

**unesp**



**UNIVERSIDADE ESTADUAL PAULISTA**  
**“JÚLIO DE MESQUITA FILHO”**  
Instituto de Ciência e Tecnologia de Sorocaba  
Campus de São João da Boa Vista

**Thaís Berrettini Bosco**

**Estudo da Engenharia de Atributos e Qualidade de  
Dados para Desagregação de Cargas**

Sorocaba  
2023

Thaís Berrettini Bosco

**Estudo da Engenharia de Atributos e Qualidade de  
Dados para Desagregação de Cargas**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica, junto ao Programa de Pós-Graduação em Engenharia Elétrica, Interunidades, entre o Instituto de Ciência e Tecnologia de Sorocaba e o Campus de São João da Boa Vista da Universidade Estadual Paulista “Júlio de Mesquita Filho”.

Orientador: Prof. Dr. Flavio Alessandro Serrão  
Gonçalves

Coorientador: Prof. Dr. Wesley Angelino de Souza

Sorocaba  
2023

B742e

Bosco, Thaís Berrettini

Estudo da engenharia de atributos e qualidade de dados para desagregação de cargas / Thaís Berrettini Bosco. -- Sorocaba, 2023

139 p.

Dissertação (mestrado profissional) - Universidade Estadual Paulista (Unesp), Instituto de Ciência e Tecnologia, Sorocaba

Orientador: Flavio Alessandro Serrão Gonçalves

Coorientador: Wesley Angelino de Souza

1. Algoritmos. 2. Inteligência artificial. 3. Aprendizado de máquinas. 4. Circuitos elétricos. 5. Banco de dados. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Ciência e Tecnologia, Sorocaba. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.



**UNIVERSIDADE ESTADUAL PAULISTA**

**Câmpus de Sorocaba**

**CERTIFICADO DE APROVAÇÃO**

**TÍTULO DA DISSERTAÇÃO:** Estudo da Engenharia de Atributos e Qualidade de Dados para Desagregação de Cargas

**AUTORA:** THAÍS BERRETTINI BOSCO

**ORIENTADOR:** FLÁVIO ALESSANDRO SERRÃO GONÇALVES

**COORIENTADOR:** WESLEY ANGELINO DE SOUZA

Aprovada como parte das exigências para obtenção do Título de Mestre em Engenharia Elétrica, área: Automação pela Comissão Examinadora:

Prof. Dr. FLÁVIO ALESSANDRO SERRÃO GONÇALVES (Participação Virtual)  
Departamento de Engenharia de Controle e Automação / Instituto de Ciência e Tecnologia UNESP - Campus de Sorocaba

Prof. Dr. MOACYR AURELIANO GOMES DE BRITO (Participação Virtual)  
Faculdade de Engenharias, Arquitetura e Urbanismo e Geografia / Universidade Federal de Mato Grosso do Sul

Prof. Dr. LUIS ARMANDO DE ORO ARENAS (Participação Presencial)  
Departamento de Engenharia Ambiental / Universidade Estadual Paulista - UNESP - Campus de Sorocaba - ICTS

Sorocaba, 15 de dezembro de 2022

## RESUMO

Devido à crescente necessidade de utilização otimizada de energia e caracterização detalhada dos perfis de consumo de eletricidade, o conceito de desagregação de cargas está ganhando atenção dentro da perspectiva de rede inteligente de energia (REI). Buscando melhorias relacionadas à conscientização do consumo de energia elétrica, este trabalho tem o intuito de contribuir com o estado do conhecimento do consumo individualizado por eletrodoméstico, apresentando estudos de aplicações e uma análise crítica sobre o uso de atributos de desagregação de cargas baseados em teorias de potência. A avaliação das estratégias propostas é efetuada considerando conjuntos de bancos de dados de referências da literatura (PLAID), que descreve comportamentos de cargas elétricas reais em funcionamento, e utilizando-se do software PSim, os mesmos são simulados, e a partir de índices de potência e de corrente calculados com base em teorias de potência, compõe-se o banco de dados teórico. As teorias de potência aplicadas são CPC, CPT e IEEE 1459-2010, que com seus conceitos modernos apresentados no decorrer do trabalho, auxiliam no estabelecimento dos atributos relevantes que são utilizados no processo de desagregação de cargas. Diferentes algoritmos de aprendizado de máquinas (k-ésimo vizinho mais próximo, Máquina de vetor de suporte e Floresta Aleatória) são aplicados nos processos de classificação e suas performances analisadas, tendo resultados com precisões maiores que 99 %. Uma otimização no banco de dados é proposta por meio dos estudos de correlação existente entre os atributos que constituem o banco de dados, possibilitando reduzir de 55 para 7 atributos mantendo a precisão de classificação de cargas em valores maiores que 99 %. Desta forma, além da contribuição relacionada à validação da criação de um banco de dados teórico a ser utilizado no conceito de desagregação de cargas elétricas, destaca-se que a otimização do mesmo traz o benefício de redução na demanda de processamento, tanto na questão dos cálculos de índices de potência e corrente, quanto na etapa de classificação das cargas.

**Palavras-chave:** algoritmo, aprendizado de máquinas, banco de dados, cargas elétricas, corrente, desagregação de cargas, inteligência artificial, potência.

## ABSTRACT

Due to the growing need for optimized energy utilization and detailed characterization of electricity consumption profiles, the concept of load unbundling is gaining attention within the smart energy grid (SG) perspective. Seeking improvements related to the awareness of electrical energy consumption, this work aims to contribute to the state of knowledge of individualized consumption per appliance, presenting application studies and critical analysis on the use of load breakdown attributes based on theories of potency. The evaluation of the proposed strategies is carried out considering sets of databases of literature references (PLAID), which describe behaviors of real electrical loads in operation and using the PSIM software they were simulated, and from indices of power and current calculated based on power theories, the theoretical database is composed. The applied power theories are CPC, CPT, and IEEE 1459-2010 with their modern concepts presented in the future in the course of the work, they help in the establishment of the relevant attributes that will be used in the load disaggregation process. Different machine learning algorithms (k-NN, SVM, and RF) are applied in the classification processes and their performances are analyzed, with results achieved with precision greater than 99 %. An optimization in the database is proposed through the existing correlation studies between the attributes that make up the database, which makes it possible to reduce from 55 to 7 attributes maintaining the load classification accuracy at values greater than 99%. Thus, in addition to the contribution related to the validation of the creation of a theoretical database to be used in the concept of electrical load disaggregation, it is highlighted that its optimization brings the benefit of reducing the processing demand, both in terms of calculations of power and current indices, as well as in the load classification step.

**Keywords:** algorithm, artificial intelligence, current, database, electrical loads, load disaggregation, machine learning, power.

## LISTA DE ILUSTRAÇÕES

Figura 1-Esquema geral de um sistema NILM	19
Figura 2-Esquema geral de funcionamento um sistema NILM	20
Figura 3- Medidor de Energia Elétrica e o Sistema NILM	25
Figura 4 - Diagrama da evolução das Teorias de Potência	27
Figura 5- Diagrama de decomposição de correntes segundo a CPC	32
Figura 6 - Diagrama de Potências segundo a IEEE 1459-2010	41
Figura 7 - Conjuntos Separáveis Linearmente	46
Figura 8 - Conjunto de dados que não são Linearmente Separáveis	48
Figura 9 - Separação não linear (Kernel Radial)	49
Figura 10 - Exemplo de conjunto a ser classificado por algoritmo k-enésimo vizinho mais próximo.	51
Figura 11 - Exemplo de Árvore de Decisão	53
Figura 12 - Componentes da Árvore de Decisão	54
Figura 13 - Exemplo de conjunto que será classificado pelo algoritmo RF	56
Figura 14 - Fluxograma do desenvolvimento do trabalho	59
Figura 15 -Etapa 1: Aquisição de dados	60
Figura 16 - Estrutura de simulação para Aquisição de Dados	62
Figura 17 -Etapa 2: Otimização do Banco de Dados	65
Figura 18 - Fluxograma de Simulações Cenário Geral	66
Figura 19 - Fluxograma de Simulações Cenário Individualizado	66
Figura 20 - Classificação das Cargas elétricas	68
Figura 21 - Oscilografias - Simulação 1: lâmpada fluorescente compacta	71
Figura 22 - Oscilografias - Simulação 2: Geladeira	72
Figura 23 - Oscilografias - Simulação 3: Secador de cabelos	73
Figura 24 - Oscilografias - Simulação 4: Micro-ondas	74
Figura 25 - Esquemático elétrico de um forno micro-ondas Electrolux ME28G / ME28X74	
Figura 26 - Oscilografias - Simulação 5: Ar-condicionado	75
Figura 27 - Oscilografias - Simulação 6: Notebook	76
Figura 28 - Oscilografias - Simulação 7: Lâmpada incandescente	77
Figura 29 - Oscilografias - Simulação 8: Ventilador	78
Figura 30 - Oscilografias - Simulação 9: Máquina de lavar roupas	78
Figura 31 - Oscilografias - Simulação 10: Aquecedor	79
Figura 32 - Fluxograma da otimização do banco de dados	87
Figura 33 - Matriz de correlação dos atributos calculados	89
Figura 34 - Matriz de correlação dos atributos calculados CPT	116
Figura 35 - Matriz de correlação dos atributos calculados CPC	116
Figura 36 - Matriz de correlação dos atributos calculados IEEE	117

## LISTA DE TABELAS

Tabela 1 - Número de equipamentos e amostras coletadas no PLAID	21
Tabela 2- Índices da CPT, como são calculados e seus significados físicos	35
Tabela 2- Índices da CPT, como são calculados e seus significados físicos (continuação)	36
Tabela 3-Índices de potência para sistemas monofásicos segundo IEEE 1459-2010	39
Tabela 4-Índices de potência para sistemas trifásicos segundo IEEE 1459-2010	40
Tabela 5 - Indicadores de conteúdo harmônico segundo IEEE 1459-2010	40
Tabela 6 - Vantagens e desvantagens do algoritmo SVM	47
Tabela 7 - Vantagens e desvantagens do algoritmo k-NN	52
Tabela 8 - Vantagens e desvantagens do algoritmo Árvore de Decisão	55
Tabela 9 - Matriz de confusão	57
Tabela 10 - Fórmulas matemáticas para cálculo dos indicadores de avaliação do algoritmo	58
Tabela 11 - Simulações realizadas	61
Tabela 12 -Índices Calculados, Levantamento dos Atributos	63
Tabela 13 -Simulações realizadas	68
Tabela 14 - Média e desvio padrão (Carga 1).	80
Tabela 15 - Média e desvio padrão (Carga 6).	81
Tabela 16 - Média e desvio padrão (Carga 7).	82
Tabela 17 - Média e desvio padrão (Carga 10).	82
Tabela 18 - Média e desvio padrão (Carga 15).	83
Tabela 19 - Média e desvio padrão (Carga 28).	84
Tabela 20 - Média e desvio padrão (Carga 57).	84
Tabela 21 - Média e desvio padrão (Carga 66).	85
Tabela 22 - Média e desvio padrão (Carga 80).	86
Tabela 23 - Média e desvio padrão (Carga 136).	86
Tabela 24 - Correlação observada a partir da Matriz de correlação e otimizações propostas	90
Tabela 24 - Correlação observada a partir da Matriz de correlação e otimizações propostas (continuação)	91
Tabela 24 - Correlação observada a partir da Matriz de correlação e otimizações propostas (continuação)	92
Tabela 24 - Correlação observada a partir da Matriz de correlação e otimizações propostas (continuação)	93
Tabela 25 - Tabela de atributos definidos à partir das correlações observadas	94
Tabela 26 - Simulações e banco de dados utilizado	95
Tabela 27 - Performance dos algoritmos - Banco de dados inicial - Simulação 1	96
Tabela 28 - Performance do k-NN para cada classe - Simulação 1	96
Tabela 29 - Performance do SVM Linear para cada classe - Simulação 1	97
Tabela 30 - Performance do SVM Radial para cada classe - Simulação 1	97



Tabela 30 - Performance do SVM Radial para cada classe - Simulação 1 (continuação)	98
Tabela 31 - Performance do DT para cada classe - Simulação 1	98
Tabela 32 - Performance do RF para cada classe - Simulação 1	99
Tabela 33 - Performance dos algoritmos-Banco de dados otimização 1 - Simulação 2	100
Tabela 34 - Performance do k-NN para cada classe - Simulação 2	100
Tabela 35 - Performance do SVM Linear para cada classe - Simulação 2	101
Tabela 36 - Performance do SVM Radial para cada classe - Simulação 2	101
Tabela 37 - Performance do DT para cada classe - Simulação 2	102
Tabela 38 - Performance do RF para cada classe - Simulação 2	102
Tabela 38 - Performance do RF para cada classe - Simulação 2 (continuação)	103
Tabela 39 - Performance dos algoritmos-Banco de dados otimização 2 - Simulação 3	103
Tabela 40 - Performance do k-NN para cada classe - Simulação 3	104
Tabela 41 - Performance do SVM Linear para cada classe - Simulação 3	105
Tabela 42 - Performance do SVM Radial para cada classe - Simulação 3	105
Tabela 43 - Performance do DT para cada classe - Simulação 3	106
Tabela 44 - Performance do RF para cada classe - Simulação 3	106
Tabela 44 - Performance do RF para cada classe - Simulação 3 (continuação)	107
Tabela 45 - Performance dos algoritmos-Banco de dados otimização 3 - Simulação 4	107
Tabela 46 - Performance do k-NN para cada classe - Simulação 4	108
Tabela 47 - Performance do SVM Linear para cada classe - Simulação 4	108
Tabela 47 - Performance do SVM Linear para cada classe - Simulação 4 (continuação)	109
Tabela 48 - Performance do SVM Radial para cada classe - Simulação 4	109
Tabela 49 - Performance do DT para cada classe - Simulação 4	110
Tabela 50 - Performance de classificação do RF para cada classe - Simulação 4	110
Tabela 51 - Performance dos algoritmos-Banco de dados otimização 4 - Simulação 5	111
Tabela 52 - Performance do k-NN para cada classe - Simulação 5	111
Tabela 53 - Performance do SVM Linear para cada classe - Simulação 5	112
Tabela 54 - Performance do SVM Radial para cada classe - Simulação 5	113
Tabela 55 - Performance do DT para cada classe - Simulação 5	113
Tabela 56 - Performance do RF para cada classe - Simulação 5	114
Tabela 57 - Banco de dados individualizado	115
Tabela 58 - Correlação observada a partir da Matriz de correlação e otimizações propostas	1187
Tabela 58 - Correlação observada a partir da Matriz de correlação e otimizações propostas (continuação)	118
Tabela 58 - Correlação observada a partir da Matriz de correlação e otimizações propostas (continuação)	119
Tabela 59 - Tabela de atributos definidos a partir das correlações observadas	119
Tabela 59 - Tabela de atributos definidos a partir das correlações observadas (continuação)	120

Tabela 60 - Simulações, banco de dados e parâmetros utilizados	121
Tabela 61 - Performance dos algoritmos - Simulação 2	122
Tabela 62 - Performance dos algoritmos - Simulação 3	122
Tabela 63 - Performance dos algoritmos - Simulação 4	123
Tabela 64 - Simulações e banco de dados utilizado	124
Tabela 65 - Comparação - Simulação 5 - Cenário geral e individual	125
Tabela 66 - Avaliação Tempo x Precisão com a Otimização do Banco de dados	1266

## LISTA DE ABREVIATURAS E SIGLAS

<b>CPC</b>	Currents Physical Components
<b>CPT</b>	Conservative Power Theory
<b>DT</b>	Decision Tree
<b>IEEE</b>	Institute of Electrical and Electronics Engineers
<b>k-NN</b>	k-Nearest Neighbor
<b>SVM</b>	Support Vector Machines
<b>RF</b>	Random Forest
<b>QEE</b>	Qualidade de Energia Elétrica
<b>NILM</b>	Non-Intrusive Load Monitoring
<b>RFE</b>	Recursive Feature Elimination
<b>USP</b>	Universidade de São Paulo
<b>PEE</b>	Programa de Eficiência Energética
<b>IEE</b>	Instituto de Energia e Ambiente
<b>Proben</b>	Programa de Bom Uso Energético

## SUMÁRIO

<b>CAPÍTULO 1 - INTRODUÇÃO</b> .....	<b>13</b>
1.1. Contextualização e Motivação.....	15
1.2. Objetivos .....	16
1.3. Organização da Dissertação.....	17
<b>CAPÍTULO 2 – FUNDAMENTAÇÃO CIENTÍFICA</b> .....	<b>18</b>
2.1. Sistema NILM .....	18
2.2. Medidor inteligente .....	24
2.3. Teorias de Potência .....	26
2.4. Inteligência artificial e aprendizado de máquinas.....	43
<b>CAPÍTULO 3 - METODOLOGIA</b> .....	<b>59</b>
3.1. Concepção Do Trabalho Desenvolvido .....	59
<b>CAPÍTULO 4 - RESULTADOS</b> .....	<b>70</b>
4.1. Cargas Elétricas - Oscilografias .....	70
4.2. Construção Do Banco De Dados E Validação Dos Resultados.....	79
4.3. Otimização Do Banco De Dados - Definição Dos Atributos Para O Banco De Dados Final - Cenário Geral.....	87
4.4. Otimização Do Banco De Dados - Definição Dos Atributos Para O Banco De Dados Final - Cenário Individual .....	114
4.5. Comparação Entre Cenários Geral e Individual .....	124
<b>CAPÍTULO 5 - CONCLUSÕES</b> .....	<b>127</b>
<b>REFERÊNCIAS</b> .....	<b>129</b>

## CAPÍTULO 1 - INTRODUÇÃO

Com a expansão tecnológica é notável o aumento de dispositivos eletrônicos de potência, microprocessadores, controladores, luminárias, entre outras cargas que inserem distúrbios elétricos em cenários residenciais, comerciais e industriais e de distribuição. As não linearidades, quando inseridas nas instalações elétricas podem gerar alto conteúdo harmônico, acarretando diversos problemas relacionados à qualidade da energia elétrica (QEE) [1 - 2].

Neste cenário, impactos manifestados por meio de oscilações de tensão, corrente ou frequência podem ser observados na forma de incômodos visuais relacionados com a variação luminosa, interferências em equipamentos eletrônicos, aumento de perdas elétricas, inserção de harmônicos, entre outros, que podem resultar em mau funcionamento do equipamento, sendo considerado um problema de QEE [1].

As redes elétricas podem minimizar os impactos ocasionados pelas condições de baixa qualidade de energia elétrica com melhorias nas técnicas de análise, medição e controle aplicadas. Conforme [3] algumas das principais dificuldades enfrentadas nas redes elétricas são:

- Incapacidade de monitorar e controlar a rede em tempo real;
- Incapacidade de fornecer ao consumidor e à rede dados de consumo confiáveis em tempo real;
- Não há gerenciamento de demanda e resposta;
- Não há conectividade suficiente entre as várias áreas da rede para balancear as cargas.

Diante deste cenário, observa-se que a medição de energia se torna um fator primordial, uma vez que compreende a base das transações comerciais de energia elétrica. Neste sentido, entende-se que a caracterização dos medidores de energia elétrica e a instrumentação utilizada para medição assumem um papel crucial para quantificar e qualificar a energia elétrica fornecida e consumida [4].

A potência elétrica é a grandeza mais comumente utilizada para a verificação da eficiência da transmissão de energia elétrica e a classificação de equipamentos. Em casos em que as tensões de fornecimento são senoidais puras em uma frequência

fundamental, esta aplicação tem forte confiabilidade, uma vez que os cálculos para as análises estão estabelecidos na literatura, sendo facilmente realizados. Entretanto, atualmente (2022), as cargas não lineares utilizadas nas redes elétricas impactam em distorções também nas potências elétricas. Para estes casos, com a finalidade de auxiliar na quantificação e qualificação da energia elétrica, em especial para situações não senoidais, algumas teorias de potência foram desenvolvidas e continuam em constante análise e evolução. Neste contexto, o presente trabalho aborda os principais conceitos e definições das teorias: Currents Physical Components [5], Conservative Power Theory [6-7] e IEEE 1459-2010 [2 - 8 - 9].

Os fatores de distorção harmônica total de tensão e corrente são alguns dos indicadores mais conhecidos para a avaliação da distorção harmônica e abordados nas teorias de potência, pois são capazes de quantificar a distorção. Apesar disto, outros índices de potência são propostos e analisados nas teorias de potência modernas mencionadas, uma vez que os fatores de distorção, por si só, não identificam a fonte causadora de tal distúrbio [10].

Assim, a necessidade de gerenciar a energia elétrica é emergente nas denominadas redes inteligentes (*Smart Grids*) em todo o mundo. A *smart grid* combina várias tecnologias de comunicação, medidores inteligentes e aparelhos de controle com o objetivo de tornar a rede elétrica inteligente e bem gerenciada, trabalhando em conjunto para otimizar a produção e o consumo de energia. Este arranjo permite tanto a gestão quanto o controle da rede na forma bidirecional e em tempo real [3].

Medidores de energia elétrica inteligentes são partes importantes deste sistema, com a função de medir e calcular os parâmetros de QEE. Neste conceito de *smart grid* é inerente que o medidor tenha a capacidade de classificar cargas para possibilitar a tomada de decisão tanto por parte do consumidor como da concessionária [11].

Neste contexto, com base em estudos existentes na literatura e utilizados como referências no desenvolvimento, conforme [12-15], o presente trabalho busca contribuir com o estado do conhecimento propondo uma otimização do banco de dados e a classificação das cargas analisada pelo algoritmo *Random Forest* ou Árvores Aleatórias.

Além disso, considerando o emprego de três teorias de potência diferentes (CPC, CPT e IEE) para o levantamento dos índices de potência e elaboração do banco de dados,

visando a otimização dos processos, pode-se verificar por meio da correlação entre os índices a real necessidade do cálculo dos mesmos e possivelmente desconsiderá-los como relevantes na composição do banco de dados.

### **1.1. Contextualização e Motivação**

O trabalho possui como motivação alguns indicadores brasileiros relacionados à conscientização do consumo de energia elétrica. Nesta conjuntura, algumas matérias publicadas no site da Procel reportam que a possibilidade de monitorar em tempo real o consumo de energia por tipo de equipamento, é um caminho para criar uma maior consciência sobre os gastos podendo gerar economia [16].

Além disso, percebe-se o expressivo incentivo a projetos voltados para eficiência energética, como por exemplo, os projetos entregues pela Enel SP a três instituições da Universidade de São Paulo (USP): o Hospital Universitário, a Escola Politécnica e o Instituto de Energia e Ambiente. Com investimento de R\$ 1,97 milhão, a concessionária instalou 522 módulos fotovoltaicos e modernizou os sistemas de iluminação por meio da substituição de 13,3 mil lâmpadas fluorescentes por modelos à LED, que são mais eficientes e têm maior durabilidade. Esta é uma iniciativa, financiada com os recursos do Programa de Eficiência Energética (PEE) regulado pela Aneel, em prédios universitários que terão uma redução total de 1.411 MWh/ano no consumo de energia, o equivalente ao consumo de energia de 588 residências por mês. A estimativa é que as obras tragam uma economia estimada de R\$ 808,8 mil por ano na conta de luz da USP. Na obra do Instituto de Energia e Ambiente (IEE/USP), a Enel SP investiu R\$ 235,4 mil na troca de 2.463 lâmpadas por modelos LED, o que reduzirá o consumo de energia em 302 MWh/ano [17].

Outro investimento recente foi feito pela Universidade Federal de Pelotas em ações de eficiência energética suportadas pelo Programa de Bom Uso Energético (Proben), que resultaram em uma economia de energia elétrica de R\$ 758.350,39 no ano de 2020. O programa desenvolvido se baseia em práticas de bom uso da energia dentro da universidade, fornecendo maior conscientização aos usuários. Outra iniciativa é a correção do fator de potência das unidades consumidoras [18].

Pelos índices apontados como exemplo, percebe-se que no Brasil, a conscientização do consumo de energia elétrica pode impactar substancialmente na redução do desperdício de energia. Assim sendo, possibilitar o desenvolvimento de ferramentas que contribuam para proporcionar conhecimento necessário ao usuário é de extrema importância. Ao informar o consumidor de possíveis pontos de melhoria em seu próprio consumo de energia, a informação auxilia o mesmo na tomada de decisão.

Dentro dos impactos e das dificuldades expostas, o trabalho contribui por prever funcionalidades adicionais ao medidor de energia elétrica inteligente, como ferramentas que possibilitem uma medição e classificação de cargas elétricas precisas, visando fornecer tanto ao consumidor quanto à concessionária relatórios completos, conscientizando quanto à QEE consumida e distribuída.

Compreender a contribuição deste estudo requer conhecer o padrão de sistema elétrico atual, bem como o instrumento de medição de energia elétrica utilizado. A confiabilidade do banco de dados teórico frente aos estudos existentes na literatura, está diretamente ligada à precisão de classificação das cargas elétricas desagregadas feita por meio de algoritmos de inteligência artificial.

Diante do exposto, é evidente a necessidade de quantificar os impactos gerados pela utilização dos equipamentos elétricos e eletrônicos em uma rede elétrica. Detectar a fonte causadora de um consumo excessivo, por exemplo, é um fator não observado atualmente, que possibilitará tanto ao consumidor quanto à concessionária, agir em favor de melhorias no consumo e na geração de energia elétrica.

## **1.2. Objetivos**

O objetivo geral do trabalho consiste em contribuir com o conhecimento envolvido na natureza de operação individualizada por equipamento, em especial considerando eletrodomésticos monofásicos, apresentando estudos de aplicações e uma análise crítica sobre o uso de atributos de desagregação de cargas baseados em três teorias de potência modernas (CPC, CPT e IEEE 1459-2010), por meio de algoritmos de aprendizado, para a geração de um banco de dados teórico e otimizado para classificação de cargas.



Neste contexto, os objetivos específicos envolvem estabelecer os atributos relevantes relacionados com três diferentes teorias de potências modernas (CPC, CPT e IEEE 1459-2010) para utilização no processo de desagregação de cargas, e ainda, a aplicação e a avaliação de desempenho de diferentes algoritmos de aprendizado de máquinas (k-ésimo vizinho mais próximo, Máquina de vetor de suporte e Floresta aleatória) nos processos de classificação.

### **1.3. Organização da Dissertação**

Partindo deste capítulo introdutório, em que são apresentadas a contextualização das motivações, a relevância da temática e os objetivos propostos, o Capítulo 2 apresenta a revisão bibliográfica das teorias para desenvolvimento do trabalho, detalhando o sistema, os equipamentos necessários, bem como as teorias de potências empregadas para desagregação de cargas elétricas. Os Capítulos 3 e 4 apresentam a metodologia e os resultados, respectivamente. E por fim, o Capítulo 5 apresenta as conclusões finais.

## CAPÍTULO 2 – FUNDAMENTAÇÃO CIENTÍFICA

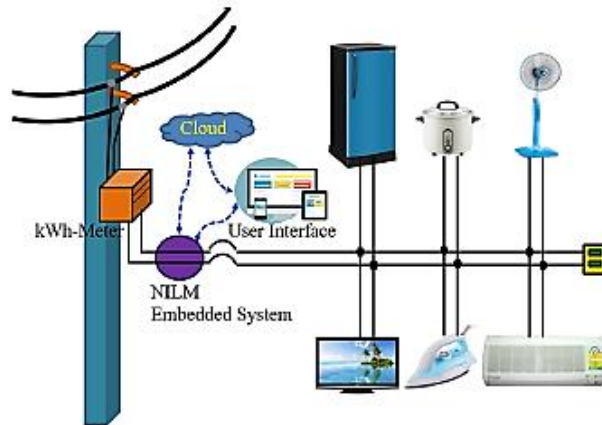
Este capítulo apresenta uma revisão bibliográfica com conceitos detalhados utilizados para entendimento da evolução do trabalho. Neste sentido, os conceitos de NILM, medidor de energia elétrica inteligente, teorias de potência, inteligência artificial são apresentados para compor o escopo técnico envolvido.

### 2.1. Sistema NILM

Nos sistemas de gerenciamento de energia doméstica existe a medição dos dados de consumo de energia em tempo real, com o objetivo de monitorar e otimizar o seu uso. Sistemas tradicionais de monitoramento intrusivos captam dados de eletricidade através de medidores de energia elétrica conectados ao ponto de acoplamento comum entre os equipamentos e a rede elétrica. Como o próprio nome sugere, por serem intrusivos, se desejado individualizar a medição do consumo de energia elétrica por equipamento, são necessários sensores de corrente e tensão destinados à cada um dos equipamentos existentes, ou seja, não é possível prever a medição individual somente pelo medidor de energia elétrica [19].

Os sistemas de monitoramento de carga não intrusivo (NILM, *Non-Intrusive Load Monitoring*) contribuem para a redução do *hardware* e dos custos de manutenção, uma vez que é necessário apenas um medidor de energia elétrica inteligente instalado ao ponto de acoplamento comum do quadro de distribuição de energia elétrica da residência. O sistema NILM permite a desagregação do consumo das cargas elétricas, como observado na Figura 1 [19 - 20].

**Figura 1-Esquema geral de um sistema NILM**



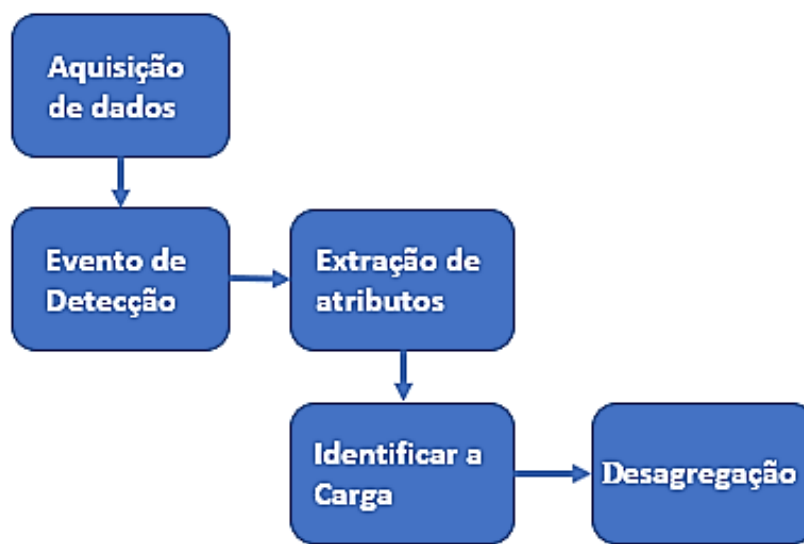
FONTE: [20]

Destaca-se como um dos primeiros estudos sobre o sistema NILM, o estudo proposto, em 1992, por Hart [21], em que, existe a classificação de aparelhos diferentes aplicando algoritmo de agrupamento nos recursos de energia ativa-reativa. Embora o método NILM tenha sido inicialmente usado apenas como um medidor de energia total para casas ou edifícios, foi visto que este método pode ajudar os usuários a economizar mais espaço físico (necessário apenas um medidor) e reduzir os impactos ao meio ambiente [14 - 19 - 20].

Desde sua concepção, o ramo de pesquisa em desagregação de carga tem sido impulsionado e discutido com novas abordagens [14 - 22]. A desagregação permite melhorar o desempenho de outras funcionalidades importantes no sistema de gestão de energia, como por exemplo, na previsão de demanda de carga. Ao gerir a energia, destaca-se como benefícios a contribuição na economia da mesma uma vez que o consumidor terá o *feedback* de seu consumo em tempo real [19 - 20].

A Figura 2 representa o esquema geral de funcionamento de um sistema NILM.

**Figura 2-Esquema geral de funcionamento um sistema NILM**



FONTE: [63]

A Figura 2 mostra o diagrama de blocos da sequência de ações características esperadas em um monitoramento de carga não intrusiva. Em “Aquisição de dados” a tensão e a corrente são coletadas no ponto de acoplamento comum, fato que o torna não intrusivo, sendo posteriormente, processadas para a contabilização das potências. Este processamento pode ser efetuado diretamente nos medidores de energia elétrica. O bloco “Evento de Detecção” é responsável por investigar variações na potência, notando se há anormalidades em comparação ao padrão da rede, por exemplo, tensões acima de 127 V para sistemas 110 V. Após, os sinais são conduzidos para o bloco “Extração de atributos” que permitirá “Identificar a Carga” elétrica medida e esta, por sua vez, será “Desagregada”. A extração de atributos se refere ao cálculo dos índices de potência e corrente. Os índices são atributos que compõe um banco de dados, e permite identificar as cargas elétricas. Uma vez identificadas, é possível desagregá-las. No estudo presente, as etapas existentes à partir do bloco “Extração de atributos” serão desenvolvidas empregando três teorias de potência modernas (CPC, CPT e IEEE 1459-2020) [2 - 5 - 9 - 14].

A “Aquisição de dados” pode ser feita de diversas formas, como por exemplo, utilizando sensores de corrente e tensão alternada. A coleta de amostras pode ser feita

em alta ou baixa frequência. Dados em alta frequência são muito utilizados para análises de consumo e possuem alguns bancos de dados já desenvolvidos na literatura, como por exemplo o PLAID (*Plug-Level Appliance Identification Dataset*) [14].

*PLAID* é um conjunto de dados público e de origem coletiva para pesquisa de identificação de cargas que fornece valores de tensão e corrente amostradas a 30 kHz de 11 diferentes tipos de aparelhos elétricos presentes em 56 famílias em Pittsburgh, Pensilvânia, EUA, coletados em 2013. O banco de dados contempla amostras tanto para a operação dos dispositivos em estado estacionário, quanto para o transiente, totalizando 1094 instâncias. A Tabela 1 apresenta o resumo do número de dispositivos e instâncias coletadas no PLAID [15].

**Tabela 1 - Número de equipamentos e amostras coletadas no PLAID**

<b>Tipo de Equipamento</b>	<b>Número de Equipamentos</b>	<b>Número de Instâncias</b>
Ar-condicionado	19	92
Lâmpada Fluorescente Compacta	35	173
Geladeira	18	46
Secador de Cabelos	31	156
Notebook	38	163
Micro-ondas	23	135
Máquina de Lavar	7	26
Lâmpada Incandescente	25	117
Aspirador de pó	7	35
Ventilador	23	114
Aquecedor	23	114
<b>Total</b>	<b>235</b>	<b>1094</b>

A etapa de “Detecção de Eventos” é uma das mais importantes para o funcionamento do sistema NILM pois é neste momento que, por exemplo, identificam-se os tempos de comutação de cada dispositivo conectado à rede elétrica. A detecção pode ser feita de 4 formas:

- Modelos Heurísticos, implementado por Hart, realiza-se a segmentação do estado transiente estável e um limite pré-determinado é utilizado [14];

- Métodos Probabilísticos, determina-se quando o consumo total de energia é alterado [14];
- Modelos de Correspondência compara a potência total com sinais de energia de algum modelo [14];
- Modelos híbridos utilizando mais de um dos três métodos mencionados para detecção de um evento [14];

Na “Extração de Atributos” obtém-se informações dos sinais da tensão e da corrente. Estas informações são responsáveis por representar a assinatura da carga elétrica utilizada. Na literatura, encontram-se diversos métodos utilizados para esta etapa. Hart sugeriu que fossem desenvolvidos estudos com base na combinação de potências ativa e reativa [14][21]. No estudo de Y. Su, K. Lian e H. Chang, a extração de recursos é realizada usando a Transformada de Fourier e a transformada *wavelet* [14][23]. Características baseadas na trajetória V-I são propostas por Lam et al. para descrever as características dos dispositivos [14][24].

No presente trabalho, esta etapa é realizada por meio dos índices de potência e corrente que serão calculados com base nas Teorias de Potência CPC, CPT e IEEE 1459-2010. Optou-se por este direcionamento, uma vez que a representação de fenômenos elétricos ainda não pode ser compreendida e mensurada por completo, sem incertezas, por meio dos métodos clássicos evidenciados na literatura. Oscilografias são utilizadas para verificação dos comportamentos das cargas e, a partir dos sinais de tensão e corrente observados, os índices de potência e corrente podem ser calculados compondo os atributos que representarão cada dispositivo elétrico do sistema NILM em questão.

Para a seleção dos atributos que serão utilizados no sistema, existem: métodos de filtro, métodos de *wrapper* e métodos incorporados.

Os métodos de filtros observam apenas as propriedades intrínsecas dos dados, calculando a relevância de cada recurso, pré-selecionando aqueles que possuem maior relevância. Possuem baixo esforço computacional, sendo não tão precisos devido ao desacoplamento do modelo computacional. Os métodos *wrapper* avaliam a precisão de classificadores ao utilizar subconjuntos de recursos sendo a construção de um modelo para cada subconjunto de recursos muitas vezes computacionalmente intangível. Os

métodos incorporados treinam um modelo para classificação dos recursos e automaticamente verificam a importância de cada recurso [25].

