

5 ESTUDO DE CASO

Neste Capítulo, se faz a avaliação do Sistema Inteligente de Identificação de Perdas Comerciais SIIPERCOM em casos reais fornecidos pela Light S.A.

São apresentados as diferentes fases e respectivos módulos desenvolvidos anteriormente, aplicados em clientes inspecionados, destacando-se os módulos de Pré-processamento, Filtragem, Classificação (Fase de *Treinamento*), e Indicação de possíveis Irregularidades (Fase de *Classificação*).

5.1. Seleção e Avaliação dos Dados na Light

As empresas distribuidoras no Brasil, conforme a Resolução 456 da ANEEL [RES456, 2000] [PATRÍCIO, 2005], Art. 21, devem organizar e manter atualizados cadastros relativos aos clientes onde conste, obrigatoriamente, para cada cliente, no mínimo as seguintes informações:

1. Identificação do consumidor;
2. Número ou código de referência da unidade consumidora;
3. Endereço da unidade consumidora, incluindo o nome do município;
4. Classe e subclasse, se houver, da unidade consumidora;
5. Data de início do fornecimento;
6. Tensão nominal do fornecimento;
7. Potência disponível e, quando for o caso, a carga instalada declarada ou prevista no projeto de instalações elétricas;
8. Valores de demanda de potência e consumo de energia elétrica ativa, expressos em contrato, quando for o caso;
9. Informações relativas aos sistemas de medição de demandas de potência e de consumos de energia elétrica ativa e reativa, de fator de potência e, na falta destas medições, o critério de faturamento;
10. Históricos de leitura e de faturamento referentes aos últimos 60 (sessenta) ciclos consecutivos e completos, arquivados em meio magnético, inclusive com as alíquotas referentes a impostos incidentes sobre o faturamento realizado;
11. Código referente à tarifa aplicável; e

12. Código referente ao pagamento de juros do Empréstimo Compulsório/ELETROBRÁS.

A base de dados utilizada nesta dissertação foi fornecida pela empresa Light S.A., a qual contém estas e muitas outras informações adicionais que formam a sua base de dados gerada ao longo dos anos - constantemente atualizada – com diversas informações sobre seus clientes. Essa base é produto da implantação do sistema de faturamento Pegasus, da própria *Electricité de France* (EDF) no ano 2000.

A partir de 2003, foi dado um passo adicional, com a implantação do Sistema de Tratamento de Fraudes (STF) e do Sistema de Cálculo de Recuperação de Energia (IEN), que possibilitaram, respectivamente, a identificação de potenciais fraudadores e o cálculo da energia recuperada por meio das inspeções motivadas pelo STF.

Os resultados obtidos com esses sistemas foram significativos, o que incentivou a expansão do escopo e da funcionalidade dessas duas ferramentas, gerando o Sistema de Tratamento de Irregularidades (STI). A base de dados utilizada nesta dissertação é originária dessas 3 bases de dados:

- Base Pegasus – base de dados principal da empresa, que contém os registros e atributos gerais de todos os clientes da Light;
- Base STI – gerada a partir da base Pegasus, contém todos os registros e atributos de clientes de baixa tensão que foram identificados como suspeitos de estarem cometendo algum tipo de irregularidade;
- Base IEN – gerada a partir da base STI, contém todos os registros e atributos de clientes de baixa tensão que foram inspecionados.

As bases de dados resultantes destes procedimentos estão divididas em 5 regionais que compõem a área de concessão da Light: Litorânea, Leste, Oeste, Baixada e Interior. Entretanto, por apresentarem maiores volumes de perdas, o estudo de casos deste trabalho se concentrou nas regionais Leste e Oeste. Portanto, o Sistema Inteligente SIIPERCOM foi avaliado nestas duas regionais.

Cada uma destas bases estão subdivididas em bases de dados de Treinamento e Avaliação utilizadas no Sistema Inteligente na Fase de *Treinamento e Classificação* respectivamente.

5.2.Fase de Treinamento na Light

5.2.1.Módulo 1: Pré-Processamento da Base de Dados da Light

5.2.1.1.Limpeza da Base de Dados

Nesta etapa foi realizada a transformação de cada uma das bases de dados de *Treinamento* originais, das Regionais Leste e Oeste, em bases de dados que possuam integridade relacional. Esta transformação foi feita a partir de uma limpeza de cada uma das bases de dados, visando a detectar e remover anomalias presentes nos dados, com o objetivo de aumentar e melhorar a sua qualidade.

A limpeza dos dados inclui uma checagem da consistência das informações, a correção de possíveis erros e o preenchimento ou a exclusão de valores nulos e redundantes. Um exemplo de limpeza de dados aplicada nesse caso de estudo foi a definição de uma faixa de valores possíveis para um determinado atributo, como [0....10], sendo que valores diferentes dos definidos no intervalo seriam extraídos. Por exemplo, o atributo qualidade pagadora está codificado com os seguintes códigos: [001, 002, 003, 004, 005, 006]; mas, na prática, se encontrou os seguintes valores existentes na base: [001, 002, 003, 004, 005, 006, 084, 999]. Assim, é necessário excluir os códigos 084 e 999.

O código [999] identifica problemas de pagamento com determinado cliente, que deveria ser futuramente estudado ou sanado. O código [084] é excluído por ser um valor inválido (erro de preenchimento da base). Os dados com problemas para o caso de estudo foram removidos, pois seu percentual é muito baixo.

5.2.1.2.Seleção de atributos de entrada

A seleção de atributos começou a ser avaliada durante as diversas reuniões técnicas com os funcionários da Light que trabalham diretamente com este problema, com o objetivo de selecionar atributos que tivessem maior relevância para o problema que se deseja modelar. Esta atividade foi bastante trabalhosa e importante, devido à grande quantidade de atributos existentes nas diversas bases de dados da empresa (mais de 500 atributos).

A Figura 27 abaixo exibe a estrutura da base de dados que foi elaborada com ajuda de especialistas da Light. Esta estrutura ajudou na correta seleção

das variáveis de entrada das Redes Neurais, para as diferentes Regionais em estudo.

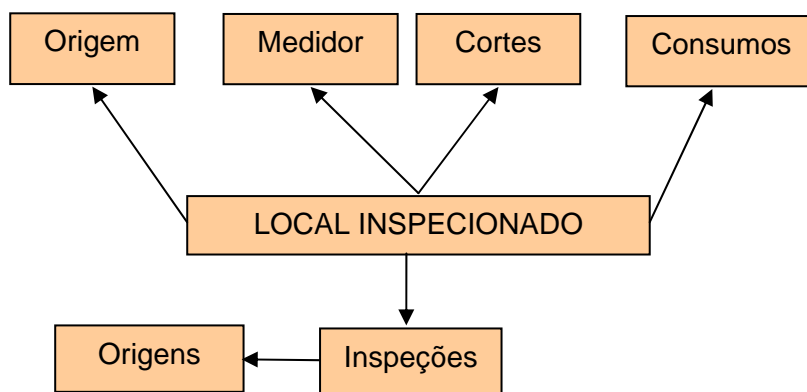


Figura 27 – Estrutura da Base de Dados Seleccionada da Light.

Os dados estão vinculados à informação *local inspecionado* (cliente suspeito). Ao local estão relacionadas todas as outras informações, tais como as inspeções sofridas dentro do histórico, as origens associadas a cada inspeção, os consumos históricos, os cortes (quando for o caso) e os medidores.

As Tabelas 18 a 24 indicam os atributos pré-seleccionados, seus formatos e a descrição dos campos relacionados ao contexto estudado da figura de acima.

Tabela 18 – Atributos seleccionados do campo Origem da Light.

Atributo	Descrição
Número do Local	Número do Local
Número do processo	Número do processo
Origem	Código da origem ¹
Data da Origem	Data da ocorrência da origem
Resultado da Inspeção	Código do resultado da Inspeção ²
Data da Normalização	Data da ocorrência da normalização
Normalização Realizada	Código da normalização ³

¹ Tipos de Origem (Códigos): ALL, API, ARG, COC, COF, COI, DCA, DCT, DIR, GCL, IEN, IPA, MAN, MAS, MET, PEG, SAT, SGL, STR, VIS. [Anexo I]

² Resultados da Inspeção (Códigos): NI, NA, FR, IR. [Anexo I]

³ Tipos de Normalização (Códigos): 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13. [Anexo I]

Tabela 19 – Atributos seleccionados do campo Origens da Light.

Atributo	Descrição
Número do processo	Número do processo
Origem	Número da origem
Data da Origem	Data da origem

Tabela 20 – Atributos selecionados do campo Corte da Light

Atributo	Descrição
Número do Local	Número do Local
Tipo de Corte	Tipo de Corte ¹
Data do Corte	Data do Corte
Tipo de Normalização	Tipo de normalização ²
Data de Re-ligação	Data da re-ligação

¹Tipos de Cortes (Códigos): CRR, CRM, CIR, CIL, CFP, CFF, CCI, CAR, C06, C02. [Anexo II]

² Tipo de Normalização (Códigos): AUC, DIC, LDM, LNC, MPD, NCD, NCS, R06, C06, RCI RCM, RDJ, RDP, REA, RIL, RIN, RME, RMS, RNC, RSM, SBM, SBR, SBS. [Anexo II]

Tabela 21 – Atributos selecionados do campo Inspeção da Light.

