b>	<b>0.733</b>	<b>0.731</b>	<b>0.799</b>	
		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class																																	
	0.802	0.341	0.714	0.802	0.756	0.799	PROCEDENTE																																		
	0.659	0.198	0.758	0.659	0.705	0.799	IMPROCEDENTE																																		
	0	0	0	0	0	0	IMPEDIDO																																		
<b>Weighted Avg.</b>	<b>0.733</b>	<b>0.272</b>	<b>0.735</b>	<b>0.733</b>	<b>0.731</b>	<b>0.799</b>																																			
<p>=== Confusion Matrix ===</p> <table border="1"> <thead> <tr> <th>a</th> <th>b</th> <th>c</th> <th>&lt;- classified as</th> </tr> </thead> <tbody> <tr> <td>3.438</td> <td>848</td> <td>0</td> <td>a = PROCEDENTE</td> </tr> <tr> <td>1.375</td> <td>2.654</td> <td>0</td> <td>b = IMPROCEDENTE</td> </tr> <tr> <td>0</td> <td>0</td> <td>0</td> <td>c = IMPEDIDO</td> </tr> </tbody> </table>	a	b	c	<- classified as	3.438	848	0	a = PROCEDENTE	1.375	2.654	0	b = IMPROCEDENTE	0	0	0	c = IMPEDIDO																									
a	b	c	<- classified as																																						
3.438	848	0	a = PROCEDENTE																																						
1.375	2.654	0	b = IMPROCEDENTE																																						
0	0	0	c = IMPEDIDO																																						

“continua”

Tabela 1C “continuação”

<i>Multilayer Perceptron</i>	<p>=== Summary ===</p> <p><b>Correctly Classified Instances</b>      <b>4.287</b>    <b>51,557%</b> %  <b>Incorrectly Classified Instances</b>    <b>4.028</b>    <b>48,443%</b> %  Kappa statistic                            0.0003  Mean absolute error                      0.3258  Root mean squared error                0.4964  Relative absolute error                 97,827%    %  Root relative squared error            121,646%  Total Number of Instances            8.315</p>																																								
	<p>=== Detailed Accuracy By Class ===</p> <table border="1"> <thead> <tr> <th></th> <th>TP Rate</th> <th>FP Rate</th> <th>Precision</th> <th>Recall</th> <th>F-Measure</th> <th>ROC Area</th> <th>Class</th> </tr> </thead> <tbody> <tr> <td></td> <td>1</td> <td>1</td> <td>0.516</td> <td>1</td> <td>0.68</td> <td>0.512</td> <td>PROCEDENTE</td> </tr> <tr> <td></td> <td>0</td> <td>0</td> <td>1</td> <td>0</td> <td>0</td> <td>0.512</td> <td>IMPROCEDENTE</td> </tr> <tr> <td></td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>IMPEDIDO</td> </tr> <tr> <td><b>Weighted Avg.</b></td> <td><b>0.516</b></td> <td><b>0.515</b></td> <td><b>0.75</b></td> <td><b>0.516</b></td> <td><b>0.351</b></td> <td><b>0.512</b></td> <td></td> </tr> </tbody> </table>		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class		1	1	0.516	1	0.68	0.512	PROCEDENTE		0	0	1	0	0	0.512	IMPROCEDENTE		0	0	0	0	0	0	IMPEDIDO	<b>Weighted Avg.</b>	<b>0.516</b>	<b>0.515</b>	<b>0.75</b>	<b>0.516</b>	<b>0.351</b>	<b>0.512</b>	
		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class																																	
	1	1	0.516	1	0.68	0.512	PROCEDENTE																																		
	0	0	1	0	0	0.512	IMPROCEDENTE																																		
	0	0	0	0	0	0	IMPEDIDO																																		
<b>Weighted Avg.</b>	<b>0.516</b>	<b>0.515</b>	<b>0.75</b>	<b>0.516</b>	<b>0.351</b>	<b>0.512</b>																																			
<p>=== Confusion Matrix ===</p> <table border="1"> <thead> <tr> <th>a</th> <th>b</th> <th>c</th> <th>&lt;- classified as</th> </tr> </thead> <tbody> <tr> <td>4.286</td> <td>0</td> <td>0</td> <td>a = PROCEDENTE</td> </tr> <tr> <td>4.028</td> <td>1</td> <td>0</td> <td>b = IMPROCEDENTE</td> </tr> <tr> <td>0</td> <td>0</td> <td>0</td> <td>c = IMPEDIDO</td> </tr> </tbody> </table>	a	b	c	<- classified as	4.286	0	0	a = PROCEDENTE	4.028	1	0	b = IMPROCEDENTE	0	0	0	c = IMPEDIDO																									
a	b	c	<- classified as																																						
4.286	0	0	a = PROCEDENTE																																						
4.028	1	0	b = IMPROCEDENTE																																						
0	0	0	c = IMPEDIDO																																						