Em [25] é utilizada a técnica denominada Eliminação Recursiva de Atributos (RFE - *Recursive Feature Elimination*). Por meio do método incorporado, retira-se de todo o conjunto de atributos aqueles menos importantes definidos pelo algoritmo. Normalmente, utiliza-se um critério de parada, que pode ser, por exemplo, um valor de erro de validação. Como conclusão, quando há alta correlação entre os recursos, este método não atinge boa performance. Neste caso, a utilização de *Random Forest* obtém resultados melhores na seleção de atributos.

A “Identificação da Carga” é feita classificando as propriedades dos dispositivos. O objetivo desta etapa é determinar uma assinatura que possa identificar exclusivamente cada equipamento elétrico. O desempenho de qualquer sistema NILM depende da exclusividade da assinatura do dispositivo em comparação com a de outros dispositivos. Portanto, a identificação de tal assinatura é de suma importância para melhorar a capacidade de determinação de carga de um sistema NILM [20 - 22].

Este processo pode ser realizado por meio de algoritmos de aprendizado de máquinas, modelos de correspondência e *Deep Learning*. [26]

Este último possui a capacidade de a partir dos sinais de entrada, quando comparados com sinais de saída correspondentes, gerar um modelo de inferência e relação entre os dados. Esta, poderia ser uma contribuição futura do tema abordado neste trabalho.

Uma vez que as oscilografias foram verificadas (entradas), os índices de potência e corrente calculados, e toda a classificação feita (saídas), o *Deep Learning*, com base nesse estudo preliminar pode, por meio deste modelo de inferência gerado a partir da relação entre entradas e saídas previamente conhecidas, ser utilizado para promover (gerar) uma Teoria de Potência estimada, pois como se baseia em um modelo de inferência, não é feito o levantamento e estudo com base em princípios físicos como em todas as Teorias de Potência conhecidas.

No que diz respeito aos algoritmos de aprendizado de máquinas, muitos são encontrados na literatura, como k-NN (Himeur, Alsalemi, Bensaali & Amira, 2021; Tsai & Lin, 2012), Redes Neurais (Diana & John Edwin, 2014), Máquina de Vetor de Suporte

(Lai, Lai, Huang & Chao, 2013), *Random Forest* (Wu, Gao & Jiao, 2019). Alguns estudos presentes na literatura, apontam RF como algoritmo mais utilizado para aplicações NILM destacando sua robustez e a tendência de obtenção de alta precisão de classificação dentro de um banco de dados com amostras limitadas [27].

A “Desagregação de Cargas” é a última etapa realizada para determinar o efeito dos dispositivos conectados à rede elétrica. Após a conclusão do processo de desagregação, é possível identificar o consumo de cada equipamento e repassar essa informação ao usuário, que pode ser no próprio medidor de energia elétrica, ou no *smartphone* [20 - 22].

Com o apresentado, entende-se que o sistema NILM então contribui para a bidirecionalidade e confiabilidade nas medições de consumo desagregadas, ou seja, permite à concessionária prever e gerenciar a demanda de fornecimento da energia elétrica, e ao consumidor, a conscientização e competência para economia da energia elétrica, evitando desperdícios e impactos negativos relacionados à qualidade da energia elétrica no sistema de distribuição.

## **2.2. Medidor inteligente**

Como elemento principal do sistema NILM, comentado na sessão anterior, destaca-se o medidor de energia elétrica. Este equipamento foi criado com o objetivo de quantificar o consumo de energia elétrica, com finalidades diversas, entre elas, a tarifação por meio das concessionárias [12 - 13].

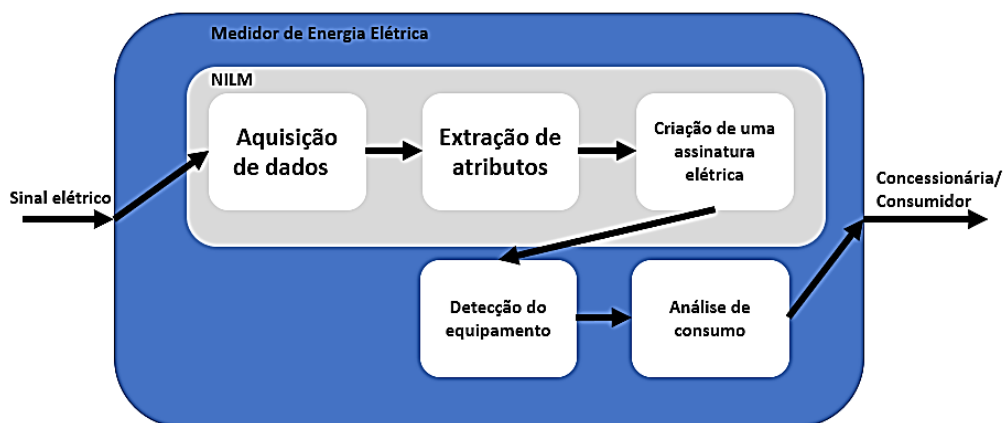
Inicialmente foram desenvolvidos diversos tipos de medidores, os quais possuíam princípios de medição, tais como: medir o peso do cátodo e ânodo da célula eletrolítica de cobre antes e após a utilização da energia (Edson, T. 1881), medir a oscilação de pêndulo provocada pelo campo magnético gerado na passagem de corrente em um condutor (Ferraris, G.; Tesla, N.; Shallenberger, O. 1885-1888), medição pela excitação de um rotor proveniente de uma tensão sob uma bobina (Gutmann, L. 1889), entre outros. Estes princípios eram utilizados principalmente em sistemas elétricos de corrente contínua, possuíam desgaste, construtivamente eram grandes e foram substituídos pelo princípio mais largamente conhecido, o indutivo [12 -13 - 28].



Com o avanço da eletrônica, os medidores elétricos por indução estão sendo substituídos pelos medidores eletrônicos. Estes novos medidores possuem o benefício de baixa manutenção, pois não utilizam de componentes mecânicos para medição, além de possibilitarem em alguns casos o armazenamento de informações de consumo e a interação do consumidor com a concessionária [12 - 13].

O fato da inclusão de elementos eletrônicos dentro de um medidor de energia elétrica, permite o aumento de sua gama de funcionalidades, uma vez que se torna capaz de processar, armazenar e transmitir dados de maneira rápida e segura. No que diz respeito ao medidor de energia elétrica inserido em um sistema NILM, entende-se que além de medir e quantificar a energia elétrica, o medidor deverá englobar as funções como a desagregação dos dados coletados, com o objetivo de criar a assinatura elétrica de cada equipamento elétrico, para então, determinar seu consumo isolado e transmitir esta informação tanto à concessionária, quanto ao consumidor, conforme pode ser visto na Figura 3.

**Figura 3- Medidor de Energia Elétrica e o Sistema NILM**



FONTE: AUTORIA PRÓPRIA

Os medidores inteligentes residenciais normalmente fornecem os componentes elétricos ativos e reativos de energia medidas em baixa frequência. Por este motivo, esses são os recursos e taxas de medição mais amplamente usados e relatados na literatura relacionada ao NILM. Por outro lado, considerando que o uso de aparelhos não lineares tem aumentado continuamente durante as últimas décadas, o conteúdo de frequência dos

sinais elétricos envolvidos pode ser explorado para melhorar a precisão no processo de identificação de carga. Isso implica que a taxa de amostragem usada para a aquisição de sinais de tensão e corrente deve ser suficientemente alta a ponto que permitam que os componentes de frequência de interesse sejam avaliados corretamente [19].

Os elementos que fornecem a aptidão de inteligência ao medidor estão baseados em algoritmos desenvolvidos para coletar os dados de tensão e corrente da linha, e posteriormente na análise de padrões de consumo. Para esta análise, podem ser utilizados índices de potência calculados por meio de teorias de potência modernas, dentre estas, as abordadas neste trabalho: CPC [5], CPT [6] e IEEE 1459-2010 [8]. Estes índices são comparados com um banco de dados por meio de técnicas de aprendizado de máquina. Tais técnicas são baseadas em algoritmos classificadores, que são capazes de identificar e selecionar a que grupo pertence determinado dado.

A classificação neste sentido é usada para individualizar os equipamentos elétrico/eletrônicos em consumo, com base nos atributos desse objeto. A identificação da carga elétrica e sua classificação desempenham papéis importantes na tomada de decisão dos sistemas de energia podendo melhorar o planejamento da distribuição e aumentar a individualização do serviço de energia para consumidores e concessionárias. Quando a precisão da classificação é alta, a análise e as decisões podem levar a uma melhor otimização no planejamento de geração e consumo de energia elétrica [11].

### **2.3. Teorias de Potência**

A qualidade da energia elétrica tem sido comprometida devido ao uso generalizado de dispositivos eletrônicos, desta forma, o estudo da mesma é crucial para a análise e compreensão do sistema elétrico, no qual o produto comercializado é a energia elétrica [29].

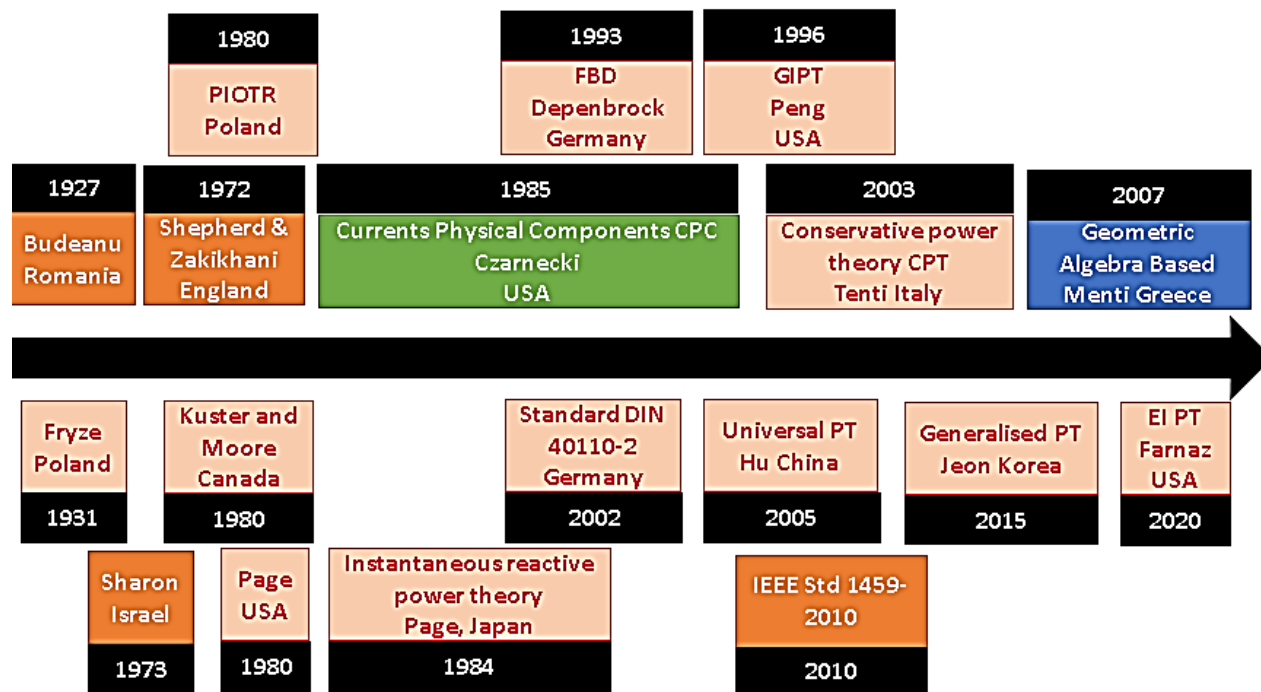
Prever melhorias nos sistemas elétricos requer uma boa descrição dos fenômenos físicos. Com o intuito de mitigar o efeito das perturbações e para manter uma boa qualidade de energia, faz-se necessário o cálculo de fatores de potência, que são índices que possibilitam realizar a desagregação de cargas calculados por meio de teorias de potência. Atualmente, existem muitas teorias que descrevem o fluxo de energia, porém

cada uma possui uma abordagem e um foco de análise para compreensão de fenômenos e cálculos de índices de potências [1, 30, 31].

O problema dos harmônicos gerados pela utilização de equipamentos eletrônicos é o principal fato que impulsiona o desenvolvimento de ferramentas simples para avaliar os níveis de poluição harmônica no ponto de acoplamento comum, bem como em toda a rede. Dentre elas, encontra-se o medidor de energia elétrica inteligente, que faz uso das teorias de potência existentes para caracterizar cargas elétricas e proporcionar identificação, quantificação e qualificação dos pontos de impacto na qualidade da energia elétrica [29].

As teorias de potência e suas contribuições têm sido documentadas há pelo menos um século. Com uma melhor compreensão dos fenômenos físicos, os conceitos abordados por cada uma foram aperfeiçoados. A Figura 4 mostra o diagrama de principais fatos históricos e descobertas científicas que impulsionaram o desenvolvimento das teorias que são empregadas neste trabalho [5 - 32].

**Figura 4 - Diagrama da evolução das Teorias de Potência**



FONTE: ADAPTADO DE: [33]

Em 1892, Steinmetz fez um experimento, testando a aplicabilidade das teorias de potências existentes desenvolvidas para circuitos elétricos lineares e invariantes no tempo, como por exemplo, a veracidade da aplicação do triângulo de potência. Para tal, o circuito linear invariante no tempo, foi substituído por um uma lâmpada de arco, demonstrando que apesar de a potência reativa ser nula, a potência aparente poderia ser maior que a potência ativa. Este fato, despertou o interesse e impulsionou o desenvolvimento de teorias de potência, a fim de obter respostas para dois fenômenos observados: a potência aparente ser maior que a potência ativa e como reduzir a diferença entre elas com uma rede de compensação [5 - 34].

Em 1927, C.I. Budeanu propôs em sua publicação “Potências reativas e fictícias” uma formulação analítica da potência reativa em regime periódico não senoidal. Além disso, o autor definiu a potência de distorção, conforme (1) [34].

$$D_B \triangleq \sqrt{\sum_{n=1}^{\infty} \sum_{m=1, m \neq n}^{\infty} [V_n^2 I_n^2 + V_m^2 I_m^2 - 2V_n I_m V_m I_n \cos(\varphi_n - \varphi_m)]} \quad (1)$$

$$S^2 = P^2 + Q^2 + D_B^2 \quad (2)$$

A potência de distorção será nula se:

$$\frac{V_n}{V_m} = \frac{I_n}{I_m} , \quad \varphi_n = \varphi_m \quad \forall n, m \quad (3)$$

Ademais, Fryze em 1932 introduziu o conceito de corrente ativa  $i_a$ , definida por (4):

$$i_a(t) = \frac{P}{\|u\|^2} u(t) = G_e u(t) \quad (4)$$

Onde  $\|u\|$  representa a tensão eficaz e  $G_e$  a condutância equivalente de uma carga simétrica que nas mesmas tensões tem a mesma potência ativa P.

Conforme mencionado em [34], Shepherd e Zakikhani introduziram o conceito de corrente reativa  $i_r(t)$ , definida a partir de uma fonte de tensão com n harmônicas de um conjunto N e na forma complexa da Série de Fourier, conforme (5):

$$i_r(t) \stackrel{\text{def}}{=} \sqrt{2} \operatorname{Re}(\sum_{n \in N} j B_n U_n e^{jn\omega_1 t}) \quad (5)$$

Onde  $B_n$  é susceptância de carga para o harmônico de enésima ordem e  $U_n$  é a tensão de pico no enésimo harmônico.

Entre 1972 e 1984, destacam-se os trabalhos de Sharon, 1973, Emanuel, 1974, Kusters e Moore, 1980, Page e Filipski, 1980 e Akagi, Kanazawa e Nabar, 1983, que contribuíram para evolução e análise da corrente reativa, no que tange ao desenvolvimento de teorias relacionadas com a observação das correntes indutiva e capacitiva e seu comportamento com relação às potências [34].

Desde 1920, foram propostas diversas teorias de potência não sinusoidais para sistemas monofásicos e trifásicos. No geral, é possível dividi-las em quatro grupos conforme Figura 4. Os blocos em vermelho representam as teorias de potência baseadas no domínio do tempo, enquanto os blocos laranjas são no domínio da frequência, os blocos verdes no domínio do tempo e frequência, e por fim, os blocos azuis representam as teorias no domínio de álgebra geométrica. Cada uma delas será mais bem utilizada de acordo com a aplicação, uma vez que fornecem melhores resultados em identificar harmônicos e potências reativas, compensar e suprimir componentes de energia, entre outros [33].

Considerando os avanços apresentados, serão utilizadas no trabalho as teorias denominadas CPC (*Currents Physical Components*), CPT (*Conservative Power Theory*) e IEEE1459 STD 2020 que serão apresentadas nas próximas sessões. Cada uma das teorias é baseada em índices de corrente ou potência que possibilitam desagregar as cargas elétricas mesmo em condições não senoidais. Uma vez que, dentro de suas definições matemáticas, existem abordagens em domínios diferenciados busca-se com a escolha destas três teorias de potência observar a influência de cada uma delas na representação das cargas elétricas.

### 2.3.1. CPC (Currents Physical Components)

A CPC (*Currents Physical Components*) é uma teoria de potência desenvolvida no domínio do tempo e da frequência, foi proposta por Czarnecki entre os anos de 1984 a 1988, e, visa separar a corrente elétrica em componentes ortogonais, as quais possuem relações com fenômenos físicos. A teoria CPC se baseia nos conhecimentos anteriores, e sendo assim, a teoria abrange os conceitos discutidos por Sharon, Fryze e Shepherd e Zakikhani, partindo da decomposição da corrente elétrica total em parcelas [34 – 35 - 36].

A teoria permite calcular várias potências em sistemas monofásicos e trifásicos. O conceito é aplicado geralmente para o cálculo e definição de potências não ativas na presença de correntes e tensões não senoidais. Os espectros de cada forma de onda são usados para definir a origem da forma de onda por meio da teoria CPC. [11]

A série de Fourier é a maneira geral de apresentar tensão harmônica não senoidal no domínio do tempo, conforme sua definição dada em (6):

$$u(t) = U_0 + \sqrt{2} \operatorname{Re}(\sum_{n \in \mathbb{N}} \bar{U}_n e^{jn\omega_1 t}) \quad (6)$$

Onde:  $U_0$  é a componente DC e  $\bar{U}_n$  é o valor eficaz da tensão no enésimo harmônico.

A admitância  $Y_n$  para o enésimo harmônico pode ser expressa em sua forma complexa (7). A mesma, multiplicada pela tensão, resulta na série de Fourier da corrente harmônica, conforme (8).

$$Y_n = G_n + jB_n \quad (7)$$

$$i(t) = \sqrt{2} \operatorname{Re}(\sum_{n \in \mathbb{N}} Y_n \bar{U}_n e^{jn\omega_1 t}) \quad (8)$$

A potência ativa  $P$  é proveniente do produto escalar entre tensão e corrente, conforme (9).

$$P = \sum_n \frac{U_n I_n}{2} \cos \varphi_n \quad (9)$$

Nesta equação o índice “n” representa a harmônica de referência, e  $\varphi_n$  é o ângulo de defasagem entre corrente e tensão considerando a mesma harmônica.

A corrente ativa, permanece com sua definição conforme apresentado por Fryze em (4). A corrente reativa por sua vez, é obtida subtraindo-se da corrente geral a corrente ativa. Seguindo esta operação, tomando como base a parte imaginária da mesma, a corrente reativa permanece baseada na definição apresentada por Shepherd e Zakikhani, em (5) [11].

Em 1984, surgiu o conceito de corrente dissipada  $i_s(t)$ , harmônica e ortogonal à tensão de alimentação não compensável por uma rede de elementos reativos, definida por (10). Esta componente de corrente existirá se  $G_n \neq G_e$  [5 – 11 - 34].

$$i_s(t) \stackrel{\text{def}}{=} (G_0 - G_e)U_0 + \sqrt{2}Re \sum_{n \in N} (G_n - G_e)U_n e^{jn\omega_1 t} \quad (10)$$

Onde  $G_n$  é a condutância no enésimo harmônico e  $G_e$  a condutância equivalente.

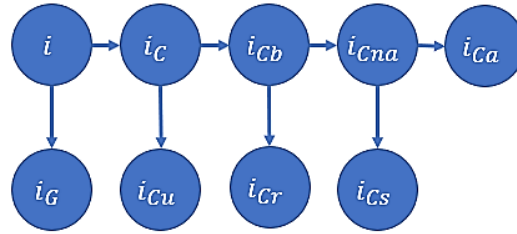
Em 1990, foi levantada a questão de como escrever uma equação de potência de harmônicos de corrente gerando pelas cargas (HGLs) e descobriu-se que a potência ativa dos harmônicos individuais pode ser negativa (segundo  $\varphi_n$ ), significando um fluxo de energia contrário, da carga para a fonte de alimentação. Neste caso, as harmônicas geradas podem ser divididas em dois grupos:  $N_c$  corresponde ao conjunto de harmônicas que foram originadas na própria fonte e,  $N_G$  por sua vez, é o conjunto de harmônicas originadas na carga. A corrente associada ao grupo  $N_G$  é chamada de corrente de carga gerada  $i_G$  [11, 34, 37].

$$i = i_a + i_r + i_s + i_G = i_C + i_G \quad (11)$$

A composição de (11) pode também ser representada de forma visual através da Figura 5, na qual, é possível, notar que a corrente  $i$ , foi dividida em  $i_C$  e  $i_G$ . Além disso, é possível notar que  $i_C$  é subdividida em outras componentes de corrente, sendo  $i_{Cu}$  a corrente útil gerada pela fonte,  $i_{Cb}$  a corrente de balanceamento gerada pela fonte,  $i_{Cr}$  a

corrente reativa gerada pela fonte,  $i_{Cna}$  a corrente ativa no  $n$ ésimo harmônio gerada pela fonte,  $i_{Cs}$  a corrente de espalhamento gerada pela fonte e  $i_{Ca}$  a corrente ativa gerada pela fonte.

**Figura 5- Diagrama de decomposição de correntes segundo a CPC**



FONTE: ADAPTADO DE: [31]

Caso o sistema seja trifásico com um desequilíbrio de carga, produz-se um efeito no consumo de  $i_C$ . Para identificar a componente relacionada a este efeito, uma admitância equilibrada equivalente  $Y_{nb}$  é calculada como em (12), em que  $S_n^*$  é o conjugado de potência aparente do complexo harmônio e  $\|U_n\|$  é o valor eficaz de  $u_n$ . A soma das componentes de corrente resulta na corrente balanceada  $i_{Cb}$ , definida em (13) e a diferença entre esta e  $i_C$  resulta na corrente de desequilíbrio  $i_{Cu}$ , conforme abaixo [37].

$$Y_{nb} = \frac{S_n^*}{\|U_n\|^2} = \frac{P_n - j \cdot Q_n}{\|U_n\|^2} = G_{nb} + j \cdot B_{nb} \quad (12)$$

$$i_{Cb} = \sqrt{2} \operatorname{Re} \sum_{n \in N_C} Y_{nb} U_n e^{jn\omega_1 t} \quad (13)$$

$$i_{Cu} = i_C - i_{Cb} \quad (14)$$

Ressalta-se que para tensões simétricas ou cargas balanceadas, a admitância equivalente é definida por (15). Se as tensões forem assimétricas e as cargas forem desequilibradas, existirá uma diferença entre as admitâncias  $Y_{nb}$  e  $Y_{ne}$ , segundo (16): [37]

$$Y_{ne} = Y_{nRS} + Y_{nST} + Y_{nTR} = \frac{Y_{nR} + Y_{nS} + Y_{nT}}{3} \quad (15)$$

$$Y_{nd} \stackrel{\text{def}}{=} Y_{ne} - Y_{nb} \quad (16)$$



Conforme definição de Shepherd e Zakikhani, a corrente balanceada é dividida em corrente reativa  $i_{Cr}$  e corrente relacionada à condutância harmônica  $i_{Cna}$ . A corrente reativa é resultado da mudança de fase entre elementos harmônicos da corrente e da tensão, e não às oscilações de energia entre carga e fonte, como comumente interpretada. Por sua vez, a corrente  $i_{Cna}$  é calculada conforme (17) e difere-se da corrente ativa convencional [37].

$$i_{Ca} = G_e u_C = \frac{P_C}{\|u_C\|^2} u_C = \frac{\sum_{n \in N_C} P_n}{\sum_{n \in N_C} \|u_C\|^2} u_C \quad (17)$$

Onde  $u_C$  é a tensão gerada pela fonte.

A última componente é a porção da soma ativa harmônica que ocorre devido ao espalhamento dos valores de condutância  $G_{nb}$  ao redor da condutância equivalente  $G_e$ , chamada de corrente dispersa  $i_{Cs}$  e definida por (18): [37]

$$i_{Cs} = i_{Cna} - i_{Ca} = \sqrt{2} Re \sum_{n \in N_C} (G_{nb} - G_e) U_n e^{jn\omega_1 t} \quad (18)$$

Tais componentes de corrente são mutualmente ortogonais e compõe a base da Teoria de Componentes físicas de Corrente, obedecendo às equações:

$$\|i\|^2 = \|i_{Ca}\|^2 + \|i_{Cr}\|^2 + \|i_{Cs}\|^2 + \|i_{Cu}\|^2 + \|i_G\|^2 \quad (19)$$

Multiplicando as componentes de corrente por  $\|u\|^2$  atinge-se os termos de potência relacionados a cada componente de corrente já definida previamente.

$$S^2 = P_C^2 + D_{Cs}^2 + Q_C^2 + D_{Cu}^2 + D_G^2 \quad \text{Sendo: } P_C = P - P_G \quad (20)$$

Onde  $D_{Cs}$  é a potência de dispersão gerada pela fonte,  $Q_C$  é a potência de reativa gerada pela fonte,  $D_{Cu}$  é a potência de dispersão útil gerada pela fonte e  $D_G$  é a potência de dispersão gerada pela carga.

O fator de potência é definido conforme (21).

$$\lambda = \frac{P_C}{S} = \frac{\|i_{Ca}\|}{\|i\|} \quad (21)$$

Pode-se observar que a CPC é a primeira teoria de potência que fornece interpretação física completa de fenômenos elétricos no que se refere à potência elétrica. Além disso, contribui para a melhoria do fator de potência com compensadores reativos em sistemas monofásico, trifásicos e polifásicos de tensão e correntes não senoidais.

### 2.3.2. CPT (Conservative Power Theory)

A Teoria da Potência Conservativa baseia-se nos cálculos de potências associadas a condições não senoidais e/ou assimétricas (em sistemas trifásicos), servindo como uma importante ferramenta de análise de sistemas elétricos em condições gerais de operação. Foi apresentada por Tenti et al. em 2003 [6], como uma proposta de análise de circuitos não lineares, no domínio do tempo. Fundamentava-se na definição de variáveis elementares de corrente e tensão relacionadas com suas respectivas derivadas e integrais. Para relacionar a corrente com seus respectivos fenômenos físicos de potência (ou energia) na carga, os autores introduziram o conceito de homo variáveis [7].

A Teoria foi reformulada e reapresentada em 2010 [38]. A nova proposta apresenta uma evolução na compreensão do comportamento dos sistemas elétricos e sua associação com correntes e potências decompostas. Suas definições são ditas conservativas por terem como base as Leis de Tensões e Correntes de Kirchhoff e o Teorema de Tellegen. Tal teoria foi chamada de Teoria de Potência Conservativa, do inglês, *Conservative Power Theory* (CPT). Nela, os autores introduziram o conceito de variáveis de derivada e *unbiased* integral ou integral imparcial [7].

A CPT proporciona uma análise abrangente dos fenômenos físicos relacionados às tensões e correntes, como alterações na frequência de operação, assimetria das tensões e desbalanceamento das cargas. O fato de lidar com grandezas e operações conservativas, fornece à CPT a capacidade de analisar condições de corrente e tensão não senoidais, de forma diversa das teorias de circuitos elétricos convencionais, permitindo a criação de metodologias de análise e controle para redes elétricas. Com isso, permite analisar, modelar e prever respostas em cargas submetidas a condições senoidais não totalmente puras. Este fato, dá a possibilidade de conhecer fatores prejudiciais a circulação de energia útil na rede [7].

Assim, é possível constatar que a CPT é de extrema valia atualmente, em que o uso racional de energia é um dos temas mais discutidos, se for considerado que através das variáveis que ela utiliza é possível o aprofundamento no estudo da operação do circuito elétrico, com a finalidade de observar sua operação em condições normais da rede e em condições adversas, tornando possível a prevenção de possíveis danos ou problemas de sinais.

A Tabela 2 mostra os índices definidos na CPT e utilizados na mesma, como são calculados, e logo abaixo, seu significado físico no sistema de energia elétrica.

**Tabela 2- Índices da CPT, como são calculados e seus significados físicos**

Índice	Representação	Cálculo
Potência Aparente	A	$A = I \cdot V$
Corrente Ativa	$i_a$	$i_a = G_e \cdot v$
Potência Ativa	P	$P = I_a \cdot V$
Corrente Reativa	$I_r$	$i_r = \frac{W_r}{\hat{V}^2} \hat{v}$
Potência Reativa	Q	$Q = V \cdot I_r$
Corrente Residual	$i_v$	$i_v = i - i_a - i_r$
Potência Residual	D	$D = I_v V$
Potência de Desbalanço	U	$U^2 = U_a^2 + U_r^2$
Fator de Potência	$\lambda$	$\lambda = \frac{P}{A}$

**Tabela 2- Índices da CPT, como são calculados e seus significados físicos (continuação)**

Índice	Representação	Cálculo
Fator de não-linearidade	$\lambda_D$	$\lambda_D = \frac{D}{A}$
Fator de Reatividade	$\lambda_Q$	$\lambda_Q = \frac{Q}{\sqrt{P^2 + Q^2}}$
Fator de Assimetria	$\lambda_U$	$\lambda_U = \frac{U}{\sqrt{P^2 + Q^2 + U^2}}$

Onde:

$G_e$ : é a condutância equivalente medida em  $ohm^{-1}$ ;

U: é a potência de desbalanço, é proveniente do desbalanceamento das cargas e não foi utilizada no trabalho por se tratar de circuitos monofásicos;

$U_a$  : é a potência de desbalanço ativa;

$U_r$  : é a potência de desbalanço reativa;

Os principais índices que podem ser obtidos ou calculados por meio da CPT são: potência instantânea coletiva, potência ativa, potência reativa, potência residual (nula), potência ativa dispersa, potência reativa dispersa, potência aparente, potência harmônica gerada, energia reativa instantânea coletiva, corrente ativa, corrente reativa, corrente residual, corrente ativa dispersa, corrente residual dispersa, corrente gerada pela carga, fator de distorção da tensão, fator de não linearidade, fator de reatividade.

Os conceitos dos índices apresentados, seguem abaixo:

- **Potência aparente:** Potência proveniente da fonte de alimentação do circuito;
- **Corrente ativa:** Corrente mínima necessária para o transporte de potência ativa útil;
- **Potência ativa:** Potência proveniente da corrente ativa;
- **Corrente reativa:** Corrente mínima necessária para transporte de energia reativa; quando o termo  $\frac{W_r}{V^2}$  é positivo, indica energia reativa proveniente de indutores, quando é negativo, indica energia reativa proveniente de capacitores;
- **Potência reativa:** Potência proveniente da corrente reativa;

- **Corrente residual:** É a corrente existente devido às distorções na corrente e na tensão; não transporta potência ativa e energia reativa;

- **Potência residual:** Potência relacionada com a absorção de corrente residual quando há não linearidade ou distorção entre tensão e corrente;

- **Potência de Desbalanço:** Potência formada pela parcela de desbalanço da potência ativa e da reativa;

- **Fator de potência:** Índice utilizado para caracterizar o uso da rede, quanto mais próximo de 1, indica menos distorções harmônicas presentes na corrente;

- **Fator de não-linearidade:** Representa a relação entre potência residual e aparente;

- **Fator de reatividade:** Representa a influência da corrente reativa e residual presente no fator de potência;

É importante notar que na teoria, as variáveis representadas em maiúsculo indicam os valores eficazes, bem como as minúsculas indicam valores instantâneos.

Além disso, algumas das variáveis se relacionam seguindo a equação (22).

$$A = \sqrt{P^2 + Q^2 + D^2 + U^2} \quad (22)$$

Em meio ao apresentado, destaca-se as principais aplicações da CPT: identificação de características de cargas elétricas e análise de comportamento, prestação de contas e medição de receita, filtro e compensação de energia usando conversor de potência [33].

Por fim, a CPT oferece uma forma de obter circuitos equivalentes de cargas elétricas e identificar suas características. Em [39] representa-se circuitos originais por meio de circuitos passivos equivalentes, com os mesmos índices de potência. Em [40] e [41] a CPT foi usada para analisar o comportamento de máquinas de raio X e máquinas de soldagem elétrica [33].

### 2.3.3. NORMA IEEE 1459-2010

A Norma IEEE 1459-2010, “Definições para a medição de quantidades de energia elétrica sob sinusoidal, condições não sinusoidais, equilibradas ou desequilibradas”, é baseado em um sinal periódico com uma forma de onda não senoidal composta por uma superposição de harmônicos que pode ser matematicamente representada pela série de Fourier [5, 42].

Visando determinar um padrão, a ser adotado pela comunidade científica e produtiva internacional, foi elaborado o documento *Institute of Electrical and Electronics Engineers - IEEE 1459*, no qual defende-se a decomposição da potência aparente em fundamental e não fundamental, desta forma, enfatizando a potência derivada de distorções harmônicas de corrente e tensão [1].

A atualização de março de 2010 da IEEE Std. 1459 é o um padrão disponível composto por definições de potências, quantificares de fluxo de energia em sistemas monofásicos e trifásicos senoidal e não senoidal, balanceado ou não, facilitando a identificação de parâmetros e fontes poluidoras, contendo harmônicos [43 - 44].

A norma IEEE 1459 é baseada na separação das componentes fundamentais à sequência direta de tensão e corrente dos fasores restantes, e por este motivo, permite a medição de grandezas tradicionais, como potência ativa, reativa, aparente, além de parâmetros e indicadores de distorção harmônica e do grau de desequilíbrio de medição [34].

O padrão IEEE 1459 define os valores eficazes de tensão e corrente ( $V$  e  $I$ ) e das componentes harmônicas na frequência fundamental ( $V_1$  e  $I_1$ ) e, ainda, os valores eficazes das componentes harmônicas ( $V_h$  e  $I_h$ ). A partir destas definições, explica-se as relações entre as potências conhecidas e a potência instantânea exemplificando o significado físico da potência aparente como máxima potência transmitida [34].

A característica principal das definições é a distinção entre a fundamental e a soma das harmônicas que compõe a tensão ou corrente. [45].

A Tabela 3 apresenta o cálculo da potência aparente, ativa e não ativa, conforme IEEE 1459-2010, considerando circuitos monofásicos. Estes cálculos serão os

empregados no decorrer do estudo, uma vez que no Brasil, as cargas elétricas encontradas nos ambientes residenciais são em maior parte monofásicas.

**Tabela 3-Índices de potência para sistemas monofásicos segundo IEEE 1459-2010**

Potência	Combinada	Fundamental	Harmônica
Aparente [VA]	$S = VI$	$S_1 = \sqrt{P_1^2 + Q_1^2}$	$S_H = V_H I_H$ $S_N = \sqrt{S^2 - S_1^2}$
Ativa [W]	$P = \sum_{h \neq 1} V_h I_h \cos \theta_h$	$P_1 = \frac{1}{kT} \int_{\tau}^{\tau+kT} v_1 i_1 dt = V_1 I_1 \cos \theta_1$	$P_H = V_0 I_0 + \sum_{h \neq 1} V_h I_h \cos \theta_h$ $= P - P_1$
Não ativa [VAR]	$N = \sqrt{S^2 - P^2}$	$Q_1 = \frac{\omega}{kT} \int_{\tau}^{\tau+kT} i_1 \left[ \int v_1 dt \right] dt$ $= V_1 I_1 \sin \theta_1$	$D_I = V_1 I_H = S_1 (THD_I)$ $D_V = V_H I_1 = S_1 (THD_V)$ $D_H = \sqrt{S_H^2 - P_H^2}$

Neste caso, para condições de estado estacionário, a tensão ou corrente instantânea periódica e não senoidal possui duas componentes distintas, sendo a parcela fundamental do sistema de energia  $v_1$  e  $i_1$  e a harmônica  $v_H$  e  $i_H$ , definidas conforme equações (23) a (26).

$$v_1 = \sqrt{2V_1} \sin(\omega t - \alpha_1) \quad (23)$$

$$i_1 = \sqrt{2I_1} \sin(\omega t - \beta_1) \quad (24)$$

$$v_H = V_0 + \sqrt{2} \sum_{h \neq 1} V_h \sin(h\omega t - \alpha_h) \quad (25)$$

$$i_H = I_0 + \sqrt{2} \sum_{h \neq 1} I_h \sin(h\omega t - \beta_h) \quad (26)$$

A potência aparente é a quantidade de potência ativa que pode ser fornecida a uma carga, ou um conjunto de cargas, sob condições ideais (sem distorções ou reativos). As cargas são compensadas por meio de dispositivos ativos ou passivos, de modo que a corrente da linha seja senoidal e em fase com a tensão. O valor eficaz da corrente é mantido igual ao valor eficaz da corrente de linha real. A tensão de carga é ajustada para um valor que produz desempenho de carga inalterado. Se o critério de desempenho for a

conversão eletrotérmica de energia elétrica por efeito Joule, então o valor eficaz da tensão nos terminais onde a medição é implementada deve ser mantido constante [3].

