DANILO SINKITI GASTALDELLO

**Desenvolvimento de metodologia para previsão da demanda de energia elétrica
residencial considerando aspectos socioeconômicos e ferramentas computacionais
inteligentes**

São Paulo
(2017)

DANILO SINKITI GASTALDELLO

**Desenvolvimento de metodologia para previsão da demanda de energia elétrica
residencial considerando aspectos socioeconômicos e ferramentas computacionais
inteligentes**

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção do
título de Doutor em Ciências

Orientador: Prof. Dr. André Nunes de Souza.

São Paulo
(2017)

DANILO SINKITI GASTALDELLO

**Desenvolvimento de metodologia para previsão da demanda de energia elétrica
residencial considerando aspectos socioeconômicos e ferramentas computacionais
inteligentes**

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção do
título de Doutor em Ciências

Área de Concentração: Engenharia Elétrica /
Sistemas de Potência

Orientador: Prof. Dr. André Nunes de Souza

São Paulo
(2017)

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 30 de JUNHO de 2017

Assinatura do autor: Daniilo Sinkiti Gastaldello

Assinatura do orientador: [Assinatura]

Catálogo-na-publicação

Sinkiti Gastaldello, Daniilo

Desenvolvimento de metodologia para previsão da demanda de energia elétrica residencial considerando aspectos socioeconômicos e ferramentas computacionais inteligentes / D. Sinkiti Gastaldello – versão corr. – São Paulo, 2017.

147 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Energia e Automação Elétricas.

1.Distribuição de energia elétrica 2.Consumo de energia elétrica
3.Inteligência artificial 4.reconhecimento de padrões I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Energia e Automação Elétricas II.t.

Dedico este trabalho com todo o carinho a minha esposa, aos meus pais e aos meus familiares que me incentivaram e me apoiaram na concretização desta conquista.

AGRADECIMENTOS

Primeiramente, agradeço a Deus por ser a base de nossas conquistas, e por estar sempre nos abençoando e nos dando energia para seguir em frente;

Agradecimentos especiais ao orientador, amigo, professor e membro da família, André Nunes de Souza, que esteve ao meu lado desde a graduação, passando pelo mestrado e agora no doutorado, me auxiliando, me ensinando e também me repreendendo, com um único objetivo, que eu fosse melhor a cada dia;

Em memória, agradeço a Prof^a. Maria Goretti Zago Nunes de Souza, que sempre me apoiou, mas que infelizmente não está hoje presente para ver a conclusão deste trabalho, mas que fez parte dele com toda certeza;

Aos meus pais, Ayrton e Luiza, aos meus irmãos, Daniel e Juliana, à família da minha esposa, Alcides, Elza, Natália e Paulo, total gratidão por sempre me apoiarem, e por entenderem a importância desta conquista;

Aos meus amigos do Laboratório de Sistemas de Potência e Técnicas Inteligentes (LSISPOTI), em especial, o Marcio Gotti, o Haroldo Amaral, o Caio Ramos, a Karila Santos e o João Paulo Papa, sem esquecer os demais que passaram durante todos os anos de pesquisa e estudo no laboratório, pela companhia, pela ajuda, pela amizade nos momentos bons e ruins para conseguir cumprir todos os cronogramas e desafios;

A CAPES, pelo incentivo e apoio financeiro durante o doutorado;

Ao Prof. Dr. Luiz Lebensztajn, pelo apoio e incentivo para conclusão deste trabalho;

Em especial, gostaria de agradecer muito à minha esposa Cássia Hiromi Yamamoto Gastaldello, pela paciência, pelo apoio, pelas noites em claro analisando os ensaios e escutando meus comentários sobre os resultados alcançados, por tudo que ela representa em minha vida e por estar ao meu lado nessa fase tão difícil, mas muito gratificante. Aqui também cito nossos filhos, os ‘birds’, o Toddy, o Ninho e a Pandora, que sempre nos recebem felizes em casa; e

A todos que colaboraram de forma direta ou indireta, na execução desta pesquisa.

Não confunda derrotas com fracasso nem vitórias com sucesso. Na vida de um campeão sempre haverá algumas derrotas, assim como na vida de um perdedor sempre haverá vitórias. A diferença é que, enquanto os campeões crescem nas derrotas, os perdedores se acomodam nas vitórias.

(Roberto Shinyashiki)

RESUMO

O aumento da demanda por energia registrado nos últimos anos preocupa, pois a construção de novas fontes geradoras é barrada, muitas vezes, por restrições ambientais. Assim, o governo e as empresas de energias estão investindo em um melhor planejamento do sistema. No entanto, para haver uma proposta mais consistente para os consumidores residenciais se faz necessário conhecer melhor o perfil de cada consumidor, que é uma tarefa um tanto quanto difícil, visto que cada consumidor possui o livre arbítrio para consumir a energia de acordo com o conforto que ele deseja, de acordo com seus padrões econômicos e conforme aspectos culturais e sociais do ambiente em que ele vive. Neste contexto, a proposta desta tese foi analisar os impactos que os aspectos socioeconômicos tinham sobre o consumo de energia da classe residencial, sendo desenvolvido um algoritmo que gera curvas de carga virtuais baseadas em dados estatísticos do IBGE e da ANEEL. A partir dados de curvas virtuais, as ferramentas computacionais inteligentes, mais especificamente, as Redes Bayesianas e a Floresta de Caminhos Ótimos, foram treinadas com intuito de avaliar a possibilidade de criação de perfis e classificação dos consumidores e de suas características. Os resultados alcançados demonstram que a consideração dos aspectos socioeconômicos em avaliação de curvas de carga são pertinentes e que devem fazer parte do planejamento do sistema. Outra constatação é que as ferramentas computacionais inteligentes estudadas podem ser exploradas para auxiliar na previsão de consumo e criação de padrões e perfis dos consumidores.

Palavras-Chave: Demanda Residencial, Curvas de Carga, Previsão de Consumo, Consumo de Energia.

ABSTRACT

The need for energy has increased in the past years, thus requiring the design of new power plants. However, the project of such new constructions has been considerably neglected, mainly due to environment constraints. However, the whole government and companies are now focusing on a better management of the national energy grid. Despite that new policy, there is a need for a better knowledge concerning the user's behavior in order to present proposals that really take into account the consumers, since each of them has the freedom to use the energy the way he wants to, as well as according to his socioeconomic habits. In this context, this thesis proposes to analyze the socioeconomic impacts of the energy consumption concerning residential consumers, being also developed an algorithm that generates virtual load curves based on statistical data from both IBGE and ANEEL. With that data on hand, the intelligent tools, e.g., Bayesian Networks and Optimum-Path Forest, were trained aiming at evaluating the possibility to create profiles for the further identification of their classes according to that information. The results obtained highlighted the importance of the socioeconomic information when evaluating the load curves, which should be part of the whole system. Another conclusion concerns the intelligent tools, which can be further used for consumer forecasting, as well as to create patterns related to the consumers' profiles.

Keywords: Residential Demand, Load Curves, Consumption Forecast, Energy Consumption.

LISTA DE FIGURAS

Figura 1 – Curva de Carga Padrão.....	31
Figura 2 – Consumo de energia elétrica por setor	32
Figura 3 – Consumo final na carga residencial	32
Figura 4 – Curva de carga de consumidor residencial.....	33
Figura 5 – Representação do consumo de energia elétrica por equipamento para o setor comercial	34
Figura 6 – Curva de carga de um consumidor comercial	34
Figura 7 – Representação do consumo de energia elétrica na indústria por setor.....	35
Figura 8 – Curvas de carga do setor industrial	35
Figura 9 – Representação da participação de eletrodomésticos no consumo residencial da região Norte	39
Figura 10 – Curva de carga média de consumidores residenciais da região Norte	39
Figura 11 – Participação dos eletrodomésticos no consumo residencial da região Nordeste ..	40
Figura 12 – Curva de carga média de consumidores residenciais na região Nordeste.....	41
Figura 13 – Participação dos eletrodomésticos no consumo residencial da região Centro-Oeste	41
Figura 14 – Curva de carga diária média de um consumidor residencial da região Centro-Oeste	42
Figura 15 – Participação dos eletrodomésticos no consumo residencial da região Sudeste	43
Figura 16 – Curva de carga diária média de consumidores residenciais da região Sudeste....	44
Figura 17 – Participação dos eletrodomésticos no consumo residencial da região Sul	45
Figura 18 – Curva de carga média de um consumidor residencial da região Sul.....	45
Figura 19 – Gráfico sobre OPF	51
Figura 20 - Exemplo de Rede Bayesiana.....	53
Figura 21 – Estrutura da Rede Neive Bayes.....	54
Figura 22 – Algoritmo de Construção de Curva de Carga, 1ºPasso: Característica da Família	57
Figura 23 – Algoritmo de Construção de Curva de Carga, 2ºPasso: Posse de Equipamento ..	58
Figura 24 – Algoritmo de Construção de Curva de Carga, 3ºPasso: Curva de Carga de cada Aparelho	59
Figura 25 – Algoritmo de Construção de Curva de Carga, 4ºPasso: Curva de Carga da Residência.....	59
Figura 26 – Curvas de carga residencial (classe C, duas pessoas) geradas pelo algoritmo – Norte	62
Figura 27 – Curvas de carga residencial (classe C, duas pessoas) geradas pelo algoritmo – Nordeste.....	62

Figura 28 – Curvas de carga residencial (classe C, duas pessoas) geradas pelo algoritmo – Centro-Oeste	62
Figura 29 – Curvas de carga residencial (classe C, duas pessoas) geradas pelo algoritmo – Sudeste.....	63
Figura 30 – Curvas de carga residencial (classe C, duas pessoas) geradas pelo algoritmo – Sul	63
Figura 31 – Comparação entre as curvas médias do algoritmo e do Procel-PNAD – Região Norte	66
Figura 32 – Comparação entre as curvas médias do algoritmo e do Procel-PNAD – Região Nordeste.....	67
Figura 33 – Comparação entre as curvas médias do algoritmo e do Procel-PNAD – Região Centro-Oeste	67
Figura 34 – Comparação entre as curvas médias do algoritmo e do Procel-PNAD – Região Sudeste.....	68
Figura 35 – Comparação entre as curvas médias do algoritmo e do Procel-PNAD – Região Sul.....	68
Figura 36 – Comparação entre as curva medida em campo e curvas geradas pelo algoritmo – Estudo 1	70
Figura 37 – Comparação entre as curva medida em campo e curvas geradas pelo algoritmo – Estudo 2	70
Figura 38 – Análise estatística da influência da variação do número de pessoas sobre o consumo – Norte.....	73
Figura 39 – Análise estatística da influência da variação do número de pessoas sobre o consumo – Nordeste	75
Figura 40 – Análise estatística da influência da variação do número de pessoas sobre o consumo – Centro-Oeste	76
Figura 41 – Análise estatística da influência da variação do número de pessoas sobre o consumo – Sudeste	77
Figura 42 – Análise estatística da influência da variação do número de pessoas sobre o consumo – Sul	78
Figura 43 – Gráfico de BoxPlot da Classe Econômica C para a Variação de Número de Pessoas.....	80
Figura 44 – Análise estatística da influência da variação da classe econômica sobre o consumo – Norte.....	81
Figura 45 – Análise estatística da influência da variação da classe econômica sobre o consumo – Nordeste	83
Figura 46 – Análise estatística da influência da variação da classe econômica sobre o consumo – Centro-Oeste	84
Figura 47 – Análise estatística da influência da variação da classe econômica sobre o consumo – Sudeste	85
Figura 48 – Análise estatística da influência da variação da classe econômica sobre o consumo – Sul	86

Figura 49 – Gráfico de BoxPlot do Número de Pessoas 3 para a Variação da Classe Econômica	87
Figura 50 – Análise estatística da influência da variação do tamanho da residência sobre o consumo – Norte.....	90
Figura 51 – Análise estatística da influência da variação do tamanho da residência sobre o consumo – Nordeste	91
Figura 52 – Análise estatística da influência da variação do tamanho da residência sobre o consumo – Centro-Oeste	92
Figura 53 – Análise estatística da influência da variação do tamanho da residência sobre o consumo – Sudeste	93
Figura 54 – Análise estatística da influência da variação do tamanho da residência sobre o consumo – Sul	94
Figura 55 – Gráfico de BoxPlot do Número de Pessoas 3 para a Variação do Tamanho da Residência.....	96
Figura 56 – Análise estatística da influência da variação da faixa etária sobre o consumo – Norte	97
Figura 57 – Análise estatística da influência da variação da faixa etária sobre o consumo – Nordeste.....	98
Figura 58 – Análise estatística da influência da variação da faixa etária sobre o consumo – Centro-Oeste	99
Figura 59 – Análise estatística da influência da variação da faixa etária sobre o consumo – Sudeste.....	99
Figura 60 – Análise estatística da influência da variação da faixa etária sobre o consumo – Sul	100
Figura 61 – Gráfico de BoxPlot do Número de Pessoas 2 para a Variação da Faixa Etária ..	102
Figura 62 – Exemplo de matriz Confusão – Classificação de animais	105
Figura 63 – Representação gráfica do ensaio de classificação de classe de consumo.	109
Figura 64 – Representação gráfica do ensaio de classificação da classe econômica.	117
Figura 65 – Representação gráfica do ensaio de classificação do tamanho da residência.	121
Figura 66 – Representação gráfica do ensaio de classificação da composição/faixa-etária...	126

LISTA DE TABELAS

Tabela 1 – Faixas para Classificação Social – Critério Brasil.....	36
Tabela 2 – Classes Sociais – Critério IBGE	36
Tabela 3 – Classificação dos Perfis de Consumo de Consumidor Residencial.....	37
Tabela 4 – Divisão da população por região e faixa etária no Brasil.	38
Tabela 5 – Percentual de Número de Habitantes por Residência.....	64
Tabela 6 – Percentual por Faixa Etária dos Habitantes da Residência.....	64
Tabela 7 – Erro percentual entre as Médias do Algoritmo e as Médias Procel.....	66
Tabela 8 – Análise Comparativa entre curvas medidas e geradas pelo algoritmo – ESTUDO 1	69
Tabela 9 – Análise Comparativa entre curvas medidas e geradas pelo algoritmo – ESTUDO 2	69
Tabela 10 – Divisão por faixa de consumo proposto	72
Tabela 11 – Consumo médio na avaliação da variação do Número de Integrantes - Norte.....	74
Tabela 12 – Consumo médio na avaliação da variação do Número de Integrantes - Nordeste.....	75
Tabela 13 – Consumo médio na avaliação da variação do Número de Integrantes – Centro- Oeste	76
Tabela 14 – Consumo médio na avaliação da variação do Número de Integrantes – Sudeste.....	78
Tabela 15 – Consumo médio na avaliação da variação do Número de Integrantes – Sul.....	79
Tabela 16 – Consumo médio na avaliação da variação da Classe Econômica – Norte	82
Tabela 17 – Consumo médio na avaliação da variação da Classe Econômica – Nordeste	83
Tabela 18 – Consumo médio na avaliação da variação da Classe Econômica – Centro-Oeste.....	85
Tabela 19 – Consumo médio na avaliação da variação da Classe Econômica – Sudeste	86
Tabela 20 – Consumo médio na avaliação da variação da Classe Econômica – Sul	87
Tabela 21 – Configurações de Tamanhos de Residência	88
Tabela 22 – Resumo da média de consumo do estudo de variação do tamanho da residência	95
Tabela 23 – Média de consumo de cada configuração da análise de faixa etária	101
Tabela 24 – Média da acurácia para estudo de classificação da classe de consumo	110
Tabela 25 – Matriz de Classificação para estudo de classificação da classe de consumo - Sudeste.....	113
Tabela 26 – Avaliação das Ferramentas Inteligentes para Classificação da Classe de Consumo - Sudeste	115
Tabela 27 – Média da acurácia para estudo de classificação da característica classe econômica	117
Tabela 28 – Matriz de Classificação para estudo de classificação da característica classe econômica - Sudeste	119
Tabela 29 – Avaliação das Ferramentas Inteligentes para Classificação da Classe Econômica - Sudeste.....	120

Tabela 30 – Média da acurácia para estudo de classificação da característica tamanho da residência	122
Tabela 31 – Matriz de Classificação para estudo de classificação da característica tamanho da residência - Sudeste	124
Tabela 32 – Avaliação das Ferramentas Inteligentes para Classificação da Tamanho da Residência - Sudeste	125
Tabela 33 – Média da acurácia para estudo de classificação da característica faixa-etária ...	126
Tabela 34 – Matriz de Classificação para estudo de classificação da característica faixa-etária - Sudeste	128

LISTA DE ABREVIATURAS E SIGLAS

ABEP	Associação Brasileira de Empresas de Pesquisa
AIE	Agência Internacional de Energia
BN	<i>Bayesians Networks</i>
CCEB	Critério de Classificação Econômica Brasil
CEB	Companhia Energética de Brasília
CEEE	Companhia Estadual de Energia Elétrica – Rio Grande do Sul
CELESC	Centrais Elétricas de Santa Catarina
CELG	Companhia Energética de Goiás
CELPA	Centrais Elétricas do Pará
CELPE	Companhia Energética de Pernambuco
CEMAR	Companhia Energética do Maranhão
CEMAT	Centrais Elétricas Matogrossenses
CEMIG	Companhia Energética de Minas Gerais
CERON	Centrais Elétricas de Rondônia
COELBA	Companhia de Eletricidade do Estado da Bahia
COELCE	Companhia Energética do Ceará
COPEL	Companhia Paranaense de Energia
COSERN	Companhia Energética do Rio Grande do Norte
CPFL	Companhia Paulista de Força e Luz
OCDE	Organização para Cooperação e Desenvolvimento Econômico
OPF	<i>Optimun-Path Forest</i>
PDE	Plano Decenal de Expansão de Energia
PNAD	Pesquisa Nacional por Amostras de Domicílios
PPH	Pesquisa de Posse e Hábitos de Consumo
PROCEL	Programa Nacional de Conservação de Energia Elétrica
SIACC	Sistema Inteligente de Análise de Curvas de Carga
SINPHA	Sistema de Informação de Posses e Hábitos de Uso de Aparelhos Elétricos
RGE	Rio Grande Energia
RNA	Redes Neurais Artificiais

SUMÁRIO

1. INTRODUÇÃO	17
1.1 Motivação	17
1.2 Objetivo.....	19
1.3 Ineditismo	19
1.4 Publicações Relacionadas a Tese	20
1.5 Estrutura da Tese	22
2. REVISÃO BIBLIOGRÁFICA.....	24
2.1 Estudo de Anterioridade e Estado da Arte.....	24
2.2 Conceitos Importantes	30
2.2.1. <i>Curvas de Carga</i>	30
2.2.2. <i>Tipos de Consumidores</i>	31
2.2.3. <i>Classes Econômicas e Classe de Consumo</i>	35
2.2.4. <i>Faixa Etária Brasileira</i>	37
2.2.5. <i>Posse de Eletrodomésticos e Hábitos de Consumo</i>	38
2.3 Sistemas Inteligentes	46
2.3.1. <i>Floresta dos Caminhos Ótimos</i>	46
2.3.2. <i>Redes Bayesianas</i>	52
3. DESENVOLVIMENTO DO ALGORITMO PARA CONSTRUÇÃO DE CURVAS DE CARGA E VALIDAÇÃO	56
3.1 Algoritmo Desenvolvido	56
3.2 Curvas Geradas pelo Algoritmo.....	61
3.3 Elaboração do Banco de Dados e Validação do Algoritmo	64
4. ANÁLISE DAS CARACTERÍSTICAS SOCIOECONÔMICAS	72
4.1 Estudo da Influência do Número de Pessoas	72
4.1.1. <i>Região Norte – Número de Pessoas</i>	73
4.1.2. <i>Região Nordeste – Número de Pessoas</i>	74
4.1.3. <i>Região Centro-Oeste – Número de Pessoas</i>	76
4.1.4. <i>Região Sudeste – Número de Pessoas</i>	77
4.1.5. <i>Região Sul – Número de Pessoas</i>	78
4.1.6. <i>Influência da variação do número de pessoas sobre o consumo</i>	79
4.2 Estudo da Influência da Classe Econômica	80
4.2.1. <i>Região Norte – Classe Econômica</i>	81
4.2.2. <i>Região Nordeste – Classe Econômica</i>	82
4.2.3. <i>Região Centro-Oeste – Classe Econômica</i>	84
4.2.4. <i>Região Sudeste – Classe Econômica</i>	85
4.2.5. <i>Região Sul – Classe Econômica</i>	86
4.2.6. <i>Influência da variação da classe econômica sobre o consumo</i>	87
4.3 Estudo da Influência do Tamanho da Residência	88
4.3.1. <i>Região Norte – Tamanho da Residência</i>	89
4.3.2. <i>Região Nordeste – Tamanho da Residência</i>	90
4.3.3. <i>Região Centro-Oeste – Tamanho da Residência</i>	92
4.3.4. <i>Região Sudeste – Tamanho da Residência</i>	93
4.3.5. <i>Região Sul – Tamanho da Residência</i>	94
4.3.6. <i>Influência da variação do tamanho da residência sobre o consumo</i>	95

4.4	Estudo da Influência da Composição/Faixa Etária Familiar	96
4.4.1.	<i>Região Norte – Faixa Etária</i>	97
4.4.2.	<i>Região Nordeste – Faixa Etária</i>	98
4.4.3.	<i>Região Centro-Oeste – Faixa Etária</i>	99
4.4.4.	<i>Região Sudeste – Faixa Etária</i>	99
4.4.5.	<i>Região Sul – Faixa Etária</i>	100
4.4.6.	<i>Influência da variação da faixa etária sobre o consumo</i>	101
5.	ESTUDO DE CLASSIFICAÇÃO UTILIZANDO FERRAMENTAS INTELIGENTES	104
5.1	Estudo de Previsão da Classe de Consumo	108
5.2	Determinação das Características das Famílias.....	116
5.2.1.	<i>Classificação da Característica: Classe Econômica</i>	116
5.2.2.	<i>Classificação da Característica: Tamanho da Residência</i>	121
5.2.3.	<i>Classificação da Característica: Composição/faixa-etária</i>	125
6.	CENÁRIOS DE APLICAÇÃO DOS RESULTADOS	130
6.1	Planejamento de Novas Cargas na Rede de Distribuição de Energia	130
6.2	Metodologia de Carregamento de Transformadores	132
6.3	Participação dos Consumidores e a Relação Concessionária/Consumidores.	133
7.	CONCLUSÕES FINAIS	135
7.1	Aspectos Gerais	135
7.2	Futuros desdobramentos da pesquisa.....	137
	REFERÊNCIAS BIBLIOGRÁFICAS	139
	APÊNDICE A – Link para Tabelas dos Resultados da Análise Estatística	146
	ANEXO A – Link para NT 2.018	147

1. INTRODUÇÃO

1.1 Motivação

Segundo o anuário World Energy Outlook 2016, divulgado pela *Energy Information Administration* (EIA), é previsto um aumento de 30% da demanda global de energia, dados mostram que muitos países, preocupados com o meio ambiente e o efeito estufa, estão fazendo altos investimentos em fontes renováveis, e que houve uma queda acentuada nos investimentos em petróleo e gás, devido à proposta de diminuição de carbono emitido ao ar, embora cerca de 60% de todo investimento de energia ainda estão relacionados com os combustíveis fósseis (IEA, 2016).

Segundo o estudo apresentado por Peduzzi (2011), os grandes responsáveis pelo crescimento da energia global são a China e a Índia. Apesar desses países apresentarem aumentos significativos na demanda por energia, o consumo *per capita* desses é bem abaixo em relação ao de países desenvolvidos, o que mostra que, além da quantidade de pessoas, esses dois países ainda tem muita desigualdade social.

Nos últimos anos, o mercado energético brasileiro tem recebido influência de um conjunto de fatores e tendências associados à economia mundial, que ditam as condições e oportunidades de comércio, fluxo de capitais e inovações tecnológicas. Essas interações diárias são determinantes para a oferta e demanda de energia (GOMES, 2010).

Em 1980, a população brasileira era de 119 milhões de habitantes e o consumo global brasileiro, ou seja, o total consumido por todas as atividades desenvolvidas no Brasil, era de 122,7 TWh, o que resulta na razão de 1.031 kWh por habitante por ano. Em 2000, essa razão atingiu 1.955 kWh, o que representa um aumento de 90% ou 4,5% ao ano. Segundo o estudo apontado pelo BEN (EPE, 2013a) o total consumido foi de 498,4 TWh, e a população brasileira segundo o estudo do IBGE (Instituto Brasileiro de Geografia e Estatística) - 2007 era de 184 milhões de habitantes com estimativas de mais de 200 milhões em 2012. Ao se considerar essa estimativa, a relação passa a ser de aproximadamente 2.492 kWh por habitante por ano, caracterizando o crescimento da atividade econômica e a melhoria das condições de vida do povo brasileiro.

O último balanço energético referente ao ano de 2015 mostra que o consumo global do Brasil caiu em relação ao de 2014, em aproximadamente 2,1%, totalizando 464,7 TWh, sendo puxado principalmente pelo declínio de consumo das indústrias (-5,3%), a classe residencial

também teve uma queda, segundo o estudo. Ao se considerar a estimativa de população brasileira em 206.492.500 habitantes, têm-se um consumo de 2.250 kWh por habitante por ano, ou seja, uma diminuição nos últimos anos, devido ao cenário econômico que o país vive (EPE, 2016).

Tendo em vista esse histórico de crescimento no consumo de energia elétrica ano após ano e desde 2010, mesmo com diminuição de consumo *per capita*, a AIE (Agência Internacional de Energia) estima que até 2035 o consumo de energia primária no Brasil deve aumentar 78%, sendo o segundo país em crescimento de consumo, perdendo apenas para a Índia.

Devido a esse crescimento na demanda por energia elétrica no Brasil, a análise de como está ocorrendo o consumo de energia pela população brasileira se faz importante e uma forma de se analisar o consumo de energia é através das curvas de carga.

As curvas de carga possuem características de acordo com o tipo de consumidor (residencial, industrial, comercial, entre outros). Sendo que o consumidor industrial, hoje no Brasil, é o que mais consome, seguido do residencial e depois o comercial e outros. Porém, é o consumidor residencial que vem apresentando nos últimos anos o maior crescimento percentual, apesar de quedas nos dois últimos anos, atrelado principalmente aos avanços da tecnologia, ao aumento do poder de aquisição da população e ao investimento em melhoria do conforto em sua residência.

Apesar de o consumidor residencial possuir um perfil característico de consumo, a curva de carga está relacionada às escolhas de cada consumidor, sendo assim, refém do livre arbítrio de cada pessoa e de suas vontades. Mas, através de características culturais, climáticas, sociais e econômicas, padrões de perfis econômicos podem ser determinados, porém, a relação entre estas características muitas vezes são interpretadas de formas erradas, devido à toda complexidade envolvida.

Relacionar variáveis de diferentes campos de estudo para melhorar a previsão de demanda de energia é o desafio desta proposta. Sabe-se que aspectos socioeconômicos e culturais influenciam na posse de equipamentos e hábitos de consumo, no entanto, entender e ponderar pesos a essas características se faz necessário, para assim, ser possível enxergar de maneira mais padronizada os consumidores residenciais.

Neste contexto, o uso de sistemas computacionais inteligentes de tratamento de dados podem ser úteis para estabelecer esta relação, auxiliando na previsão da demanda e melhorando a análise do planejamento do sistema elétrico.

1.2 Objetivo

O objetivo desta tese foi desenvolver uma metodologia para previsão da demanda de energia de consumidores residenciais considerando aspectos socioeconômicos e ferramentas computacionais inteligentes, mostrando a importância em relacionar dados técnicos com dados econômicos e sociais para melhorar no reconhecimento de padrões de consumo da classe residencial.

Neste contexto foi avaliada a influência que cada característica socioeconômica tem sobre o consumo de energia residencial, avaliando cada uma das regiões do país de forma individualizada.

Posteriormente foi realizado um estudo com ferramentas computacionais inteligentes para melhorar a classificação e a criação de padrões de consumo relacionando as variáveis técnicas com as socioeconômicas.

Para finalizar, foi proposta a avaliação de dois possíveis cenários de aplicação desta nova metodologia desenvolvida, considerando o carregamento de transformadores e a utilização de padrão de consumo em um ambiente de *demand response*.

1.3 Ineditismo

Foi desenvolvido um algoritmo para construção de curvas de carga virtuais de consumidores residenciais que leva em consideração aspectos cultural, social e econômico, tais como, região, número de pessoas, faixa etária, tamanho da residência, classe econômica/social, sazonalidade de dias de semana e até mesmo datas comemorativas. Tratar tantos dados juntos para o estudo de curvas de carga é inovador e muito complexo, mas os resultados atingidos foram satisfatórios. Esse algoritmo desenvolvido em MATLAB® busca representar o livre arbítrio de cada consumidor, porém possui uma média representativa quando comparada com as médias medidas pelo país, além disso, foi comparado com medidas de campo. A partir desse algoritmo foi criado um banco de dados com diversas curvas de carga. Esse banco de dados foi utilizado para realizar uma análise da influência que cada uma das características tem sobre o consumo de energia, avaliando-se cada uma das cinco regiões do país separadamente.

Outro banco de dados foi criado para realizar estudos de classificação de perfis de consumo e de características das famílias, com o auxílio de ferramentas computacionais inteligentes. O ineditismo está em realizar a classificação através de características de diferentes campos (cultural, social, econômico). Primeiramente foi verificado a possibilidade de se obter perfis de consumo, apenas conhecendo as características das famílias, ou seja, prever qual a classe de consumo de uma residência apenas conhecendo suas características culturais, sociais e econômicas. Posteriormente foi avaliado a possibilidade de se conseguir extrair as características das famílias ao se conhecer o perfil de consumo da residência, uma vez que com a intensificação dos medidores inteligentes e do *Smart Grid*, há a possibilidade de armazenamento de dados do consumo da residência.

1.4 Publicações Relacionadas a Tese

Seguem artigos provenientes de resultados parciais da pesquisa em diversos fóruns e também os trabalhos submetidos à revistas, relacionados ao tema e/ou às ferramentas computacionais.

A) Aprovados em Eventos Científicos

- Danilo S. Gastaldello; André N. Souza; Haroldo L. M. Amaral; Marcio A. Gotti; João P. Papa; Juliana A. G. Maginador, **Estudo e Análise dos Padrões de Curvas de Carga de Consumidores Residenciais**. Congresso Brasileiro de Automática CBA, 2016, Vitória, Espírito Santo.
- Danilo Sinkiti Gastaldello; André Nunes de Souza; Haroldo Luiz Moretti do Amaral; Maria Goretti Zago, **Classification of Residential Load Curves in Consumption Classes Using Intelligent Techniques**. Congresso Latino-Americano de Geração e Transmissão de Energia Elétrica CLAGTEE 2015, São José dos Campos, São Paulo.
- Danilo Sinkiti Gastaldello; Haroldo Luiz Moretti do Amaral; André Nunes de Souza; Caio César Oba Ramos, **Income Classification of Residential Consumers Through Intelligent Techniques and Load Curves**. Innovative Smart Grid Technologies Latin America IEEE ISGT, 2015, Montevideo, Uruguai.
- Danilo Sinkiti Gastaldello; Haroldo Luiz Moretti do Amaral; André Nunes de Souza; Caio César Oba Ramos; Filipe Fernandes; Zita Vale, **Intelligent Techniques Application for Consumption Class Classification Through Residential Load**

Curves. Intelligent Systems Applications to Power Systems ISAP, 2015 Porto, Portugal.

