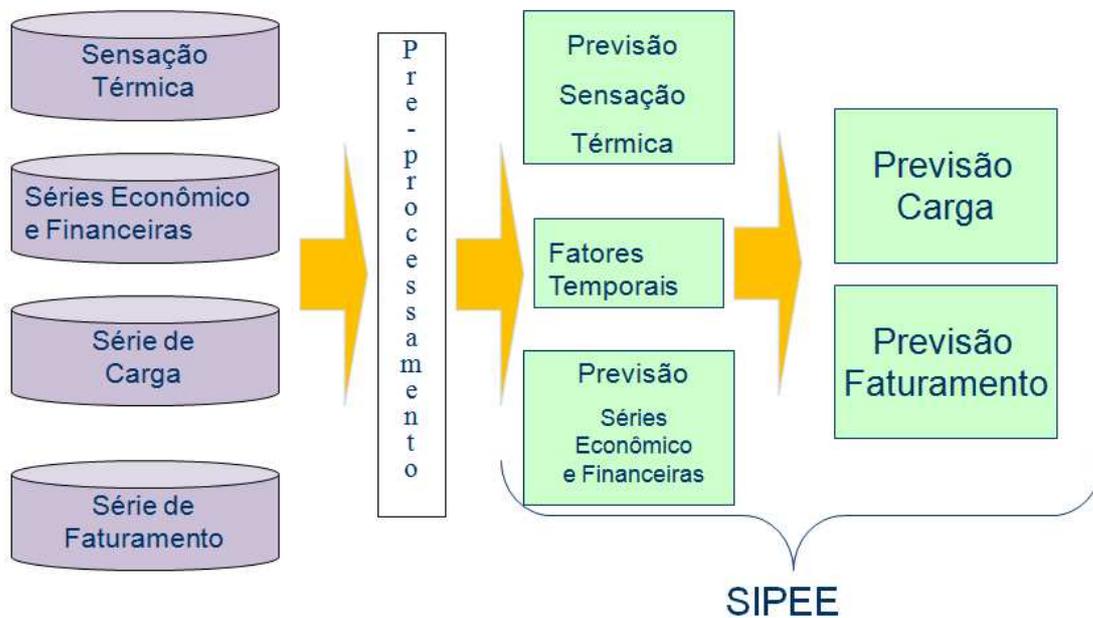


### 3. SIPEE – Sistema Inteligente para previsão de energia elétrica.

Este modelo foi desenvolvido no MATLAB para uma previsão de médio prazo, isto é 15 meses à frente, fazendo com que a Concessionária possa programar a geração, avaliar a segurança do sistema, planejando assim, os serviços de manutenção preventiva necessários, minimizando a ocorrência de defeitos e aumentando seu grau de acerto na previsão da compra de energia elétrica, evitando assim as multas impostas pela ANEEL.

A figura 3.1 mostra o diagrama do sistema, que descreve a metodologia utilizada para sua elaboração:



.Figura 3.1: Diagrama do sistema

#### 3.1 Variáveis de Entrada

O objetivo de se usar métodos de seleção de variáveis é reconhecer quais entradas, ou atributos, desempenham um papel importante na definição das saídas e eliminar entradas que porventura sejam irrelevantes ou não

contribuam para um melhor desempenho do sistema a ser utilizado. Dessa forma, a seleção de variáveis reduz a dimensão do vetor de entrada do modelo, eliminando entradas de pouca importância, o que contribui para um melhor desempenho dos algoritmos de previsão (SOUZA, 2010). Além disso, ao se verificar quais atributos de entrada exercem maior influência na variável de saída, também se obtém informação valiosa sobre o problema estudado.

A seleção dos atributos de entrada foi feita através de entrevistas com especialistas da empresa distribuidora de energia, sendo assim, para esta previsão, os valores de entrada do sistema são as séries:

- Série de Sensação Térmica – Segundo a EPE, o principal fator de estímulo ao consumo de energia elétrica no Sudeste (região onde está localizada a Concessionária) foi a temperatura. Comparados a 2008, os valores registrados nos últimos meses de 2009 foram mais elevados nos quatro estados da região, considerando que o ciclo de faturamento inclui dias de novembro e dezembro. Os efeitos foram mais intensos na capital do Rio de Janeiro, onde, na média dos dois últimos meses do ano, a temperatura foi 2,4 °C mais alta que a de 2008. Isso levou a uma alta no consumo residencial de 26%, além de um acréscimo de 18% no consumo comercial. Devido a importância deste atributo de entrada, os dados desta série foram adquiridos junto a empresa Climatempo e apesar de serem fornecidos os valores diários de sensação térmica, temperatura mínima e máxima, umidade e velocidade do vento, neste trabalho, optou-se em utilizar somente a sensação térmica, pois esta, acordo com o INMET (Instituto Nacional de Meteorologia), é a temperatura que o “nosso corpo sente”. Para esses dados foram calculadas as médias mensais da sensação térmica, de modo que todas as entradas do SIPEE fossem síncronas com os dados mensais fornecidos pela Concessionária, os valores de carga e faturamento. A granularidade escolhida pela empresa foi bairro para o município do Rio de Janeiro e municípios para todos os outros municípios da área de concessão da Light e o período foi de janeiro de 1998 até dezembro de 2009, gerando assim, 144 dados para cada bairro e município da área de concessão desta Concessionária.
- Série Econômica e financeira – Sobre o aumento do consumo de energia no começo do ano de 2011, Tolmasquim (presidente da EPE) afirmou que a expansão de 6,5% em janeiro está, em parte, relacionada à queda provocada pela crise econômica internacional de 2008/2009. Segundo ele, o consumo da indústria nacional cresceu muito no primeiro mês deste ano, sobretudo na Região Sudeste (9,6%). Isto significa que são diversos os fatores que

contribuem para a expansão do setor elétrico, merecendo destaque a solidez do sistema financeiro, o mercado de trabalho, a queda da taxa média de juros cobrados de pessoas físicas e a ampliação do crédito. Neste trabalho, foram pesquisadas e avaliadas séries que tivessem relação com cada classe de faturamento e seus valores foram retirados dos sites do IBGE, Banco Central do Brasil e do DIEESE. Cabe salientar que algumas dessas séries não possuem valores mensais e sim trimestrais, que foram trabalhados para estar em sincronia com os valores de carga e faturamento fornecidos pela Concessionária. O período utilizado de cada série foi de março/2002 a dezembro/2009.

- Série de Carga – O sistema elétrico brasileiro é constituído basicamente por usinas geradoras de energia, estações de elevação de tensão, linhas de transmissão, estações abaixadora de tensão e redes de distribuição. Através desse conjunto grandes blocos de energia são gerados, transformados e depois distribuídos pelas Distribuidoras de Energia para os consumidores conectados em sua rede. Esse bloco de energia distribuído pela Concessionária é medido através de equipamentos de medição instalados em suas subestações, sendo assim, essa série foi composta pelos resultados dessas medições compreendidas entre abril/2002 e junho/2010. Os dados fornecidos pela Light S.E.S.A. já são mensais, porém tiveram que ser divididos em bairros e municípios para acompanhar a granularidade da série de sensação térmica. Esta divisão foi feita através do conhecimento da área de abrangência de cada subestação e de sua carga total instalada.
- Série de Faturamento – Os dados foram fornecidos pela Concessionária, porém, de acordo com a Resolução 414 da ANEEL, em vigor desde setembro/2010, a leitura dos equipamentos de medição, que dará origem ao consumo faturado, deverá ser feita em intervalos de, aproximadamente, 30 dias, de acordo com o seu calendário de leitura. Então se uma determinada unidade consumidora estiver cadastrada no calendário de leitura como lote 10, seu medidor será lido no 10º dia útil de cada mês e seu consumo será a diferença de dois meses subseqüentes, conforme exemplo abaixo:

Unidade Consumidora – lote 10

Data da 1ª Leitura – 14/04/2011 – Leitura 5000

Data da 2ª Leitura – 16/05/2011 – Leitura 6000

Consumo = 6000 – 5000 = 1000 kWh

Esse fato gera o segundo problema encontrado, pois essa variável não é mensal e sim por lote. No exemplo anterior, pode-se ver que o consumo foi

registrado em dois meses. Adiante, será mostrado o procedimento que fará com que essa variável se torne mensal.

### **3.2 Pré processamento das bases de dados**

Em uma rede neural, o tempo de processamento está diretamente relacionado à quantidade atributos de entradas na base de dados. Um atributo de entrada pode ser visto como uma propriedade ou característica que agrega valor a informação processada pela rede. Entretanto, algumas dessas características podem não estar contribuindo para melhorar o desempenho durante o treinamento da rede neural, pois seus valores podem não ter um relacionamento que contribua de forma significativa com os exemplos de saída, podendo inclusive resultar em um processamento com menores taxas de acerto (SOUSA, 2010).

Um atributo de entrada desnecessário pode ser perturbador. Além de não classificar a informação adequadamente e aumentar a quantidade de informação na base de dados, aumentando o tempo de processamento durante o treinamento, a eliminação deste atributo poderia resultar em um treinamento mais rápido e até com desempenho significativamente melhor. (SOUSA, 2010)

#### **3.2.1 Série de Sensação Térmica**

Como a área de concessão da Light engloba uma grande parte do estado do Rio de Janeiro, a empresa forneceu os dados da seguinte maneira:

- Por município – totalizando 30 municípios;
- Para o município do Rio de Janeiro – a empresa forneceu a sensação térmica por bairros, totalizando 173 bairros;

A partir desses dados, foi feita uma avaliação com o objetivo de agrupar bairros e municípios em microclimas, de forma a evitar uma quantidade proibitiva de modelos. Além disso, devido à semelhança climática de algumas regiões, era esperado que muitos modelos apresentassem comportamentos semelhantes. Assim, a previsão da sensação térmica na área de concessão da Light passou por um processo de agrupamento, definindo-se um conjunto de microclimas por meio do processo de clusterização. Aproveitando a maneira com que os dados

foram fornecidos e a diversidade do relevo do Município do Rio de Janeiro, partiu-se para a divisão de dois grupos principais: um com os bairros do Município do Rio de Janeiro e outro para os Municípios restantes. Para esse agrupamento foi utilizado o modelo K-Means, que foi primeiramente usado por James MacQueen (MacQueen, 1967), embora a idéia remonte a Hugo Steinhaus (Steinhaus 1957).

O objetivo do método K-Means (também chamado de K-Médias) é separar as informações diferentes e aproximar informações semelhantes de acordo com os próprios dados. Esta “divisão” e “agregação” são baseadas em análise e comparações entre os valores numéricos da base de dados utilizada. Desta maneira, o algoritmo automaticamente fornece uma clusterização sem a necessidade de nenhuma supervisão humana, ou seja, sem nenhuma pré-classificação existente. Devido a esta característica, o K-Means é considerado como um algoritmo de mineração de dados não supervisionado.