A Tabela 4, mostra o cálculo da potência aparente, ativa e não ativa, conforme IEEE 1459-2010 para circuitos trifásicos, considerando as parcelas combinada, fundamental e não fundamental.

**Tabela 4-Índices de potência para sistemas trifásicos segundo IEEE 1459-2010**

Potência	Combinada	Fundamental	Harmônica
Aparente [VA]	$S_e = 3V_e I_e$	$S_{e1} = 3V_{e1} I_{e1}$ $S_1^+ = 3V_1^+ I_1^+$ $S_{U1} = \sqrt{S_{e1}^2 - S_1^{+2}}$	$S_{eH} = 3V_{eH} I_{eH}$ $S_{eN} = \sqrt{S_e^2 - S_{e1}^2}$
Ativa [W]	$P = \sum_{a,b,c} \sum_{h=1}^n V_h I_h \cos\theta_h$	$P^+ = 3V_1^+ I_1^+ \cos\theta_1^+$	$P_H = \sum_{a,b,c} \sum_{h=2}^n V_h I_h \cos\theta_h$ $= P - P_1$
Não ativa [VAR]	$N = \sqrt{S_e^2 - P^2}$	$Q^+ = 3V_1^+ I_1^+ \sin\theta_1^+$	$D_{eI} = 3V_{e1} I_{eH}$ $D_{eV} = 3V_{eH} I_{e1}$ $D_{eH} = \sqrt{S_{eH}^2 - P_{eH}^2}$

A Tabela 5 apresenta os indicadores relacionados ao conteúdo harmônico tanto para sistemas monofásicos, quanto para trifásicos.

**Tabela 5 - Indicadores de conteúdo harmônico segundo IEEE 1459-2010**

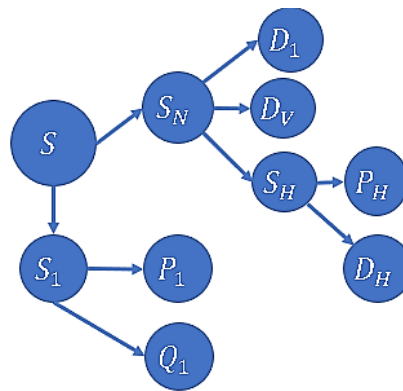
Indicadores	Monofásico	Trifásico
Fator de potência (Utilização da linha)	$PF = \frac{P}{S}, PF_1 = \frac{P_1}{S_1}$	$PF_e = \frac{P}{S_e}, FP_1^+ = \frac{P_1^+}{S_1^+}$
Poluição harmônica	$\frac{S_N}{S_1}$	$\frac{S_{eN}}{S_{e1}}$
Desbalanceamento de carga		$\frac{S_{U1}}{S_1^+}$



Como pode-se notar, os indicadores são definidos como sendo relações entre as quantidades de energia, lembrando o conceito comum de fator de potência normalmente utilizado para compensação de potência reativa.

As medições das quantidades de energia determinadas por meio do padrão IEEE 1459 possuem simplicidade no entendimento, uma vez que são baseadas na separação das componentes fundamentais do conteúdo harmônico remanescente de tensão e corrente. Vários instrumentos comerciais atuais permitem medir as potências fundamental e total ativo, não ativo e aparente. A Figura 6 mostra o diagrama de blocos das potências calculadas por meio da norma IEEE 1459 e as subdivisões em componentes de potências harmônicas e fundamentais.

**Figura 6 - Diagrama de Potências segundo a IEEE 1459-2010**



FONTE: ADAPTADO DE: [4]

#### 2.3.4. Principais diferenças entre as teorias de potência

Fundamentando-se nas revisões bibliográficas e nas informações e discussões expostas neste capítulo, é possível perceber que, embora as teorias de potência abordadas se baseiem nos mesmos estudos previamente definidos, cada uma delas possui um ponto de partida e um foco de observação diverso. Todas são utilizadas para auxiliar no cálculo da qualidade de energia elétrica, seja identificando os pontos poluidores de harmônicos na rede ou quantificando as distorções. Além disso, utilizam-se da decomposição de componentes de corrente ou potência afim de permitir a

desagregação das cargas elétricas. Por outro lado, existem divergências entre as teorias, pois algumas foram criadas voltadas para os procedimentos de tarifação, enquanto outras voltadas para projetos de compensadores e de compensação de energia. Outra grande diferença é a nomenclatura dos mesmos significados físicos ou matemáticos [1].

Além das abordagens, nomenclaturas e dos cálculos serem diferentes, alguns pontos que distinguem a CPC, CPT e IEEE Std. 1459-2010 são apresentadas a seguir:

- Na CPC as harmônicas de potência reativa são somadas geometricamente. Na IEEE não existe esta componente de potência reativa correspondente definida. Ambas utilizam Fourier. Na norma IEEE 1459-2010 dispositivos capacitivos ou indutivos podem ser determinados pelo sinal da potência reativa [30];

- A IEEE Std. 1459-2010 possui simplicidade na decomposição, bem como o destaque de potência fundamental. A CPC alinha-se com a essência da pesquisa de energia, uma vez que cada componente de corrente tem sua interpretação física correspondente [46];

- Em simulações com sistemas elétricos híbridos monofásicos de compensação passivo-ativo, a CPC possui maior eficácia na compensação em comparação à IEEE Std. 1459-2010 [46];

- A CPT é vista como uma técnica interessante para a análise, controle e regulação de diferentes circuitos de potência, desde o caso dos sinais de tensão e corrente senoidais e balanceados tradicionais até circuitos de carga não linear com tensões deterioradas [1];

- A CPT tem o conceito da ortogonalidade entre as correntes. Além disso, os antecedentes físicos e matemáticos apresentam o mesmo processo de potência ativa. Portanto, potência reativa está atrelada à energia reativa e à frequência, que leva em consideração a energia armazenada no indutor e no capacitor. Assim, a definição de potência reativa não é conservativa para sistemas cuja frequência varia e possui distorções de tensão, ao contrário da teoria proposta por Budeanu e Emanuel que somam todos os componentes harmônicos, desconsiderando as características indutivas e capacitivas de cada ordem [1];

- Ao contrário das definições de energia recomendadas pelo IEEE Std.1459-2010 com base na decomposição simultânea de tensão e corrente, a teoria CPC é apenas

sobre a decomposição de correntes. Na teoria da CPC, cada componente de corrente tem seu fenômeno físico correspondente [46].

Por fim, embora uma abordagem unificada para as teorias de potência minimize as contradições e erros, podendo ser implementada tanto na medição e tarifação de energia, muitos estudos ainda são necessários para aprimorar os conceitos atuais e suas eventuais deficiências, a fim de que possam ser aplicadas com eficiência a qualquer condição de circuitos elétricos [1].

## **2.4. Inteligência artificial e aprendizado de máquinas**

Ainda não existe uma única definição para Inteligência Artificial. Pode-se dizer que é uma junção de quatro campos: pensar como humano, pensar racionalmente, agir como seres humanos e agir racionalmente [1].

Um sistema de Inteligência Artificial deve ser capaz de realizar três tarefas principais: armazenar o conhecimento, aplicá-lo para resolver problemas e adquirir mais conhecimento por meio da experiência, sendo composto por três ferramentas principais: representação, racionalidade e aprendizado. A representação está ligada ao fato do uso de símbolos para representar conhecimentos gerais sobre problemas e suas soluções. Racionalidade está relacionada com a habilidade e controle para resolver uma larga escala de tipos de problemas. Aprendizado se relaciona com a capacidade de adquirir conhecimento a partir de experiências e utilizar estes aprendizados para realizar novas tarefas [2].

### **2.4.1 Inteligência artificial e contextualização**

Normalmente, em sistemas NILM a desagregação de cargas elétricas inclui cinco partes: Aquisição de dados, Processamento de dados, Detecção de Eventos, Extração de atributos e Identificação de Cargas. Esta última etapa é caracterizada pela utilização de algoritmos de classificação para caracterização das cargas elétricas com base nos atributos previamente levantados e destaca-se a utilização de algoritmos de aprendizado de máquinas supervisionado e não supervisionado [47 - 48].

Na literatura encontram-se muitos artigos aplicando diferentes métodos para a classificação de cargas em sistemas NILM [47 - 48]. Sistemas supervisionados requerem um banco de dados com informações para treinar o algoritmo. Sistemas não supervisionados são projetados para extrair toda a informação para operar diretamente da medição dos perfis de consumo. Devido à performance observada nos trabalhos existentes, a maior parte das abordagens se baseia em algoritmos supervisionados [49].

Desta forma, entende-se que o padrão de classificação de cargas depende dos atributos previamente extraídos. O objetivo de identificar as cargas pode ser alcançado à medida que se aprende os atributos provenientes de cada tipo de carga elétrica. Uma máquina de aprendizado deve ser capaz de melhorar suas tarefas baseado em suas experiências. Sendo assim, o sistema poderá utilizar os atributos como dados de treinamento, permitindo ao sistema de aprendizado fornecer saídas conforme o esperado. Para a aplicação em questão, são comumente utilizados os algoritmos *k-Nearest Neighbor* (k-NN), *Artificial Neural Network* (ANN), *Support Vector Machine* (SVM), *Random Forest* (RF), entre outros [48].

#### 2.4.2 Algoritmos de classificação

A inteligência a ser desenvolvida, deve proporcionar ao medidor a capacidade de reconhecer qual o tipo de carga por meio dos índices calculados com auxílio das teorias de potência, e ainda quantificá-la de modo a verificar as contribuições no consumo total.

Sendo assim, neste trabalho serão utilizados algoritmos classificadores, especificamente serão estudados e comparados os algoritmos denominados Máquina de Vetores de Suporte (SVM), o k-ésimo Vizinho Mais Próximo (k-NN), *Random Forest* (RF), *Decision Tree* (DT) e *Deep Learning*.

##### 2.4.2.1 SVM

O SVM (*Support Vector Machines*) é um conjunto de métodos de aprendizado supervisionados universalmente utilizado para classificação, regressão e detecção de *outliers*, sendo criado por Vapnik e aplicada em 1992 [50].

Máquina de Vetores de Suporte é uma ferramenta matemática de aprendizagem estatística aplicável às tarefas de classificação e regressão de padrões que possui como base o princípio de minimização de risco estrutural, utilizada para separação de conjuntos de elementos. A separação é feita através da definição de um hiperplano em um espaço característico de alta dimensão, a fim de solucionar problemas linearmente separáveis. No caso em que o problema seja não linear, altera-se a função kernel, para que métodos lineares possam ser aplicados. A equação (27), representa uma função de classificação na qual K é o parâmetro de kernel, utilizada no estudo [50 - 51].

$$f(x) = \sum_{i=1}^N \alpha_i y_i K(x, x_i) + b \quad (27)$$

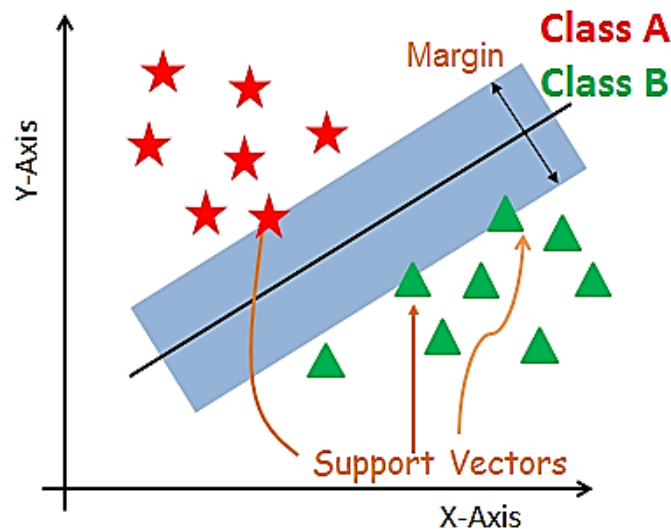
O kernel é uma maneira de calcular o produto escalar de dois vetores x e y em algum espaço de características (de dimensão muito alta), e é por isso que as funções do kernel são às vezes chamadas de “produto escalar generalizado” [52].

Os classificadores de SVM são ferramentas de aprendizado de máquina criadas para prever a classe ou categoria à qual um determinado objeto pertence como uma função de um vetor de recurso n-dimensional. São construídos ajustando os parâmetros da equação (27) para obter uma classificação ótima de uma série de vetores de características de classes correspondentes. O objetivo da fase de treinamento (ajuste) é encontrar um hiperplano no espaço de recursos que separa as classes de destino, o que fornece ao modelo SVM a capacidade de generalizar e prever a classe de amostras fora de treinamento. Obviamente, se o conjunto de treinamento não contém amostras representativas com informações relevantes, nenhum modelo empírico pode ser construído [51].

O SMV é utilizado para classificar os pontos em um estilo segmentado, e parece ser perfeitamente adequada para pequenos conjuntos de dados com alto número de dimensões. Em particular, as análises baseadas em SVM foram aplicadas com sucesso a diferentes dados biológicos em áreas como ciência ambiental, neurociência, bioinformática e diagnóstico médico entre outras [51 - 53].

O funcionamento do algoritmo pode ser compreendido por meio da Figura 7, na qual é possível notar dois conjuntos separados pela reta e mostrados pelas cores vermelho e verde.

**Figura 7 - Conjuntos Separáveis Linearmente**



FONTE: [54]

As etapas de desenvolvimento do algoritmo podem ser explicadas considerando a Figura 7, conforme descrito a seguir:

- 1) Cria-se um hiperplano de separação dos elementos (reta) [55];
- 2) Os elementos mais próximos do hiperplano são denominados Vetores de Suporte [55];
- 3) A margem do hiperplano de separação é a distância mínima entre os dois conjuntos. O limite de decisão mais ideal é aquele que tem margem máxima dos pontos mais próximos de todas as classes. Os pontos mais próximos do limite de decisão que maximizam a distância entre o limite de decisão e os pontos são chamados de vetores de suporte. O limite de decisão no caso de máquinas de vetores de suporte é chamado classificador de margem máxima ou plano de máxima margem [55 - 56];

Estes entendimentos, implicam que somente alguns elementos são importantes para o aprendizado da máquina, sendo estes os Vetores de Suporte. Os demais itens,

apresentam peso zero na determinação dos vetores de suporte, e, portanto, são desconsiderados [55].

Um dos principais problemas dos métodos de aprendizagem SVM é incorrer em *overfitting*, ou seja, mesmo que todos os vetores de recursos no conjunto de treinamento possam ser bem classificados, o algoritmo Máquina de Vetor de Suporte pode perder sua capacidade de generalizar fora do conjunto de treinamento. Por um lado, os classificadores de SVM têm a capacidade de aprender limites de decisão complicados quando os dados apresentam distribuições complexas [51].

SVM difere dos outros algoritmos de classificação na maneira que escolhe o limite de decisão que maximiza a distância dos pontos de dados mais próximos de todas as classes. Um SVM não apenas encontra um limite de decisão; ele encontra o limite de decisão mais próximo do ideal.

A Tabela 6 elenca as principais vantagens e desvantagens do algoritmo SVM, algumas delas expostas anteriormente.

**Tabela 6 - Vantagens e desvantagens do algoritmo SVM**

<p><b>Vantagens</b></p>	<ul style="list-style-type: none"> <li>• Efetividade com espaços de alta dimensão</li> <li>• Precisão quando o número de dimensões é maior que o número de amostras</li> <li>• Eficiência em questão de memória</li> <li>• Versátil (funções de definição do Kernel)</li> <li>• Geração do plano de maneira iterativa, buscando minimizar o erro</li> </ul>
<p><b>Desvantagens</b></p>	<ul style="list-style-type: none"> <li>• Problemas com ajustes da função Kernel quando o número de <i>features</i><sup>1</sup> é maior que o de amostras</li> <li>• Não fornece estimativa de probabilidade, é calculada utilizando <i>five-fold validation</i><sup>2</sup></li> <li>• <i>Overfitting</i><sup>3</sup></li> </ul>

<sup>1</sup> *Features*: são os atributos levantados com base no cálculo dos índices de potência e corrente. São os atributos que caracterizam uma classe dentro de um banco de dados.

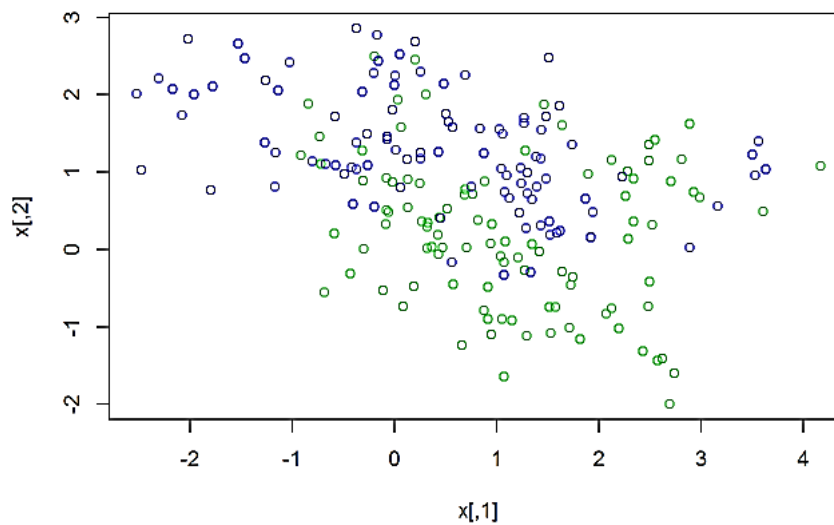
<sup>2</sup> *Five-fold Validation*: Teste de validação do algoritmo por meio de validação cruzada.

<sup>3</sup> *Overfitting*: quando os algoritmos de classificação se tornam muito ajustados ao banco de dados. Ao sofrer *overfitting*, o algoritmo perde a capacidade de generalizar a classificação por se especializar demais nos dados do banco de dados.

Muitas vezes, os separadores lineares e os limites falham devido às interações não lineares nos dados e à dependência não linear entre os recursos no espaço de recursos. A máquina de vetor de suporte radial do kernel é uma boa abordagem quando os dados não são linearmente separáveis [57].

A Figura 8 abaixo, mostra um exemplo de conjunto de dados que não podem ser separados por meio de um hiperplano reto.

**Figura 8 - Conjunto de dados que não são Linearmente Separáveis**



FONTE: [57]

Pode-se notar que para este caso, os dados não são linearmente separados com boa precisão, pois há uma mistura entre os dados verdes e azuis. Para que eles possam ser separados corretamente, é necessário que sejam reorganizados e separados por um hiperplano polinomial.

O kernel RBF (Função de Base Radial) é outro método Kernel popular cujo valor depende da distância da origem ou de algum ponto. Também chamado de kernel gaussiano, esta função pode ser definida por (28). [52]

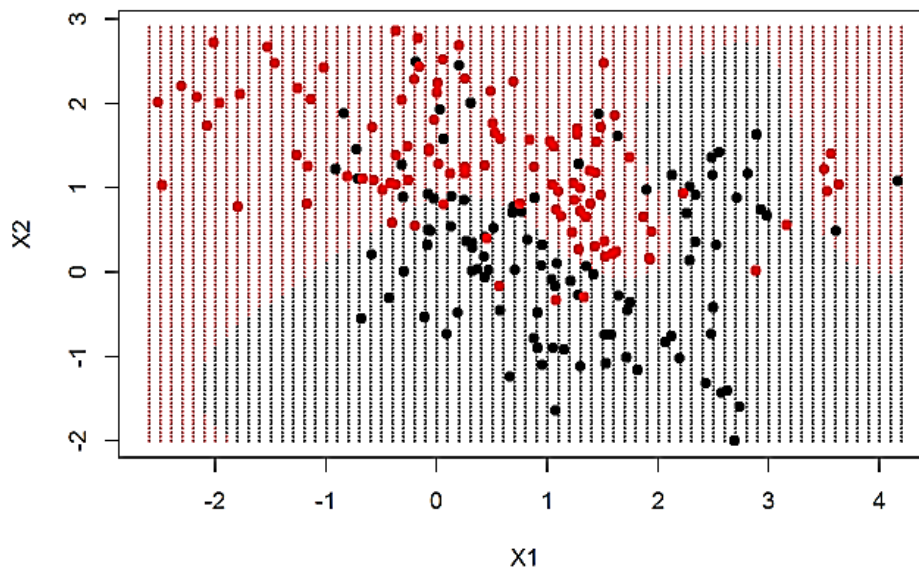
$$K(x, x_i) = \text{exponent}(-\gamma \|x - x_i\|^2) \quad (28)$$



Onde:  $\gamma$  é um hiperparâmetro ou um parâmetro de ajuste que explica a suavidade do limite de decisão e controla a variância do modelo. Desta forma altos valores de  $\gamma$ , acarretam limites de decisão flutuantes e ondulados, o que explica a alta variância e o *overfitting*. Por sua vez, se  $\gamma$  for pequeno, a linha de decisão ou limite é mais suave e tem baixa variância [57].

Utilizando-se de uma separação a partir da criação de um hiperplano baseada no kernel radial, os mesmos dados mostrados na Figura 8 podem ser separados considerando a Figura 9.

**Figura 9 - Separação não linear (Kernel Radial)**



FONTE: [57]

É possível notar que há um hiperplano criado para separar as regiões destacadas em “preto” e em “vermelho”. Todavia, mesmo conferindo maior adaptação na separação dos dados, ainda existe uma porção pequena de dados misturados, ou seja, ainda há dados “pretos” na região “vermelha” e vice-versa.

#### 2.4.2.2 k-enésimo vizinho mais próximo

O algoritmo *k-Nearest Neighbor* (k-NN) é um método não paramétrico usado para classificação. A entrada consiste nos k exemplos de treinamento mais próximos no espaço de recurso. A saída na classificação k-NN é uma associação de classe. A classificação é dada pelo rótulo existente pertencente à maior quantidade de vizinhos mais próximos [58].

O k-enésimo vizinho mais próximo é um algoritmo de classificação de padrões que não possui fase de treinamento. É baseado em analisar elementos próximos de um ponto de partida e classificados por uma função heurística [12].

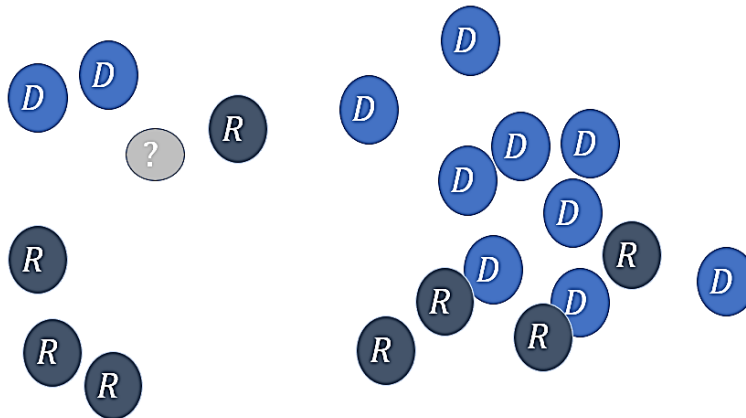
O k-NN é encontrado dentro de alguns algoritmos de aprendizado supervisionados. Informalmente, isso significa que o pacote de dados rotulados é considerado como entrada, tendo em mente as observações de treinamento  $(x, y)$ , com o objetivo de capturar a relação entre  $x$  e  $y$ , onde  $x$  são dados de entrada e  $y$  os rótulos, inclinando a função heurística:  $X \rightarrow Y$  para que, dada uma observação invisível  $x$ ,  $h(x)$  seja capaz de prever com precisão a saída correspondente  $y$ .

A principal intenção do algoritmo é classificar um determinado dado com base em seus vizinhos mais próximos. Essa classificação é feita por semelhança que é definida com base na métrica de distância entre dois pontos de dados, normalmente utilizada a distância euclidiana dada pela equação (29), mas outras medidas podem ser mais adequadas a depender do cenário, por exemplo: a distância de Manhattan, Chebyshev e Hamming.

$$d(x, x') = \sqrt{((x_1) - (x'_1))^2 + ((x_2) - (x'_2))^2 + \dots + ((x_n) - (x'_n))^2} \quad (29)$$

A Figura 10, mostra um conjunto que deverá ser classificado a partir do algoritmo do k-enésimo vizinho mais próximo. Na sequência, tomando o conjunto apresentado na Figura 10, o método é descrito de forma resumida, conforme [59].

**Figura 10 - Exemplo de conjunto a ser classificado por algoritmo k-ésimo vizinho mais próximo.**



FONTE: ADAPTADO DE: [59]

Com relação a metodologia envolvida, inicialmente define-se um ponto de partida, definido como sendo o sinal de interrogação (?) [59].

Examinam-se os K vizinhos mais próximos deste ponto, e os classificam. Na imagem, para  $K=1$ , observe que o item em destaque é o único vizinho mais próximo. À medida que K aumenta, para 3 por exemplo, há vizinhos próximos classificados como D, além do R em destaque [59].

Por meio da função heurística, se determina a classe que prevalece dentre os vizinhos. A função heurística determina uma regra para a classificação dos vizinhos, e segue sempre pelo maior número dentre todos os vizinhos identificados [59].

O classificador k-NN atribui a entrada à classe com mais exemplos entre os k vizinhos da entrada. Todos os vizinhos têm voto igual e a classe que tem o número máximo de representantes entre os k vizinhos é escolhida. As convergências são arbitrariamente divididas ou uma pontuação ponderada é adotada. Assim, k geralmente é considerado um número ímpar para minimizar os empates nas condições em que: existe uma confusão geralmente entre duas classes vizinhas [60].

Além disso, é considerado também um algoritmo de aprendizado não paramétrico e baseado em instâncias. Como não paramétrico, entende-se que não são feitas suposições explícitas. A aprendizagem baseada em instâncias significa que o algoritmo não aprende explicitamente um modelo, mas memoriza as instâncias de treinamento que

são posteriormente usadas como "conhecimento" para a fase de previsão. Concretamente, isso significa que somente quando uma consulta é feita ao banco de dados, o algoritmo usará as instâncias de treinamento para obter uma resposta.

No processo, é importante considerar apenas os vizinhos para os quais uma classificação correta já exista, isto é, conjunto previamente treinado. Todos os objetos são considerados presentes no espaço de recurso multidimensional e são representados por vetores de posição os quais representam a distância entre o objeto e seus vizinhos. O espaço multidimensional é dividido em regiões utilizando os locais e rótulos dos dados de treinamento. Um objeto neste espaço será classificado com a classe que possuir dentre os k vizinhos mais próximos o maior número [58].

Uma das dificuldades mais importantes encontradas nesta técnica é a velocidade de classificação. Especificamente, o tempo que o classificador leva para encontrar os vizinhos mais próximos na tarefa de consulta. Um algoritmo de busca K-ésimo vizinho mais próximo realiza diversas iterações através de todo o conjunto de dados, medindo a distância entre o ponto da consulta e todos os demais presentes no conjunto de dados para, então determinar os vizinhos mais próximos adequadamente. Este fato, pode tornar inviável o uso desta técnica quando os conjuntos dos dados for muito grande. Por outro lado, há simplicidade na elaboração do algoritmo facilitando a implementação. Além desta vantagem, o k-ésimo vizinho mais próximo apresenta elevada precisão na classificação [58, 61].

A Tabela 7 elenca de forma resumida as principais vantagens e desvantagens discutidas sobre o algoritmo k-NN.

**Tabela 7 - Vantagens e desvantagens do algoritmo k-NN**

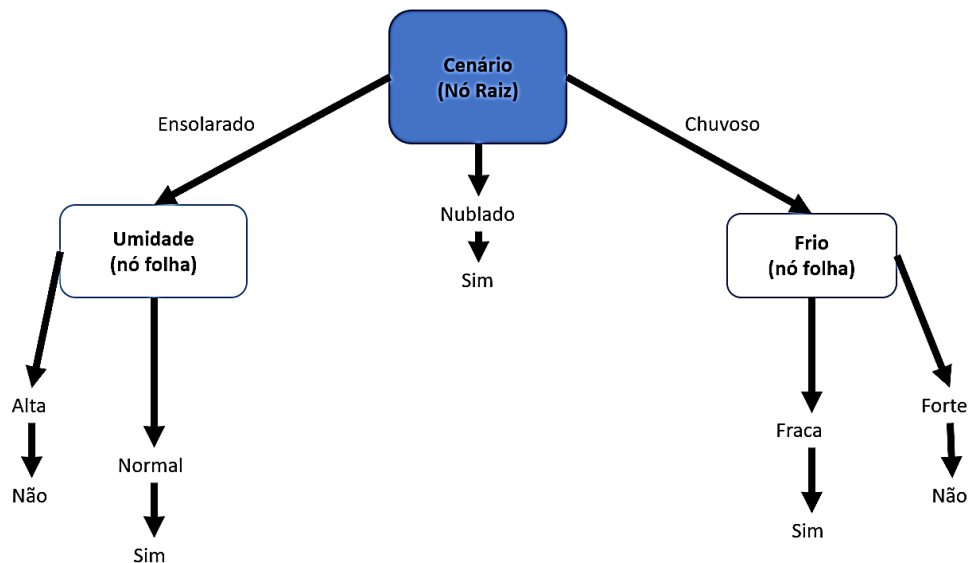
<b>Vantagens</b>	<ul style="list-style-type: none"> <li>• Simplicidade Computacional</li> <li>• Alta precisão nos resultados</li> <li>• Não requer pré-processamento das amostras</li> </ul>
<b>Desvantagens</b>	<ul style="list-style-type: none"> <li>• Não distinção dos vetores quanto à sua importância</li> <li>• Não atribuição de pesos para o vetor de classificação</li> <li>• Velocidade</li> <li>• Alta dimensionalidade</li> </ul>

### 2.4.2.3. Decision Tree (Árvore de Decisão)

A árvore de decisão (Decision Tree) é uma das ferramentas mais populares utilizada para classificação e predição. Uma árvore de decisão é um fluxograma com estrutura de árvore, em que cada nó interno denota um teste em um atributo, cada ramo representa um resultado do teste e cada nó folha (nó terminal) contém um rótulo de classe [62].

Uma árvore de decisão normalmente começa com um único nó, que se ramifica em possíveis resultados. O algoritmo aprende dividindo o banco de dados em subconjuntos com base em um teste de valor de atributo. Esse processo é repetido em cada subconjunto derivado de uma maneira recursiva chamada particionamento recursivo. A recursão é concluída quando o subconjunto em um nó tem o mesmo valor da variável de destino ou quando a divisão não adiciona mais valor às predições. Uma instância é classificada iniciando no nó raiz da árvore, testando o atributo especificado por este nó, depois descendo o galho da árvore correspondente ao valor do atributo conforme mostrado na Figura 11. Este processo é então repetido para a subárvore enraizada no novo nó [62 - 63].

**Figura 11 - Exemplo de Árvore de Decisão**



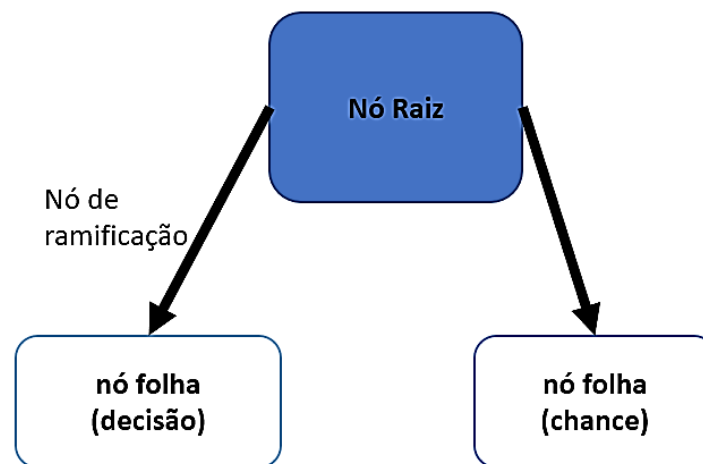
FONTE: ADAPTADO DE: [62][64]

A árvore de decisão na Figura 11 classifica uma determinada manhã de acordo com se é adequada para jogar tênis e retorna à classificação associada à determinada folha (neste caso Sim ou Não) [62].

Conforme [84], as árvores de decisão geralmente consistem em três elementos diferentes:

- **Nó raiz:** O nó de nível superior representa o objetivo final ou a grande decisão que você está tentando tomar.
- **Galhos:** As ramificações, que se originam da raiz, representam diferentes opções - ou cursos de ação - disponíveis ao tomar uma decisão específica. Estes são mais comumente indicados com uma linha de seta e geralmente incluem custos associados, bem como a probabilidade de ocorrer.
- **Nó da folha:** Os nós folha - que são anexados no final dos ramos - representam resultados possíveis para cada ação. Normalmente, existem dois tipos de nós folha. Os primeiros denominados nós folha quadrada, que indicam outra decisão a ser tomada, e nós folha circular, que indicam um evento aleatório ou resultado desconhecido.

**Figura 12 - Componentes da Árvore de Decisão**



FONTE: ADAPTADO DE: [64][ 65]

A construção do classificador de árvore de decisão não requer nenhum conhecimento de domínio ou configuração de parâmetros e, portanto, é apropriado para

a descoberta de conhecimento exploratório. As árvores de decisão podem lidar com dados de alta dimensão. Em geral, o classificador de árvore de decisão tem boa precisão. A indução da árvore de decisão é uma abordagem indutiva típica para aprender o conhecimento sobre classificação [62].

O resumo das principais vantagens e desvantagens discutidas sobre o algoritmo de Árvore de Decisão é apresentado na Tabela 8.

**Tabela 8 - Vantagens e desvantagens do algoritmo Árvore de Decisão**

<b>Vantagens</b>	<ul style="list-style-type: none"> <li>• Simplicidade Computacional para execução</li> <li>• Capazes de gerar regras compreensíveis</li> <li>• Não requer configuração de parâmetros</li> <li>• Lidam com variáveis contínuas e categóricas</li> </ul>
<b>Desvantagens</b>	<ul style="list-style-type: none"> <li>• São propensas a erros em problemas de classificação com muitas classes e um número relativamente pequeno de exemplos de treinamento.</li> <li>• O processo de crescimento de uma árvore de decisão é computacionalmente caro</li> </ul>

#### 2.4.2.4. Random Forest (RF)

O *Random Forest* é um algoritmo de aprendizado de máquinas proposto por Breinman (2001) que possui a capacidade de combinar uma floresta de classificadores de árvore de decisão em várias subamostras do conjunto de dados, com subconjuntos aleatórios dos recursos para divisões de nós. A utilização de múltiplas árvores elimina o risco na previsão de classificações se comparado a utilização de apenas uma árvore de decisão e diminui a ocorrência de *overfitting* [66 - 67 - 68].