- Danilo Sinkiti Gastaldello; André Nunes de Souza; Haroldo Luiz Moretti do Amaral; Maria Goretti Zago, **Técnicas Inteligentes para Classificação de Curvas de Carga Residenciais em Relação à Classe de Consumo e Faixa Salarial.** Conferência Brasileira sobre Qualidade da Energia Elétrica CBQEE, 2015, Campina Grande Paraíba.
- Danilo Sinkiti Gastaldello; André Nunes de Souza, **Estudo de Curvas de Carga e a Influência de Características Sociais, Econômicas e Culturais.** Congresso Brasileiro de Automática CBA, 2014, Belo Horizonte, MG.
- Danilo Sinkiti Gastaldello; André Nunes de Souza; Haroldo Luiz Moretti do Amaral; Pedro Faria; Zita Vale, **Study of Load Curves Concerning the Influence of Socio-economic and Cultural Issues.** IEEE PES Innovative Smart Grid Technologies, Europe. ISGT 2014, Istanbul, Turkey.
- Karila Martins dos Santos; Andre Nunes de Souza; Danilo Sinkiti Gastaldello; Caio César Oba Ramos. **Avaliação do Uso de Ferramentas Inteligentes no Estudo de Padrões em Curvas de Carga.** In: X Conferência Brasileira sobre Qualidade da Energia Elétrica, 2013, Araxá, Minas Gerais.
- Marcio Augusto Gotti; André Nunes de Souza ; Danilo Sinkiti Gastaldello. **Caracterização dos Padrões de Hábitos de Consumo Residencial e seus Impactos sobre as Curvas de Carga.** In: Simpósio Internacional de Iniciação Científica da USP, 2013, São Carlos. 21º SIICUSP, 2013.
- Marcio Augusto Gotti; André Nunes de Souza ; Danilo Sinkiti Gastaldello. **Estudos de Padrões de Hábitos de Consumo Residencial e seus Impactos no Sistema Elétrico.** In: Congresso de Iniciação Científica da Unesp, 2013, Bauru. CIC UNESP 2013.
- Karila Martins dos Santos; André Nunes de Souza ; Danilo Sinkiti Gastaldello . **Aplicação de Arquiteturas Inteligentes no Estudo de Padrões em Curvas de Carga de Consumidores Residenciais.** In: Congresso de Iniciação Científica da UNESP, 2013, Bauru. CIC UNESP 2013.
- Karila Martins dos Santos, André Nunes de Souza, Danilo Sinkiti Gastaldello, Caio César Oba Ramos, Pedro da Costa Junior. **Desenvolvimento de um Algoritmo para Determinar Curvas de Carga a partir da Posse e dos Hábitos de Consumo para**

Consumidores: Residenciais, Comerciais e Industriais. SBSE – Simpósio Brasileiro de Sistemas Elétricos 2012, Goiânia, Goiás, 2012.

- Karila M. Santos, André N. de Souza, Danilo S. Gastaldello, Pedro da Costa Junior. **Uma Ferramenta Computacional Aplicada no Estudo de Curvas de Carga em Sistema de Energia.** Conferência Brasileira de Dinâmica, Controle e Aplicações, DINCON 2011, Águas de Lindóia, São Paulo, Brasil, 2011.
- Karila M. Santos, André N. de Souza, Danilo S. Gastaldello. **Uma Discussão sobre Hábito de Consumo de Energia e seus Impactos na Tarifa.** Simpósio Internacional de Iniciação Científica, SIICUSP 2011, São Carlos, São Paulo, Brasil, 2011.
- Karila M. Santos, André N. de Souza, Danilo S. Gastaldello. **Estudo de Hábitos de Consumo Residencial para o Planejamento de Sistemas de Distribuição de Energia.** Congresso de Iniciação Científica da UFSCAR, CICUFSCAR 2011, São Carlos, São Paulo, Brasil, 2011.
- André N. de Souza, Renato de Oliveira, Pedro da Costa Junior, Maria G. Zago, João P. Papa, Danilo S. Gastaldello. **Estimação de Carga como Suporte do Planejamento de Sistemas de Distribuição de Energia Elétrica.** Conferência Brasileira de Dinâmica, Controle e Aplicações, DINCON 2010, Serra Negra, São Paulo, Brasil, 2010.

B) Artigos submetidos à revistas, tanto com o tema de curvas de cargas, quanto sobre ferramentas inteligentes

- Danilo Sinkiti Gastaldello; Haroldo Luiz Moretti do Amaral; André Nunes de Souza; João Paulo Papa, **A New Algorithm for Building Virtual Load Curves of Residential Consumers.** IEEE Latin America Transactions, submetido em 2017.
- Silas Evandro Nachif Fernandes; André Nunes de Souza; Danilo Sinkiti Gastaldello; Danilo Roberto Pereira; João Paulo Papa. **Pruning Optimun-Path Forest Ensembles Using Meta-heuristic Optimization for Land-Cover Classification.** International Journal of Remote Sensing. Submetido em 2016.

1.5 Estrutura da Tese

Esta tese está estruturada em sete capítulos:

- O Capítulo 2 – REVISÃO BIBLIOGRÁFICA. É apresentada um estudo de anterioridade com conceitos importantes sobre curvas de carga, demanda de energia, tipos de consumidores e diferenciação de classe econômica e classe de consumo, além disso, são apresentados dados relevantes para essa pesquisa com um levantamento de

dados de características do Brasil que influenciam no perfil de curva de carga dos consumidores residenciais. Nesse capítulo também é apresentado o estado da arte sobre o tema de curvas de carga. Por fim, é apresentada a teoria de contextualização de técnicas inteligentes utilizadas no trabalho;

- O Capítulo 3 – DESENVOLVIMENTO DO ALGORITMO PARA CONSTRUÇÃO DE CURVAS DE CARGA E VALIDAÇÃO. Nesse capítulo será abordada a metodologia utilizada para desenvolvimento do algoritmo de construção de curvas de carga virtuais de consumidores residenciais no Brasil, levando-se em consideração as variáveis expostas no capítulo anterior, esses dados são considerados inovadores, pois agregam informações de diferentes campos, como, cultural, social e econômico. Posteriormente é apresentado como foi realizado a validação desse algoritmo e os resultados atingidos;
- O Capítulo 4 – SIMULAÇÕES COM O ALGORITMO PARA ANÁLISE DAS CARACTERÍSTICAS ESCOLHIDAS. O capítulo apresenta simulações realizadas com o algoritmo e uma análise numérica para apontar quais as características que mais influenciam na construção de curvas de carga em cada uma das regiões do país;
- O Capítulo 5 – ESTUDO DE CLASSIFICAÇÃO DE PERFIS DE CONSUMO E DAS CARACTERÍSTICAS DAS FAMÍLIAS UTILIZANDO FERRAMENTAS INTELIGENTES. Nesse capítulo é apresentado mais uma inovação desta tese, aplicando-se ferramentas computacionais inteligentes, mais especificamente as Redes Bayesianas e o OPF, para classificação de curvas de carga virtuais, que foram construídas com dados de aspectos culturais, sociais e econômicos. Houve a avaliação da possibilidade de utilização dos estudos para auxiliar no planejamento de energia.
- O Capítulo 6 – CENÁRIOS DE APLICAÇÃO. Nesse capítulo é apresentado propostas de cenários de aplicação do conhecimento adquirido, através da discussão sobre planejamento de carregamento de transformadores, bem como, do conhecimento das características dos consumidores residenciais. São apresentadas as possibilidades de melhorias alcançadas com a utilização dos dados apresentados no estudo dessa tese; e
- O Capítulo 7 – CONCLUSÕES FINAIS. Serão apresentadas nesse capítulo as conclusões finais. Será também apresentado o levantamento de possibilidades de desdobramentos da pesquisa.

2. REVISÃO BIBLIOGRÁFICA

2.1 Estudo de Anterioridade e Estado da Arte

A energia é essencial nos dias atuais, ela é necessária para todas as atividades no mundo e em nossas vidas. Mais de 7 bilhões de pessoas na Terra utilizam energia diariamente para tornar suas vidas mais confortáveis, mais produtivas, seguras e saudáveis. Sendo assim, o crescimento populacional e a necessidade de conforto é que impulsionam o aumento da demanda por energia. Até 2040, o crescimento populacional e econômico levará a uma demanda maior, e por isso, a busca por melhora no uso e na geração da energia se torna essencial.

Um estudo da ExxonMobil (2012) mostrou uma projeção segundo a qual a demanda por energia apresentará, em 2040, um aumento de 35% em relação ao verificado em 2010. De acordo com o “Panorama Energético: Perspectivas para 2040”, a elevação será sustentada por práticas e tecnologias de economia de energia, aumento de combustíveis com baixa emissão de carbono e pelo desenvolvimento de novas fontes. Dentre as causas apontadas pelo estudo para a elevação da demanda energética estão o aumento da população mundial, que, em 2040, deverá atingir 9 bilhões, e o crescimento da economia global, principalmente nos países subdesenvolvidos (EXXONMOBIL, 2012).

Segundo o mesmo estudo, as residências e empresas representam uma parcela significativa da demanda por energia global, especialmente quando o uso da eletricidade é considerado. Até 2040, estima-se que devido à prosperidade econômica e o aumento contínuo no número de residências fará com que a demanda cresça em torno de 25% no setor residencial e comercial.

Nos países que não fazem parte da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), incluindo o Brasil, a maioria em desenvolvimento, as taxas de crescimento do consumo energético e de eletricidade foram de 3,1% e 4,8%, respectivamente, bem superiores aos valores mundiais, refletindo o crescimento econômico acelerado desses países quando comparados com os da OCDE, com taxas de apenas 1,0% e 2,6%, respectivamente, no período de 1970 a 2011.

Na comparação com o crescimento do PIB, verifica-se que, no mesmo período considerado, a elasticidade entre renda e consumo de eletricidade foi de 1,09 no mundo, 1,12 nos países não OCDE e de 1,00 nos países da OCDE. A elasticidade quase unitária demonstra a forte relação da energia e da eletricidade com a economia. Um país que não tenha acesso a

fontes de energia competitivas e tecnologia de transformação apropriada tem sérias dificuldades para promover seu desenvolvimento.

No Brasil, o consumo de eletricidade cresceu a uma taxa média de 5,8% ao ano, de 1973 a 2011, enquanto a demanda total energética (eletricidade, combustíveis fósseis, entre outras) foi de 3,2%, e o PIB, de 3,4%, valores bem superiores aos verificados no mundo. O consumo residencial, no Brasil, evoluiu, em média, 6,3%, enquanto o industrial, 4,0%, evidenciando um maior uso social da energia. No entanto, essas duas classes registraram quedas nos dois últimos anos, devido principalmente ao cenário econômico de crise que o país está vivendo.

Nos próximos dez anos, o planejamento energético do MME, considerando os estudos do Plano Decenal de Expansão de Energia (PDE) 2021, indica uma taxa média de crescimento de 4,9% ao ano para o consumo de eletricidade e para a demanda total de energia, num cenário de PIB de 4,7% ao ano (VENTURA FILHO, 2013).

O consumo de energia pelo setor de comércio e serviços no ano de 2013 cresceu 5,7% em relação a 2012. A região Sudeste foi responsável por metade deste resultado. Já o consumo de energia pela indústria em 2013 foi 0,6% maior que o registrado em 2012. Ainda segundo a EPE, o desempenho da indústria no Centro-Oeste e no Sul do país compensou o menor consumo dos seguimentos eletro intensivos (EPE, 2013b). No entanto, tanto as indústrias, quanto a classe residencial, apresentaram quedas nos anos subsequentes, devido principalmente ao cenário econômico desfavorável que o país está passando.

Devido ao crescente aumento e a tendência de continuar crescendo a demanda por energia, aliado às preocupações ambientais referentes à novas construções de fontes de energia, as companhias estão procurando investir na melhoria da qualidade de fornecimento e uso da energia. Muitos estudos e avanços tecnológicos ajudaram na eficiência dos equipamentos, investimentos foram feitos nas indústrias durante os últimos anos, e bons resultados foram atingidos, com novas tecnologias e planos de comércio de energia diferenciados.

Contudo, a classe residencial nunca foi estudada com tanto afínco, mesmo que ela represente, no caso brasileiro, cerca de 27% do consumo de energia. Trabalhos realizados buscavam encontrar padrões gerais/médios de consumo, levando a um planejamento muito subjetivo, podendo ocasionar investimentos além dos necessários. Além disso, com uma nova tendência de mercado de energia mundial, através da adoção do conceito de *Smart Grid* e *Demand Response*, a classe de consumidor residencial deverá ser muito mais investigada, pois

irá existir maior controle das cargas, bem como novos fluxos de energia, além de contratos diferenciados para esses consumidores, que poderão ajudar o sistema como um todo.

A grande dificuldade de se estudar essa classe de consumidor de uma forma mais profunda está relacionado ao livre arbítrio que cada consumidor tem de utilizar a energia conforme as suas necessidades. Mesmo que haja um padrão, o consumidor pode simplesmente mudar a forma de consumir energia devido a algum fato que ocorreu na sua vida. Assim sendo, ao estudar essa classe de consumo, deve-se levar em consideração aspectos sociais, econômicos, culturais e humanos, além de vislumbrar possibilidades de combinações existentes e padrões médios já estudados anteriormente.

O trabalho de Jardini et al. (2000) apresenta a caracterização de curvas de cargas, destacando os consumidores residenciais, comerciais e industriais ligados em baixa tensão. Nesse estudo, foram utilizados medidores eletrônicos para levantar a utilização real da energia dos consumidores. Para um dado consumidor ou transformador calculou-se as curvas médias e de desvio-padrão, considerando intervalos de 15 minutos. Essas curvas foram denominadas: média $\{\mu(t)\}$ e desvio-padrão $\{\sigma(t)\}$.

O passo seguinte foi a elaboração das curvas $\mu(t)$ e $\sigma(t)$ para os seguimentos do mercado analisados (residencial, comercial e industrial de baixa tensão). Para isso, os consumidores residenciais foram agrupados por faixa de consumo e os consumidores comerciais e industriais por ramo de atividade. Para obter as curvas de carga dos transformadores basta agregarem as curvas de carga dos consumidores a ele conectados.

Em Armstrong et al. (2009) é proposta a criação de perfis de carga para serem usados na simulação de dispositivos de cogeração residencial para analisar questões do sistema como desempenho, eficiência e potencial redução na emissão de poluentes.

A construção da simulação do desempenho, no entanto, carece de prever a capacidade da ocupação orientada de cargas elétricas (como iluminação e aparelhos), ou seja, esses perfis levam em consideração o quanto de carga ainda pode ser instalada nos locais estudados, para não haver uma sobrecarga no sistema de distribuição. Nesse estudo foram usados perfis de representação com uma resolução de cinco minutos com o intuito de diminuir o erro nos resultados das simulações feitas dos perfis.

O estudo feito por Oliveira et al. (2010) apresenta uma ferramenta, o software SIACC – Sistema Inteligente de Análise de Curvas de Carga. É um sistema computacional georeferenciado criado para análise de grandezas elétricas provenientes de medições de curvas de carga, sejam essas medidas em subestações de distribuição ou em consumidores de baixa e média tensão. Após tratamento dessas informações de medição, as mesmas são

armazenadas na base de dados do software proporcionando análises de crescimento espacial de demanda e consequente avaliação do carregamento e da perda de vida útil em transformadores de distribuição.

De maneira geral, o SIACC busca atender às necessidades da análise geográfica da carga que viabilize estudos voltados para análise de redes e previsão de mercado em diversos níveis, sendo eles: global, por subestação de distribuição e espacial.

Já nos estudos realizados por Santos et al. (2011, 2012 e 2013) é proposta a criação de um algoritmo para construção de curvas de carga, que leve em consideração os estudos realizados sobre a posse e os hábitos de consumo e posteriormente a utilização de redes neurais para a classificação dos consumidores em relação ao consumo de energia. Santos realizou inicialmente estudos com as três classes de consumidores (industrial, comercial e residencial), porém só aplicou o estudo com RNA na classe residencial, com um foco nos consumidores da região Sudeste do Brasil.

Outro estudo interessante está em avaliar como os consumidores se comportariam em relação ao consumo de energia se houvesse incentivos por parte do governo para que eles reduzissem o consumo, ou mesmo se os consumidores pudessem participar da negociação da energia como micro fornecedores e como consumidores. Nesse contexto, o conceito de *demand response* surge muito fortemente pelo mundo. Muitos avanços e investimentos ainda devem ser feitos, porém já se iniciaram pelo o mundo e estão atingindo ótimos resultados de eficiência energética (GHEYDI; FARHADI; GHAFARI, 2016).

A implantação de medidores de energia capazes de fornecer as informações requeridas através de *displays* é necessária de maneira que se possa proporcionar uma aplicação de tarifas diferenciadas em tempo real, monitoramento da energia, redução de carga, planejamento de interrupções de fornecimento, detecção de perdas de energia, previsão de demanda e serviços de planejamento, aumentando ainda mais os custos de operação (CHEN; HWANG, 2006).

Devido a grande dificuldade em se prever o consumo residencial, em Chrysopoulos et al. (2013) é proposta uma metodologia de incentivos a consumidores residenciais para que eles consumam de acordo com um padrão estabelecido, e é verificada a influência que esse consumidor causará na rede como um todo. Além disso, a metodologia proposta sugere a criação de cenários para a avaliação desse novo controle sobre os consumidores.

Já em Cao et al. (2013) é proposta a construção de padrões de consumo por família, ou seja, cada família deve ser estudada individualmente em sua região, pois o planejamento dos dias atuais utilizam curvas médias, que muitas vezes não caracterizam bem uma cidade, ou

um bairro específico. Assim, esse trabalho procura estabelecer novos padrões de consumo residencial para cada dia da semana e para cada família individualmente.

O trabalho apresentado em Ullah *et al.* (2013) está mais relacionado ao controle das cargas por parte dos fornecedores de energia, que conseguiriam a partir deste controle melhorar o fornecimento de energia e diminuiriam o desgaste de seus equipamentos, porém junto a esse controle é necessário que haja um incentivo aos consumidores para que os mesmos permitam que os fornecedores assumam algumas cargas chave para o controle durante o consumo de pico. Esse trabalho está diretamente atrelado aos conceitos de *demand response*.

O conceito de *demand response* é definido como a mudança dos hábitos de consumo de energia elétrica em resposta às alterações feitas nas tarifas ou aos detalhados dados de consumo expostos ao cliente, tendo como objetivo manter baixas flutuações da demanda e garantir o equilíbrio energético entre a fonte e a carga (ALBADI; EL-SAADANY, 2007). Nesse contexto, a aplicação da eletrônica de potência como ferramenta de controle da velocidade de motores e bombas por parte dos consumidores é considerada uma tecnologia chave para tornar possível as mudanças de hábitos de consumo nos períodos de pico (STEIMER, 2010).

Para Chaouch et al. (2014) o aumento da preocupação ambiental ligado ao aumento de demanda por energia faz com que investimentos sejam realizados em fontes renováveis, que podem ser utilizadas em momentos de picos de carga, porém, a conexão desses dispositivos pode acarretar em problemas para o sistema de distribuição, assim, no trabalho é proposta a existência de armazenadores de energia em clientes de baixa tensão, para que esses consigam suportar seus próprios picos de energia de forma isolada, esse trabalho trata com conceitos de *smart grid*, *demand response* e ilhamento.

A eficiência energética tem sido o foco de diversos estudos e seus resultados trazem benefícios diretos ao sistema elétrico. Segundo estudiosos e consultores da área, a eficiência energética é a solução mais rápida, barata e clara contra um sistema elétrico sobrecarregado (CANAVAGH; HOWAT, 2012). A busca pela melhoria da eficiência energética do sistema ocorre através da substituição, modernização e manutenção dos equipamentos para que o uso da energia seja feito de uma forma mais eficiente. Em vista disso, as concessionárias de energia elétrica tentam desenvolver ações que promovem maior eficiência energética do sistema, como por exemplo, melhoras nos sistemas de controle de consumo e mudanças nos períodos de uso e hábitos de consumo dos clientes (MAHESWARAN *et al.*, 2012). Além

disso, no Brasil, elas são obrigadas a investir um percentual de seus lucros em projetos de eficiência energética, de acordo com o projeto PEE, da ANEEL.

Dessa forma, a necessidade de implementação de ferramentas computacionais inteligentes que são capazes de processar esses dados está crescendo cada vez mais, principalmente no âmbito de determinação da topologia das curvas de carga (AL-OTAIBI *et al.*, 2016). Isso porque, a medição eletrônica possibilita que cargas possam ser deslocadas do período de ponta para os períodos fora de ponta, além de poder ser utilizada como uma ferramenta de localização de fraudes e furtos (LAMIN, 2013).

No âmbito residencial, Temneanu et al. (2014) expõe que a desagregação do consumo total de energia elétrica em contribuições individuais dos equipamentos que compõem a residência representa uma importante diretriz em pesquisas para o uso mais eficiente da eletricidade. Além disso, o autor explica a diferença entre os dois tipos de sistemas que podem ser implementados para a desagregação do consumo: o sistema intrusivo e o sistema não intrusivo. O sistema intrusivo é caracterizado pela instalação direta de sensores nos equipamentos, os quais farão medições e enviarão esses dados para um servidor. O sistema não-intrusivo é capaz de realizar a desagregação do consumo apenas analisando o consumo total do cliente. O algoritmo desenvolvido monitora os níveis de tensão e corrente continuamente. Caso a potência ativa ou reativa tenha uma variação que ultrapasse o patamar estipulado, o algoritmo extrai a forma de onda do equipamento que gerou essa variação, e faz uma análise espectral do sinal para determinar a assinatura do equipamento. Por fim, essa assinatura é armazenada em um banco de dados e será utilizada para identificação de eletrodomésticos em outras curvas de carga.

Em Kuruppu e Kulatunga (2012) destaca-se que para uma efetiva redução no consumo de energia elétrica, o consumidor deve obter informações precisas com relação ao consumo individual de cada equipamento da residência. Contudo, os métodos de detecção de consumo baseados em sensores externos são muito caros e, portanto, o desenvolvimento de um algoritmo que realize esse procedimento se faz necessário para que uma ampla aplicação se torne possível. Sendo assim, foi desenvolvido um algoritmo capaz de destacar os horários em que um equipamento é ligado ou desligado através do rastreamento em tempo real de modificações na demanda, fator de potência e ciclos de operação, destacando o período em que o mesmo esteve em funcionamento. Dessa forma, o consumidor consegue estar ciente do gasto energético do equipamento e realizar projeções da demanda.

Já no trabalho apresentado por Marques (2014) é proposta a utilização de redes neurais artificiais para modelagem de curvas de cargas típicas, considerando medições em campo e

características climáticas na região de Alegrete, Rio Grande do Sul. Com bons resultados de previsão de demanda e formação de curvas de potência ativa e reativa, é possível verificar que o clima possui forte influência no consumo e que os sistemas inteligentes podem interpretar bem esse problema de caracterização e construção de curvas de carga.

O uso de ferramentas computacionais inteligentes vem ganhando destaque na área da engenharia, atuando como classificador, previsor ou mesmo para desagregar informações de um dado com ruídos. Muito difundida no tratamento de dados *data mining*, auxiliando na identificação de inadimplentes de cartão de crédito, e também muito aplicada em tratamento de imagens, seja para usos de segurança, mapeamento de áreas (agricultura, topologia), seja na área da medicina, entre outros. A engenharia elétrica passou a entender que o tratamento de dados pode trazer informações pertinentes aos problemas mais complexos, e que a resposta mais rápida dos algoritmos computacionais podem auxiliar no planejamento do sistema, porém é necessário um especialista para interpretar os dados advindos das ferramentas inteligentes.

Em Ramos (2014) , é proposta a análise da escolha de melhores características para melhorar a taxa de acerto da identificação e classificação de consumidores que realizam fraudes ou não, diminuindo a taxa de perdas comerciais. Através de ferramentas híbridas baseadas em técnicas evolutivas, o trabalho verificou que a escolha dos dados para o estudo de classificação pode ajudar na acurácia do algoritmo, bem como diminuir o tempo de processamento da ferramenta.

Nesse contexto, a presente tese objetiva o desenvolvimento de uma metodologia de previsão de demanda de energia residencial considerando aspectos socioeconômicos e ferramentas computacionais inteligentes, mais precisamente as Redes Bayesianas e o OPF.

O tópico a seguir irá apresentar alguns conceitos importantes, que servirão de base para o desenvolvimento dessa tese.

2.2 Conceitos Importantes

2.2.1. Curvas de Carga

Define-se curva de carga como a curva que apresenta a demanda em função do tempo, $D(t)$, para um dado período T , conforme a Figura 1. Na verdade, ela é constituída por patamares, no entanto, é mais comum apresentá-la como uma curva, resultando da união dos pontos médios das bases superiores dos retângulos de largura Δt . Para um período T , a

ordenada máxima da curva define a demanda máxima, DM . A energia total consumida no período, εT , é medida pela área entre a curva e o eixo dos tempos, ou seja,

$$\varepsilon_T = \int_0^T D(t) \cdot dt \quad (1)$$

A demanda média, D_m , é definida como a altura do retângulo cuja base é o período T e cuja área é a energia total εT . Ou seja,

$$D_m = \frac{\varepsilon_T}{T} \quad (2)$$

Figura 1 – Curva de Carga Padrão

Caracterizar a carga elétrica é, acima de tudo, conhecer a forma como os clientes utilizam a energia, posto que a carga seja uma variável muito dependente de seu consumidor, seja ele, residencial, comercial ou industrial. A carga sofre influências de muitos fatores e pode variar com o tempo.

2.2.2. Tipos de Consumidores

A classificação dos consumidores de energia elétrica leva em consideração a finalidade para qual o usuário consome a energia. As principais classificações dos consumidores são:

- Residenciais;
- Comerciais; e
- Industriais.

Esses critérios de classificação são importantes para análises de hábitos de consumo de cada tipo de consumidor.

A Figura 2 representa a distribuição do consumo de energia elétrica por cada tipo de consumidor no Brasil:

Figura 2 – Consumo de energia elétrica por setor (Fonte: EPE 2015 – adaptado).

a) Consumidores residenciais

A curva de carga dos consumidores residenciais é caracterizada por uma carga com crescimento constante, apresentando pico de demanda entre as 18h e às 22h, sensível a mudanças meteorológicas, econômicas, tarifárias, entre outras.

Um estudo feito pelo PROCEL (2007) mostrava que este setor de consumidor representava 22,2% de todo o consumo de energia elétrica nacional, porém estudos recentes já apontam um aumento para aproximadamente 28%, conforme Figura 2. Já a Figura 3 representa o consumo de energia de um consumidor residencial dividido entre os eletrodomésticos.

Figura 3 – Consumo final na carga residencial (Fonte: PROCEL, 2007 - adaptado).

Sabe-se que o consumo final na carga residencial sofreu algumas alterações de 2007 até os dias atuais, contudo, nenhum estudo foi publicado nessa área de contabilização do percentual de consumo de cada equipamento na residência. Apesar de os chuveiros elétricos, os condicionadores de ambiente e as geladeiras continuarem sendo os principais equipamentos relacionados ao consumo, é crescente o consumo dos computadores e das televisões.

A Figura 4 representa a curva de carga diária de um consumidor residencial classificado na faixa de consumo de 330 kWh/mês.

Figura 4 – Curva de carga de consumidor residencial (FONTE: FRANCISQUINI, 2006).

b) Consumidores comerciais

Os consumidores comerciais são caracterizados por uma carga sensível a sazonalidade como datas comemorativas, às condições meteorológicas, susceptível às mudanças econômicas e à qualidade de energia entregue pelas companhias. A curva desse tipo de consumidor é aproximadamente constante no horário de serviço, entre as 8h e às 18h.

Esse setor de consumidor representa uma demanda de aproximadamente 19% da energia elétrica produzida no Brasil. São geralmente classificados por ramo de atividade como, por exemplo, hotéis, supermercados, bancos, hospitais, instituições de ensino, varejos, restaurantes, locais de entretenimento, entre outros. Fora do horário comercial a demanda desses consumidores deve-se, praticamente, à refrigeração e iluminação. A Figura 5 apresenta a distribuição do consumo de energia elétrica por aparelho do setor comercial (PROCEL, 2007).

Figura 5 – Representação do consumo de energia elétrica por equipamento para o setor comercial (Fonte: PROCEL, 2007 - adaptado).

A Figura 6 representa a curva de carga característica de um consumidor comercial.

Figura 6 – Curva de carga de um consumidor comercial (FONTE: FRANCISQUINI, 2006).

c) Consumidores industriais

Os consumidores industriais têm suas curvas de carga bem diferentes entre si. Isso porque, para cada tipo de atividade desenvolvida pela indústria, uma curva de carga característica é formada. Os consumidores industriais correspondem por aproximadamente 38% da demanda total, sendo que desta porcentagem a força motriz corresponde a 68,3% da demanda. A Figura 7 apresenta a distribuição do consumo energético por setores da indústria (PROCEL, 2007).

Figura 7 – Representação do consumo de energia elétrica na indústria por setor (Fonte: PROCEL, 2007 - adaptado).

A Figura 8 apresenta as curvas de carga características das indústrias levando em consideração diferentes ramos, com consumos diferentes de acordo com a atividade da indústria.

Figura 8 – Curvas de carga do setor industrial (FONTE: FRANCISQUINI, 2006).

2.2.3. Classes Econômicas e Classe de Consumo

Há vários critérios para definir classes sociais/econômicas, sem uma predileção específica na literatura. Pesquisas sobre o consumo costumam relatar hábitos segundo as classes sociais. Entretanto, a metodologia de cálculo e caracterização de cada uma das classes é bastante difusa e há, atualmente, pelo menos duas visões relevantes:

- (1) ABEP - Associação Brasileira de Empresas de Pesquisa.
- (2) IBGE - Instituto Brasileiro de Geografia e Estatística, utilizado no censo populacional.

A visão da ABEP é a utilizada pela maioria dos institutos de pesquisa. Segundo a ABEP, o Critério de Classificação Econômica Brasil (CCEB) é um instrumento de

segmentação econômica que utiliza o levantamento de características domiciliares (presença e quantidade de alguns itens domiciliares de conforto e grau escolaridade do chefe de família) para diferenciar a população. O critério atribui pontos em função de cada característica domiciliar e realiza a soma desses pontos. É feita então uma correspondência entre faixas de pontuação do critério e faixas de classificação econômica definida por A1, A2, B1, B2, C1, C2, D, E. A Tabela 1 apresenta a faixa (*range*) de dados para cada uma das classes de classificação deste método.

Tabela 1 – Faixas para Classificação Social – Critério Brasil

Classe	Pontos
A1	42-46
A2	35-41
B1	29-34
B2	23-28
C1	18-22
C2	14-17
D	8-13
E	0-7

Fonte: ABEP - Associação Brasileira de Empresas de Pesquisa, 2011.

A visão do IBGE, baseada no número de salários mínimos, é mais simples e divide em apenas cinco faixas de renda ou classes sociais, conforme a Tabela 2 válida para o ano de 2016 (salário mínimo em R\$ 880,00). Essa tabela foi obtida a partir de vários artigos sobre classes sociais nas pesquisas do IBGE divulgados na imprensa e é parecida com a visão da FGV (Fundação Getúlio Vargas).

Tabela 2 – Classes Sociais – Critério IBGE

Classe	Número de Salários Mínimos	Renda Familiar (R\$)
A	Acima de 20	17.600,00 ou mais
B	10 a 20	8.800,00 a 17.599,99
C	4 a 10	3.520,00 a 8.799,99
D	2 a 4	1.760,00 a 3.519,99
E	Até 2	Até 1.759,99

Fonte: IBGE - Instituto Brasileiro de Geografia e Estatística, 2013 – adaptado para 2017.

Nesta tese optou-se por utilizar os critérios do IBGE, pois esses não dependem de características dos integrantes da residência, como por exemplo a formação escolar do chefe da família. Cabe salientar que independente do critério utilizado, o aumento de classe social/econômica está relacionado com o aumento de consumo de bens e posses de equipamentos e consequentemente com a variação do consumo de energia elétrica.

Já a classe de consumo de energia elétrica, especificamente dos consumidores residenciais, é dividida de acordo com a quantidade de energia consumida. A Tabela 3 apresenta a classificação de consumo comumente utilizada pelas concessionárias e grupos de pesquisa, porém uma nova tabela será proposta neste trabalho de doutorado.