<i>Decorate</i>	<p>=== Summary ===</p> <p><b>Correctly Classified Instances</b>      <b>6.649</b>    <b>79,964%</b> %  <b>Incorrectly Classified Instances</b>    <b>1.666</b>    <b>20,036%</b> %  Kappa statistic                            0.598  Mean absolute error                      0.3034  Root mean squared error                0.351  Relative absolute error                 91,101%    %  Root relative squared error            86,022%  Total Number of Instances            8.315</p>																																								
	<p>=== Detailed Accuracy By Class ===</p> <table border="1"> <thead> <tr> <th></th> <th>TP Rate</th> <th>FP Rate</th> <th>Precision</th> <th>Recall</th> <th>F-Measure</th> <th>ROC Area</th> <th>Class</th> </tr> </thead> <tbody> <tr> <td></td> <td>0.844</td> <td>0.247</td> <td>0.794</td> <td>0.844</td> <td>0.813</td> <td>0.876</td> <td>PROCEDENTE</td> </tr> <tr> <td></td> <td>0.753</td> <td>0.156</td> <td>0.819</td> <td>0.753</td> <td>0.785</td> <td>0.879</td> <td>IMPROCEDENTE</td> </tr> <tr> <td></td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> <td>IMPEDIDO</td> </tr> <tr> <td><b>Weighted Avg.</b></td> <td><b>0.8</b></td> <td><b>0.203</b></td> <td><b>0.801</b></td> <td><b>0.8</b></td> <td><b>0.799</b></td> <td><b>0.877</b></td> <td></td> </tr> </tbody> </table>		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class		0.844	0.247	0.794	0.844	0.813	0.876	PROCEDENTE		0.753	0.156	0.819	0.753	0.785	0.879	IMPROCEDENTE		0	0	0	0	0	0	IMPEDIDO	<b>Weighted Avg.</b>	<b>0.8</b>	<b>0.203</b>	<b>0.801</b>	<b>0.8</b>	<b>0.799</b>	<b>0.877</b>	
		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class																																	
	0.844	0.247	0.794	0.844	0.813	0.876	PROCEDENTE																																		
	0.753	0.156	0.819	0.753	0.785	0.879	IMPROCEDENTE																																		
	0	0	0	0	0	0	IMPEDIDO																																		
<b>Weighted Avg.</b>	<b>0.8</b>	<b>0.203</b>	<b>0.801</b>	<b>0.8</b>	<b>0.799</b>	<b>0.877</b>																																			
<p>=== Confusion Matrix ===</p> <table border="1"> <thead> <tr> <th>a</th> <th>b</th> <th>c</th> <th>&lt;- classified as</th> </tr> </thead> <tbody> <tr> <td>3.616</td> <td>670</td> <td>0</td> <td>a = PROCEDENTE</td> </tr> <tr> <td>995</td> <td>3.033</td> <td>0</td> <td>b = IMPROCEDENTE</td> </tr> <tr> <td>0</td> <td>0</td> <td>0</td> <td>c = IMPEDIDO</td> </tr> </tbody> </table>	a	b	c	<- classified as	3.616	670	0	a = PROCEDENTE	995	3.033	0	b = IMPROCEDENTE	0	0	0	c = IMPEDIDO																									
a	b	c	<- classified as																																						
3.616	670	0	a = PROCEDENTE																																						
995	3.033	0	b = IMPROCEDENTE																																						
0	0	0	c = IMPEDIDO																																						

“continua”

Tabela 1C “conclusão”

LogitBoost							
<b>=== Summary ===</b>							
Correctly Classified Instances	5.715	68,731%					
Incorrectly Classified Instances	2.600	31,269%					
Kappa statistic	0.3722						
Mean absolute error	0.2647						
Root mean squared error	0.3665						
Relative absolute error	79,478%						
Root relative squared error	89,808%						
Total Number of Instances	8.315						
<b>=== Detailed Accuracy By Class ===</b>							
	<i>TP Rate</i>	<i>FP Rate</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>	<i>ROC Area</i>	<b>Class</b>
	0.743	0.372	0.68	0.743	0.71	0.723	<b>PROCEDENTE</b>
	0.628	0.257	0.697	0.628	0.661	0.723	<b>IMPROCEDENTE</b>
	0	0	0	0	0	0	<b>IMPEDIDO</b>
<b>Weighted Avg.</b>	<b>0.687</b>	<b>0.316</b>	<b>0.688</b>	<b>0.687</b>	<b>0.686</b>	<b>0.723</b>	
<b>=== Confusion Matrix ===</b>							
<b>a</b>	<b>b</b>	<b>c</b>	<b>&lt;- classified as</b>				
3.183	1.103	0	a = PROCEDENTE				
1.497	2.532	0	b = IMPROCEDENTE				
0	0	0	c = IMPEDIDO				

## APÊNDICE D

Quadro D *Data mining*: técnica de associação – Algoritmos e suas características, usados no estudo

<b>Algoritmo</b>	<b>Característica</b>
<i>Apriori</i>	Classe que implementa algoritmo do tipo <i>apriori</i> . Interativamente, a classe reduz ao mínimo o apoio até que se encontre o número necessário de regras, com o mínimo determinado confiança.
<i>FilteredAssociator</i>	Classe que executa uma associação arbitrária sobre os dados que foram transmitidos por um filtro arbitrário. A estrutura do filtro é baseada exclusivamente no treinamento dos dados e testa casos que serão processados pelo filtro sem alterar sua estrutura.
<i>PredictiveApriori</i>	Classe que implementa o algoritmo de predição <i>apriori</i> , minando as regras de associação. Procura apoio limiar crescente para as ' <i>n</i> ' regras relativas, dando base ao intervalo de confiança corrigido.