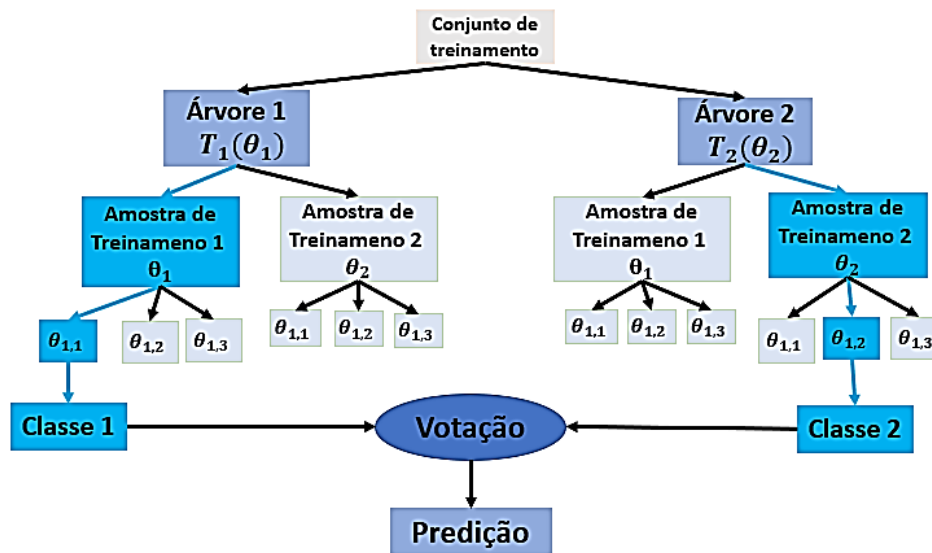
Este algoritmo emprega um método aprimorado de *bootstrapping* e *boosting*. O termo *Bagging* é a abreviação de agregação *bootstrap* foi apresentado por Breiman como uma ferramenta para reduzir os erros de previsão em algoritmos de máquina de aprendizagem, evitando assim o *overfitting*. Este processo consiste em treinar cada árvore com um conjunto *bootstrap* das amostras de treinamento com reposição, geradas independentemente das amostras anteriores. Em contraste, *boosting* usa retreinamento

iterativo, em que amostras classificadas recebem peso aumentado conforme o progresso das iterações [69].

Uma floresta aleatória é um classificador que consiste no conjunto de árvores estruturadas com classificadores  $\{T_1(\theta_1), T_2(\theta_2), \dots, T_B(\theta_B)\}$  onde os  $T_i$  representa cada árvore e  $\theta_i$  são as amostras bootstrap [66 - 70].

Basicamente parte-se do princípio de funcionamento de uma árvore de decisão simples, na qual as sementes são ramificadas e, na busca pela resposta, atinge-se a folha. Porém para o *Random Forest*, cada árvore da floresta é construída e testada independentemente de outras árvores, assim, cada aprendizado e teste pode ser feito de maneira paralela. Durante o treinamento cada nó de decisão da árvore cria um conjunto de testes aleatórios e, em seguida, seleciona o melhor entre eles de acordo com alguma medição de qualidade pré-definida. Para este fim, o índice de Gini é um parâmetro muito utilizado. As árvores são ramificadas até seu tamanho real sem poda e ao atingir a folha de cada árvore, o algoritmo deve avaliar o voto majoritário das mesmas, para então, apresentar a classificação final, como pode ser visto na Figura 13. [71 - 72]

**Figura 13 - Exemplo de conjunto que será classificado pelo algoritmo RF**



FONTE: ADAPTADO DE: [64]

Pode-se dizer que o algoritmo *Random Forest* é uma boa escolha quando é necessário lidar com dados de alta dimensão. O algoritmo usa um grande número de



árvores em conjunto que somado ao fato de existir a seleção aleatória de variáveis para uma divisão minimiza a correlação entre as árvores no conjunto, resultando em maior precisão na classificação.

Como o algoritmo trabalha com amostras aleatórias, é considerado estável e robusto, não sendo afetado pela adição de novos dados para serem classificados. Além disso, *Random Forest* é considerado um algoritmo preciso e de fácil desenvolvimento.

Um fato de desvantagem estabelecida na literatura é a demanda de esforço para treinamento do algoritmo, que se comparado a outros algoritmos classificadores, como por exemplo k-NN, levam mais tempo para se ajustar.

### 2.4.3 Técnicas de avaliação do algoritmo

Em todo processo de classificação os resultados possíveis são: falsos positivos, falsos negativos, verdadeiros positivos e verdadeiros negativos. Portanto, é possível que existam dados classificados de forma errônea, isto é, foram incluídos em uma das classes presentes no *dataset* mas realmente não pertencem às mesmas (Falso positivos), ou então, foram excluídos de uma classe, mas pertencem à mesma (Falsos Negativos). Os resultados de classificação podem ser representados por meio de uma matriz denominada matriz de confusão, conforme exemplo na Tabela 9.

**Tabela 9 - Matriz de confusão**

	Predição	
	Chuveiro	Televisão
Realidade	Chuveiro	40
	Televisão	30
	Chuveiro	15
	Televisão	15

Por meio da matriz de confusão, pode-se inferir a performance do algoritmo, uma vez que é possível verificar quantos dados foram corretamente classificados. Como exemplo, de 100 dados, 40 eram chuveiros e foram classificados como chuveiros (Verdadeiro Positivo). Porém, 30 eram chuveiros, mas foram classificados como televisão

(Falso negativo). Desta forma, a matriz é capaz de quantificar também os erros de classificação existentes [76].

A classificação possui alguns indicadores para determinar a qualidade do algoritmo empregado com a finalidade de classificar os dados de entrada, são eles: acurácia, precisão, recall e F1-Score.

As fórmulas matemáticas utilizadas para calcular estes indicadores (legenda: TP - Verdadeiro Positivo, TN - Verdadeira Negativo, FN - Falso Negativo e FP - Falso Positivo) são apresentadas na Tabela 10.

**Tabela 10 - Fórmulas matemáticas para cálculo dos indicadores de avaliação do algoritmo**

Indicador	Fórmula
Acurácia	$\frac{TP + TN}{TP + TN + FP + FN}$
Precisão	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F1-Score	$2 * \frac{Precisão * Recall}{Precisão + Recall}$

Acurácia refere-se a porcentagem do total de amostras que foram classificadas corretamente. Enquanto a definição de precisão refere-se a probabilidade de que todas as amostras previstas como positivas sejam realmente positivas. O recall refere-se a probabilidade de que uma amostra realmente positiva seja considerada uma amostra positiva. F1-Score é a média harmônica ponderada de precisão e recall. Os valores de precisão e recall geralmente não são consistentes. Assim, quando um é mais alto, o outro geralmente é mais baixo. O método mais comum para combinar os dois índices é usar o indicador denominado F1-Score [48 - 73].

## CAPÍTULO 3 - METODOLOGIA

### 3.1. Concepção Do Trabalho Desenvolvido

Conforme explicações dos capítulos anteriores, o trabalho consiste em levantar um banco de dados da desagregação de cargas elétricas comprovando sua confiabilidade para que uma ferramenta de classificação de cargas possa ser adicionada ao medidor de energia elétrica inteligente. Desta forma, neste capítulo são descritos os passos desenvolvidos para definição e levantamento de atributos para uso na classificação de cargas elétricas, conforme procedimento evidenciado no fluxograma da Figura 14.

**Figura 14 - Fluxograma do desenvolvimento do trabalho**

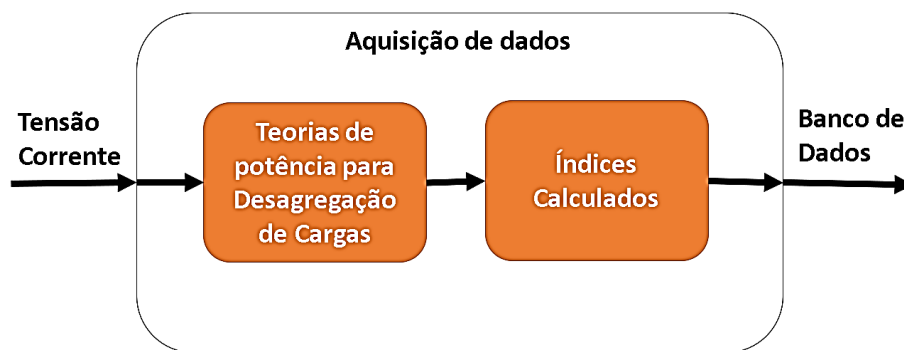


FONTE: AUTORIA PRÓPRIA [59]

### 3.1.1. Panorama geral das simulações

Conforme apresentado, neste momento, desenvolve-se a etapa 1 relacionada ao primeiro bloco de atividades do medidor de energia elétrica inteligente, denominado como “Aquisição de dados”. A Figura 15 mostra o diagrama de blocos expandido relacionado com a etapa 1.

**Figura 15 -Etapa 1: Aquisição de dados**



FONTE: AUTORIA PRÓPRIA

A tensão e corrente verificadas na Figura 15, são coletadas a partir de cargas elétricas em funcionamento. Para representar esta etapa, foram utilizados banco de dados de alta confiabilidade de cargas elétricas em operação chamados PLAID, disponíveis para acesso em [74 - 75]. Estes bancos de dados possuem pontos de tensão e corrente no decorrer do tempo de equipamentos residenciais em funcionamento. Nas simulações realizadas nesta etapa do trabalho, não foram testadas todas as cargas presentes nos bancos de dados, uma vez que o foco é a geração de atributos para a criação do banco de dados que poderá ser aprimorado no futuro.

Para as simulações foram escolhidas algumas das cargas presentes nos bancos de dados PLAID, especificamente: lâmpada fluorescente compacta, geladeira, secador de cabelos, micro-ondas, ar-condicionado, notebook, lâmpada incandescente, ventilador, máquina de lavar roupas e aquecedor, compreendendo um total de 10 cargas elétricas.

A Tabela 11, mostra o panorama de todas as simulações realizadas.

**Tabela 11 - Simulações realizadas**

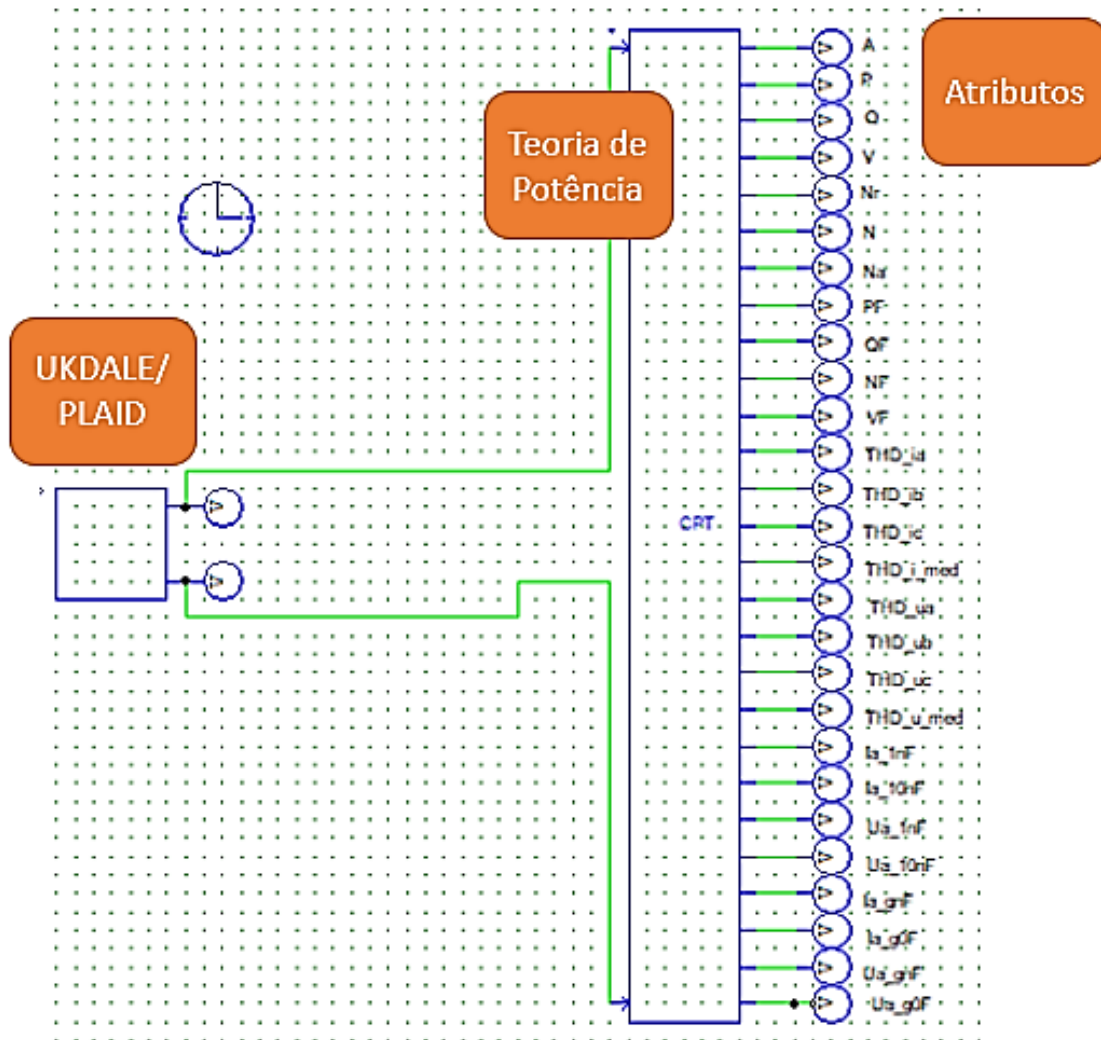
<b>SIMULAÇÃO</b>	<b>EQUIPAMENTO</b>	<b>CPT</b>	<b>IEEE</b>	<b>CPC</b>
1	Lâmpada fluorescente compacta	X	X	X
2	Geladeira	X	X	X
3	Secador de cabelos	X	X	X
4	Micro-ondas	X	X	X
5	Ar-condicionado	X	X	X
6	Notebook	X	X	X
7	Lâmpada incandescente	X	X	X
8	Ventilador	X	X	X
9	Máquina de lavar roupas	X	X	X
10	Aquecedor	X	X	X

Dentro da literatura levantada, percebe-se que muitos desenvolvimentos se baseiam em cargas elétricas específicas para análise das teorias de potência utilizadas para a desagregação de cargas, conforme mencionado em [34], sem que a metodologia proposta seja expandida para uma gama maior de comportamentos elétricos a serem analisados. Buscando o levantamento de banco de dados, utiliza-se uma diversificação de cargas elétricas simuladas, bem como, são verificados os parâmetros de três diferentes teorias de potência.

Primeiramente, as cargas foram simuladas e verificadas as oscilografias das cargas elétricas, que são representações gráficas de seu comportamento adquiridas com taxa de amostragem em alta frequência. Pode-se por meio delas, verificar se as cargas possuem um comportamento indutivo, ou capacitivo, por exemplo, para posteriormente, verificar a validade dos índices calculados.

A Figura 16 mostra os blocos (recurso do PSIM disponível para programações em linguagem C) desenvolvidos para o cálculo dos índices provenientes das Teorias de Desagregação de Cargas Elétricas (CPC, CPT e IEEE 1459) com o auxílio do *software* PSIM. Esta é a etapa definida como “Extração de atributos”. Simulações computacionais foram efetuadas com a finalidade de validação dos cálculos providos pelos algoritmos desenvolvidos e/ou aplicados. A Tabela 12, mostra os parâmetros que foram calculados por cada uma das teorias de potência avaliadas neste trabalho.

**Figura 16 - Estrutura de simulação para Aquisição de Dados**



FONTE: AUTORIA PRÓPRIA, DESENVOLVIDO NO SOFTWARE PSIM

Na Figura 16, tem-se o bloco C inicial representando a carga elétrica em funcionamento, responsável por acessar os arquivos .csv coletando as informações de nuvem de pontos de tensão e corrente de cada carga escolhida para simulação, dados que estão diretamente ligados às saídas do bloco. Após, o bloco C destina-se a cada teoria de desagregação de cargas, neste caso CPT, recebe as informações de tensão e corrente e realiza os cálculos necessários para apresentar em suas saídas os índices conforme disposto na Tabela 12.

**Tabela 12 - Índices Calculados, Levantamento dos Atributos**

<b>CPT</b>	
<b>SAIDAS</b>	<b>NOME</b>
A	Potência Aparente
P	Potência Ativa
Q	Potência Reativa
V	Potência Nula
PF	Fator de Potência
QF	Fator de Reatividade
VF	Fator de Distorção
I_efe	Corrente inativa
I.v_RMS	Corrente eficaz
U.v_RMS	Tensão Eficaz
Nr	Potência de Desbalanço Reativa
N	Potência de Desbalanço
Na	Potência de Desbalanço Ativa
NF	Fator de Desbalanço
THD_ia	Distorção Harmônica
THD_ib	Distorção Harmônica
THD_ic	Distorção Harmônica
THD_i_med	Distorção Harmônica
THD_ua	Distorção Harmônica
THD_ub	Distorção Harmônica
THD_uc	Distorção Harmônica
THD_u_med	Distorção Harmônica

<b>CPC</b>	
<b>SAIDAS</b>	<b>NOME</b>
iacR, iacS, iacT	Corrente Ativa fase R,S,T
iscR, iscS, iscT	Corrente Dissipada fase R, S, T
ircR, ircS, ircT	Corrente Reativa fase R, S, T
iucR, iucS, iucT	Corrente de Desequilíbrio fase R,S,T
igR, igS, igT	Corrente Gerada fase R,S,T
Re_Sc	Parte real da Potência Aparente Gerada pela carga
Im_Sc	Parte Imaginária da Potência Aparente Gerada pela carga
S	Potência Aparente
Dsc	Potência Dissipada
Duc	Potência de Desequilíbrio
Dus	Potência Gerada
pf	Fator de potência

<b>IEEE 1459</b>	
<b>SAIDAS</b>	<b>NOME</b>
FP1p	Fator de Potência Fundamental Sequência Positiva
ia1	Corrente Fase A
Ie	Corrente Efetiva
P	Potência Ativa
PF1	Fator de Potência Fundamental
PFe	Fator de Potência Trifásico
Q1p	Potência Reativa
Se	Potência Aparente Efetiva
THDia	Distorção Harmônica
va1	Tensão Fase A
Ve	Tensão Efetiva

Ao final desta etapa, um banco de dados inicial contendo 55 atributos foi formulado para o desenvolvimento das demais etapas existentes no fluxograma da Figura 14. Destaca-se que os atributos levantados podem ser aplicados às cargas trifásicas, porém, no estudo em questão, são somente discutidas e analisadas cargas monofásicas e por este motivo, algumas variáveis são futuramente descartadas.

### 3.1.2. Panorama geral da definição dos atributos

Em [7] é apresentado um método para definição de atributos para classificação de cargas elétricas, em que as teorias de desagregação de cargas apresentadas, torna possível uma única representação de tensão e corrente em um ambiente não senoidal. Ambas utilizam do conceito da Transformada de Fourier para extrair os harmônicos que caracterizam as formas de onda e permitem o cálculo de índices de potência que, por sua vez, serão utilizados como dados para classificação de cargas elétricas.

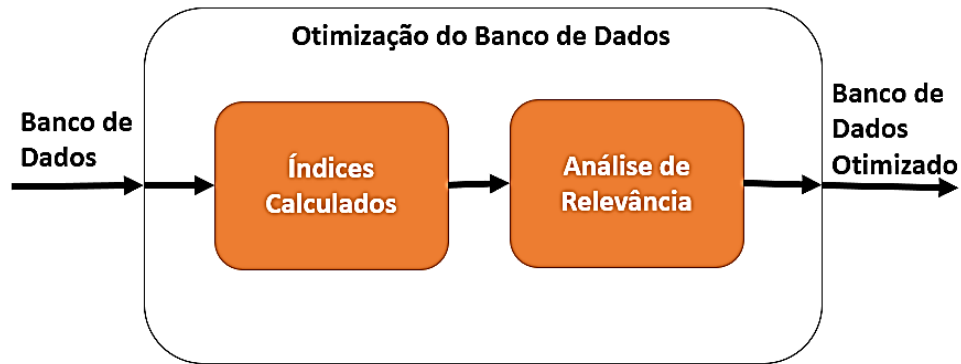
Neste contexto, este trabalho se baseia no mesmo procedimento e conceito, todavia, utilizando-se de cargas e teorias de potência distintas.

Empregando as 10 cargas elétricas escolhidas e três teorias de potência para desagregação de cargas, previamente comentadas, a definição de atributos para posterior classificação é realizada com base no cálculo dos índices de potência determinados pelas teorias de potência fazendo uso das oscilografias simuladas no PSIM.

A Figura 17, representa a etapa de Otimização do Banco de Dados, responsável por analisar os dados levantados que compõe o banco de dados atual e definir os atributos que serão utilizados para classificação das cargas elétricas.



**Figura 17 -Etapa 2: Otimização do Banco de Dados**



FONTE: AUTORIA PRÓPRIA

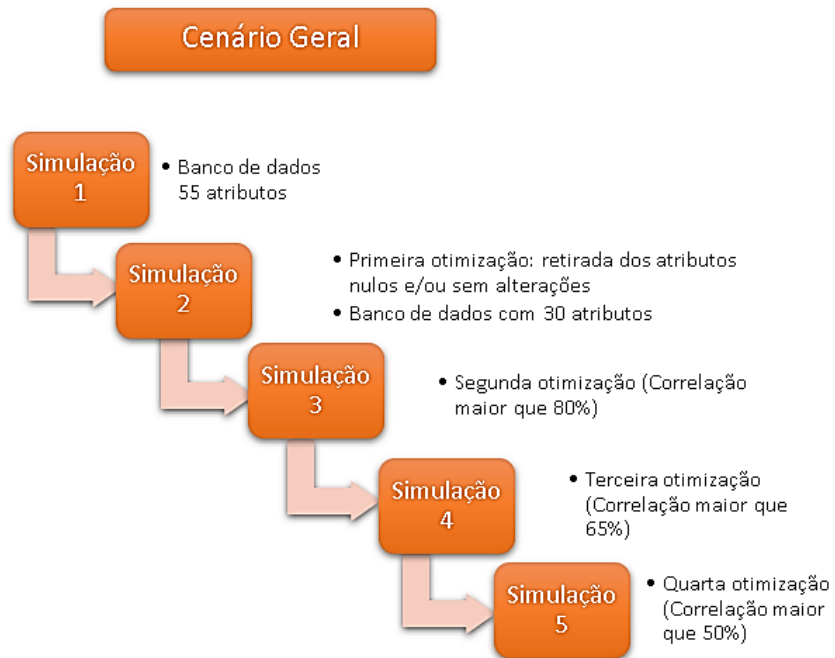
Observa-se que todos os índices calculados para as três teorias, fazem parte dos atributos levantados nas simulações, porém, para uma futura classificação de cargas de forma mais eficiente, faz-se uma análise de relevância para as cargas utilizadas. Sendo assim, espera-se que o banco de dados final proposto neste trabalho, seja enxuto, conferindo aos demais passos, confiabilidade e redução de processamento. Além desta otimização, é proposta também uma análise da classificação considerando os atributos levantados a partir das teorias de potência analisados individualmente.

Conforme mostrado na Tabela 12, no cenário geral, inicialmente existe um total de 55 atributos. O desvio padrão para cada um deles foi utilizado como grandeza de definição de relevância, uma vez que é responsável por representar a dispersão entre os dados. Sendo assim, quanto maior o desvio padrão entre as amostras calculadas, espera-se que maior seja o potencial em diferenciar cargas elétricas, e, portanto, poderá definir os atributos de maior relevância que constituirão o banco de dados.

Como verificação final da relevância dos atributos levantados, faz-se a matriz de correlação entre as variáveis levantadas. De forma geral, a matriz de correlação determina o grau pelo qual um par de variáveis está linearmente relacionado. Neste sentido, uma segunda análise para a relevância das variáveis foi realizada com a intenção de aumentar a confiabilidade nos dados que compõe o banco de dados otimizado que será utilizado para o desenvolvimento das etapas futuras.

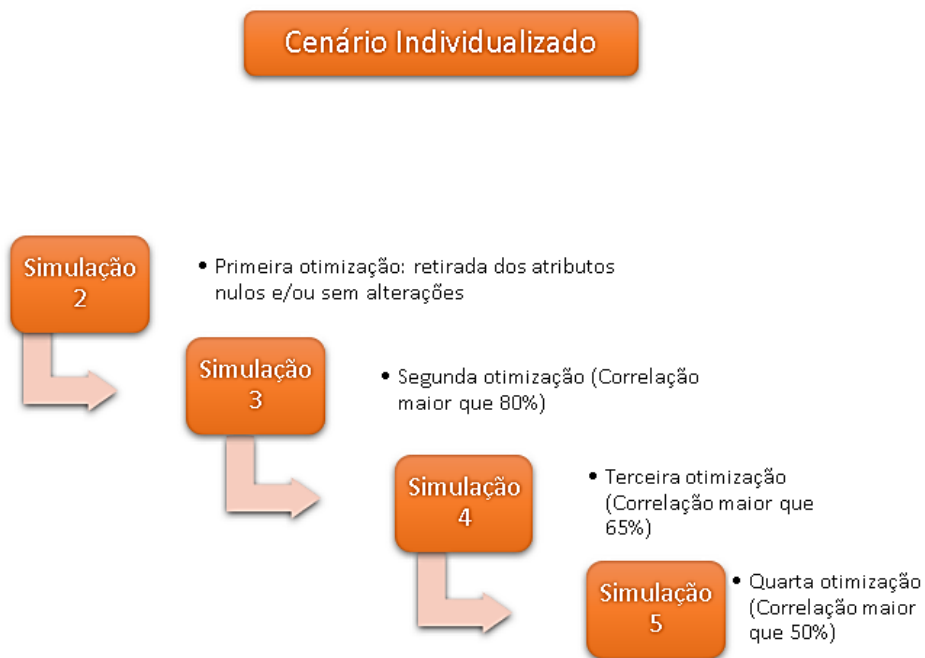
Desta forma foram realizados os procedimentos de teste e análise expostos no fluxograma das Figuras 18 e 19:

**Figura 18 - Fluxograma de Simulações Cenário Geral**



FONTE: AUTORIA PRÓPRIA

**Figura 19 - Fluxograma de Simulações Cenário Individualizado**



FONTE: AUTORIA PRÓPRIA

É possível notar a partir do fluxograma que no cenário geral, têm-se:

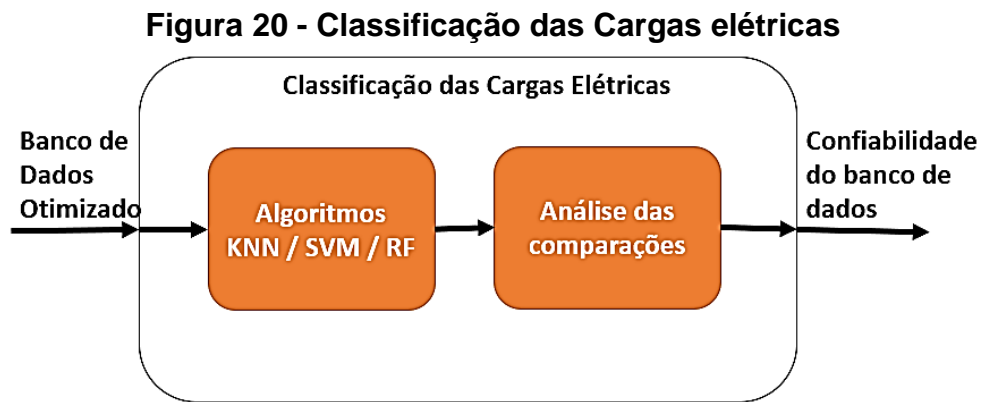
- 1) Com o banco de dados inicial (com 55 atributos) foi realizada a primeira simulação para verificar a performance de classificação dos algoritmos de inteligência artificial
- 2) Primeira otimização do banco de dados: Retirado os atributos nulos e/ou sem variação conforme a carga elétrica simulada. Nesta etapa, tem-se um banco de dados composto de 30 atributos. Uma nova simulação de performance de classificação é realizada.
- 3) Segunda otimização: retirados atributos com correlação maior ou igual a 80%.
- 4) Terceira otimização: retirados atributos com correlação maior ou igual a 65%.
- 5) Quarta otimização: retirados atributos com correlação maior ou igual a 50%.

Em todas as etapas descritas, para verificação da performance dos algoritmos, foi feita uma *SearchGrid* para definição dos melhores parâmetros de simulação dos mesmos e, com isso, otimizar a execução do algoritmo. Para o algoritmo k-NN, definiu-se os valores de K para cada caso, para o SVM linear e RBF foram definidos margem (C) e *gamma*, assim como para Árvore de Decisão foi definido o critério de ramificação e a profundidade da árvore e para RF foram definidos o número de árvores e o estado randômico.

Como análise adicional, um cenário individualizado é apresentado e estão previstos os mesmos passos de simulações comentados, porém, verificando individualmente cada teoria de potência, ou seja, o banco de dados geral, composto por índices de potência e corrente provenientes da CPT, CPC e IEEE foi, neste momento, dividido em 3 bancos de dados e todo o procedimento de verificação da relevância dos atributos, otimização do banco de dados, e simulações de performance foram realizadas individualmente.

### 3.1.3. Validação do banco de dados por meio da classificação das cargas elétricas

Para garantir que o banco de dados levantado seja confiável, os algoritmos determinados para classificar as cargas são aplicados seguindo procedimento apresentado no fluxograma da Figura 18 e 19, e as definições tomadas se baseiam na lógica apresentada na Figura 20.



FONTE: AUTORIA PRÓPRIA

Na definição inicial do banco de dados otimizado, uma primeira avaliação de precisão dos algoritmos de inteligência artificial foi realizada visando comprovar a confiabilidade da otimização. A comparação entre a precisão das classificações feitas entre o banco de dados inicial e o banco de dados otimizado confere confiabilidade na otimização do banco de dados. Neste processo, foram realizadas as simulações descritas na Tabela 13, e posteriormente as performances dos algoritmos foram avaliadas.

**Tabela 33 -Simulações realizadas**

<b>Simulação</b>	<b>Alteração Banco de dados</b>
1	Banco de dados inicial 55 atributos
2	Banco de dados inicial sem os atributos nulos
3	Banco de dados após correlação 80%
4	Banco de dados após correlação 65%
5	Banco de dados após correlação 50%

Com o banco de dados otimizado definido, busca-se comparar o funcionamento de cada algoritmo proposto (k-NN, SVM Linear, SVM Radial, DT e RF), com base na performance dos mesmos variando-se seus principais parâmetros, detectando o cenário de melhor desempenho a partir de uma *SearchGrid*.

Esta etapa permite que sejam definidos pontos de objetivo deste trabalho: confiabilidade de um banco de dados teórico, definição de um melhor algoritmo e seus principais parâmetros neste tipo de aplicação. A relação observada com a precisão obtida na otimização combinada à performance obtida no melhor cenário de classificador garante o alcance ao objetivo esperado.

## **CAPÍTULO 4 - RESULTADOS**

Este capítulo, apresenta as simulações realizadas e comentadas no capítulo 3, incluindo os principais resultados obtidos.

Inicialmente foram simuladas as cargas elétricas e coletadas as oscilografias correspondentes aos equipamentos apresentados na Tabela 11. Estas foram analisadas quanto ao comportamento e tipo de carga elétrica, inclusão de harmônicas na rede elétrica devido ao seu funcionamento, entre outros fatores.

A partir das verificações realizadas, as teorias de potência foram aplicadas com o objetivo de calcular os índices de potência e corrente apresentados no Capítulo 2, visando a desagregação das cargas elétricas estudadas. Com isso, o levantamento do banco de dados teórico é apresentado neste momento, bem como as melhorias propostas no mesmo.

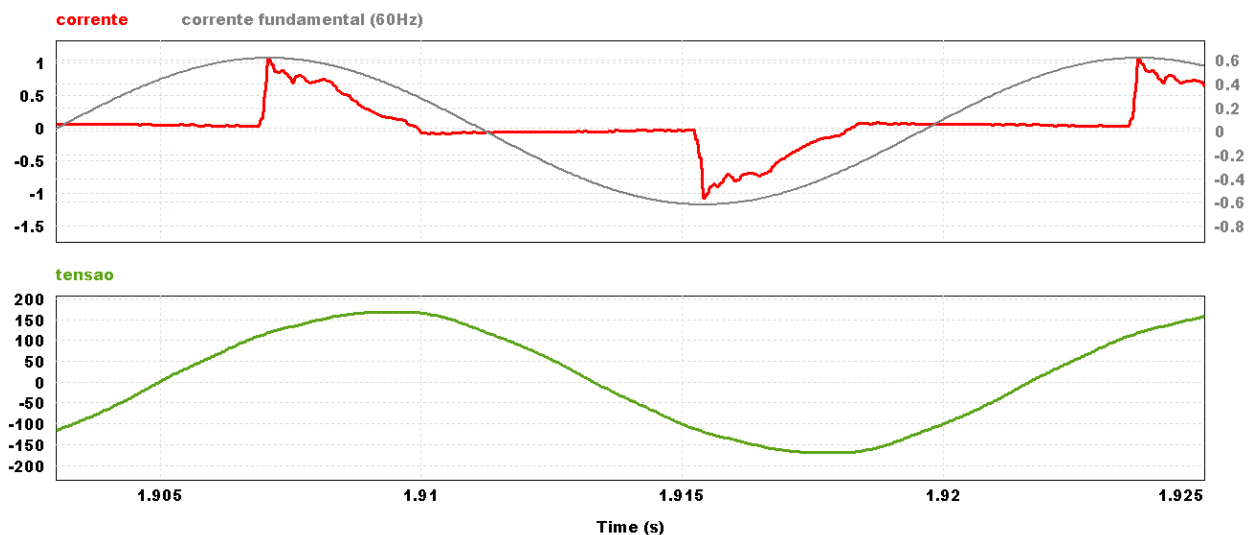
As melhorias e otimizações realizadas no banco de dados inicial, foram analisadas segundo o ponto de vista de desempenho obtida na classificação das cargas elétricas feita por meio dos algoritmos de inteligência artificial.

Ao final, os resultados obtidos para cada análise e otimização realizada são apresentados, no qual é possível verificar os melhores algoritmos de classificação voltados para este tipo de aplicação.

### **4.1. Cargas Elétricas - Oscilografias**

As oscilografias são formas de onda adquiridas com amostragem de 30 kHz que caracterizam as grandezas elétricas simuladas neste trabalho. Os comportamentos das cargas foram caracterizados levando-se em consideração o padrão de comportamentos de cargas indutivas, capacitivas e resistivas, considerando o atraso ou adiantamento da corrente em relação à tensão. De acordo com a Tabela 11 do capítulo 3, as Figuras 21 a 31 representam as oscilografias das cargas elétricas utilizadas nas simulações já informadas.

**Figura 21 - Oscilografias - Simulação 1: lâmpada fluorescente compacta**



FONTE: AUTORIA PRÓPRIA, DESENVOLVIDO NO SOFTWARE PSIM

A Figura 21 representa a oscilografia da lâmpada fluorescente compacta, na qual é possível verificar que na forma de onda de tensão a mesma permanece senoidal e sem grandes distorções, caso que não ocorre com a corrente. A forma de onda da corrente, pode-se notar periódica, porém distorcida ou fora do formato senoidal, sendo assim, espera-se a presença de elevado conteúdo harmônico. Além disso, nota-se que o comportamento deste equipamento se mantém capacitivo conforme esperado segundo a literatura, devido à tensão estar atrasada em relação à corrente.

Em [77] faz-se o estudo dos impactos e injeções de harmônicos provenientes do extenso uso de lâmpadas fluorescentes compactas nos sistemas de distribuição de energia elétrica atuais. Neste estudo e em [78], é possível observar que as formas de ondas de corrente obtidas estão equivalentes se comparadas com a Figura 20. Além disso, o estudo de referência comprova um aumento de mais de 30% na distorção harmônica de correntes para análises feitas em 520 pontos de medição em um condomínio residencial. Tais dados, comprovam que a oscilografia apresentada no trabalho atual, corresponde às análises reais evidenciadas em outras pesquisas específicas no tema. Em [78] verifica-se a porcentagem de distorção harmônica de tensão em torno de 3% e de corrente 90%.