Como já foi mencionado anteriormente, o algoritmo analisa todos os dados e cria, segundo um padrão desejado de acurácia, grupos. Nesse método deve-se fornecer ao algoritmo a quantidade de cluster que se deseja separar os dados. Este número de clusters que deve ser informado para o algoritmo é chamado de  $k$  e é daí que vem a primeira letra do algoritmo: K-Means.

Cabe ressaltar que para o agrupamento das regiões da área de concessão da Light, considerando a característica do clima, foram investigadas várias variáveis climáticas (temperatura máxima, temperatura média, sensação térmica e sensação térmica no verão), porém a sensação térmica foi a que apresentou o resultado mais consistente. Já o parâmetro  $k$ , ficou entre 2 e 12. Para se determinar o melhor  $k$  (número de clusters) foi utilizada a variância dos dados (sensação térmica) dos clusters. Assim, o critério de escolha do número  $k$  de clusters para os bairros e municípios considerou como parâmetro a variância média no ano em torno de 1 grau centígrado, ou seja, os bairros ou municípios que pertencessem a um cluster poderiam ter, em média ao longo do ano, variância em torno de 1 grau centígrado. Por questões de otimização do tempo, este foi o único algoritmo estudado.

O método K-Means identificou 17 microclimas, divididos em 2 grupos:

- 8 microclimas para os bairros do município do Rio de Janeiro;
- 9 microclimas para os municípios da área de concessão da Light S.E.S.A.,

Pode-se ver nas tabelas 3.1 e 3.2 a identificação dos microclimas:

MICROCLIMA B1	MICROCLIMA B2	MICROCLIMA B3	MICROCLIMA B4
ACARI, ANCHIETA, BARROS FILHO, COELHO NETO, COLEGIO, COSTA BARROS, DEODORO, GUADALUPE, HONORIO GURJEL, PAVUNA, RICARDO DE ALBUQUERQUE, ROCHA MIRANDA;	ANIL, BARRA DA TIJUCA, BOTAFOGO, CATETE, CATUMBI, CENTRO(RJ), CIDADE DE DEUS, CIDADE NOVA, JURICICA, ESTACIO, FREGUESIA, GAMBOA, GARDENIA AZUL, GLORIA, GRAJAU, JACAREPAGUA, PECHINCHA, PEDRA DE GUARATIBA, PRACA DA BANDEIRA, RIO COMPRIDO, SANTA TERESA, SAUDE e TIJUCA;	BANGU, CAJU, GALEAO, PADRE MIGUEL, SANTA CRUZ, SANTISSIMO, SANTO CRISTO, SAO CRISTOVAO, SENADOR CAMARA e SENADOR VASCONCELOS	ALTO DA BOA VISTA, BARRA DE GUARATIBA e GRUMARI;
MICROCLIMA B5	MICROCLIMA B6	MICROCLIMA B7	MICROCLIMA B8
AGUA SANTA, ANDARAÍ, BENFICA, CAMPO GRANDE, COMPLEXO DA MARE, COSMOS, ENGENHO NOVO, ILHA CIDADE UNIVERSITARIA, INHOAIBA, LINS DE VASCONCELOS, MANGUEIRA, MANGUINHOS, MARACANA, MARE, PACIENCIA, RIACHUELO, ROCHA, SAC FRANCISCO XAVIER, VASCO DA GAMA e VILA ISABEL;	COPACABANA, HUMAITA, IPANEMA, LAGOA, LEBLON, LEME, RECREIO DOS BANDEIRANTES, SEPETIBA e URCA;	BANCARIOS, CACUIA, CAMORIM, COCOTA, COSME VELHO, FLAMENGO, FREGUESIA IG, GAVEA, GUARATIBA, ILHA DO GOVERNADOR, ITANHANGA, JARDIM BOTANICO, JARDIM CARICCA, JARDIM GUANABARA, JOA LARANJEIRAS, MONERO, PAQUETA, PITANGUEIRAS, PORTUGUESA, PRAIA DA BANDEIRA, RIBEIRA, ROCINHA, SAO CONRADO, TALUA, VARGEM GRANDE, VARGEM PEQUENA, VIDIGAL e ZUMBI;	ABOLICAO, BENTO RIBEIRO, BONSUCESSO, BRAZ DE PINA, CACHAMBI, CAMPINHO, CAMPO DOS AFONSO, CASCADURA, CAVALCANTI, CORDOVIL, DEL CASTILHO, ENCANTADO, ENGENHEIRO LEAL, ENGENHO DA RAINHA, ENGENHO DE DENTRO, HIGIENOPOLIS, INHAUVA, IRAJA, JACARE, JARDIM AMERICA, JARDIM SULACAP, MADUREIRA, MARECHAL HERMES, MARIA DA GRACA, MEIER, OLARIA, OSWALDO CRUZ, PARADA DE LUCAS, PENHA, PENHA CIRCULAR, PIEDADE, PILARES, PRACA SECA, QUINTINO BOCAIUVA, RAMOS, SAMPAIO, TANQUE, TAQUARA, TODOS OS SANTOS, TOMAS COELHO, TURIACU, VAZ LOBO, VICENTE DE CARVALHO, VIGARIO GERAL, VILA DA PENHA, VILA COSMOS, VILA MILITAR, VILA VALQUEIRE, VISTA ALEGRE, MAGALHAES BASTOS, PARQUE ANCHIETA e REALENGO.

Tabela 3.1: Microclimas de Bairros

MICROCLIMA M1	MICROCLIMA M2	MICROCLIMA M3	MICROCLIMA M4	MICROCLIMA M5
ENG. PAULO DE FRONTIN e ITAGUAÍ	PARACAMBI e TRÊS RIOS	JAPERI, PIRAI e SEROPÉDICA	BARRA DO PIRAI e PINHERAL	MIGUEL PEREIRA e PATY DO ALFERES
MICROCLIMA M6	MICROCLIMA M7	MICROCLIMA M8	MICROCLIMA M9	
CARMO, RIO CLARO, RIO DAS FLORES, RIO PRETO (MG), SAPUCAIA E VALENÇA	BELFORD ROXO, DUQUE DE CAXIAS, MESQUITA, NILÓPOLIS, NOVA IGUAÇU, QUEIMADOS e SÃO JOÃO DE MERITI.	BELMIRO BRAGA (MG), CDOR LEVY GASPARIAN, CHIADOR (MG), MENDES, PARAÍBA DO SUL, QUATIS, SANTANA DO DESERTO (MG), SIMÃO PEREIRA (MG) e VASSOURAS	BARRA MANSA e VOLTA REDONDA	

Tabela 3.2: Microclimas de Municípios

As figuras 3.2 a 3.18 mostram a curva de temperatura dos microclimas.