Tabela 3 – Classificação dos Perfis de Consumo de Consumidor Residencial

Perfil	Consumo Mensal (kWh)	Consumo Diário (kWh)
A	Maior que 500	Maior que 16,667
B	300 - 500	10,001 – 16,666
C	200 - 300	6,667 – 10
D	100 - 200	3,334 – 6,666
E	0 – 100	0 – 3,333

2.2.4. Faixa Etária Brasileira

No Brasil, em consequência da desaceleração do aumento médio anual da população (crescimento demográfico) e o envelhecimento populacional, está havendo a mudança do desenho da pirâmide etária. Até os anos 1980, a pirâmide era afunilada na ponta e larga na base, setor em que estão representadas as camadas de jovens. No ano de 2000, a pirâmide mostrou uma base mais estreita e formato menos afunilado, indicando a tendência de crescimento da população adulta.

Nas últimas décadas, aumentou o percentual de idosos e adultos e diminuiu a porcentagem de jovens. Esse fato é o resultado da diminuição das taxas de mortalidade e natalidade e aumento da expectativa de vida.

Em 1950, a distribuição era a seguinte: idosos, 4,6%; adultos, 43,1%; e jovens, 52,3%. Conforme dados do último recenseamento geral da população, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2016), a faixa etária dos jovens abrange 40,2%, a dos adultos 50,5%, e a dos idosos, 9,3% do total da população.

A média da expectativa de vida da população brasileira só não é maior devido à grande quantidade de assassinatos entre os indivíduos da faixa etária composta por jovens.

Esse panorama nacional está relacionado ao consumo de energia elétrica na medida em que a faixa etária influencia nos costumes do uso e também na posse de equipamentos, assim, entender como é composta as famílias brasileiras no que diz respeito a faixa etárias dos participantes de uma residência é de fundamental importância.

A Tabela 4 apresenta a distribuição da população por região e por faixa etária.

Tabela 4 – Divisão da população por região e faixa etária no Brasil.

FAIXA (ANOS)	REGIÕES – POPULAÇÃO RESIDENTE x 1000 PESSOAS				
	NORTE	NORDESTE	CENTRO-OESTE	SUDESTE	SUL
0-1	263	721	208	969	342
1-4	1210	3359	839	4008	1400
5-6	662	1837	425	2161	738
7-9	1020	2835	687	3434	1139
10-14	1746	5316	1249	6649	2232
15-17	1089	3286	757	3984	1464
18-19	649	1928	484	2545	865
20-24	1450	4471	1274	6569	2263
25-29	1418	4555	1278	6796	2245
30-49	4354	14801	4463	24566	8140
50-59	1336	4904	1449	9490	3387
60-64	418	1846	520	3453	1203
65-69	338	1465	366	2563	890
70 ou +	546	2902	579	4882	1565

Fonte: Síntese de Indicadores Sociais – Uma Análise das Condições de Vida da População Brasileira, 2012.

2.2.5. *Posse de Eletrodomésticos e Hábitos de Consumo*

Os estudos e os levantamentos de dados sobre as características de demanda e eletrodomésticos utilizados pelos consumidores residenciais foram baseados nas Pesquisas de Posse e Hábitos de Consumo (PPH) realizadas pelo PROCEL em 2007, que corresponde a última publicação desses dados.

As PPH's foram realizadas através da aplicação de questionários contemplando dezesseis estados e o Distrito Federal, com a participação de 21 empresas: Celpa, Ceron e Manaus Energia (região Norte); Celpe, Cemar, Coelba, Coelce e Cosern (região Nordeste); Ampla “ex-Cerj”, Cemig, CPFL, Eletropaulo, Light (região Sudeste); CEEE, RGE, Copel e Celesc (região Sul) e CEB, Celg e Cemat (região Centro-Oeste). Foram aplicados 9.847 questionários nessa pesquisa. Os dados da Elektro também foram coletados com o mesmo questionário das demais, porém com ampliação da amostra.

Dessa pesquisa de Posse e Hábitos de Consumo (PPH) foi possível verificar quais são os eletrodomésticos que proporcionam o maior consumo de energia em cada região, uma pequena descrição dos motivos é apresentada a seguir.

A) Região Norte

Os aparelhos eletrodomésticos mais utilizados são os condicionadores de ambiente, devido ao clima quente da região, seguidos de geladeiras e lâmpadas. Outra consequência das altas temperaturas é o baixo consumo do chuveiro elétrico. A Figura 9 representa a participação dos eletrodomésticos no consumo residencial na região Norte, destacando os condicionadores de ambiente com cerca de 40%.

Figura 9 – Representação da participação de eletrodomésticos no consumo residencial da região Norte
(Fonte: PROCEL, 2007- adaptado).

A Figura 10 mostra que os maiores responsáveis pelos picos de demanda nas curvas de carga dos consumidores residenciais da região Norte são os condicionadores de ambiente, que nesse caso são utilizados como ar condicionado. Observa-se também que durante o horário de pico, entre as 18h e 22h, a TV e as lâmpadas influenciam bastante no consumo das residências desta região. Outra característica importante que deve ser destacada é em relação ao consumo de energia durante a madrugada, com os condicionadores de ar que permanecem ligados devido às altas temperaturas.

Figura 10 – Curva de carga média de consumidores residenciais da região Norte
(Fonte: PROCEL, 2007).

B) Região Nordeste

Há três tipos de clima que agem sobre a região: tropical, semiárido e equatorial úmido. O primeiro possui elevadas temperaturas e duas estações bem definidas, sendo uma seca e uma chuvosa, as temperaturas variam entre 24°C e 26°C. O segundo possui temperaturas elevadas e chuvas irregulares, essa característica climática faz com que as áreas influenciadas sejam secas devido aos longos períodos de estiagem e no terceiro existe a predominância de uma grande umidade relativa do ar, além disso, demonstra elevadas temperaturas com chuvas regulares durante todo o ano. A Figura 11 representa a participação dos eletrodomésticos no consumo residencial da região Nordeste.

Figura 11 – Participação dos eletrodomésticos no consumo residencial da região Nordeste
(Fonte: PROCEL, 2007 - adaptado).

Devido às altas temperaturas nos três tipos de clima presentes na região Nordeste, assim como na região Norte, os condicionadores de ambiente utilizados como ar condicionado são os aparelhos que possuem alta demanda por energia elétrica, no entanto na região Nordeste, a geladeira é considerada a maior vilã, com cerca de 29%. Observa-se que há o uso de chuveiro elétrico em alguns estados da região, pois a sensação térmica é menor do que no Norte, além disso, uma diminuição percentual do uso dos condicionadores de ar, devido uma variação maior do número de frentes um pouco mais frias nos estados do litoral ocasionando uma redução de aquisição desses aparelhos.

Na Figura 12 é possível verificar a inclusão dos chuveiros durante o horário de pico, mantendo-se as TVs e as lâmpadas durante este período. Durante a madrugada os condicionadores de ar são utilizados como no Norte, porém com menor potência média,

devido a menor quantidade de pessoas que possuem o aparelho e também devido às sensações menores de temperatura.

Figura 12 – Curva de carga média de consumidores residenciais na região Nordeste
(Fonte: PROCEL, 2007).

C) Região Centro-Oeste

O clima característico da região é denominado tropical semiúmido, com frequentes chuvas de verão. Na primavera/verão, são comuns temperaturas elevadas, sendo que a média do mês mais quente varia de 24° a 26 °C. Já no inverno, em virtude da invasão polar, é comum a ocorrência de temperaturas mais baixas, as médias ficam em torno de 15 °C a 24 °C.

Devido a essa tipologia climática da região Centro-Oeste, com frequentes chuvas de verão e inverno com baixas temperaturas, os chuveiros elétricos passam a ter a maior participação na demanda por energia elétrica, seguidos pelos condicionadores de ambientes que são bastante utilizados nas estações mais quentes. Essa divisão da participação no consumo de energia elétrica pelos eletrodomésticos pode ser visualizada na Figura 13.

Participação dos Eletrodomésticos no Consumo - Centro-Oeste

Figura 13 – Participação dos eletrodomésticos no consumo residencial da região Centro-Oeste
(Fonte: PROCEL, 2007 - adaptado).

A Figura 14 demonstra que o chuveiro elétrico também é responsável pelo grande aumento da demanda em horários de pico na região Centro-Oeste, algo que não vinha acontecendo nas regiões anteriormente analisadas. Os televisores também são ligados, porém possuem uma pequena contribuição comparada aos chuveiros.

Além disso, o horário em que acontece a demanda máxima na região Centro-Oeste é em torno das 19 horas, sendo também diferente das regiões Norte e Nordeste onde o horário de demanda máxima ocorre por volta das 22 horas, onde os condicionadores de ambiente são mais acionados.

Cabe ressaltar um novo momento de pico durante as manhãs devido o uso dos chuveiros elétricos, salienta-se também o valor do pico durante as 19 h que chega próximo a 400 Wh/h, enquanto que na região Nordeste era de 300 Wh/h e no Nordeste de 350 Wh/h, aproximadamente.

Figura 14 – Curva de carga diária média de um consumidor residencial da região Centro-Oeste (Fonte: PROCEL, 2007).

D) Região Sudeste

Essa Região do Brasil situa-se na parte mais elevada do planalto Atlântico, onde estão localizadas as serras da Mantiqueira, do Mar e do Espinhaço. Nessas localidades predominam o clima tropical úmido e o tropical semiúmido, com geadas ocasionais.

Dessa forma, diferentes médias de temperatura estão presentes em toda a região. No limite de São Paulo e Paraná, a temperatura média anual situa-se próximo de 20 °C, enquanto ao norte de Minas Gerais a média é 24 °C, e nas áreas mais elevadas das serras do Espinhaço,

Mantiqueira e do Mar, a média pode ser inferior a 18 °C, devido ao efeito conjugado da latitude com a frequência das correntes polares.

Com temperaturas mais amenas, assim como na região Centro-Oeste, os condicionadores de ambiente são menos utilizados pela população residencial da região. Em contrapartida, temperaturas mais baixas levam a uma maior utilização dos chuveiros elétricos e, por isso, eles têm a maior parcela de contribuição na demanda total de energia elétrica nas residências, por volta de 26%, como mostrado pela Figura 15.

Figura 15 – Participação dos eletrodomésticos no consumo residencial da região Sudeste
(Fonte: PROCEL, 2007- adaptado).

Um fator que chama bastante atenção na Figura 15 é o de que, analisando a participação dos eletrodomésticos na demanda por energia elétrica em todas as regiões, é na região Sudeste em que se tem a maior contribuição das lâmpadas. Isso porque, verificou-se que a posse média de lâmpadas incandescentes é superior ao de lâmpadas fluorescentes. Os consumidores residenciais dessa região possuem a maior média de lâmpadas incandescentes por domicílio, 5,36 unidades (PROCEL, 2007).

Observando a curva de carga característica da região Sudeste, representada pela Figura 16, pode-se notar que os chuveiros elétricos são os maiores responsáveis pela formação dos picos de carga que ocorrem no período da manhã (das 6 horas até as 8 horas) e a noite (das 18 horas as 23 horas).

A Figura 16 mostra que a demanda máxima ocorre por volta das 20 horas com um consumo de aproximadamente 500 Wh/h. Esse pico é o segundo maior entre todas as curvas de carga médias analisadas, perdendo somente para a região Sul. Isso pode ser explicado

devido essas regiões (Sudeste e Sul) serem as mais desenvolvidas economicamente e, portanto, terem a população com o maior percentual de posse de eletrodomésticos.

Figura 16 – Curva de carga diária média de consumidores residenciais da região Sudeste
(Fonte: PROCEL, 2007).

E) Região Sul

Nela predomina o clima temperado, responsável pelas temperaturas mais baixas registradas no Brasil durante o inverno. A única exceção é o norte do Paraná, onde se faz presente o clima tropical. Durante o inverno, na região central do Paraná e no planalto serrano de Santa Catarina e Rio Grande do Sul, os termômetros registram temperaturas negativas com possíveis ocorrências de geada e neve. A temperatura média anual situa-se entre 14 e 22 °C, sendo que nos locais com altitudes acima de 1.100 m, cai para aproximadamente 10 °C.

Devido à região Sul possuir a menor média de temperatura anual de todo o país, a demanda por energia elétrica se concentra em mais da metade pelo uso de chuveiros elétricos e condicionadores de ambiente que trabalham em ciclo reverso funcionando como aquecedores. Segundo pesquisa da PROCEL, as residências localizadas nessa região possuem a maior média de chuveiros elétricos, 1,17 unidades por domicílio. (PROCEL 2007).

Dessa forma, a contribuição de cada aparelho eletrodoméstico no consumo de energia elétrica na região Sul é representada na Figura 17.

Participação dos Eletrodomésticos no Consumo - Sul

Figura 17 – Participação dos eletrodomésticos no consumo residencial da região Sul
(Fonte: PROCEL, 2007- adaptada).

Um aspecto importante a ser analisado é o de que essa contribuição dos chuveiros elétricos na demanda por energia na região Sul tende a não se modificar. Isso porque, mais da metade da população, 58,8%, não mudariam o sistema de aquecimento atual.

Figura 18 – Curva de carga média de um consumidor residencial da região Sul
(Fonte: PROCEL, 2007).

Portanto, assim como nas regiões Sudeste e Centro-Oeste, o chuveiro elétrico é o maior responsável pela formação dos picos de demanda em horários críticos. Como mostrado pela Figura 18, os horários críticos da região Sul ocorrem das 6 horas às 9 horas da manhã e das 17 horas à meia noite. A demanda máxima é anotada por volta das 19 horas e possui o maior consumo entre todas as regiões analisadas, aproximadamente 700 Wh/h.

Apesar do consumo de energia diminuir a partir da meia noite, ele ainda continua alto devido ao uso dos condicionadores de ambiente. A demanda por energia só é realmente baixa das 10 horas às 17 horas, pois nesse período a maioria da população não está em casa.

2.3 Sistemas Inteligentes

Diante de todo o estudo já desenvolvido na área e buscando um trabalho inédito propõe-se um estudo mediante a aplicação de sistemas inteligentes para auxiliar na construção de curvas de carga e no auxílio do planejamento do sistema de distribuição de energia. Entre as técnicas de inteligência computacional mais utilizadas para a classificação de padrões, de maneira a solucionar os problemas nos sistemas de energia, estão as Redes Neurais Artificiais e as Máquinas de Vetores de Suporte (HAYKIN, 1998) e (HEARST *et al.*, 1998).

Técnicas evolutivas têm deixado clara a importância da concepção de algoritmos baseados em dinâmica social e no comportamento coletivo, sendo de grande interesse no contexto da seleção de características. A facilidade de implementação e a convergência rápida em algumas aplicações têm atraído muita atenção para essas abordagens, que cobrem uma vasta gama de fenômenos naturais para modelar problemas de otimização, como um bando de pássaros voando em conjunto em busca de comida na Otimização por Enxames de Partículas (*Particle Swarm Optimization* - PSO) (KENNEDY; EBERHART, 2001), a evolução natural em Algoritmos Genéticos (*Genetic Algorithm* - GA) (KOZA, 2012), improvisação musical em Busca Harmônica (*Harmony Search* - HS) (GEEM, 2009) e caminhos ótimos em Otimização de Colônia de Formigas (*Ant Colony Optimization* - ACO) (DORIGO; MANIEZZO; COLORNI, 1996), Algoritmo do Morcego (*Bat Algorithm* - BA) (YANG, 2011), Algoritmo do Vaga-Lume (*FireFly Algorithm* - FFA) (YANG, 2010), Busca do Sistema Carregado (*Charged System Search* - CSS) (KAVEH; TALATAHARI, 2010), Floresta dos Caminhos Ótimos (*Optimum-Path Forest* - OPF) (PAPA; FALCÃO, A. X.; SUZUKI, 2009) e Redes Bayesianas (*Bayesian Networks* – BN) (CHARNIAK, 1991).

Nesta tese, em específico, serão utilizadas duas técnicas (Floresta dos Caminhos Ótimos e Redes Bayesianas) para estudo de classificação e padronização, devido às suas velocidades de aprendizado com bases de dados grandes e boa acurácia de resultados. A seguir é apresentado um resumo de cada uma dessas técnicas.

2.3.1. Floresta dos Caminhos Ótimos

Abordagens baseadas em OPF tratam as amostras como sendo os nós de um grafo, sendo os arcos definidos por uma relação de adjacência e ponderados por alguma métrica de distância aplicada em seus vetores de atributos, e diferem dos métodos tradicionais por não utilizar a ideia de geometria do espaço de atributos conseguindo, assim, melhores resultados

em bases com *outliers* (amostras de uma determinada classe que estão presentes em uma região de outra classe no espaço de atributos) e sobreposição de classes.

a) Classificador OPF

A técnica de classificação supervisionada baseada em florestas de caminhos ótimos apresentada modela as amostras como sendo os nós de um grafo completo. Os elementos mais representativos de cada classe do conjunto de treinamento, isto é, os protótipos, são escolhidos como sendo os elementos pertencentes às regiões de fronteira entre as classes. Os protótipos participam de um processo de competição disputando as outras amostras oferecendo-lhes caminhos de menor custo e seus respectivos rótulos. Ao final deste processo, obtém-se um conjunto de treinamento particionado em árvores de caminhos ótimos, sendo que a união das mesmas nos remete a uma floresta de caminhos ótimos. Esta abordagem apresenta vários benefícios em relação a outros métodos de classificação de padrões supervisionados: (i) é livre de parâmetros, (ii) possui tratamento nativo de problemas multiclases e (iii) não faz alusão sobre forma e/ou separabilidade das classes.

b) Fundamentação Teórica

Seja Z uma base de dados λ -rotulada de Z_1 e Z_2 os conjuntos de treinamento e teste, respectivamente, com $|Z_1|$ e $|Z_2|$ amostras, tal que $Z = Z_1 \cup Z_2$. Seja $\lambda(s)$ uma função que associa o rótulo correto i , $i = 1, 2, \dots, c$ da classe i a qualquer amostra $s \in Z_1 \cup Z_2$. Seja $S \in Z_1$ um conjunto de protótipos de todas as classes (isto é, amostras que melhor representem as classes). Seja $\vec{o}$ um algoritmo que extrai n atributos de qualquer amostra $s \in Z_1 \cup Z_2$, e retorna um vetor de atributos $\vec{o}(s) \in \mathbb{R}^n$. A distância $d(s, t)$ entre duas amostras, s e t , é dada pela distância entre seus vetores de atributos $\vec{o}(s)$ e $\vec{o}(t)$. O problema consiste em usar S , $(\vec{v}, d)$ e Z_1 para projetar um classificador ótimo, o qual pode prever o rótulo correto $\lambda(s)$ de qualquer amostra $s \in Z_2$. Assim sendo, OPF é um classificador que cria uma partição discreta ótima, a qual é uma floresta de caminhos ótimos computada em $\mathbb{R}^n$ pelo algoritmo da transformada imagem floresta (FALCÃO, A. X.; STOLFI e LOTUFO., 2004)

Seja (Z_1, A) um grafo completo cujos nós são amostras em Z_1 , onde qualquer par de amostras define um arco A (isto é, $A = Z_1 \times Z_1$) (Figura 19 - a). Note que os arcos não precisam ser armazenados e o grafo não precisa ser explicitamente representado.

Tem-se, também, que um caminho em um grafo é uma sequência de amostras $\pi_{sk} = \langle s_1, s_2, \dots, s_k \rangle$, onde $(s_i, s_{i+1}) \in A$ para $1 \leq i \leq k - 1$. Um caminho é dito ser trivial

se $\pi_s = \langle s \rangle$. Associa-se a cada caminho π_s um valor dado por uma função de valor de caminho f , denotada $f(\pi_s)$. É dito que um caminho π_s é ótimo se $f(\pi_s) \leq f(\tau_s)$ para qualquer caminho τ_s , onde π_s e τ_s terminam na mesma amostra s , independente de sua origem. Também denota-se $\pi_s \cdot \langle s, t \rangle$ a concatenação do caminho π_s com término em s e o arco (s, t) . O algoritmo OPF pode ser utilizado com qualquer função de valor de caminho suave. Uma função de caminho f é suave quando, para qualquer amostra t , existe um caminho ótimo π_t o qual é trivial ou possui a forma $\pi_s \cdot \langle s, t \rangle$, onde:

- $f(\pi_s) \leq f(\pi_t)$;
- π_s é ótimo; e
- para qualquer caminho ótimo $\tau_s, f(\tau_s \cdot \langle s, t \rangle) = f(\pi_t)$.

Na versão OPF com grafo completo, a função de custo abordada foi a $f_{\max}$, a qual é definida como:

$$f_{\max}(\langle s \rangle) = \begin{cases} 0 & \text{se } s \in S \\ +\infty & \text{caso contrário} \end{cases}$$

$$f_{\max}(\pi_s \cdot \langle s, t \rangle) = \max\{f_{\max}(\pi_s), d(s, t)\}, \quad (3)$$

onde $f_{\max}(\pi_s)$ computa a distância máxima entre amostras adjacentes em π_s , quando π_s não é um caminho trivial.

O algoritmo baseado em OPF associa um caminho ótimo $P^*(s)$ de S a toda amostra $s \in Z_1$, formando uma floresta de caminhos ótimos P (uma função sem ciclos, a qual associa a todo $s \in Z_1$ seu predecessor $P(s)$ em $P^*(s)$, ou uma marca *nil* quando $s \in S$, como mostrado na Figura 19-d). Seja $R(s) \in S$ a raiz de $P^*(s)$ a qual pode ser alcançada por $P(s)$. O algoritmo computa, para cada $s \in Z_1$, o custo $V(s)$ de $P^*(s)$, o rótulo $L(s) = \lambda(R(s))$ e o seu predecessor $P(s)$, como segue (Algoritmo1).

Algoritmo 1 – Classificador Supervisionado Baseado em Floresta dos Caminhos Ótimos Usando Grafo Completo

Entrada: Um conjunto de treinamento Z_1 λ -rotulado, protótipos $S \subset Z_1$ e o par (v, d) para vetor de atributos e cálculo das distâncias.

Saída: Floresta dos caminhos ótimos P , mapa de valores de custo de caminhos V e mapa de rótulos L .

Auxiliares: Fila de prioridades Q , e variável cst .

1. **Para todo** $s \in Z_1$, **faça** $P(s) \leftarrow nil$ e $V(s) \leftarrow +\infty$.
2. **Para todo** $s \in S$, **faça**
3. $V(s) \leftarrow 0$, $P(s) \leftarrow nil$, $L(s) = \lambda(s)$
4.   Insira s em Q
5. **Enquanto** Q é não vazia, **faça**
6.   Remova de Q uma amostra s tal que $V(s)$ é mínimo.
7.   **Para cada** $t \in Z_1$ tal que $s \neq t$ e $V(t) > V(s)$, **faça**
8.   Calcule $cst \leftarrow \max \{ V(s), d(s, t) \}$.
9.   **Se** $cst < V(t)$, **então**
10.   **Se** $V(t) \neq +\infty$, **então** remova t de Q .
11. $P(t) \leftarrow s$, $L(t) \leftarrow L(s)$ e $V(t) \leftarrow cst$.
12.   Insira t em Q .

As linhas 1-4 inicializam os mapas e inserem protótipos em Q . O laço principal calcula um caminho ótimo de S para cada amostra $s \in Z_1$ em uma ordem não decrescente de custos (Linhas 5-12). A cada iteração um caminho de custo ótimo $V(s)$ é obtido em P (Linha 6). Empates são resolvidos em Q utilizando a política FIFO (*first-in-first-out*), ou seja, quando dois caminhos atingem uma determinada amostra s com o mesmo custo mínimo, s é associado ao primeiro caminho que o atingiu. O restante das linhas avalia se o caminho que atinge uma amostra adjacente t através de s é mais barato que o caminho que termina em t . Caso positivo, atualiza Q , $P(t)$, $L(t)$ e $V(t)$. No final do algoritmo, V armazena o valor do custo do caminho ótimo de S a cada amostra $s \in Z_1$ de acordo com $f_{\max}$.

c) *Treinamento*

A fase de treinamento do classificador baseado em floresta dos caminhos ótimos usando grafo completo, consiste, basicamente, em encontrar o conjunto S de protótipos, ou seja, os elementos mais representativos de cada classe e iniciar o processo de competição entre eles no conjunto de treinamento. Várias heurísticas poderiam ser adotadas como, por exemplo, uma escolha aleatória de protótipos. Entretanto, tal escolha pode prejudicar o desempenho do classificador, tornando-o instável e com um alto grau de sensibilidade com relação aos protótipos escolhidos. Deseja-se, assim, estimar protótipos nas regiões de sobreposição de amostras e nas fronteiras entre as classes, visto que são regiões muito susceptíveis a erros de classificação.

Computando uma árvore de Espalhamento Mínima (*Minimim Spanning Tree – MST*) no grafo completo (Z_1, A) , é obtido um grafo conexo acíclico cujos nós são todas as amostras em Z_1 , e os arcos são não direcionados e ponderados (Figura 19-b). Seus pesos são dados pela distância d entre os vetores de atributos de amostras adjacentes. Esta árvore de espalhamento é ótima no sentido em que a soma dos pesos de seus arcos é mínima se comparada a outras árvores de espalhamento no grafo completo. Os protótipos a serem escolhidos são os elementos conectados na MST com diferentes rótulos em Z_1 , isto é, elementos mais próximos de classes diferentes (Figura 19-c). Removendo-se os arcos entre classes diferentes, tais amostras adjacentes tornam-se protótipos em S e o Algoritmo 1 pode computar uma floresta de caminhos ótimos em Z_1 (Figura 19-d). Note que uma dada classe pode ser representada por múltiplos protótipos (isto é, árvores de caminhos ótimos) e deve existir pelo menos um protótipo por classe (PAPA; FALCÃO, A. X.; SUZUKI, 2009).

Figura 19 – Gráfico sobre OPF, (a) Grafo completo ponderado nas arestas para um determinado conjunto de treinamento. (b) MST do grafo completo. (c) Protótipos escolhidos como sendo os elementos adjacentes de classes diferentes na MST (nós circutados). (d) Floresta de caminhos ótimos resultante para a função de valor de caminho $f_{\max}$ e dois protótipos. Os identificadores (x,y) acima dos nós são, respectivamente, o custo e o rótulo dos mesmos. A seta indica o nó predecessor no caminho ótimo. (e) Uma amostra de teste (triângulo) da classe 2 e suas conexões (linhas pontilhadas) com os nós do conjunto de treinamento. (f) O caminho ótimo do protótipo mais fortemente conexo, seu rótulo 2 e o custo de classificação 0,4 são associados a amostra teste. Note, que mesmo a mostra de teste estando mais próxima de um nó da classe 1, ela foi classificada como sendo classe 2. (Fonte: PAPA; FALCÃO, A. X.; SUZUKI, 2009)

d) Classificação

Para qualquer amostra $t \in Z_2$, consideram-se todos os arcos conectando t com amostras $s \in Z_1$, tornando t como se fosse parte do grafo original (ver Figura 19 (e) onde a amostra t é representada pelo triângulo no grafo). Considerando todos os possíveis caminhos entre S e t , deseja-se encontrar o caminho ótimo $P^*(t)$ de S até t com a classe $\lambda(R(t))$ de seu protótipo $R(t) \in S$ mais fortemente conexo. Este caminho pode ser identificado incrementalmente, avaliando o valor do custo ótimo $V(t)$ como segue:

$$V(t) = \min\{\max\{V(s), d(s, t)\}\}. \forall s \in Z_1. \quad (4)$$

Seja $s^* \in Z_1$ o nó que satisfaz a equação anterior (isto é, o predecessor $P(t)$ no caminho ótimo $P^*(t)$). Dado que $L(s^*) = \lambda(R(t))$, a classificação simplesmente associa $L(s^*)$ como a classe de t (Figura 19 (f)). Um erro ocorre quando $L(s^*) \neq \lambda(t)$.

2.3.2. Redes Bayesianas

As Redes Bayesianas, também conhecidas como Redes Casuais, Rede de crença e Gráficos de dependência probabilística, surgiram na década de 80 e têm sido aplicadas em uma grande variedade de atividades do mundo real (BOBBIO *et al.*, 2001). Algumas aplicações atuais se estendem às áreas como finanças (CHANG, 2000), saúde (ABICALAFFE; AMARAL; DIAS, 2004) e (KORB; NICHOLSON, 2010), desenvolvimento de jogos (VIEIRA FILHO; ALBUQUERQUE, 2007), entre outras.

As Redes Bayesianas vêm sendo bastante utilizadas em áreas financeiras para a estimação de risco operacional e *credit scoring* (ex: Sistema BayesCredit, um sistema criado por Nykredit, uma das principais empresas no mercado dinamarquês de financiamento imobiliário) e possui vários programas específicos disponíveis como, por exemplo, os softwares Netica (www.norsys.com) e Hugin (www.hugin.com). Segundo Neapolitan (2004), a técnica de Redes Bayesianas surgiu no contexto no qual há um grande número de variáveis e o objetivo de verificar qual a influência probabilística não direta de uma variável para as demais. Assim, a teoria de Redes Bayesianas combina princípios de Teoria de Grafos, Teoria de Probabilidades, Ciência da Computação e Estatística (BEN-GAL, 2008). Além disso, as Redes Bayesianas podem ser consideradas como uma representação visual e informativa da tabela de probabilidade conjunta de todas as variáveis que envolvem o domínio do problema.

A representação das variáveis e das suas relações são realizadas utilizando a Teoria dos Grafos, onde os nós são as variáveis e os arcos são as relações entre elas. Uma rede Bayesiana é composta pelos seguintes itens:

- Um conjunto de variáveis e um conjunto de grafos ligando as variáveis;
- Cada variável tem estados finitos e mutualmente exclusivos;
- As variáveis e os grafos de ligação formam um gráfico direcionado; e
- Cada variável A com pais, $B_1, B_2, \dots, B_n$ possui uma tabela de probabilidades condicionais, $P(A|B_1, B_2, \dots, B_n)$ associada.

A estrutura e os parâmetros de uma rede Bayesiana podem ser determinados manualmente, com apoio de especialistas, ou aprendidos a partir de bases de dados utilizando algoritmos de aprendizado de estrutura e de treinamento.

Um exemplo de fácil compreensão está ilustrado na Figura 20, onde uma determinada doença gera um determinado sintoma e tem como indicadores de risco a idade, a profissão e o clima em que vive o paciente (KARCHER, 2009).

O nó “sintoma” tem o nó “doença” como pai e três ancestrais “idade”, “profissão” e “clima”. Isso representa que a probabilidade de o “sintoma” se manifestar é diretamente dependente da “doença” ocorrer, que é o seu nó pai. Já a probabilidade dessa ocorrer é proporcional a três fatores (nós pais) que são “profissão”, o “clima” e a “idade”. Como esses três parâmetros não possuem nós pais, pode-se afirmar que o “clima” é independente da “idade”, que é independente da “profissão”.

Pode-se inferir ainda que o “sintoma” é indiretamente dependente dos nós ancestrais “idade”, “clima”, e “profissão” através de sua ligação com o nó “doença”. A equação que representa a probabilidade dessa rede Bayesiana utilizada como exemplo é apresentada a seguir:

$$P(i, p, c, d, s) = P(i) * P(p) * P(c) * P(d) * P(d|i, p, c) * P(s|d) \quad (5)$$

onde i = idade, p = profissão, c = clima, d = doença e s = sintoma.

Figura 20 - Exemplo de Rede Bayesiana
(Fonte: KARCHER, 2009)

a) Aprendizagem Bayesiana

A aprendizagem Bayesiana pode ser visto como uma forma de obter a representação interna da rede que define um dado domínio de modo a facilitar a extração do conhecimento. Dentro do processo de aprendizagem, é necessário calcular as distribuições de probabilidade (parâmetros numéricos) e identificar a estrutura da rede, ou seja, identificar variáveis e as relações de interdependência dadas pelos arcos (HRUSCHKA JÚNIOR, 2003).