## APÊNDICE E

Tabela 1E *Data mining*: técnica de associação – Resumo das estatísticas dos algoritmos, usados no estudo

Algoritmo	Resumo Estatístico
<i>Apriori</i>	<b>Apriori</b> =====  Minimum support: 0.5 (4157 instances) Minimum metric <confidence>: 0.9 Number of cycles performed: 10  <b>Generated sets of large itemsets:</b>  Size of set of large itemsets L(1): 4  Size of set of large itemsets L(2): 4  Size of set of large itemsets L(3): 1  <b>Best rules found:</b>
	1. Classe=RESIDENCIAL 5841 ==> Tensao=BT 5840 conf:(1)
	2. Ramo_Atividade=RA_380 5673 ==> Tensao=BT 5672 conf:(1)
	<b>3. Classe=RESIDENCIAL Ramo_Atividade=RA_380 5634 ==&gt; Tensao=BT 5633 conf:(1)</b>
	4. Retorno_SVC=PROCEDENTE 4286 ==> Tensao=BT 4274 conf:(1)
	5. Ramo_Atividade=RA_380 5673 ==> Classe=RESIDENCIAL 5634 conf:(0.99)
	<b>6. Tensao=BT Ramo_Atividade=RA_380 5672 ==&gt; Classe=RESIDENCIAL 5633 conf:(0.99)</b>
	7. Ramo_Atividade=RA_380 5673 ==> Tensao=BT Classe=RESIDENCIAL 5633 conf:(0.99)
	8. Classe=RESIDENCIAL 5841 ==> Ramo_Atividade=RA_380 5634 conf:(0.96)
	<b>9. Tensao=BT Classe=RESIDENCIAL 5840 ==&gt; Ramo_Atividade=RA_380 5633 conf:(0.96)</b>
10. Classe=RESIDENCIAL 5841 ==> Tensao=BT Ramo_Atividade=RA_380 5633 conf:(0.96)	

“continua”

Tabela 1E “continuação”

<i>FilteredAsso ciator</i>	<b>FilteredAssociator</b> =====  Minimum support: 0.5 (4157 instances) Minimum metric <confidence>: 0.9 Number of cycles performed: 10  <b>Generated sets of large itemsets:</b>  Size of set of large itemsets L(1): 4  Size of set of large itemsets L(2): 4  Size of set of large itemsets L(3): 1  <b>Best rules found:</b>
	1. Classe=RESIDENCIAL 5841 ==> Tensao=BT 5840 conf:(1)
	2. Ramo_Atividade=RA_380 5673 ==> Tensao=BT 5672 conf:(1)
	<b>3. Classe=RESIDENCIAL Ramo_Atividade=RA_380 5634 ==&gt; Tensao=BT 5633 conf:(1)</b>
	4. Retorno_SVC=PROCEDENTE 4286 ==> Tensao=BT 4274 conf:(1)
	5. Ramo_Atividade=RA_380 5673 ==> Classe=RESIDENCIAL 5634 conf:(0.99)
	<b>6. Tensao=BT Ramo_Atividade=RA_380 5672 ==&gt; Classe=RESIDENCIAL 5633 conf:(0.99)</b>
	7. Ramo_Atividade=RA_380 5673 ==> Tensao=BT Classe=RESIDENCIAL 5633 conf:(0.99)
	8. Classe=RESIDENCIAL 5841 ==> Ramo_Atividade=RA_380 5634 conf:(0.96)
	<b>9. Tensao=BT Classe=RESIDENCIAL 5840 ==&gt; Ramo_Atividade=RA_380 5633 conf:(0.96)</b>
10. Classe=RESIDENCIAL 5841 ==> Tensao=BT Ramo_Atividade=RA_380 5633 conf:(0.96)	