Atributo	Descrição
Número do Local	Número do Local
Número do processo	Número do processo
Data da Inspeção	Data da Inspeção
Indicador de Irregularidade	Código indicativo do resultado da inspeção ¹
Motivo da não inspeção	Código do motivo da não inspeção ²
Equipamento Vital	Se o cliente possui equipamento vital ou não
Atividade	Atividade do cliente
Ação	Código da Ação ³
Qualidade Pagadora	Código da qualidade pagadora ⁴
Origem	Código do tipo de origem ⁵
Tipo de Irregularidade Encontrada	Códigos com os tipos de irregularidades encontradas ⁶

¹ Indicador de Irregularidade (Códigos): A, O, I, S, N. [Anexo III]

² Motivo da Não Inspeção (Códigos): 1, 2. [Anexo III]

³ Códigos do tipo de Ação (Códigos): AP, DIR, DIS, GC, MAS, ST. [Anexo IV]

⁴ Código Qualidade Pagadora (Códigos): 1, 2, 3, 4, 5, 6. [Anexo III]

⁵ Tipos de Origem (Códigos): ALL, API, ARG, COC, COF, COI, DCA, DCT, DIR, GCL, IEN, IPA, MAN, MAS, MET, PEG, SAT, SGL, STR, VIS. [Anexo III]

⁶ Códigos dos Tipos de Irregularidade encontrados (Códigos): RL, DNM, DNR, CDD, SFT, DCX, SVT, MC, SHT, F, BD, RT, PT, SVO, CD, SVD, DCR, SVB, DCR, SVC, FI, PT, DCX, LC, DSR, BDD, RD, PI, LI, DSR, BD, DPD, DCR, PT, DP, DSR, MN, LD, SVM. [Anexo III]

Tabela 22 – Atributos selecionados do campo Local da Light.

Atributo	Descrição
Número do Local	Número do Local
ZEI	Localização geográfica da unidade consumidora
Bairro	Bairro
Município	Município
Data em que foi exteriorizado	Medidor exteriorizado se for o caso
Número do Fornecimento	Indica o cliente do local
Tipo de Normalização	Código da normalização ¹
Número do Lote	Lote
Número do Livro	Livro (informação geográfica)
Número da Instalação	Número da Instalação
Nome da Regional	Regional (informação geográfica)
Código do Transformador	Código do transformador que a unidade está relacionada
Tipo de Ação	Ação ²
Data da Inspeção	Data da realização da Inspeção
Situação	Campo relacionado ao controle do consumo em unidades que sofreram ação e não se encontrou irregularidade
Indicador de Irregularidade	Indica sim ou não
Indicador de Área de Risco	Indica sim ou não
Sub-Bairro	Sub-Bairro
Data da Normalização	Data da normalização
Código de Linha	Linha (Distribuição aérea, distribuição subterrânea ou ramal aéreo)
Código Subestação	Código da subestação
Descrição Titular	Descrição do Local
Tipo de cliente*	Residencial, comercial, rural, industrial ³

¹ Tipo de Normalização (Códigos): CRR, CIR, SBM, EMD, CCI, RSM, RCM, CFF, RME, RMS, SBS, NCD e MPD. [Anexo IV]

² Tipo de Ação (Códigos): CL, AC, LC, ST, GC, AP, DI e MA. [Anexo IV]

³ Tipo de Clientes (Códigos): 100, 101, 102, 140, 141, 142, 180, 181, 182, 200, 240, 300, 340, 400, 500, 700, 800 e 900. [Anexo IV]

Tabela 23 – Atributos selecionados do campo Medidor da Light.

Atributo	Descrição
Número do Local	Número do Local
Número do Medidor	Número do Medidor
Tipo de Medidor	Tipo do Medidor
Constante Medidor	Especifica medição indireta e direta ¹
Data da Instalação	Data da Instalação do medidor
Número TC1	Número Transformador de Corrente C1
Relação TC1	Relação Transformador de Corrente 1
Modelo TC1	Modelo Transformador de Corrente 1
Número TC2	Número Transformador de Corrente 2
Relação TC2	Relação Transformador de Corrente 2
Modelo TC2	Modelo Transformador de Corrente 2
Número TC3	Número Transformador de Corrente 3
Relação TC3	Relação Transformador de Corrente 3
Modelo TC3	Modelo Transformador de Corrente 3
Tipo Medição [Direta/Indireta]	Tipo Medição ²
Mono Bi Tri	Número de Fases da Instalação ³

¹ Tipo de Constante Medidor (Códigos): 1, 3, 10, 12, 15, 16, 20, 25, 30, 40, 60, 80, 100, 120, 200 400. [Anexo V]

² Tipos de Medição (Códigos): 0, 1, 2, 9. [Anexo V]

³ Mono Bi Tri (Códigos): 1, 2, 3. [Anexo V]

Tabela 24 – Atributos selecionados do campo Consumo da Light.

Atributo	Descrição
Número do Local	Número do Local
Mês de Referência	Mês utilizado como referencia (mês anterior ao mês da realização da inspeção)
Consumo no mês	Consumo de energia faturado do cliente no mês de referência
Consumo no ano anterior	Consumo no mesmo mês de referencia, porem no ano anterior ao de referência
Consumo no ano base	Consumo no mesmo mês de referencia, porem no ano base [2001]
Quantidade em Fraude	Quantidade de fraudes já registradas
Cód. da Irregularidade consumo	Código de irregularidade de consumo
Cód. da Irregularidade Leitura	Código de irregularidade de leitura

Também foram identificados outros atributos relevantes que permitem compreender o comportamento dos clientes consumidores em diferentes estações do ano (verão e inverno), e que são importantes para nosso estudo:

- Temperatura média, máxima e mínima mensal na região metropolitana do Rio de Janeiro;
- Curva de Carga própria da Light mensal;

Alguns outros atributos foram calculados a partir dos valores existentes nas bases de dados. Esses atributos foram criados com base nas metodologias (métricas heurísticas) utilizadas pela Light, e são inspirados no atual procedimento de detecção de irregularidade. É importante mencionar que a Light usa essas métricas para avaliá-las com relação a uma referência, por exemplo, todos os clientes que tiveram variação percentual maior de consumo de 30% entre os consumos dos 3 últimos meses eram separados para futura investigação. Esses novos atributos calculados são apresentados abaixo:

- **Média 3 últimos meses**

É o valor médio dos 3 últimos faturamentos (valor 1, valor 2 e valor 3).

$$ValorMedio3 = \frac{Val1 + Val2 + Val3}{3} \quad \text{eq. (25)}$$

Onde:

Val1: Consumo do último mês (kWh);

Val2: Consumo do penúltimo mês (kWh);

Val3: Consumo do antepenúltimo mês (kWh).

- **Média 6 últimos meses**

É o valor médio dos 6 últimos faturamentos (valor 1 ao valor 6).

$$ValorMedio6 = \frac{Val1 + Val2 + Val3 + Val4 + Val5 + Val6}{6} \quad \text{eq. (26)}$$

Onde:

Val1 ... Val 6: Consumo médio dos últimos 6 meses (kWh).

- **Média dos meses 1 ao 12**

É o valor médio dos 12 últimos faturamentos (valor 1 ao valor 12), contando a partir do mês de referência.

$$ValorMedio1_12 = \frac{\sum_{i=1}^{i=12} Val_i}{12} \quad \text{eq. (27)}$$

Onde:

$$\sum_{i=1}^{i=12} Val_i : \text{Consumo dos últimos 12 meses (kWh).}$$

Exemplo: mês de referência: julho 2007 → meses para cálculo da média: julho 2006 a junho 2007

- **Média dos meses 13 ao 24**

É o valor médio dos faturamentos entre o 13º e o 24º meses (valor 13 ao valor 24), contando a partir do mês do ano anterior ao de referência.

$$ValorMedio13_24 = \frac{\sum_{i=13}^{i=24} Val_i}{12} \quad \text{eq. (28)}$$

Onde:

$$\sum_{i=13}^{i=24} Val_i : \text{Consumo dos últimos 12 meses anteriores aos últimos (kWh).}$$

Exemplo: mês de referência julho 2007 → meses para cálculo da média: julho 2005 a junho 2006.

- **Indicador 1_2**

É o indicador que mostra a queda percentual do penúltimo para o último faturamento.

$$Indicador1_2 = \frac{Val2 - Val1}{Val2} \quad \text{eq. (29)}$$

Onde:

Val1: Consumo do último mês (kWh);

Val2: Consumo do penúltimo mês (kWh).

- **Indicador 2_3**

É o indicador que mostra a queda percentual do antepenúltimo para o penúltimo faturamento.

$$Indicador2_3 = \frac{Val3 - Val2}{Val3} \quad \text{eq. (30)}$$

Onde:

Val2: Consumo do penúltimo mês (kWh);

Val3: Consumo do antepenúltimo mês (kWh).