O processo de aprendizagem é dividido em duas partes:

- aprendizagem da estrutura (e relações entre as variáveis); e
- aprendizagem dos parâmetros numéricos (distribuição de probabilidades).

Alguns algoritmos foram propostos na literatura de Redes Bayesianas, com o objetivo de encontrar a estrutura que represente fielmente o domínio modelado; e algoritmos que determinam as distribuições de probabilidade, considerando aprendizagem indutiva.

Para a aprendizagem da estrutura, (HRUSCHKA JÚNIOR, 2003) observa que existem várias metodologias na literatura, sendo que cada uma aplica-se melhor em um tipo de problema. Por serem bastante específicas, não é possível definir qual a melhor.

b) Classificação Naive Bayes

Uma rede bayesiana pode ser modelada como um classificador, calculando a probabilidade de $P(C|V)$, onde C representa a classe analisada e V o conjunto de variáveis que descrevem os padrões.

O classificador mais importante dentre os classificadores Bayesianos é o *Naive Bayes*, descrito em (DUDA; HART, 1973). É um modelo simples que se destaca pelos sucessos obtidos na aplicação de diversos problemas, mesmo comparado a classificadores mais complexos.

Este modelo descreve um caso particular de uma rede Bayesiana, o qual considera que as variáveis do domínio são condicionalmente independentes, ou seja, uma característica não é relacionada a outra. Em decorrência desta restrição, utiliza-se o termo “naive”. A Figura 21 mostra a estrutura da rede *Naive Bayes*, considerando sua restrição.

Figura 21 – Estrutura da Rede *Naive Bayes*
(Fonte: HRUSCHKA JÚNIOR, 2003)

A classificação é então feita aplicando o teorema de Bayes para calcular a probabilidade da classe C dada uma particular instância de $A_1, A_2, \dots, A_n$ e então predizer a classe com a maior probabilidade *a posteriori* (FRIEDMAN; GEIGER; GOLDSZMIDT, 1997):

$$\text{classificador}(A_1, A_2, \dots, A_n) = \arg \max_c P(c) \prod_i P(A_i|c) \quad (6)$$

O processo de aprendizagem do *Naive Bayes* é feito de maneira indutiva, apresentando um conjunto de dados de treinamento e calculando a probabilidade condicional de cada atributo A_i , dada a classe C (FRIEDMAN; GEIGER; GOLDSZMIDT, 1997). O algoritmo 2, baseado em (CARVALHO, 2014), identifica as etapas do treinamento do *Naive Bayes*.

Algoritmo 2 – Algoritmo de aprendizagem do *Naive Bayes*

Input exemplos para treinamento

1. **Para** cada classe C_j , **faça**
2.   Obtenha a probabilidade incondicional $P(C_j)$;
3.   **Para** cada atributo A_i de um exemplo, **faça**
4.   Obtenha a probabilidade estimada $P(A_i|C_j)$;
5.   **Fim**
6.   **Fim**

A probabilidade incondicional das classes C_j pode ser obtida por meio do conhecimento de um especialista ou atribuindo probabilidades iguais para todas as classes.

Vários outros algoritmos foram desenvolvidos como melhorias do *Naive Bayes*, tais como o *Tree Augmented Naive Bayes* (TAN) (FRIEDMAN; GEIGER; GOLDSZMIDT, 1997), o *BN Augmented Naive Bayes* (BAN), o *General Bayesian Network* (GBN) (CHENG; GREINER, 1999), entre outros.

3. DESENVOLVIMENTO DO ALGORITMO PARA CONSTRUÇÃO DE CURVAS DE CARGA E VALIDAÇÃO

3.1 Algoritmo Desenvolvido

Análises dos dados, tanto do IBGE – características da população, quanto do PROCEL – características de consumo, serviram de base para o desenvolvimento de um programa simulador de curvas de cargas utilizando-se o software MATLAB®. Com o algoritmo foi possível criar um banco de dados consistente para avaliação da influência de cada uma das características da família (número de pessoas, classe econômica, tamanho da residência e faixa etária) no perfil de consumo de cada residência e também propor o estudo de classificação através de sistemas inteligentes.

Devido à dificuldade em se obter dados do consumo relacionados a dados econômicos e sociais, pois as concessionárias devem manter em sigilo essas informações, foi desenvolvido esse algoritmo para gerar as curvas de carga virtuais mensal de consumidores residenciais.

O desenvolvimento do algoritmo de construção de curvas de carga passou por uma ampla análise dos dados apresentados pelo PROCEL sobre os hábitos de consumo de energia dos consumidores residenciais, bem como dos dados socioeconômicos das famílias e a posse de equipamentos das residências, apresentados pelo IBGE no Censo Demográfico de 2010 (IBGE, 2011) e Pesquisa Nacional por Amostra de Domicílios (IBGE, 2013a). Basicamente ele pode ser explicado através de quatro passos, que se seguem:

1º Passo – Característica da Família: são selecionadas características da família de acordo com a região em que ela reside (Norte, Nordeste, Centro-Oeste, Sudeste ou Sul), a quantidade de pessoas que moram na residência, variando de 1 a 5 pessoas. Com esse dado de número de pessoas é montado a composição familiar, podendo ter indivíduos abaixo de quatro anos (neném), de cinco a onze anos (criança), de doze a dezoito anos (adolescente), de dezenove a cinquenta e nove anos (adulto) e acima de sessenta anos (idoso), tomando o cuidado para que bebês e crianças não ficassem sozinhos em casa.

Outra característica selecionada é a faixa econômica da família. Para isso, é realizado um sorteio das pessoas que podem trabalhar, quantas realmente trabalham, determinando uma faixa salarial familiar, classificadas em cinco classes (A, B, C, D e E). Além dessas características, o tamanho da residência também é selecionado de acordo com a área de construção que é dividida em 12 possíveis tamanhos. Por fim, o período em que a família se encontra em casa (manhã, tarde, noite e/ou madrugada) é determinado.

A seleção dessas características é realizada através de sorteios usando a função ‘randi’ da biblioteca do MATLAB, assim, qualquer combinação pode ocorrer. No entanto, esse sorteio pode ser controlado, para que possa ser atingido valores médios próximos das pesquisas realizadas pelo IBGE e PROCEL, como o que foi realizado nessa tese.

Figura 22 – Algoritmo de Construção de Curva de Carga, 1ºPasso: Característica da Família

2º Passo – Posse de Equipamentos: foram selecionados um total de 43 combinações de equipamento/local, como TV da sala e TV do quarto, ou Lâmpada do banheiro, Lâmpada da cozinha, ferro de passar, chuveiro, entre outros. Para determinar a posse ou não do equipamento, são realizados para cada um dos aparelhos, sorteios de pontos para as características que influenciam na posse de equipamentos (região, número de pessoas, composição, faixa econômica e tamanho da residência), e através da soma desses pontos e a comparação com um parâmetro pré-determinado, é indicado se o equipamento existe ou não na residência. Esses parâmetros pré-determinados se baseiam em dados de posse de equipamentos apresentados no estudo do Procel, SINPHA – Sistema de Informação de Posses

e Hábitos de Uso de Aparelhos Elétricos (PROCEL, 2007). Por fim, é gerado um vetor com todos os equipamentos que a família possui. Com esse vetor, segue-se para o terceiro passo.

Figura 23 – Algoritmo de Construção de Curva de Carga, 2ºPasso: Posse de Equipamento

3º Passo – Curva de Carga de cada Aparelho: para cada aparelho que a família possui e de acordo com as características que influenciam nos hábitos familiares, são sorteados tempo de uso e horário de uso, formando a curva de carga de cada equipamento. Tanto o sorteio de tempo de uso e horário do uso, além das características familiares, eles seguem o padrão dos estudos médios apresentados pelo Procel, no SINPHA (PROCEL, 2007), ou seja, a probabilidade de ser sorteado um horário próximo da média é maior do que outro horário, porém, como cada indivíduo possui o livre arbítrio, o sorteio pode definir um horário e um tempo totalmente diferente da média. A curva de carga de cada aparelho, nada mais é que um vetor de noventa e seis posições no qual cada um representa a potência média consumida nos últimos quinze minutos, totalizando as vinte e quatro horas do dia. Após ser gerado a curva de cada aparelho têm-se o quarto passo.

Figura 24 – Algoritmo de Construção de Curva de Carga, 3ºPasso: Curva de Carga de cada Aparelho

4º Passo – Curva de Carga da Residência: soma-se a curva de carga de cada aparelho, formando a curva de carga diária da residência. Para obter a curva de carga mensal, basta somar as curvas de carga de cada dia do mês, levando-se em consideração o dia da semana e as diferenças entre dia da semana e finais de semana.

Figura 25 – Algoritmo de Construção de Curva de Carga, 4ºPasso: Curva de Carga da Residência

A seguir, está descrito um resumo de como imagina-se que cada uma das variáveis influenciará na construção de curva de carga individualmente.

a) Região do País

Cada região do país possui uma característica média única de consumo de energia que foi levada em consideração no desenvolvimento do algoritmo. Além disso, estudos apontam quais são os equipamentos mais utilizados em cada uma das cinco regiões e o que influencia

na escolha dos aparelhos que farão parte de cada residência simulada. Essa variável se relacionará com todas as outras, pois é a região que aponta os costumes, os incentivos do governo, o clima, a cultura, entre outros.

b) Dia da Semana

O dia da semana está diretamente relacionado com uso da energia nas residências, pois o consumo de energia está relacionado com o tempo de uso dos equipamentos. Assim, em geral, o consumo é bem parecido nos dias da semana, e diferentes nos finais de semana. Essa variável estará mais relacionada com o modo do uso da energia de cada residência. É possível estabelecer uma relação dessa variável com a faixa etária dos integrantes da família, uma vez que os hábitos das pessoas variam de acordo com a faixa etária.

c) Número de Pessoas

O número de pessoas está ligado tanto à posse de equipamentos como o modo de uso dos equipamentos. As residências possuem diferentes médias de número de pessoas em cada região, além disso, o aumento do número de pessoas faz com que diferentes medidas sejam tomadas na escolha de aquisição de novos aparelhos de acordo com a região. Ou seja, observa-se uma grande relação dessa variável com a variável da Região, e também pode-se considerar que ela se relacionará com a classe econômica.

d) Faixa Etária

Esta variável está relacionada com os hábitos de consumo das residências e de certa forma também com a aquisição dos equipamentos. Possui ligação com a variável de número de pessoas, e estudos do IBGE mostram a distribuição por faixa etária por região do país, dados que foram considerados no algoritmo.

Para utilizar e avaliar esta variável no algoritmo foi pressuposto que em uma residência deve existir ao menos uma pessoa com renda, e que possa viver sozinho. Foi considerado a possibilidade de adolescentes morarem sozinhos, visto que é grande o número de jovens que vão morar sozinhos nos primeiros anos de faculdade.

Assim, foi estabelecido um total de 42 combinações possíveis para serem avaliadas.

Neste primeiro momento não foram diferenciados os sexos das pessoas, porém sabe-se que existem diferenças de perfil de consumo entre homens e mulheres, justificadas pela forma que cada um consome energia, e os equipamentos que um usa mais que o outro.

e) Classe Social/Econômica

A classe social/econômica está relacionada tanto com a posse de equipamentos, como com aos hábitos de uso. Através de dados do IBGE, que mostram as classes econômicas em cada região, a posse de equipamentos e o consumo médio de uma residência, é possível relacionar essa variável com todas as outras variáveis estudadas.

f) Tamanho da Residência

Esta variável possui relação com a classe econômica e com o número de pessoas de forma direta. Mas relações podem ser estabelecidas em relação a posse de equipamentos e também a região, pois dados apontam as médias de construções em cada região. Outra relação importante que essa variável irá fornecer ao algoritmo diz respeito ao número de cômodos e consequentemente a iluminação necessária na residência.

3.2 Curvas Geradas pelo Algoritmo

Para demonstrar o funcionamento do algoritmo é apresentado curvas de cada uma das cinco regiões do país, levando-se em consideração características como classe econômica C, número de pessoas igual a dois. As características: permanência em casa, faixa etária e tamanho da residência, serão aleatórios. É apresentado dez curvas diárias aleatórias e a média destas dez curvas (GASTALDELLO *et al.*, 2014).

É importante destacar que apesar de existirem duas características fixas (classe econômica e número de pessoas), as outras características (faixa etária, tamanho da residência e período em casa) influenciam bastante na posse de equipamentos e hábitos de consumo, fazendo com que existam curvas de diferentes perfis, e com diferentes classes de consumo.

Para se obter as curvas mensais, foi implementado no algoritmo a função de calendário, onde ao se escolher um mês e um ano, o algoritmo gera vinte e oito ou vinte e nove (fevereiro), trinta (abril, junho, setembro e novembro), trinta e uma (janeiro, março, maio, julho, agosto, outubro e dezembro), curvas e tira a média, diferenciando curvas de dia de semana e finais de semana. As Figuras 26 a 30 apresentam curvas de demanda diária das cinco regiões do país, são apresentadas dez curvas geradas aleatoriamente e a curva média destas dez curvas.

Figura 26 – Curvas de carga residencial (classe C, duas pessoas) geradas pelo algoritmo – Norte

Figura 27 – Curvas de carga residencial (classe C, duas pessoas) geradas pelo algoritmo – Nordeste

Figura 28 – Curvas de carga residencial (classe C, duas pessoas) geradas pelo algoritmo – Centro-Oeste

Figura 29 – Curvas de carga residencial (classe C, duas pessoas) geradas pelo algoritmo – Sudeste

Figura 30 – Curvas de carga residencial (classe C, duas pessoas) geradas pelo algoritmo – Sul

Destaca-se nas curvas geradas que existem diferentes tipos de perfis de consumo, que representam o livre arbítrio de cada consumidor. Porém, as regiões possuem semelhanças, como por exemplo, as regiões Norte e Nordeste, possuem consumo alto nas madrugadas, devido ao uso do ar condicionado. As regiões Centro-Oeste e Sudeste, possuem dois picos bem definidos de manhã e no início da noite, devido principalmente ao uso do chuveiro elétrico, e o Sul, apresenta semelhanças com os dois grupos citados, possuindo picos de manhã e de noite, e também consumo de madrugada, devido ao uso de condicionador de ambiente, principalmente no sentido inverso, para aquecer a residência (GASTALDELLO *et al.*, 2015).

3.3 Elaboração do Banco de Dados e Validação do Algoritmo

Para validação do algoritmo foi proposta uma comparação entre médias de curvas geradas pelo algoritmo, com as médias apresentadas pela pesquisa do Procel e também comparando com medições em campo.

O banco de dados gerados pelo algoritmo contem dez mil curvas de carga de consumidores residenciais para cada uma das cinco regiões do país. Esse banco de dados foi gerado considerando dados apresentados pelo PNAD - Pesquisa Nacional por Amostra de Domicílios, referentes à quantidade de pessoas na residência e faixa etária dos moradores, conforme as Tabelas 5 e 6.

Como a proposta é comparar médias, foi necessário levar em consideração esses aspectos de número de habitantes e faixa etária, pois assim, o banco de dados fica proporcional as características existentes em cada uma das regiões. O intuito é que as médias geradas pelo algoritmo fiquem próximas das médias criadas pela extrapolação de dados medidos no estudo realizado pelo Procel.

Tabela 5 – Percentual de Número de Habitantes por Residência

Região	Valor Percentual de nº de Habitantes				
	1	2	3	4	5 ou +
Norte	9,7	18,2	22,5	21,7	27,9
Nordeste	11,2	21,2	24,9	21,6	21,1
Centro-Oeste	13,8	24,5	25,1	21,8	14,8
Sudeste	13,7	25,5	26,1	20,6	14,0
Sul	13,2	26,5	27,7	20,3	12,4

Fonte: IBGE. “PNAD 2013 – Pesquisa Nacional por Amostra de Domicílios”, Rio de Janeiro, 2013.

Tabela 6 – Percentual por Faixa Etária dos Habitantes da Residência

Região	Valor Percentual de habitantes por faixa etária			
	0-24	25-39	40-59	+60
Norte	49	23,5	19,6	7,9
Nordeste	43,7	23,2	21,5	11,5
Centro-Oeste	40,6	25,6	23,7	10,0
Sudeste	37	24,1	25,6	13,3
Sul	37,5	22,9	26,5	13,1

Fonte: IBGE. “PNAD 2013 – Pesquisa Nacional por Amostra de Domicílios”, Rio de Janeiro, 2013.

Desse banco de dados com dez mil curvas foi calculada uma média de consumo diária para cada uma das cinco regiões, essa média é comparada levando-se em consideração cinco

aspectos, mínimo erro percentual, máximo erro percentual, erro médio relativo, erro médio absoluto e o erro percentual de consumo final (mensal).

As equações utilizadas para cálculos de erros percentuais de máximo e mínimo, além de erro de consumo percentual mensal estão descritos a seguir:

$$E_{\min\%} = \left(\min_{i=1:96} \left| \frac{V_a(i) - V_P(i)}{V_a(i)} \right| \right) \times 100 \quad (7)$$

$$E_{\max\%} = \left(\max_{i=1:96} \left| \frac{V_a(i) - V_P(i)}{V_a(i)} \right| \right) \times 100 \quad (8)$$

$$E_{\text{méd},\text{Rel}} = \left(\sum \frac{V_a(i) - V_P(i)}{V_a(i)} \right) \times 100 \quad (9)$$

$$E_{\text{méd},\text{Abs}} = \left(\sum \left| \frac{V_a(i) - V_P(i)}{V_a(i)} \right| \right) \times 100 \quad (10)$$

$$C_m = \frac{(\sum_{i=1}^{96} V_d) \times 30}{4} \quad (11)$$

$$E_{\text{consumo}\%} = \left| \frac{(C_{m,a} - C_{m,P})}{C_{m,a}} \right| \times 100 \quad (12)$$

Onde:

- $E_{\min\%}$ = erro percentual mínimo;
- $E_{\max\%}$ = erro percentual máximo;
- $E_{\text{méd},\text{Rel}}$ = erro percentual médio relativo;
- $E_{\text{méd},\text{Abs}}$ = erro percentual médio absoluto;
- $E_{\text{consumo}\%}$ = erro percentual consumo;
- $V_a(i)$ = i-ésimo valor do vetor de consumo do algoritmo;
- $V_P(i)$ = i-ésimo valor do vetor de consumo do Procel;
- V_d = vetor diário;
- C_m = consumo mensal;
- $C_{m,a}$ = consumo mensal do algoritmo; e
- $C_{m,P}$ = consumo mensal do Procel.

Os resultados atingidos dessa comparação são apresentados na Tabela 7:

Tabela 7 – Erro percentual entre as Médias do Algoritmo e as Médias Procel

Região	Erro Consumo(%)	Mínimo	Máximo	Erro Médio	Erro Médio
		Erro(%)	Erro(%)	Rel. (%)	Abs. (%)
Norte	1,58	0,17	33,91	1,34	13,52
Nordeste	2,07	0,06	57,02	-1,81	15,08
Centro-Oeste	0,58	0,11	58,73	-1,10	19,83
Sudeste	1,67	0,54	87,32	0,74	19,69
Sul	4,68	0,33	63,65	6,45	19,38

Observa-se que apesar de um erro percentual máximo alto, o erro percentual de consumo é baixo, assim como os erros médios absolutos e relativos, o que significa que as curvas médias estão próximas. A seguir são apresentadas as Figuras 31 a 35 de comparação entre as médias do Procel e do banco de dados, para cada uma das regiões.

Figura 31 – Comparação entre as curvas médias do algoritmo e do Procel-PNAD – Região Norte

Para a região Norte o algoritmo de geração de curvas de carga conseguiu acompanhar a curva média do estudo do Procel. Tendo alguns pontos de maior consumo e outros de menor consumo.

Figura 32 – Comparação entre as curvas médias do algoritmo e do Procel-PNAD – Região Nordeste

Já para a região Nordeste o algoritmo de geração de curvas de carga conseguiu acompanhar a curva média do estudo do Procel na maior parte do tempo, tendo um pequeno desvio no final do dia, porque o pico máximo do algoritmo foi atingido depois do pico da média do Procel, causando um erro relativo alto.

Figura 33 – Comparação entre as curvas médias do algoritmo e do Procel-PNAD – Região Centro-Oeste

Na região Centro-Oeste observa-se um erro maior no início do dia. Enquanto o algoritmo estima um alto consumo nas primeiras horas do dia, a média do Procel considera um consumo menor. Essa diferença foi ocasionada, possivelmente, devido os condicionadores de ambiente, que são utilizados durante a madrugada.

Figura 34 – Comparação entre as curvas médias do algoritmo e do Procel-PNAD – Região Sudeste

A região Sudeste é a que teve o maior erro percentual, porém observa-se que o algoritmo segue uma tendência de acompanhamento da média do estudo do Procel.

Figura 35 – Comparação entre as curvas médias do algoritmo e do Procel-PNAD – Região Sul

E por fim, a região Sul que apresenta o segundo maior erro percentual máximo, mas que também segue a tendência de acompanhar o estudo do Procel.

Finalizando a validação com dados médios, foi proposto avaliar a comparação com dados medidos em campo. Duas residências participaram do estudo de comparação com medidas em campo. A primeira residência (Estudo 1) residiam dois adultos de classe econômica C. Já na segunda residência (Estudo 2) residia uma família de classe econômica B, composta por dois idosos e um adulto, porém, essa residência foi considerada como atípica,

pois as medidas foram realizadas em cargas não usuais, localizadas no quarto do fundo da residência, aonde se localizava um laboratório de mecânica.

Para o estudo com curvas medidas em campo, foi gerado um banco com dez mil curvas com as mesmas características da residência em que foram realizados os ensaios. Esses dados são comparados e são selecionadas as cem curvas (1% do banco de dados) mais próximas da curva medida em relação à soma dos desvios absolutos ponto a ponto. Desse novo banco selecionado é tirado a média. Por fim, os dados selecionados são comparados com a curva medida nas residências através de erros percentuais máximos, mínimos, médios relativos, médios absolutos e erro percentual de consumo. São apresentados os resultados da comparação entre a média, curva de menor desvio, curva de menor diferença de potência em relação à curva medida.

Os erros apresentados pela curva média dos dados selecionados, a curva de desvio mínimo, e a curva de diferença de potência mínima, em relação às curvas medidas podem ser vistos na Tabelas 8 e 9.

Tabela 8 – Análise Comparativa entre curvas medidas e geradas pelo algoritmo – ESTUDO 1

Estudo 1					
	Erro Consumo	Máximo Desvio	Mínimo Desvio	Erro Médio	Erro Médio
	(%)	(%)	(%)	Rel. (%)	Abs. (%)
Média	5,554	550,8	0,492	-35,78	53,01
Menor Desvio	8,413	509,6	0,184	-28,45	46,55
Menor Pot.	0,141	592,1	2,045	-34,36	64,58

Tabela 9 – Análise Comparativa entre curvas medidas e geradas pelo algoritmo – ESTUDO 2

Estudo 1					
	Erro Consumo	Máximo Desvio	Mínimo Desvio	Erro Médio	Erro Médio
	(%)	(%)	(%)	Rel. (%)	Abs. (%)
Média	13,81	81,14	0,274	-95,76	112,25
Menor Desvio	34,09	80,12	0,518	-67,17	89,34
Menor Pot.	8,853	87,73	3,326	-55,19	108,13

A seguir são apresentadas as Figuras 36 e 37 das análises realizadas para cada um dos dois estudos.

Figura 36 – Comparação entre as curva medida em campo e curvas geradas pelo algoritmo – Estudo 1

Observa-se que as curvas selecionadas acompanham a tendência da curva medida, porém, elevados erros percentuais são registrados devido principalmente a forma que foi registrada as medidas. Enquanto o algoritmo segue a regra de consumo de quinze em quinze minutos, as medidas foram realizadas de minuto em minuto, assim, muitas oscilações ocorrem, causando erros significativos. Outra explicação é de que o algoritmo utiliza-se da potência máxima das cargas quando essas estão ligadas, já nas medições é representado o consumo real, ou seja, na maioria das cargas ocorre uma oscilação durante o uso. Esses erros podem ser solucionados com a modelagem das cargas, bem como, com o conhecimento exato das cargas existentes na residência, porém, não é esse o intuito apresentado por nessa tese.

Figura 37 – Comparação entre as curva medida em campo e curvas geradas pelo algoritmo – Estudo 2

Para o segundo estudo o erro de consumo final foi maior, porém os erros máximos registrados foram menores, no entanto, os erros médios foram maiores, o que mostra que as curvas geradas pelo algoritmo não conseguiram acompanhar essa curva atípica medida. Como o algoritmo segue uma média, e possui equipamentos comuns, essa residência com cargas atípicas, e com consumo diferenciado obteve resultados ruins, como era de se esperar.

4. ANÁLISE DAS CARACTERÍSTICAS SOCIOECONÔMICAS

Para o estudo e a análise de cada uma das variáveis de formação familiar (número de pessoas, faixa etária/composição, classe econômica e tamanho da residência) foi criado um banco de dados com curvas de carga mensais de consumidores gerados pelo algoritmo. Essas curvas são classificadas de acordo com o consumo em classes de consumo. Comumente as concessionárias utilizam cinco classes de consumo (A, B, C, D e E), correspondente à média de consumo da residência, conforme a Tabela 3 apresentada anteriormente. Contudo, para esse estudo é proposto a utilização de treze classes (A+, A-, B+, B-, C+, C-, D+, D-, E+ e E-), com intuito de melhorar a visualização de mudança de faixa de consumo. A Tabela 10 apresentam as treze classes de consumo propostas (GASTALDELLO, *et al.*, 2016).

Tabela 10 – Divisão por faixa de consumo proposto

Consumo Mensal (kWh)	Perfil (Classe)
< 80	E-
≥ 80 e < 100	E+
≥ 100 e < 130	D-
≥ 130 e < 160	D
≥ 160 e < 200	D+
≥ 200 e < 230	C-
≥ 230 e < 260	C
≥ 260 e < 300	C+
≥ 300 e < 370	B-
≥ 370 e < 440	B
≥ 440 e < 500	B+
≥ 500 e < 700	A-
≥ 700	A+

A seguir são apresentados os resultados dos estudos da influência de cada uma das variáveis estudadas sobre a classificação de consumo de energia elétrica.

4.1 Estudo da Influência do Número de Pessoas

Nesse estudo, o algoritmo foi configurado para gerar, para cada região, cinco bancos de dados. Cada banco de dados é composto de cinco mil curvas, sendo que mil dessas são de um integrante familiar, mil são de dois integrantes, mil são de três integrantes, mil de quatro e mil de cinco integrantes. Optou-se por avaliar qual a influência da variação do número de pessoas sobre a classe/perfil de consumo em função da classe econômica, ou seja, cada um

dos cinco bancos de dados para cada região representa uma classe econômica (A, B, C, D e E).

Cada uma das curvas mensais é classificada conforme a Tabela 10 de perfil de consumo apresentada anteriormente. Posteriormente é analisada a quantidade de curvas existentes em cada patamar e com isso é possível avaliar um comportamento devido à variação do número de pessoas.

4.1.1. Região Norte – Número de Pessoas

A Figura 38 apresenta os resultados obtidos da Região Norte, optou-se por colocar todas as classes em uma mesma figura para melhorar a avaliação da influência do número de pessoas no consumo médio da residência para esta região. O eixo das abscissas representa as classes de consumo e o eixo das ordenadas representa a quantidade de curvas.

Figura 38 – Análise estatística da influência da variação do número de pessoas sobre o consumo – Norte

É possível observar na figura anterior que o aumento do número de pessoas, aumenta o consumo e consequentemente, a mudança de classes de consumo. Porém, é interessante observar, que cada classe econômica possuiu uma distribuição diferente na medida em que se varia o número de pessoas.

A classe econômica E, por exemplo, possui na ordem de 600 curvas nos perfis de consumo (E-, E+, D- e D), respectivamente para 1, 2, 3 e 5 integrantes na família. Ou seja, praticamente o aumento de um integrante, faz com que se suba um nível de classe de consumo.

Já na classe econômica D há basicamente dois picos (500 e 400), sendo o primeiro para uma pessoa, classe E+ e de cinco pessoas, classe A-. As demais configurações se diluem em todas as classes.

A tendência é que à medida que se aumenta as classes econômicas, as residências com cinco integrantes passem a consumir muito mais, como é demonstrado na classe econômica A, onde 800 curvas são classificadas como classe de consumo A+ para cinco integrantes.

A Tabela 11 apresenta a média de consumo total de cada classe econômica estudada, e a média de acordo com o número de pessoas.

Tabela 11 – Consumo médio na avaliação da variação do Número de Integrantes - Norte

	Média Total (kWh)	Média (kWh) n = 1	Média (kWh) n = 2	Média (kWh) n = 3	Média (kWh) n = 4	Média (kWh) n = 5
Classe Econômica A	579,66	256,88	444,02	591,95	750,42	855,03
Classe Econômica B	531,42	177,56	377,19	539,19	719,29	843,88
Classe Econômica C	373,86	147,13	215,16	333,58	534,35	639,05
Classe Econômica D	278,90	117,75	168,35	221,37	386,94	500,08
Classe Econômica E	112,65	78,65	98,98	113,33	131,84	140,48

Observando a tabela fica mais fácil perceber que em classes econômicas maiores, a mudança de número de integrantes impacta bastante no consumo. Isso pode ser explicado devido ao poder aquisitivo de uma família maior e com melhores condições econômicas, eles pagam por melhorar o conforto da família.

Os valores elevados apresentados na tabela, não significa que o consumo médio da região é o mais alto no país, essa é uma análise estatística distribuída igualmente em número de integrantes, e classes econômicas. Pois, sabe-se que no Norte há famílias maiores que no Sudeste, mas que existem mais famílias com menores condições econômicas.

4.1.2. Região Nordeste – Número de Pessoas

A Figura 39 mostra os resultados alcançados para a região Nordeste.

Figura 39 – Análise estatística da influência da variação do número de pessoas sobre o consumo – Nordeste

Observa-se que as mudanças são bem parecidas com a região Norte, onde a classe econômica E possui perfis na ordem de 600 curvas, a classe D bem distribuída, tendo uma pequena diferença de perfil nas classes econômicas C e B, enquanto que no Norte tem picos de 600, o Nordeste tem patamares de 400.

Pressupõe-se que essas diferenças existem devido ao clima diferenciado e também ao poder de compra e o mercado diversificado do Nordeste, o que influencia no aumento da média de consumo e distribuição das curvas.

A Tabela 12 apresenta a média de consumo total de cada classe econômica estudada, e a média de acordo com o número de pessoas.

Tabela 12 – Consumo médio na avaliação da variação do Número de Integrantes - Nordeste

	Média Total (kWh)	Média (kWh) n = 1	Média (kWh) n = 2	Média (kWh) n = 3	Média (kWh) n = 4	Média (kWh) n = 5
Classe Econômica A	633,90	279,14	503,59	632,65	826,49	927,64
Classe Econômica B	584,75	192,16	433,79	582,57	805,47	909,76
Classe Econômica C	426,86	162,97	245,45	374,93	613,66	737,27
Classe Econômica D	322,57	132,02	192,62	258,89	442,95	586,36
Classe Econômica E	131,21	87,55	117,35	136,31	152,99	161,83

Observa-se que as médias do Nordeste realmente são maiores que a do Norte, como foi pressuposto.

4.1.3. Região Centro-Oeste – Número de Pessoas

Para a região Centro-Oeste têm-se a Figura 40 para análise da variação do número de pessoas:

Figura 40 – Análise estatística da influência da variação do número de pessoas sobre o consumo – Centro-Oeste

Observa-se uma mudança de comportamento de distribuição das curvas já na classe econômica E, pois para esta região Centro-Oeste os picos são maiores, ou seja, a distribuição é menor, além disso, já está deslocada para a esquerda, ou seja, maiores consumos em classes econômicas mais baixas.