## APÊNDICE E

Tabela 1E Conclusão

*PredictiveApriori*

PredictiveApriori									
*****									
Best rules found:									
1.	LOCAL_LEITURA=L_000	MOT_ANTIGO=BI	GR_EXEC_INSPECAO=EMPREGATEIRA	RETORNO_SVC=PROCEDENTE	295	***	TENSAO=BT	295	acc:(0.999)
2.	LOCAL_LEITURA=L_0030	MOT_ANTIGO=D	RETORNO_SVC=IMPEDEDO	294	***	TENSAO=BT	294	acc:(0.999)	
3.	LOCAL_LEITURA=L_0031	MOT_ANTIGO=GC	RETORNO_SVC=IMPEDEDO	294	***	TENSAO=BT	294	acc:(0.999)	
4.	LOCAL_LEITURA=L_0013	MOT_ANTIGO=GE	RETORNO_SVC=PROCEDENTE	293	***	TENSAO=BT	293	acc:(0.999)	
5.	LOCAL_LEITURA=L_0018	MOT_ANTIGO=GE	RETORNO_SVC=PROCEDENTE	292	***	TENSAO=BT	292	acc:(0.999)	
6.	MOT_ANTIGO=FM	RETORNO_SVC=PROCEDENTE	291	***	GR_EXEC_INSPECAO=SERVICO_PROPPIO		291	acc:(0.999)	
7.	LOCAL_LEITURA=L_0108	MOT_ANTIGO=CJ	RETORNO_SVC=PROCEDENTE	290	***	TENSAO=BT	290	acc:(0.999)	
8.	LOCAL_LEITURA=L_0530	MOT_ANTIGO=GE	RETORNO_SVC=PROCEDENTE	290	***	TENSAO=BT	290	acc:(0.999)	
9.	LOCAL_LEITURA=L_0120	MOT_ANTIGO=GC	GR_EXEC_INSPECAO=EMPREGATEIRA	289	***	TENSAO=BT	289	acc:(0.99999)	
10.	LOCAL_LEITURA=L_0128	GR_EXEC_INSPECAO=SERVICO_PROPPIO	289	***	TENSAO=BT		289	acc:(0.99999)	
11.	LOCAL_LEITURA=L_0014	GR_EXEC_INSPECAO=EMPREGATEIRA	RETORNO_SVC=PROCEDENTE	289	***	TENSAO=BT	289	acc:(0.99999)	
12.	LOCAL_LEITURA=L_0021	RETORNO_SVC=PROCEDENTE	289	***	TENSAO=BT		289	acc:(0.99999)	
13.	LOCAL_LEITURA=L_0017	RETORNO_SVC=PROCEDENTE	279	***	TENSAO=BT		279	acc:(0.99999)	
14.	LOCAL_LEITURA=L_0530	MOT_ANTIGO=GC	GR_EXEC_INSPECAO=SERVICO_PROPPIO	278	***	TENSAO=BT	278	acc:(0.99999)	
15.	LOCAL_LEITURA=L_0104	RETORNO_SVC=IMPEDEDO	275	***	TENSAO=BT		275	acc:(0.99999)	
16.	LOCAL_LEITURA=L_0019	MOT_ANTIGO=GC	RETORNO_SVC=IMPEDEDO	274	***	TENSAO=BT	274	acc:(0.99999)	
17.	LOCAL_LEITURA=L_0011	MOT_ANTIGO=GC	GR_EXEC_INSPECAO=SERVICO_PROPPIO	273	***	TENSAO=BT	273	acc:(0.99999)	
18.	LOCAL_LEITURA=L_0011	MOT_ANTIGO=GE	RETORNO_SVC=PROCEDENTE	273	***	TENSAO=BT	273	acc:(0.99999)	
19.	LOCAL_LEITURA=L_0108	RETORNO_SVC=IMPEDEDO	272	***	TENSAO=BT		272	acc:(0.99999)	
20.	LOCAL_LEITURA=L_0031	RETORNO_SVC=PROCEDENTE	272	***	TENSAO=BT		272	acc:(0.99999)	
21.	LOCAL_LEITURA=L_0021	MOT_ANTIGO=D	269	***	TENSAO=BT		269	acc:(0.99999)	
22.	LOCAL_LEITURA=L_0108	MOT_ANTIGO=GC	267	***	TENSAO=BT		267	acc:(0.99999)	
23.	LOCAL_LEITURA=L_0017	GR_EXEC_INSPECAO=EMPREGATEIRA	RETORNO_SVC=PROCEDENTE	267	***	TENSAO=BT	267	acc:(0.99999)	
24.	LOCAL_LEITURA=L_0012	MOT_ANTIGO=GE	RETORNO_SVC=PROCEDENTE	266	***	TENSAO=BT	266	acc:(0.99999)	
25.	LOCAL_LEITURA=L_0001	MOT_ANTIGO=GC	GR_EXEC_INSPECAO=EMPREGATEIRA	RETORNO_SVC=PROCEDENTE	264	***	TENSAO=BT	264	acc:(0.99499)
26.	LOCAL_LEITURA=L_0101	GR_EXEC_INSPECAO=SERVICO_PROPPIO	RETORNO_SVC=PROCEDENTE	263	***	TENSAO=BT	263	acc:(0.99499)	
27.	MOT_ANTIGO=OS	RETORNO_SVC=PROCEDENTE	263	***	TENSAO=BT		263	acc:(0.99499)	
28.	LOCAL_LEITURA=L_0108	RETORNO_SVC=PROCEDENTE	262	***	TENSAO=BT		262	acc:(0.99499)	
29.	LOCAL_LEITURA=L_0230	MOT_ANTIGO=GC	GR_EXEC_INSPECAO=EMPREGATEIRA	262	***	TENSAO=BT	262	acc:(0.99499)	
30.	MOT_ANTIGO=9	GR_EXEC_INSPECAO=EMPREGATEIRA	RETORNO_SVC=PROCEDENTE	262	***	TENSAO=BT	262	acc:(0.99499)	
31.	LOCAL_LEITURA=L_0017	MOT_ANTIGO=D	261	***	TENSAO=BT		261	acc:(0.99499)	
32.	LOCAL_LEITURA=L_0021	RETORNO_SVC=PROCEDENTE	261	***	TENSAO=BT		261	acc:(0.99499)	
33.	LOCAL_LEITURA=L_0021	GR_EXEC_INSPECAO=SERVICO_PROPPIO	261	***	TENSAO=BT		261	acc:(0.99499)	
34.	MOT_ANTIGO=13	RETORNO_SVC=IMPEDEDO	261	***	TENSAO=BT		261	acc:(0.99499)	
35.	LOCAL_LEITURA=L_0018	GR_EXEC_INSPECAO=EMPREGATEIRA	RETORNO_SVC=PROCEDENTE	249	***	TENSAO=BT	249	acc:(0.99499)	

**APÊNDICE F**Quadro F1 *Data mining*: técnica de agrupamento – Algoritmos e suas características, usados no estudo

<b>Algoritmo</b>	<b>Característica</b>
<i>SimpleKMeans</i>	Forma <i>cluster</i> dos dados usando o algoritmo <i>k-means</i> . Pode usar tanto a distância Euclidiana (padrão) quanto à distância Manhattan.
<i>FarthestFirst</i>	Aloca os dados em <i>clusters</i> usando o algoritmo <i>FarthestFirst</i> . Aproxima-se do cluster modelado pelo algoritmo <i>k-means</i> .

## APÊNDICE G

Tabela G1 *Data mining*: técnica de agrupamento – Resumo das estatísticas dos algoritmos, usados no estudo

Algoritmo	Característica						
<i>SimpleK Means</i>	=== kMeans ===						
	Number of iterations: 3						
	Within cluster sum of squared errors: 27280.0						
	Missing values globally replaced with mean/mode						
	=== Cluster centroids ===						
			CLUSTERS				
		Attribute Full Data	0	1	2	3	
			-8315	-19307	-5488	-443	-732
	Tensao	BT	BT	BT	BT	BT	
	Dia_Leitura	13_D	13_D	13_D	03_D	15_D	
Local_Leitura	L_8015	L_8015	L_8015	L_0104	L_0530		
Rota_Leitura	R_30	R_30	R_30	R_33	R_25		
Classe	RESIDENCIAL	RESIDENCIAL	RESIDENCIAL	RESIDENCIAL	RESIDENCIAL		
Ramo_Atividade	RA_380	RA_380	RA_380	RA_380	RA_380		
Retorno_SVC	PROCEDENTE	PROCEDENTE	PROCEDENTE	PROCEDENTE	PROCEDENTE		
=== Clustered instances ===							
	0	5488	66%				
	1	443	5%				
	2	732	9%				
	3	1.652	20%				