- **Indicador 1_3**

É o indicador que mostra queda percentual do antepenúltimo para o último faturamento.

$$Indicador1_3 = \frac{Val3 - Val1}{Val3} \quad \text{eq. (31)}$$

Onde:

Val1: Consumo do último mês (kWh);

Val3: Consumo do antepenúltimo mês (kWh).

- **Indicador Anual**

É o indicador que mostra queda percentual da média do faturamento dos 12 meses defasados de um ano do mês de referência para a média do faturamento dos 12 meses exatamente anteriores ao mês de referência.

$$IndicadorAnual = \frac{ValorMedio13_24 - ValorMedio1_12}{ValorMedio13_24} \quad \text{eq. (32)}$$

$$IndicadorAnual = \frac{\frac{\sum_{i=13}^{i=24} Val_i}{12} - \frac{\sum_{i=1}^{i=12} Val_i}{12}}{\frac{\sum_{i=13}^{i=24} Val_i}{12}} \quad \text{eq. (33)}$$

Onde:

ValorMedio1_12: Valor médio dos 12 últimos faturamentos (kWh);

ValorMedio13_24: É o valor médio dos faturamentos entre o 13º e o 24º meses (kWh).

Exemplo: mês de referência: Agosto de 2006.

ValorMedio1_12: Agosto de 2005 a Julho de 2006.

ValorMedio13_24: Agosto de 2004 a Julho de 2005.

- **Indicador Ajuste**

É o indicador que mostra queda percentual do último faturamento para o faturamento do mesmo mês do ano anterior.

$$IndicadorAjuste = \frac{Valor_12 - Val1}{Valor_12} \quad \text{eq. (34)}$$

Onde:

Val1: Consumo do último mês (kWh);

Valor_12: Consumo do mesmo mês do ano anterior (kWh).

Exemplo:

Val1: Julho de 2006.

Valor_12: Julho de 2005.

- **Indicador de Tendência**

É o indicador que mostra queda percentual do último faturamento para o faturamento do mês no ano base (2001).

$$IndicadorTendencia = \frac{ValorBase - Val1}{ValorBase} \quad \text{eq. (35)}$$

Onde:

Val1: Consumo do último mês (kWh);

ValorBase: Consumo do mês no ano base (2001).

No ano de 2001 aconteceu a 'Crise de Energia' no Brasil, a qual influenciou sobremaneira o comportamento do consumo nesse ano até a atualidade.

Exemplo:

Val1: Julho de 2006.

ValorBase: Julho de 2001.

A seleção de atributos foi realizada a partir de entrevistas não estruturadas as quais permitiu ao(s) especialista(s) da Light a discussão do problema a ser resolvido de modo natural. O objetivo dessas entrevistas foi adquirir uma visão geral do domínio em questão com a finalidade de determinar quais atributos são relevantes para nosso trabalho.

Também foi possível obter uma maior compreensão dos conceitos mais importantes a respeito do domínio em estudo, auxiliar na identificação de conceitos e objetivos; e conhecer que estratégias o(s) especialista(s) da Light utilizam para resolver o problema das Perdas Comerciais.

Após a seleção desses atributos, a partir das entrevistas com os especialistas nas bases da Light, passou-se a uma análise mais minuciosa de cada atributo, interpretando seus diversos códigos. Além disso, foram realizados diversos testes estatísticos, verificando-se a média, a variância e a correlação entre os atributos, para avaliar a consistência destes dados, permitindo maior conhecimento contida em cada atributo [DASH E LIU, 1997]. Estas estatísticas foram úteis para identificar e remover registros que têm baixa relevância em relação ao objetivo final desta dissertação. Após análise dos atributos, detectou-se que alguns apresentavam inconsistências, e portanto não foram considerados.

Fez-se, também, uma análise de frequência para todas as variáveis qualitativas (categóricas); esse procedimento é importante pois permite definir como deverão ser agrupados e codificados alguns atributos que apresentem um número grande de valores diferentes.

5.2.1.3.Codificação da Base de Dados

A partir dos atributos encontrados na etapa anterior, foram construídas as bases de dados codificadas, necessárias para o correto treinamento dos módulos de filtragem e de classificação do sistema inteligente. Para isso é necessário identificar quais atributos da entrada precisam ser codificados. Na Tabela 25, a seguir, apresenta-se a codificação dos atributos categóricos selecionados na etapa anterior.

Tabela 25 – Codificação de Atributos de Entrada.

Atributo	Descrição	Codificação	Tipo de Atributo
Equipamento Vital	Indica se há equipamento vital na unidade consumidora	'0', se não existe equipamento vital no local (valor de equip. vital = 'N') '1', se existe equipamento vital no local (valor de equip. vital = 'S')	Numérico Discreto
Origem	Origem principal da unidade consumidora até a inspeção [Anexo VI]	'00', para o tipo 'denúncia' (valores de origem = [ALL; AUD; COC; COM; SGL; VIR; MAN 01; MAN 02; MAN M01 até MAN M08; MAN DIR; MAN P02; MAN SAT]) '01', para o tipo 'agrupamento' (valores de origem = [API; IPA; GCL; MAS; MET; STR; VIS; MAN M09; MAN MAS]) '10', para o tipo 'inteligência' (valores de origem = [ARG; COF; COI; DCA; DCT; DIR; IEN; MAN M11]) '11', para o tipo 'sistema' (valores de origem = [PEG; CCS])	Numérico Discreto
Medidor	Identificador de modelo/fabricante do medidor [Anexo VII]	'000', se o medidor estiver obsoleto e for do tipo 'MMD' '001', se o medidor estiver obsoleto e for do tipo 'MMI' '010', se o medidor estiver obsoleto e for do tipo 'MED' '011', se o medidor não estiver obsoleto e for do tipo 'MMD' '100', se o medidor não estiver obsoleto e for do tipo 'MMI' '101', se o medidor não estiver obsoleto e for do tipo 'MED'	Numérico Discreto
Monobitri	Indica se a instalação é mono, bi ou trifásica	'00', se a instalação for monofásica (valor de monobitri = '1') '01', se a instalação for bifásica (valor de monobitri = '2'). 10', se a instalação for trifásica (valor de monobitri = '3')	Numérico Discreto
Qualidade da Unidade Pagadora	Quantificação do atraso no pagamento das contas [Anexo III]	'000', se a qualidade pagadora for do tipo '1' '001', se a qualidade pagadora for do tipo '2' '010', se a qualidade pagadora for do tipo '3' '011', se a qualidade pagadora for do tipo '4' '100', se a qualidade pagadora for do tipo '5' '101', se a qualidade pagadora for do tipo '6'	Numérico Discreto
Tipo Medição	Indica se a medição feita no local é direta ou indireta.	'0', se a medição é direta '1', se a medição é indireta	Numérico Discreto
Código de Irregularidade de Consumo (C.I.C.)	É gerado a partir de verificação dos consumos passados do cliente.	'0', se a unidade consumidora não possuir nenhum código de irregularidade de consumo '1', se a unidade consumidora possuir algum código de irregularidade de consumo	Numérico Discreto
Código de Irregularidade de Leitura (C.I.L.)	É preenchido pelo leiturista no momento da leitura, ou seja, é uma indicação baseada em avaliação informal e superficial.	'0', se a unidade consumidora não possuir nenhum código de irregularidade de leitura '1', se a unidade consumidora possuir algum código de irregularidade de consumo	Numérico Discreto
Indicador de Irregularidade	Indica se o inspetor encontrou uma fraude, irregularidade ou normalidade. Atributo utilizado como a saída (alvo) para o treinamento das redes neurais do comitê.	'0', se não foi encontrada nenhuma irregularidade ou fraude na UC (valor de indicador fraude = 'N') '1', se foi encontrada alguma irregularidade técnica na UC (valor de indicador fraude = 'I') '2', se foi encontrada alguma irregularidade não técnica (fraude) na UC (valor de indicador fraude = 'S')	Numérico Discreto

5.2.1.4. Normalização da Base de Dados

O passo seguinte é a preparação e adequação dos dados ao formato requerido pela rede neural, ou seja, a formatação dos dados. Neste processo cria-se uma escala, estabelecendo um novo intervalo válido dentro do qual todos os dados são distribuídos.

Para isso são selecionadas as variáveis numéricas contínuas, para sua normalização, criando um intervalo de amplitude similar [0 e 1]. No caso desta dissertação são aplicados 2 tipos de normalizações aos seguintes atributos, os quais são descritos a continuação na Tabela 26.

Tabela 26 – Identificação de Atributos de Entrada por tipo de normalização.