A distribuição da classe econômica D e C são mais dispersas que nas regiões anteriores, enquanto que nas classes econômicas B e A, a variação das pessoas distribui um pouco mais as curvas entre as classes de consumo, criando menos picos do que visto nas regiões Norte e Nordeste. A Tabela 13 apresenta a média de consumo total de cada classe econômica estudada, e a média de acordo com o número de pessoas.

Tabela 13 – Consumo médio na avaliação da variação do Número de Integrantes – Centro-Oeste

	Média Total (kWh)	Média (kWh) n = 1	Média (kWh) n = 2	Média (kWh) n = 3	Média (kWh) n = 4	Média (kWh) n = 5
Classe Econômica A	581,36	209,13	458,78	606,30	770,47	862,11
Classe Econômica B	496,93	184,21	363,75	479,45	670,09	787,18
Classe Econômica C	334,85	171,08	236,85	290,22	432,71	543,39
Classe Econômica D	269,12	149,93	217,39	247,79	336,85	393,64
Classe Econômica E	173,94	114,42	159,35	179,10	201,55	215,31

A diferença apresentada na tabela dessa região em relação as anteriores está atrelada ao fato de o consumo ser maior nas classes econômicas menores, e a média menor nas classes econômicas maiores.

4.1.4. Região Sudeste – Número de Pessoas

O estudo da influência da variação do número de pessoas sobre o consumo de consumidores residências para a região Sudeste é apresentada na Figura 41:

Figura 41 – Análise estatística da influência da variação do número de pessoas sobre o consumo – Sudeste

Com o clima mais ameno do que as outras regiões estudadas, a região Sudeste apresenta maior distribuição das curvas nas classes de consumo devido a variação do número de pessoas. Como o maior “vilão” dessa região são os chuveiros elétricos, a tendência é de que o aumento do número de pessoas, aumente o número de banhos e consequentemente o aumento o consumo, em uma relação mais direta.

É possível observar que em todas as classes econômicas o aumento/variação do número de pessoas faz com que as curvas se distribuam entre as classes, todas as classes de consumo possuem curvas em todas faixas de integrantes.

A Tabela 14 apresenta a média de consumo total de cada classe econômica, e a média de acordo com o número de pessoas.

Tabela 14 – Consumo médio na avaliação da variação do Número de Integrantes – Sudeste

	Média Total (kWh)	Média (kWh) n = 1	Média (kWh) n = 2	Média (kWh) n = 3	Média (kWh) n = 4	Média (kWh) n = 5
Classe Econômica A	583,27	243,09	446,99	587,68	752,88	885,69
Classe Econômica B	529,25	219,57	386,96	507,11	689,37	843,28
Classe Econômica C	402,80	211,99	290,04	352,01	497,15	662,81
Classe Econômica D	349,94	194,25	273,59	320,24	429,35	532,29
Classe Econômica E	220,16	131,38	190,43	221,69	266,42	290,89

É importante mencionar que a distribuição das curvas devido à mudança do número de integrantes ocorre em todas as classes econômicas de forma rápida, apresentando alargamento da distribuição, ou seja, o número de pessoas tem bastante impacto nessa região.

4.1.5. Região Sul – Número de Pessoas

Com um clima bem definido entre as estações, a região Sul possui características comuns tanto ao norte, nordeste, quanto ao sudeste e centro-oeste, referente aos equipamentos que mais consomem. Pois nessa região há o uso em grande escala de condicionadores de ambiente (para esquentar) e os chuveiros elétricos, e por isso, esta região é a que mais consome, em média, no Brasil. Conforme é observado na Figura 42:

Figura 42 – Análise estatística da influência da variação do número de pessoas sobre o consumo – Sul

Observa-se que os picos de quantidades de curvas são maiores que as demais apresentadas, e que estão distribuídas, bem parecido com o Sudeste, devido principalmente ao uso do chuveiro, porém ao contrário da região Sudeste, quanto mais alta a classe econômica,

as famílias investem também nos condicionadores de ambiente, aumentando assim significativamente a quantidade de curvas nos perfis de consumo A, A+, B+, B-.

A Tabela 15 mostra a média de consumo total de cada classe econômica e a média de acordo com o número de pessoas.

Tabela 15 – Consumo médio na avaliação da variação do Número de Integrantes – Sul

	Média Total (kWh)	Média (kWh) n = 1	Média (kWh) n = 2	Média (kWh) n = 3	Média (kWh) n = 4	Média (kWh) n = 5
Classe Econômica A	853,21	358,49	684,91	857,35	1115,85	1249,42
Classe Econômica B	793,72	250,21	591,60	787,72	1090,70	1248,35
Classe Econômica C	589,10	179,36	344,43	549,78	859,89	1015,06
Classe Econômica D	454,28	153,44	245,02	366,44	656,62	849,88
Classe Econômica E	197,63	120,19	173,83	211,57	225,17	257,37

Verifica-se que a média desta região é maior do que as outras e que com aumento da classe econômica, a variação do número de integrantes aumenta drasticamente o consumo, uma vez que, não é apenas o número de banhos que aumentam, mas também o condicionamento de ambientes.

4.1.6. *Influência da variação do número de pessoas sobre o consumo*

Essa característica é mais impactante nas regiões onde o uso de chuveiros elétricos é mais presente, pois a tendência é de que o aumento do número de pessoas seja proporcional ao aumento do número de banhos.

Para mostrar a influência dessa característica é apresentado um gráfico conhecido como *BoxPlot*. Nele estão representados os valores, mínimo, máximo, primeiro quartil, terceiro quartil, mediana e média. Com isso, é possível identificar o comportamento de mudança de consumo de acordo com a variação do número de pessoas, em cada região.

A Figura 43 apresenta esse gráfico para cada uma das cinco regiões, sendo levado em consideração a classe econômica C, que é a mais representativa do país.

Figura 43 – Gráfico de BoxPlot da Classe Econômica C para a Variação de Número de Pessoas

Conforme discutido anteriormente, com o aumento do número de pessoas, em todas as regiões há um aumento de consumo, observado pelo aumento da média. Mas analisando o gráfico é possível perceber que para as regiões Norte e Nordeste esse aumento tem características semelhantes, com crescimento mais acentuado na variação de 4 para 5 integrantes, possivelmente devido às melhores condições econômicas conforme aumenta o número de integrantes. Já na região Sudeste e Centro-Oeste tem característica de crescimento constante em todas variações de mudança de integrantes. E por fim, no Sul é composta, isto é observado pelas médias e também pela distribuição das curvas, vista pelo tamanho dos retângulos formados entre o primeiro e o terceiro quartil, que representam a concentração de 50% das amostras (entre 25%-primeiro quartil e 75%-terceiro quartil).

Para essa variável, as regiões Norte e Nordeste possuem bastante semelhanças nas distribuições de curvas, enquanto as regiões Sudeste e Centro-Oeste se assemelham um pouco, porém devido ao clima mais seco no Centro-Oeste, para classes econômicas maiores a influência da variação do número de integrantes é um pouco diferenciada da Sudeste, e por fim, a região Sul possui características das quatro outras regiões no que se refere a distribuição de curvas. Pressupõe-se que essa característica impacta mais diretamente as regiões Sudeste e Centro-Oeste, pois o “vilão” de consumo nas residências é o chuveiro elétrico, e esse é um bem individual.

4.2 Estudo da Influência da Classe Econômica

Neste estudo, o algoritmo foi configurado para gerar, para cada região, cinco bancos de dados. Cada banco de dados é composto de cinco mil curvas, sendo que mil destas são de

famílias com classe econômica E, mil são de classe econômica D, mil são de classe econômica C, mil classe D e mil E. Optou-se por avaliar qual a influência da variação da classe econômica sobre a classe/perfil de consumo em função do número de pessoas, ou seja, cada um dos cinco bancos de dados para cada região representa uma quantidade de integrantes da família (1, 2, 3, 4 e 5).

Cada uma das curvas mensais é classificada conforme a Tabela 10 de perfil de consumo. Posteriormente é analisada a quantidade de curvas existentes em cada patamar e com isto é possível avaliar um comportamento devido à variação da classe econômica.

4.2.1. Região Norte – Classe Econômica

Esse estudo apresenta os gráficos juntos para cada um dos números de integrantes, com intuito de verificar a influência da variação da classe econômica sobre o consumo de energia em cada uma das configurações.

Diferentemente da variação do número de integrantes, a variação da classe econômica para a região Norte é um pouco mais impactante. Conforme citado anteriormente, o eletrodoméstico que mais consome energia nesta região, segundo as pesquisas do Procel, são os condicionadores de ar. Esse equipamento é caro, e por isso é necessário ter melhores condições econômicas para adquiri-los. A Figura 44 apresenta a análise de quantidade de curvas classificadas em classes de consumo em função da variação da classe econômica em cada configuração de número de pessoas.

Figura 44 – Análise estatística da influência da variação da classe econômica sobre o consumo – Norte

Observa-se que o aumento da classe econômica, independentemente do número de pessoas faz com que o consumo cresça de forma bem distribuída, com um acúmulo de curvas nas classes de consumo altas para a configuração de cinco integrantes na família.

Um detalhe importante é que famílias de classe econômica E se concentram na faixa de consumo D em todas configurações.

A Tabela 16 apresenta os valores de consumo médio para cada uma das configurações.

Tabela 16 – Consumo médio na avaliação da variação da Classe Econômica – Norte

	Média Total (kWh)	Média (kWh) Classe E	Média (kWh) Classe D	Média (kWh) Classe C	Média (kWh) Classe B	Média (kWh) Classe A
Número de Pessoas = 1	155,31	78,38	120,57	146,82	178,18	252,56
Número de Pessoas = 2	259,37	99,13	162,89	216,91	376,23	441,72
Número de Pessoas = 3	356,49	114,27	217,87	322,54	530,24	597,51
Número de Pessoas = 4	503,01	132,85	382,85	535,78	712,96	750,63
Número de Pessoas = 5	594,66	140,86	503,73	642,42	830,86	855,96

Verifica-se que ao se aumentar a classe econômica, aumenta-se a classe de consumo, consequentemente a média de consumo. A variação da média de consumo total se difere da variação média total apresentada para o número de integrantes, os extremos, ou seja, um integrante e cinco integrantes nessa tabela, mostrando que o impacto da classe econômica são significativos nessa configuração.

4.2.2. Região Nordeste – Classe Econômica

A Figura 45 apresenta os resultados deste estudo de variação de classe econômica para a região Nordeste:

Figura 45 – Análise estatística da influência da variação da classe econômica sobre o consumo – Nordeste

Assim como na avaliação da variação do número de integrantes, é possível verificar uma semelhança entre as regiões Norte e Nordeste no estudo da variação da Classe Econômica. A variação da classe econômica é mais impactante, na medida em que o aumento econômico proporciona a possibilidade de se ter um equipamento de condicionamento de ar, que é o principal “vilão” da região em relação ao consumo de energia.

Constata-se que em todas as classes econômicas E, não importando o número de pessoas, a concentração das curvas está em classes de consumo baixas, o que intensifica a justificativa de que a classe econômica influencia bastante na compra de equipamentos mais caros. Observa-se que a partir de três pessoas com classe econômica acima de D, o número de curvas em classes de consumo baixas estão abaixo de 200, o que mostra que foi investido em algo para melhorar a qualidade de conforto para a família, que nessas regiões geralmente são os condicionadores de ambiente.

A Tabela 17 apresenta os valores médios de consumo para cada configuração.

Tabela 17 – Consumo médio na avaliação da variação da Classe Econômica – Nordeste

	Média Total (kWh)	Média (kWh) Classe E	Média (kWh) Classe D	Média (kWh) Classe C	Média (kWh) Classe B	Média (kWh) Classe A
Número de Pessoas = 1	171,12	88,25	131,37	161,39	194,59	279,98
Número de Pessoas = 2	297,87	116,96	190,13	246,27	436,88	499,14
Número de Pessoas = 3	398,53	136,05	257,64	372,84	590,12	635,99
Número de Pessoas = 4	556,43	153,98	436,28	611,97	792,46	837,43
Número de Pessoas = 5	663,49	160,83	576,21	719,48	919,94	941,01

A mudança brusca a partir de três integrantes da classe E para classe D, o que pode representar por exemplo a existência e o uso de condicionadores de ambiente.

4.2.3. Região Centro-Oeste – Classe Econômica

A Figura 46 representa o estudo da influência da classe econômica:

Figura 46 – Análise estatística da influência da variação da classe econômica sobre o consumo – Centro-Oeste

O comportamento da distribuição das curvas é diferente das duas anteriores. Para a análise de um integrante, a mudança de classe econômica, não impactou tanto no aumento de consumo, haja visto que as classes econômicas mais altas, a maior parte das curvas se concentram na classe de consumo C- e D+.

As demais configurações apresentam picos em todas as classes econômicas mais equilibradas, não apresentando picos exagerados como enxergado nas configurações da região Norte e Nordeste.

O equipamento que mais consome nessa região é o chuveiro elétrico (segundo estudos do Procel), assim sendo, ele sofre maior influência da variação do número de pessoas. A Tabela 18 ilustra as médias de consumo dessa região.

Tabela 18 – Consumo médio na avaliação da variação da Classe Econômica – Centro-Oeste

	Média Total (kWh)	Média (kWh) Classe E	Média (kWh) Classe D	Média (kWh) Classe C	Média (kWh) Classe B	Média (kWh) Classe A
Número de Pessoas = 1	166,86	114,01	150,84	174,08	184,89	210,47
Número de Pessoas = 2	283,64	158,59	209,65	239,64	353,28	457,02
Número de Pessoas = 3	359,86	179,27	245,53	288,01	486,91	599,63
Número de Pessoas = 4	481,27	201,31	337,81	430,78	675,19	761,26
Número de Pessoas = 5	561,52	215,90	388,89	536,13	797,06	869,61

As variações não são acentuadas como na análise feita na região Nordeste, a não ser pela configuração com cinco integrantes, que possui bastante variação de consumo de acordo com a variação da classe econômica, as demais classes mudam de forma mais amena.

4.2.4. Região Sudeste – Classe Econômica

O estudo da influência da classe para a Região Sudeste pode ser visto na Figura 47:

Figura 47 – Análise estatística da influência da variação da classe econômica sobre o consumo – Sudeste

Destaca-se que em todas as configurações os picos ocorrem um pouco mais a esquerda do que no Centro-Oeste, ou seja, o consumo é um pouco maior. Isso pode ter ocorrido porque a região Sudeste é a região mais desenvolvida do país, logo, o mercado tecnológico é mais aquecido, fazendo com que o aumento econômico da família proporcione a possibilidade de ter mais equipamentos em casa e conseqüentemente maior consumo.

Tabela 19 – Consumo médio na avaliação da variação da Classe Econômica – Sudeste

	Média Total (kWh)	Média (kWh) Classe E	Média (kWh) Classe D	Média (kWh) Classe C	Média (kWh) Classe B	Média (kWh) Classe A
Número de Pessoas = 1	200,09	131,30	195,02	210,88	220,49	242,76
Número de Pessoas = 2	317,98	191,43	276,87	289,76	384,99	446,88
Número de Pessoas = 3	397,57	211,05	323,21	357,08	499,70	586,80
Número de Pessoas = 4	524,45	266,35	419,78	501,63	688,11	746,40
Número de Pessoas = 5	643,10	292,85	540,43	662,39	832,78	887,06

Observa-se um comportamento bem semelhante com a região Centro-Oeste, ou seja, mudança não tão bruscas devido à variação da classe econômica, porém com uma média um pouco maior em todas as configurações, conforme explicado anteriormente.

4.2.5. Região Sul – Classe Econômica

Por fim, para a região Sul, Figura 48 apresenta a avaliação da classe econômica:

Figura 48 – Análise estatística da influência da variação da classe econômica sobre o consumo – Sul

Diferentemente das demais regiões, o Sul se caracteriza por diferentes climas. Esta peculiaridade faz com que o fator econômico seja bem impactante nesta região, pois o conforto em épocas mais frias é primordial, ou seja, se há condições, a prioridade das famílias é ter um aquecedor em suas residências. O consumo geralmente é mais alto devido a esse fator, o que pode ser comprovado pela Figura 48 com aumentos bruscos da quantidade de curvas na classe de consumo mais alta de acordo com a variação da classe econômica.

As classes com menor poder aquisitivo não conseguem aumentar seu patamar de consumo, não importando a quantidade de pessoas, pois equipamentos que trazem conforto e que consomem mais nesta região são mais caros e dependem diretamente do poder aquisitivo da família.

Tabela 20 – Consumo médio na avaliação da variação da Classe Econômica – Sul

	Média Total (kWh)	Média (kWh) Classe E	Média (kWh) Classe D	Média (kWh) Classe C	Média (kWh) Classe B	Média (kWh) Classe A
Número de Pessoas = 1	210,94	119,08	153,11	183,25	244,61	354,66
Número de Pessoas = 2	408,66	173,48	248,99	335,98	596,46	688,39
Número de Pessoas = 3	549,75	211,38	356,52	539,60	783,39	857,87
Número de Pessoas = 4	787,32	225,48	666,74	854,86	1074,25	1115,24
Número de Pessoas = 5	917,38	256,95	858,94	1013,45	1227,05	1230,50

Assim como na avaliação da variação do número de integrantes, a região Sul possui semelhanças com a região Norte e Nordeste, e também com a região Sudeste e Centro-Oeste. Seu consumo sofre impactos diretos da variação tanto do número de integrantes quanto de classe econômica, fazendo com que o consumo das residências do Sul sejam maiores do que as demais regiões em média.

4.2.6. Influência da variação da classe econômica sobre o consumo

É apresentado um gráfico de *BoxPlot*, para melhorar a interpretação dos dados apresentados pelo algoritmo desenvolvido.

Figura 49 – Gráfico de BoxPlot do Número de Pessoas 3 para a Variação da Classe Econômica

Esse gráfico de BoxPlot mostra um comportamento um pouco diferente do anterior para as regiões Sudeste e Centro-Oeste, pois agora ele não possui crescimento constante nem para a média, nem para o tamanho da concentração. Já para as regiões Norte e Nordeste, percebe-se que as curvas se concentraram um pouco acima das curvas vistas no estudo de número de pessoas, observado pelo deslocamento da mediana. Isso mostra que o consumo sofreu uma influência um pouco maior para essa característica nessas regiões.

Pressupõe-se que para essa variável, as regiões Norte e Nordeste possuem bastante semelhanças nas distribuições de curvas, enquanto as regiões Sudeste e Centro-Oeste se assemelham um pouco, porém devido ao maior desenvolvimento do mercado tecnológico na região Sudeste, essa apresenta um consumo maior nas classes econômicas mais elevadas, e por fim, a região Sul possui características das quatro outras regiões no que se refere à distribuição de curvas, com mudanças acentuadas devido à variação da classe econômica.

Essa característica influencia um pouco mais nas regiões em que o condicionador são os “vilões”, no entanto, diferentemente da característica número de pessoa, esta característica não pode ser considerada de impacto direto, pois elas influenciam até certo ponto de forma mais direta e depois passam a influenciar de forma mais amena.

4.3 Estudo da Influência do Tamanho da Residência

A configuração da residência e o seu tamanho está conforme a Tabela 21.

Tabela 21 – Configurações de Tamanhos de Residência

Nome	Quantidade de Cômodos	Tamanho em m ²
Tamanho 1	4	0 a 30
Tamanho 2	4	31 a 45
Tamanho 3	5	31 a 45
Tamanho 4	5	46 a 70
Tamanho 5	6	46 a 70
Tamanho 6	7	46 a 70
Tamanho 7	7	71 a 90
Tamanho 8	8	71 a 90
Tamanho 9	9	71 a 90
Tamanho 10	9	Mais de 91
Tamanho 11	10	Mais de 95
Tamanho 12	11	Mais de 100

Neste estudo, o algoritmo foi configurado para gerar, para cada região, três bancos de dados. Cada banco de dados é composto de doze mil curvas, sendo que mil destas são de famílias que moram em casas de tamanho 1, mil são de tamanho 2, e assim por diante até o tamanho 12. Optou-se por avaliar qual a influência da variação do tamanho da residência sobre a classe/perfil de consumo em função do número de pessoas, ou seja, cada um dos três bancos de dados de cada região representa uma quantidade de integrantes da família (2, 3, 4).

4.3.1. Região Norte – Tamanho da Residência

Assim como nos estudos anteriormente apresentados, nesse, todas as configurações de pessoas estão expostas na mesma figura, porém para melhorar a avaliação, devido à quantidade de tamanhos de residência, a Figura 50, foi feita em formato tridimensional onde o eixo horizontal representa as classes de consumo, o eixo vertical é a quantidade de curvas e eixo longitudinal são os tamanhos da residência.

Pressupõe-se que o tamanho da residência está atrelado a classe econômica, ou seja, quanto maior a classe, espera-se que a família tenha mais condições, então para casas maiores o consumo fica maior não somente pelo aumento de ambientes na casa, mas também porque o poder aquisitivo da família aumentou.

Destaca-se que o aumento de integrantes na família influencia pouco no aumento da classe de consumo, haja visto a similaridade da distribuição das curvas das três figuras apresentadas na Figura 50. Onde existe concentração de curvas em classes de consumo menores para residências menores, e concentração de curvas em classes maiores para residência maiores, nas três configurações de número de integrantes.

Figura 50 – Análise estatística da influência da variação do tamanho da residência sobre o consumo – Norte

4.3.2. Região Nordeste – Tamanho da Residência

O estudo da variação de tamanho da residência para a região Nordeste se aproxima bastante com o da região Norte, devido suas similaridades climáticas.

A Figura 51 apresenta os resultados da análise da variação do tamanho da residência para a região Nordeste:

Figura 51 – Análise estatística da influência da variação do tamanho da residência sobre o consumo – Nordeste

Observa-se os acentuados picos nas maiores residências, possivelmente por existirem ar-condicionado. Outro fator importante a se destacar é a pouca mudança de consumo nas residências menores, não importando o número de integrantes da família (primeiras curvas azul, vermelha e verde, de cada configuração), isso comprova que o fator mais impactante desta região é a classe econômica e não o número de integrantes.

4.3.3. Região Centro-Oeste – Tamanho da Residência

O estudo estatístico para a região Centro-Oeste é apresentado na Figura 52:

Figura 52 – Análise estatística da influência da variação do tamanho da residência sobre o consumo – Centro-Oeste

Observa-se que nesta região o comportamento da distribuição das curvas de acordo com a variação do tamanho da residência é um pouco diferente da anterior quando se altera o número de integrantes, o aumento desloca as curvas para a esquerda. Outro fator de destaque é a concentração das curvas com picos maiores do que os dados das regiões Norte e Nordeste.

4.3.4. Região Sudeste – Tamanho da Residência

Agora para a região Sudeste, a Figura 53 é que apresenta o resultado do estudo:

Figura 53 – Análise estatística da influência da variação do tamanho da residência sobre o consumo – Sudeste

Assim como a região Centro-Oeste, a região Sudeste é bastante influenciada pela alteração do número de integrantes e por isso o deslocamento da distribuição das curvas para esquerda, mesmo em tamanhos de casas menores.

Pode-se diferenciar mais uma vez essas duas regiões em relação à quantidade de curvas com maior consumo na região Sudeste.

A concentração de curvas de acordo com o tamanho das casas também é diferente e isso poderá ser analisado com os gráficos de BoxPlot, que serão apresentados a seguir.

4.3.5. Região Sul – Tamanho da Residência

A Figura 54 apresenta o resultado para a região Sul:

Figura 54 – Análise estatística da influência da variação do tamanho da residência sobre o consumo – Sul

Percebe-se que a região Sul se diferencia das demais. Ela sofre influência direta tanto do número de integrantes, visto no deslocamento para esquerda das residências menores, quanto da classe econômica, visto pelas mudanças bruscas, com picos acentuados devido ao aumento da residência e consequentemente da classe econômica.

Observa-se que o número de pessoas faz com que o consumo cresça rapidamente, devido ao aumento do número de banhos com chuveiros elétricos, bem como, com a possibilidade de se obter um bem comum mais caro e necessário na região, como os condicionadores de ambiente. Isso é observado com o crescimento da classe de consumo A+.

Destaca-se que a distribuição da quantidade de curvas não é nem tão distribuída como nas regiões Norte e Nordeste, e nem tão concentradas como nas regiões Centro-Oeste e Sudeste, o que comprova que a região Sul possui similaridades com todas regiões, devido ao seu clima característico e a tendência de posse de equipamentos.

4.3.6. *Influência da variação do tamanho da residência sobre o consumo*

Essa característica está mais relacionada ao crescimento econômico da família, assim, tem maior influência no poder aquisitivo, impactando conforme discutido anteriormente mais nas Regiões Norte, Nordeste e Sul. Porém, essa característica tem certa relação com o número de integrantes, pois geralmente famílias maiores tendem a construir casas maiores, assim impactando também nas regiões Sudeste e Centro-Oeste.

Portanto, com maior impacto, ou menor impacto, essa variável, tamanho da residência se torna muito importante para o estudo de perfil de consumo, pois delimita a previsão de consumo de acordo com os pontos de energia e iluminação, bem como pode apresentar informações relevantes em relação ao tamanho da família ou suas características financeiras.

A Tabela 22 ilustra a média de consumo geral dos doze tamanhos de residência para cada configuração de diferentes números de integrantes.

Tabela 22 – Resumo da média de consumo do estudo de variação do tamanho da residência

	Consumo Médio (kWh)			Variação das Médias (kWh)		Variação Percentual (%)	
	N = 2	N = 3	N = 4	De 2 para 3	De 3 para 4	De 2 para 3	De 3 para 4
Norte	268,81	362,42	494,90	93,61	132,48	34,82	36,55
Nordeste	303,34	388,82	528,49	85,48	139,67	28,18	35,92
Centro-Oeste	326,87	406,53	521,95	79,66	115,42	24,37	28,39
Sudeste	361,70	442,56	556,97	80,86	114,41	22,36	25,85
Sul	461,53	614,99	856,21	153,46	241,22	33,25	39,22

Os números apresentados na Tabela 22 mostram que a influência da variação do tamanho da residência é mais impactante nas regiões que sofrem mais impacto da característica econômica, ou seja, as regiões Norte, Nordeste e Sul.

Para verificar como estão as concentrações das curvas, e a relação entre os quartis e a mediana, é apresentado a seguir o gráfico de BoxPlot desse estudo.

Figura 55 – Gráfico de BoxPlot do Número de Pessoas 3 para a Variação do Tamanho da Residência

Analisando esse gráfico é possível verificar a semelhança entre as regiões em relação à concentração dos dados, basicamente bem concentrados nas menores residências, se espelhando mais a partir das residência de tamanho 6, 7 e 8, nas Regiões Norte e Nordeste.

Nas demais regiões o crescimento tem característica com crescimento mais acentuado a partir de tamanhos de casas 6 e 7, ou seja, 7 cômodos, variando de 46 a 90 m², sendo que a maior dispersão para Norte e Nordeste é na casa 7 e 8, e nas regiões Sudeste e Centro-Oeste, a maior dispersão ocorre na casa de tamanho 9.

4.4 Estudo da Influência da Composição/Faixa Etária Familiar

Nesse estudo, o algoritmo foi configurado para gerar, para cada região, quatro bancos de dados, o primeiro para famílias com um integrante, o segundo com famílias de dois integrantes, o terceiro com três e o quarto com quatro integrantes.

No primeiro banco de dados, de um integrante, há três mil curvas, mil para adulto, mil para adolescente e mil para idosos. O segundo banco de dados, com famílias de dois integrantes possui seis configurações possíveis (referentes a combinações entre adultos,

neném, crianças, adolescentes e idosos, tomando o cuidado para não haver criança e/ou neném sozinhos na casa), para cada uma foram geradas mil curvas.

No terceiro, há quatro configurações e no quarto banco de dados mais quatro configurações. A classe econômica dos quatro banco de dados foi mantida iguais a C, já o tamanho da residência mudou entre os bancos, cinco, seis, sete e novamente sete, respectivamente para os bancos um, dois, três e quatro.

Todas as curvas foram classificadas quanto à classe de consumo, sendo classificadas individualmente em uma das treze classes pré-estabelecidas anteriormente.

4.4.1. Região Norte – Faixa Etária

Mais uma vez optou-se em colocar todas as configurações juntas em uma mesma figura para tentar melhorar a análise da influência da variação do número de pessoas e da faixa etária dos integrantes da família. A Figura 56 apresenta o resultado para a região Norte:

Figura 56 – Análise estatística da influência da variação da faixa etária sobre o consumo – Norte

Para esse estudo é importante comentar que a faixa etária influencia bastante na posse de equipamentos da residência e também nos hábitos de consumo, e é esse comportamento que determina se haverá aumento ou não no consumo de energia e consequentemente mudança nas classes de consumo.

Destaca-se que diferentemente dos outros estudos não será observado um aumento contínuo da distribuição das curvas e que portanto uma análise com números também se faz necessário para obter a real diferenciação entre as regiões.

Primeiramente, analisando a Figura 56 é possível verificar que na configuração de um integrante, um adolescente e um idoso consomem mais que um adulto, pois existem mais

curvas classificadas na classe D+. Possivelmente isto se deve a permanência maior destes integrantes na residência.

Nas configurações de dois e três integrantes, é possível verificar que a combinação adulto com adolescente consome mais do que qualquer outra combinação, possivelmente porque eles possuem maior diversificação de gostos e autonomia de escolhas dos equipamentos que querem se utilizar.

Na configuração de quatro integrantes, a combinação adultos e adolescentes consomem bastante, no entanto, começa-se a destacar a combinação adultos, neném e criança, possivelmente por que um dos adultos fica mais em casa para cuidar do neném, aumentando o consumo da residência.

4.4.2. Região Nordeste – Faixa Etária

A Figura 57 apresenta este estudo para a região Nordeste:

Figura 57 – Análise estatística da influência da variação da faixa etária sobre o consumo – Nordeste

Diferentemente da região Norte, na região Nordeste com número de pessoas igual à um, quem mais consome é o adulto, seguido dos idosos e por fim os adolescentes, podendo ser explicado pelo tempo em casa, e/ou pela possibilidade de escolha destes integrantes, haja visto que os adolescentes precisam de autorização para realizar um investimento em um novo produto.

Nas demais configurações a combinação de adultos e adolescentes seguem tendo destaque, pois os adultos é que decidem os horários das crianças e nenéns, e os idosos geralmente gastam menos.

4.4.3. Região Centro-Oeste – Faixa Etária

O estudo da influência da faixa etária para a região Centro-Oeste é apresentada na Figura 58:

Figura 58 – Análise estatística da influência da variação da faixa etária sobre o consumo – Centro-Oeste

A figura apresenta novamente que na configuração de um integrante o adulto é que consome mais nesta região. Nas demais configurações a combinação com adolescente é mais significativa, sendo seguida de forma mais próxima do que na região Nordeste, pela combinação de adulto e idoso, nas configurações de dois e três integrantes, e pela combinação adultos neném e criança, na configuração de quatro integrantes.

4.4.4. Região Sudeste – Faixa Etária

Já para a região Sudeste, a Figura 59 apresenta o estudo proposto:

Figura 59 – Análise estatística da influência da variação da faixa etária sobre o consumo – Sudeste

Observa-se uma diferença em relação ao idoso nesta região. O adulto continua sendo o que mais consome, porém o idoso perde espaço nas duas primeiras configurações de um e dois integrantes.

A partir da segunda configuração, a combinação adulto e adolescente passam a ser os que consomem mais, já na quarta configuração todas as combinações consomem bem parecido.

Lembrando que essa é uma análise qualitativa e que uma melhor análise é feita a partir dos dados das médias de consumo de cada configuração que serão apresentadas a seguir.

4.4.5. Região Sul – Faixa Etária

Por fim, para a região Sul a Figura 60 mostra o estudo da variação da faixa etária e sua influência sobre o consumo de energia.