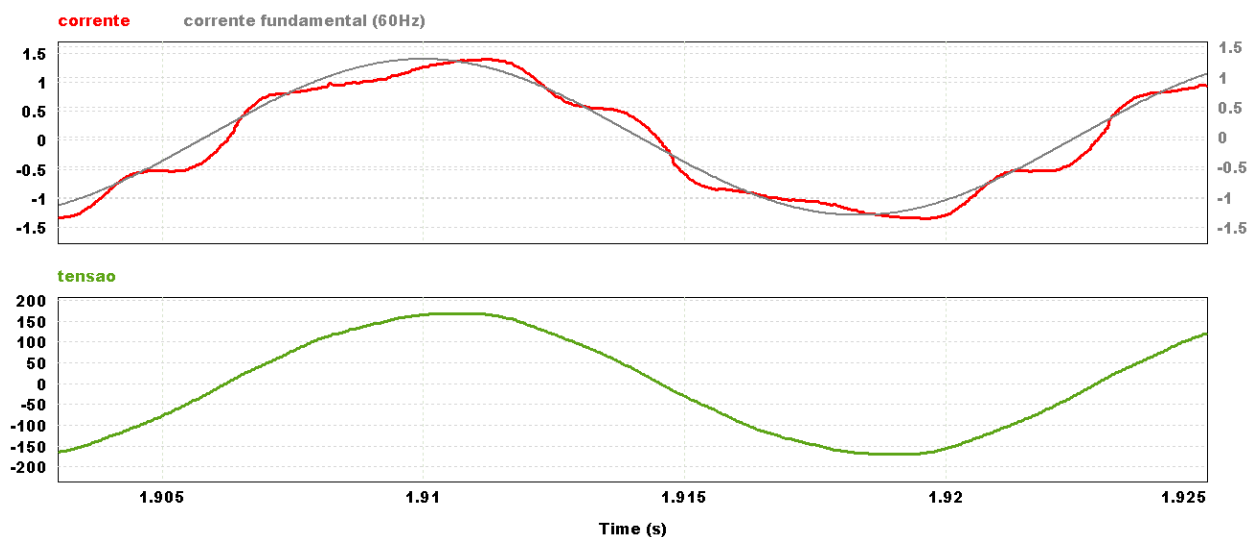
Em [79] os resultados obtidos para distorção harmônica de corrente foram ainda mais expressivos, chegando em até 171% para sistemas com tensão de 110V.

Na referência [80] é possível observar as formas de onda de tensão e corrente do funcionamento de uma lâmpada fluorescente compacta. Além disso, neste trabalho é muito discutido o efeito da capacitância existente no circuito do equipamento, e das capacitâncias parasitas presentes no mesmo. Neste sentido, comprova-se o observado na Figura 21, o comportamento capacitivo da carga simulada.

A partir dos estudos analisados, é possível concluir que a distorção harmônica esperada para as lâmpadas fluorescentes compactas é considerada alta, e como os trabalhos de referência apresentam, podem variar de 30 % a 170 %, dependendo das marcas utilizadas, sistemas analisados, tensão de referência, fonte de tensão com ou sem presença de harmônicos, precisão de equipamentos, modelos de simulação computacional, dentre outros fatores. A distorção harmônica, com grande variação pode ser um fator de diferenciação deste tipo de carga no momento da classificação.

A oscilografia da Geladeira pode ser vista na Figura 22. Nesta representação, verifica-se que tanto a tensão quanto a corrente possuem tendências senoidais, embora seja observado na corrente, uma distorção evidente.

**Figura 22 - Oscilografias - Simulação 2: Geladeira**

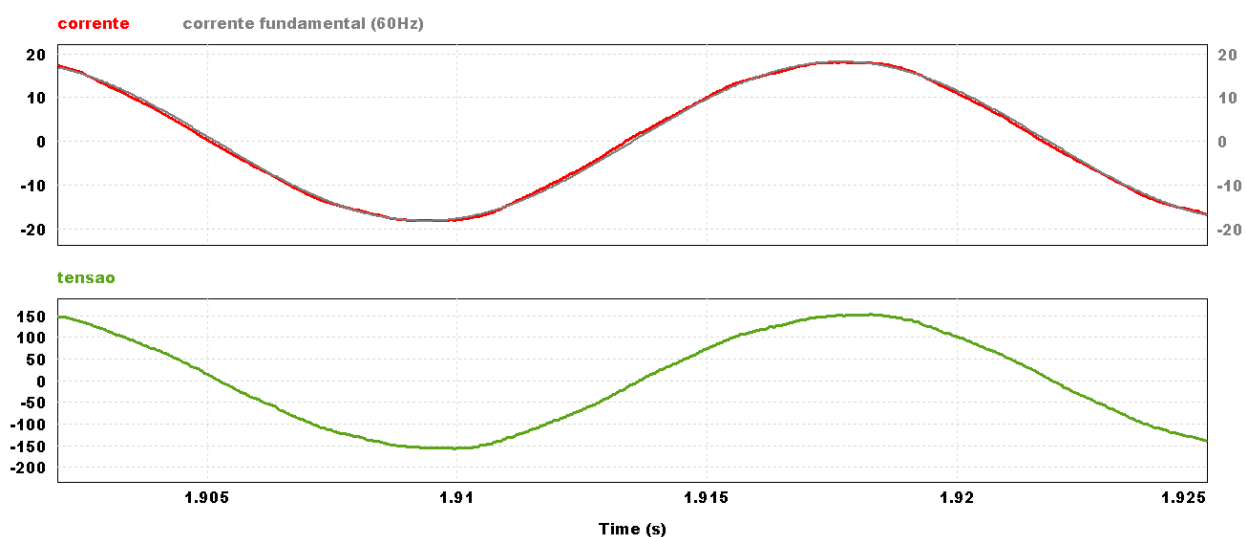


FONTE: AUTORIA PRÓPRIA, DESENVOLVIDO NO SOFTWARE PSIM



A Figura 23 apresenta a oscilografia do Secador de cabelo. Ao estabilizar o funcionamento, caracteriza-se o comportamento do secador de cabelos como resistivo visto que a corrente está praticamente em fase com relação à tensão na oscilografia observada. Sabe-se que seu funcionamento basicamente é representado por um ventilador e uma resistência para aquecimento do fluxo de ar gerado. Mais adiante é apresentada a oscilografia de um ventilador, que se comparado com a secador de cabelos, pode-se dizer possuir os mesmos comportamentos senoidais, sem muita inserção de harmônicos observada.

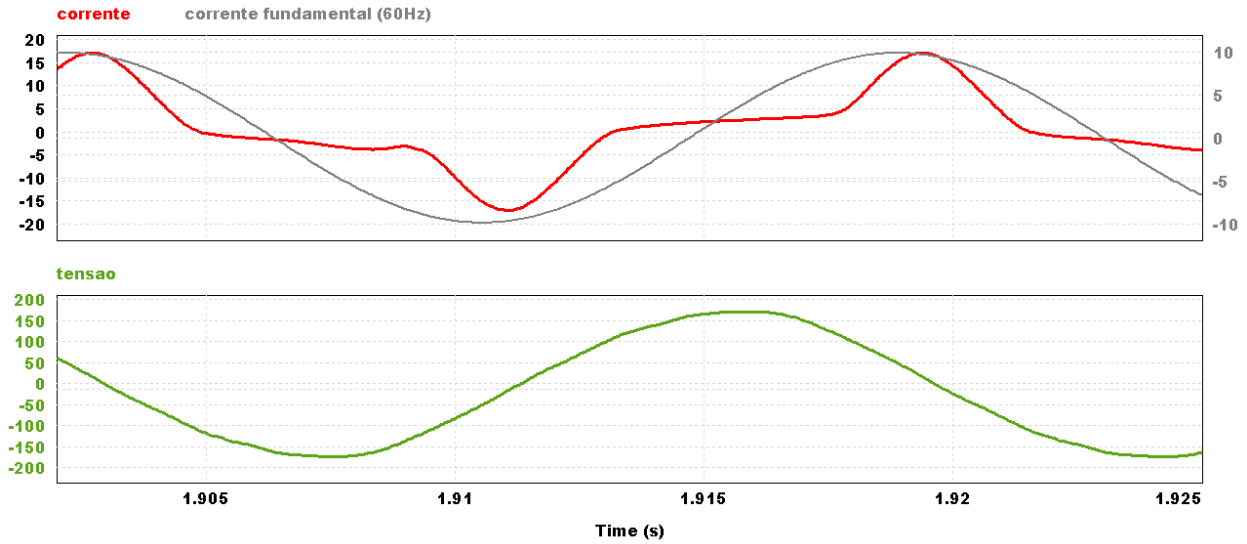
**Figura 23 - Oscilografias - Simulação 3: Secador de cabelos**



FORTE: AUTORIA PRÓPRIA, DESENVOLVIDO NO SOFTWARE PSIM

A oscilografia do micro-ondas é exibida na Figura 24, em que, é possível notar que há grande distorção na corrente.

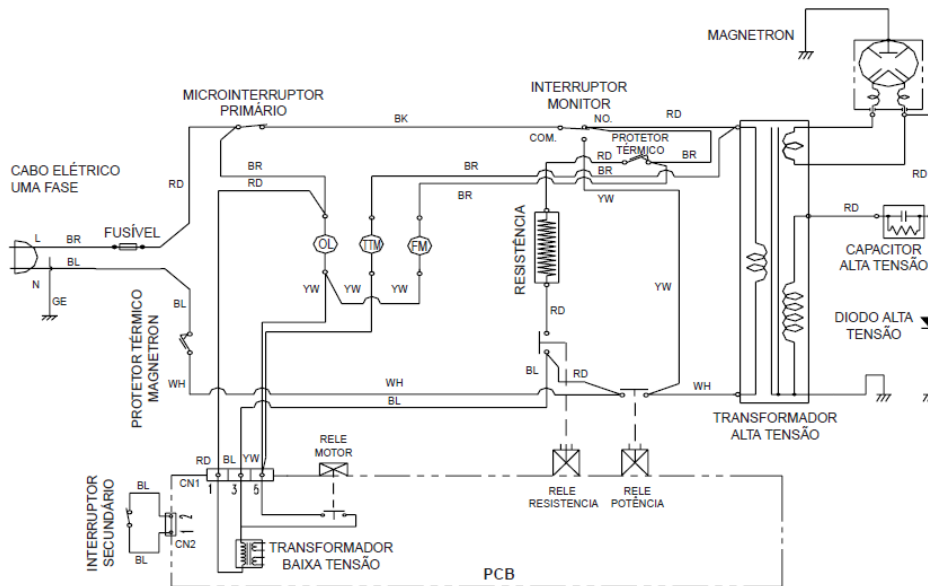
**Figura 24 - Oscilografias - Simulação 4: Micro-ondas**



FONTE: AUTORIA PRÓPRIA, DESENVOLVIDO NO SOFTWARE PSIM

A Figura 25 representa o esquemático elétrico de um forno micro-ondas Electrolux ME28G / ME28X.

**Figura 25 - Esquemático elétrico de um forno micro-ondas Electrolux ME28G / ME28X**



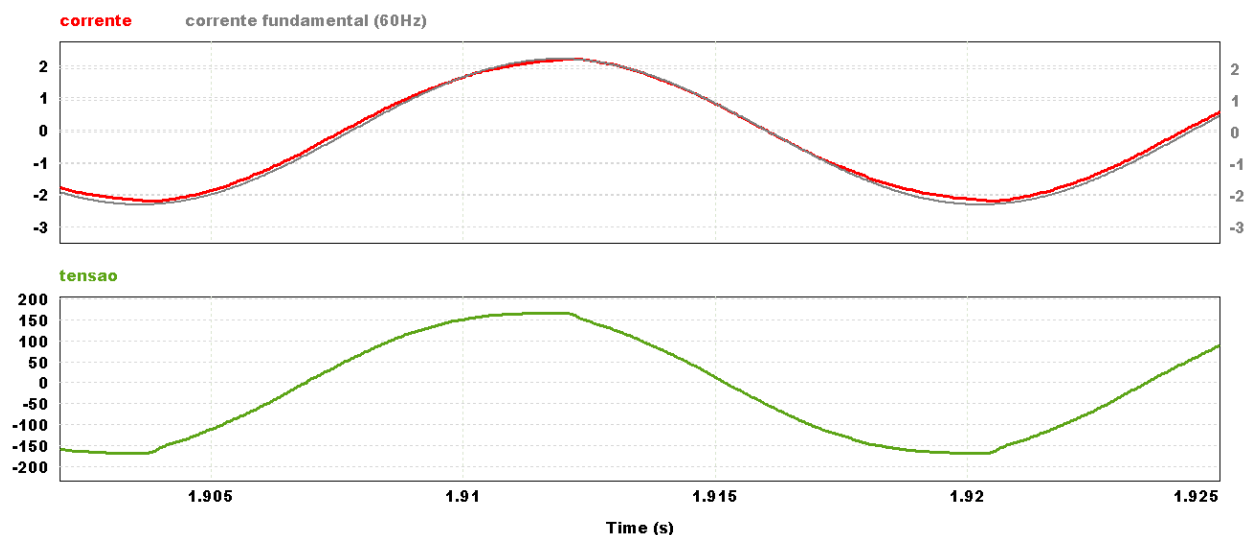
FONTE: [82]

Por meio deste diagrama, é possível justificar o comportamento indutivo deste equipamento. Nota-se que grande parte dos componentes elétricos do circuito são transformadores compostos por bobinas responsáveis pela indução de campos elétricos. Além disso, como principal destaque, tem-se o magnetron, responsável pela emissão de radiação eletromagnética através da vibração dos elétrons nas cavidades ressonantes, esse componente é constituído por um anodo cilíndrico com cavidades que de acordo com sua especificação possui um design específico. A vibração dos elétrons nesse componente acontece pelo fato de que são submetidos a uma alta tensão na ordem de kV em um curto espaço de tempo, entre o anodo e o catodo do magnetron. [83]

Segundo [78] o conteúdo harmônico para a 3ª harmônica da corrente está em aproximadamente 30%. Além disso, a oscilografia obtida, é muito similar às de referência avaliadas.

A Figura 26 representa a oscilografia do ar-condicionado. No destaque, é possível notar que há um comportamento indutivo e uma pequena distorção na forma de onda de tensão.

**Figura 26 - Oscilografias - Simulação 5: Ar-condicionado**



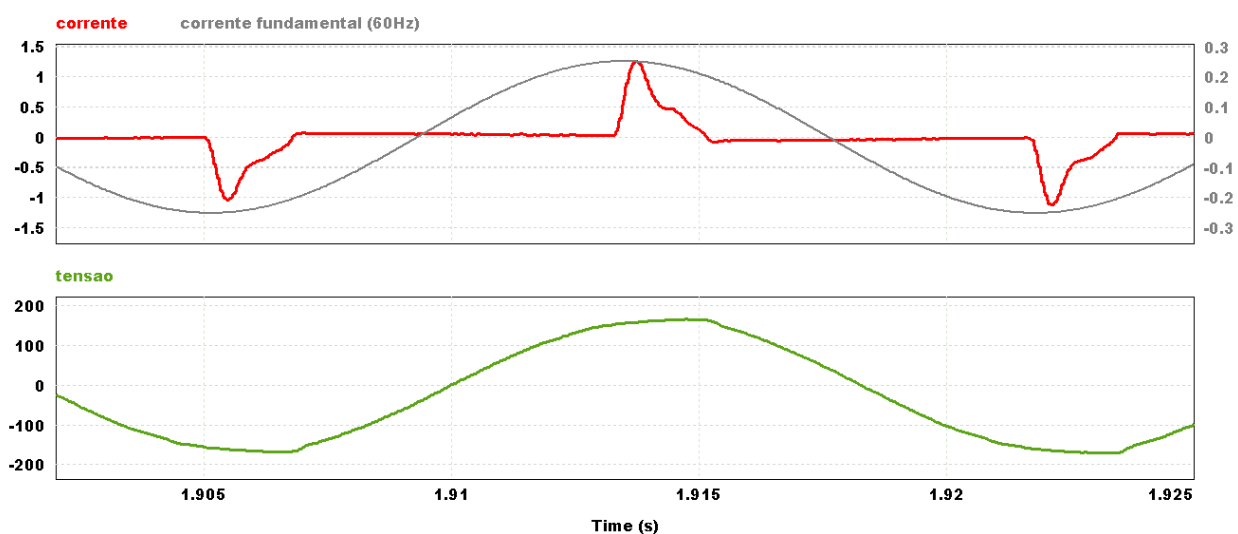
FONTE: AUTORIA PRÓPRIA, DESENVOLVIDO NO SOFTWARE PSIM

O circuito elétrico deste equipamento é marcado pelo motor de indução. Em residências, o comum é existir condicionadores de ar compostos por motores de indução

monofásicos, fato que justifica seu comportamento indutivo observado [84]. Por meio de [85] é possível observar que o funcionamento do ar-condicionado insere na tensão alguns harmônicos, com referências entre 1,5 a 2,3%.

A oscilografia obtida para o Notebook pode ser vista na Figura 27. Este tipo de carga também apresenta comportamento capacitivo, como observado na oscilografia. A forma de onda da corrente é altamente distorcida, conferindo uma grande alteração da forma de onda senoidal e presença de harmônicos notáveis.

**Figura 27 - Oscilografias - Simulação 6: Notebook**

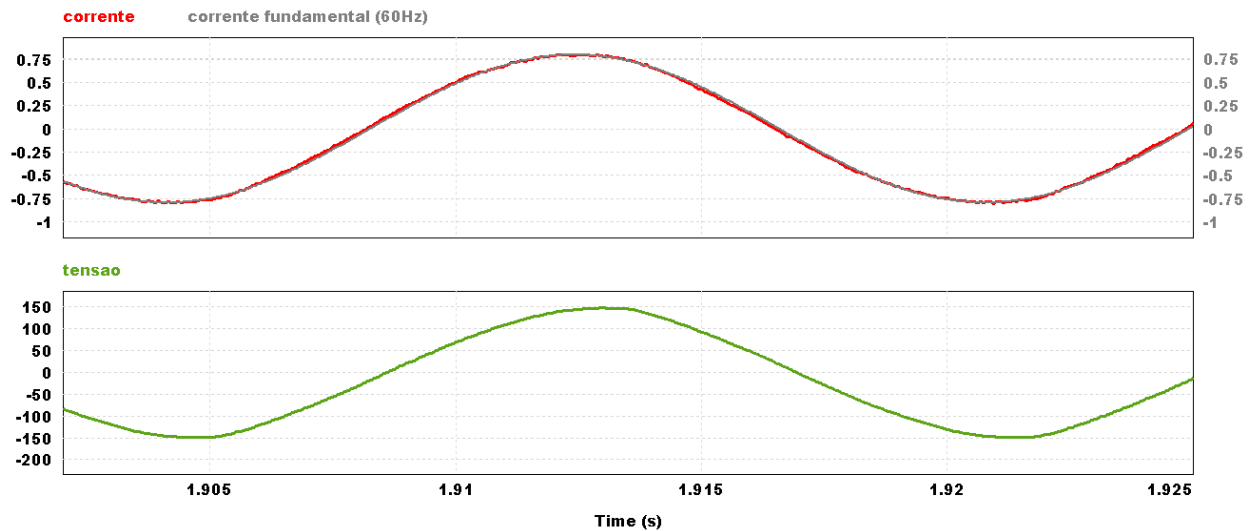


FORNTE: AUTORIA PRÓPRIA, DESENVOLVIDO NO SOFTWARE PSIM

Em [78] a distorção harmônica da tensão é elevada resultando aproximadamente em 100%. A oscilografia observada na referência é similar à Figura de referência.

Para a lâmpada incandescente, obteve-se a oscilografia vista na Figura 28. Este tipo de carga apresenta comportamento capacitivo e com formas de ondas senoidais.

**Figura 28 - Oscilografias - Simulação 7: Lâmpada incandescente**

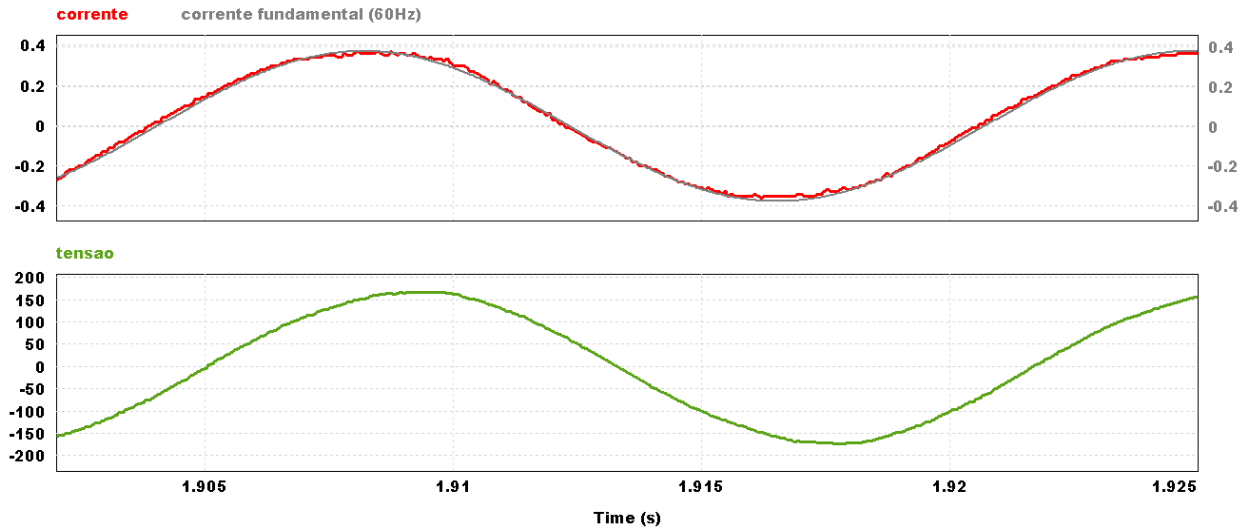


FORTE: AUTORIA PRÓPRIA, DESENVOLVIDO NO SOFTWARE PSIM

Como observado em [86], as distorções tanto na tensão quanto na corrente, são insignificantes, conforme observado na oscilografia obtida. As porcentagens de distorção harmônica apresentadas nesta referência assumem os valores de 0,1% para a tensão e 0,6% para a corrente. A oscilografia vista em [78] se assemelha ao observado neste trabalho.

A Figura 29 apresenta a oscilografia do Ventilador. Verifica-se algumas distorções presentes na forma de onda da corrente. Este tipo de carga apresenta comportamento capacitivo. Se comparado com o secador de cabelos, possuem similaridade no comportamento elétrico, porém, com valores de corrente e porcentagem de harmônicos incidentes na forma de onda da corrente ligeiramente maiores.

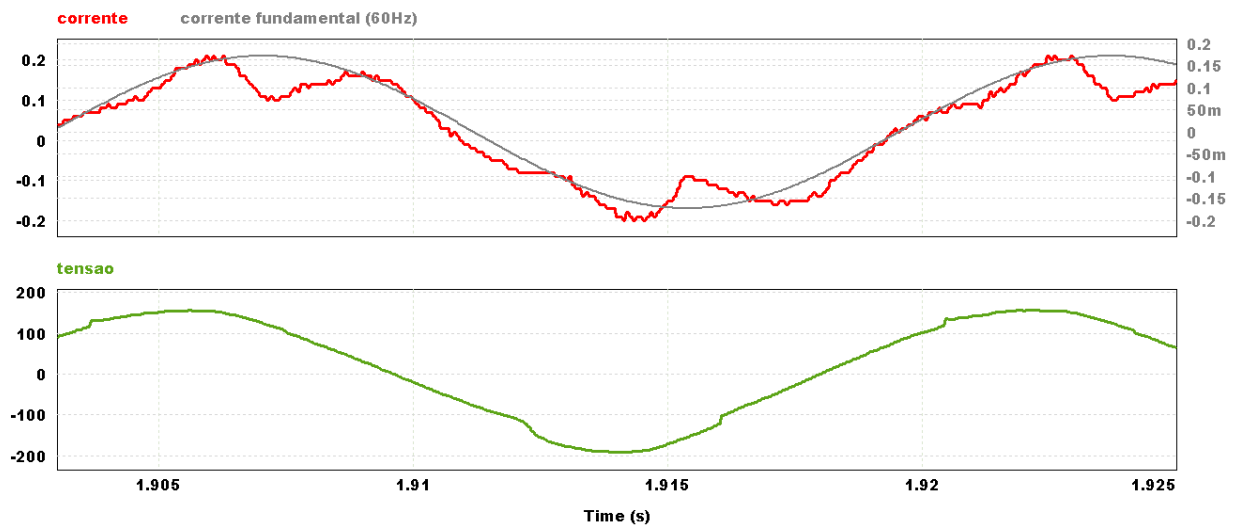
**Figura 29 - Oscilografias - Simulação 8: Ventilador**



FONTE: AUTORIA PRÓPRIA, DESENVOLVIDO NO SOFTWARE PSIM

A oscilografia da máquina de lavar roupas pode ser vista na Figura 30. Em destaque, nota-se a distorção da forma de onda, que deixa de ser perfeitamente senoidal, principalmente presente na corrente, mas também, existente na tensão deformando o formato próximos aos picos. Segundo [78] o conteúdo harmônico observado na corrente é expressivo e resulta em torno de 40%. Este tipo de carga apresenta comportamento ligeiramente indutivo.

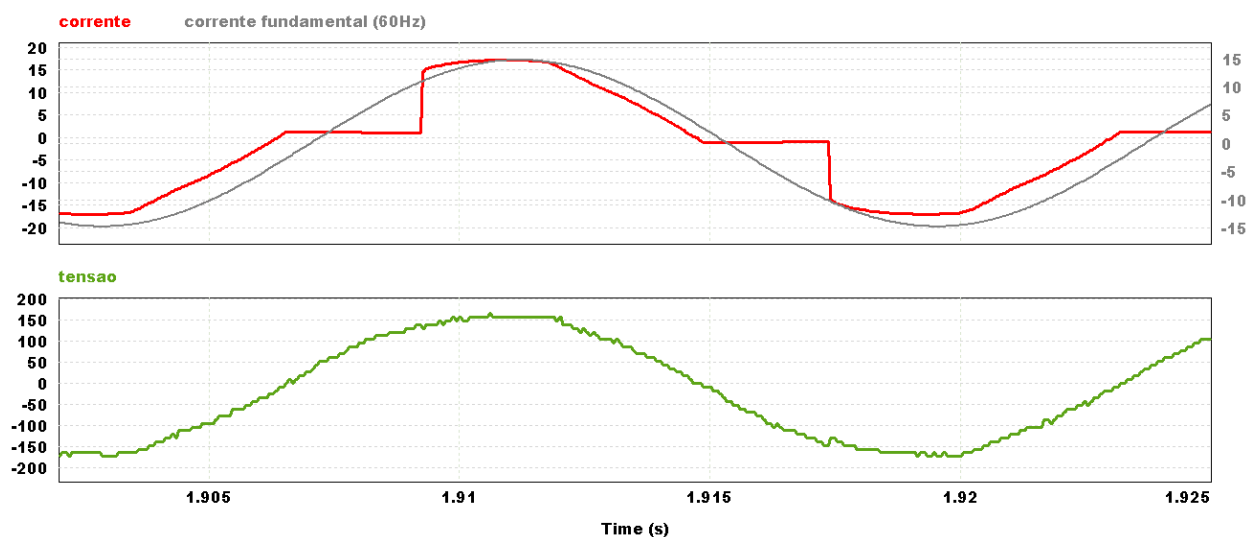
**Figura 30 - Oscilografias - Simulação 9: Máquina de lavar roupas**



FONTE: AUTORIA PRÓPRIA, DESENVOLVIDO NO SOFTWARE PSIM

A Figura 31 apresenta a oscilografia do Aquecedor. No detalhamento, é possível notar que a tensão possui um número significativo de distorções, o que acarretará significativamente em componentes harmônicas. Além disso, percebe-se que a corrente não possui a forma de onda senoidal. Esta característica pode ser utilizada como separação na classificação entre os tipos de cargas observadas.

**Figura 31 - Oscilografias - Simulação 10: Aquecedor**



FONTE: AUTORIA PRÓPRIA, DESENVOLVIDO NO SOFTWARE PSIM

## 4.2. Construção Do Banco De Dados E Validação Dos Resultados

O Banco de dados foi construído considerando dados de tensão e corrente provenientes das simulações apresentadas anteriormente, sendo composto por 590 instâncias calculadas, 10 classes (com base nas oscilografias apresentadas) e 55 atributos (índices de potência e corrente calculados com base nas teorias de potência apresentadas) que serão discutidos na seção 4.2.1 e 4.3.

#### 4.2.1. Banco de dados Inicial

As Tabelas 14 a 23 mostram os atributos para cada uma das 10 classes considerados, as médias dos valores obtidos e o desvio padrão para cada uma delas.

Como referência para todas as tabelas, as potências ativas são medidas em Watts [W], as potências aparentes em Volt Ampère [VA] e as potências não ativas como a reativa por exemplo, são medidas em Volt Ampère Reativo [VAR]. As correntes são medidas em Ampère [A], enquanto os índices de distorção harmônica, fator de potência, entre outros, são adimensionais.

**Tabela 14 - Média e desvio padrão (Carga 1).**

CLASSE 1											
CPT	Atributo	Média	Desvio	CPC	Atributo	Média	Desvio	IEEE	Atributo	Média	Desvio
		A	42.445		1.733		iacR		-0.286	0.017	
	P	24.200	1.307		iscR	-0.006	0.061		Q1p	17.920	0.513
	Q	-18.050	1.363		ircR	0.012	0.061		Se	42.446	1.738
	V	29.896	0.974		igR	0.141	0.154		Ve	119.998	0.034
	PF	0.570	0.006		iucR	0.000	0.000		Ie	0.354	0.014
	QF	-0.598	0.006		iacS	0.000	0.000		PF1p	0.000	0.000
	VF	0.705	0.005		iscS	0.000	0.000		PF1	0.804	0.012
	I_efe	0.249	0.008		ircS	0.000	0.000		Pfe	0.570	0.006
	I_eficaz	0.354	0.014		igS	0.000	0.000		THDia	96.896	5.244
	U_eficaz	119.998	0.034		iucS	0.000	0.000		vaf	-164.748	7.419
	N	0.000	0.000		iacT	0.000	0.000		iaf	-0.281	0.052
	Nr	0.000	0.000		iscT	0.000	0.000				
	Na	0.000	0.000		ircT	0.000	0.000				
	NF	0.000	0.000		igT	0.000	0.000				
	THD_ia	0.000	0.000		iucT	0.000	0.000				
	THD_ib	0.000	0.000		S	42.737	1.729				
	THD_ic	0.000	0.000		DS	6.298	0.198				
	THD_i_med	0.000	0.000		DU	0.000	0.000				
	THD_ua	0.000	0.000		DG	28.631	0.515				
	THD_ub	0.000	0.000		PF	0.582	0.005				
	THD_uc	0.000	0.000		Re_Sc	24.856	1.278				
	THD_u_med	0.000	0.000		Im_Sc	-18.229	1.329				

Para a classe 1 (lâmpada fluorescente compacta), verifica-se que os atributos com maior desvio padrão é vaf e THDia calculados por meio da norma IEEE 1459. Verificando as oscilografias, é esperada a existência de alta distorção harmônica na corrente para este caso.



O fator de potência calculado é de fato muito próximo do esperado para este tipo de equipamento conforme dados disponibilizados em análise de consumo para a compra do mesmo.

**Tabela 15 - Média e desvio padrão (Carga 6).**

CLASSE 6											
	CPT				CPC			IEEE	IEEE		
	Atributo	Média	Desvio		Atributo	Média	Desvio		Atributo	Média	Desvio
	A	113.921	1.082		iacR	-1.208	0.100		P_mean_ou	112.436	1.134
	P	112.435	1.133		iscR	0.128	0.047		Q1p	8.094	0.713
	Q	-8.177	0.163		ircR	-0.046	0.028		Se	113.922	1.082
	V	16.398	0.330		igR	0.010	0.020		Ve	119.980	0.034
	PF	0.987	0.001		iucR	0.000	0.000		Ie	0.950	0.009
	QF	-0.073	0.002		iacS	0.000	0.000		PF1p	0.000	0.000
	VF	0.144	0.004		iscS	0.000	0.000		PF1	0.997	0.000
	I_efe	0.137	0.003		ircS	0.000	0.000		Pfe	0.987	0.001
	I_eficaz	0.950	0.009		igS	0.000	0.000		THDia	19.124	16.197
	U_eficaz	119.980	0.034		iucS	0.000	0.000		vaf	-155.053	13.759
	N	0.000	0.000		iacT	0.000	0.000		iaf	-1.243	0.099
	Nr	0.000	0.000		iscT	0.000	0.000				
	Na	0.000	0.000		ircT	0.000	0.000				
	NF	0.000	0.000		igT	0.000	0.000				
	THD_ia	0.000	0.000		iucT	0.000	0.000				
	THD_ib	0.000	0.000		S	113.948	1.084				
	THD_ic	0.000	0.000		DS	11.753	0.292				
	THD_i_med	0.000	0.000		DU	0.000	0.000				
	THD_ua	0.000	0.000		DG	2.948	0.230				
	THD_ub	0.000	0.000		PF	0.987	0.001				
	THD_uc	0.000	0.000		Re_Sc	112.446	1.135				
	THD_u_med	0.000	0.000		Im_Sc	-8.145	0.170				

Considerando classe 6 (geladeira), o atributo com maior desvio padrão observado também foi a THDia calculada por meio da norma IEEE 1459. Devido às oscilações de funcionamento do motor comentada anteriormente, realmente espera-se uma inserção de harmônicos na corrente. Outra observação importante sobre a carga refere-se ao alto fator de potência calculado.

**Tabela 16 - Média e desvio padrão (Carga 7).**

CLASSE 7											
	CPT				CPC				IEEE		
	Atributo	Média	Desvio		Atributo	Média	Desvio		Atributo	Média	Desvio
	A	1275.875	475.175		iacR	15.937	5.461		P_mean_out	1271.305	473.746
	P	1271.281	473.788		iscR	0.147	0.152		Q1p	105.630	44.156
	Q	-103.207	39.099		ircR	-0.070	0.103		Se	1275.882	475.179
	V	34.729	88.355		igR	0.026	0.039		Ve	118.524	17.708
	PF	0.868	0.352		iucR	0.000	0.000		le	11.359	4.202
	QF	-0.135	0.169		iacS	0.000	0.000		PF1p	0.000	0.000
	VF	0.141	0.319		iscS	0.000	0.000		PF1	0.768	0.626
	I_efe	0.300	0.709		ircS	0.000	0.000		Pfe	0.868	0.353
	I_eficaz	11.359	4.202		igS	0.000	0.000		THDia	212.826	621.427
	U_eficaz	118.522	17.706		iucS	0.000	0.000		vaf	164.785	22.731
	N	0.000	0.000		iacT	0.000	0.000		iaf	15.471	5.833
	Nr	0.000	0.000		iscT	0.000	0.000				
	Na	0.000	0.000		ircT	0.000	0.000				
	NF	0.000	0.000		igT	0.000	0.000				
	THD_ia	0.000	0.000		iucT	0.000	0.000				
	THD_ib	0.000	0.000		S	1295.321	444.913				
	THD_ic	0.000	0.000		DS	29.462	19.559				
	THD_i_med	0.000	0.000		DU	0.000	0.000				
	THD_ua	0.000	0.000		DG	4.390	1.343				
	THD_ub	0.000	0.000		PF	0.897	0.299				
	THD_uc	0.000	0.000		Re_Sc	1290.399	443.525				
	THD_u_med	0.000	0.000		Im_Sc	-1028.237	380.608				

Para os dados observados na Tabela 16 referentes ao secador de cabelos, nota-se o que já foi observado com relação ao atributo THDia calculado por meio da norma IEEE 1459.

**Tabela 47 - Média e desvio padrão (Carga 10).**

CLASSE 10											
	CPT				CPC				IEEE		
	Atributo	Média	Desvio		Atributo	Média	Desvio		Atributo	Média	Desvio
	A	956.932	388.419		iacR	1.524	1.310		P_mean_out	266.850	186.079
	P	266.871	186.267		iscR	-1.147	0.436		Q1p	750.499	201.193
	Q	765.021	227.936		ircR	8.209	2.542		Se	956.958	388.400
	V	503.539	330.371		igR	1.338	3.578		Ve	121.522	0.512
	PF	0.306	0.150		iucR	0.000	0.000		le	7.888	3.283
	QF	0.909	0.153		iacS	0.000	0.000		PF1p	0.000	0.000
	VF	0.496	0.107		iscS	0.000	0.000		PF1	0.362	0.149
	I_efe	4.155	2.804		ircS	0.000	0.000		Pfe	0.306	0.150
	I_eficaz	7.888	3.284		igS	0.000	0.000		THDia	68.688	76.443
	U_eficaz	121.523	0.512		iucS	0.000	0.000		vaf	84.626	8.881
	N	0.000	0.000		iacT	0.000	0.000		iaf	8.765	2.343
	Nr	0.000	0.000		iscT	0.000	0.000				
	Na	0.000	0.000		ircT	0.000	0.000				
	NF	0.000	0.000		igT	0.000	0.000				
	THD_ia	0.000	0.000		iucT	0.000	0.000				
	THD_ib	0.000	0.000		S	965.417	435.129				
	THD_ic	0.000	0.000		DS	125.263	34.770				
	THD_i_med	0.000	0.000		DU	0.000	0.000				
	THD_ua	0.000	0.000		DG	467.144	331.763				
	THD_ub	0.000	0.000		PF	0.315	0.146				
	THD_uc	0.000	0.000		Re_Sc	278.103	198.967				
	THD_u_med	0.000	0.000		Im_Sc	766.797	236.990				

Neste caso, representando o micro-ondas, o maior desvio padrão é observado no atributo A (potência aparente). Este ponto era esperado, conforme visto nas oscilografias, há uma mudança significativa da forma de onda da corrente ocasionada pela inserção de harmônicos.