Atributo	Tipo de Normalização
Consumo no mês	Normalização Horizontal pela faixa de variação
Consumo no ano anterior	Normalização Horizontal pela faixa de variação
Consumo no ano base	Normalização Horizontal pela faixa de variação
Média 3 meses	Normalização Horizontal pela faixa de variação
Média 6 meses	Normalização Horizontal pela faixa de variação
Média dos meses 1 ao 12	Normalização Horizontal pela faixa de variação
Média dos meses 13 ao 24	Normalização Horizontal pela faixa de variação
Quantidade de Cortes	Normalização Vertical pela variável de entrada
Temperatura Mínima	Normalização Vertical pela variável de entrada
Temperatura Máxima	Normalização Vertical pela variável de entrada
Carga	Normalização Vertical pela variável de entrada

Na tabela 27 são apresentados os 26 atributos finais, os quais representam 31 entradas e 1 saída para a RN (devido à codificação de alguns atributos em 2 ou 3 bits), assim como sua origem, que foram utilizados para o Sistema Inteligente SIIPERCOM.

Tabela 27 – Entradas da Rede Neural da Light.

Nro	Atributo	Origem
1.	Equipamento Vital	Extraído
3.	Origem ¹	Extraído
6.	Medidor ²	Extraído
8.	Monobitri ¹	Extraído
11.	Qualidade Pagadora ²	Extraído
12.	Tipo Medição	Extraído
13.	Código de Irregularidade de Consumo	Extraído
14.	Código de Irregularidade de Leitura	Extraído
15.	Consumo no mês	Extraído
16.	Consumo no ano anterior	Extraído
17.	Consumo no ano base	Extraído
18.	Média 3 meses	Calculado
19.	Média 6 meses	Calculado
20.	Média dos meses 1 ao 12	Calculado
21.	Média dos meses 13 ao 24	Calculado
22.	Indicador 1_2	Calculado
23.	Indicador 2_3	Calculado
24.	Indicador 1_3	Calculado
25.	Indicador Anual	Calculado
26.	Indicador Ajuste	Calculado
27.	Indicador Tendência	Calculado
28.	Quantidade de Cortes	Extraído
29.	Temperatura Mínima	Extraído
30.	Temperatura Máxima	Extraído
31.	Carga	Extraído
32.	Indicador de Irregularidade	Extraído

¹ Codificação Numérico Discreto em dois bits

² Codificação Numérico Discreto em três bits

Um exemplo da base de dados utilizada pode ser visualizado na tabela 28, ela contém os diferentes atributos seleccionados pré-processados para a utilização na Rede Neural.

Tabela 28 – Exemplo de atributos finais utilizados no SIIPERCOM.

0	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0,103	0,206	0,103	0,34	0,485	0,289	0,196	0,231	0,071	0,231	0	0,091	0	0	0,777	0,674	0,57	0
0	1	1	0	1	1	1	0	0	1	0	0	1	0	0	0	0,811	0,822	0	0	0,235	0,822	0	0	0	0,714	1	0	0	0,444	0,459	0,429	0
0	0	0	0	1	1	0	1	0	0	1	0	0	0	0	0,693	0,561	0,601	0,766	0,473	0,538	0,484	0,134	0	0,134	0	0	0	0,777	0,674	0,57	0	
0	0	1	1	0	0	1	0	0	1	0	1	0	1	0,881	0,034	0,034	0,655	0,528	0,242	0,034	0	0	0	0	0	0	0	0,726	0,648	0,773	0	
0	0	0	0	1	1	0	0	0	0	0	0	1	0	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0	0	0	0	0	0	0,365	0,445	0,34	1	
0	0	1	0	1	1	1	0	0	0	1	0	0	1	0,004	0,761	0,079	0,429	0,619	0,69	0,768	0,937	0,118	0,937	0,097	0,95	0,348	0	0,444	0,459	0,429	0	
0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0,225	0,146	0	0	0,183	0,219	0	0	0	0,165	1	0	0	0,365	0,445	0,34	0	
0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0,055	0,561	0,041	0,151	0,465	0,534	0,506	0	0,434	0	0	0,544	0	0,365	0,445	0,34	0	
0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0,785	0,6	0	0	0,354	0,701	0	0	0	0,495	1	0	0,777	0,674	0,57	1	

5.2.2. Módulo 2: Filtragem de dados na Light

A empresa Light possui cerca de 452 mil clientes inspecionados em baixa tensão com perfis de consumo de energia diferentes, distribuídos nas 2 regionais em estudo (Leste e Oeste).

O Módulo 2 ‘*Filtragem de Dados*’ permite a filtragem, mediante a aplicação de redes neurais, entre os clientes verdadeiramente *Fraudadores/Irregulares* e os clientes verdadeiramente *Normais*. Isto é necessário devido à baixa qualidade das bases de dados originais, com muitos dados conflitantes em relação ao perfil do consumidor indicado na base de dados da Light.

Na tabela 29 é apresentada a quantidade de Unidades Consumidoras, assim como os clientes inspecionados⁷, nas 2 regionais em estudo (Leste e Oeste). Os dados relativos à quantidade de clientes inspecionados foram coletados no período de Janeiro de 2001 até agosto do 2006.

Tabela 29 – Quantidade de Usuários Inspeccionados Validados por Regional.

Regional	Unidades Consumidoras	Quantidade de Clientes Inspeccionados
Leste	888508	254286
Oeste	766760	197249
Total	1655268	451535

As Regionais estão divididas em Zeis, onde cada Zei representa uma área de assentamento habitacional de população com proximidades geográficas, a qual forma parte da área de concessão da Light, no Rio de Janeiro. Nas Tabelas 30 e 31 exibe-se a lista de Zeis distribuídas por suas respectivas regionais em estudo, com suas respectivas localidades e número de clientes inspecionados.

⁷ Não serão considerados os clientes que estão em áreas de risco operacional, nas quais não se consegue trabalhar, cortar e atender a chamados dos clientes.

Tabela 30 Relação de Zeis e Localidades na Regional Leste.

Regional	Nome da Localidade	Zeí	Quantidade de Clientes Inspeccionados
Leste	Méier	535	23491
Leste	Pilares	538	37824
Leste	Madureira	540	45495
Leste	Guadalupe	544	47110
Leste	Irajá	545	35979
Leste	Olaria	550	8969
Leste	Maré	551	735
Leste	Ramos	552	8141
Leste	Penha	555	27600
Leste	Ilha do Governador	565	18781
Leste	Paquetá	566	161
Total			254286

Tabela 31 – Relação de Zeis e Localidades na Regional Oeste.

Regional	Nome da Localidade	Zeí	Quantidade de Clientes Inspeccionados
Oeste	Jacarepaguá	539	49598
Oeste	Realengo	580	0
Oeste	Bangu	585	54691
Oeste	Campo Grande	590	65849
Oeste	Santa Cruz	593	27111
Total			197249

Devido à grande quantidade de clientes inspeccionados nessas categorias, decidiu-se aplicar, num primeiro momento, um processo de clusterização para determinar grupos (*clusters*) de clientes que apresentassem perfis de consumo similares. Para realizar esta clusterização foi empregado o algoritmo *Fuzzy C-Means* [PAL, 1995]. Entretanto, em função da baixa qualidade das bases de dados originais, não foi possível identificar agrupamentos que tivessem características comuns.

Deste modo, optou-se por utilizar a própria separação dos clientes por regiões, identificados pelo seu código de ZEIS, conforme descrito a seguir.

5.2.2.1. Construção da base de dados para sua Filtragem na Light

Conforme mencionado, devido ao grande número de clientes de baixa tensão, a base foi dividida em Residenciais e Não-Residenciais, com o objetivo de criar sistemas classificadores específicos para cada uma dessas categorias. Entretanto, o número de clientes em cada categoria ainda é muito grande, com o agravante de terem perfis de consumo distintos, em função da classe social, etc. No intuito de especializar mais ainda o modelo inteligente, as bases foram subdivididas em função dos códigos das ZEIS.

A Tabela 32 apresenta o número de clientes Residenciais que compõem a Regional Leste, classificados segundo os resultados da inspeção que representa 3 classes: clientes Fraudadores, clientes com Irregularidade, e clientes Normais (sem irregularidades), por código de ZEI. É importante destacar que, devido à pouca quantidade de clientes inspecionados em algumas ZEIS, estas foram agrupadas em um único classificador, conforme exemplificado no caso das zeis 535, 505, 551, 552, 555, 565 e 566.

Tabela 32 – Número de Clientes Residenciais da Regional Leste classificados segundo os resultados da inspeção.

Base Treinamento	Fraude	Irregular	Normal
538	3972	3847	20708
540	4630	3182	28414
544	6698	7330	24946
545	4139	3476	21392
535/550/551/552/555 /565/566	8659	6950	45171
Total	28098	24785	140631

Na Tabela 33 é apresentado o número de clientes Não Residenciais (industrial e comercial) que compõem a Regional Leste, classificados também segundo os resultados da inspeção que representa 3 classes: clientes Fraudadores, clientes com Irregularidade, e clientes Normais (sem irregularidades). Neste caso, devido à pouca quantidade de clientes inspecionados, todas as zeis que compõem essa regional foram agrupadas em um único classificador.

Tabela 33 – Número de Clientes Não Residenciais da Regional Leste classificados segundo os resultados da inspeção.