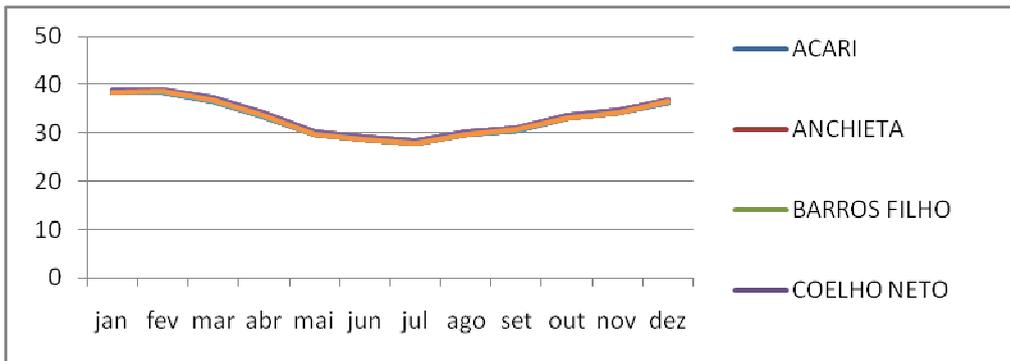


Figura 3.2 – Microclima Bairro (B1) com sensação térmica média de 36.39

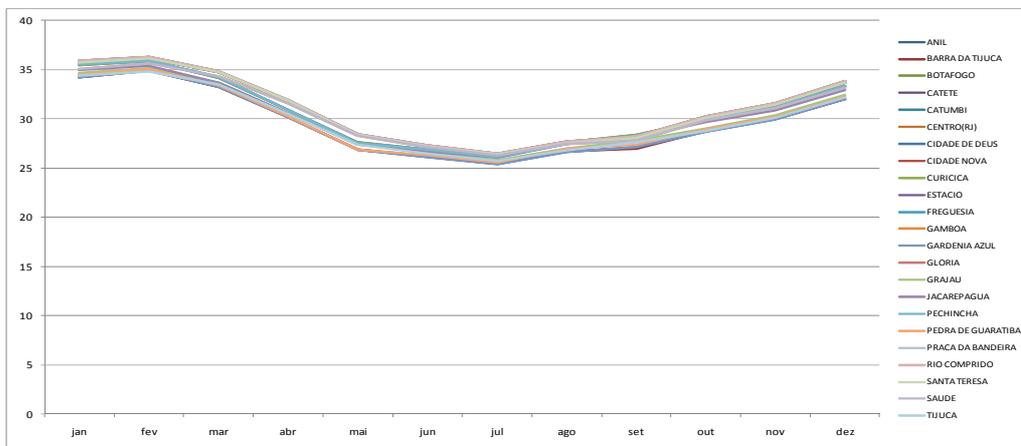


Figura 3.3– Microclima Bairro (B2) com sensação térmica média de 33.19

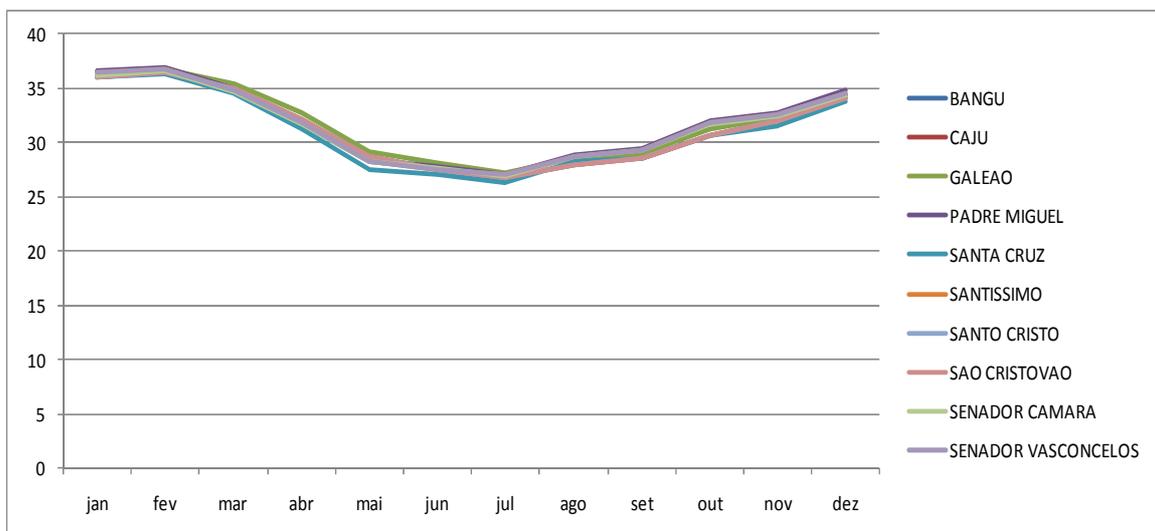


Figura 3.4 – Microclima Bairro (B3) com sensação térmica média de 34.25

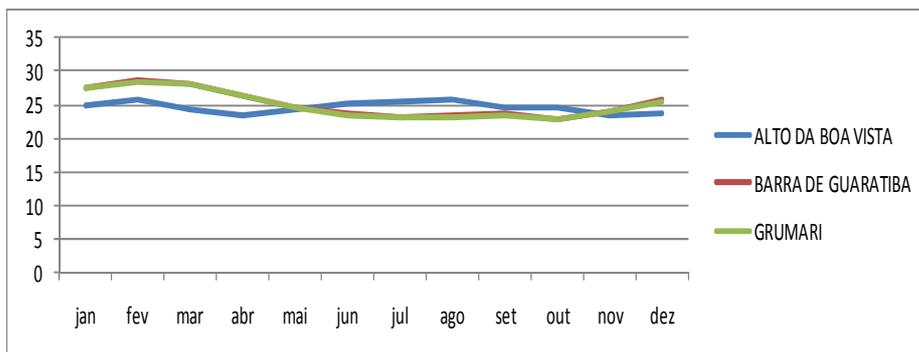


Figura 3.5 – Microclima Bairro (B4) com sensação térmica média de 25.56

Na figura 3.5, pode-se observar que os bairros desse microclima, possuem características climáticas diferentes. No Alto da Boa Vista se registram as menores temperaturas da cidade do Rio de Janeiro, em todas as épocas do ano, sendo que todos os anos as temperaturas caem para abaixo de 11 graus. Seu clima ameno, em comparação com o restante da cidade, deve-se, tanto ao fato de estar situado a uma altitude média de 300 metros acima do nível do mar, quanto ao fato de estar totalmente cercado pela Mata Atlântica, já os outros dois bairros fazem parte do litoral carioca. Para que o sistema não fosse prejudicado com uma quantidade de modelos excessivo, optou-se em deixar esses bairros dentro desse microclima

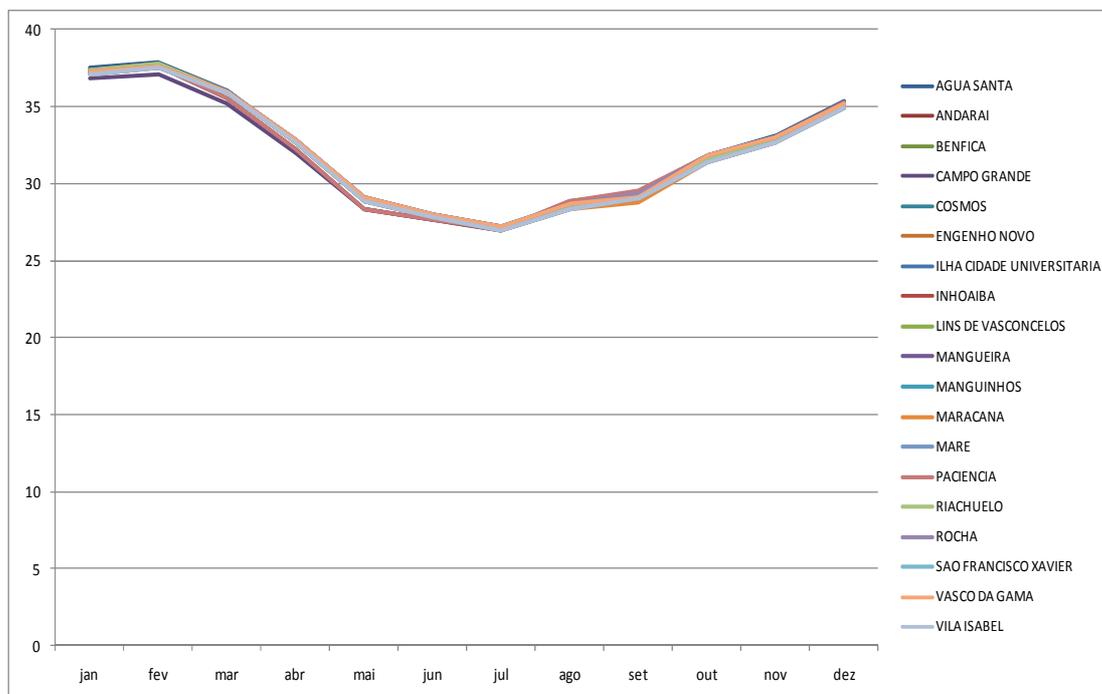


Figura 3.6 – Microclima Bairro (B5) com sensação térmica média de 35.07 . Foi considerado que a area do Complexo da Maré pertence a nesse microclima (B5).

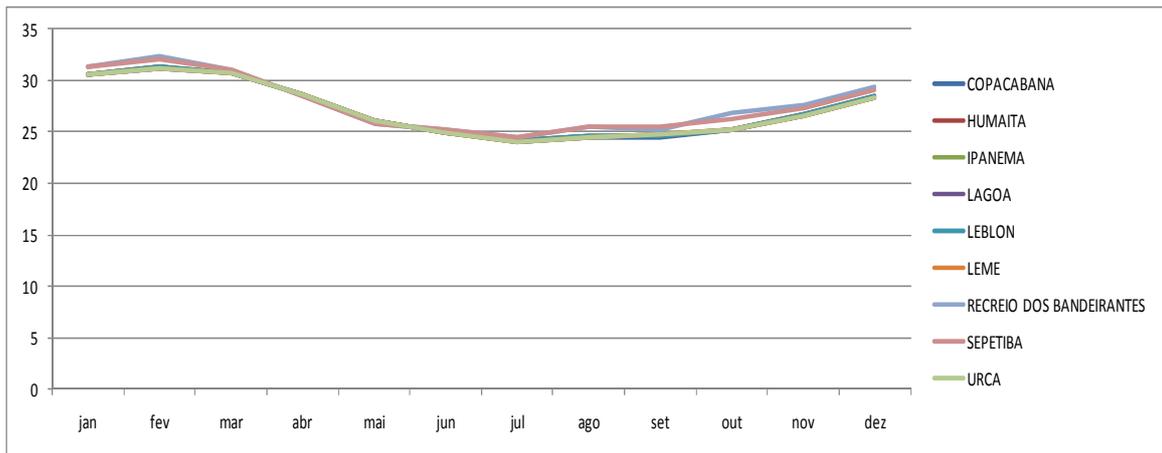


Figura 3.7 – Microclima Bairro (B6) com sensação térmica média de 28.98. Foi considerado que o bairro Fonte da Saudade pertence a nesse microclima (B6).

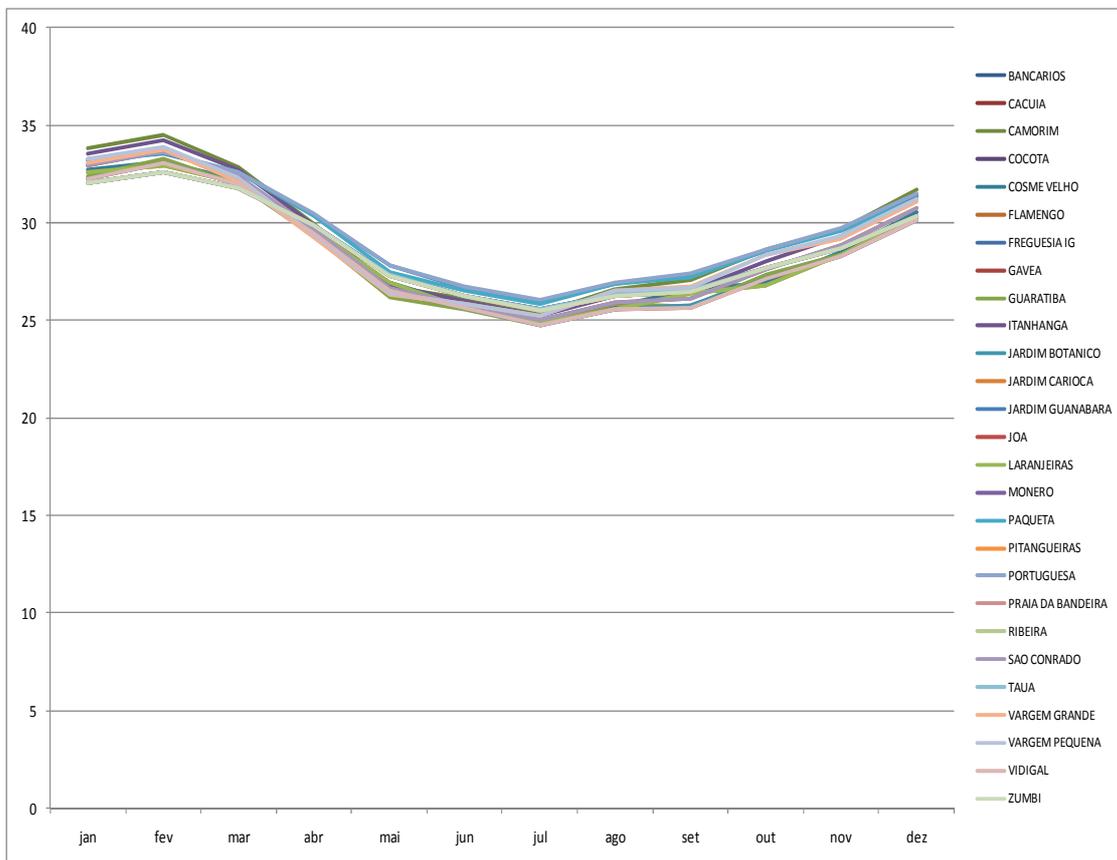


Figura 3.8 – Microclima Bairro (B7) com sensação térmica média de 30.82. Foi considerado que os bairros Ilha do Governador e Rocinha pertencem a esse microclima (B7).