“continua”

Tabela G1 “conclusão”

*Farthest  
First*

=== FarthestFirst ===				
=== Cluster centroids ===				
Atribute Full Data 8315	0	1	2	3
Tensao	BT	MT	AT	BT
Dia_Leitura	03_D	44_D	90_D	01_D
Local_Leitura	L_0631	L_0109	L_0126	L_0101
Rota_Leitura	R_12	R_02	R_01	R_13
Classe	RESIDENCIAL	OUTROS_SERVICOS_E_OUTRAS_ATIV	INDUSTRIAL	COMERCIAL
Ramo_Atividade	RA_380	RA_999	RA_999	RA_136
Retorno_SVC	IMPROCEDENTE	PROCEDENTE	IMPROCEDENTE	PROCEDENTE
=== Clustered Instances ===				
<b>0</b>	6 626	80%		
<b>1</b>	846	10%		
<b>2</b>	401	5%		
<b>3</b>	442	5%		

## APÊNDICE H

Figura H1 Perdas DSS: tela principal da aplicação

Cálculos registrados				
Nome	Data de Início	Data Final	Estatísticas	Excluir
2011 - 2008 a 2011 Calculo Completo	01/02/2008	28/02/2011	141744	

Fonte: Resultados da pesquisa – *software* Perdas DSS

## APÊNDICE I

Figura 11 Perdas DSS: SEGURANÇA – telas de cadastro e consulta/edição de usuários

Cadastro/Edição de Usuários

Última mensagem do sistema:

Nome:

Login:

Senha:

Confirmar senha:

Nível de acesso:  Super Usuário  Administrador  Usuário Padrão

Concluir

Fonte: Resultados da pesquisa – *software* Perdas DSS

Consulta de Usuários					
 Última mensagem do sistema: Usuário criado com sucesso					
Análises registradas					
Nome	Login	Ativo?	Editar	Trocar Senha	Excluir
Administrador	admin	<input checked="" type="checkbox"/>			
gustavo	gustavo	<input checked="" type="checkbox"/>			
Luiz Gustavo	admin01	<input checked="" type="checkbox"/>			
André Medeiros	andremedeiros	<input checked="" type="checkbox"/>			

Fonte: Resultados da pesquisa – *software* Perdas DSS

## APÊNDICE J

Figura 1J Perdas DSS: IMPORTAÇÃO – tela de importação das inspeções realizadas

The screenshot displays the 'Importação de Inspeções' screen. On the left, there is a sidebar with the following menu items:

- Segurança
  - Cadastro de Usuários
  - Consulta/Edição de Usuários
- Importação
  - Importação de Inspeções Realizadas
  - Importação de Instalações Suspeitas
  - Importação de Classes de Atividade
- Operacional
  - Gerar estatísticas das inspeções realizadas
  - Analisar as estatísticas geradas
  - Análise das Instalações Suspeitas (Rel.)
  - Seleção de alvos (Gráficos)
- Planejamento
  - Alocação de equipes por gerência/complexidade
  - Configurar Locais
- RD112
  - Página Inicial
  - Encerrar sessão

The main content area is titled 'Importação de Inspeções' and contains the instruction: 'Selecione o arquivo que contenha as análises a serem importadas:'. Below this is a file selection area with an 'Adicionar' button. A yellow box highlights a portion of the interface below the file selection area.

At the bottom of the screen, there is a table titled 'Inspeções registradas' with the following data:

Data da análise	Quantidade	Excluir
20/01/2011	187	<input type="checkbox"/>
21/01/2011	215	<input type="checkbox"/>
22/01/2011	13	<input type="checkbox"/>
23/01/2011	3	<input type="checkbox"/>
24/01/2011	195	<input type="checkbox"/>
25/01/2011	260	<input type="checkbox"/>
26/01/2011	186	<input type="checkbox"/>
27/01/2011	221	<input type="checkbox"/>
28/01/2011	215	<input type="checkbox"/>
29/01/2011	1	<input type="checkbox"/>
30/01/2011	5	<input type="checkbox"/>

Fonte: Resultados da pesquisa – *software* Perdas DSS

## APÊNDICE K

Figura 1K Perdas DSS: IMPORTAÇÃO – tela de importação das instalações suspeitas

The screenshot shows the 'Importação das Instalações Suspeitas' screen. On the left is a sidebar menu with categories: Segurança, Importação, Operacional, and Planejamento. The 'Importação' category is expanded, showing options like 'Importação de Inspeções Realizadas', 'Importação de Instalações Suspeitas', and 'Importação de Classes de Atividade'. The main area contains a text prompt 'Selecione o arquivo que contenha as análises para liberação a serem importadas:' and an 'Adicionar' button. Below this is a table titled 'Instalações registradas' with columns for 'Data de importação', 'Quantidade', and 'Excluir'.

Instalações registradas		
Data de importação	Quantidade	Excluir
09/06/2011	403655	✘
10/06/2011	788027	✘
12/06/2011	362254	✘
13/06/2011	221074	✘
14/06/2011	531637	✘

Fonte: Resultados da pesquisa – *software* Perdas DSS

## APÊNDICE L

Figura L1 Perdas DSS: OPERACIONAL – tela de geração das estatísticas das insp. realizadas

The screenshot displays the 'Gerar Estatísticas' (Generate Statistics) screen in the PERDASS DSS - 20011 application. The interface is divided into a left sidebar menu and a main content area.