Figura 60 – Análise estatística da influência da variação da faixa etária sobre o consumo – Sul

O adulto segue consumindo mais na configuração de um integrante. Nas demais, a combinação com adolescente consome mais. Cabe destacar a última configuração com quatro integrantes onde todas as combinações estão classificadas em A+, se fazendo necessária a análise quantitativa através da média de consumo para poder fazer uma análise mais conclusiva.

4.4.6. *Influência da variação da faixa etária sobre o consumo*

Toda discussão anterior foi baseada nas Figuras 56 a 60, que apresentam dados qualitativos, através de barras. Assim sendo, algumas conclusões não podem ser tomadas em relação a quem consome mais ou menos e com qual exatidão essa informação é dada. Para isso, será utilizada a Tabela 23, com os valores médios de consumo de cada uma das combinações, para cada configuração:

Tabela 23 – Média de consumo de cada configuração da análise de faixa etária

	Consumo Médio (kWh)				
	Norte	Nordeste	Centro-Oeste	Sudeste	Sul
Adulto	149,97	179,15	193,35	226,61	205,21
Adolescente	156,77	159,28	169,44	202,65	178,45
Idoso	157,47	160,04	165,59	188,77	170,04
1 adulto e 1 neném	237,08	295,65	295,13	353,62	447,57
1 adulto 1 criança	241,52	293,67	295,11	354,25	446,76
1 adulto e 1 adolescente	270,39	320,90	319,68	393,41	493,33
1 adulto 1 idoso	273,04	339,30	330,35	385,07	490,48
2 adultos	224,07	281,53	286,61	342,62	423,26
2 idosos	232,69	245,73	234,57	264,56	314,55
2 adultos e 1 neném	429,12	485,86	330,22	403,04	654,23
2 adultos e 1 criança	434,37	482,39	331,36	401,64	645,91
2 adultos e 1 adolescente	481,41	540,31	365,29	449,10	721,05
2 adultos e 1 idoso	460,42	538,66	367,23	437,10	684,65
2 adultos e 2 nenéns	548,16	627,14	466,37	515,57	875,77
2 adultos e 2 crianças	543,51	620,15	461,76	512,46	869,41
2 adultos e 2 adolescentes	606,60	701,95	511,14	569,76	977,01
2 adultos e 1 neném e 1 criança	607,99	685,15	485,88	547,78	951,08

Por meio dos dados apresentados na Tabela 23, observa-se que a combinação adulto e idoso e adulto e adolescente consomem mais do que a combinação com crianças e recém nascidos, uma hipótese é que muitas vezes as crianças ficam fora das residências em escolas e creches, no entanto, esta hipótese não foi programada, apenas foi pressuposto que crianças e nenéns não ficam sozinhos sem um responsável em casa.

Em média a combinação de adultos com adolescentes e idosos, consomem em torno de 20 kWh a 50 kWh a mais do que as demais combinações, isso é considerado grande, na medida em que a média de consumo está em torno de 300 kWh mês.

Na configuração de um integrante o idoso têm média maior na região Norte, nas demais o adulto é que têm a maior média de consumo. Exatamente como mostrado nos gráficos anteriores da distribuição das curvas.

Considerando a configuração de duas pessoas a combinação adulto e idoso consomem mais nas regiões Norte, Nordeste e Centro-Oeste, enquanto que na região Sudeste e Sul a combinação de maior consumo médio é a adulto com adolescente.

Para a combinação de três indivíduos a combinação de dois adultos com o adolescente prevalece como maior consumidora média nas regiões Norte, Nordeste, Sudeste e Sul, ficando em segundo, por muito pouco, na região Centro-Oeste, perdendo para dois adultos e um idoso.

E por fim, no estudo de quatro integrantes, a combinação de maior consumo segue sendo a relação adultos e adolescentes, menos na região Norte, que perde por muito pouco para a relação adultos e neném e criança.

Apesar dos resultados serem representativos nesse estudo, a faixa etária e a composição da família é uma variável bem complexa, pois a forma de interação entre os indivíduos pode impactar muito no consumo, na medida em que os indivíduos podem conviver juntos ou em ambientes separados.

A seguir o gráfico BoxPlot da configuração com dois integrantes para verificar se existe semelhanças de distribuição de curvas de demanda de energia entre as regiões é apresentado na Figura 61.

Figura 61 – Gráfico de BoxPlot do Número de Pessoas 2 para a Variação da Faixa Etária

Observa-se que novamente as regiões Norte e Nordeste possuem distribuições bem parecidas em relação a variação da faixa etária, bem como, as regiões Centro-Oeste e Sudeste, tanto de comportamento, quanto de tamanhos de distribuição. E também que a região Sul se difere das demais.

Outro destaque nessa figura é a concentração de curvas nas regiões Sudeste e Centro-Oeste, isso representa que para cada uma das configurações de faixa-etária, existe um perfil bem característico de consumo.

Comparando os gráficos do Norte, Nordeste e Sul, percebe-se que a mediana nos dois primeiros está mais embaixo, enquanto que no Sul está mais no alto, isso significa que a concentração de curvas no Norte e Nordeste estão em baixo consumo, e a concentração na região Sul está em altos consumos.

No Apêndice A está o link de acesso de todas as tabelas de distribuição das curvas de cada um dos estudos realizados na pesquisa.

5. ESTUDO DE CLASSIFICAÇÃO UTILIZANDO FERRAMENTAS INTELIGENTES

As ferramentas computacionais inteligentes estão sendo utilizadas atualmente para auxiliar no tratamento de grandes massas de dados que as empresas possuem, dados esses que não são mais mensurados em *giga* ou *terabytes*, mas sim em *peta*, *exa*, e até *zettabytes*, que respectivamente denotam os fatores 10^{15} , 10^{18} e 10^{21} . Com as ferramentas inteligentes as empresas conseguem traçar semelhanças, ou criar padrões e perfis que nem mesmo os mais experientes na área conseguiriam, ou se conseguissem, demorariam muito tempo.

As curvas de carga virtuais geradas pelo algoritmo desenvolvido, considerando dados socioeconômicos, podem ser analisadas de diversas maneiras. Uma delas é o estudo de previsão de classe de consumo conhecendo-se as características da família. Outro estudo é o sentido contrário, ou seja, sabendo-se o perfil da curva de carga, determinar as características da família. Além destes, é possível auxiliar na busca de fraudes de energia, ou mesmo realizar uma desagregação da curva de carga para determinar os equipamentos que existem na residência. Entretanto, é necessário melhorar os dados, como por exemplo, deve-se considerar a realização de inspeções em campo para caracterizar um possível fraudador, e uma modelagem das cargas para o caso de desagregação das curvas.

No entanto, nessa tese são propostas duas análises utilizando ferramentas computacionais inteligentes:

- a) prever qual a classe de consumo de uma residência, quando se conhece as características da família; e
- b) determinar as características das famílias quando se conhece o perfil de consumo da residência.

Essas propostas foram feitas para verificar a possibilidade da utilização de ferramentas inteligentes para auxiliar, primeiro, no planejamento de novas instalações, pois se for possível prever a classe de consumo das residências, poderá melhorar o dimensionamento dos equipamentos das novas instalações e estudos de carregamentos de transformadores. E segundo, tendo o perfil de consumo de cada residência (com aplicação dos *smart-meters*), poder identificar as características da família, pode possibilitar uma melhoria dos planos de relação com os clientes, na medida em que estima-se que os conceitos de *smart grid* e *demand response* estejam sendo aplicados em um futuro próximo, onde os clientes terão participação

mais ativa no mercado de energia, esse estreitamento entre a concessionária e os clientes se tornam vital.

Para a avaliação dos ensaios realizados com as ferramentas computacionais inteligentes serão avaliados dois parâmetros, a acurácia e a matriz de classificação. A acurácia, muitas vezes chamada de exatidão, mede a proporção de predições corretas, sem levar em consideração o que é positivo e o que é negativo. Essa medida é altamente suscetível a desbalanceamentos do conjunto de dados e pode facilmente induzir a uma conclusão errada sobre o desempenho do sistema, por isso a proposta da utilização da matriz de classificação.

Essa matriz, também chamada de matriz de confusão é uma tabela que permite a visualização do desempenho de um algoritmo de aprendizado, tipicamente em um algoritmo supervisionado. Cada coluna da matriz representa as instâncias de uma classe prevista, enquanto as linhas representam os casos de uma classe real. O nome é originado do fato de que a matriz torna mais fácil verificar se o algoritmo está confundindo duas ou mais classes.

Por definição, na diagonal principal da matriz de classificação, encontram-se os valores que foram corretamente classificados em suas respectivas classes, enquanto que nas demais posições da matriz encontram-se os valores de classificação errada.

Tomando como exemplo uma tabela de classificação de animais apresentado na Figura 62, observa-se os principais conceitos que nortearão os resultados e discussão apresentados nessa tese.

		Classe Prevista		
		Gato	Cão	Coelho
Classe Real	Gato	5	3	0
	Cão	2	3	1
	Coelho	0	2	11

Figura 62 – Exemplo de matriz Confusão – Classificação de animais (Fonte: REVISTABW, 2015).

As linhas da matriz representam as classes reais, ou seja, as classes que foram apresentadas para os sistemas inteligentes de classificação. As colunas representam as classes que o sistema inteligente classificou as amostras.

A interpretação do gráfico é a seguinte: somando os valores da linha tem-se a quantidade de elementos de classe, ou seja, observe na primeira linha os valores ‘5’ e ‘3’, isso significa que existiam ‘8’ gatos para serem classificados, mas o sistema classificou ‘5’ como

gato e '3' como cão. Continuando a análise, na segunda linha existiam '6' cães para serem classificados ('2' + '3' + '1'), no entanto o sistema classificou, '2' como gato, '3' como cão e '1' como coelho. E por fim, na última linha eram '13' coelhos, mas a ferramenta classificou '2' como cão e '11' como coelho. A soma total dos valores é a quantidade de elementos que a ferramenta inteligente teria que classificar (total 27 elementos), e a diagonal principal a quantidade que ele acertou (total de 19 elementos certos).

A acurácia nada mais é que a divisão entre o total de acerto, pelo total de elementos, que nesse caso é de aproximadamente 70% ('19' dividido por '27'). A discussão se esse valor de acurácia é bom ou ruim, depende muito da proposta do problema, e cada situação deve ser avaliado de forma individual.

Outra forma de apresentar essa matriz, é através de percentuais de acerto, ou seja, analisando a primeira linha, foram classificados um total de '8' animais, sendo '5' gatos e '3' cachorros. De forma percentual, teríamos na primeira coluna '0,625', na segunda coluna '0,375' e a terceira '0,00'. Assim, a interpretação seria: acertou '62,5%' os que eram gatos e foram classificados como tal, e confundiu gatos classificando como cachorros '37,5%'; não classificando nenhum como coelho.

Mais uma vez, a percepção se a classificação foi boa ou ruim depende muito da aplicação e do problema, bem como da análise que será feita sobre os dados, no entanto, a matriz de classificação pode direcionar a avaliação dos dados de maneira mais profunda.

Para esse estudo de classificação, proposto nessa tese, foram selecionadas duas ferramentas computacionais inteligentes, o classificador OPF, e o classificador de Redes Bayesianas. Os detalhes de cada uma estão descritos no Capítulo 2.

A escolha dessas ferramentas basicamente está relacionado ao esforço computacional exigido no treinamento e teste (tempo de processamento), que segundo estudos na literatura, em ambos esse esforço é menor do que as demais técnicas utilizadas geralmente como, Redes Neurais Artificiais e Máquinas de Vetor de Suporte (PAPA; FALCÃO, A. X.; SUZUKI, 2009).

Além disso, as Redes Bayesianas têm como vantagens, fácil visualização das assertividades, a boa exploração da relação entre as variáveis, tornam o processo de inferência eficiente, o que permite a análise de grande quantidade de dados. Elas se baseiam em conceitos de grafos e de probabilidade.

Já o classificador OPF além da velocidade de treinamento, ele não depende da quantidade de dados de cada classe, ou seja, trabalha bem com bancos desbalanceados, uma vez que ele trabalha com conceito de grafos e vai formando, durante a fase de treinamento, as árvores de classes de acordo com os dados e ajustando os pesos das ligações entre as instâncias daquela classe. Uma nova amostra é classificada de acordo com os pesos da árvore através de uma competição entre os elementos mais próximos da amostra.

Para cada uma das ferramentas inteligentes tomou-se o cuidado de realizar testes com dados normalizados e não-normalizados, pois segundo a literatura a performance de classificação está ligada a qualidade de dados de entrada e cada problema reage de maneira diferenciada em relação ao tratamento de dados. Geralmente os dados normalizados têm resposta melhor, mas deve-se avaliar os dois casos, ainda mais nessa aplicação que é inédita, onde estão sendo relacionados dados socioeconômicos com técnicos.

Como dito anteriormente, os bons resultados de classificação dependem da qualidade dos dados e a relação existente entre as variáveis, tanto é que, atualmente muitas pesquisas estão sendo feitas para selecionar as melhores características/variáveis para determinado problema, melhorando ainda mais a performance da classificação.

Geralmente o uso de normalização de dados é feita para ajustar a dispersão existente entre as variáveis, devido a diferença de unidade e heterogeneidade física das variáveis, ou então devido à distorções de dados (PAPA et al., 2010).

Nessa pesquisa também foram considerados diferentes percentuais de treinamento e teste (25/75%, 50/50%, 75/25%), respectivamente, pois estudos apontam que esses diferentes ensaios devem ser aplicados e avaliados para identificar se o problema possui uma tendência de aprendizado e se este parâmetro irá impactar na performance da classificação. Se a performance for muito semelhante, geralmente opta-se pelo menor esforço computacional.

As ferramentas computacionais inteligentes ainda foram programadas para rodarem ao menos dez vezes com resultados acima de um patamar estabelecido (acurácia 50%), pois a seleção de dados foi de forma aleatória. Dessas dez rodadas é tirada uma média, com o intuito de diminuir o problema de desbalanceamento de dados.

Toda a complexidade desse estudo pode ser medido através do esforço computacional exigido pelas ferramentas durante o processo de treinamento e teste. Em todos os ensaios o tempo entre treinamento e teste foi medido em dias, ou muitas horas do dia, ou seja,

estabelecer uma relação entre essas variáveis socioeconômicas e técnicas de consumo de energia é difícil.

Assim sendo, os estudos de previsão de classe de consumo e o de classificação das características das famílias são apresentados a seguir.

5.1 Estudo de Previsão da Classe de Consumo

A palavra previsão pode ser trocada neste estudo por classificação, pois é como a ferramenta computacional inteligente entende o problema e classifica as curvas de carga em faixas/classes de consumo. Foi criado um banco de dados contendo cinquenta mil dados, para cada região, com as características da família (composição/faixa etária dos integrantes, classe econômica, número de pessoas, tamanho da residência, horário que está em casa e pessoas que trabalham) e a classe de consumo. Esses dados foram gerados de acordo com as características da região, ou seja, com número de pessoas na residência e a classe de consumo proporcional aos estudos socioeconômicos do IBGE; assim como foi feito na validação do algoritmo.

Esses dados foram apresentados a dois sistemas inteligentes, o classificador OPF e as Redes Bayesianas, que foram escolhidas devido ao seu menor custo computacional durante o treinamento de dados. A ideia principal desse primeiro ensaio é verificar a possibilidade de prever a classe de consumo de uma família, conhecendo-se seus dados socioeconômicos (composição/faixa etária dos integrantes, classe econômica, número de pessoas, tamanho da residência, horário que está em casa e pessoas que trabalham).

Ressalta-se que a complexidade desse primeiro ensaio está em haver treze classes de consumo e de que a criação de um perfil de consumo, que relaciona todas as variáveis é muito difícil, pois, conforme mencionado anteriormente, o livre arbítrio sobre como é utilizada a energia elétrica, permite que em duas famílias com as mesmas características exista um consumo totalmente distinto.

A Figura 63 apresenta uma representação gráfica da proposta estudada neste primeiro ensaio, onde existem duas fases, o treinamento e o teste. No treinamento as ferramentas recebem os dados de entrada e a saída desejada, criando uma lógica de relacionamento entre as variáveis (treinamento supervisionado). No teste novos dados são apresentados para que o as ferramentas os classifiquem, medindo-se a qualidade dessa classificação.

Figura 63 – Representação gráfica do ensaio de classificação de classe de consumo.

Cabe ressaltar que se trata de um banco de dados desbalanceado, ou seja, existe variação de número de perfis por classe, optando-se por realizados três ensaios referentes à quantidade de dados de treinamento e teste. Foram eles: 25% de treinamento para 75% de dados de teste; 50% de treinamento, para 50% de teste; e 75% de treinamento para 25% de teste. Foi proposta também a análise com dados normalizados e não normalizados. Além disso, foram realizadas dez rodadas de testes e calculada a média, tudo isso, visando a diminuição dos possíveis erros devido ao desbalanceamento dos dados.

A seguir, são apresentados os resultados de acurácia média de cada configuração, para cada região do país, sendo que as siglas T.N.N, e T.N. significam, treinamento não normalizado e treinamento normalizado, respectivamente, ou seja, os primeiros valores são para 25% de treinamento com dados não normalizados e 75% de teste, e assim sucessivamente.

Tabela 24 – Média da acurácia para estudo de classificação da classe de consumo

		25% T. N. N.	25% T. N.	50% T. N. N.	50% T. N.	75% T. N. N.	75% T. N.
Norte	OPF	68,54%	70,32%	71,47%	66,14%	64,82%	71,85%
	RB	68,25%	68,53%	68,68%	65,86%	64,65%	71,99%
Nordeste	OPF	69,02%	73,54%	51,57%	67,58%	50,01%	63,60%
	RB	81,07%	70,22%	64,49%	64,37%	64,28%	63,75%
Centro-Oeste	OPF	67,95%	65,46%	54,49%	67,82%	70,80%	68,95%
	RB	74,72%	70,22%	54,50%	67,94%	51,58%	67,16%
Sudeste	OPF	72,36%	71,92%	71,60%	72,26%	50,64%	51,13%
	RB	72,05%	68,59%	70,94%	71,13%	67,54%	51,14%
Sul	OPF	69,61	67,61%	68,52%	65,12%	50,28%	50,15%
	RB	68,71	69,00%	67,36%	67,30%	67,98%	58,46%

Observa-se através da Tabela 24 que as ferramentas computacionais inteligentes tiveram resultados semelhante e que nem a quantidade de dados de treinamento e nem se esses dados são normalizados ou não, influenciam de forma impactante na performance da classificação, referente a acurácia.

O destaque positivo da Região Norte foi a configuração de 75% de treinamento normalizado, em torno de 72% para as duas ferramentas, enquanto que o destaque negativo foi a configuração de 75% de treinamento não normalizados em torno de 65% para as duas ferramentas. Os demais resultados dessa região foram entre esses dois valores, ou seja, bem próximos uns dos outros, no entanto, quando se trata de uma classificação de milhares, ou em um caso real, de milhões de dados, 1% de melhoria de acurácia pode significar bastante. Assim, a escolha das melhores configurações se torna imprescindível.

Já para a Região Nordeste as ferramentas divergiram entre as melhores e as piores configurações. Enquanto que para o OPF a melhor classificação foi a configuração de 25% de treinamento normalizado com 73,54%, a RB foi a de 25% de treinamento não normalizado com 81,07%, enquanto que o pior resultado do OPF foi de 75% não normalizado com 50,01%, e a RB foi de 75% de treinamento normalizado com 63,75%. Isso mostra que dados normalizados e não normalizados não impactam diretamente na performance dos classificadores nesse estudo.

Agora para a Região Centro-Oeste a melhor configuração do OPF é a de 75% de treinamento não normalizado com 70,80% de acurácia, enquanto que a pior é de 50% não normalizado com 54,49% de acurácia. Já para a RB a melhor configuração é a de 25% não

normalizada com 74,72% e a pior é a de 75% de treinamento não normalizado com 51,58%. Nessa região fica claro que os métodos utilizados para o treinamento das ferramentas são diferentes, pois a mesma configuração 75% de treinamento não normalizado foi a melhor configuração para o OPF e a pior para as Redes Bayesianas.

Analisando a Região Sudeste, assim como a Região Norte, as ferramentas convergiram para as melhores e piores configurações. No caso, a melhor configuração para ambas foi a de 25% de treinamento não normalizado (OPF 72,36% e RB 72,05%) e a pior configuração de ambas foi a de 75% normalizado (OPF 51,13% e RB 51,14%). É importante ressaltar que o valor de aproximadamente 50% de acurácia para esse caso de dados desbalanceados pode ser considerado bom, pois a ferramenta pode ter aprendido muito bem algumas classes, e outras nem tanto, fazendo a acurácia diminuir. Portanto, é importante avaliar outros fatores além da acurácia e também a aplicação em que a ferramenta será utilizada.

Por fim, a Região Sul onde tanto o OPF quanto as Redes Bayesianas tiveram a mesma configuração (75% de treinamento normalizado) para a pior acurácia, respectivamente, 50,15% e 58,46%, enquanto que para as melhores configurações tem-se o OPF com 25% de treinamento não normalizado (69,61% de acurácia) e a RB com 25% de treinamento normalizado com 69,00% de acurácia.

Resumidamente, percebe-se que os sistemas inteligentes conseguiram um bom resultado, dado a complexidade da proposta, com valores de acurácia, em média, acima de 65%. Assim sendo, é possível concluir que existe realmente relação entre as características da família e a classe de consumo, neste banco de dados gerados.

Contudo, para melhorar esses valores é necessário identificar novas características ou então remover características que estejam atrapalhando no entendimento dos sistemas inteligentes. Algumas acurácias próximas de 50%, em algumas configurações, representam que a classificação pode ter ocorrido de forma equivocada em determinadas classes. No entanto, vale ressaltar que esses casos devem ser avaliados de acordo com o problema proposto. Devido ao desbalanceamento de dados, e a escolha aleatória de dados de treinamento e teste, pode ter sido a causa desse baixo índice de acurácia.

O valor de acurácia pode apresentar uma interpretação errada do problema, principalmente em banco de dados desbalanceados, como é o caso desse estudo. Note que a classificação errada de uma classe de menor quantidade de perfil, pode ajudar no crescimento

de acurácia, e o caso contrário também é válido, ou seja, uma classificação errada de uma classe com muitos perfis, pode colocar a acurácia em níveis baixíssimos.

Assim sendo, optou-se em analisar melhor a região Sudeste, que é a região de maior impacto de consumo e de população, utilizando-se da matriz de classificação, onde é possível analisar as tendências dos erros, e enxergar melhor os perfis criados, possibilitando melhorar a análise dos dados e enxergar aonde se encontra a confusão criada.

Para melhorar a visualização da matriz de classificação, foi proposto colorir as suas células, da seguinte maneira:

- cor verde para os maiores valores percentuais das linhas;
- tom de azul mais escuro, para valores que estão entre 60% do maior até 99% do maior valor, representando confusão forte;
- tom de azul médio, para valores entre 30% do maior até 60% do maior valor, indicando confusão média;
- tom azul claro, para valores diferentes de zero e menores que 30% do maior, representando baixa confusão; e
- cor Branca, para células sem porcentagem, representando que não houve confusão.

A classificação perfeita é obtida quando na diagonal principal estão os maiores valores (100%), e esta fica pintada de verde, e as demais células ficam brancas com 0%. Pode ocorrer de algumas classes serem classificadas em perfeição e outras não, normalmente isto mostra que a classe em destaque se difere das demais classes. Isto poderá ser verificado na matriz de classificação apresentada a seguir.

Boas classificações são obtidas quando acerta-se a diagonal principal e não há confusões fortes (azul escuro), mostrando que o sistema inteligente conseguiu definir os perfis das diferentes classes. Outra coisa que deve-se levar em consideração é a proximidade da confusão, pois como as classes selecionadas equivalem às faixas de consumo, as classes mais próximas podem ser confundidas. Já se o sistema inteligente tiver confusão com classes muito distintas, significa que o aprendizado foi muito baixo; ou por falta de dados, devido o desbalanceamento do banco de dados, ou pela complexidade do problema e falta de melhores dados de classificação. A seguir encontram-se as matrizes de confusão da Rede Bayesiana e do OPF, para este estudo de classificação de classes de consumo (Tabela 25).

Matriz Confusão Rede Bayesiana (Sudeste – Classe de Consumo)

[illegible]

Matriz Confusão OPF (Sudeste – Classe de Consumo)

[illegible]

A seguir têm-se a legenda:

Legenda

- 1 – Maior Valor Percentual – Verde
- 2 - Igual a zero - Branco
- 3 - Se maior que 60% do maior - intensidade maior de azul
- 4 - Se está entre 30% até 60% do maior - intensidade média azul
- 5 - Se menor que 30% do maior - intensidade fraca do azul

Destaca-se que nenhum deles errou bruscamente a classificação, ou seja, os erros foram em suas vizinhanças de classe. Isso mostra que as ferramentas entenderam que as características mostram as tendências de consumo, e consequentemente das classes de consumo. As Redes Bayesianas apresentam percentuais de classificação correta em algumas classes maiores do que o OPF.

Entretanto, existem classes em que o percentual de maior classificação está errado como, por exemplo, a classe B+ onde se confundiu com as classes B e A-, sendo que o maior percentual ficou em B. Outro exemplo é a classe E+, que ele classificou totalmente como sendo D-. Já o classificador OPF acertou todos os maiores percentuais das linhas em suas posições corretas de coluna (verdes em toda a diagonal principal), o que mostra que o OPF acertou maior valor percentual das classes certas.

Para as duas ferramentas de classificação, a classe B+ teve confusão elevada. A Rede Bayesiana, por exemplo, errou a posição do maior percentual, impactando bastante nos valores verdadeiros, e assumindo bastante valores falsos. Já para o OPF, apesar de acertar a posição do maior percentual, seus erros na vizinhança são elevados. Assim pode-se concluir que essa é uma classe que deve ter atenção redobrada, podendo ser modificada seus limites.

Observando os dados utilizados para treinamento e teste, percebe-se que as classes A-, B+ e B, são bem semelhantes em posse de equipamentos, em horário em casa, número de pessoas, geralmente acima de três e tamanho de residência, para esta região Sudeste. Essa semelhança, faz com que as ferramentas encontrem dificuldades de estabelecer as ‘fronteiras’ entre as classes, prejudicando em seu aprendizado e consequentemente na classificação. No segundo caso, a classificação errada da Rede Bayesiana, ocorreu porque haviam poucos dados dessa classe de consumo, assim, poucos dados de treinamento.

Outro ponto de destaque é que a Rede Bayesiana classificou todos da classe E+ como D-, e o OPF acertou 91,9% dessa classe. Isto pode ter ocorrido porque a Rede Bayesiana criou

uma regra totalmente diferente do OPF para essa classe devido ao pouco número de dados dessa classe (desbalanceamento do banco de dados).

Destaca-se também a classificação perfeita da classe E- nas duas ferramentas. Apesar de poucos dados desta classe de consumo, eles possuem características bem fortes, como, baixo número de pessoas, poucos equipamentos, fora de casa, e casas pequenas.

Para avaliar qual das duas ferramentas computacionais se saiu melhor, é proposto uma comparação através da matriz de classificação, que leva em consideração o percentual de acertos na diagonal principal, a quantidade de confusões em diferentes níveis, e a ‘qualidade’ dessa confusão.

O percentual de acerto da diagonal principal talvez seja o mais importante dos índices avaliados, pois indica qual ferramenta entendeu o problema e acertou mais as classificações em cada uma das classes exigidas. No entanto, medir a quantidade de erros e/ou confusões e a ‘qualidade’ dessa confusão é importante.

Então, foi proposto quantizar em números quantos erros ocorreram em cada um dos níveis de confusão estabelecidos (níveis de azul), e posteriormente tirar uma média do percentual que representa cada um desses níveis de azul, e chamados de ‘qualidade’ da confusão. Assim, quanto maior a porcentagem de acerto da diagonal, melhor é o sistema, e para a quantidade e ‘qualidade’ da confusão, quanto menor o valor é melhor. Outros critérios poderiam ser levados em consideração, como por exemplo a distância do erro, que mostraria o quão errado a ferramenta aprendeu, ou que as classes escolhidas estão confusas, e que novas características devam ser buscadas para diferenciar bem as classes.

A Tabela 26 apresenta esta comparação entre as Redes Bayesianas com o OPF para a classificação da classe de consumo, tendo como entrada as características da família.

Tabela 26 – Avaliação das Ferramentas Inteligentes para Classificação da Classe de Consumo - Sudeste

	% Acerto Diagonal Principal	Quantidade de Confusão (n°)		‘Qualidade da confusão (%)	
RB	66,9	Azul Escuro	5	Azul Escuro	37,2
		Azul Médio	5	Azul Médio	22,2
		Azul Claro	39	Azul Claro	3,4
OPF	100,0	Azul Escuro	7	Azul Escuro	29,5
		Azul Médio	11	Azul Médio	19,6
		Azul Claro	38	Azul Claro	4,1

Observa-se na Tabela 26 de comparação, que o classificador OPF se saiu melhor do que o classificador por Rede Bayesiana para o estudo de classificação da classe de consumo através das características da família, uma vez que, a quantidade de acertos na diagonal principal está maior e a ‘qualidade’ de confusão está menor em todos os níveis. Apesar de ter quantidades de confusão nos níveis mais altos (azul escuro e azul médio-‘5’ a ‘7’; e ‘5’ a ‘11’, respectivamente) mais alto do que as Redes Bayesianas, o OPF mostrou maior ‘qualidade’, com médias percentuais abaixo da Rede Bayesiana. Os maiores índices percentuais de azul escuro da Rede Bayesiana é de 42% e do OPF é de 36,4%.

O resultado desse estudo pode ser utilizado para realizar uma previsão de demanda e carregamento de transformador através do conhecimento das características da família, auxiliando assim, no planejamento feito hoje em dia, que leva em consideração apenas a potência instalada projetada e calculada durante o projeto da obra e a demanda média da região.

5.2 Determinação das Características das Famílias

Para esta análise, foram criados três banco de dados distintos, um para determinar a composição familiar-faixa etária, um para a classe econômica e outro para o tamanho da residência. Novamente cada um dos banco de dados continha cinquenta mil dados, agora com as características do perfil de consumo (classe de consumo, potência instalada, demanda máxima, demanda média, fator de carga, fator de demanda, número de picos, número de vales, picos acima da média e vales acima da média), e a característica que quer se determinar (faixa etária-composição, classe econômica ou tamanho da residência). Assim, cada região possui três bancos de dados (classificação de classe econômica, classificação de tamanho de residência, e classificação de faixa etária).

5.2.1. *Classificação da Característica: Classe Econômica*

A Figura 64 é uma representação gráfica de como foi feito esse estudo de classificação da classe de consumo, ressaltando que há uma fase de treinamento (supervisionado) e uma fase de teste.

Figura 64 – Representação gráfica do ensaio de classificação da classe econômica.

Novamente, os sistemas inteligentes OPF e Redes Bayesianas foram treinados e testados com 25/75%, 50/50% e 75/25%, utilizando dados normalizados e não normalizados seguindo a lógica de dez rodadas de testes para ser calculada a média dos resultados.

A Tabela 27 apresenta os resultados de acurácia das ferramentas para esse estudo.

Tabela 27 – Média da acurácia para estudo de classificação da característica classe econômica

		25% T. N. N.	25% T. N.	50% T. N. N.	50% T. N.	75% T. N. N.	75% T. N.
Norte	OPF	74,51%	50,01%	58,02%	71,95%	74,16%	72,82%
	RB	74,80%	58,83%	58,16%	72,34%	74,53%	73,31%
Nordeste	OPF	73,98%	57,96%	69,64%	70,02%	64,87%	69,57%
	RB	74,30%	67,89%	67,96%	70,17%	68,57%	69,91%
Centro-Oeste	OPF	50,02%	66,53%	72,98%	67,13%	73,20%	67,53%
	RB	56,24%	66,93%	73,47%	67,57%	73,87%	67,82%
Sudeste	OPF	63,85%	66,52%	73,99%	67,24%	74,01%	57,50%
	RB	65,94%	66,89%	74,36%	67,57%	74,56%	61,02%
Sul	OPF	66,63%	67,41%	74,15%	67,97%	60,33%	57,64%
	RB	67,01%	67,69%	74,49%	68,25%	62,68%	60,58%

Observando a Tabela 27 para o estudo de classificação de classe econômica percebe-se novamente que a quantidade de dados de treinamento e se esses estão ou não normalizados, não afeta diretamente na performance do classificador, quanto sua acurácia.