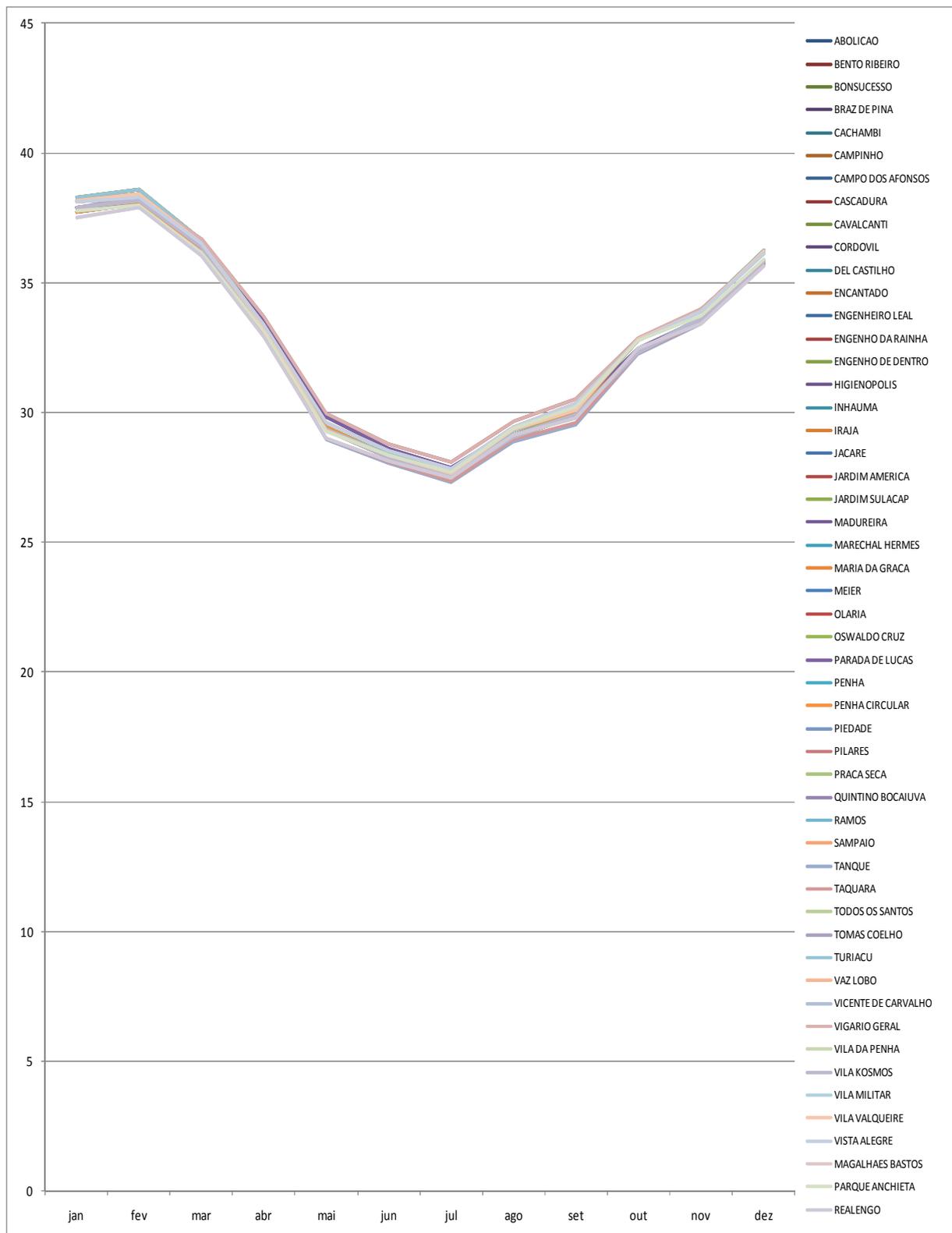


Figura 3.9 – Microclima Bairro (B8) com sensação térmica média de 35.78. Foi considerado que o Complexo do Alemão, Jacarezinho e Parque Columbia pertencem a esse microclima (B8).