**Sidebar Menu:**

- Segurança
  - Cadastro de Usuários
  - Consulta/Edição de Usuários
- Importação
  - Importação de Inspeções Realizadas
  - Importação de Instalações Suspeitas
  - Importação de Classes de Atividade
- Operacional
  - Gerar estatísticas das inspeções realizadas
  - Analisar as estatísticas gerais
  - Análise das Instalações Suspeitas (Res.)
  - Seleção de alvos (Gráficos)
- Planejamento
  - Alocação de equipes por gerência/complexidade
  - Configurar Local
- PERDASS
  - Página Inicial
  - Encerrar sessão

**Main Content Area:**

- Última mensagem do sistema:** Cálculo: PERDASS DSS - 20011 gerado com sucesso.
- Nome do Cálculo:** PERDASS DSS - 20011
- Data de início:** 03/04/2011
- Data de Término:** 04/02/2011
- Detalhamento:** ESTADÍSTICA DAS INSPEÇÕES REALIZADAS | BIM DE 2011.
- Dimensões para a geração das estatísticas:**
  - Motivo
  - Ramo de Atividade
  - Unidade de Lettura
  - Classe
  - Composto
  - Número de Fases
  - Rota
  - Local
- Botão:** Calcular Probabilidades

Fonte: Resultados da pesquisa – *software* Perdas DSS

## APÊNDICE M

Figura 1M Perdas DSS: OPERACIONAL – tela de análise das estatísticas geradas (visão 1)

The screenshot displays the 'Perdas DSS: OPERACIONAL' software interface. On the left is a navigation menu with categories: 'Ingeniería', 'Inspección', 'Operaciones', 'Planamiento', and 'HISTÓRICO'. The main area contains analysis settings:

- Selecciona a dimensão a ser analisada:**
  - Motivo
  - Razo de Atividade
  - Unidade de Letra
  - Classe
  - Conjunto
  - Numero de Fases
  - Rota
  - Local
- Selecione um dos cálculos gerados:** 2011 - 2000 a 2011 Cálculo Completo - Data de Início: 01/02/2008 - Data de Término: 20/03/2011
- Utilizar filtros avançados?**
- Configurações extras para dimensão: Conjunto:**
  - 1989 o caracter (-) para separar os parâmetros de busca
- Filtros avançados de busca:**
  - Mostrar análises:  procedentes  e  tratadas de qual  que  %
- Botão:** Gerar estatísticas

Below the settings is a table titled 'Análises registradas' with the following data:

Combinações *	Elem. Inaprovechados *	Elem. Provedentes *	Prob. Inaprovechada (%) *	Prob. Provedente (%) *	Tam. da Amostra *
S QC DS	0	26	0,00	100,00	26
QC S DS	0	21	0,00	100,00	21
S CP DS QC	0	19	0,00	100,00	19
BFASCO DS QC 2003 1790 01 2001	0	15	0,00	100,00	15
BFASCO DS					

Fonte: Resultados da pesquisa – *software* Perdas DSS

## APÊNDICE N

Figura 1N Perdas DSS: OPERACIONAL – tela de análise das estatísticas geradas (visão 2)

Análise registrada					
Combinações *	Elem. Inaproveitadas †	Elem. Provedentes ‡	Prot. Inaproveitadas (%) †	Prot. Provedentes (%) ‡	Tam. de Amostra †
X907	35	0	100,00	0,00	35
X905	20	2	90,91	9,09	22
X904	115	4	96,54	3,46	119
X903	20	47	29,26	70,74	67
X902	1481	388	81,67	18,33	1869
X901	110471	58895	85,24	14,76	169366
TRFASCO Q5 6357 1485 18 3064	1	0	100,00	0,00	1
TRFASCO Q5 6358 6361 08 3064	1	0	100,00	0,00	1
TRFASCO Q5 6371 6361 01 3064	1	0	100,00	0,00	1
TRFASCO Q5 6321 4361 27 3064	1	0	100,00	0,00	1
TRFASCO Q5 6311 6361 18 3064	1	0	100,00	0,00	1
TRFASCO Q5 6321 6361 15 3064	1	0	100,00	0,00	1
TRFASCO Q5 6171 6361 05 3064	1	0	100,00	0,00	1

Fonte: Resultados da pesquisa – *software* Perdas DSS

## APÊNDICE O

Figura 10 Perdas DSS: OPERACIONAL – tela de análise instalações suspeitas

Segurança

- ▣ Cadastro de Usuários
- ▣ Consulta/Edição de Usuários

Operacional

- ▣ Gerar estatísticas das inspeções realizadas
- ▣ Analisar as estatísticas geradas
- ▣ Análise das Instalações Suspeitas (Rel.)
- ▣ Seleção de alvos (Gráficos)

Planejamento

- ▣ Alocação de equipes por gerência/complexidade
- ▣ Configurar Locais

PD212

- ▣ Página Inicial
- ▣ Encerrar sessão

Análise das instalações suspeitas

Selecione um dos cálculos gerados: PERDASS DSS - 20011 - Data de Inicio: 03/01/2011 - Data de Término: 28/02/2011

Data de Inicio:

Data de Término:

Nome:

Gerar Cálculo de Relatório

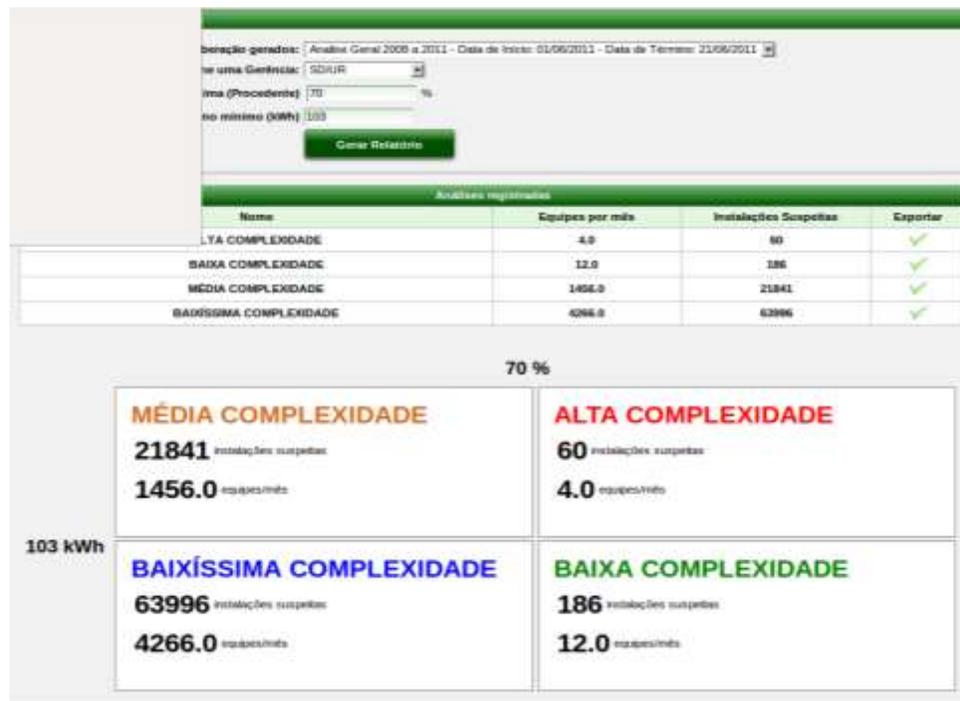
Cálculos de Liberação registrados

Nome	Data de Inicio	Data Final	Estatísticas	Excluir	Visualizar
Análise Geral 2006 a 2011	01/06/2011	21/06/2011	2306647		

Fonte: Resultados da pesquisa – *software* Perdas DSS

## APÊNDICE P

Figura 1P Perdas DSS: OPERACIONAL – tela de com o gráfico de alocação de equipes de campo



Fonte: Resultados da pesquisa – *software* Perdas DSS

## APÊNDICE Q

Figura 1Q Perdas DSS: OPERACIONAL – tela para a seleção de alvos



Fonte: Resultados da pesquisa – *software* Perdas DSS

## APÊNDICE R

Figura 1R Perdas DSS: PLANEJAMENTO – tela para alocação de equipes por complexidade

The screenshot shows the 'Análise das instalações suspeitas' screen. On the left is a sidebar menu with the following items:

- Segurança
  - Cadastro de Usuários
  - Consulta/Edição de Usuários
- Importação
  - Importação de Inspeções Realizadas
  - Importação de Instalações Suspeitas
  - Importação de Classes de Atividade
- Operacional
  - Gerar estatísticas das inspeções realizadas
  - Analisar as estatísticas geradas
  - Análise das Instalações Suspeitas (Rel.)
  - Seleção de alvos (Gráficos)
- Planejamento
  - Alocação de equipes por gerência/complexidade
  - Configurar Locais
- PD212
  - Página Inicial
  - Encerrar sessão

The main content area is titled 'Análise das instalações suspeitas' and contains the following form elements:

- Seleção um dos cálculos de liberação gerados:
- Seleção uma Gerência:
- Probabilidade mínima (Procedente):  %
- Retorno mínimo (KWh):
- Gerar Relatório (button)

Fonte: Resultados da pesquisa – *software* Perdas DSS

## APÊNDICE S

Figura 1S Perdas DSS: PLANEJAMENTO – tela para configuração de locais

	Código #	Número de Inspeções	Inspeções por Equipe	Atualizar
<b>Segurança</b>	DESCONHECIDO	3000	15	✓
▣ Cadastro de Usuários	SD/DV	3000	15	✓
▣ Consulta/Edição de Usuários	SD/IGV	3000	15	✓
<b>Importação</b>	SD/MP	3000	15	✓
▣ Importação de Inspeções Realizadas	SD/JF	3000	15	✓
▣ Importação de Instalações Suspeitas	SD/MC	3000	15	✓
▣ Importação de Classes de Atividade	SD/MP	3000	15	✓
<b>Operacional</b>	SD/MT	3000	15	✓
▣ Gerar estatísticas das inspeções realizadas	SD/PA	3000	15	✓
▣ Analisar as estatísticas geradas	SD/PA	3000	15	✓
▣ Análise das Instalações Suspeitas (Rel.)	SD/PM	3000	15	✓
▣ Seleção de alvos (Gráficos)	SD/PR	3000	15	✓
<b>Planejamento</b>	SD/PS	3000	15	✓
▣ Alocação de equipes por gerência/complexidade	SD/SJ	3000	15	✓
▣ Configurar Locais	SD/TO	3000	15	✓
<b>PO212</b>	SD/UL	3000	15	✓
▣ Página Inicial	SD/UR	3000	15	✓
▣ Encerrar sessão	SD/VR	3000	15	✓

Fonte: Resultados da pesquisa – *software* Perdas DSS

**ANEXO**

## ANEXO A

Quadro 1A DEE: Plano Operacional Padrão (POP) usado na seleção e liberação de inspeção dos consumidores suspeitos de irregularidade