Nesse estudo, somente a Região Norte teve uma divergência entre os classificadores em relação à pior configuração. Enquanto o OPF a pior configuração é a de 25% T.N. (50,01%), na RB a pior configuração é de 50% T.N.N. com 58,16% de acurácia. A melhor configuração para as duas ferramentas foi a de 25% T.N.N. com 74,51% para o OPF e 74,80% para as Redes Bayesianas.

Na Região Nordeste a melhor configuração foi de 25% T.N.N. com 73,98% para o OPF e 74,30% para a RB. E a pior configuração foi a de 25% T.N. com 57,96% para o OPF e 67,89% para a RB. Analisando os dados normalizados e não normalizados dessa região percebe-se que para cada configuração, ora os dados normalizados são melhores, ora eles são piores, ou seja, cada caso deve ser avaliado e estudado de forma individualizado.

Para a Região Centro-Oeste a melhor configuração para as duas ferramentas é a de 75% de treinamento não normalizado, onde o OPF apresentou acurácia média de 73,20% e a RB de 73,87%, enquanto que a pior configuração para ambas foi a de 25% T.N.N com 50,02% e 56,24% de acurácia, respectivamente, OPF e Redes Bayesianas.

No Sudeste a melhor configuração foi de 75% T.N.N. para ambas e a pior foi a de 75% T.N. também para ambas, com 74,01% e 57,50% para o OPF e 74,56% e 61,02% para as Redes Bayesianas.

Por fim, na Região Sul, a classificação de classe econômica teve como melhor configuração do OPF e da RB a 50% de treinamento de dados não normalizados, e a pior configuração 75% de treinamento normalizados, tendo 74,15% e 57,64% de acurácia para o OPF, e 74,49% e 60,58% de acurácia para a RB.

Observa-se que para este problema de caracterização de classe econômica em todas as regiões, há pelo menos uma configuração de treinamento e teste acima de 70% de acurácia, dando a entender que as ferramentas aprenderam a realizar essa classificação. Como os valores estão bem semelhantes, é a partir da matriz de classificação que poderá ser definido qual obteve melhor resultado.

Salienta-se que, para esse estudo, apesar de haver apenas cinco classes, menos da metade do estudo de classe de consumo, que eram treze classes, há mais variáveis de entrada.

No total foram utilizadas dez características das curvas de consumo geradas pelo algoritmo desenvolvido.

Novamente foi escolhida a Região Sudeste para realizar a avaliação da matriz de classificação, sendo mais uma vez utilizadas as cores para mapeamento dos acertos e erros, além do valor percentual das classificações estipuladas pela ferramenta. Ao verificar as cores, é possível identificar tendências dos perfis estipulados, além de verificar se as ferramentas estão encontrando e mapeando perfis dos consumidores. Já com o valor percentual, é possível verificar a intensidade de relacionamento entre as classes, e a intensidade de confusão da ferramenta, auxiliando na análise dos dados.

A matriz de classificação dessa análise é apresentada na Tabela 28.

Tabela 28 – Matriz de Classificação para estudo de classificação da característica classe econômica - Sudeste
Matriz de Confusão Rede Bayesiana (Sudeste - Classe Econômica)

A	49,40%	46,00%	4,30%	0,20%	0,00%
B	5,60%	62,60%	25,80%	6,00%	0,00%
C	0,70%	24,60%	46,10%	27,70%	0,80%
D	0,00%	7,80%	29,00%	51,80%	11,30%
E	0,00%	0,00%	1,30%	11,60%	87,10%

Matriz de Confusão OPF (Sudeste - Classe Econômica)

A	48,20%	47,00%	4,60%	0,20%	0,00%
B	6,00%	61,40%	26,30%	6,30%	0,00%
C	0,80%	24,70%	45,80%	27,80%	0,90%
D	0,00%	8,10%	29,50%	51,10%	11,20%
E	0,00%	0,00%	1,20%	12,60%	86,10%

Legenda

- 1 – Maior Valor Percentual – Verde
- 2 - Igual a zero - Branco
- 3 - Se maior que 60% do maior - intensidade maior de azul
- 4 - Se está entre 30% até 60% do maior - intensidade média azul
- 5 - Se menor que 30% do maior - intensidade fraca do azul

Ao analisar as matrizes de classificação, a classe com maior acerto foi a classe E (87,1% com Rede Bayesiana e 86,10% OPF). Essa classe econômica é bem característica, pois a posse de equipamentos é menor. A classe de maior confusão para as duas ferramentas foram a classe A e a classe C, sendo que o OPF teve o maior erro percentual. Observa-se também que as confusões estão relacionadas às suas vizinhanças, isto indica que a ferramenta está entendendo o problema, e está criando tendências de criação de perfis.

Para entender o porquê das confusões das ferramentas uma análise mais crítica dos dados apresentados para seu treinamento se faz necessária. Sabe-se que o problema é muito complexo, pois cada pessoa possui o livre arbítrio de consumir da forma que lhe convém, porém a proposta do trabalho é identificar tendências de características comuns dos consumidores que possuem semelhanças culturais, econômicas e sociais.

Dos dados, pode-se verificar que 90% das amostras de classe econômica E estão classificadas como classe econômicas C-, enquanto que apenas 25% das amostras de classe econômica D estão classificadas desta forma, ou seja, há uma separação clara de tendência de que esta característica classe de consumo influencie na classe econômica para estas duas classes citadas D e E. Porém, verifica-se, por exemplo a classe econômica C em conjunto com a D e a B, onde ocorre grande confusão de classificação, que os dados de classe de consumo não são suficientes para determinar um perfil, pois a maior parte dos dados se concentram nas classes de consumo B e C, fazendo com que a ferramenta confunda sua classificação.

Analisando as demais características, verificou-se que elas também estão um tanto quanto difusas devido a toda complexidade do problema, fazendo com que a confusão apresentada pela ferramenta se torne plausível. Uma saída para essa confusão seria identificar e estudar novas variáveis que auxiliem nestas zonas de confusão como, por exemplo, a posse de equipamentos, ou incluir ainda alguma característica da família conhecida previamente.

Agora, para comparar as ferramentas é apresentada na Tabela 29 a classificação conforme explicado no estudo anterior, ou seja, com a quantidade percentual de acertos da diagonal principal, a quantidade e ‘qualidade’ das confusões geradas para a classificação da classe econômica, tendo como entrada as características de consumo.

Tabela 29 – Avaliação das Ferramentas Inteligentes para Classificação da Classe Econômica - Sudeste

	% Acerto Diagonal Principal	Quantidade de Confusão (n°)		‘Qualidade da confusão (%)	
RB	100,0	Azul Escuro	2	Azul Escuro	36,85
		Azul Médio	3	Azul Médio	26,47
		Azul Claro	10	Azul Claro	4,96
OPF	100,0	Azul Escuro	2	Azul Escuro	37,40
		Azul Médio	3	Azul Médio	26,83
		Azul Claro	10	Azul Claro	5,19

Observando-se a Tabela 29, é possível indicar que, para esse estudo de classificação de classe econômica, as Redes Bayesianas se saíram um pouco melhores do que o OPF, uma

vez os acertos da diagonal são iguais, e a quantidade de erros também, avaliando-se a ‘qualidade’ de confusão nota-se que essa é menor na da RB do que na do OPF, nos três níveis (‘36,85’ a ‘37,40’; ‘26,47’ a ‘26,83’; e ‘4,96’ a ‘5,19’), respectivamente, azul escuro, azul médio e azul claro.

Com esse estudo, pode-se estabelecer novas campanhas das concessionárias, direcionadas para públicos econômicos específicos. Além disso, conhecendo, ou tendo a tendência da classe econômica da família, pode-se também determinar tipos de contratos diferenciados para cada um dos perfis de consumidores.

5.2.2. Classificação da Característica: Tamanho da Residência

Nesse estudo são utilizadas as características das curvas de carga para o treinamento e classificação considerando-se o tamanho da residência. A Figura 65 é uma representação gráfica do que consiste esse segundo estudo de classificação de característica da família.

Figura 65 – Representação gráfica do ensaio de classificação do tamanho da residência.

Os sistemas inteligentes OPF e Redes Bayesianas foram treinados e testados com 25/75%, 50/50% e 75/25%, com dados normalizados e não normalizados, seguindo a lógica

de dez rodadas de testes para ser tirado a média dos resultados. A Tabela 30 apresenta os resultados de média de acurácia das ferramentas computacionais inteligentes.

Tabela 30 – Média da acurácia para estudo de classificação da característica tamanho da residência

		25% T. N. N.	25% T. N.	50% T. N. N.	50% T. N.	75% T. N. N.	75% T. N.
Norte	OPF	80,26%	67,89%	73,05%	67,74%	80,81%	67,41%
	RB	80,50%	68,75%	73,24%	67,88%	68,02%	67,58%
Nordeste	OPF	78,39%	67,69%	72,58%	67,10%	68,58%	67,55%
	RB	78,55%	67,84%	73,44%	67,55%	69,47%	67,43%
Centro-Oeste	OPF	70,02%	71,76%	68,25%	72,27%	67,44%	72,41%
	RB	66,88%	72,04%	66,02%	72,73%	67,71%	72,41%
Sudeste	OPF	81,32%	71,88%	80,92%	72,52%	68,42%	71,40%
	RB	81,41%	72,08%	81,20%	72,89%	68,60%	71,77%
Sul	OPF	71,77%	73,01%	80,90%	73,35%	80,94%	71,80%
	RB	70,22%	73,36%	80,96%	73,56%	81,18%	72,07%

Verifica-se na Tabela 30 que para a Região Norte as ferramentas computacionais inteligentes possuem a mesma configuração para a pior média de acurácia em 75% T.N., com 67,41% para OPF e 67,58% para RB. Já a melhor configuração do OPF ocorreu em 75% T.N.N. e da Rede Bayesiana foi de 25% T.N.N., respectivamente, com valores de acurácia de 80,81% e 80,50%.

A configuração 25% de treinamento não normalizado foi a melhor na Região Nordeste e na Sudeste para as duas ferramentas inteligentes, sendo no Nordeste 78,39% OPF e 78,55% RB e no Sudeste 81,32% OPF e 81,41% RB. As piores configurações no Nordeste foram 50% T.N. para OPF (67,10% de acurácia) e 75% T.N. para RB (67,43% de acurácia). Enquanto que a pior configuração do Sudeste foi de 75% T.N.N. com 68,42% do OPF e 68,60% da RB.

Um detalhe interessante é que para a Região Sul as configurações foram opostas à Sudeste, ou seja, a melhor foi a de 75% T.N.N. para as duas ferramentas e a pior foi a de 25% T.N.N. também para as duas ferramentas, com 80,94% e 71,77% para o OPF, e 81,18% e 70,22 para as Redes Bayesianas.

Por fim, na Região Centro-Oeste as ferramentas inteligentes tiveram comportamento distintos, enquanto para o OPF a melhor configuração foi 75% T.N. (72,41%) e a pior foi 75% T.N.N. (67,44%), as Redes Bayesianas teve melhor configuração em 50% de

treinamento normalizado (72,73% de acurácia) e a pior em 50% de treinamento não normalizado com 66,02% de acurácia.

Destaca-se nesse estudo que o último valor de acurácia, 66,02% da Região Centro-Oeste é o menor de todos dessa Tabela 30, e diferentemente dos estudos anteriores, passou mais distante dos 50,01% do estudo de classe de consumo e classe econômica. Espera-se que esses bons resultados se reflitam na matriz de classificação que será avaliada adiante.

Mais uma vez fica claro que cada caso deve ser analisado de maneira individualizada e a escolha da melhor configuração de tamanho de dados de treinamento e se os dados devem ser tratados anteriormente dependem dos dados de cada região e do objetivo final da classificação, se o ideal é acertar todas as classes, ou se possuem algumas que devem estar mais certas do que outras, uma vez que o banco de dados desse estudo é bastante desbalanceado.

Analisando-se somente a acurácia, verifica-se que os sistemas inteligentes entenderam o problema de classificação da característica tamanho da residência tendo o perfil de consumo como entrada. Todas as regiões possuem ao menos uma configuração acima de 70% de acurácia, sendo então os resultados satisfatórios. Entretanto, uma análise mais apurada através da matriz de classificação será realizada para determinar as tendências e as confusões dos sistemas para esse estudo.

Assim sendo, a Região Sudeste por ser a mais impactante no número de pessoas, e também na parte econômica, será avaliada através da matriz de classificação. Salienta-se que está sendo apresentada a matriz de classificação com melhor resultado, 25% de treinamento, 75% de teste, não normalizado. A complexidade dessa segunda análise está na quantidade de dados de entrada e saída que as ferramentas tem que relacionar. No total são dez características de entrada (classe de consumo, potência instalada, demanda máxima, demanda média, fator de carga, fator de demanda, número de picos, número de vales, picos acima da média e vales acima da média) e doze tamanhos de residência (classes para classificação).

Tabela 31 – Matriz de Classificação para estudo de classificação da característica tamanho da residência - Sudeste

Matriz de Confusão Bayesiana (Sudeste - Tamanho)

T1	97,9%	2,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
T2	0,2%	55,6%	43,8%	0,2%	0,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
T3	0,0%	28,3%	71,7%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
T4	0,0%	0,2%	0,0%	49,5%	50,3%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
T5	0,0%	0,0%	0,0%	44,4%	55,6%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
T6	0,0%	0,0%	0,0%	0,0%	0,0%	56,7%	43,3%	0,0%	0,0%	0,0%	0,0%	0,0%
T7	0,0%	0,0%	0,0%	0,0%	0,0%	46,0%	54,0%	0,0%	0,0%	0,0%	0,0%	0,0%
T8	0,0%	0,0%	0,0%	0,0%	0,0%	0,1%	0,1%	88,4%	11,4%	0,0%	0,0%	0,0%
T9	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	45,5%	54,5%	0,0%	0,0%	0,0%
T10	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,1%	99,9%	0,0%	0,0%
T11	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,1%	53,6%	46,3%
T12	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	39,4%	60,6%

Matriz de Confusão OPF (Sudeste - Tamanho)

T1	97,9%	2,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
T2	0,2%	55,6%	43,9%	0,2%	0,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
T3	0,0%	28,9%	71,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
T4	0,0%	0,2%	0,0%	49,5%	50,3%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
T5	0,0%	0,0%	0,0%	44,8%	55,2%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
T6	0,0%	0,0%	0,0%	0,0%	0,0%	56,4%	43,6%	0,0%	0,0%	0,0%	0,0%	0,0%
T7	0,0%	0,0%	0,0%	0,0%	0,0%	45,6%	54,4%	0,0%	0,0%	0,0%	0,0%	0,0%
T8	0,0%	0,0%	0,0%	0,0%	0,0%	0,1%	0,1%	87,7%	12,1%	0,0%	0,0%	0,0%
T9	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	46,1%	53,9%	0,0%	0,0%	0,0%
T10	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,1%	99,9%	0,0%	0,0%
T11	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,1%	53,0%	46,9%
T12	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	38,4%	61,6%

Legenda

- 1 – Maior Valor Percentual – Verde
- 2 - Igual a zero - Branco
- 3 - Se maior que 60% do maior - intensidade maior de azul
- 4 - Se está entre 30% até 60% do maior - intensidade média azul
- 5 - Se menor que 30% do maior - intensidade fraca do azul

Ao analisar as matrizes de classificação, a classe com maior acerto foi a de tamanho 1 e tamanho 10 (97,9% e 99,9% na Rede Bayesiana e no OPF), e a de maior confusão foi a classe de tamanho 4, para as duas ferramentas. Observa-se, assim como no estudo anterior, que as confusões estão relacionadas às suas vizinhanças, indicando que a ferramenta está entendendo o problema, e está criando tendências de criação de perfis.

Comparando-se os tamanhos de residência 1 e 2, observa-se que mais de 90% dos dados de T1 estão classificados como classe D, D- e E+, enquanto que apenas 20% dos dados

de T2 estão nessa condição. Assim, a confusão entre estas classes é menor. Já a diferença entre as classes de tamanho 9, 10 e 11 não está presente na classe de consumo, mas sim na potência instalada. Cada uma é bem diferente da outra, pois nessas configurações há mudança de cômodos e consequentemente, mudança de potência instalada.

As confusões das demais classes se dá pela falta de diferenciação entre as variáveis selecionadas. Entretanto, a confusão é considerado somente com seus vizinhos. A característica mais marcante que determina esse fato é a potência instalada, que está relacionada com o número de ambientes que há na casa e consequentemente, a posse de equipamentos, número de tomadas e número de lâmpadas.

Observando-se a matriz de classificação, pode-se concluir que as duas ferramentas tiveram comportamentos bem semelhantes, com pequenas diferenças percentuais entre algumas classes. Assim, para se compará-las novamente, é proposto a Tabela 32 de classificação quanto a quantidade de acertos, a quantidade e ‘qualidade’ das confusões criadas para a classificação do tamanho da residência, tendo como entrada as características de consumo.

Tabela 32 – Avaliação das Ferramentas Inteligentes para Classificação da Tamanho da Residência - Sudeste

	% Acerto Diagonal Principal	Quantidade de Confusão (n°)		‘Qualidade da confusão (%)’	
RB	91,67	Azul Escuro	8	Azul Escuro	44,77
		Azul Médio	1	Azul Médio	28,30
		Azul Claro	10	Azul Claro	1,46
OPF	91,67	Azul Escuro	8	Azul Escuro	44,85
		Azul Médio	1	Azul Médio	28,90
		Azul Claro	10	Azul Claro	1,53

Novamente, percebe-se que as Redes Bayesianas são um pouco melhores do que o classificador OPF devido a ‘qualidade’ de confusão menor. Esse estudo possibilita estipular, ao se conhecer o perfil de consumo da residência, além do tamanho da casa, uma previsão de potência instalada. Assim, é possível melhorar o dimensionamento e a instalação de novas residências.

5.2.3. Classificação da Característica: Composição/faixa-etária

Para este estudo de classificação de faixa etária, mais uma vez são utilizadas as características das curvas de carga para o treinamento e classificação. A Figura 66 é uma representação gráfica do que consiste esse terceiro estudo de classificação de característica da família.

Figura 66 – Representação gráfica do ensaio de classificação da composição/faixa-etária.

Mais uma vez, os sistemas inteligentes OPF e Redes Bayesianas foram treinados e testados com 25/75%, 50/50% e 75/25%, com dados normalizados e não normalizados seguindo a lógica de dez rodadas de testes para ser tirado a média dos resultados.

A Tabela 33 apresenta os resultados de acurácia das ferramentas para esse estudo.

Tabela 33 – Média da acurácia para estudo de classificação da característica faixa-etária

		25% T. N. N.	25% T. N.	50% T. N. N.	50% T. N.	75% T. N. N.	75% T. N.
Norte	OPF	53,91%	54,42%	71,65%	54,66%	73,35%	54,82%
	RB	50,41%	54,45%	71,86%	54,67%	74,05%	54,83%
Nordeste	OPF	51,28%	51,24%	71,74%	51,18%	58,24%	51,23%
	RB	51,28%	51,23%	71,68%	51,19%	58,35%	51,24%
Centro-Oeste	OPF	54,64%	64,79%	55,22%	57,22%	50,79%	57,36%
	RB	54,66%	64,74%	55,26%	57,28%	55,50%	57,46%
Sudeste	OPF	81,25%	57,66%	55,88%	58,18%	51,31%	69,91%
	RB	80,67%	57,73%	55,92%	58,29%	51,32%	69,98%
Sul	OPF	56,87%	57,96%	50,00%	58,28%	56,88%	59,01%
	RB	51,62%	58,06%	56,37%	58,44%	56,96%	59,12%

Avaliando-se a Tabela 33 pode-se destacar a configuração 25% T.N.N da Região Sudeste com acurácia de 81,25% para o OPF e 80,67% para as Redes Bayesianas. Tanto que um estudo mais aprofundado dessa configuração será realizado através da matriz de classificação.

Novamente percebe-se que a quantidade de dados de treinamento e se esses foram ou não tratados antes de serem utilizados não influencia na performance das ferramentas inteligentes quanto a acurácia.

As melhores configurações foram: para Norte 75% T.N.N., Nordeste 50% T.N.N., Centro-Oeste 25% T.N., Sudeste 25% T.N.N., e Sul 75% T.N., com acurácias de 73,35%, 74,05% no Norte para OPF e RB, No Nordeste 71,74% e 71,68%, no Centro-Oeste 64,79% e 64,74%, Sudeste 81,25% e 80,67% e no Sul com 59,01% e 59,12% de acurácia, respectivamente OPF e Redes Bayesianas.

Já as piores configurações devido a média de acurácia para cada uma das regiões é: Norte 25% T.N.N (53,91 OPF e 50,41% RB), Nordeste 50% T.N. (51,18% OPF e 51,19% RB), Centro-Oeste 75% T.N.N (50,79% o OPF) e 25% T.N.N (54,66% a RB), no Sudeste a pior foi 75% T.N.N. (51,31% OPF e 51,32% RB) e por fim no Sul o OPF foi para a 50% T.N.N. (50,00% de acurácia) e RB foi para a 25% T.N.N (51,62% de acurácia).

Ou seja, para esse estudo deve-se estudar cada região separadamente, pois os dados estão atrelados a região, seus costumes, sua cultura e seus aspectos socioeconômicos.

Pela análise de acurácia, esse estudo de composição apresentou uma classificação um pouco pior. Isto se deve pela escolha das classes. Uma vez que há apenas cinco tipos de integrantes e as classes são a combinação entre eles, existe muita confusão entre as classes, ou seja, tanto a acurácia quanto a matriz de classificação terão resultados aparentemente ruins. No entanto, deve-se obter uma discussão muito boa para algumas classes na medida em que, para algumas configurações de análise, obteve uma boa resposta de acurácia.

A complexidade deste estudo de faixa etária está em avaliar todas as combinações juntas, totalizando quarenta e duas classes de combinações, sendo que existem cinco tipos de integrantes (adulto, idoso, adolescente, nenê e criança), ou seja, os perfis de consumo se misturam entre as combinações. É apresentado a seguir a matriz confusão desse estudo. Como esta matriz é muito grande, ela foi colocada apenas para mostrar os erros das classificações através das cores.

Tabela 54 – Matriz de Classificação para estudo de Classificação da característica faixa-etária - Sudeste

Matriz Confusão Bayesiana (Sudeste - Composição)

1	0.000	0.008	0.026	0.043	0.034	0.047	0.043	0.04	0.04	0.007	0.012	0.01	0.007	0.009	0.007	0.021	0.014	0.018	0.003	0.006	0	0.005	0.012	0.005	0.005	0.005	0.002	0	0	0	0.002	0.002	0.005	0.002	0.003	0.003	0.003	0.003	0.002
2	0.010	0.005	0.009	0.022	0.023	0.029	0.057	0.033	0.05	0.033	0.012	0.012	0.005	0.012	0.007	0.003	0.012	0.007	0.005	0.003	0.003	0.005	0.012	0.005	0.003	0.009	0.002	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
3	0.010	0.015	0.010	0.041	0.043	0.043	0.061	0.040	0.040	0.031	0.011	0.012	0.009	0.004	0.009	0.020	0.014	0.011	0.002	0.002	0.004	0.002	0	0.004	0.007	0.005	0.007	0	0.004	0.002	0.002	0.004	0	0	0.005	0	0.002	0.002	0.005
4	0.002	0.002	0.004	0.030	0.030	0.030	0.040	0.030	0.030	0.030	0.001	0.002	0.006	0.025	0.030	0.02	0.025	0.022	0.009	0.02	0.005	0.011	0.007	0.011	0.007	0.002	0.007	0.005	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	
5	0.004	0.014	0.007	0.040	0.040	0.040	0.050	0.040	0.040	0.040	0.001	0.002	0.025	0.025	0.025	0.028	0.028	0.025	0.009	0.01	0.005	0.011	0.009	0.011	0.009	0.010	0.009	0.007	0.007	0.005	0.002	0.005	0.002	0.005	0.002	0.005	0.014		
6	0.041	0.045	0.040	0.084	0.084	0.084	0.095	0.084	0.084	0.084	0.021	0.041	0.025	0.014	0.03	0.023	0.023	0.028	0.027	0.011	0.009	0.011	0.011	0.009	0.005	0.010	0.009	0.007	0.007	0.002	0.002	0.007	0.007	0	0.005	0.002	0.011		
7	0.041	0.045	0.040	0.084	0.084	0.084	0.095	0.084	0.084	0.084	0.021	0.041	0.025	0.014	0.03	0.023	0.023	0.028	0.027	0.011	0.009	0.011	0.011	0.009	0.005	0.010	0.009	0.007	0.007	0.002	0.002	0.007	0.007	0	0.005	0.002	0.011		
8	0.041	0.045	0.040	0.084	0.084	0.084	0.095	0.084	0.084	0.084	0.021	0.041	0.025	0.014	0.03	0.023	0.023	0.028	0.027	0.011	0.009	0.011	0.011	0.009	0.005	0.010	0.009	0.007	0.007	0.002	0.002	0.007	0.007	0	0.005	0.002	0.011		
9	0.041	0.045	0.040	0.084	0.084	0.084	0.095	0.084	0.084	0.084	0.021	0.041	0.025	0.014	0.03	0.023	0.023	0.028	0.027	0.011	0.009	0.011	0.011	0.009	0.005	0.010	0.009	0.007	0.007	0.002	0.002	0.007	0.007	0	0.005	0.002	0.011		
10	0.041	0.045	0.040	0.084	0.084	0.084	0.095	0.084	0.084	0.084	0.021	0.041	0.025	0.014	0.03	0.023	0.023	0.028	0.027	0.011	0.009	0.011	0.011	0.009	0.005	0.010	0.009	0.007	0.007	0.002	0.002	0.007	0.007	0	0.005	0.002	0.011		
11	0.041	0.045	0.040	0.084	0.084	0.084	0.095	0.084	0.084	0.084	0.021	0.041	0.025	0.014	0.03	0.023	0.023	0.028	0.027	0.011	0.009	0.011	0.011	0.009	0.005	0.010	0.009	0.007	0.007	0.002	0.002	0.007	0.007	0	0.005	0.002	0.011		
12	0.022	0.01	0.025	0.036	0.036	0.036	0.045	0.036	0.036	0.036	0.001	0.002	0.025	0.025	0.025	0.028	0.028	0.025	0.009	0.01	0.005	0.011	0.009	0.011	0.009	0.010	0.009	0.007	0.007	0.005	0.002	0.005	0.002	0.005	0.002	0.005	0.014		
13	0.017	0.017	0.007	0.034	0.034	0.034	0.043	0.034	0.034	0.034	0.001	0.002	0.034	0.034	0.034	0.037	0.037	0.034	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
14	0.021	0.021	0.009	0.036	0.036	0.036	0.045	0.036	0.036	0.036	0.001	0.002	0.036	0.036	0.036	0.040	0.040	0.036	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
15	0.017	0.017	0.007	0.034	0.034	0.034	0.043	0.034	0.034	0.034	0.001	0.002	0.034	0.034	0.034	0.037	0.037	0.034	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
16	0.01	0.007	0.016	0.026	0.026	0.026	0.032	0.026	0.026	0.026	0.001	0.002	0.026	0.026	0.026	0.029	0.029	0.026	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
17	0.01	0.013	0.016	0.026	0.026	0.026	0.032	0.026	0.026	0.026	0.001	0.002	0.026	0.026	0.026	0.029	0.029	0.026	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
18	0.018	0.009	0.018	0.034	0.034	0.034	0.043	0.034	0.034	0.034	0.001	0.002	0.034	0.034	0.034	0.037	0.037	0.034	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
19	0.009	0.008	0.012	0.026	0.026	0.026	0.032	0.026	0.026	0.026	0.001	0.002	0.026	0.026	0.026	0.029	0.029	0.026	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
20	0.002	0.002	0.002	0.023	0.023	0.023	0.028	0.023	0.023	0.023	0.001	0.002	0.023	0.023	0.023	0.026	0.026	0.023	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
21	0.001	0.007	0.001	0.02	0.02	0.02	0.023	0.02	0.02	0.02	0.001	0.002	0.02	0.02	0.02	0.023	0.023	0.02	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
22	0.001	0.007	0.001	0.02	0.02	0.02	0.023	0.02	0.02	0.02	0.001	0.002	0.02	0.02	0.02	0.023	0.023	0.02	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
23	0.001	0.007	0.001	0.02	0.02	0.02	0.023	0.02	0.02	0.02	0.001	0.002	0.02	0.02	0.02	0.023	0.023	0.02	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
24	0.001	0.007	0.001	0.02	0.02	0.02	0.023	0.02	0.02	0.02	0.001	0.002	0.02	0.02	0.02	0.023	0.023	0.02	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
25	0.003	0.01	0.007	0.017	0.017	0.017	0.023	0.017	0.017	0.017	0.001	0.002	0.017	0.017	0.017	0.023	0.023	0.017	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
26	0.017	0.01	0.003	0.017	0.017	0.017	0.023	0.017	0.017	0.017	0.001	0.002	0.017	0.017	0.017	0.023	0.023	0.017	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
27	0.003	0.003	0.003	0.016	0.016	0.016	0.02	0.016	0.016	0.016	0.001	0.002	0.016	0.016	0.016	0.02	0.02	0.016	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
28	0.01	0.013	0.013	0.017	0.017	0.017	0.023	0.017	0.017	0.017	0.001	0.002	0.017	0.017	0.017	0.023	0.023	0.017	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
29	0.007	0.01	0.01	0.021	0.021	0.021	0.028	0.014	0.014	0.014	0.017	0.014	0.017	0.014	0.017	0.014	0.017	0.014	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
30	0.007	0.007	0	0.014	0.021	0.007	0.014	0.014	0.014	0.007	0	0.008	0.021	0.008	0.021	0.007	0.042	0.008	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
31	0.007	0.007	0	0.014	0.021	0.007	0.014	0.014	0.014	0.007	0	0.008	0.021	0.008	0.021	0.007	0.042	0.008	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004	
32	0.007	0.014	0.014	0.007	0.026	0.01	0.007	0.007	0.007	0.007	0.021	0.01	0.007	0.026	0.01	0.007	0.042	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.002	0.013	0.004	0.002	0.011	0.007	0.009	0	0.004	0	0.002	0.004	0.004		
33	0.007	0.007	0	0.013	0	0.02	0.007	0.027	0	0	0.02	0.013	0.02	0.013	0.02	0.013	0.027	0.013	0.007	0.009	0.015	0.016	0.013	0.009	0.007	0.													

Matriz Confusao UPF (Sudeste - Composicao)

[illegible]

Apenas avaliando a distribuição das cores nas células, é possível verificar que a classificação em muitas classes semelhantes é muito confusa. As ferramentas inteligentes não conseguiram determinar um padrão, nem uma tendência entre as classes.

Portanto, para definir melhor as composições, e realizar novos estudos, a ideia é diminuir o número de classes, e definir exatamente quais variáveis afetam mais nesse problema, ou mesmo, avaliar separadamente as combinações de um integrante, depois as de dois integrantes, de três integrantes e de quatro integrantes.

6. CENÁRIOS DE APLICAÇÃO DOS RESULTADOS

Com a chegada de um novo cenário de *smart grid* e *demand response*, conhecer os consumidores e a sua participação no mercado de energia se tornam essenciais, assim, os estudos apresentados nessa tese comprovam que uma análise mais profunda, considerando aspectos sociais, culturais e econômicos são importantes indicadores de consumo e podem auxiliar no estreitamento de relacionamento entre concessionárias e consumidores.