Base Treinamento	Fraude	Irregular	Normal
538/540/544/545/535 /550/551/552/555/565/566	2471	2902	28988
Total	2471	2902	28988

As Tabelas 34 e 35 apresentam o número de clientes Residenciais e Não-Residenciais, respectivamente, que compõem a Regional Oeste, classificados novamente segundo os resultados da inspeção que representa 3 classes: clientes Fraudadores, clientes com Irregularidade, e a de clientes Normais (sem irregularidades).

Tabela 34 - Número de Clientes Residenciais da Regional Oeste classificados segundo os resultados da inspeção.

Base Treinamento	Fraude	Irregular	Normal
539	4544	6505	25305
580	0	0	0
585	7458	7796	24761
590	10205	7548	32596
593	5397	4116	10347
Total	27604	25965	93009

Tabela 35 – Número de Clientes Não Residenciais da Regional Oeste classificados segundo os resultados da inspeção.

Base Treinamento	Fraude	Irregular	Normal
539 / 580 / 585 / 590 / 593	1928	2705	18159
Total	1928	2705	18159

Conforme apresentado nas tabelas anteriores, a quantidade de clientes normais é bastante superior à das outras classes (Fraude e Irregular). Para um melhor treinamento das redes neurais, optou-se por criar uma base de Treinamento/Validação equilibrada no que se refere ao número de padrões em cada classe a ser distinguida pelo modelo. Deste modo, a quantidade de clientes *Normais* selecionados da base existente foi igual à soma de clientes

Fraudadores e Irregulares (estas 2 bases estão grupadas, porque apresentam características semelhantes no consumo).

Esta base nova, homogeneizada e embaralhada, foi então dividida em 5 bases para a '*Filtragem de dados*', cada uma delas sub-dividida em 2 partes: Treinamento (75%) e Validação (25%). Os dados relativos ao Treinamento/Validação foram coletados de dois períodos distintos: de Janeiro de 2001 a dezembro de 2005 e os meses de março até junho do 2006. A seleção desses períodos teve como objetivo avaliar o SIIPERCOM em períodos distintos (verão e inverno de 2006).

Nas Tabelas 36, 37, 38 e 39 são apresentados os números de registros contidos nas bases de Treinamento/Validação que serão utilizadas no Sistema Inteligente para as Regionais Leste e Oeste, Residencial e Não Residencial respectivamente.

Tabela 36 – Divisão de amostras de Clientes Residenciais da Regional Leste.

			538	540	544	545	535/550/551/552/ 555/565/566
Base Treinamento	Base1	Treinam1	2432	2430	4368	2370	4856
		Validação1	694	694	1244	676	1388
	Base2	Treinam2	2432	2430	4368	2370	4856
		Validação2	694	694	1244	676	1388
	Base3	Treinam3	2432	2430	4368	2370	4856
		Validação3	694	694	1244	676	1388
	Base4	Treinam4	2432	2430	4368	2370	4856
		Validação4	694	694	1244	676	1388
	Base5	Treinam5	2438	2432	4364	2368	4856
		Validação5	696	696	1244	678	1386

Tabela 37 – Divisão de amostras de Clientes Não Residenciais da Regional Leste.

			538 / 540 / 544 / 545 / 535 / 550 / 551 / 552 / 555 / 565 / 566
Base Treinamento	Base1	Treinam1	1670
		Validação1	478
	Base2	Treinam2	1670
		Validação2	478
	Base3	Treinam3	1670
		Validação3	478
	Base4	Treinam4	1670
		Validação4	478
	Base5	Treinam5	1676
		Validação5	478

Tabela 38 – Divisão de amostras de Clientes Residenciais da Regional Oeste.

			539	585	590	593
Base Treinamento	Base1	Treinam1	3438	4748	5526	2958
		Validação	982	1354	1576	846
	Base1	Treinam2	3438	4748	5526	2958
		Validação	982	1354	1576	846
	Base1	Treinam3	3438	4748	5526	2958
		Validação	982	1354	1576	846
	Base1	Treinam4	3438	4748	5526	2958
		Validação	982	1354	1576	846
	Base1	Treinam5	3436	4748	5526	2958
		Validação	982	1354	1576	846

Tabela 39 – Divisão de amostras de Clientes Não Residenciais da Regional Oeste.

			539 / 585 / 590 / 593
Base Treinamento	Base1	Treinam1	1442
		Validação	412
	Base1	Treinam2	1442
		Validação	412
	Base1	Treinam3	1442
		Validação	412
	Base1	Treinam4	1442
		Validação	412
	Base1	Treinam5	1440
		Validação	410

A topologia de cada Rede Neural que compõe o comitê do Módulo de 'Filtragem de dados' é apresentada na Figura 28, contendo as 31 entradas e 1 saída da RN (26 atributos), uma camada oculta de neurônios e 1 neurônio na camada de saída.

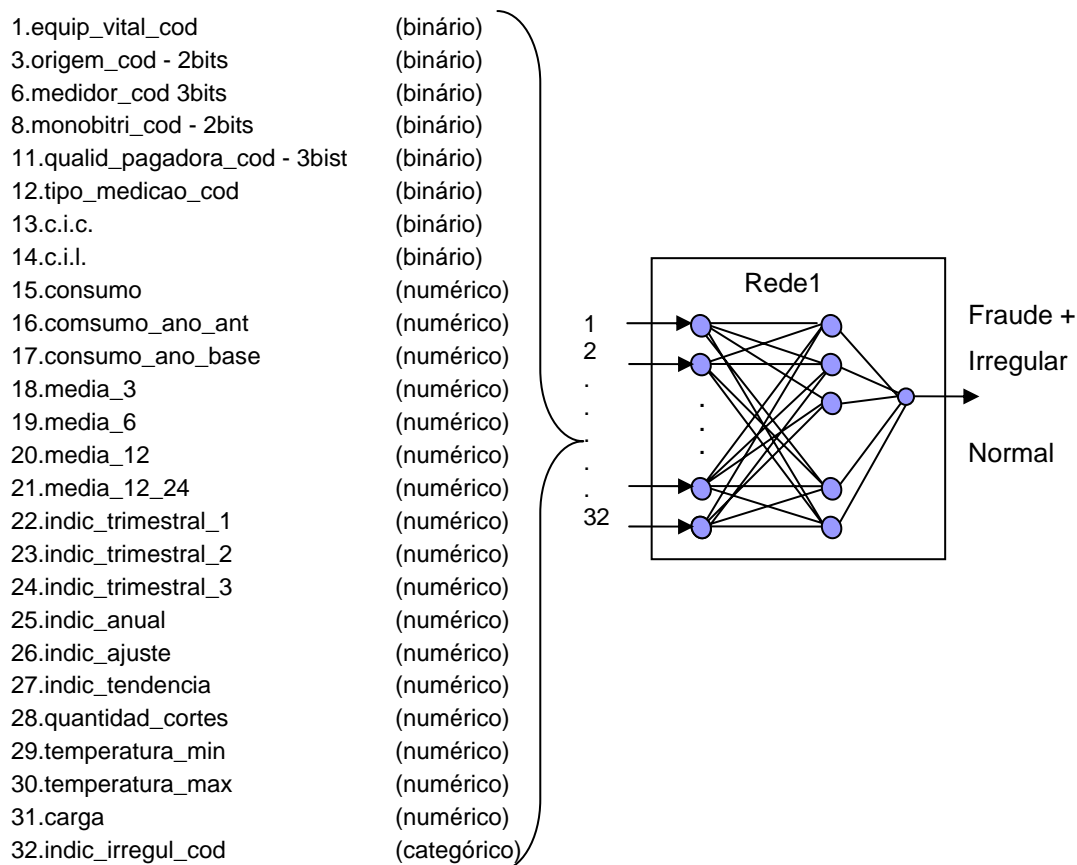


Figura 28 - Topologia de Rede Neural MLP com 1 saída (Módulo de Filtragem de dados) – Caso de Estudo Light.

Para o treinamento das redes foi escolhido o software *Matlab*. Este software oferece pacotes prontos para utilização das RNA's. A topologia empregada foi a MLP (*Multilayer Perceptron*) *feedforward* com uma camada de entrada, consistindo em 32 nós, uma camada oculta, e uma camada de saída com 1 neurônio. O treinamento foi feito variando-se o número de neurônios na camada oculta, de 4 a 32 neurônios, fazendo-se incremento de 4 em 4.

Para este caso de estudo, a melhor avaliação ficou com 8 neurônios na camada oculta. Buscando a arquitetura mais exata, também se avaliou em torno de 8 neurônios, ou seja 6 e 10. Essa estratégia visava a reduzir a busca pela melhor estrutura, já que o número de redes testadas seria bastante grande.

Devido ao neurônio único na camada de saída, os clientes suspeitos de apresentar algum tipo de *Fraude/Irregularidade* são representados pelo código 1, e os clientes *Normais* são representados pelo código 0. Neste estudo, foi empregada na camada oculta e na camada de saída a função de ativação não-linear do tipo sigmoideal (*logsig*), que assume um intervalo contínuo de valores entre 0 e 1.

Esta função de ativação sigmoideal foi produto de múltiplas avaliações preliminares testando outras combinações no desenvolvimento deste caso de estudo.