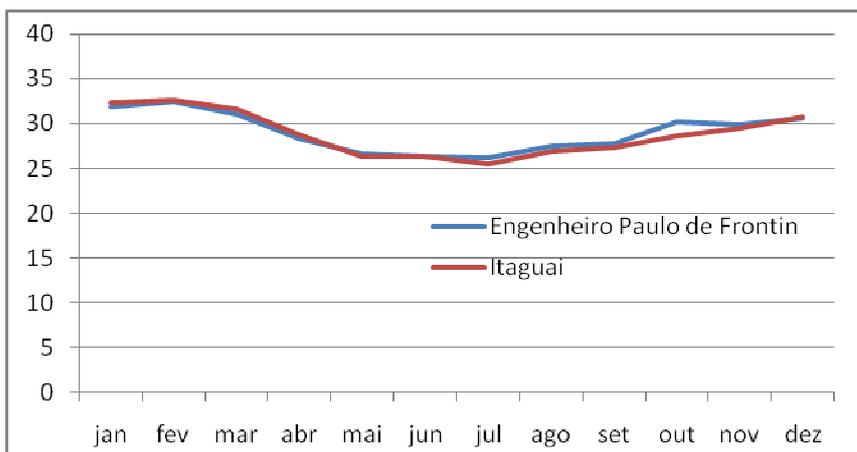


Figura 3.10 – Microclima Municípios 1 (M1) com sensação térmica média 30.9488

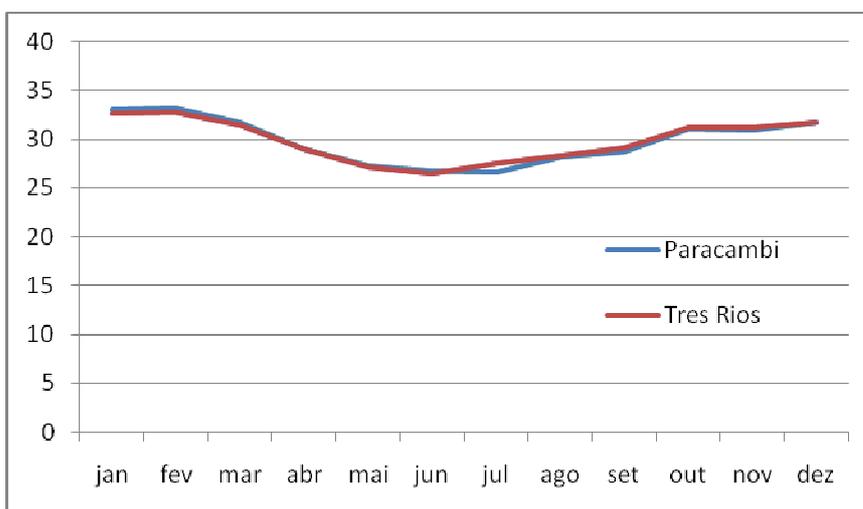


Figura 3.11 – Microclima Municípios 2 (M2) com sensação térmica média 31.8511

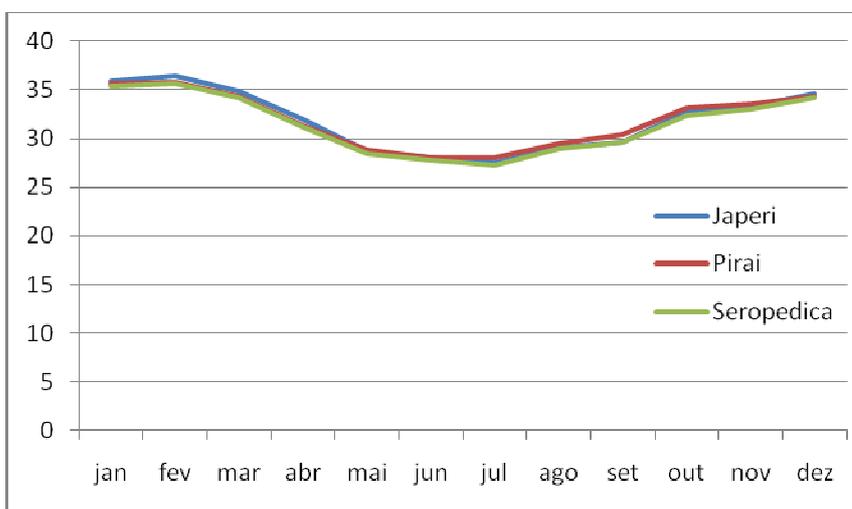


Figura 3.12 – Microclima Municípios 3 (M3) com sensação térmica média 34.4043

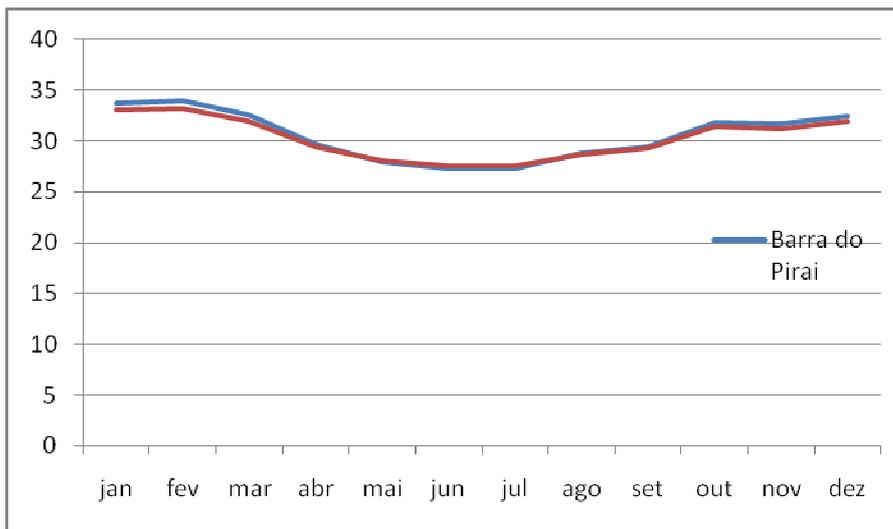


Figura 3.13 – Microclima Municípios 4 (M4) com sensação térmica média 32.3992

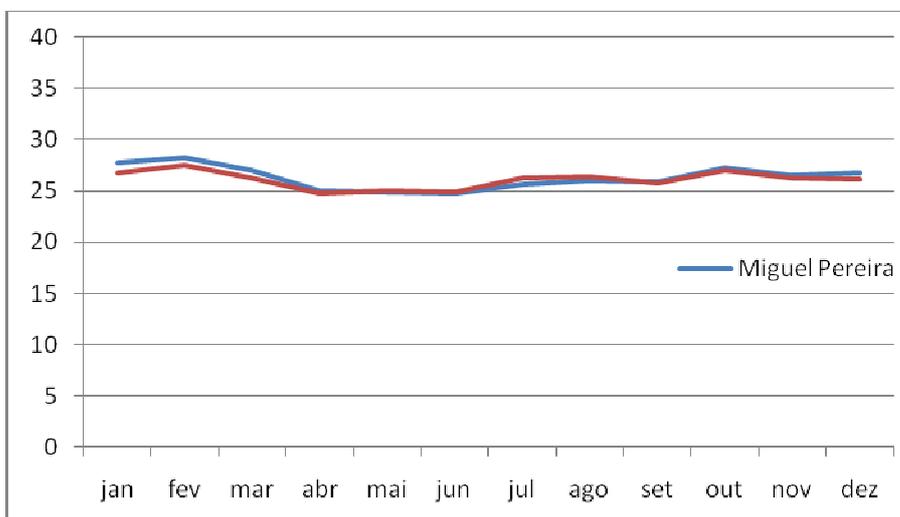


Figura 3.14 – Microclima Municípios 5 (M5) com sensação térmica média de 26.9617

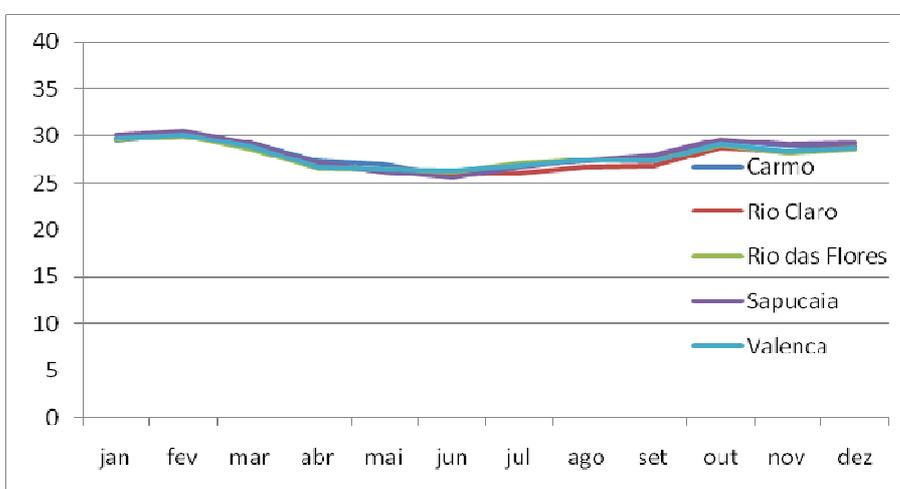


Figura 3.15 – Microclima Municípios 6 (M6) com sensação térmica média de 29.2599

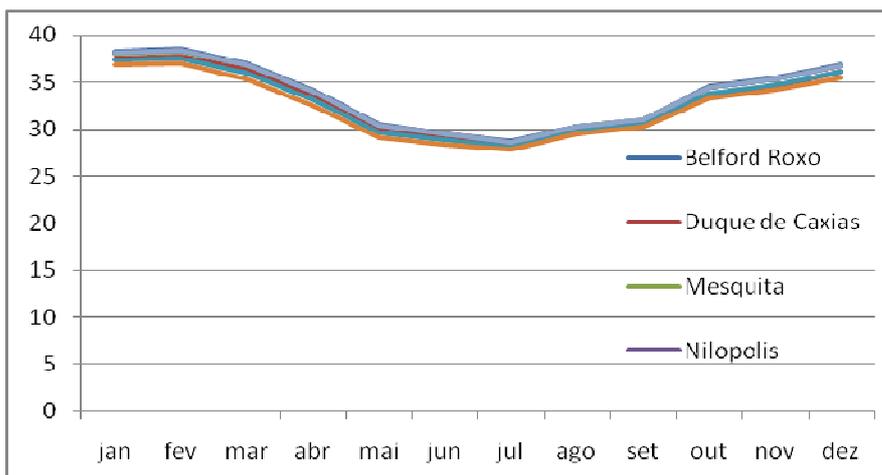


Figura 3.16 – Microclima Municípios 7 (M7) com sensação térmica média de 36.2735

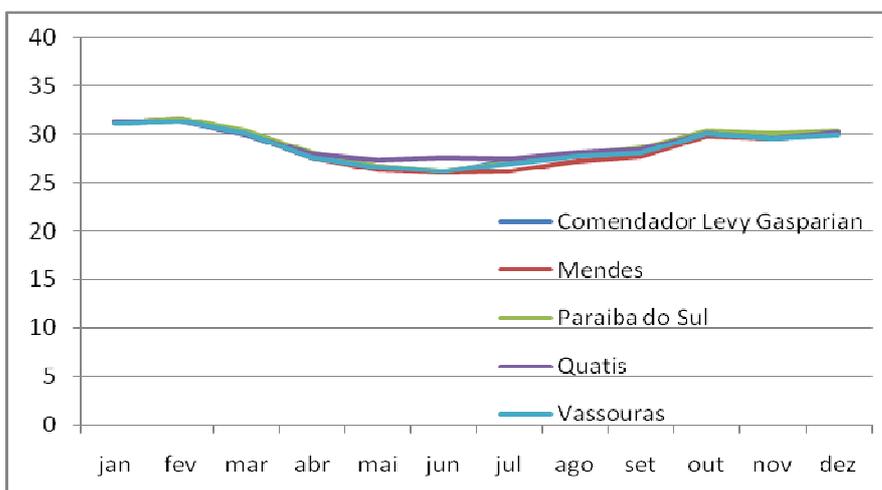


Figura 3.17 – Microclima Municípios 8 (M8) com sensação térmica média de 30.4184. Nesse microclima foram incluídos, posteriormente, os municípios de Belmiro Braga (MG), Chiador (MG), Santana do Deserto (MG), Simão Pereira (MG).

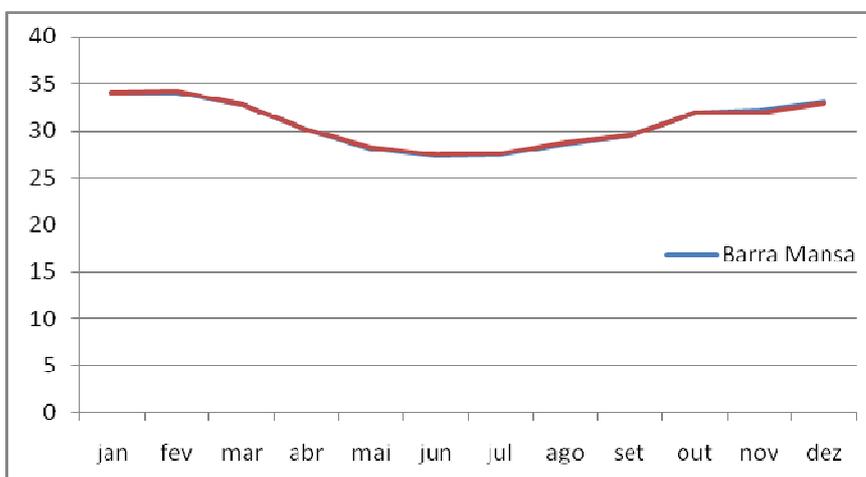


Figura 3.18 – Microclima Municípios 9 (M9) com sensação térmica média de 32.98359

Para dividir as sensações térmicas em microclima, não se preocupou em localizar os outliers.

### 3.2.2 Série Econômico e Financeira

As séries abaixo foram obtidas nos sites do IBGE, BCB e DIEESE, porém os dados mensais da série de PIB foram adquiridos diretamente com a Concessionária, pois esta já faz o seu tratamento. Essas séries não foram tratadas e nem divididas em microclimas.

- PIB: PIB - Mar/02 a Dez/09 (fornecido pela Light)
- API: Fabricação Alimentos (RJ) (do CNAE da indústria. Fonte: IBGE) – Mar/02 a Dez/09
- BEP: Fabricação de Borracha e plástico (RJ) (do CNAE da indústria. Fonte: IBGE) – Mar/02 a Dez/09
- BPF: Fabricação Bebidas (RJ) (do CNAE da indústria. Fonte: IBGE) – Mar/02 a Dez/09
- FPQF: Fabricação de produtos químicos + farmacêuticos (RJ) (do CNAE da indústria. Fonte: IBGE) – Mar/02 a Dez/09
- ITSA: Indústria da transformação sem ajuste sazonal (RJ) (Fonte: IBGE) - Mar/02 a Dez/09
- IOPC :Inadimplência em op. de crédito do sistema financeiro privado – total (BR) (Fonte: BCB-Depec) – Mar/02 a Dez/09
- IVRN : Índice de receita nominal de vendas no comércio varejista (RJ) (Fonte: IBGE) – Mar/02 a Dez/09
- OPC : Operações de crédito totais do sistema financeiro - Ao setor comercial (BR) (Fonte: BCB-Depec) - Mar/02 a Dez/09
- RTP: Rendimento mediano real do trabalho principal da Região Metropolitana do RJ (Fonte: IBGE) – Mar/02 a Dez/09
- VCV: Volume de Vendas do comercio varejista com ajuste sazonal (RJ) (Fonte: IBGE) – Mar/02 a Dez/09
- CBN: Cesta Básica Nacional - Gasto mensal – Região Metropolitana do Rio de Janeiro (Fonte: DIEESE) - Mar/02 a Dez/09
- ICVG: Índice de Custo de Vida - Índice Geral (BR)(Fonte: DIEESE) - Mar/02 a Dez/09 (Série Nova)
- OPCRF: Operações de crédito totais do sistema financeiro - A pessoas físicas (BR) (Fonte: BCB-Depec) - Mar/02 a Dez/09

- OPCRH: Operações de crédito totais do sistema financeiro – Habitacionais (BR) (Fonte: BCB-Depec) - Mar/02 a Dez/09
- OPCRural : Operações de crédito totais do sistema financeiro (BR) (Fonte: BCB-Depec) - Mar/02 a Dez/09
- RMRH: Rendimento médio Real habitual da Região Metropolitana do RJ (Fonte: IBGE) – Mar/02 a Dez/09
- OPCPP: Operações de crédito totais do sistema financeiro- Ao setor público e estadual (BR) (Fonte: BCB-Depec) - Mar/02 a Dez/09

### 3.2.3 Série de Carga

A Concessionária possui várias subestações espalhadas por sua área de concessão, sendo que cada uma delas é responsável pelo fornecimento de energia para uma determinada área. Após a definição dos microclimas, providenciou-se ao somatório das cargas das subestações que o atendiam. Algumas das subestações possuem medições de energia fornecida e em outras se chegou a esse valor somando a potência nominal dos transformadores instalados. Na figura 3.19, pode-se verificar como ficou divisão da carga da Concessionária por microclima.