**Tabela 185 - Média e desvio padrão (Carga 15).**

CLASSE 15											
	CPT				CPC			IEEE	IEEE		
	Atributo	Média	Desvio		Atributo	Média	Desvio		Atributo	Média	Desvio
	A	119.451	95.375		iacR	-1.188	0.953		P_mean_out	112.110	89.932
	P	112.116	89.932		iscR	-0.007	0.007		Q1p	38.604	31.642
	Q	39.077	32.028		ircR	0.177	0.149		Se	119.448	95.374
	V	4.747	14.318		igR	0.002	0.019		Ve	120.059	0.356
	PF	0.873	0.087		iucR	0.000	0.000		Ie	0.997	0.796
	QF	0.136	0.247		iacS	0.000	0.000		PF1p	0.000	0.000
	VF	0.268	0.301		iscS	0.000	0.000		PF1	0.954	0.049
	I_efe	0.040	0.119		ircS	0.000	0.000		Pfe	0.873	0.087
	I_eficaz	0.997	0.796		igS	0.000	0.000		THDia	43.399	70.357
	U_eficaz	120.059	0.356		iucS	0.000	0.000		vaf	-147.062	11.580
	N	0.000	0.000		iacT	0.000	0.000		iaf	-1.000	0.815
	Nr	0.000	0.000		iscT	0.000	0.000				
	Na	0.000	0.000		ircT	0.000	0.000				
	NF	0.000	0.000		igT	0.000	0.000				
	THD_ia	0.000	0.000		iucT	0.000	0.000				
	THD_ib	0.000	0.000		S	119.643	95.456				
	THD_ic	0.000	0.000		DS	1.404	0.933				
	THD_i_med	0.000	0.000		DU	0.000	0.000				
	THD_ua	0.000	0.000		DG	2.866	14.364				
	THD_ub	0.000	0.000		PF	0.870	0.091				
	THD_uc	0.000	0.000		Re_Sc	112.120	89.914				
	THD_u_med	0.000	0.000		Im_Sc	39.079	32.018				

Para o ar-condicionado, observa-se o maior desvio padrão no atributo A (potência aparente). O fator de potência observado para este caso também pode ser considerado alto e representa um equipamento de alta eficiência de motores.

**Tabela 19 - Média e desvio padrão (Carga 28).**

CLASSE 28											
CPT	Atributo	Média	Desvio	CPC	Atributo	Média	Desvio	IEEE	Atributo	Média	Desvio
		A	35.911		8.112		iacR		0.053	0.044	
	P	17.686	4.569		iscR	0.001	0.037		Q1p	5.764	1.918
	Q	-5.918	1.965		ircR	-0.062	0.050		Se	35.911	8.112
	V	30.656	6.453		igR	-0.040	0.059		Ve	119.817	4.480
	PF	0.488	0.021		iucR	0.000	0.000		Ie	0.299	0.064
	QF	-0.311	0.026		iacS	0.000	0.000		PF1p	0.000	0.000
	VF	0.857	0.018		iscS	0.000	0.000		PF1	0.952	0.009
	I_efe	0.255	0.051		ircS	0.000	0.000		Pfe	0.488	0.021
	I_eficaz	0.299	0.064		igS	0.000	0.000		THDia	170.693	33.753
	U_eficaz	119.817	4.480		iucS	0.000	0.000		vaf	54.207	39.841
	N	0.000	0.000		iacT	0.000	0.000		iaf	-0.012	0.059
	Nr	0.000	0.000		iscT	0.000	0.000				
	Na	0.000	0.000		ircT	0.000	0.000				
	NF	0.000	0.000		igT	0.000	0.000				
	THD_ia	0.000	0.000		iucT	0.000	0.000				
	THD_ib	0.000	0.000		S	36.120	8.365				
	THD_ic	0.000	0.000		DS	6.195	1.901				
	THD_i_med	0.000	0.000		DU	0.000	0.000				
	THD_ua	0.000	0.000		DG	29.242	6.249				
	THD_ub	0.000	0.000		PF	0.498	0.023				
	THD_uc	0.000	0.000		Re_Sc	18.136	4.819				
	THD_u_med	0.000	0.000		Im_Sc	-5.822	1.991				

Por meio dos atributos calculados, entende-se que o notebook possui uma baixa eficiência, uma vez que os fatores de potência calculados são baixos.

**Tabela 20 - Média e desvio padrão (Carga 57).**

CLASSE 57											
CPT	Atributo	Média	Desvio	CPC	Atributo	Média	Desvio	IEEE	Atributo	Média	Desvio
		A	70.420		90.926		iacR		-0.291	0.120	
	P	67.948	78.965		iscR	-0.012	0.021		Q1p	10.144	4.759
	Q	-9.568	3.315		ircR	-0.113	0.029		Se	70.420	90.923
	V	12.144	75.331		igR	0.008	0.015		Ve	118.271	37.377
	PF	0.917	0.252		iucR	0.000	0.000		Ie	0.553	0.244
	QF	-0.169	0.126		iacS	0.000	0.000		PF1p	0.000	0.000
	VF	0.114	0.261		iscS	0.000	0.000		PF1	0.919	0.252
	I_efe	0.046	0.206		ircS	0.000	0.000		Pfe	0.917	0.252
	I_eficaz	0.553	0.244		igS	0.000	0.000		THDia	305.788	1160.348
	U_eficaz	118.271	37.378		iucS	0.000	0.000		vaf	-57.975	14.010
	N	0.000	0.000		iacT	0.000	0.000		iaf	-0.396	0.135
	Nr	0.000	0.000		iscT	0.000	0.000				
	Na	0.000	0.000		ircT	0.000	0.000				
	NF	0.000	0.000		igT	0.000	0.000				
	THD_ia	0.000	0.000		iucT	0.000	0.000				
	THD_ib	0.000	0.000		S	71.239	89.672				
	THD_ic	0.000	0.000		DS	4.875	25.967				
	THD_i_med	0.000	0.000		DU	0.000	0.000				
	THD_ua	0.000	0.000		DG	1.137	0.925				
	THD_ub	0.000	0.000		PF	0.937	0.208				
	THD_uc	0.000	0.000		Re_Sc	68.705	77.215				
	THD_u_med	0.000	0.000		Im_Sc	-11.002	10.706				

A Tabela 20 representa os atributos calculados para a lâmpada incandescente. Os valores de potência reativa calculados para ambas teorias, apresentam ligeira variação nos cálculos das médias.

**Tabela 21 - Média e desvio padrão (Carga 66).**

CLASSE 66											
CPT			CPC			IEEE			IEEE		
Atributo	Média	Desvio	Atributo	Média	Desvio	Atributo	Média	Desvio	Atributo	Média	Desvio
A	27.948	10.907	iacR	-0.300	0.121	p_mean_out	25.811	10.309			
P	25.809	10.314	iscR	-0.005	0.003	Q1p	10.247	4.175			
Q	-10.403	4.143	ircR	0.022	0.017	Se	27.949	10.906			
V	1.224	0.928	igR	0.004	0.006	Ve	120.002	0.081			
PF	0.804	0.312	iucR	0.000	0.000	Ie	0.233	0.091			
QF	-0.404	0.227	iacS	0.000	0.000	PF1p	0.000	0.000			
VF	0.171	0.332	iscS	0.000	0.000	PF1	0.847	0.257			
I_efe	0.010	0.008	ircS	0.000	0.000	Pfe	0.803	0.313			
I_eficaz	0.233	0.091	igS	0.000	0.000	THDia	339.262	897.292			
U_eficaz	120.002	0.081	iucS	0.000	0.000	vaf	-164.923	7.347			
N	0.000	0.000	iacT	0.000	0.000	iaf	-0.276	0.115			
Nr	0.000	0.000	iscT	0.000	0.000						
Na	0.000	0.000	ircT	0.000	0.000						
NF	0.000	0.000	igT	0.000	0.000						
THD_ia	0.000	0.000	iucT	0.000	0.000						
THD_ib	0.000	0.000	S	27.972	10.892						
THD_ic	0.000	0.000	DS	0.555	0.224						
THD_i_med	0.000	0.000	DU	0.000	0.000						
THD_ua	0.000	0.000	DG	0.900	0.098						
THD_ub	0.000	0.000	PF	0.806	0.308						
THD_uc	0.000	0.000	Re_Sc	25.822	10.314						
THD_u_med	0.000	0.000	Im_Sc	-10.397	4.146						

Para o ventilador, era esperado obter atributos similares ao do secador de cabelo, porém, é possível notar que os mesmos são distintos. Os valores de potência ativa, reativa e aparente apresentam-se elevados no secador de cabelos se comparados ao do ventilador. O fator de potência é próximo se comparada as duas cargas, fato que era esperado, segundo o estudo levantado sobre o funcionamento do equipamento.

**Tabela 226 - Média e desvio padrão (Carga 80).**

CLASSE 80											
	CPT				CPC				IEEE		
	Atributo	Média	Desvio		Atributo	Média	Desvio		Atributo	Média	Desvio
	A	15.160	0.206		iacR	0.008	0.014		P_mean_out	12.409	0.171
	P	12.410	0.171		iscR	-0.002	0.007		Q1p	7.540	0.672
	Q	7.441	0.146		ircR	-0.113	0.003		Se	15.160	0.206
	V	4.753	0.080		igR	-0.006	0.013		Ve	120.663	1.602
	PF	0.819	0.001		iucR	0.000	0.000		le	0.126	0.000
	QF	0.514	0.003		iacS	0.000	0.000		PF1p	0.000	0.000
	VF	0.314	0.007		iscS	0.000	0.000		PF1	0.851	0.001
	I_efe	0.039	0.001		ircS	0.000	0.000		Pfe	0.819	0.001
	I_eficaz	0.126	0.000		igS	0.000	0.000		THDia	28.151	15.018
	U_eficaz	120.663	1.603		iucS	0.000	0.000		vaf	5.027	19.050
	N	0.000	0.000		iacT	0.000	0.000		iaf	-0.085	0.016
	Nr	0.000	0.000		iscT	0.000	0.000				
	Na	0.000	0.000		ircT	0.000	0.000				
	NF	0.000	0.000		igT	0.000	0.000				
	THD_ia	0.000	0.000		iucT	0.000	0.000				
	THD_ib	0.000	0.000		S	15.187	0.205				
	THD_ic	0.000	0.000		DS	0.874	0.505				
	THD_i_med	0.000	0.000		DU	0.000	0.000				
	THD_ua	0.000	0.000		DG	2.918	0.482				
	THD_ub	0.000	0.000		PF	0.824	0.001				
	THD_uc	0.000	0.000		Re_Sc	12.448	0.175				
	THD_u_med	0.000	0.000		Im_Sc	7.501	0.126				

Considerando a máquina de lavar roupas, espera-se um valor de distorção harmônica alto, conforme observado para THDia. Percebe-se que os desvios padrões neste caso, são baixos se comparados aos demais equipamentos presentes neste estudo.

**Tabela 23 - Média e desvio padrão (Carga 136).**

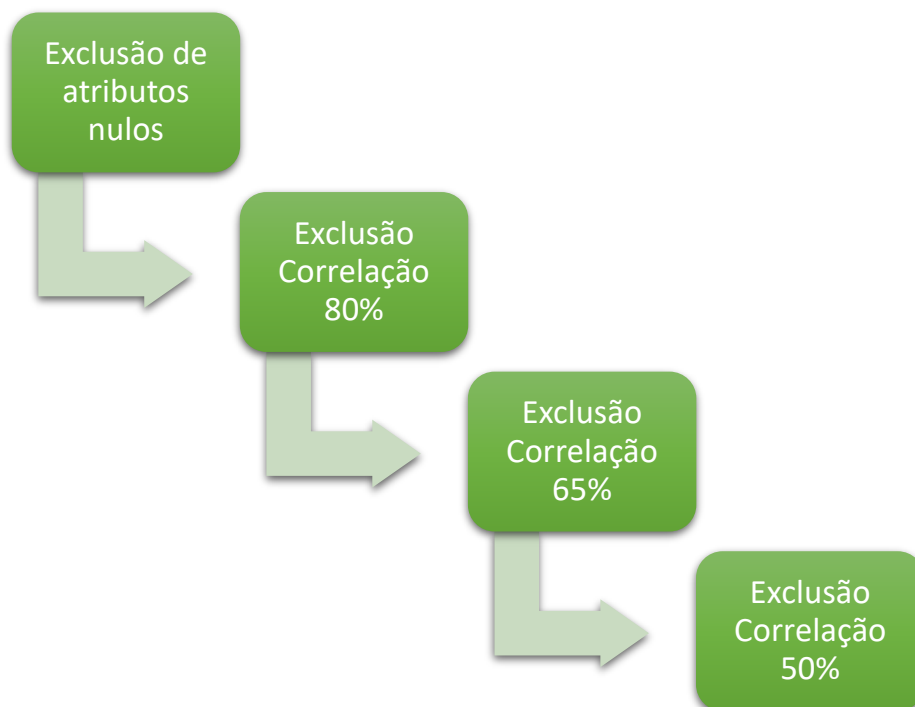
CLASSE 136											
	CPT				CPC				IEEE		
	Atributo	Média	Desvio		Atributo	Média	Desvio		Atributo	Média	Desvio
	A	1038.811	531.341		iacR	-10.158	5.413		P_mean_out	952.312	492.443
	P	952.374	492.468		iscR	-0.011	0.723		Q1p	191.017	98.802
	Q	186.844	99.931		ircR	1.022	1.469		Se	1038.775	531.323
	V	365.696	190.302		igR	0.394	3.468		Ve	119.987	2.520
	PF	0.776	0.282		iucR	0.000	0.000		le	8.749	4.478
	QF	-0.041	0.463		iacS	0.000	0.000		PF1p	0.000	0.000
	VF	0.422	0.148		iscS	0.000	0.000		PF1	0.838	0.273
	I_efe	3.078	1.597		ircS	0.000	0.000		Pfe	0.776	0.282
	I_eficaz	8.750	4.478		igS	0.000	0.000		THDia	60.502	82.811
	U_eficaz	119.987	2.520		iucS	0.000	0.000		vaf	-135.007	24.394
	N	0.000	0.000		iacT	0.000	0.000		iaf	-8.552	4.820
	Nr	0.000	0.000		iscT	0.000	0.000				
	Na	0.000	0.000		ircT	0.000	0.000				
	NF	0.000	0.000		igT	0.000	0.000				
	THD_ia	0.000	0.000		iucT	0.000	0.000				
	THD_ib	0.000	0.000		S	1041.513	531.531				
	THD_ic	0.000	0.000		DS	82.872	44.762				
	THD_i_med	0.000	0.000		DU	0.000	0.000				
	THD_ua	0.000	0.000		DG	314.504	169.681				
	THD_ub	0.000	0.000		PF	0.783	0.284				
	THD_uc	0.000	0.000		Re_Sc	961.152	496.203				
	THD_u_med	0.000	0.000		Im_Sc	190.531	99.320				

Para o aquecedor, percebe-se que os cálculos das potências possuem bastante dispersão nos dados, uma vez que em ambas as teorias há valores expressivos de desvio padrão para os atributos relacionados.

#### 4.3. Otimização Do Banco De Dados - Definição Dos Atributos Para O Banco De Dados Final - Cenário Geral

Após levantamento do banco de dados inicial, análise das oscilografias e índices calculados considerando o cenário geral, observou-se que somente a média e o desvio padrão não seriam suficientes para determinar a relevância dos atributos. Desta forma, optou-se por seguir as etapas descritas no fluxograma da Figura 32, a fim de otimizar o banco de dados levantado.

**Figura 32 - Fluxograma da otimização do banco de dados**



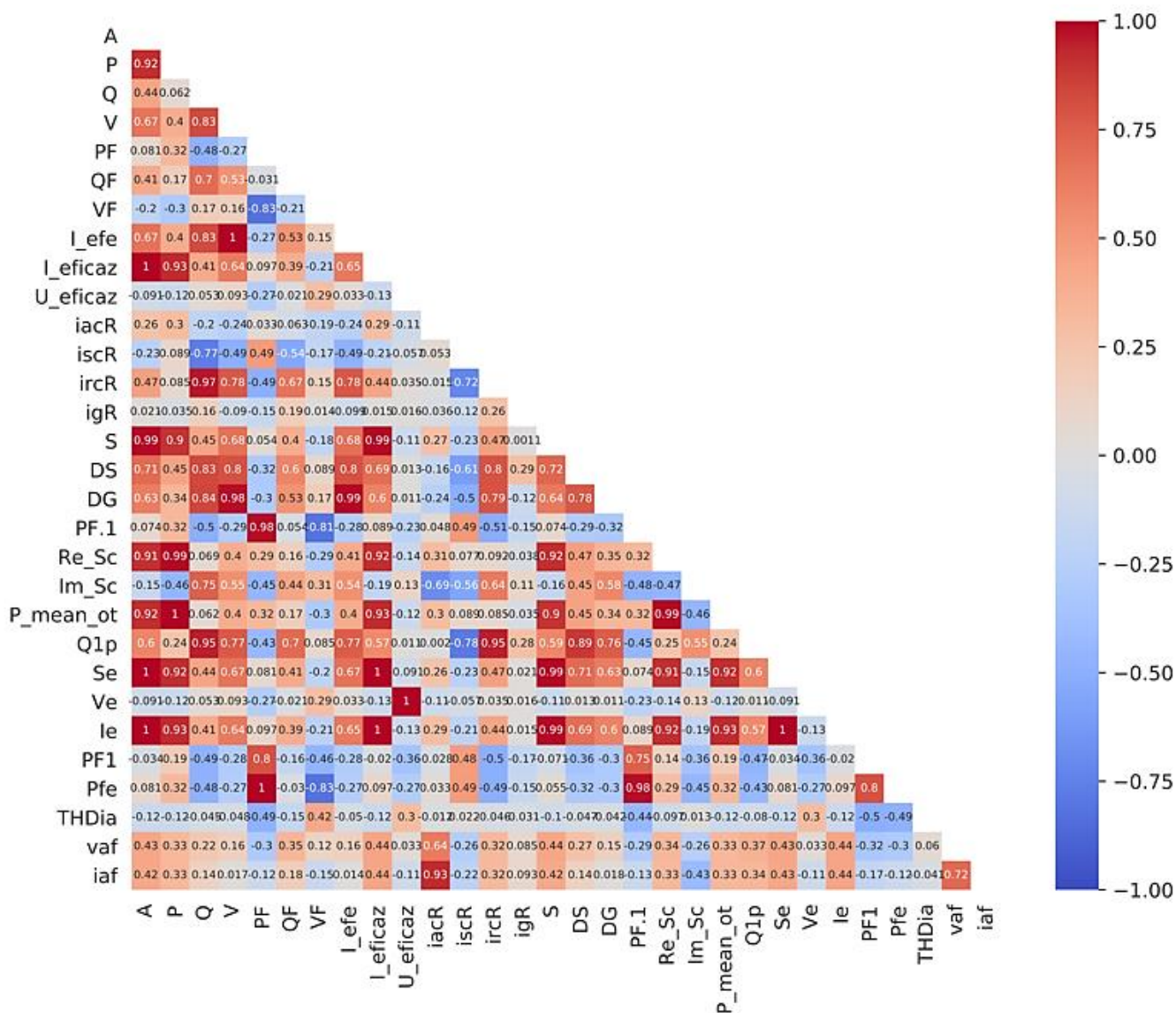
FONTE: AUTORIA PRÓPRIA

Primeiramente, observou-se no banco de dados levantado a variação dos índices calculados e constatou-se que alguns dos 55 atributos levantados não sofriam variação por algumas razões. Como por exemplo, não estavam sendo calculados, tinham sempre valores nulos, ou então apresentavam sempre o mesmo valor pois são índices correspondentes à cargas trifásicas e no trabalho são analisadas apenas cargas monofásicas. Desta forma, houve a primeira redução no número de dados para serem analisados. Assim, foram excluídos 25 atributos, sendo eles: N, Nr, Na, NF, THD\_ia, THD\_ib, THD\_ic, THD\_i\_med, THD\_ua, THD\_ub, THD\_uc, THD\_u\_med, iucR, iacS, iscS, ircS, igS, iucS, iacT, iscT, ircT, igT, iucT, DU, PF1p.

Após a primeira análise, a elaboração da matriz de correlação, contribuiu para uma segunda verificação da composição dos atributos que estarão presentes na elaboração do banco de dados. Desta forma, observa-se na Figura 33 o resultado da matriz de correlação para CPT, CPC e IEEE STD 1459-2010.



Figura 33 - Matriz de correlação dos atributos calculados



FONTE: AUTORIA PRÓPRIA, DESENVOLVIDO NO SOFTWARE SPYDER (ANACONDA)

Confirma-se que o desvio padrão não reflete a relevância das variáveis. Desta forma, a Figura 33 será utilizada para definição dos atributos que serão de fato analisados e mantidos na composição do banco de dados.

Como determinado pela matriz de correlação, os índices que possuem correlação forte positiva (índice de correlação entre +0.5 e +1) e correlação forte negativa (índice de correlação entre -0.5 e -1) serão verificados para a eliminação de uma das variáveis que compõe o par de parâmetros correlacionados. A Tabela 24 evidencia o panorama de

correlações observadas, os valores dos índices de correlação correspondentes e os atributos que foram retirados do banco de dados em cada fase das análises.

**Tabela 74 - Correlação observada a partir da Matriz de correlação e otimizações propostas**

ÍNDICE DE CORRELAÇÃO	PAR DE ATRIBUTOS	ATRIBUTO DEFINIDO	ATRIBUTO EXCLUÍDO
+/- 1	I_ eficaz e A	I_ eficaz	A
	S e A	S	A
	Se e A	Se	A
	le e A	le	A
	Re_Sc e P	P	Re_Sc
	P_mean.out e P	P	P_mean.out
	ircR e Q	ircR	Q
	Q1p e Q	Q1p	Q
	I_ efe e V	V	I_ efe
	DG e V	V	DG
	PF.1 e PF	PF	PF.1
	Pfe e PF	PF	Pfe
	I_ eficaz e Se	Se	I_ eficaz
	I_ eficaz e S	S	I_ eficaz
	le e I_ eficaz	le	I_ eficaz
	Ve e U_ eficaz	Ve	U_ eficaz
	Q1p e ircR	Q1p	ircR
	PF.1 e Pfe	Pfe	PF.1
	Se e S	Se	S
	le e S	le	S
	le e Se	Se	le
P_mean.out e Re_Sc	Ambos excluídos	Ambos excluídos	

**Tabela 24 - Correlação observada a partir da Matriz de correlação e otimizações propostas (continuação)**

ÍNDICE DE CORRELAÇÃO	PAR DE ATRIBUTOS	ATRIBUTO DEFINIDO	ATRIBUTO EXCLUÍDO
+/- 0.8 a +/- 0.9	P e A	P	A (Já excluído)
	Re_Sc e A	Ambos excluídos	Ambos excluídos
	P_mean.out e A	Ambos excluídos	Ambos excluídos
	I_eficaz e P	P	I_eficaz (Já excluído)
	S e P	P	S (Já excluído)
	Se e P	P	Se
	le e P	P	le (Já excluído)
	V e Q	V	Q (Já excluído)
	I_efe e Q	Ambos excluídos	Ambos excluídos
	iscR e Q	iscR	Q (Já excluído)
	DS e Q	DS	Q (Já excluído)
	DG e Q	Ambos excluídos	Ambos excluídos
	ircR e V	V	ircR (Já excluído)
	DS e V	V	DS
	Q1p e V	V	Q1p
	ircR e V	V	ircR (Já excluído)
	VF e PF	PF	VF
	PF1 e PF	PF	PF1 (Já excluído)
	PF1 e VF	Ambos excluídos	Ambos excluídos
	Pfe e VF	Ambos excluídos	Ambos excluídos
	ircR e I_efe	Ambos excluídos	Ambos excluídos
	DS e I_efe	Ambos excluídos	Ambos excluídos
	Q1p e I_efe	Ambos excluídos	Ambos excluídos
	Re_Sc e I_eficaz	Ambos excluídos	Ambos excluídos
	P.mean_out e I_eficaz	Ambos excluídos	Ambos excluídos
	iaf e iacR	iaf	iacR
	Q1p e iscR	iscR	Q1p (Já excluído)
	DS e ircR	Ambos excluídos	Ambos excluídos
	DG e ircR	Ambos excluídos	Ambos excluídos
	DS e PF1	Ambos excluídos	Ambos excluídos
	Q1p e DS	Ambos excluídos	Ambos excluídos
	Se e P_mean-out	Ambos excluídos	Ambos excluídos
	Se e Re_Sc	Ambos excluídos	Ambos excluídos
	le e P_mean-out	Ambos excluídos	Ambos excluídos
le e Re_Sc	Ambos excluídos	Ambos excluídos	
Pfe e PF1	Ambos excluídos	Ambos excluídos	

**Tabela 24 - Correlação observada a partir da Matriz de correlação e otimizações propostas (continuação)**

ÍNDICE DE CORRELAÇÃO	PAR DE ATRIBUTOS	ATRIBUTO DEFINIDO	ATRIBUTO EXCLUÍDO
+/- 0.6 a +/- 0.7	V e A	V	A (Já excluído)
	I_efe e A	Ambos excluídos	Ambos excluídos
	DS e A	Ambos excluídos	Ambos excluídos
	DG e A	Ambos excluídos	Ambos excluídos
	Q1p e A	Ambos excluídos	Ambos excluídos
	QF e Q	QF	Q (Já excluído)
	Im_Sc e Q	Im_Sc	Q (Já excluído)
	Ie e V	V	Ie (Já excluído)
	Se e V	V	Se (Já excluído)
	S e V	V	S (Já excluído)
	I_eficaz e V	V	I_eficaz (Já excluído)
	Q1p e QF	QF	Q1p (Já excluído)
	DS e QF	QF	DS (Já excluído)
	ircR e QF	QF	ircR (Já excluído)
	Ie e I_efe	Ambos excluídos	Ambos excluídos
	Se e I_efe	Ambos excluídos	Ambos excluídos
	S e I_efe	Ambos excluídos	Ambos excluídos
	I_eficaz e I_efe	Ambos excluídos	Ambos excluídos
	DS e I_eficaz	Ambos excluídos	Ambos excluídos
	DG e I_eficaz	Ambos excluídos	Ambos excluídos
	Q1p e I_eficaz	Ambos excluídos	Ambos excluídos
	Im_Sc e iacR	Ambos excluídos	Ambos excluídos
	vaf e iacR	vaf	iacR (Já excluído)
	DS e iscR	iscR	DS (Já excluído)
	Im_Sc e iscR	iscR	Im_Sc
	ircR e iscR	iscR	ircR (Já excluído)
	Im_Sc e ircR	Ambos excluídos	Ambos excluídos
	S e ircR	Ambos excluídos	Ambos excluídos
	S e DS	Ambos excluídos	Ambos excluídos
	S e DG	Ambos excluídos	Ambos excluídos
	S e Q1p	Ambos excluídos	Ambos excluídos
	Se e DS	Ambos excluídos	Ambos excluídos
	Ie e Ds	Ambos excluídos	Ambos excluídos
	DG e Im_Sc	Im_Sc	DG (Já excluído)
DG e Ie	Ambos excluídos	Ambos excluídos	
DG e Se	Ambos excluídos	Ambos excluídos	
Ie e Q1p	Ambos excluídos	Ambos excluídos	

**Tabela 24 - Correlação observada a partir da Matriz de correlação e otimizações propostas (continuação)**

ÍNDICE DE CORRELAÇÃO	PAR DE ATRIBUTOS	ATRIBUTO DEFINIDO	ATRIBUTO EXCLUÍDO
+/- 0.6 a +/- 0.7	Se e Q1p	Ambos excluídos	Ambos excluídos
	iaf e vaf	vaf	iaf
+/- 0.5	ircR e A	Ambos excluídos	Ambos excluídos
	Im_Sc e P	P	Im_Sc (Já excluído)
	PF e Q	PF	Q (Já excluído)
	PF1 e Q	Ambos excluídos	Ambos excluídos
	Pfe e Q	Ambos excluídos	Ambos excluídos
	Im_Sc e V	V	Im_Sc (Já excluído)
	Im_Sc e I_efe	Ambos excluídos	Ambos excluídos
	Im_Sc e PF1	Ambos excluídos	Ambos excluídos
	Im_Sc e Re_Sc	Ambos excluídos	Ambos excluídos
	Im_Sc e P_mean.out	Ambos excluídos	Ambos excluídos
	Im_Sc e Q1P	Ambos excluídos	Ambos excluídos
	iscR e V	V	iscR
	QF e V	V	QF
	iscR e PF	PF	iscR (Já excluído)
	ircR e PF	PF	ircR (Já excluído)
	THDia e PF	PF	THDia
	DG e QF	Ambos excluídos	Ambos excluídos
	iscR e DG	Ambos excluídos	Ambos excluídos
	iscR e QF	Ambos excluídos	Ambos excluídos
	I_efe e QF	Ambos excluídos	Ambos excluídos
	PF1 e VF	Ambos excluídos	Ambos excluídos
	Re_Sc e DS	Ambos excluídos	Ambos excluídos
	iscR e I_efe	Ambos excluídos	Ambos excluídos
	PF1 e iscR	Ambos excluídos	Ambos excluídos
	Pfe e iscR	Ambos excluídos	Ambos excluídos
	Pfe e ircR	Ambos excluídos	Ambos excluídos
	PF1 e ircR	Ambos excluídos	Ambos excluídos
	Se e ircR	Ambos excluídos	Ambos excluídos
	Q1p e PF1	Ambos excluídos	Ambos excluídos
	THDia e PF1	Ambos excluídos	Ambos excluídos
THDia e Pfe	Ambos excluídos	Ambos excluídos	

A Tabela 25 exemplifica a redução proposta identificando os atributos que permanecem no banco de dados em cada uma das simulações realizadas com base na

colinearidade. Considerando primeiro a retirada de pares com colinearidade maior ou igual a 80% (Simulação 3), num segundo momento, colinearidade maior ou igual a 65% (Simulação 4) e por fim, pares com colinearidade maiores ou iguais a 50% (Simulação 5).

**Tabela 25 - Tabela de atributos definidos à partir das correlações observadas**

SIMULAÇÃO 3			SIMULAÇÃO 4			SIMULAÇÃO 5		
CPT	CPC	IEEE	CPT	CPC	IEEE	CPT	CPC	IEEE
P	iscR	Ve	P	iscR	Ve	P	igR	Ve
V	igR	THDia	V	igR	THDia	V	PF	vaf
PF	PF	vaf	PF	PF	vaf	PF		
QF	Im_Sc	iaf	QF			3	2	2
4	4	4	4	3	3	Total atributos: 7		
Total atributos: 12			Total atributos: 10					

#### 4.3.1. Validação da performance dos algoritmos de classificação

A Tabela 26 apresenta as informações dos diferentes cenários utilizados empregados para a avaliação e validação das propostas dos algoritmos de classificação, descrevendo o banco de dados com as especificações dos atributos que os compõe.

Como os algoritmos são sensíveis à variação de seus parâmetros principais para cada caso, a *Gridsearch* é realizada considerando-se cada algoritmo de inteligência artificial, facilmente implementada utilizando-se *python*. Esta busca inteligente é utilizada para definir os parâmetros dos algoritmos que possibilitarão o melhor desempenho. A busca testa diversos valores para os parâmetros (definidos pelo usuário) calculando a precisão do algoritmo, com isso, é possível verificar os parâmetros que apresentarão melhor precisão.

A Tabela 26 apresenta todas as análises que foram realizadas com base nas simulações descritas na Tabela 25 para definição dos melhores parâmetros com objetivo final de obter alta performance nas simulações realizadas.

**Tabela 26 - Simulações e banco de dados utilizado**

<b>SIMULAÇÃO</b>	<b>BANCO DE DADOS</b>	<b>ATRIBUTOS</b>	<b>QUANTIDADE</b>	<b>PARÂMETROS</b>
1	Inicial	<b>CPT:</b> A, P, Q, V, PF, QF, VF, I_efe, I_eficaz, U_eficaz, N, Nr, Na, NF, THD_ia, THD_ib, THD_ic, THD_i_med, THD_ua, THD_ub, THD_uc, THD_u_med (22) <b>CPC:</b> iacR, iscR, ircR, igR, iucR, iacS, iscS, ircS, igS, iucS, iacT, iscT, ircT, igT, iucT, S, DS, DU, DG, PF, Re_Sc, Im_Sc (22) <b>IEEE:</b> P_mean_out, Q1p, Se, Ve, Ie, PF1p, PF1, Pfe, THDia, vaf, iaf (11)	55	<b>k-NN :</b> k= 3 <b>SVM :</b> C=1 , gamma = 0,0001 <b>SVM radial:</b> C=10 , gamma = 0,0001 <b>RF:</b> trees=40, random_state=50 <b>DT:</b> critério=gini , depth = 10
2	Retirada de Valores Nulos / Invariantes	<b>CPT:</b> A, P, Q, V, PF, QF, VF, I_efe, I_eficaz, U_eficaz (10) <b>CPC:</b> iacR, iscR, ircR, igR, S, DS, DG, PF, Re_Sc, Im_Sc (10) <b>IEEE:</b> P_mean_out, Q1p, Se, Ve, Ie, PF1, Pfe, THDia, vaf, iaf (10)	30	<b>k-NN :</b> k= 1 <b>SVM :</b> C=1 , gamma = 0,0001 <b>SVM radial:</b> C=10 , gamma = 0,0001 <b>RF:</b> trees=40, random_state=15 <b>DT:</b> critério=gini , depth = 12
3	Retirada de Correlação maior ou igual a 80%	<b>CPT:</b> P, V, PF, QF (4) <b>CPC:</b> iscR, igR, PF, Im_Sc (4) <b>IEEE:</b> Ve, THDia, vaf, iaf (4)	12	<b>k-NN :</b> k= 1 <b>SVM :</b> C=1 , gamma = 0,0001 <b>SVM radial:</b> C=10 , gamma = 0,0001 <b>RF:</b> trees=30, random_state=15 <b>DT:</b> critério=gini , depth = 12
4	Retirada de Correlação maior ou igual a 65%	<b>CPT:</b> P, V, PF, QF (4) <b>CPC:</b> iscR, igR, PF(3) <b>IEEE:</b> Ve, THDia, vaf (3)	10	<b>k-NN :</b> k= 1 <b>SVM :</b> C=1 , gamma = 0,0001 <b>SVM radial:</b> C=10 , gamma = 0,0001 <b>RF:</b> trees=30, random_state=20 <b>DT:</b> critério=entropy , depth = 10
5	Retirada de Correlação maior ou igual a 50%	<b>CPT:</b> P, V, PF (3) <b>CPC:</b> igR, PF(2) <b>IEEE:</b> Ve, vaf (2)	7	<b>k-NN :</b> k= 1 <b>SVM :</b> C=1 , gamma = 0,0001 <b>SVM radial:</b> C=10 , gamma = 0,0001 <b>RF:</b> trees=20, random_state=20 <b>DT:</b> critério=gini , depth = 12

#### 4.3.1.1. Validação da performance - SIMULAÇÃO 1

A Tabela 27 apresenta informações sobre a performance dos algoritmos considerando os parâmetros indicados e tomando como base o banco de dados inicial composto por 55 atributos.

**Tabela 27 - Performance dos algoritmos - Banco de dados inicial - Simulação 1**

Medições	k-NN	SVM (Llinear)	SVM (RBF)	DT	RF
<i>Melhores Parâmetros</i>	<i>k=3</i>	<i>C = 1 gamma= 0,0001 kernel = Linear</i>	<i>C = 10 gamma= 0,0001 kernel = RBF</i>	<i>Critério = Gini Depth = 10</i>	<i>trees = 40 random state = 50</i>
Precisão	1	0,97	0,97	0,98	0,98
Recall	1	0,97	0,96	0,98	0,98
F-Score	1	0,97	0,96	0,98	0,98
Ttest(ms)	15,62	23,56	32,92	126,32	328,1

Neste caso, observa-se que o menor tempo para teste do algoritmo se encontra no k-NN, uma vez que para este algoritmo não é necessário um treinamento e por isso é conhecido como “algoritmo preguiçoso”. Além disso, a melhor precisão na classificação também é atribuída ao k-NN, chegando a 100%.