Para o treinamento, foi utilizado o algoritmo de '*Levenberg Marquardt*' (LM), por meio da função '*trainlm*' pertencente ao pacote de RNA do *Matlab*. Durante a etapa de treinamento, o erro do conjunto de validação determina o melhor ponto de parada (*early stopping*), conforme descrito no capítulo 3.

Com base no Treinamento anterior, foram selecionadas 5 novas bases de dados, uma de cada membro do comitê de filtragem, as quais serão utilizadas no seguinte módulo de '*Classificação*'.

A seleção do melhor conjunto de clientes foi realizada em duas etapas: análise do *Limiar de Filtragem*, o qual permite que se binarize a saída de ativação de cada rede em somente 2 valores [1 0]; e análise do *Comitê de filtragem* (resultado da votação do comitê), o qual determina o número de votos mínimo para classificar cada cliente em uma das duas categorias possíveis (*Irregular/Fraudador* e *Normal*).

Para o *Limiar de Filtragem* foram testados dois procedimentos: limiar único e limiar duplo. No caso do limiar único, usa-se apenas um valor, acima do qual determina-se o cliente como irregular/fraudador e abaixo do qual se classifica como cliente normal. Já no caso de limiar duplo, dois valores distintos são utilizados para determinar o valor '1' e o valor '0'. Valores de saída da rede no intervalo entre esses dois valores distintos são considerados como clientes não classificados e, obviamente, não são selecionados para o treinamento do módulo de classificação. Os critérios utilizados neste Limiar de Filtragem são descritos na tabela 40 a seguir:

Tabela 40 – Critérios utilizados no Limiar de Filtragem.

Limiar de Filtragem			
Limiar Simples		Limiar duplo	
Fraudador/ Irregular	Normal	Fraudador/ Irregular	Normal
$\geq 0,55$	$< 0,55$	$\geq 0,75$	$< 0,45$
$\geq 0,65$	$< 0,65$	$\geq 0,75$	$< 0,55$
$\geq 0,75$	$< 0,75$	$\geq 0,75$	$< 0,65$
$\geq 0,85$	$< 0,85$	$\geq 0,80$	$< 0,40$

Para a etapa de análise do *Comitê de filtragem* determinou-se outro *Limiar do Comitê de filtragem*: para selecionar clientes fraudadores/irregulares, uma votação mínima de 4 redes do comitê; e para seleção de clientes normais, optou-se pela unanimidade das redes, isto é, as 5 redes do comitê devem especificar o cliente como normal para que o mesmo seja selecionado para o treinamento do Módulo de Classificação.

Após a realização de testes com os diferentes limiares apresentados na Tabela 40, avaliou-se a *Matriz de Confusão* resultante de cada caso, selecionando-se o limiar mais adequado para cada Zei das duas categorias *Residencial* e *Não-Residencial*.

As Tabelas 41 a 51 apresentam as Matrizes de Confusão do Comitê de Filtragem, estas tabelas apresentam como melhor critério de voto do comitê (*Limiares do Comitê de filtragem*) para selecionar os verdadeiros clientes Fraudadores/Irregulares e Normais Além disso cada Zei terá seu respectivo *Limiar de Filtragem*, assim como o número de clientes selecionados (filtrados) pelo Módulo de Filtragem.

Tabela 41 – Comitê de Filtragem vs Clientes Inspeccionados – Zei 538 Residencial.

Zei 538 Residencial Leste			Limiar de Filtragem	Clientes Inspeccionados da Light (Target)	
				Fraud/Irreg (1)	Normal (0)
REDE NOVA1	Comitê de Filtragem de Redes Neurais	Fraud/Irreg (1)	$\geq 0,75$	180	37
		Normal (0)	$< 0,65$	473	961
REDE NOVA2		Fraud/Irreg (1)	$\geq 0,75$	191	41
		Normal (0)	$< 0,65$	430	954
REDE NOVA3		Fraud/Irreg (1)	$\geq 0,75$	182	41
		Normal (0)	$< 0,65$	442	1004
REDE NOVA4		Fraud/Irreg (1)	$\geq 0,75$	172	38
		Normal (0)	$< 0,65$	452	987
REDE NOVA5		Fraud/Irreg (1)	$\geq 0,75$	202	31
		Normal (0)	$< 0,65$	409	1006
Total				2023	4914

Tabela 42 – Comitê de Filtragem vs Clientes Inspeccionados – Zei 540 Residencial.

Zei 540 Residencial Leste			Limiar de Filtragem	Clientes Inspeccionados da Light (Target)	
				Fraud/Irreg (1)	Normal (0)
REDE NOVA1	Comitê de Filtragem de Redes Neurais	Fraud/Irreg (1)	$\geq 0,8$	159	25
		Normal (0)	$< 0,4$	86	286
REDE NOVA2		Fraud/Irreg (1)	$\geq 0,8$	165	28
		Normal (0)	$< 0,4$	89	296
REDE NOVA3		Fraud/Irreg (1)	$\geq 0,8$	168	26
		Normal (0)	$< 0,4$	88	299
REDE NOVA4		Fraud/Irreg (1)	$\geq 0,8$	169	19
		Normal (0)	$< 0,4$	95	304
REDE NOVA5		Fraud/Irreg (1)	$\geq 0,8$	160	27
		Normal (0)	$< 0,4$	80	294
Total				821	1479

Tabela 43 – Comitê de Filtragem vs Clientes Inspeccionados – Zei 544 Residencial.

Zei 544 Residencial Leste			Limiar de Filtragem	Clientes Inspeccionados da Light (Target)	
				Fraud/Irreg (1)	Normal (0)
REDE NOVA1	Comitê de Filtragem de Redes Neurais	Fraud/Irreg (1)	$\geq 0,75$	58	10
		Normal (0)	$< 0,55$	469	1110
REDE NOVA2		Fraud/Irreg (1)	$\geq 0,75$	75	16
		Normal (0)	$< 0,55$	499	1161
REDE NOVA3		Fraud/Irreg (1)	$\geq 0,75$	59	10
		Normal (0)	$< 0,55$	485	1131
REDE NOVA4		Fraud/Irreg (1)	$\geq 0,75$	57	10
		Normal (0)	$< 0,55$	501	1109
REDE NOVA5		Fraud/Irreg (1)	$\geq 0,75$	77	7
		Normal (0)	$< 0,55$	466	1140
Total				326	5651

Tabela 44 – Comitê de Filtragem vs Clientes Inspeccionados – Zei 545 Residencial.

Zei 545 Residencial Leste			Limiar de Filtragem	Clientes Inspeccionados da Light (Target)	
				Fraud/Irreg (1)	Normal (0)
REDE NOVA1	Comitê de Filtragem de Redes Neurais	Fraud/Irreg (1)	$\geq 0,75$	179	39
		Normal (0)	$< 0,55$	312	750
REDE NOVA2		Fraud/Irreg (1)	$\geq 0,75$	160	25
		Normal (0)	$< 0,55$	311	811
REDE NOVA3		Fraud/Irreg (1)	$\geq 0,75$	163	41
		Normal (0)	$< 0,55$	292	824
REDE NOVA4		Fraud/Irreg (1)	$\geq 0,75$	128	29
		Normal (0)	$< 0,55$	326	801
REDE NOVA5		Fraud/Irreg (1)	$\geq 0,75$	149	23
		Normal (0)	$< 0,55$	302	821
Total				1543	4007

Tabela 45 – Comitê de Filtragem vs Clientes Inspeccionados – Zei 535/550/551/552/555/565/566 Residencial.

Zei 535/550/551/552/555/565/566 Residencial Leste			Limiar de Filtragem	Clientes Inspeccionados da Light (Target)	
				Fraud/Irreg (1)	Normal (0)
REDE NOVA1	Comitê de Filtragem de Redes Neurais	Fraud/Irreg (1)	$\geq 0,75$	538	106
		Normal (0)	$< 0,75$	1641	2633
REDE NOVA2		Fraud/Irreg (1)	$\geq 0,75$	505	123
		Normal (0)	$< 0,75$	1674	2589
REDE NOVA3		Fraud/Irreg (1)	$\geq 0,75$	548	139
		Normal (0)	$< 0,75$	1661	2537
REDE NOVA4		Fraud/Irreg (1)	$\geq 0,75$	539	101
		Normal (0)	$< 0,75$	1662	2643
REDE NOVA5		Fraud/Irreg (1)	$\geq 0,75$	570	108
		Normal (0)	$< 0,75$	1620	2626
Total				2700	13028

Tabela 46 – Comitê de Filtragem vs Clientes Inspeccionados – Zei 590 Residencial.

Zei 590 Residencial Oeste			Limiar de Filtragem	Clientes Inspeccionados da Light (Target)	
				Fraud/Irreg (1)	Normal (0)
REDE NOVA1	Comitê de Filtragem de Redes Neurais	Fraud/Irreg (1)	$\geq 0,8$	206	42
		Normal (0)	$< 0,4$	284	1465
REDE NOVA2		Fraud/Irreg (1)	$\geq 0,8$	192	39
		Normal (0)	$< 0,4$	275	1493
REDE NOVA3		Fraud/Irreg (1)	$\geq 0,8$	213	32
		Normal (0)	$< 0,4$	278	1529
REDE NOVA4		Fraud/Irreg (1)	$\geq 0,8$	209	35
		Normal (0)	$< 0,4$	285	1465
REDE NOVA5		Fraud/Irreg (1)	$\geq 0,8$	197	25
		Normal (0)	$< 0,4$	264	1492
Total				1386	7444

Tabela 47 – Comitê de Filtragem vs Clientes Inspeccionados – Zei 593 Residencial.