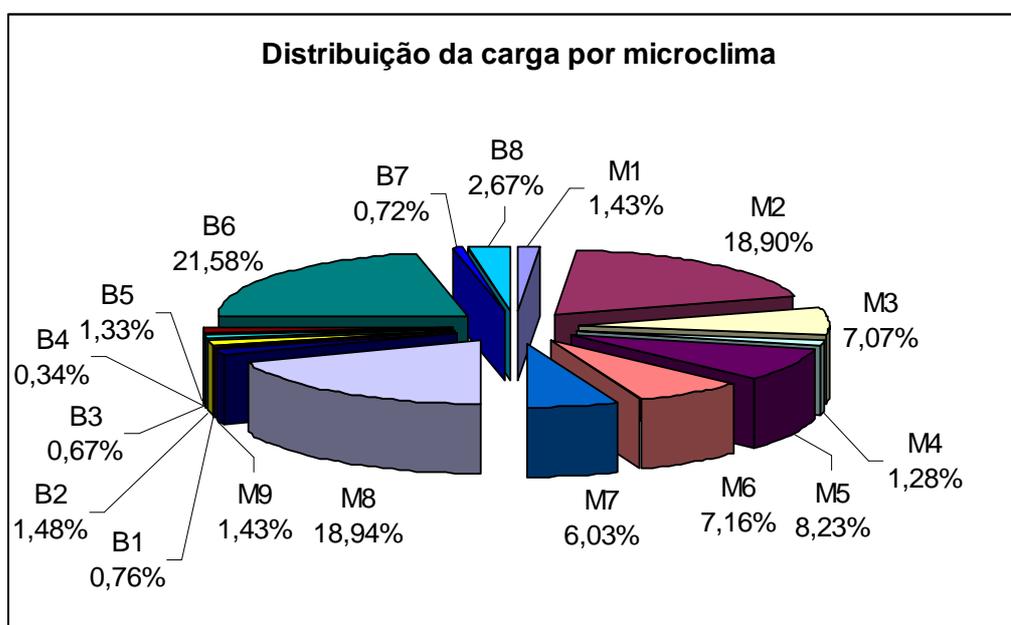


Figura 3.19: Distribuição da carga da concessionária por microclimas. Dados relativos ao ano de 2009.

Pode-se observar que a maior carga pertence ao Microclima Bairros 6 (B6). Esse microclima possui bairros que sofrem grande influência dos

fenômenos climáticos, pois estão localizados no litoral, e no verão o consumo de energia elétrica se eleva em virtude dos aparelhos de ar condicionados nas residências. Na classe comercial, tem-se um aumento desse consumo em virtude da presença de turistas na cidade.

### 3.2.4 Série de Faturamento

Primeiramente, os dados de faturamento foram divididos em classes, porém para essa previsão não foi levado em consideração o consumo do sistema de iluminação pública, pois para esses casos a Concessionária não é obrigada a instalar equipamentos de medição, logo os valores de consumo de energia elétrica e/ou demanda de potência ativa são estimados, para fins de faturamento, com base nos seguintes itens:

- Tempo de acendimento das lâmpadas (12 h para todos os municípios da área de concessão, menos para o Rio de Janeiro – Quando o logradouro público necessite de iluminação permanente, o tempo será de 24h);
- O cálculo da energia consumida pelos equipamentos auxiliares de iluminação pública, como: relés, reatores, células fotoelétricas..... Esses valores são fixados com base em critérios das normas da Associação Brasileira de Normas Técnicas (ABNT), em dados do fabricante dos equipamentos ou em ensaios realizados em laboratórios credenciados.
- Tipo e Quantidade de Lâmpadas.

Após a divisão do faturamento por classe e clima, verificou-se a necessidade de calcular o faturamento mensal, uma vez que os valores de energia faturada dos clientes considerados no sistema proposto são lidos conforme o cronograma de cada um dos 20 lotes existentes, aos quais os clientes estão associados, essas energias não se encontram restritas a um mês de referência (mês do faturamento), e sim distribuídos ao longo de quase 2 meses, como pode ser visto na figura 3.20.

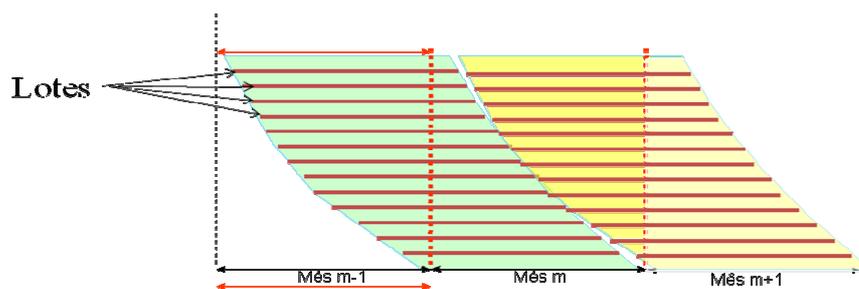


Figura 3.20 - Esquemática de leitura das energias faturadas

Deste modo, para reduzir os erros produzidos por esse tipo de medida, o sistema proposto, em uma primeira modelagem, utilizou-se do ponderador de faturamento por mês/lote/classe, considerando a média mensal dos últimos 3 anos fornecido pela Light. A partir da aplicação desse ponderador foram gerados os faturamentos por lote/mês/classe. De posse desses valores, foi feito o processo de “encaixe” do faturamento baseado nos dados de escala da leitura dos lotes (datas que indicam o início e o fim do lote e a data do final do mês de referência da energia faturada, de forma a ser aplicado, lote a lote). O cálculo do encaixe é feito somando-se as partes dos meses adjacentes que se encaixam cronologicamente no mês de referência. Assim, o faturamento do mês  $m$  é dado pela soma das partes dos faturamentos dos meses  $m$  e  $m-1$  que se encaixam, cronologicamente, no mês  $m$ . Essas partes são estimadas através de ponderadores de encaixe, cujo cálculo é feito através da carga diária total da Light.

Após o conhecimento do faturamento “encaixado” os valores dos lotes foram somados por classe de consumo, resultando em um dado definido por classe/mês de referência. Ao final, os resultados obtidos para previsão de faturamento passaram por uma função para serem “desencaixados”. Para isso, foram criados ponderadores a partir dos valores dos lotes encaixados.

No entanto, os resultados de previsão obtidos a partir desse processo não foram considerados satisfatórios. Considerou-se que os erros devido ao encaixe e ao desencaixe poderiam ser os responsáveis pelo mau desempenho. No processo de previsão acredita-se que os erros aumentavam ainda mais, já que para os meses a serem previstos, não se conhece antecipadamente as cargas, que devem ser utilizadas para o cálculo do encaixe, tornando esse cálculo ainda mais vulnerável.

Assim, outra metodologia foi proposta, para todos os microclimas/classes existentes, onde não se considerou o faturamento encaixado. Como os resultados obtidos foram melhores do que os anteriores, o sistema final desenvolvido na ferramenta computacional não considerou o encaixe na sua metodologia.

Os dados históricos (2002 a 2009) de energia faturada são informados na ferramenta por classe/mês/bairro. Como a previsão de energia faturada é mensal, por microclima e classe, foram somados os valores de faturamento de todos os bairros que pertencem a um dado microclima conforme a expressão abaixo:

$$Faturamento\_Microclima_{i,m,c} = \sum_{b=1}^B faturamento_{b,i,m,c} \quad (\text{eq.1})$$

Onde:

$Faturamento\_Microclima_{i,m,c}$  é o valor total de energia faturada da classe  $c$  em um dado microclima  $i$  no mês  $m$ ;

$faturamento_{b,i,m,c}$  é o faturamento do bairro  $b$  da classe  $c$  do microclima  $i$  do mês  $m$ ;

$b$  é um dos bairros do microclima  $i$  da classe  $c$  sendo considerada;

$B$  é o número total de bairros da classe  $c$  do microclima  $i$ .

Dessa forma, foram desenvolvidos 119 modelos (7 classes de faturamento x 17 microclimas = 119 modelos) que consideraram aspectos climáticos, econômico-financeiros, e outras informações que estão atreladas ao consumo de energia elétrica para a previsão de 15 meses a frente, mês a mês. Assim, foram avaliados 112000 por classe, ou 13328000 de modelos distintos considerando todas as classes.

Cabe ressaltar que os dados fornecidos pela concessionária não contemplam os consumos registrados pelos clientes livres, os enquadrados no Subgrupo A2 (tensão de fornecimento de 138 kV) e o consumo registrado a partir de serviços oferecidos pela Concessionária, como a energia plus.

### 3.2.5 Localização dos Outliers

Frequentemente, são encontradas em observações discordantes em séries temporais. A sua presença pode facilmente afetar os procedimentos convencionais de análise. Uma das estratégias mais utilizadas no tratamento dos outliers consiste na identificação e na da localização dos mesmos como também na escolha de modelos que permitam acomodar os seus efeitos (RIBEIRO, 2002).

Neste trabalho, a identificação dos outliers foi feita por meio da média e do desvio padrão calculados para uma janela de 36 meses com superposição de 12 meses. Cada amostra contida nessas janelas foi comparada com os limites aceitáveis: 2 desvios-padrão para cima e para baixo da média da janela. A definição de outlier é dada sempre que a amostra apresentar outliers em metade ou mais do número total de pontos da janela.

Seu tratamento é feito de forma seqüencial, na ordem em que os outliers aparecem, sendo que estes são substituídos pela média das amostras vizinhas, com intervalo de vizinhança de 18 antes e depois. A diferença entre o valor original, considerado outlier e o seu novo valor é redistribuído nas próximas 3 amostras de forma igual, fazendo com que o resultado do somatório das amostras não seja alterado depois do tratamento.

### **3.2.6 Seleção de Variáveis**

Uma das etapas mais importantes na modelagem com as RNAs é a definição das variáveis de entrada, que constituirão os neurônios da camada de entrada. É interessante pensar no problema que se quer modelar e no modo como as redes funcionam, para escolher quais informações devem ser fornecidas a esses sistemas, de modo a obter o melhor desempenho (PAZ & BUENO & TUCCI, 2002)

A fim de melhorar o processo de escolha do melhor conjunto de variáveis de entrada optou-se por usar o próprio modelo da Rede Neural. Este é o modelo ideal, já que os métodos baseados na execução de testes estatísticos entre os subconjuntos das variáveis de entrada e as saídas desejadas no modelo, que existem são métodos lineares ou pseudo-não lineares, que parecem um contra-censo ao se utilizar métodos não lineares como redes neurais. Dessa forma, diversos modelos de redes neurais foram avaliados e o melhor resultado obtido indicou o melhor conjunto de variáveis de entrada. Embora esse procedimento seja mais custoso computacionalmente, ele gerou os melhores resultados, uma vez que os modelos de seleção de variáveis normalmente não captam a não linearidade que os modelos baseados em redes neurais são capazes de identificar.