<b>Gerência de Coordenação da Proteção da Receita e Utilização de Energia – RC/PR</b>		<b>POP RC/PR-043</b>	
<b>PROCESSO:</b> Gestão das Perdas		<b>Data:</b> 20/05/2008	
<b>Tarefa:</b> Liberação de Inspeções		<b>Revisado em:</b>	
<b>DESCRIÇÃO RESUMIDA</b>	<b>QUEM É O RESPONSÁVEL</b>	<b>O QUE FAZER</b>	<b>COMO EXECUTAR</b>
1) Estabelecer as localidades e a quantidade de inspeções a serem geradas.	Selecionador de Inspeção - BT e MT	Avaliar os recursos disponíveis para a execução de inspeções na área.	Ajustando com o SvC a capacidade de execução de inspeções, considerando o acordo de nível de serviço, inspeções pendentes, nível de tensão, complexidade das inspeções, entre outros.
2) Abrir o Relatório.	Selecionador de Inspeção - BT e MT	Abrir o Relatório de Seleção de Alvos para Inspeção.	Através do SGC – ambiente BW Gestão da Informação – transação RRMX. Ver Infopak Nota: Pode-se acessar também por Iniciar => Business Explorer => Analyser => BW Gestão da Informação. Ver Infopak
3) Acessar as Queries ou as Visões Gravadas do “Seleção”.	Selecionador de Inspeção - BT e MT	Entrar nas Queries ou Visões Gravadas do “Seleção”.	Clicando em Queries => Seleção de Alvos para acessar o Relatório na versão padrão ou clicando em Visões gravadas => Seleção de Alvos, quando quiser acessar o Relatório já direcionado para uma estratégia específica. Acessar pelo modelo “Mod Seleccion Alvos – Final”

“continua”

Quadro 1A “continuação”

<b>Gerência de Coordenação da Proteção da Receita e Utilização de Energia – RC/PR</b>		<b>POP RC/PR-043</b>	
<b>PROCESSO:</b> Gestão das Perdas		<b>Data:</b> 20/05/2008	
<b>Tarefa:</b> Liberação de Inspeções		<b>Revisado em:</b>	
<b>DESCRIÇÃO RESUMIDA</b>	<b>QUEM É O RESPONSÁVEL</b>	<b>O QUE FAZER</b>	<b>COMO EXECUTAR</b>
4) Se desejar gerar inspeção para instalações que tiveram Notas AP.	Selecionador de Inspeção - BT e MT	Entrar com as instalações que tiveram notas AP (Priorização de Inspeção) para terem inspeções geradas	Acessando planilha salva pelo próprio selecionador anteriormente, conforme critérios estabelecidos no POP RC/PR 033 – Tratamento de Notas AP. Entrar a data dos motivos e a Hierarquia por Base Operativa, copiar as instalações da planilha e colar no filtro inicial, no campo “Instalação”, clicando em  e em  . Nota: Não usar “ctrl v” para “colar” as instalações.
5) Se instalações de notas AP, gerar notas de inspeção.	Selecionador de Inspeção - BT e MT	Gerar inspeções para as instalações pretendidas	Conforme passo 7.

“continua”

Quadro 1A “continuação”

Gerência de Coordenação da Proteção da Receita e Utilização de Energia – RC/PR		POP RC/PR-043	
PROCESSO: Gestão das Perdas		Data: 20/05/2008	
Tarefa: Liberação de Inspeções		Revisado em:	
DESCRIÇÃO RESUMIDA	QUEM É O RESPONSÁVEL	O QUE FAZER	COMO EXECUTAR
6) Definir e aplicar estratégia para a seleção de alvos.	Selecionador de Inspeção - BT e MT	Aplicar a estratégia mais adequada para a seleção, considerando critérios definidos pela Gestão das Perdas, recursos disponíveis, prioridades da área, a possibilidade de acerto, de recuperação e de incremento.	Preenchendo os campos de filtragem e escolhendo a forma mais adequada de exibição do relatório (visão gravada ou mesmo uma query), tal como pelo peso dos Motivos, o motivo da suspeita, a Unidade de Leitura ou Objeto de Ligação com maior número de instalações, pelo Ramo de Atividade, por Denúncias. 
7) Gerar Notas de Inspeção.	Selecionador de Inspeção - BT e MT	Gerar inspeções para execução pelo SVC.	Selecionando todas as instalações que devem realmente ter inspeção gerada, clicando com o botão direito do mouse nessas células e em “Manter valor do filtro”. Se desejar inserir observações para o SvC, clicar em “Gerar obs para notas”, inseri-las e depois clicar em “Gerar Notas de Serviço”. Ver Infopak

“continua”

Quadro 1A “conclusão”

Gerência de Coordenação da Proteção da Receita e Utilização de Energia – RC/PR		POP RC/PR-043	
PROCESSO: Gestão das Perdas		Data: 20/05/2008	
Tarefa: Liberação de Inspeções		Revisado em:	
DESCRIÇÃO RESUMIDA	QUEM É O RESPONSÁVEL	O QUE FAZER	COMO EXECUTAR
8) Salvar Planilha.	Selecionador de Inspeção - BT e MT	Salvar planilha com as Notas de Inspeção e instalações, para acompanhamento da execução das inspeções.	Clicando em  , em  e nomeando o arquivo indicando o local e período das inspeções (ex.: Inspeções Divinópolis semana 28_04 a 02_05_08). Nota: Antes de salvar a planilha, excluir as instalações em que no campo “Mensagem”, o texto for diferente de “Nota gerada com sucesso!”. Dessa forma, não serão salvas instalações que já possuem notas de serviços ou de inspeção já geradas. Ver Infopak
9) Encaminhar planilha com as instalações com notas geradas no período.	Selecionador de Inspeção - BT e MT	Encaminhar planilha com as Notas de Inspeção para o SvC e Agente de Negócios, nos casos de clientes de MT ou Corporativos.	Acessando a planilha e a encaminhando por e-mail, para o responsável pelo SvC e Agente de Negócios, indicando, quando necessário, a necessidade de priorização de determinado serviço.
Preparado por: Colaborador X/DEE, Colaborador A/DEE.		Aprovado: Coordenador/DEE.	

Fonte: Arquivo DEE.