Zei 593 Residencial Oeste			Limiar de Filtragem	Clientes Inspeccionados da Light (Target)	
				Fraud/Irreg (1)	Normal (0)
REDE NOVA1	Comitê de Filtragem de Redes Neurais	Fraud/Irreg (1)	$\geq 0,75$	49	15
		Normal (0)	$< 0,75$	1172	1575
REDE NOVA2		Fraud/Irreg (1)	$\geq 0,75$	52	14
		Normal (0)	$< 0,75$	1207	1593
REDE NOVA3		Fraud/Irreg (1)	$\geq 0,75$	58	5
		Normal (0)	$< 0,75$	1222	1614
REDE NOVA4		Fraud/Irreg (1)	$\geq 0,75$	67	15
		Normal (0)	$< 0,75$	1225	1561
REDE NOVA5		Fraud/Irreg (1)	$\geq 0,75$	62	10
		Normal (0)	$< 0,75$	1217	1607
Total				288	7950

Tabela 48 – Comitê de Filtragem vs Clientes Inspeccionados – Zei 539 Residencial.

Zei 539 Residencial Oeste			Limiar de Filtragem	Clientes Inspeccionados da Light (Target)	
				Fraud/Irreg (1)	Normal (0)
REDE NOVA1	Comitê de Filtragem de Redes Neurais	Fraud/Irreg (1)	$\geq 0,75$	479	89
		Normal (0)	$< 0,65$	394	1186
REDE NOVA2		Fraud/Irreg (1)	$\geq 0,75$	464	94
		Normal (0)	$< 0,65$	347	1271
REDE NOVA3		Fraud/Irreg (1)	$\geq 0,75$	446	88
		Normal (0)	$< 0,65$	404	1270
REDE NOVA4		Fraud/Irreg (1)	$\geq 0,75$	447	97
		Normal (0)	$< 0,65$	409	1243
REDE NOVA5		Fraud/Irreg (1)	$\geq 0,75$	479	94
		Normal (0)	$< 0,65$	415	1201
Total				2245	6171

Tabela 49 – Comitê de Filtragem vs Clientes Inspeccionados Zei 585 Residencial.

Zei 585 Residencial Oeste			Limiar de Filtragem	Clientes Inspeccionados da Light (Target)	
				Fraud/Irreg (1)	Normal (0)
REDE NOVA1	Comitê de Filtragem de Redes Neurais	Fraud/Irreg (1)	$\geq 0,75$	298	57
		Normal (0)	$< 0,75$	1319	2337
REDE NOVA2		Fraud/Irreg (1)	$\geq 0,75$	301	64
		Normal (0)	$< 0,75$	1297	2350
REDE NOVA3		Fraud/Irreg (1)	$\geq 0,75$	312	57
		Normal (0)	$< 0,75$	1278	2326
REDE NOVA4		Fraud/Irreg (1)	$\geq 0,75$	320	63
		Normal (0)	$< 0,75$	1279	2325
REDE NOVA5		Fraud/Irreg (1)	$\geq 0,75$	327	65
		Normal (0)	$< 0,75$	1291	2324
Total				1558	11662

As seguintes bases novas de Treinamento, que serão utilizados no módulo de 'Classificação', para as zeis com clientes Não Residenciais.

Tabela 50 – Comitê de Filtragem vs Clientes Inspeccionados Zeis 538/540/544/545/535/550/551/552/555/565/566 Não Residencial.

Zeis 538/540/544/545/535/550/551/552/555/565/566 Não Residencial Leste		Limiar de Filtragem	Clientes Inspeccionados da Light (Target)		
			Fraud/Irreg (1)	Normal (0)	
REDE NOVA1	Comitê de Filtragem de Redes Neurais	Fraud/Irreg (1)	$\geq 0,75$	69	28
		Normal (0)	$< 0,75$	582	869
REDE NOVA2		Fraud/Irreg (1)	$\geq 0,75$	87	19
		Normal (0)	$< 0,75$	574	890
REDE NOVA3		Fraud/Irreg (1)	$\geq 0,75$	74	16
		Normal (0)	$< 0,75$	591	867
REDE NOVA4		Fraud/Irreg (1)	$\geq 0,75$	77	20
		Normal (0)	$< 0,75$	598	892
REDE NOVA5		Fraud/Irreg (1)	$\geq 0,75$	80	13
		Normal (0)	$< 0,75$	586	875
Total			387	4393	

Tabela 51 – Comitê de Filtragem vs Clientes Inspeccionados Zeis 539/585/590/593 Não Residencial.

Zeis 539/585/590/593 Não Residencial Oeste		Limiar de Filtragem	Clientes Inspeccionados da Light (Target)		
			Fraud/Irreg (1)	Normal (0)	
REDE NOVA1	Comitê de Filtragem de Redes Neurais	Fraud/Irreg (1)	$\geq 0,75$	71	13
		Normal (0)	$< 0,75$	365	662
REDE NOVA2		Fraud/Irreg (1)	$\geq 0,75$	63	15
		Normal (0)	$< 0,75$	343	660
REDE NOVA3		Fraud/Irreg (1)	$\geq 0,75$	79	15
		Normal (0)	$< 0,75$	342	658
REDE NOVA4		Fraud/Irreg (1)	$\geq 0,75$	63	14
		Normal (0)	$< 0,75$	343	640
REDE NOVA5		Fraud/Irreg (1)	$\geq 0,75$	70	20
		Normal (0)	$< 0,75$	318	663
Total			346	3283	

5.2.3. Módulo 3: Classificação na Light

Neste terceiro módulo de treinamento utilizando Redes Neurais, se procura, com a utilização de novas bases de dados filtradas, uma melhor caracterização entre os clientes verdadeiramente *Fraudadores/Irregulares* e clientes verdadeiramente *Normais* na Light.

Em função das configurações obtidas no módulo anterior, onde o número de neurônios na camada oculta foi próximo de 8, o treinamento foi feito variando

o número de neurônios de 4 a 10 neurônios, fazendo-se incremento de 2 em 2 neurônios. Da mesma forma que no módulo anterior, o neurônio único na saída representa os clientes suspeitos de apresentar algum tipo de *Irregularidade* pelo código 1, e os clientes *Normais* pelo código 0. A estrutura da Rede Neural utilizada no módulo de 'Classificação' é a mesma apresentada na Figura 28, contendo os atributos de entrada, 1 camada oculta e 1 neurônio na camada de saída.

5.3.Fase de Classificação na Light

5.3.1.Módulo 4: Indicação de Possíveis Irregularidades na Light

O Módulo 4 pertencente à Fase de *Classificação*, é responsável pela indicação dos clientes, de modo a detectar os clientes que são suspeitos de algum tipo de *irregularidade*, em diferentes níveis de prioridade, conforme descrito no capítulo 4. De forma a avaliar o desempenho do SIIPERCOM na indicação de irregularidades, foram considerados registros de clientes investigados nunca apresentados a nosso sistema inteligente SIIPERCOM dos meses de janeiro, fevereiro, julho e agosto do 2006, com o objetivo de verificar o desempenho do sistema em diferentes estações do ano (verão e inverno). As métricas de avaliação foram as apresentadas no capítulo 4.

As Tabelas 52 e 53 apresentam o número de clientes Residenciais e Não-Residenciais pré-processadas, respectivamente, que compõem a Base de Dados de Avaliação da Regional Leste.

Tabela 52 – Número de Clientes da Base de Dados de Avaliação Residenciais da Regional Leste classificados segundo os resultados da inspeção

Base de Dados	Fraude	Irregular	Normal
538	545	250	1906
540	776	272	1893
544	1096	435	3132
545	576	291	2027
535/550/551/552/555 /565/566	1764	905	6203
Total	4757	2153	15161

Tabela 53 – Número de Clientes da Base de Dados de avaliação Não Residenciais da Regional Leste classificados segundo os resultados da inspeção.

Base de Dados	Fraude	Irregular	Normal
538/540/544/545/535 /550/551/552/555/565/566	289	292	3759
Total	289	292	3759

Já as Tabelas 54 e 55, a seguir, apresentam o número de clientes Residenciais e Não-Residenciais pré-processadas, respectivamente, que compõem a Base de Dados de Avaliação da Regional Oeste.

Tabela 54 - Número de Clientes da Base de Dados de avaliação Residenciais da Regional Oeste classificados segundo os resultados da inspeção.

Base de Dados	Fraude	Irregular	Normal
539	1073	1510	4029
580	0	0	0
585	1240	1944	4252
590	1498	2131	4995
593	1029	771	2087
Total	4840	6356	15363

Tabela 55 – Número de Clientes da Base de Dados de avaliação Não Residenciais da Regional Oeste classificados segundo os resultados da inspeção.