### **3.2.7 Normalização**

O projeto das redes neurais artificiais consistiu em uma etapa inicial de normalização (NASCIMENTO, 2009), que tem como finalidade adaptar os dados de entrada à faixa dinâmica das funções de ativação da rede neural, neste caso,

tangente hiperbólica (na camada intermediária) e função linear (na camada de saída). O processo de normalização dos dados foi feito da seguinte forma:

$$\bar{x} = \frac{(y_{max} - y_{min}) * (x - x_{min})}{(x_{max} - x_{min})} + y_{min} \quad \text{eq. (2)}$$

Onde:

$y_{min}$ : valor mínimo definido para a normalização (padrão é -1)

$y_{max}$ : valor máximo definido para a normalização (padrão é +1)

$x_{max}$ : valor máximo da série original;

$x_{min}$ : valor mínimo da série original;

$\bar{x}$ : valor normalizado.

Nesse projeto utilizou-se a faixa de valores [-1,1] para a normalização das variáveis segundo a equação 1.

### 3.3. Arquitetura da Rede Neural Artificial Utilizada

O sistema possui 172 redes neurais MLP, sendo que 17 são para previsão de sensação térmica, 18 para prever as séries econômica-financeiras, 119 para previsão de faturamento por classe e 18 para prever carga por microclima para 15 meses à frente. O período de dados para todas essas séries compreende o período de janeiro/2002 a dezembro/2009. Os últimos 15 meses das séries de dados foram reservados para validar os modelos, e o restante da base de dados foi usada para ajuste dos modelos de previsão desenvolvidos.

Cada previsão possui uma rede com uma arquitetura própria, porém em todas elas o algoritmo utilizado foi o back propagation com os neurônios variando de 1 a 10. Nos próximos itens as redes MLP utilizadas em cada previsão poderão ser vistas.

#### 3.3.1. Rede Neural utilizada na previsão da série de sensação térmica

Definidos os microclimas, os dados climáticos são então analisados para a previsão dos 15 meses à frente através de uma rede neural MLP (Multi-Layer Perceptron) em processo *multistep* (previsão de múltiplos passos à frente).

Dessa forma, foram investigados os modelos de previsão *multistep* para 15 meses à frente, um modelo para cada um dos 17 microclimas.

As redes neurais *multistep* são modelos de previsão através de realimentação, cujos valores previstos são utilizados recursivamente para estimar os meses subsequentes.

Várias abordagens são possíveis, sendo adotada (para todos os modelos de previsão) neste projeto aquela a seguir descrita.

Através da calibragem de uma função não linear de domínio multidimensional, a rede é capaz de prever um mês a frente. A fim de ajustar este modelo para a previsão dos 15 passos a frente, os valores previstos são utilizados como entrada para a previsão do próximo valor, isto é, para a previsão do mês seguinte. Estima-se o erro da rede neural para a previsão dos valores conhecidos como conjunto de validação, contendo 15 valores da série a ser prevista. O erro obtido com a previsão dos valores desse conjunto é utilizado para a escolha da topologia da rede. A que apresentar o menor erro é a rede escolhida. Ressalta-se, entretanto, que o conjunto de validação não está presente no treinamento da função de regressão: a rede escolhida é, então, genérica o suficiente para lidar com valores ou trechos de série desconhecidos. Daí seu uso para previsões com longos horizontes (neste caso, 15 meses).

A equação abaixo (equação 3) apresenta a função de regressão genérica dos modelos de previsão das séries temporais deste projeto.

$$y_s = \sum_{j=1}^h w_{sj} \cdot \tanh\left(\sum_{i=1}^n x_i \cdot w_{ji} + \theta_j\right) + \theta_s \quad (\text{eq.3})$$

Onde:

- $y_s$  = saída do modelo de previsão;
- $w_{sj}$  = peso sináptico  $j$  do processador  $s$ ;
- $x_i$  =  $i$ -ésimo atributo de entrada do modelo de previsão;
- $\theta_j$  = parâmetro livre do neurônio  $j$ ;
- $h$  = número de neurônios na camada escondida;
- $n$  = número de atributos de entrada.

O vetor  $x$  de entrada é formado por diversos valores adquiridos a partir da série a se prever, a fim de fornecer informação relevante para a rede prever corretamente os valores futuros da série em questão. Além disso, pode-se aplicar pré-processamento aos dados de entrada, antes de gerar o vetor  $x$ .

Foram investigadas, como possíveis variáveis para o modelo de previsão mensal de sensação térmica, considerando a série mensal para a previsão um passo a frente, as seguintes variáveis de entrada:

- Previsão da sensação térmica um passo a frente ( $S_M$ );
- 4 últimos valores da série em relação ao valor que se deseja prever ( $S_{M-4}$ ,  $S_{M-3}$ ,  $S_{M-2}$ ,  $S_{M-1}$ );
- Valor correspondente a um ano antes em relação ao mês que se deseja prever ( $S_{M-12}$ );
- Diferença entre o último e o penúltimo valor em relação ao mês que se deseja prever ( $DM1=S_{M-1}-S_{M-2}$ );
- Diferença entre o último e o antepenúltimo valor em relação ao mês que se deseja prever ( $DM3=S_{M-1}-S_{M-3}$ );
- Diferença entre o último e o valor correspondente a 7 meses em relação ao mês que se deseja prever ( $DM7=S_{M-1}-S_{M-7}$ );
- Média dos últimos valores em relação ao mês que se deseja prever ( $MM2=(S_{M-2}+S_{M-1})/2$ );
- Média dos últimos valores em relação ao mês que se deseja prever ( $MM3=(S_{M-3}+S_{M-2}+S_{M-1})/3$ );
- Média dos últimos 6 meses anteriores ao que se deseja prever (MM06)
- Média dos últimos 12 meses anteriores ao que se deseja prever (MM12)

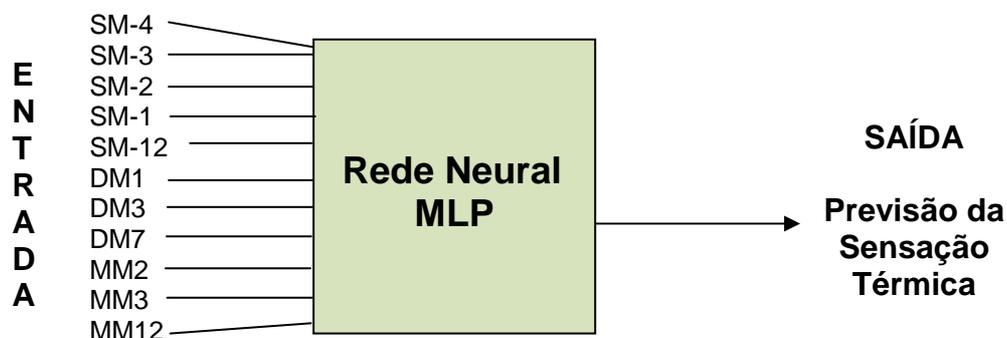


Figura 3.21: Rede Neural MLP utilizada na previsão da série de sensação térmica

Cada microclima possui uma rede própria, com diferentes vetores de entradas e quantidade de neurônios na camada escondida, que foi escolhida com base no menor erro encontrado.

### 3.3.2. Rede Neural utilizada na previsão da série econômico financeira

A introdução de séries exógenas no modelo de previsão de energia faturada ou carga visa melhorar o resultado através de informações financeiras, pois se sabe que o consumo de energia elétrica está fortemente atrelado ao desempenho financeiro dos diversos setores da economia ou mesmo ao poder aquisitivo dos seus clientes. Assim, foram pesquisadas e avaliadas séries de cunho econômico-financeiro que tivessem relação com séries de energia faturada e carga. Nesta previsão foram desenhadas 18 redes diferentes, isto é, uma rede para cada variável econômico-financeira. Essas variáveis foram escolhidas de acordo com sua relevância para cada classe de faturamento e são descritas na figura 3.22.

Industrial	Comercial	Residencial	Rural	Poder Público
•PIB	•PIB	•PIB	•PIB	•PIB
•API	•IOPC	•CBN	•OPCRural	•OPCPP
•BEP	•IVRN	•ICVG	•RMRH	
•BPF	•OPC	•OPCRF	Consumo Próprio	
•FRQF	•RMRH	•OPCRH	•PIB	
•ITSA	•RTP	•RMRH	Serviço Público	
	•VCV		•PIB	

Figura 3.22: Divisão das séries econômica e financeira de acordo com as classes de faturamento.

Para esta previsão variou-se o número de neurônio da camada escondida entre 1 e 10. As funções de ativação e saída utilizadas foram, respectivamente, tangente hiperbólica e linear. Também nesse caso, utilizou-se, além da respectiva série histórica, a seguinte combinação entre seus dados (figura 3.23).

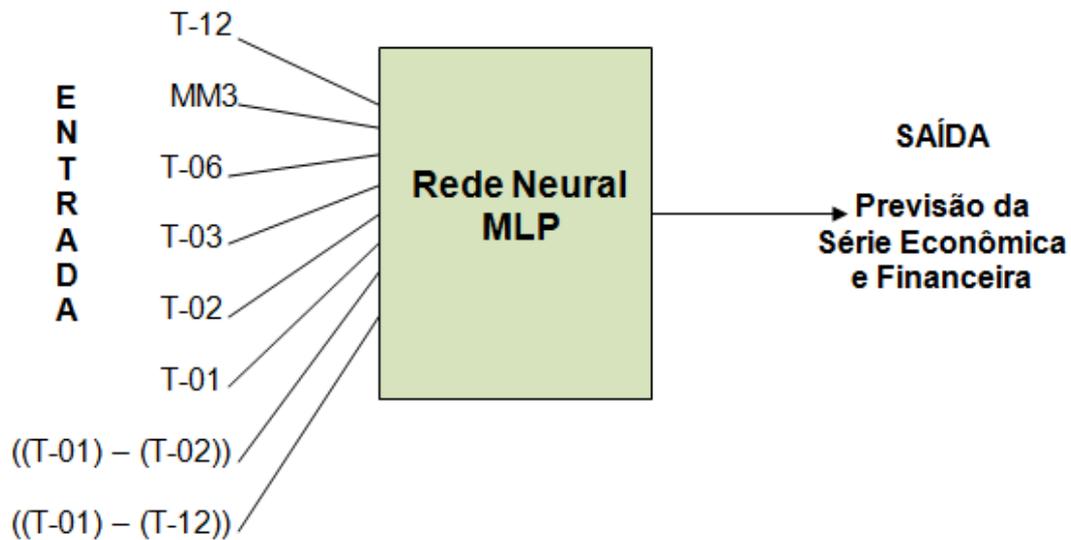


Figura 3.23: Rede Neural MLP utilizada na previsão das séries econômica e financeira

Onde:

- 12º mês anterior em relação ao mês (T) que se deseja prever (T-12)
- Média Móvel dos 3 últimos meses em relação ao mês (T) que se deseja prever (MM3)
- 6º mês anterior em relação ao mês (T) que se deseja prever (T-06)
- 3º mês anterior em relação ao mês (T) que se deseja prever (T-03)
- 2º mês anterior em relação ao mês (T) que se deseja prever (T-02)
- 1º mês anterior em relação ao mês (T) que se deseja prever (T-01)
- Diferença entre T-01 e T-02 ((T-01)-(T-02)) em relação ao mês (T) que se deseja prever
- Diferença entre T-01 e T-12 ((T-01)-(T-12)) em relação ao mês (T) que se deseja prever.