Considerando individualmente cada algoritmo, pode-se analisar a performance individual obtida para cada classe de carga elétrica utilizada como referência. A Tabela 28, representa a performance coletada a partir da Simulação 1, realizada utilizando o algoritmo k-NN, com k definido no valor de 3.

Percebe-se que todos os equipamentos são classificados com precisão de 100%, ou seja, neste momento, o k-NN é uma excelente recomendação como algoritmo para realizar a classificação de cargas considerando o banco de dados inicial apresentado.

**Tabela 28 - Performance do k-NN para cada classe - Simulação 1**

k-NN (k=3)				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	12
6	1	1	1	12
7	1	1	1	11
10	1	1	1	12
15	1	1	1	12
28	1	1	1	12
57	1	1	1	12
66	1	1	1	12
80	1	1	1	11
136	1	1	1	12



Analogamente, a Tabela 29 mostra informações da performance coletada a partir da Simulação 1, realizada utilizando-se o algoritmo SVM Linear, com C definido no valor de 1 e gamma como 0,0001.

**Tabela 29 - Performance do SVM Linear para cada classe - Simulação 1**

SVM Linear				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	0,9	1	0,95	18
7	1	0,91	0,95	11
10	1	0,86	0,92	14
15	1	1	1	13
28	0,89	1	0,94	8
57	1	0,91	0,95	11
66	1	1	1	9
80	1	1	1	9
136	0,94	1	0,97	16

É possível notar que neste caso, temos precisões acima de 90% para todas as classes, com exceção da classe 28 (Notebook) que foi classificada com 89% de precisão. No total geral, o algoritmo SVM linear possui uma precisão de 97% na classificação das cargas elétricas apresentadas, considerando o banco de dados inicialmente levantado e analisado. Isto indica que as instâncias são linearmente separáveis entre si.

O mesmo se repete para o SMV Radial e a Tabela 30, representa a performance do algoritmo considerando-se a Simulação 1, definindo C no valor de 10 e gamma como 0,0001.

**Tabela 30 - Performance do SVM Radial para cada classe - Simulação 1**

SVM RBF				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	1	1	1	18
7	0,69	1	0,81	11
10	1	0,79	0,88	14
15	1	1	1	13

**Tabela 30 - Performance do SVM Radial para cada classe - Simulação 1 (continuação)**

SVM RBF				
Classe	Precisão	Recall	f-Score	support
28	1	1	1	8
57	1	0,82	0,9	11
66	1	1	1	9
80	1	1	1	9
136	1	1	1	16

Para este caso, destaca-se uma baixa precisão de classificação da carga 7 (Secador de cabelos). Dentre as performances globais, o SVM Radial é o que obteve menor média de precisão. Por este motivo, entende-se que não seria o mais adequado para o banco de dados considerado.

A análise também foi realizada para o algoritmo *DT* representada por meio da Tabela 31, considerando critério de ramificação *gini* e profundidade da árvore igual a 10 (parâmetros definidos a partir da *Search grid*). Nota-se que todas as classes são bem classificadas pelo DT, tendo como menor precisão apresentada a classe 28 (Notebook). É importante notar que o notebook já apresentou dificuldade para ser classificado com o SVM Linear, e por este motivo, entende-se que os atributos podem estar sendo confundidos com as demais classes.

**Tabela 31 - Performance do DT para cada classe - Simulação 1**

DT				
Classe	Precisão	Recall	f-Score	support
1	0,9	1	0,95	9
6	1	1	1	18
7	1	1	1	11
10	1	1	1	14
15	1	1	1	13
28	0,89	1	0,94	8
57	1	0,82	0,9	11
66	1	1	1	9
80	1	1	1	9
136	1	1	1	16

Por fim, apresenta-se na Tabela 32 os resultados obtidos para o *Random Forest* na Simulação 1, realizada considerando número de árvores igual a 40 e estado randômico igual a 50.

**Tabela 32 - Performance do RF para cada classe - Simulação 1**

RF				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	1	1	1	18
7	0,85	1	0,92	11
10	1	1	1	14
15	1	1	1	13
28	1	1	1	8
57	1	0,82	0,9	11
66	1	1	1	9
80	1	1	1	9
136	1	1	1	16

É possível verificar que, assim como os demais algoritmos, o RF foi capaz de classificar com alta precisão todas as classes utilizadas como referência no estudo, exceto a classe 7 (Secador de Cabelos). Dentre todos os algoritmos observados, para o banco de dados considerado, o RF é o algoritmo que possui maior tempo de treinamento observado, porém, numa análise geral, após o k-NN, a melhor performance global observada está atrelada ao mesmo.

#### 4.3.1.2. Validação da performance - SIMULAÇÃO 2

As informações da performance dos algoritmos considerando os parâmetros indicados e tomando como base o banco de dados otimizado com a exclusão de atributos nulos e/ou invariantes composto no total por 30 atributos são apresentadas na Tabela 33.

**Tabela 33 - Performance dos algoritmos-Banco de dados otimização 1 - Simulação 2**

Medições	k-NN	SVM (Linear)	SVM (RBF)	DT	RF
<i>Melhores Parâmetros</i>	<i>k=1</i>	<i>C = 1 gamma= 0,0001 kernel = Linear</i>	<i>C = 10 gamma= 0,0001 kernel = RBF</i>	<i>Critério = Gini Depth = 12</i>	<i>tress = 40 random state = 15</i>
Precisão	0,99	0,97	0,97	0,98	0,98
Recall	0,99	0,97	0,96	0,98	0,98
F-Score	0,99	0,97	0,96	0,98	0,98
Ttest(s)	15,08	28,62	40,03	129,12	384,79

O k-NN se mantém como algoritmo com menor tempo de teste e com uma performance geral alta no valor de 99%.

A Tabela 34 mostra os resultados obtidos a partir da Simulação 2, realizada utilizando-se o algoritmo k-NN, com k definido no valor de 1.

**Tabela 34 - Performance do k-NN para cada classe - Simulação 2**

k-NN (k=1)				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	12
6	1	1	1	12
7	1	1	1	11
10	0,92	1	0,96	12
15	1	0,91	0,95	11
28	1	1	1	12
57	1	1	1	12
66	1	1	1	12
80	1	1	1	12
136	1	1	1	12

Se comparado à simulação 1, houve uma pequena perda na precisão de 100% para 99%, porém, entende-se que não é uma redução significativa que impactará na aplicação. Neste momento, a classe 10 (micro-ondas) passa a ser classificada com uma precisão de 92%.

A Tabela 35, representa a performance coletada à partir da Simulação 2, realizada utilizando-se o algoritmo SVM Linear, com C definido no valor de 1 e gamma como 0,0001.

Para o SVM Linear, não é observado uma melhora significativa de performance do algoritmo se comparado com a Simulação 1 realizada anteriormente.

**Tabela 35 - Performance do SVM Linear para cada classe - Simulação 2**

SVM Linear				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	0,9	1	0,95	18
7	1	0,91	0,95	11
10	1	0,86	0,92	14
15	1	1	1	13
28	0,89	1	0,94	8
57	1	0,91	0,95	11
66	1	1	1	9
80	1	1	1	9
136	0,94	1	0,97	16

De forma análoga, o comportamento se repete para o SMV Radial e a Tabela 36, apresenta a performance do algoritmo considerando-se a Simulação 2, definindo C no valor de 10 e gamma como 0,0001.

**Tabela 36 - Performance do SVM Radial para cada classe - Simulação 2**

SVM RBF				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	1	1	1	18
7	0,69	1	0,81	11
10	1	0,79	0,88	14
15	1	1	1	13
28	1	1	1	8
57	1	0,82	0,9	11
66	1	1	1	9
80	1	1	1	9
136	1	1	1	16

Assim como para o SVM Linear, para o SVM Radial também não é observado uma melhora significativa de performance do algoritmo se comparado com a Simulação 1. Neste cenário, infere-se que a redução não contribuiu para a performance do algoritmo SVM, seja ele Linear ou Radial.

Posteriormente, para o algoritmo *DT* a Tabela 37 representa sua performance obtida, considerando critério de ramificação *gini* e profundidade da árvore 12.

**Tabela 37 - Performance do DT para cada classe - Simulação 2**

DT				
Classe	Precisão	Recall	f-Score	support
1	0,9	1	0,95	9
6	1	1	1	18
7	1	1	1	11
10	1	1	1	14
15	1	1	1	13
28	0,89	1	0,94	8
57	1	0,82	0,9	11
66	1	1	1	9
80	1	1	1	9
136	1	1	1	16

De igual forma, assim como para o SVM Linear e RBF, para o DT não é observado alterações na classificação se comparado à Simulação 1 realizada anteriormente.

Por fim, apresenta-se na Tabela 38 os resultados obtidos para o *Random Forest* na Simulação 2, realizada considerando número de árvores 40 e estado randômico 15.

**Tabela 38 - Performance do RF para cada classe - Simulação 2**

RF				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	1	1	1	18
7	0,85	1	0,92	11
10	1	1	1	14
15	1	1	1	13
28	1	1	1	8
57	1	0,82	0,9	11

**Tabela 38 - Performance do RF para cada classe - Simulação 2 (continuação)**

RF				
Classe	Precisão	Recall	f-Score	support
66	1	1	1	9
80	1	1	1	9
136	1	1	1	16

Assim, comprova-se que a retirada dos atributos nulos não apresenta melhorias na atividade de classificação. Em uma visão geral, nota-se que não há alterações significativas. O tempo de teste dos algoritmos é um pouco mais elevado, exceto para o k-NN que é ligeiramente menor.

Assim como na simulação 1, na simulação 2, entende-se que o k-NN seria mais bem aplicado ao banco de dados em questão para a classificação de cargas com maior precisão observada e com o menor tempo de teste do algoritmo.

#### 4.3.1.3. Validação da performance - SIMULAÇÃO 3

A Tabela 39 considera a performance dos algoritmos e seus respectivos parâmetros definidos, simulando a classificação de cargas feita a partir do banco de dados otimizado com a exclusão de atributos com correlação maior ou igual a 80% composto no total por 12 atributos.

**Tabela 39 - Performance dos algoritmos-Banco de dados otimização 2 - Simulação 3**

Medições	k-NN	SVM (Linear)	SVM (RBF)	DT	RF
<i>Melhores Parâmetros</i>	<i>k=1</i>	<i>C = 1 gamma= 0,0001 kernel = Linear</i>	<i>C = 10 gamma= 0,0001 kernel = RBF</i>	<i>Critério = Gini Depth = 12</i>	<i>tress = 30 random state = 15</i>
Precisão	0,99	0,98	0,96	0,98	0,98
Recall	0,99	0,98	0,96	0,98	0,98
F-Score	0,99	0,98	0,96	0,98	0,98
Ttest(ms)	< 1	9,92	10,09	5,02	369,82

Para mais um dos casos observados, o k-NN se mantém como algoritmo com menor tempo de teste e apresenta a melhor performance global observada, em 99%. Neste caso, é possível concluir que a redução do banco de dados pode de fato, trazer melhorias à aplicação sem afetar a performance de classificação das cargas elétricas.

Fazendo uma análise individual para cada algoritmo, observa-se nas tabelas posteriormente apresentadas, as vantagens e desvantagens observadas para a otimização do banco de dados proposta neste momento. A Tabela 40, representa a performance coletada a partir da Simulação 3, realizada utilizando-se o algoritmo k-NN, com k definido no valor de 1.

**Tabela 40 - Performance do k-NN para cada classe - Simulação 3**

k-NN (k=1)				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	12
6	1	1	1	12
7	1	1	1	12
10	1	1	1	12
15	1	1	1	11
28	1	1	1	11
57	0,92	1	0,96	12
66	1	0,92	0,96	12
80	1	1	1	12
136	1	1	1	12

Neste momento o algoritmo é capaz de classificar 9 das 10 classes apresentadas ao estudo. A classe que possui menor precisão na classificação é a 57 (Lâmpada Incandescente) com 92%.

A performance obtida para o SVM Linear é observada na Tabela 41, considerando a Simulação 3, com C definido no valor de 1 e gamma como 0,0001.



**Tabela 41 - Performance do SVM Linear para cada classe - Simulação 3**

SVM Linear				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	0,95	1	0,97	18
7	1	0,91	0,95	11
10	1	0,93	0,96	14
15	1	1	1	13
28	0,89	1	0,94	8
57	1	1	1	11
66	1	1	1	9
80	1	1	1	9
136	1	1	1	16

Para o SVM Linear, se comparado as Simulações 1 e 2, o processo tem uma melhora significativa, passando classificar com precisão de 100% oito das dez classes apresentadas. De forma análoga, a constatação se repete para o SMV Radial, em que a Tabela 42 representa a performance do algoritmo considerando-se a Simulação 2, definindo C no valor de 10 e gamma como 0,0001.

**Tabela 42 - Performance do SVM Radial para cada classe - Simulação 3**

SVM RBF				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	1	1	1	18
7	1	0,91	0,95	11
10	1	0,86	0,92	14
15	1	1	1	13
28	1	1	1	8
57	1	0,82	0,9	11
66	0,64	1	0,78	9
80	1	1	1	9
136	1	1	1	16

Assim como na Simulação 2, para o SVM Radial não é observado uma melhora significativa de performance do algoritmo se comparado com a Simulação 1 e 2 realizada anteriormente. Neste momento, ele passa a ter baixa precisão de classificação na classe

66 (Ventilador) ao invés da classe 7 (secador de cabelos) conforme observado anteriormente.

O algoritmo *DT* também foi verificado para este cenário e sua performance é representada por meio da Tabela 43, considerando critério de ramificação *gini* e profundidade da árvore 12.

**Tabela 43 - Performance do DT para cada classe - Simulação 3**

DT				
Classe	Precisão	Recall	f-Score	support
1	0,9	1	0,95	9
6	1	1	1	18
7	1	1	1	11
10	1	1	1	14
15	1	1	1	13
28	0,89	1	0,94	8
57	1	0,82	0,9	11
66	1	1	1	9
80	1	1	1	9
136	1	1	1	16

Comparando-se com a Simulação 2, não há alteração nas precisões de classificação para este algoritmo. Observa-se uma diminuição no tempo de teste considerando um banco de dados consideravelmente reduzido.

Por fim, apresenta-se na Tabela 44 os resultados obtidos para o *Random Forest* na Simulação 3, realizada considerando número de árvores 30 e estado randômico 15.

**Tabela 44 - Performance do RF para cada classe - Simulação 3**

RF				
Classe	Precisão	Recall	f-Score	support
1	0,9	1	0,95	9
6	1	1	1	18
7	1	1	1	11
10	1	0,93	0,96	14
15	1	1	1	13
28	0,89	1	0,94	8

**Tabela 44 - Performance do RF para cada classe - Simulação 3 (continuação)**

RF				
Classe	Classe	Classe	Classe	Classe
57	1	0,91	0,95	11
66	1	1	1	9
80	1	1	1	9
136	1	1	1	16

Comparando-se à Simulação 2, tem-se uma ligeira diminuição do tempo de teste, e 2 classes passam a ter precisão de classificação menores: 1 (lâmpada fluorescente compacta) e 28 (notebook).

#### 4.3.1.4. Validação da performance - SIMULAÇÃO 4

A Tabela 45 expõe a performance dos algoritmos já apresentados anteriormente, considerando seus respectivos parâmetros definidos conforme orientado, na classificação feita ao banco de dados otimizado com a exclusão de atributos com correlação maior ou igual a 65% composto no total por 10 atributos.

**Tabela 45 - Performance dos algoritmos-Banco de dados otimização 3 - Simulação 4**

Medições	k-NN	SVM (Linear)	SVM (RBF)	DT	RF
<i>Melhores Parâmetros</i>	<i>k=1</i>	<i>C = 1 gamma= 0,0001 kernel = Linear</i>	<i>C = 10 gamma= 0,0001 kernel = RBF</i>	<i>Critério = Entropy Depth = 10</i>	<i>trees = 30 random state = 20</i>
Precisão	0,99	0,99	0,96	0,98	0,98
Recall	0,99	0,99	0,96	0,98	0,98
F-Score	0,99	0,99	0,96	0,98	0,98
Ttest(ms)	< 1	189,77	7	2	280,25

Observa-se que o k-NN se mantém com melhor precisão de classificação com menor tempo de teste se comparado aos demais algoritmos estudados. À medida que as otimizações são realizadas, reduz-se o número de atributos existentes no banco de dados, bem como, o esforço necessário do *hardware* para o momento dos cálculos dos índices de potência e corrente. Por este motivo, além da precisão, é preciso verificar o

tempo necessário para a realização da classificação, pois este índice compõe a performance como um todo a ser analisada, e impactará a aplicação no sentido de obter a informação em tempo real.

Conforme simulações anteriores, a análise individual dos algoritmos foi também realizada para este caso. A Tabela 46, representa a performance coletada à partir da Simulação 4, realizada utilizando-se o algoritmo k-NN, com k definido no valor de 1.

**Tabela 46 - Performance do k-NN para cada classe - Simulação 4**

k-NN (k=1)				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	12
6	1	1	1	12
7	1	1	1	11
10	1	1	1	12
15	1	1	1	12
28	1	1	1	12
57	0,92	1	0,96	12
66	1	0,92	0,96	12
80	1	1	1	11
136	1	1	1	12

Comparando-se com a Simulação 3, não houve alteração nas precisões observadas.

A Tabela 47 representa a performance coletada a partir da Simulação 4, realizada utilizando-se o algoritmo SVM Linear, com C definido no valor de 10 e gamma como 0,0001.

**Tabela 47 - Performance do SVM Linear para cada classe - Simulação 4**

SVM Linear				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	1	1	1	18
7	1	1	1	11
10	1	0,93	0,96	14
15	1	1	1	13
28	0,89	1	0,94	8
57	1	1	1	11

**Tabela 47 - Performance do SVM Linear para cada classe - Simulação 4  
(continuação)**

SVM Linear				
Classe	Precisão	Recall	f-Score	support
66	1	1	1	9
80	1	1	1	9
136	1	1	1	16

Em relação às simulações anteriores, observa-se que a precisão de classificação para o SVM Linear aumentou. Este fato é interessante, pois comprova que quanto mais reduzido o banco de dados apresentado, os dados estão se tornando linearmente separáveis dado a boa precisão do algoritmo.

Repete-se a análise ao SMV Radial e a Tabela 48, representa a performance do algoritmo considerando-se a Simulação 4, definindo C no valor de 10 e gamma como 0,0001.

**Tabela 48 - Performance do SVM Radial para cada classe - Simulação 4**

SVM RBF				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	1	1	1	18
7	1	0,91	0,95	11
10	1	0,86	0,92	14
15	1	1	1	13
28	1	1	1	8
57	1	0,82	0,9	11
66	0,64	1	0,78	9
80	1	1	1	9
136	1	1	1	16

Comparado com a Simulação 3, não há alterações observadas na performance de classificação do algoritmo. A redução de oito atributos não impactou o mesmo.

Estende-se a análise para o algoritmo *DT* representada por meio da Tabela 49, considerando critério de ramificação *entropy* (definido pela *Search grid*) e profundidade da árvore 10.

**Tabela 49 - Performance do DT para cada classe - Simulação 4**

DT				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	1	1	1	18
7	1	1	1	11
10	1	1	1	14
15	1	1	1	13
28	1	1	1	8
57	1	0,82	0,9	11
66	1	1	1	9
80	0,9	1	0,95	9
136	0,94	1	0,97	16

Comparando-se com a Simulação 3, o algoritmo não perde em precisão global e reduz o tempo de teste em aproximadamente, metade. O algoritmo passa a classificar com menores precisões as classes 80 (máquina de lavar roupas) e 136 (aquecedor) neste caso.

A Tabela 50 expõe os resultados obtidos para o *Random Forest* na Simulação 4, realizada considerando número de árvores 30 e estado randômico 20.

Comparando-se à Simulação 3, tem-se uma redução no tempo de teste e o número de classes classificadas com alta precisão se mantém o mesmo de 8 entre 10.

**Tabela 50 - Performance de classificação do RF para cada classe - Simulação 4**

RF				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	0,97	1	1	18
7	1	1	1	11
10	1	1	0,93	14
15	1	1	1	13
28	1	0,89	1	8
57	0,9	1	0,91	11
66	1	1	1	9
80	1	1	1	9
136	1	0,94	1	16

#### 4.3.1.5. Validação da performance - SIMULAÇÃO 5

A Tabela 51 apresenta a performance dos algoritmos considerando a última otimização do banco de dados proposta com a exclusão de atributos com correlação maior ou igual a 50%, composto no total por 7 atributos.

**Tabela 51 - Performance dos algoritmos-Banco de dados otimização 4 - Simulação 5**

Medições	k-NN	SVM (Linear)	SVM (RBF)	DT	RF
<i>Melhores Parâmetros</i>	<i>k=1</i>	<i>C = 1 gamma= 0,0001 kernel = Linear</i>	<i>C = 10 gamma= 0,0001 kernel = RBF</i>	<i>Critério = gini Depth = 12</i>	<i>trees = 20 random state = 20</i>
Precisão	1	0,99	0,98	0,99	0,99
Recall	1	0,99	0,99	0,99	0,99
F-Score	1	0,99	0,98	0,99	0,99
Ttest(ms)	< 1	124,96	9,97	< 1	260,66

Observa-se que para este caso, o RF apresenta melhor precisão na classificação, porém com tempo de teste 260 vezes maior que o k-NN.

Novamente, realizando-se a verificação individual da performance dos algoritmos, a Tabela 52, representa a performance coletada a partir da Simulação 5, realizada utilizando-se o algoritmo k-NN, com k definido no valor de 1.

**Tabela 52 - Performance do k-NN para cada classe - Simulação 5**

k-NN (k=1)				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	12
6	1	1	1	12
7	1	1	1	11
10	1	1	1	12
15	1	1	1	12
28	1	1	1	12
57	1	1	1	12
66	1	1	1	11
80	1	1	1	12
136	1	1	1	12

Nota-se que para o k-NN, tanto na simulação 1 quanto na simulação 5, há a precisão de classificação de 100%, ou seja, reduzindo o banco de dados de 55 atributos para 7, não há perdas na performance deste algoritmo. Além disso, devido à redução significativa, o esforço computacional para teste do algoritmo é substancialmente reduzido. Neste sentido, considerando todas as vantagens estudadas no k-NN, e comprovando-as nos testes, esta pode ser definida como uma boa opção para a aplicação.

Deste mesmo modo, mantém-se a análise utilizando-se o algoritmo SVM Linear, com C definido no valor de 1 e gamma como 0,0001 e os resultados são apresentados na Tabela 53.

Não se observa alterações na classificação entre as Simulações 4 e 5.

**Tabela 53 - Performance do SVM Linear para cada classe - Simulação 5**

SVM Linear				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	1	1	1	18
7	1	1	1	11
10	1	0,93	0,96	14
15	1	1	1	13
28	0,89	1	0,94	8
57	1	1	1	11
66	1	1	1	9
80	1	1	1	9
136	1	1	1	16

A simulação foi também realizada para o SMV Radial, a performance do algoritmo definindo C no valor de 10 e gamma como 0,0001 é apresentada na Tabela 54.



**Tabela 54 - Performance do SVM Radial para cada classe - Simulação 5**

SVM RBF				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	1	1	1	18
7	0,92	1	0,96	11
10	1	0,86	0,92	14
15	1	1	1	13
28	0,89	1	0,94	8
57	1	1	1	11
66	1	1	1	9
80	1	1	1	9
136	1	1	1	16

O algoritmo neste momento passa a ter precisões mais baixas em 2 das 10 classes apresentadas, porém, estas precisões não são inferiores a 89%.

Seguindo a sequência de verificações realizadas, a performance do algoritmo *DT* pode ser vista na Tabela 55, considerando critério de ramificação *gini* e profundidade da árvore 12.

**Tabela 55 - Performance do DT para cada classe - Simulação 5**

DT				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	1	1	1	18
7	1	1	1	11
10	1	1	1	14
15	1	1	1	13
28	1	1	1	8
57	1	0,91	0,95	11
66	1	1	1	9
80	0,9	1	0,95	9
136	1	1	1	16

Se comparado com a Simulação 4, o algoritmo passa a ter um aumento na precisão de classificação sendo capaz de classificar 9 das 10 classes com alta precisão.

Para finalizar, a Tabela 56 apresenta os resultados obtidos para o *Random Forest* na Simulação 5, realizada considerando número de árvores 40 e estado randômico 5.

**Tabela 56 - Performance do RF para cada classe - Simulação 5**

RF				
Classe	Precisão	Recall	f-Score	support
1	1	1	1	9
6	1	1	1	18
7	0,92	1	0,96	11
10	1	1	1	14
15	1	1	1	13
28	1	1	1	8
57	1	0,91	0,95	11
66	1	1	1	9
80	1	1	1	9
136	1	1	1	16

O algoritmo classifica com alta precisão 9 entre 10 classes. Entre a simulação 4 e 5, nota-se um aumento de performance, com precisão de 99% e tempo de teste de 260 ms.

#### **4.4. Otimização Do Banco De Dados - Definição Dos Atributos Para O Banco De Dados Final - Cenário Individual**

A fim de investigar a real necessidade de se considerar mais de uma teoria de potência para representar os fenômenos elétricos advindos dos comportamentos das cargas elétricas anteriormente estudadas, propõe-se uma análise individual, feita por cada teoria de potência, mantendo-se os procedimentos e análises aplicadas ao primeiro banco de dados geral levantado.

##### **4.4.1. Levantamento dos bancos de dados individualizados**

Assim como foi realizado o levantamento do banco de dados geral, compostos pelos índices de potência e corrente das teorias CPT, CPC e IEEE, subdivide-se o mesmo

em três novos bancos de dados individualizados por teoria de potência conforme pode ser observado na Tabela 57. Pode-se notar que neste primeiro momento, já está sendo considerado os dados não nulos, e para cada teoria, há um total de 10 atributos. Seguindo o procedimento já utilizado, o estudo inicia com base na simulação 2.

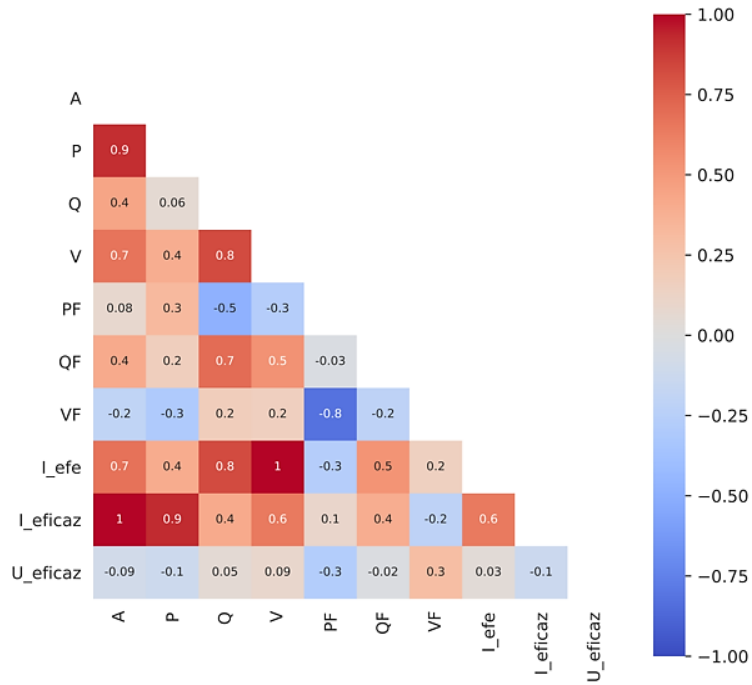
**Tabela 57 - Banco de dados individualizado**

SIMULAÇÃO 2 (sem nulos)		
CPT	CPC	IEEE
A	iacR	P_mean_ot
P	iscR	Q1p
Q	ircR	Se
V	igR	Ve
PF	S	le
QF	DS	PF1
VF	DG	Pfe
I_efe	PF	THDia
I_eficaz	Re_Sc	vaf
U_eficaz	Im_Sc	iaf
Total de atributos: 10	Total de atributos: 10	Total de atributos: 10

#### 4.4.2. Análise de relevância dos atributos do banco de dados

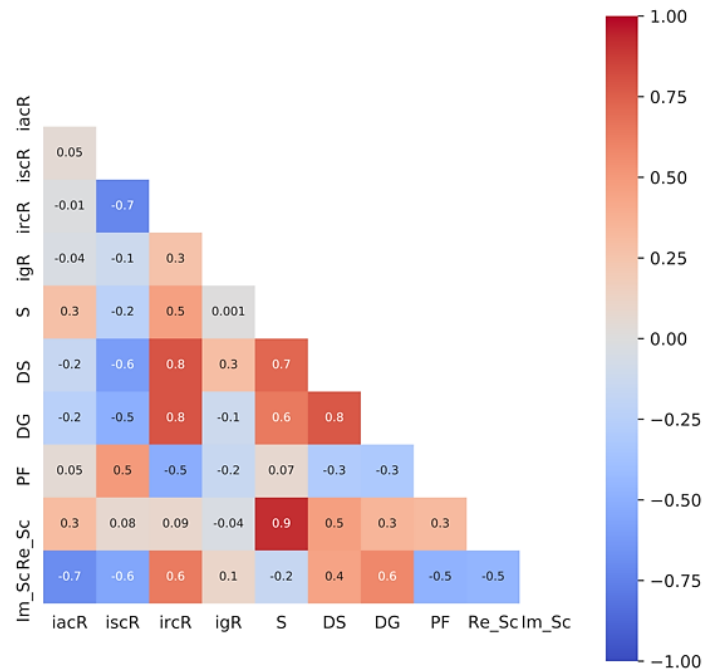
Após levantamento dos bancos de dados considerando os atributos não nulos, realizou-se a análise de relevância utilizando-se da matriz de correlação, seguindo o procedimento já utilizado. As figuras 34, 35 e 36, mostram a matriz de correlação observada para as teorias CPT, CPC e IEEE, respectivamente.

**Figura 34 - Matriz de correlação dos atributos calculados CPT**



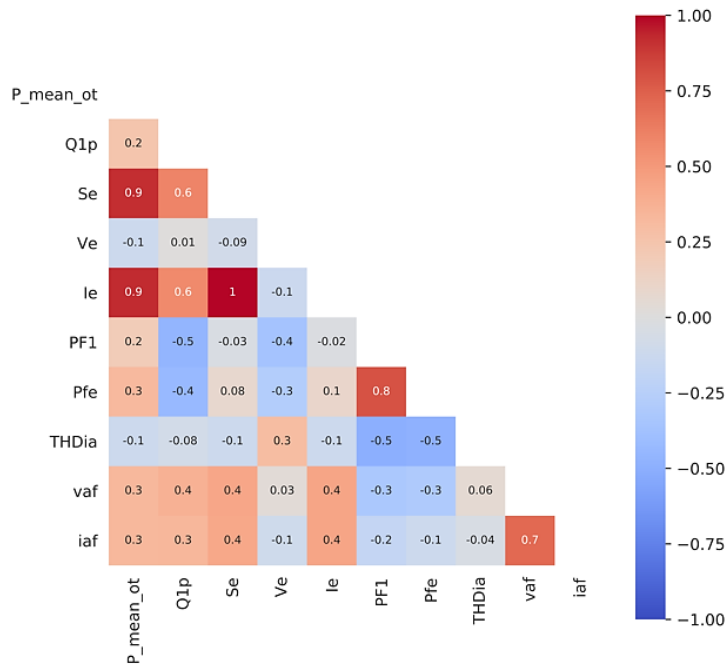
FONTE: AUTORIA PRÓPRIA, DESENVOLVIDO NO SOFTWARE SPYDER (ANACONDA)

**Figura 35 - Matriz de correlação dos atributos calculados CPC**



FONTE: AUTORIA PRÓPRIA, DESENVOLVIDO NO SOFTWARE SPYDER (ANACONDA)

**Figura 36 - Matriz de correlação dos atributos calculados IEEE**



FONTE: AUTORIA PRÓPRIA, DESENVOLVIDO NO SOFTWARE SPYDER (ANACONDA)

A partir das matrizes de correlação, foi individualmente analisado os pares com correlação superiores a 80%, 65% e 50%. A Tabela 58, evidencia por teoria de potência os pares e seus índices de correlação, bem como os atributos definidos como pertencentes ao banco de dados, ou retirados do mesmo.

**Tabela 58 - Correlação observada a partir da Matriz de correlação e otimizações propostas**

CPT			
ÍNDICE DE CORRELAÇÃO	PAR DE ATRIBUTOS	ATRIBUTO DEFINIDO	ATRIBUTO EXCLUÍDO
+/- 1	I_eficaz e A	A	I_eficaz
	I_efe e A	A	I_efe
+/- 0.8 a +/- 0.9	P e A	A	P
	I_eficaz e P	Ambos excluídos	Ambos excluídos
	V e Q	Q	V
	I_efe e Q	Q	I_efe(Já excluído)
	VF e PF	PF	VF

**Tabela 58 - Correlação observada a partir da Matriz de correlação e otimizações propostas (continuação)**

<b>CPT</b>			
<b>ÍNDICE DE CORRELAÇÃO</b>	<b>PAR DE ATRIBUTOS</b>	<b>ATRIBUTO DEFINIDO</b>	<b>ATRIBUTO EXCLUÍDO</b>
+/- 0.6 a +/- 0.7	V e A	A	V (Já excluído)
	I_efe e A	A	I_efe (Já excluído)
	QF e Q	Q	QF
	I_eficaz e V	Ambos excluídos	Ambos excluídos
	I_eficaz e I_efe	Ambos excluídos	Ambos excluídos
+/- 0.5	PF e Q	PF	Q
	QF e V	QF	V (Já excluído)
	I_efe e QF	Ambos excluídos	Ambos excluídos
<b>CPC</b>			
<b>ÍNDICE DE CORRELAÇÃO</b>	<b>PAR DE ATRIBUTOS</b>	<b>ATRIBUTO DEFINIDO</b>	<b>ATRIBUTO EXCLUÍDO</b>
+/- 1	N.A		
	N.A		
+/- 0.8 a +/- 0.9	DS e ircR	ircR	DS
	DG e ircR	ircR	DG
	Re_Sc e S	S	Re_Sc
	DG e DS	Ambos excluídos	Ambos excluídos
+/- 0.6 a +/- 0.7	Im_Sc e iacR	Im_Sc	iacR
	ircR e iscR	iscR	ircR
	DS e iscR	iscR	DS (Já excluído)
	im_Sc e iscR	Im_Sc	iscR
	Im_Sc e ircR	Im_Sc	ircR (Já excluído)
	DS e S	S	DS (Já excluído)
	DG e S	S	DG (Já excluído)
	Im_Sc e DG	Im_Sc	DG (Já excluído)
+/- 0.5	PF e iscR	PF	iscR (Já excluído)
	DG e iscR	Ambos excluídos	Ambos excluídos
	S e ircR	S	ircR (Já excluído)
	PF e ircR	PF	ircR (Já excluído)
	Re_Sc e DS	Ambos excluídos	Ambos excluídos
	Im_Sc e PF	PF	Im_Sc
	Im_Sc e Re_Sc	Ambos excluídos	Ambos excluídos

**Tabela 58 - Correlação observada a partir da Matriz de correlação e otimizações propostas (continuação)**

IEEE			
ÍNDICE DE CORRELAÇÃO	PAR DE ATRIBUTOS	ATRIBUTO DEFINIDO	ATRIBUTO EXCLUÍDO
+/- 1	le e Se	Se	le
+/- 0.8 a +/- 0.9	Se e P_mean.out	Se	P_mean.out
	le e P_mean.out	Ambos excluídos	Ambos excluídos
	Pfe e PF1	Pfe	PF1
+/- 0.6 a +/- 0.7	Se e Q1p	Se	Q1p
	le e Q1p	Ambos excluídos	Ambos excluídos
	iaf e vaf	vaf	iaf
+/- 0.5	PF1 e Q1p	Ambos excluídos	Ambos excluídos
	ThDia e PF1	ThDia	PF1 (Já excluído)
	ThDia e Pfe	ThDia	Pfe

A Tabela 59 exemplifica a redução proposta identificando os atributos que permanecem no banco de dados em cada uma das simulações realizadas com base na colinearidade, considerando primeiro a retirada de pares com colinearidade maior ou igual a 80% (Simulação 3), num segundo momento, colinearidade maior ou igual a 65% (Simulação 4) e por fim, pares com colinearidade maiores ou iguais a 50% (Simulação 5).