Base de Dados	Fraude	Irregular	Normal
539/580/585/590/593	139	170	1011
Total	139	170	1011

Da mesma forma que no *Módulo de Filtragem*, a indicação de clientes irregulares é realizada em duas etapas: análise do *Limiar de Classificação*, para binarizar a saída de ativação de cada rede; e análise do *Comitê de Classificação*, o qual verifica os votos do comitê de redes neurais.

Para o *Limiar de Classificação* utilizou-se apenas o critério de limiar único, onde usa-se apenas um valor de limiar, acima do qual determina-se o cliente como irregular/fraudador e abaixo do qual classifica-se o cliente como normal. Os critérios utilizados com ambos limiares são apresentados nas tabela 56 e 57 a seguir:

Tabela 56 – Limiares de Classificação utilizados no módulo de Classificação da Light.

Limiares de Classificação	
Limiar Simples	
Irregular	Normal
$\geq 0,40$	$< 0,40$
$\geq 0,45$	$< 0,45$
$\geq 0,5$	$< 0,5$
$\geq 0,55$	$< 0,55$
$\geq 0,60$	$< 0,60$
$\geq 0,65$	$< 0,65$
$\geq 0,70$	$< 0,70$

Tabela 57 – Limiares do Comitê de Classificação utilizados no módulo de Classificação da Light.

Limiares do Comitê de Classificação		
Primeira Avaliação	Irregular ≥ 5	5 redes votem 1
	Normal ≥ 1	1 ou mais redes votem 0
Segunda Avaliação	Irregular ≥ 4	4 ou 5 redes votem 1
	Normal ≥ 2	2 ou mais redes votem 0
Terceira Avaliação	Irregular ≥ 3	3 ou mais redes votem 1
	Normal ≥ 3	3 ou mais redes votem 0

A partir do produto da votação do *Comitê de Classificação* se faz a comparação com os clientes inspecionados reais da Base de Dados de Avaliação mediante uma matriz de confusão. A análise do desempenho, apresentada nas próximas seções, foi realizada utilizando as quatro métricas básicas:

- Acertividade VPP_{Light};
- Acertividade VPP_{SIIPERCOM};
- Sensibilidade_{SIIPERCOM};
- Erro de classificação.

Para uma melhor comparação dos resultados do Sistema Inteligente SIIPERCOM utilizou-se a Base de Dados de Avaliação e com ajuda do *Comitê*

de *Classificação* na fase de *Treinamento* e *Classificação*, se determinou o grau de irregularidade (ou normalidade) de cada cliente.

Os resultados obtidos para cada Zei, em cada uma das duas categorias de clientes, são apresentados nos anexos VIII ao XVIII, e o resumo final para as regionais Leste e Oeste Residências e Não Residenciais é mostrado nas tabelas 58, 59, 60 e 61.

5.4. Resumo Resultados de Treinamento Residencial Leste

A Tabela 58 apresenta o resumo final dos resultados do Módulo de Filtragem para a Regional Leste Residências e Não Residenciais.

Observa-se que a melhor Acertividade_{SIIPERCOM} é de 55,4%; com uma Acertividade_{Light} de 30,1%. A Sensibilidade tem como valor 24,5%. O erro de classificação apresenta um valor de 28,6%.

Tabela 58 – Resumo de Resultados Regional Leste (Módulo de Filtragem).

Zeis	Limiar Classificação	Limiar do Comitê de Classificação	VPP _{SIIPERCOM}	Sensibilidade	Erro Class	VPP _{Light}
Residencial						
538	0,60	Prim ¹	43,6	26,8	31,7	29,4
540	0,55	Prim ¹	51,9	41,6	34,5	35,6
544	0,50	Prim ¹	46,9	38,8	34,5	32,8
545	0,70	Prim ¹	43,2	9,6	30,9	30,0
535/550/551/552/555/565/566	0,65	Prim ¹	55,4	24,5	28,6	30,1
Não Residencial						
538/540/544/545/535/550/551/552/555/565/566	0,40	Prim ¹	22,7	56,8	31,6	13,4

¹ 5 redes votem de 1 e 1 ou mais redes votem de 0

² 4 ou 5 redes votem de 1 e 2 ou mais redes votem de 0

³ 3 ou mais redes votem de 1 e 3 ou mais redes votem de 0

A Tabela 59 apresenta o resumo final dos resultados do Módulo de Classificação para a Regional Leste Residências e Não Residenciais.

Observa-se que a melhor Acertividade_{SII PERCOM} é de 62,6%; com uma Acertividade_{Light} de 32,8%. A Sensibilidade tem como valor 9,4%. O erro de classificação apresenta um valor de 31,3%.

Tabela 59 – Resumo de Resultados Regional Leste (Módulo de Classificação).

Zeis	Limiar Classificação	Limiar do Comitê de Classificação	VPP _{SII PERCOM}	Sensibilidade	Erro Class	VPP _{Light}
Residencial						
538	0,60	Prim ¹	47,8	27,2	30,2	29,4
540	0,55	Prim ¹	55,6	36,3	33,0	35,6
544	0,50	Prim ¹	62,6	9,4	31,3	32,8
545	0,70	Prim ¹	44,2	29,0	32,2	30,0
535/550/551/552/555/565/566	0,65	Prim ¹	56,3	24,4	28,4	30,1
Não Residencial						
538/540/544/545/535/550/551/552/555/565/566	0,40	Prim ¹	24,9	11,7	16,5	13,4

¹ 5 redes votem de 1 e 1 ou mais redes votem de 0

² 4 ou 5 redes votem de 1 e 2 ou mais redes votem de 0

³ 3 ou mais redes votem de 1 e 3 ou mais redes votem de 0

Verifica-se pelas tabelas 58 (Módulo de Filtragem) e 59 (Módulo de Classificação), um desempenho superior da Acertividade_{SII PERCOM} em cada uma das Zeis que compoem a Regional Leste Residenciais e Não Residenciais, frente à Acertividade_{Light}. O qual permite uma seleção mais precisa dos clientes de baixa tensão a serem inspecionados.

5.5. Resumo Resultados de Treinamento Residencial Oeste

A Tabela 60 apresenta o resumo final dos resultados do Módulo de Filtragem para a Regional Oeste Residências e Não Residenciais.

Observa-se que a melhor Acertividade_{SII PERCOM} é de 61,4%; com uma Acertividade_{Light} de 46,3%. A Sensibilidade tem como valor 4,3%. O erro de classificação apresenta um valor de 45,1%.

Tabela 60 – Resumo de Resultados Regional Oeste (Módulo de Filtragem).

Zeis	Limiar Classificação	Limiar do Comitê de Classificação	VPP _{SII} PERCOM	Sensibilidade	Erro Class	VPP Light
Residencial						
590	0,70	Tercer ³	44,2	33,6	45,8	42,1
593	0,70	Prim ¹	61,4	4,3	45,1	46,3
539	0,70	Segun ²	43,5	35,3	43,2	39,1
585	0,65	Prim ¹	54,2	12,9	42,0	42,8
Não Residencial						
539/585/590/593	0,40	Prim ¹	31,0	73,8	44,6	23,4

¹ 5 redes votem de 1 e 1 ou mais redes votem de 0

² 4 ou 5 redes votem de 1 e 2 ou mais redes votem de 0

³ 3 ou mais redes votem de 1 e 3 ou mais redes votem de 0

A Tabela 61 apresenta o resumo final dos resultados do Módulo de Classificação para a Regional Oeste Residências e Não Residenciais.

Observa-se que a melhor Acertividade_{SII}PERCOM é de 63,9%; com uma Acertividade_{Light} de 46,3%. A Sensibilidade tem como valor 8,6%. O erro de classificação apresenta um valor de 44,6%.

Tabela 61 – Resumo de Resultados Regional Oeste (Módulo de Classificação).

Zeis	Limiar Classificação	Limiar do Comitê de Classificação	VPP _{SII} PERCOM	Sensibilidade	Erro Class	VPP Light
Residencial						
590	0,70	Tercer ³	46,1	58,4	46,2	42,1
593	0,70	Prim ¹	63,9	8,6	44,6	46,3
539	0,70	Segun ²	45,6	49,8	42,8	39,1
585	0,65	Prim ¹	57,4	21,4	40,5	42,8
Não Residencial						
539/585/590/593	0,40	Prim ¹	35,6	34,3	29,9	23,4

¹ 5 redes votem de 1 e 1 ou mais redes votem de 0

² 4 ou 5 redes votem de 1 e 2 ou mais redes votem de 0

³ 3 ou mais redes votem de 1 e 3 ou mais redes votem de 0

Verifica-se pelas tabelas 60 (Módulo de Filtragem) e 61 (Módulo de Classificação), um desempenho superior da Acertividade_{SUPERCOM} em cada uma das Zeis que compoem a Regional Oeste Residenciais e Não Residencias, frente à Acertividade_{Light} . O qual permite uma seleção mais precisa dos clientes de baixa tensão a serem inspecionados.