### 3.3.3. Fatores Temporais

O fator temporal utilizado neste trabalho foi a relação percentual do número de dias úteis sobre o total de dias no mês, isto é, dividiu-se o número de dias úteis pelo número total de dias do mês de referência. Para diminuir a

quantidade de valores dessa entrada, optou-se em categorizá-los, de acordo, com o percentual de dias úteis, conforme a tabela 3.3.

Valor	Percentual correspondente
-2	$\leq 65,5\%$
-1	$\geq 65,5\%$ e $\leq 67,7\%$
0	$\geq 67,7\%$ e $\leq 69,8\%$
1	$\geq 69,8\%$ e $\leq 72,0\%$
2	$\geq 72\%$

Tabela 3.3: Categorização ou discretização dos valores de entrada da variável “dias úteis”.

Após a previsão da sensação térmica e da série econômico e financeira e a categorização dos dias úteis (fatores temporais), a primeira parte do sistema SIPEE está pronto, já que esses dados, fazem parte do conjunto de entrada para a rede neural MLP que irá prever a carga e faturamento da Light S.E.S.A. (Figura 3.24).

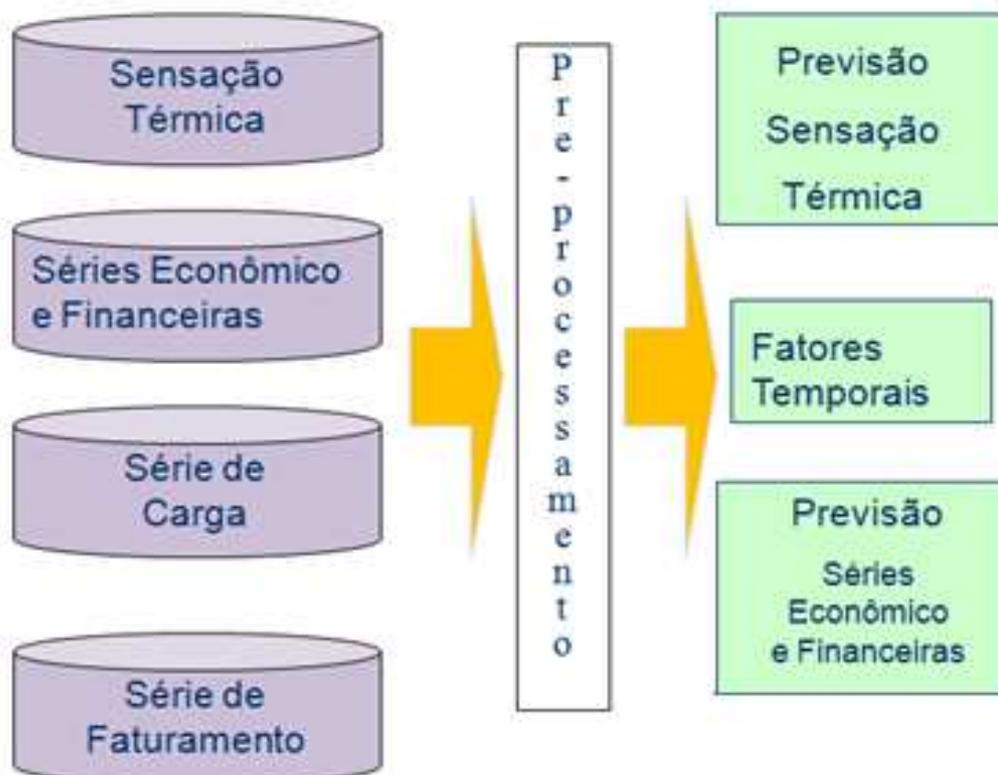


Figura 3.24: 1ª parte do SIPEE

### 3.3.4. Rede Neural utilizada na previsão da série de faturamento

Nesta previsão foram utilizadas 119 redes (7 classes de faturamento x 17 microclimas = 119 redes). A figura 3.25 apresenta o diagrama da rede utilizada nessa previsão.

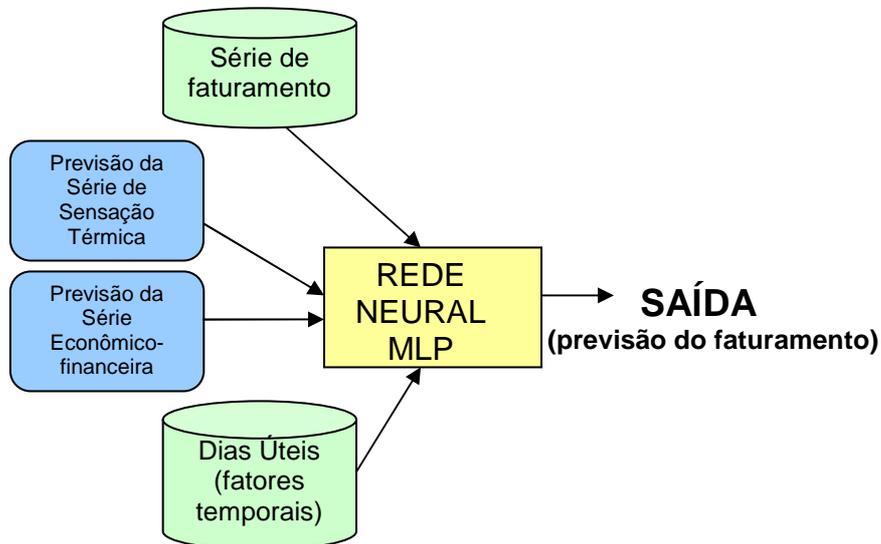


Figura 3.25: Diagrama da rede utilizada na previsão de faturamento.

Como ocorreu com as previsões anteriores, as funções de ativação e saída utilizadas foram, respectivamente, tangente hiperbólica e linear. Os dados da série de faturamento foram combinados da seguinte maneira:

- Previsão da sensação térmica um passo a frente ( $S_M$ );
- 4 últimos valores da série em relação ao valor que se deseja prever ( $S_{M-4}$ ,  $S_{M-3}$ ,  $S_{M-2}$ ,  $S_{M-1}$ );
- Valor correspondente a um ano antes em relação ao mês que se deseja prever ( $S_{M-12}$ );
- Diferença entre o último e o penúltimo valor em relação ao mês que se deseja prever ( $DM1=S_{M-1}-S_{M-2}$ );
- Diferença entre o último e o antepenúltimo valor em relação ao mês que se deseja prever ( $DM3=S_{M-1}-S_{M-3}$ );
- Diferença entre o último e o valor correspondente a 7 meses em relação ao mês que se deseja prever ( $DM7=S_{M-1}-S_{M-7}$ );
- Média dos últimos valores em relação ao mês que se deseja prever ( $MM2=(S_{M-2} + S_{M-1})/2$ );
- Média dos últimos valores em relação ao mês que se deseja prever ( $MM3=(S_{M-3} + S_{M-2} + S_{M-1})/3$ );

- Média dos últimos 6 meses anteriores ao que se deseja prever (MM6)
- Média dos últimos 12 meses anteriores ao que se deseja prever (MM12)

### 3.3.5. Rede Neural utilizada na previsão da série de carga por microclima

Para esta previsão utilizou-se uma rede neural MLP para cada microclima, totalizando 17 redes. Analisando os erros encontrados, foram escolhidas as redes, que podem apresentar diferentes entradas e quantidades de neurônios na camada escondida. O vetor de entrada foi composto com os atributos abaixo e como nas previsões anteriores a série de carga foi trabalhada de modo a formar novos atributos de entrada:

- Previsão da sensação térmica um passo a frente ( $S_M$ );
- 4 últimos valores da série em relação ao valor que se deseja prever ( $S_{M-4}$ ,  $S_{M-3}$ ,  $S_{M-2}$ ,  $S_{M-1}$ );
- Valor correspondente a um ano antes em relação ao mês que se deseja prever ( $S_{M-12}$ );
- Diferença entre o último e o penúltimo valor em relação ao mês que se deseja prever ( $DM1=S_{M-1}-S_{M-2}$ );
- Diferença entre o último e o antepenúltimo valor em relação ao mês que se deseja prever ( $DM3=S_{M-1}-S_{M-3}$ );
- Diferença entre o último e o valor correspondente a 7 meses em relação ao mês que se deseja prever ( $DM7=S_{M-1}-S_{M-7}$ );
- Média dos últimos valores em relação ao mês que se deseja prever ( $MM2=(S_{M-2} + S_{M-1})/2$ );
- Média dos últimos valores em relação ao mês que se deseja prever ( $MM3=(S_{M-3} + S_{M-2} + S_{M-1})/3$ );
- Média dos últimos 6 meses anteriores ao que se deseja prever (MM6)
- Média dos últimos 12 meses anteriores ao que se deseja prever (MM12)
- Fatores temporais;
- Previsão da série de sensação térmica;
- Previsão da série de PIB;

- Previsão da série de Rendimento médio real habitual da Região Metropolitana do RJ;
- Previsão da série Fabricação de Alimentos (RJ);
- Previsão da série de Operações de créditos totais do sistema financeiro (BR);
- Previsão da série de Operações de créditos totais do sistema financeiro (BR).

Deve-se observar que os 5(cinco) últimos itens pertencem à previsão das séries econômico-financeiras.

As métricas de erro utilizadas em todas as redes utilizadas estão a seguir:

- MAPE – média dos erros percentuais absolutos;

$$MAPE = \frac{\sum_{t=1}^n \left| e_t / Y_t * 100 \right|}{n} \quad (\text{eq.4})$$

Onde:

$e_t$  - Erro apresentado (diferença entre a saída encontrada e a real);

$Y_t$  - Valor da saída real;

$n$  - Número de dados.

- RMSE – Raiz do erro quadrático médio.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n e_t^2}{n}} \quad (\text{eq.5})$$