**Tabela 59 - Tabela de atributos definidos a partir das correlações observadas**

SIMULAÇÃO 3 (>80%)		
CPT	CPC	IEEE
A	iacR	Q1p
Q	iscR	Se
PF	ircR	Ve
QF	igR	Pfe
U_eficaz	S	THDia
Total de atributos: 5	PF	vaf
	Im_Sc	iaf
	Total de atributos: 7	Total de atributos: 7

**Tabela 59 - Tabela de atributos definidos a partir das correlações observadas (continuação)**

<b>SIMULAÇÃO 4 (&gt;65%)</b>		
<b>CPT</b>	<b>CPC</b>	<b>IEEE</b>
A	igR	Se
Q	S	Ve
PF	PF	Pfe
U_eficaz	Im_Sc	THDia
Total de atributos: 4	Total de atributos: 4	vaf
		Total de atributos: 5
<b>SIMULAÇÃO 5 (&gt;50%)</b>		
<b>CPT</b>	<b>CPC</b>	<b>IEEE</b>
A	igR	Se
PF	S	Ve
U_eficaz	PF	THDia
Total de atributos: 3	Total de atributos: 3	vaf
		Total de atributos: 4

Nota-se que, devido à correlação ter sido aplicada à um banco de dados menor, as correlações também são ligeiramente alteradas. Desta forma, os atributos retirados e finais a cada simulação realizada por teoria de potência, não são os mesmos observados na sessão anterior, na qual estava sendo analisado o banco de dados geral.

#### 4.4.3. Validação da performance dos algoritmos de classificação

A Tabela 60 apresenta os bancos de dados considerados para cada teoria de potência, executando as simulações de 2 a 5, analisando as performances apenas dos algoritmos k-NN e RF. Na tabela 60, é possível identificar também, por teoria de potência, quais os melhores parâmetros sugeridos pela *Search grid* para cada banco de dados baseado nas teorias de potências individuais.



**Tabela 60 - Simulações, banco de dados e parâmetros utilizados**

<b>SIMULAÇÃO</b>	<b>ATRIBUTOS (quantidade)</b>	<b>CPT PARÂMETROS</b>	<b>CPC PARÂMETROS</b>	<b>IEEE PARÂMETROS</b>
2	<b>CPT:</b> A, P, Q, V, PF, QF, VF, I_efe, I_eficaz, U_eficaz (10) <b>CPC:</b> iacR, iscR, ircR, igR, S, DS, DG, PF, Re_Sc, Im_Sc (10) <b>IEEE:</b> P_mean_out, Q1p, Se, Ve, Ie, PF1, Pfe, THDia, vaf, iaf (10)	<b>k-NN :</b> k= 1  <b>RF:</b> trees=40, random_state=15	<b>k-NN :</b> k= 1  <b>RF:</b> trees=30, random_state=15	<b>k-NN :</b> k= 1  <b>RF:</b> trees=30, random_state=15
3	<b>CPT:</b> A, Q, PF, QF, U_eficaz (5) <b>CPC:</b> iacR, iscR, ircR, igR, S, PF, Im_Sc (7) <b>IEEE:</b> Q1p, Se, Ve, Pfe, THDia, vaf, iaf (7)	<b>k-NN :</b> k= 1  <b>RF:</b> trees=20, random_state=20	<b>k-NN :</b> k= 3  <b>RF:</b> trees=20, random_state=15	<b>k-NN :</b> k= 1  <b>RF:</b> trees=40, random_state=50
4	<b>CPT:</b> A, Q, PF, U_eficaz (4) <b>CPC:</b> igR, S, PF, Im_Sc (4) <b>IEEE:</b> Se, Ve, Pfe, THDia, vaf (5)	<b>k-NN :</b> k= 3  <b>RF:</b> trees=10, random_state=10	<b>k-NN :</b> k= 1  <b>RF:</b> trees=20, random_state=20	<b>k-NN :</b> k= 1  <b>RF:</b> trees=40, random_state=50
5	<b>CPT:</b> A, PF, U_eficaz (3) <b>CPC:</b> igR, S, PF (3) <b>IEEE:</b> Se, Ve, THDia, vaf (4)	<b>k-NN :</b> k= 1  <b>RF:</b> trees=30, random_state=20	<b>k-NN :</b> k= 1  <b>RF:</b> trees=40, random_state=10	<b>k-NN :</b> k= 1  <b>RF:</b> trees=40, random_state=20

Como é possível notar, a Simulação 1 realizada anteriormente, não foi seguida neste caso, uma vez que não foram observadas grandes alterações na retirada dos atributos nulos do banco de dados. Como a intenção desta análise é verificar as precisões de banco de dados otimizados, considerando apenas uma das teorias de potência, esta etapa do procedimento foi descartada.

#### 4.4.3.1 Validação da performance - SIMULAÇÃO 2

Na simulação 2, foi realizada para cada uma das teorias de potência a classificação considerando os algoritmos k-NN e RF. A Tabela 61, mostra um resumo dos parâmetros determinados para os algoritmos e as performances observadas para cada um dos casos.

**Tabela 61 - Performance dos algoritmos - Simulação 2**

SIMULAÇÃO 2						
CPT			CPC		IEEE	
Medições	k-NN	RF	k-NN	RF	k-NN	RF
<i>Melhores Parâmetros</i>	<i>k=1</i>	<i>trees = 40 random state = 15</i>	<i>k=1</i>	<i>trees = 30 random state = 15</i>	<i>k=1</i>	<i>trees = 30 random state = 15</i>
Precisão	0,99	0,98	0,98	0,98	0,99	0,98
Recall	0,99	0,98	0,97	0,98	0,99	0,98
F-Score	0,99	0,98	0,97	0,98	0,99	0,98
Ttest(ms)	< 1	249,97	< 1	257,45	< 1	247,4

Nota-se que, dentre as três teorias de potência, para o banco de dados considerado de 10 atributos, ambos k-NN e RF possuem boa precisão de classificação. Em uma observação mais criteriosa, a CPC é a teoria que apresentou precisão menor, mesmo considerando-se o k-NN.

#### 4.4.3.2. Validação da performance - SIMULAÇÃO 3

Na simulação 3, o banco de dados da CPT passa a ser reduzido pela metade, contendo um total de 5 atributos. Por sua vez, para o banco de dados da CPC e IEEE, têm-se 7 atributos em cada um deles. A Tabela 62, mostra os parâmetros utilizados para cada algoritmo e os resultados obtidos.

**Tabela 62 - Performance dos algoritmos - Simulação 3**

SIMULAÇÃO 3						
CPT			CPC		IEEE	
Medições	k-NN	RF	k-NN	RF	k-NN	RF
<i>Melhores Parâmetros</i>	<i>k=1</i>	<i>trees = 20 random state = 20</i>	<i>k=3</i>	<i>trees = 20 random state = 15</i>	<i>k=1</i>	<i>trees = 40 random state = 50</i>
Precisão	0,99	0,98	0,98	0,98	0,98	0,99
Recall	0,99	0,98	0,98	0,98	0,98	0,99
F-Score	0,99	0,98	0,98	0,98	0,98	0,99
Ttest(s)	< 1	234,12	< 1	238,27	< 1	222,74

Neste caso, a CPC ainda se mantém como a teoria com a menor precisão de classificação. A CPT é mais bem classificada pelo algoritmo k-NN e, por sua vez, a IEEE é mais bem classificada pelo RF. Pensando em performance, neste momento, entende-se que devido à relação entre precisão e tempo de teste, o k-NN e a CPT poderiam ser utilizados nesta aplicação, sem impactos. Como ressalva, a CPT neste caso é a teoria que possui o menor número de atributos no banco de dados.

#### 4.4.3.3. Validação da performance - SIMULAÇÃO 4

A simulação 4, considera um número de 4 atributos para o banco de dados da CPT, bem como no da CPC e 5 constituindo o banco de dados da IEEE. Na Tabela 63, é possível observar os parâmetros utilizados para cada algoritmo e as performances obtidas.

**Tabela 63 - Performance dos algoritmos - Simulação 4**

SIMULAÇÃO 4						
CPT			CPC		IEEE	
Medições	k-NN	RF	k-NN	RF	k-NN	RF
Melhores Parâmetros	k=3	trees = 10 random state = 10	k=1	trees = 20 random state = 20	k=1	trees = 40 random state = 50
Precisão	1	0,98	0,98	0,97	1	0,99
Recall	1	0,98	0,97	0,97	1	0,99
F-Score	1	0,98	0,97	0,97	1	0,99
Ttest(ms)	<1	231,81	<1	222,7	<1	234,75

Mais uma vez, a CPC não é a teoria com melhor precisão na classificação. Desta forma, ao verificar as demais teorias, percebe-se que na simulação 4, ambas CPT e IEEE são classificadas com precisão de 100% considerando-se o k-NN. Neste cenário, como a CPT possui menos atributos no banco de dados e maior simplicidade no cálculo dos atributos, entende-se que computacionalmente, na aplicação real, exigirá menos esforços para os cálculos dos atributos em tempo real. De qualquer forma, esta premissa, deve ser futuramente estudada.

#### 4.4.3.4. Validação da performance - SIMULAÇÃO 5

Por fim, na simulação 5, foram simulados os bancos de dados finais para cada teoria de potência. A Tabela 64 resume as simulações e resultados obtidos.

Nesta última simulação, é possível notar que apenas 3 atributos advindos da CPT poderiam ser utilizados para representar as cargas elétricas com boa performance de classificação, com precisão em torno de 99% e tempo de teste muito abaixo de 1 ms. Nestas condições, percebe-se que a potência aparente, o fator de potência e a tensão eficaz das cargas elétricas são suficientes para obtenção do resultado satisfatório.

**Tabela 64 - Simulações e banco de dados utilizado**

SIMULAÇÃO 5						
CPT			CPC		IEEE	
Medições	k-NN	RF	k-NN	RF	k-NN	RF
<i>Melhores Parâmetros</i>	<i>k=1</i>	<i>trees = 30 random state = 20</i>	<i>k=1</i>	<i>trees = 40 random state = 10</i>	<i>k=1</i>	<i>trees = 40 random state = 20</i>
Precisão	0,99	0,99	0,96	0,96	0,98	0,98
Recall	0,99	0,99	0,96	0,96	0,98	0,99
F-Score	0,99	0,99	0,96	0,96	0,98	0,98
Ttest(ms)	<1	216,19	<1	220,65	<1	234,76

#### 4.5. Comparação Entre Cenários Geral e Individual

Como análise final, verifica-se entre o cenário geral e individual, as simulações mais otimizadas (simulação 5) e uma comparação entre os resultados é realizada. A Tabela 65 mostra uma comparação entre os resultados obtidos para os dois cenários comentados.

Nota-se que em questão de tempo para teste do algoritmo, não se observa melhorias expressivas, uma vez que para ambos os casos, este tempo é considerado muito baixo. Por sua vez, com relação à precisão, percebe-se que no cenário individual,

a classe 7 (secador de cabelos) é classificada com precisão de 92%, caso que não ocorre no cenário geral.

**Tabela 65 - Comparação - Simulação 5 - Cenário geral e individual**

CENÁRIO GERAL				CENÁRIO INDIVIDUAL			
<b>Banco de dados:</b> CPT: P, V, PF (3) CPC: igR, PF(2) IEEE: Ve, vaf (2)				<b>Banco de dados:</b> CPT: A, PF, U_eficaz (3)			
k-NN (k=1)				k-NN (k=1)			
Classe	Precisão	Recall	support	Classe	Precisão	Recall	support
1	1	1	12	1	1	1	12
6	1	1	12	6	1	1	12
7	1	1	11	7	0,92	1	11
10	1	1	12	10	1	1	12
15	1	1	12	15	1	1	12
28	1	1	12	28	1	1	11
57	1	1	12	57	1	0,92	12
66	1	1	11	66	1	1	12
80	1	1	12	80	1	1	12
136	1	1	12	136	1	1	12
<b>Ttest &lt; 1 ms</b>				<b>Ttest &lt; 1 ms</b>			

Apesar de uma precisão melhor na classificação, o cenário geral pode apresentar um esforço computacional maior, ao considerar 2 teorias adicionais nos cálculos, e um total de 4 atributos adicionais ao banco de dados. Por existir uma redução significativa no banco de dados, sem perdas expressivas na precisão de classificação, acredita-se que o cenário individual pode ser utilizado na aplicação.

A Tabela 66 resume e comprova a melhoria no tempo para teste do algoritmo sem comprometer a precisão de classificação com a otimização realizada.

**Tabela 66 - Avaliação Tempo x Precisão com a Otimização do Banco de dados**

<b>SIMULAÇÃO</b>	<b>BANCO DE DADOS</b>	<b>k-NN</b>	<b>RF</b>
1	Inicial (55 Atributos)	Precisão: 100% $t_{test}$ : 15,62 ms	Precisão: 98% $t_{test}$ : 328 ms
5	Retirada de Correlação maior ou igual a 50% (7 Atributos)	Precisão: 100% $t_{test}$ : < 1 ms	Precisão: 99% $t_{test}$ : 260 ms

## CAPÍTULO 5 - CONCLUSÕES

A engenharia dos atributos é uma etapa fundamental em qualquer aplicação que envolva aprendizado de máquinas. Com base na classificação de cargas elétricas, este trabalho apresentou a engenharia do banco de dados realizada a partir da extração de índices de potência, tensão e correntes provenientes das teorias CPT, CPC e IEEE STD 1459-2010, que foram analisados e utilizados para a classificação feita por 5 tipos de algoritmos de aprendizado de máquinas: k-NN, SVM Linear, SVM Radial, *DT* e *Random Forest*.

Dentre as 3 teorias de desagregação de cargas, percebe-se que a variedade de índices pode contribuir para uma melhoria na elaboração do banco de dados a ser usado no momento da classificação de cargas. Foi possível verificar que alguns índices possuem relevância em ambas as teorias, como por exemplo, Potência Aparente e Potência Ativa.

Com a redução de 55 atributos para 7, procura-se diminuir o tempo de processamento para a execução do algoritmo de inteligência artificial responsável pela caracterização das cargas.

Se comparado puramente os resultados entre precisão e tempo de teste, entende-se que o melhor seria utilizar-se do algoritmo k-NN. Porém, como o tempo de execução depende da capacidade de processamento do *hardware*, destaca-se a precisão obtida com o algoritmo RF. Diante disto, o RF pode ser recomendado como algoritmo de classificação para o conjunto de dados observado, incluindo a desagregação de carga e uma linha de base para comparação posterior.

Quando verificado o tempo de teste a partir da redução do banco de dados feita entre as simulações 1 e 5, o RF não apresenta grande redução, porém, apresenta uma melhoria na classificação para 99%, na qual, classifica com 92% apenas a classe 7 (secador de cabelos). Análises futuras, considerando um número maior de cargas, pode ser determinante para de fato ser possível recomendar o melhor algoritmo para esta aplicação, entendendo-se estar definido entre k-NN e RF.

Fazendo-se uma análise individual por teoria de potência, a Tabela 65, representa uma análise comparativa, que reflete para o k-NN a mesma observação do cenário global feita para o RF: o k-NN passa a classificar com 92% a classe 7 (secador de cabelos).

Com esta observação, conclui-se que um banco de dados utilizando somente de 3 atributos (A, PF, U\_eficaz) advindos de uma única teoria de potência (CPT), pode ser utilizado para obter os mesmos resultados na classificação feita com base em um banco de dados levantado por 3 diferentes teorias de potência, constituído por um total de 7 atributos, com um tempo de teste menor.

Conforme observado anteriormente, os atributos da CPC não permitiram uma boa precisão se comparado às precisões obtidas nas demais teorias. Desta forma, entende-se ser possível descartar esta teoria em análises futuras, caso a inserção de novas cargas não traga benefícios ao resultado já obtido neste trabalho.

Toda a contribuição apresentada poderá fortalecer a criação de um banco de dados levantado uma vez que se verifica a confiabilidade do mesmo para utilização em futuros trabalhos de pesquisa e, com tudo, possibilitar a criação de uma ferramenta adicional de diagnósticos destinadas à aplicação em medidores de energia elétrica inteligente com a capacidade de comunicação eficaz e precisa com a concessionária e consumidor.

A otimização do banco de dados, possibilita, uma vez que a tarefa de classificação seja adicionada ao medidor de energia elétrica inteligente, uma otimização física de componentes especificados para composição do mesmo, devido à simplicidade de processamento.



## REFERÊNCIAS

- [1] Schäffer, G. J. et al. Comparative Analysis between IEEE 1459 and Conservative Power Theory Through the Development of Power Meters. In: International Conference on Renewable Energies and Power Quality (ICREPQ'18). 2018.
- [2] Mikulovic, Jovan, Tomislav Sekara, and Bojana Skrbic. "Currents' physical component (CPC) power theory for three-phase four-wire systems." Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion (MedPower 2016). IET, 2016.
- [3] Calamero, Nezah, Yuval Beck, and Doron Shmilovich. "Adaptation of power transport theory current physical components to electric network diagnostics." 2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel. IEEE, 2012.
- [4] LIPARI, ALESSANDRO, ANTONIO CATALIOTTI, And PIETRO LUCIO BUCCHERI. "STUDIO E SVILUPPO DI NUOVE STRATEGIE PER LA REALIZZAZIONE DI CAMPIONI DI POTENZA ED ENERGIA SIA IN REGIME SINUSOIDALE CHE IN PRESENZA DI DISTORSIONE ARMONICA." (2011)
- [5] Czarnecki, Leszek S. "From Steinmetz to currents' physical components (CPC): History of power theory development." 2016 International Conference on Applied and Theoretical Electricity (ICATE). IEEE, 2016.
- [6] Tenti, Paolo, Paolo Mattavelli, and Helmo K. Morales Paredes. "Conservative power theory, sequence components and accountability in smart grids." 2010 International school on nonsinusoidal currents and compensation. IEEE, 2010.
- [7] Paredes, Helmo Kelis Morales. "Teoria de potência conservativa: uma nova abordagem para o controle cooperativo de condicionadores de energia e considerações sobre atribuição de responsabilidades." UNICAMP (2011): 29-48.

- [8] Emanuel, Alexander Eigeles. "Summary of IEEE standard 1459: definitions for the measurement of electric power quantities under sinusoidal, nonsinusoidal, balanced, or unbalanced conditions." IEEE transactions on industry applications 40.3 (2004): 869-876.
- [9] Guo, Jing, Xiangning Xiao, and Tao Shun. "Discussion on instantaneous reactive power theory and currents' physical component theory." 2012 IEEE 15th International Conference on Harmonics and Quality of Power. IEEE, 2012.
- [10] Artale, Giovanni, et al. "Measurement of Simplified Single-and Three-Phase Parameters for Harmonic Emission Assessment Based on IEEE 1459-2010." IEEE Transactions on Instrumentation and Measurement 70 (2020): 1-10.
- [11] Beck, Yuval, and Ram Machlev. "Harmonic loads classification by means of currents' physical components." Energies 12.21 (2019): 4137.
- [12] Souza, Wesley Angelino de. "Estudos de técnicas de análise e tecnologias para o desenvolvimento de medidores inteligentes de energia residenciais." (2016).
- [13] Souza, Wesley A., et al. "A nilm dataset for cognitive meters based on conservative power theory and pattern recognition techniques." Journal of Control, Automation and Electrical Systems 29 (2018): 742-755.
- [14] Gurbuz, Fethi Batincan, Ramazan Bayindir, and Halil Ibrahim Bulbul. "A brief review of non-intrusive load monitoring and its impact on social life." 2021 9th International Conference on Smart Grid (icSmartGrid). IEEE, 2021.
- [15] Gao, Jingkun, et al. "Plaid: a public dataset of high-resolution electrical appliance measurements for load identification research: demo abstract." proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings. 2014.
- [16] Procel "Neoenergia desenvolve tecnologia para monitorar consumo de residências em tempo real" 14.04.2021. Disponível em: <http://www.procelinfo.com.br/main.asp?View=%7BF5EAADD6-CCB0-4E29-A0C4482D3D66BB65%7D&Team=&params=itemID=%7BE30A3D69-125E-498E-83C->

F7057C6E5F91%7D;&UIPartUID=%7BD90F22DB-05D4-4644-A8F2-FAD4803C8898%7D . Acesso em: 20 de março de 2021.

[17] Procel “Enel SP investe R\$ 1,97 milhão em obras de eficiência energética” 07.04.2021. Disponível em:

<http://www.procelinfo.com.br/main.asp?View=%7BF5EAADD6-CCB0-4E29-A0C4-482D3D66BB65%7D&Team=&params=itemID=%7BD4BB9CC5-D2D2-49EB-9244-E240C3C7C0DA%7D;&UIPartUID=%7BD90F22DB-05D4-4644-A8F2-FAD4803C8898%7D> . Acesso em: 20 de março de 2021.

[18] Procel “ Com eficiência energética, UFPel reduz em R\$ 750 mil os gastos com energia elétrica” 01.04.2021. Disponível em:

<http://www.procelinfo.com.br/main.asp?View=%7BF5EAADD6-CCB0-4E29-A0C4-482D3D66BB65%7D&Team=&params=itemID=%7B0E7BDBC7-B344-4D2F-B4CD-7746341644DF%7D;&UIPartUID=%7BD90F22DB-05D4-4644-A8F2-FAD4803C8898%7D>. Acesso em: 20 de março de 2021.

[19] Cannas, Barbara, et al. "Selection of features based on electric power quantities for non-intrusive load monitoring." Applied Sciences 11.2 (2021): 533.

[20] Biansoongnern, Somchai, and Boonyang Plangklang. "Nonintrusive load monitoring (NILM) using an Artificial Neural Network in embedded system with low sampling rate." 2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). IEEE, 2016.

[21] Hart, George William. "Nonintrusive appliance load monitoring." Proceedings of the IEEE 80.12 (1992): 1870-1891.

[22] Sadeghianpourhamami, Nasrin, et al. "Comprehensive feature selection for appliance classification in NILM." Energy and Buildings 151 (2017): 98-10

[23] Su, Yi-Ching, Kuo-Lung Lian, and Hsueh-Hsien Chang. "Feature selection of non-intrusive load monitoring system using STFT and wavelet transform." 2011 IEEE 8th international conference on e-business engineering. IEEE, 2011.

- [24] Lam, Hong Yin, G. S. K. Fung, and W. K. Lee. "A novel method to construct taxonomy electrical appliances based on load signaturesof." *IEEE Transactions on Consumer Electronics* 53.2 (2007): 653-660.
- [25] Sadeghianpourhamami, Nasrin, et al. "Comprehensive feature selection for appliance classification in NILM." *Energy and Buildings* 151 (2017): 98-106.
- [26] Kelly, Jack, and William Knottenbelt. "Neural nilm: Deep neural networks applied to energy disaggregation." *Proceedings of the 2nd ACM international conference on embedded systems for energy-efficient built environments*. 2015. doi:10.1145/2821650.2821672.
- [27] Xiao, Ziwei, et al. "Cooling load disaggregation using a NILM method based on random forest for smart buildings." *Sustainable Cities and Society* 74 (2021): 103202.
- [28] UNIVERSIDADE FEDERAL DE SANTA CATARINA - DEPARTAMENTO DE ENGENHARIA ELÉTRICA - Aula 10: Medição de Energia Elétrica. Disponível em: <http://www.labspot.ufsc.br/circuitos1/Aula10E.pdf> . Acesso em: 16 de Julho de 2020.
- [29] Artale, Giovanni, et al. "Measurement uncertainty of harmonic emission indicators based on IEEE Std. 1459-2010." 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC). IEEE, 2020.
- [30] Kukacka, Leos, et al. "CPC and IEEE power theory-Application for offline waveform data analysis." (2013): 0434-0434.
- [31] Manimala, D., and D. Sundararajan. "Evaluation of IEEE standard 1459-2010 power components using stationary wavelet transform." 2016 3rd International Conference on Advanced Computing and Communication Systems (ICACCS). Vol. 1. IEEE, 2016.
- [32] Rens, Johan, Tian Van Rooyen, and Francois De Jager. "Where is the power of the IEEE 1459-2010?." 2014 IEEE International Workshop on Applied Measurements for Power Systems Proceedings (AMPS). IEEE, 2014.

- [33] Ding, Yong, Meiqin Mao, and Liuchen Chang. "Conservative power theory and its applications in modern smart grid: Review and prospect." *Applied Energy* 303 (2021): 117617.
- [34] Muscas, Carlo, and Antonio Marongiu. "Sulle definizioni di potenze in regime non sinusoidale."
- [35] Marcelo, Alexandre Borges, et al. "Currents' physical components (CPC): Case studies in single phase systems." 2018 Simposio Brasileiro de Sistemas Eletricos (SBSE). IEEE, 2018.
- [36] Czarnecki, Leszek S. "Physical reasons of currents RMS value increase in power systems with nonsinusoidal voltage." *IEEE transactions on power delivery* 8.1 (1993): 437-447.
- [37] Marcelo, Alexandre Borges, et al. "Currents' physical components (CPC): Case studies in single phase systems." 2018 Simposio Brasileiro de Sistemas Eletricos (SBSE). IEEE, 2018.
- [38] Tenti, Paolo, Helmo Kelis Morales Paredes, and Paolo Mattavelli. "Conservative power theory, a framework to approach control and accountability issues in smart microgrids." *IEEE Transactions on Power Electronics* 26.3 (2010): 664-673.. <https://doi.org/10.1109/TPEL.2010.2093153>.
- [39] Tenti, P., and P. Mattavelli. "Third-order passive load identification under non-sinusoidal conditions." *European transactions on electrical power* 12.2 (2002): 93-100. [https://doi.org/10.1002/\(ISSN\)1546-310910.1002/etep.v12:210.1002/etep.4450120203](https://doi.org/10.1002/(ISSN)1546-310910.1002/etep.v12:210.1002/etep.4450120203).
- [40] Moreira, Alexandre C., Helmo KM Paredes, and Luiz CP Da Silva. "Applying conservative power theory for analyzing three-phase X-ray machine impact on distribution systems." *Electric Power Systems Research* 129 (2015): 114-125. <https://doi.org/10.1016/j.epsr.2015.07.019>.

- [41] Moreira, Alexandre C., Luiz CP da Silva, and Helmo KM Paredes. "Power quality study and analysis of different arc welding machines." *Journal of Control, Automation and Electrical Systems* 29.2 (2018): 163-176. <https://doi.org/10.1007/s40313-017-0363-6>.
- [42] Emanuel, A. E., R. Langella, and A. Testa. "Power definitions for circuits with nonlinear and unbalanced loads—The IEEE standard 1459-2010." 2012 IEEE Power and Energy Society General Meeting. IEEE, 2012.
- [43] Cataliotti, Antonio, and Valentina Cosentino. "Harmonic sources detection in power systems via nonactive power measurements according to IEEE Std. 1459-2010: Theoretical approach and experimental results." 2010 IEEE International Workshop on Applied Measurements for Power Systems. IEEE, 2010.
- [44] Poljak, Predrag D., Miodrag D. Kusljevic, and Josif J. Tomic. "Power components estimation according to IEEE standard 1459-2010 under wide-range frequency deviations." *IEEE Transactions on Instrumentation and Measurement* 61.3 (2011): 636-644.
- [45] Alobaidi, Abdulraheem, and Hamzeh Davarikia. "The Importance of Utilizing the Current's Physical Components Power Theory in Nonsinusoidal Power Systems." 2019 IEEE Green Technologies Conference (GreenTech). IEEE, 2019.
- [46] Bai, Zhixuan, Wenjia Jiang, and Xianyong Xiao. "Analysis of Reactive Power Compensation in Hybrid Active Power Filter Based on IEEE Std. 1459-2010 and CPC Theory." *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2020.
- [47] Ancelmo, Hellen Cristina. *Desenvolvimento de base de dados e aplicação do método de Prony para extração de características e classificação de cargas elétricas*. MS thesis. Universidade Tecnológica Federal do Paraná, 2020.
- [48] Liu, Hui. "Non-intrusive load monitoring." *Theory, Technologies and Applications*. Science Press Beijing. Springer. (2020).

- [49] Cannas, B., et al. "Real-time monitoring system of the electricity consumption in a household using NILM techniques." Proceedings of the 24th IMEKO TC4 International Symposium and 22nd International Workshop on ADC and DAC Modelling and Testing, Palermo, Virtual, Italy. 2020.
- [50] Zhang, Jixian, Xiangguo Lin, and Xiaogang Ning. "SVM-based classification of segmented airborne LiDAR point clouds in urban areas." Remote sensing 5.8 (2013): 3749-3775.
- [51] Sánchez-González, Beatriz, et al. "Support vector machines for explaining physiological stress response in Wood mice (*Apodemus sylvaticus*)." Scientific reports 8.1 (2018): 1-14.
- [52] SUPPORT VECTOR MACHINES(SVM), Published in Towards Data Science, 2018. Disponível em: <https://towardsdatascience.com/support-vector-machines-svm-c9ef22815589> . Acesso em 23 de Agosto de 2022.
- [53] Vapnik, Vladimir N. "The nature of statistical learning theory, 840 Springer-Verlag New York." Inc., New York, NY, USA 841 (1995): 842.
- [54] Support Vector Machines with Scikit-learn Tutorial, published in Datacamp, 2019. Disponível em: <https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python> . Acesso em: 16 de Julho de 2022.
- [55] Support Vector Machines (Macchine a vettori di support(. Disponível em: <https://slideplayer.it/slide/570429/> . Acesso em: 16 de Julho de 2022.
- [56] Implementing SVM and Kernel SVM with Python's Scikit-Learn, published in Stackabuse. Disponível em: <https://stackabuse.com/implementing-svm-and-kernel-svm-with-pythons-scikit-learn>
- [57] Radial Kernel SVM. Disponível em: [https://rstudio-pubs-static.s3.amazonaws.com/296261\\_b5acc05b2b0e41879c917033b1497543.htmls](https://rstudio-pubs-static.s3.amazonaws.com/296261_b5acc05b2b0e41879c917033b1497543.htmls) . Acesso em: 16 de Julho de 2022.

[58] Sarfraz, Memuna, Fadi Abu-Amara, and Ikhlas Abdel-Qader. "A computer aided detection framework for mammographic images using fisher linear discriminant and nearest neighbor classifier." (2012).

[59] MacCormick, J. , "9 ALGORITIMI CHE HANNO CAMBIATO IL FUTURO", Apogeo, Edição 9788850331697. Disponível em: [https://books.google.com.br/books?id=uC5YwCyBXboC&pg=PT98&lpg=PT98&dq=reti+neurali+vicino+piu+prossimo&source=bl&ots=DyPQ9IVq4j&sig=UXF9NbOK1r6BiH7qO\\_9V6t2TEwo&hl=pt-BR&sa=X&ved=2ahUKEwjfrvm0qTcAhUNuVMKHV4HADEQ6AEwAHoECAAQAQ#v=onepage&q=reti%20neurali%20vicino%20piu%20prossimo&f=false](https://books.google.com.br/books?id=uC5YwCyBXboC&pg=PT98&lpg=PT98&dq=reti+neurali+vicino+piu+prossimo&source=bl&ots=DyPQ9IVq4j&sig=UXF9NbOK1r6BiH7qO_9V6t2TEwo&hl=pt-BR&sa=X&ved=2ahUKEwjfrvm0qTcAhUNuVMKHV4HADEQ6AEwAHoECAAQAQ#v=onepage&q=reti%20neurali%20vicino%20piu%20prossimo&f=false). Acesso em 16 de Julho de 2022

[60] Alpaydin, Ethem. Introduction to machine learning. Fourth edition. Cambridge, Massachusetts: The MIT press, 2020.

[61] Tsigkritis, Theocharis, George Groumas, and Moti Schneider. "On the Use of k-NN in Anomaly Detection." Journal of Information Security 9.01 (2018): 70.

[62] Decision Tree, published in GeeksforGeeks, 2022. Disponível em: <https://www.geeksforgeeks.org/decision-tree/> . Acesso em 16 de Julho de 2022

[63] What are your decision tree needs?, published in Lucid Chart. Disponível em : <https://www.lucidchart.com/pages/decision-tree> . Acesso em 16 de Julho de 2022

[64] Random Forest Algorithm for Machine Learning, 2019. Disponível em: <https://www.tutorialandexample.com/random-forest-algorithm/> . Acesso em 16 de Julho de 2022

[65] What is a Decision Tree and How to Make One [Templates + Examples], published in Venngage, 2021. Disponível em: <https://venngage.com/blog/what-is-a-decision-tree/> . Acesso em 16 de Julho de 2022

[66] Iwendi, Celestine, et al. "COVID-19 patient health prediction using boosted random forest algorithm." Frontiers in public health 8 (2020): 357.



- [67] Nguyen, Hoang, and Xuan-Nam Bui. "Predicting blast-induced air overpressure: a robust artificial intelligence system based on artificial neural networks and random forest." *Natural Resources Research* 28.3 (2019): 893-907.
- [68] Sipper, Moshe, and Jason H. Moore. "Conservation machine learning: a case study of random forests." *Scientific Reports* 11.1 (2021): 3629.
- [69] Gislason, Pall Oskar, Jon Atli Benediktsson, and Johannes R. Sveinsson. "Random forests for land cover classification." *Pattern recognition letters* 27.4 (2006): 294-300.
- [70] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001 Kluwer Academic Publishers. Manufactured in The Netherlands.
- [71] Leistner, Christian, et al. "Semi-supervised random forests." 2009 IEEE 12th international conference on computer vision. IEEE, 2009.
- [72] Schonlau, Matthias, and Rosie Yuyan Zou. "The random forest algorithm for statistical learning." *The Stata Journal* 20.1 (2020): 3-29.
- [73] Medeiros, Aérton Pedra. "Classificação de eventos em monitoramento nilm de cargas elétricas residenciais utilizando rede neural convolucional." (2019).
- [74] Plaid Dataset, 2013 and 2014 (Composition: 537 devices, 11 classes). Disponível em: <http://www.timeseriesclassification.com/description.php?Dataset=PLAID> . Acesso em 16 de Julho de 2022
- [75] UK-Dale Dataset, final version 2017. Disponível em: <https://jack-kelly.com/data/> . Acesso em 16 de Julho de 2022
- [77] Teixeira, MATEU D., et al. "Análise do Impacto de lâmpadas Fluorescente Compactas na Rede de Dstribuição da COPEL." VI SBQEE'05 (2005): 309-315.
- [78] Carvalho, Flavio Gomes de. "Caracterização do conteúdo harmônico em ambientes residenciais: estudo de caso." (2015).

[79] da Silva, Cíntia Gonçalves Mendes, Hédio Tatizawa, and Kleiber Tadeu Soletto. "A Utilização De Lâmpadas Fluorescentes Compactas Nas Redes Elétricas." Revista Do Centro Federal De Educação Tecnológica De São Paulo.

[80] Siqueira, Gabriela de Paiva. Fluorescentes, Lâmpadas. Estudo Do Comportamento Elétrico De. 2011. Tese de Doutorado. Universidade Estadual de Campinas

[81] Azevedo, Ana Claudia de. "Desempenho de aparelhos refrigeradores no contexto da qualidade da energia elétrica." (2002).

[82] Datasheet Electrolux ME21G. Disponível em: <https://www.electronica-pt.com/esquema/microondas/electrolux/electrolux-me21g-7844/>. Acesso em 16 de Julho de 2022.

[83] Como Funciona o Forno Microondas, publicado em Saber Elétrica. Disponível em: <https://www.sabereletrica.com.br/funcionamento-forno-microondas/> . Acesso em 16 de Julho de 2022.

[84] Sanhueza, Sergio Manuel Rivera. "Modelagem e análise de condicionadores de ar dos tipos convencional e inteligente e influência no comportamento dinâmico das redes elétricas." (2007).

[85] Rangel, Maurício Silva. "Análise do consumo e qualidade de energia em condicionadores de ar convencional e inverter." (2020).

[86] Miquelin, Juliana Zimbardi. "Análise de desempenho de lâmpadas LEDs e LFCs como alternativas em iluminação residencial em função da implementação da portaria ministerial n 1007 que limita o uso de lâmpadas incandescentes no Brasil." (2011).

[87] Bosco, Thais Berrettini, Flavio Alessandro Serrão Gonçalves, and Wesley Angelino de Souza. "A Comparative Study of Machine Learning Classifiers for Electric Load Disaggregation based on an extended NILM dataset." 2021 14th IEEE International Conference on Industry Applications (INDUSCON). IEEE, 2021.

[88] BOSCO, T. B. ; SOUZA, W. A. ; GONCALVES, F.A.S. . Random Forest Applied To An Extended NilM Dataset To Classify Disaggregated Electric Loads. In: International Liberty Interdisciplinary Studies Conference, 2022, New York - USA. Proceedings of International Liberty Interdisciplinary Studies Conference. New York, USA: Liberty Academic Books, 2022. v. 1. p. 243-251.

[89] Souza, Wesley A., et al. "Selection of features from power theories to compose NILM datasets." *Advanced Engineering Informatics* 52 (2022): 101556.