Neste sentido, será apresentado dois cenários onde seria possível utilizar os resultados desse estudo para melhorar o planejamento de distribuição de energia de baixa tensão:

- Metodologia de carregamento de transformadores; e
- Participação dos consumidores e a relação Concessionária/Consumidores.

Primeiramente será apresentado a metodologia utilizada, por uma concessionária, para o planejamento de novas cargas e o estudo de carregamento dos transformadores.

6.1 Planejamento de Novas Cargas na Rede de Distribuição de Energia

O dimensionamento de transformadores na rede de distribuição de baixa tensão está fundamentado na aplicação de curvas típicas de carga para consumidores residenciais, comerciais e industriais. As descrições de planejamento de dimensionamento de transformadores exposto está baseado na NT-2.018 (Nota Técnica – Cálculo de Demanda Rede para Acréscimo de Novas Cargas na Rede de Distribuição de Baixa Tensão).

Neste sentido, quanto mais precisa for a estimativa para atendimento das novas cargas, melhor será o resultado da estimativa de demanda. A maior parte das concessionárias no Brasil utilizam de parâmetros, como a carga declarada pelo cliente, o fator de demanda e fatores e curvas de carga típica de cada classe (JARDINI *et al.*, 2000).

A definição da potência nominal do transformador, a ser instalado em nova estação transformadora, é feita em função da demanda calculada, que é baseada na carga declarada pelo cliente, acrescida de fatores de correção.

Estudos de segurança de dimensionamento nas estações transformadoras são feitas através do sistema GIS – Sistema de Informações Geográficas, aonde são cadastrados todos os ativos elétricos da empresa ou através de medidas e planilhas de cálculo de estimativa de demanda com base em energia consumida (kWh) (ROSSETI, 2015).

A demanda a ser considerada no carregamento de transformadores para atendimento de novos consumidores, deve ser calculada conforme as seguintes etapas:

- Determinação do consumo (kWh):

Com base nas informações de projeto, da carga instalada, da atividade, da classe de consumo e de fatores de demanda e de carga, são estimados os consumos das novas unidades.

Os fatores de demanda típicos e fator de carga típico para consumidores residenciais possuem valores tabelados, e seus valores estão atrelados a potência instalada, que por sua vez dependem do projetista e do cliente, que muitas vezes sobre dimensionam a carga.

- Determinação da curva de carga diária:

O consumo estimado deve ser convertido em kVA médio e multiplicado por valores tabelados de hora em hora, de acordo com a classe de consumo correspondente ao consumo estimado, assim, obtendo uma curva de carga média.

Esses valores tabelados são obtidos de estudos anteriores e devem ser constantemente atualizados para manter coerência com a substituição e atualização das cargas devido ao desenvolvimento de novas tecnologias e novas tendências de consumo.

- Acréscimo de carga em uma estação transformadora existente:

O valor de pico da curva de carga resultante, representa a demanda estimada que será acrescida na rede de distribuição de baixa tensão. Estes valores devem ser estudados no sistema GIS para verificar se há a necessidade de alterações ou não na rede secundária. Além disso, os parâmetros de queda de tensão e ampacidade do secundário devem sempre ser verificados.

A demanda da nova ligação por transformador depende do tipo de transformador e de suas ligações, se é monofásico, ou está ligando em delta aberto, delta fechado, ou trifásico.

Ao analisar a metodologia proposta nessa nota técnica, observa-se que para consumidores residenciais há uma dependência total do valor de potência instalada, que dependem do projeto, que muitas vezes são feitos sem o conhecimento das famílias que irão residir na casa, ou seja, não representam a correta demanda. E por isso, geralmente o planejamento da demanda de consumidores residenciais são sobre dimensionados.

Mesmo que, para as estações elevadoras, a inclusão de clientes residenciais não afetem na demanda dos transformadores de forma significativa, é importante ter conhecimento prévio

da demanda de novas cargas o mais próximo da realidade. Com isso, o dimensionamento de transformadores mais próximos aos clientes de baixa tensão será mais preciso e a concessionária poderá traçar planos e estratégias de participação dos clientes, ainda mais quando a tendência é de as redes assumirem conceitos de *smart grids* e *demand response*.

6.2 Metodologia de Carregamento de Transformadores

O dimensionamento de novos transformadores devido a entrada de novas cargas dependem da demanda estipulada, que para a classe residencial, está diretamente atrelada a potência instalada, que por sua vez dependem do projeto de instalações, e que normalmente não contam com as especificações das pessoas que irão morar na casa, fazendo com que muitas vezes a estipulação seja de forma pouco precisa. Também foi comentado que as tabelas de conversão de curva de carga devem sempre estar sendo atualizadas devido ao avanço tecnológico e devido às mudanças de hábitos das pessoas.

Assim sendo, a utilização das ferramentas inteligentes para previsão de consumo pode auxiliar na metodologia de carregamento dos transformadores.

Conforme discutido no Capítulo 5, apenas com as características da família já é possível verificar a tendência de classe de consumo da família. Assim, com esse perfil é possível melhorar as estimativas que antes só levavam em consideração a potência instalada definida pelo projetista.

Utilizando o exemplo apresentado na nota técnica 2.108 (VILELLA, CARVALHO, 2007), página 12 (link em anexo), onde uma residência de aproximadamente 200 m², com potência instalada de 10,40 kW, foi estipulada como tendo um consumo mensal de 248,8 kWh, e sua curva de demanda é montada através de tabelas para classe de consumo entre 201 kWh até 500 kWh (Anexo III da NT2.108).

Ao se considerar que para ter uma residência de 200 m² a família q deve ter condições econômicas acima de D, ou seja, entre as faixas A, B e C, e considerando que a média de habitantes do Brasil por residência está entre dois e três, ao utilizar as ferramentas inteligentes propostas, pode-se indicar diferentes classes de consumo de acordo com a classe econômica e o número de pessoas, podendo variar entre aproximadamente 289 kWh (duas pessoas classe C) até aproximadamente 586 kWh (três pessoas classe A).

Levando-se em conta os valores apresentados na nota técnica, quase todo *range* (289-586 kWh) seria classificado em uma única classe, pois a metodologia aplicada na nota possui apenas cinco classes. Portanto, para atender todos os possíveis clientes propostos pela ferramenta, seria necessário ter ao menos duas curvas de carga (201 a 500 kWh e outra de 501 a 1000 kWh). O que poderia ocasionar um sobre dimensionamento para casos de menor consumo e um sub dimensionamento nos casos de maior consumo.

Contudo, considerando as classes propostas, seriam necessárias cinco perfis de consumo (260 a 300 kWh, 300 a 370 kWh, 370 a 440 kWh, 440 a 500 kWh e 500 a 700 kWh) de acordo com a configuração e os dados da família, diminuindo assim o estudo de carregamento dos transformadores.

6.3 Participação dos Consumidores e a Relação Concessionária/Consumidores.

A participação dos consumidores no mercado de energia está sendo muito discutida mundialmente. Com o avanço da tecnologia, e dos conceitos de *smart grid* e *demand response*, os consumidores deverão interagir mais com as concessionárias e para isso as concessionárias deverão ter um maior conhecimento sobre seus consumidores.

A intensificação de sensores na rede de energia, e o comércio bidirecional, ou seja, hora o consumidor estará consumindo, hora estará fornecendo energia para a rede, são conceitos de *smart grid* e *demand response*. Atrelado a isso, a competição do mercado, e as constantes crises econômicas vivenciadas no Brasil nos últimos anos, além do aumento de informação proporcionado pela internet, faz com que a população estude mais sobre os assuntos que envolvam dinheiro, e passem a exigir melhores qualidades de fornecimento de energia, melhores condições de pagamento e briguem por seus direitos devido a falta de energia, ou mesmo um mal atendimento de serviços prestados.

A medição do consumo de energia individualiza através de medidores inteligentes pode conter informações muito importante para que as concessionárias, sendo possível identificar os equipamentos que existem na residência, podem haver controles de cargas específicas durante o dia, para que os transformadores não trabalhem tanto no vazio, e não sobrecarreguem em horários de pico, ou mesmo, conforme apresentado no capítulo anterior, seja possível identificar algumas características da família através da análise do seu comportamento de consumo.

Quando esses dados são relacionados entre si, através de um estudo sistemático, eles se transformam em informações preciosas. Se inseridas em um contexto determinado, elas tornam-se fonte de conhecimento para ser utilizado como base de um processo decisório. Dessa forma, evitam-se muitos riscos de mercado e consegue ser mais assertivo e seguro nos planos de ações a serem tomadas pelos setores da empresa.

Os bancos de dados poderão ser utilizados para determinar várias iniciativas da empresa, como, por exemplo, uma política de promoções personalizada, ações de *marketing* mais bem segmentadas e até os projetos de expansão dos negócios em outras áreas. Isso porque essas informações, permitem identificar quem é cada cliente, quais são as suas preferências e o ímpeto de consumo.

7. CONCLUSÕES FINAIS

7.1 Aspectos Gerais

Alguns levantamentos feitos sobre o crescimento da demanda de energia elétrica nos próximos anos no Brasil reforça a importância do conhecimento dos hábitos de consumo e da posse de equipamentos dos diversos tipos de consumidores frente ao planejamento do sistema elétrico. Aliado a isso, as dificuldades econômicas e ambientais na construção de novas fontes de energia, incentivam estudos para se melhorar a eficiência energética do sistema e conscientizar mais a população sobre economia de energia, ou mesmo o seu uso de forma racional.

A demanda de energia da classe residencial vem crescendo, e já atinge em torno de 28% do total de energia elétrica consumida, tornando-se uma classe a ser estudada de forma mais intensiva. Sabe-se que esse estudo é considerado complexo, devido principalmente, ao livre arbítrio de cada integrante em escolher a posse de equipamento e seus hábitos de consumo.

Devido à dificuldade imposta pelas concessionárias e pelos direitos dos consumidores, em disponibilizar informações sobre aspectos econômicos, sociais, culturais e técnicos, relacionando-os com o perfil de consumo, o desenvolvimento do algoritmo que simule essas características foi extremamente importante, sendo a base para essa pesquisa. Pois com os dados virtuais foi possível, mesmo que de forma sucinta, analisar a influência que cada característica tem sobre o consumo final, melhorando o planejamento de novas redes de distribuição, bem como estreitando o conhecimento das empresas em relação aos seus clientes.

O algoritmo desenvolvido para gerar curvas de carga, que leva em consideração aspectos sociais como número de pessoas, e faixa etária dos integrantes da família, aspectos culturais como região, hábitos de consumo e tempo em casa, e aspectos econômicos como classe econômica e tamanho da residência, obteve resultados satisfatórios ao se comparar com as médias estudadas pelo PROCEL, bem como, com medidas realizadas em campo. Tendo erros percentuais satisfatórios.

Após a “validação” do algoritmo, a proposta de se tentar entender como as características das famílias selecionadas influenciavam na classificação de consumo teve resultados significativos, mostrando que no Brasil, devido sua grande extensão territorial possui padrões de consumo bem distintos. As regiões Norte e Nordeste, por exemplo, sofrem

maior impacto da característica econômica, pois o equipamento que mais consome energia nessas regiões são os condicionadores de energia, que é considerado um bem caro e de característica coletiva. Já na região Sudeste e Centro-Oeste, a característica que afeta de forma mais incisiva é o número de pessoas, pois o “vilão” de consumo são os chuveiros elétricos, que são considerados bens individuais. A região Sul possui características das outras duas duplas citadas, possuindo chuveiros e também aquecedores de ambiente.

Além das características econômicas e números de pessoas, o estudo de tamanho de residência apresentou o resultado esperado, que a relação é direta, ou seja, aumentando o tamanho, aumenta-se o consumo, e que nas regiões mais desenvolvidas no país (Sudeste e Sul) esse impacto é um pouco maior.

Em relação ao estudo de faixas etárias / composição familiar, ficou evidente que a combinação entre adulto e idoso, e adulto adolescente, consome mais do que a combinação com crianças e nenéns, pois muitas vezes os menores ficam em escolas e creches.

Para todas as análises realizadas foi proposto a utilização de classificação de consumo com treze classes ao invés de cinco, que normalmente é utilizado pelas concessionárias, com intuito de se melhorar a interpretação da migração entre as classes, e melhorar a definição de padronização dos perfis de consumo. Essa etapa é considerada inovadora e inédita, não tendo sido aplicada em nenhuma pesquisa anteriormente com essa tônica.

A proposta também contemplou a análise de ferramentas computacionais inteligentes para classificação, sendo utilizadas duas ferramentas, as Redes Bayesianas e o classificador OPF. Esse estudo apresentou resultados muito promissores para sua utilização em planejamento de distribuição e também em ensaios de carregamento de transformadores. Foram realizados dois estudos, um que relacionava as características das famílias com a classe de consumo, e o outro que relacionava dados técnicos do perfil de consumo com as características das famílias.

No primeiro, os bons resultados de acurácia, atrelados a uma matriz de classificação bem definida, comprovam que o estudo é promissor, e que pode ajudar bastante no planejamento de distribuição, através da previsão de consumo de novos consumidores de energia. As concessionárias não utilizam dados econômicos e sociais para estudar o carregamento dos transformadores, a previsão de demanda é realizada através do cálculo médio da potência instalada, e está diretamente relacionada com o planejamento e

documentos de cada uma das casas. A boa classificação apresentada nesse estudo possibilita uma melhoria na previsão de classe de consumo das residências.

No segundo estudo foi proposto avaliar três características separadamente, sendo que para a característica de classe econômica, as ferramentas inteligentes tiveram um bom resultado, para o tamanho da residência um resultado ainda melhor, já para a composição familiar o resultado foi razoável, devido toda a complexidade de análise. Entretanto, observa-se que esse estudo pode ser aprofundado, através da alteração do banco de dados, e melhoria de treinamento da ferramenta inteligente. Haja visto que, esses dados das características do consumidor, vão ser muito importante para as concessionárias em um cenário de *demand response*, onde a participação do cliente no mercado de energia, exige que a concessionária estreite sua relação com seus clientes.

Por fim, foram propostos dois cenários, um para mostrar que uma metodologia que utilize as ferramentas inteligentes para prever a classe de consumo, levando em consideração as características das famílias, pode ser mais precisa do que a metodologia que está sendo aplicada hoje, uma vez que considera apenas a potência instalada. E outro cenário, levando em conta a concretização dos conceitos de *smart grid* e *demand response*, as concessionárias precisarão estreitar seu relacionamento com os clientes, onde as ferramentas inteligentes podem ser grandes aliadas na identificação de características das famílias, através de dados de medidas dos *smart meters* do perfil de consumo da residência.

Diante do exposto, pode-se vislumbrar que esse tipo de estudo auxiliará de forma consistente na análise do planejamento do sistema elétrico.

7.2 Futuros desdobramentos da pesquisa

Vislumbra-se como uma possibilidade de melhoria no algoritmo de construção de curvas de carga, a utilização de modelagem dos equipamentos da residência, assim, os resultados se aproximariam com as curvas de carga medidas em campo.

Outro possível desdobramento é a utilização da previsão de demanda para auxiliar no estudo de perdas não-técnicas. Ou então, adaptar o algoritmo para gerar curvas virtuais de fraudes de energia, com intuito de treinamento de uma ferramenta computacional inteligente para identificar os perfis de consumidores fraudulentos.

Com o objetivo de melhorar a eficiência do sistema e a relação das concessionárias com seus clientes, um estudo interessante seria a desagregação das curvas de carga virtuais, com o intuito de se conhecer os equipamentos que estão sendo utilizados na residência, e também monitorar quais são os equipamentos que mais consomem energia. Esse estudo também ajudaria no controle das carga em um ambiente de *demand response*.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABEP, **Critério de Classificação Econômica do Brasil**, 2011.
- ABICALAFFE, C.; AMARAL, V. F.; DIAS, J.S. **Aplicação da Rede Bayesiana na Prevenção da Gestão de Alto Risco**. In: Congresso Brasileiro de Informática Médica, Ribeirão Preto. Anais do Congresso Brasileiro de Informática Médica, v. 1. p. 1-1, 2004.
- ALBADI M. H. AND EL-SAADANY E. F., **Demand Response in Electricity Markets: An Overview**, Power Engineering Society General Meeting, 2007. IEEE, Tampa, FL, 2007, pp. 1-5.
- AL-OTAIBI R., JIN N., WILCOT. X AND FLACH P., **Feature Construction and Calibration for Clustering Daily Load Curves from Smart-Meter Data**, in IEEE Transactions on Industrial Informatics, vol. 12, no. 2, pp. 645-654, April 2016.
- ARMSTRONG, M. M.; SWINTON, M. C.; RIBBERINK, H.; BEAUSOLEIL-MORRISON, I.; MILLETTE, J. **Synthetically derived profiles for representing occupant-driven electric loads in Canadian housing**. Journal of Building Performance Simulation, v. 2, n. 1, p. 15–30, 2009.
- BEN-GAL, I. **Bayesian Networks**. Encyclopedia of Statistics in Quality and Reliability, John Wiley & Sons, Ltd, 2015.
- BOBBIO, A.; PORTINALE, L.; MINICHINO, M.; CIANCAMERLA, E. **Improving the Analysis of Dependable Systems by Mapping Fault Trees into Bayesian Networks**. Reliability Engineering & Systems Safety, V.71, p 249-260, 2001.
- CANAVAGH, R.; HOWAT, J. **Finding Common Ground Between Consumer and Environmental Advocates**. Electricity Policy.com 2012.
- CAO H., BECKEL, C. ; STAAKE, T. **Are domestic load profiles stable over time? An attempt to identify target households for demand side management campaigns**. Industrial Electronics Society, IECON 2013 - 39th Annual Conference of the IEEE, 2013.
- CARVALHO, F. **Aprendizagem Bayesiana**. Apresentação de slides. Acesso em: 05 de maio de 2015. Disponível em: <http://www.cin.ufpe.br/~compint/aulas-IAS/kdd-011-/Bayes.ppt>.

- CHANG, K. C.; FUNG, R.; LUCAS, A.; OLIVER R.; SHIKALOFF, N. **Bayesian networks applied to credit scoring**. IMA Journal of Mathematics Applied in Business and Industry. London: Oxford University Press, N. 11, p. 1-18, 2000.
- CHAOUCH M. **Clustering-Based Improvement of Nonparametric Functional Time Series Forecasting: Application to Intra-Day Household-Level Load Curves**, IEEE Transactions on Smart Grid, vol.5, no.1, pp.411-419, January 2014.
- CHARNIAK, E. **Bayesians Networks without Tears**. IA Magazine, 1991.
- CHEN, Y.; HWANG, J. K. **A Reliable Energy Information System for Promoting Voluntary Energy Conservation Benefits**. IEEE Transactions on Power Delivery, vol.21, pp.102-107. January 2016.
- CHENG, J.; GREINER, R. **Comparing bayesian network classifiers**. In: Morgan Kaufmann Publishers, 1990, p. 101-108.
- CHRYSOPOULOS, A., DIOU, C., SYMEONIDIS, A. L., MITKAS, P. A. **Agent-based small-scale energy consumer models for energy portfolio management**, 2013 IEEE/WIC/ACM International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT), 2013.
- DORIGO, M. MANIEZZO, V. COLORNI, A. **The ant system: Optimization by a colony of cooperating agents**, IEEE Transactions on Systems, Man, and Cybernetics-Part B, vol. 26, pp. 29–41, 1996.
- DUDA, R.O.; HART, P. E. **Pattern Classification and Scene Analysis**. [S.I]: John Wiley Sons Inc, 1973. Hardcover.
- EMPRESA DE PESQUISA ENERGÉTICA (EPE). **Balanço Energético Nacional – BEN 2016**. Julho, 2016.
- EMPRESA DE PESQUISA ENERGÉTICA (EPE). **Anuário Estatístico de Energia Elétrica 2015 – ano base 2014**. Setembro de 2015.
- EMPRESA DE PESQUISA ENERGÉTICA (EPE). **Balanço Energético Nacional – BEN 2013**. Junho, 2013.
- EMPRESA DE PESQUISA ENERGÉTICA (EPE). **Caracterização do Cenário Macroeconômico: para os próximos 10 anos (2013 – 2022)**. Nota Técnica DEA 06/13. Série – Estudos Econômicos, Junho, 2013.

- EMPRESA DE PESQUISA ENERGÉTICA (EPE). **Projeção da Demanda de Energia Elétrica. Série – Estudos de Demanda**, Dezembro, 2012.
- EMPRESA DE PESQUISA ENERGÉTICA (EPE) E MINISTÉRIO DE MINAS E ENERGIA (MME). **Plano Nacional de Energia 2030 - Projeções**. Setembro, 2007.
- ENERGY INFORMATION ADMINISTRATION (EIA). **World Energy Outlook**. 2016.
- EXXONMOBIL. **Panorama Energético: Perspectivas para 2040**, 2012.
- FALCÃO, A. X., STOLFI, J., LOTUFO, R. A. **The image foresting transform: Theory, algorithms, and applications**. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(1):19-29, 2004.
- FRANCISQUINI, A. A. **Estimação de Curvas de Carga em Pontos de Consumo e em Transformadores de Distribuição**. Dissertação de Mestrado, Universidade Estadual Paulista, Campus de Ilha Solteira, Departamento de Engenharia Elétrica, 2006.
- FRIEDMAN, N.; GEIGER, D.; GOLDSZMIDT, M. **Bayesian network classifier**. Machine Learning, v. 29, n.2-3, p. 131-163, 1997.
- GASTALDELLO, D. S.; SOUZA, A. N.; AMARAL, H. L. M.; GOTTI, M. A., PAPA, J. P. **Estudo e Análise dos Padrões de Curvas de Carga de Consumidores Residenciais**. Congresso Brasileiro de Automática CBA 2016, Vitória, 2016.
- GASTALDELLO, D. S.; SOUZA, A. N.; AMARAL, H. L. M.; ZAGO, M. G. **Classification of Residential Load Curves in Consumption Classes Using Intelligent Techniques**. Latin-American Congress Electricity Generation and Transmission CLAGTEE 2015, Uruguay, 2015.
- GASTALDELLO, D. S.; SOUZA, A. N.; AMARAL, H. L. M.; FARIA, P.; VALE, Z. **Study of load curves concerning the influence of socioeconomic and cultural issues**. Innovative Smart Grid Technologies Conference Europe ISGT Europe 2014, Turkey, 2014.
- GEEM, W. **Music-Inspired Harmony Search Algorithm: Theory and Applications**, 1st ed. Springer Publishing Company, Incorporated, 2009.
- GHEYDI, M.; FARHADI, P.; GHAFARI, R. **The effect of demand response on operation of smart home energy system with renewable energy resources**. Fundamentals of Electrical Engineering (ISFEE), 2016.

- GOMES, L. S. F. **A demanda por energia elétrica residencial no Brasil: 1999-2006 Uma estimativa das elasticidades-preço e renda por meio de painel**. Dissertação de Mestrado, Universidade de São Paulo, Campus de Ribeirão Preto, 2010.
- HAYKIN, S. **Neural Networks – A Comprehensive Foundation**, New York: Macmillan, 1994.
- HEARST, M.A. ; BERKELEY, CA ; DUMAIS, S.T. ; OSMAN, E. ; PLATT, J. . **Support vector machines**. IEEE Intelligent Systems, IEEE Computer Society, Los Alamitos, CA, USA, v. 13, n. 4, p. 18-28, 1998. ISSN 1094-7167.
- HRUSCHKA JR., E. R. **Imputação Bayesiana no contexto de Mineração de Dados**. Tese de Doutorado, Universidade Federal do Rio de Janeiro, 2003.
- IBGE. **Estimativa da População**, 2013.
- IBGE. **PNAD 2013 – Pesquisa Nacional por Amostra de Domicílios**, Rio de Janeiro, 2013.
- IBGE. **Síntese de Indicadores Sociais – Uma Análise de Vida da População**, 2012.
- IBGE. **Censo Demográfico 2010**, 2011.
- IEA. **World Energy Outlook**. International Energy Agency, 2016.
- JARDINI, J. A., TAHAN, C. M. V., GOUVEA, M. R., AHN, S. U., FIGUEIREDO, F. M. **Daily Load Profiles for Residential, Commercial and Industrial Low Voltage Consumers**. IEEE Transactions on Power Delivery, vol 15 no. 1, 2000.
- KARCHER, C. **Redes Bayesianas aplicadas à análise do crédito**. Dissertação de Mestrado, Escola Politécnica da Universidade de São Paulo, 2009.
- KAVEH A, TALATAHARI S. **A novel heuristic optimization method:charged system search**. Acta Mechanica, 213(3):267–289, 2010.
- KENNEDY J., EBERHART R.C. **Swarm Intelligence**. M. Kaufman, 2001.
- KORB, K. B.; NICHOLSON, A.E. **Bayesian artificial intelligence**. London: Chapman & Hall/CRC Press UK, 2004.
- KOZA, J. R. **Genetic programming: on the programming of computers by means of natural selection**. Cambridge, MA: The MIT Press, 1992.

- KURUPPU S. S. and KULATUNGA N. A., **Smart meter based non-intrusive appliance detection algorithm for local real time feedback purposes**, *IEEE PES Innovative Smart Grid Technologies*, Tianjin, 2012, pp. 1-5.
- LAMIN, H. **Análise de Impacto Regulatório da Implantação de Redes Inteligentes no Brasil**. 2013. 300 f.. (Doutorado). Departamento de Engenharia Elétrica, Universidade de Brasília - Faculdade de Tecnologia, Brasília/DF.
- MAHESWARAN D., RANGARAJ V., KAILAS K. K. J. AND KUMAR W. A., **Energy efficiency in electrical systems**, 2012 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), Bengaluru, 2012, pp. 1-6.
- MARQUES, M. S. **Metodologia para modelagem de curvas típicas de demanda elétrica utilizando redes neurais artificiais considerando variáveis climáticas**. Dissertação de Mestrado, Universidade Federal do Pampa, Campus Alegrete, 2014.
- NEAPOLITAN, R. E. **Learning Bayesian Networks**. Upper Saddle River: Pearson, 2004.
- OLIVEIRA, C. C. B.; FILHO, M. M.; SILVEIRA, C. S.; MORENO, R. G.; USHIMARA, E. A.; JUNIOR, P. F. B.; MAQUES, E. L. C.; FERREIRA, M. A. S.; GUIMARÃES, R. O. **SIACC - Sistema Inteligente de Análise de Curvas de Carga para Estudos de Previsão de Demanda e de Análise de Redes**. In: XIX Seminário Nacional de Distribuição de Energia Elétrica, 2010, São Paulo. Anais... 2010.
- PAPA, J. P., FALCÃO A. X., SUZUKI C. T. N. **Supervised pattern classification based on optimum-path forest**. *International Journal of Imaging Systems and Technology*, 19(2):120–131, 2009.
- PAPA, J. P. FALCÃO, A. X., FREITAS, G.M., AVILA, A. M. H. **Robust Pruning of Training Patterns for Optimun-Path Forest Classification Applied to Satellite-Based Rainfall Occurrence Estimation**. *IEEE Geoscience and Remote Sensing Letters*, 2010, pp 396-400, DOI:10.1109.
- PEDUZZI, P. **Demanda de Energia no Brasil Crescerá 78% até 2035**. Disponível em: <<http://planetasustentavel.abril.com.br/noticia/energia/demanda-energia-brasil-crescera-78-2035-648158.shtml>>. Acesso em: 12 de janeiro 2013
- PROCEL. **Avaliação do Mercado de Eficiência Energética no Brasil: Classe Residencial, Relatório Brasil**. Julho, 2007.

- RAMOS, C. C. O. **Caracterização de Perdas Comerciais em Sistemas de Energia Através de Técnicas Inteligentes**. Tese de Doutorado, Universidade de São Paulo USP, 2014.
- REVISTABW. **Matriz de Confusão**. Revista Brasileira de Web: Tecnologia. Disponível em <<http://www.revistabw.com.br/revistabw/matriz-de-confusao/>>. Acesso em: 25 de julho de 2016.
- ROSSETI, G. J. S. **Técnicas para restabelecimento de sistemas de distribuição de energia elétrica**. Tese de Doutorado, Univerdidade Federal de Juiz de Fora UFJF, 2015.
- SANTOS, K. M.dos; SOUZA, A. N. de; GASTALDELLO, D. S.; RAMOS, C. C. O. **Avaliação do uso de ferramentas inteligentes no estudo de padrões em curvas de carga**. CBQEE 2013 - X Conferência Brasileira sobre Qualidade da Energia Elétrica, 2013. Araxá, MG, Brasil
- SANTOS, K. M.dos; SOUZA, A. N. de; GASTALDELLO, D. S. **Desenvolvimento de um algoritmo para determinar curvas de carga a partir da posse e dos hábitos de consumo para consumidores: residenciais, comerciais e industriais**. SBSE 2012 - IV Simpósio Brasileiro de Sistemas Elétricos, 2012. Goiânia, GO, Brasil.
- SANTOS, K. M.dos; SOUZA, A. N. de; GASTALDELLO, D. S.; COSTA JR, P. **Uma ferramenta computacional aplicada no estudo de curvas de carga em sistemas de energia**. Dincon 2011 - 10ª Conferência Brasileira de Dinâmica, Controle e Aplicações, 2011. Águas de Lindóia, SP, Brasil.
- STEIMER P. K., **Enabled by high power electronics - Energy efficiency, renewables and smart grids**, Power Electronics Conference (IPEC), 2010 International, Sapporo, 2010, pp. 11-15.
- TEMNEANU M., **Appliance characterization based on spectral components analysis**, *Electrical and Power Engineering (EPE), 2014 International Conference and Exposition on*, Iasi, 2014, pp. 710-714.
- ULLAH M. N., JAVAID N., KHAN I., MAHMOOD A., FAROOQ M. U. **Residential Energy Consumption Controlling Techniques to Enable Autonomous Demand Side Management in Future Smart Grid Communications**. Eighth International Conference on Broadband, Wireless Computing, Communication and Applications, 2013.

- VENTURA FILHO, A. **Energia Elétrica no Brasil: Contexto Atual e Perspectivas**. Revista Interesse Nacional, Ano 6 número 21, jun 2013.
- VIEIRA FILHO, V.; ALBUQUERQUE, M. T. C. F. **Abordagem Bayesiana para Simulação de Jogos Complexos**. In: SBGames, 2007.
- VILLELA, A., G.; CARVALHO, V. A. **Cálculo de Demanda Rede para Acréscimo de Novas Cargas na Rede de Distribuição de Baixa Tensão**. Nota Técnica, NT-2.018 da Eletropaulo, novembro de 2007.
- YANG X.-S. **Firefly algorithm, stochastic test functions and design optimisation**. International Journal Bio-Inspired Computing, 2(2):78–84, 2010.
- YANG X.-S. **Bat algorithm for multi-objective optimisation**. International Journal of Bio-Inspired Computation, 3(5):267–274, 2011.

APÊNDICE A – Link para Tabelas dos Resultados da Análise Estatística

As tabelas encontradas no link a seguir apresentam a distribuição das curvas no estudo estatístico de cada uma das quatro variáveis estudadas e apresentadas no capítulo 4:

<https://www.dropbox.com/s/2myfq47xhg1f48q/AP%C3%80NDICE%20A%20-%20Desenvolvimento%20de%20ferramentas%20computacionais%20evolutivas%20para%20estudos%20de%20curvas%20de%20carga%20e%20planejamento%20do%20sistema%20de%20distribui%C3%A7%C3%A3o.pdf?dl=0>

ANEXO A – Link para NT 2.018

A Nota Técnica NT 2.018 - Cálculo de Demanda Rede para Acréscimo de Novas Cargas na Rede de Distribuição de Baixa Tensão, pode ser visualizada no link a seguir:

<https://www.dropbox.com/s/f51i9vx6b4tdrnl/NT-2018%20-%20eletropaulo.pdf?dl=0>