



UNIVERSIDADE FEDERAL DO MARANHÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENERGIA E AMBIENTE
MESTRADO PROFISSIONAL EM ENERGIA E AMBIENTE

ROBERTO ARTURO QUEZADA SALES

PROPOSTA DE METODOLOGIA BASEADA EM
SISTEMA DE SUPORTE À DECISÃO ESPACIAL E
APRENDIZADO NÃO SUPERVISIONADO PARA
ANÁLISE E MITIGAÇÃO DE FALTAS EM REDES DE
DISTRIBUIÇÃO MODERNAS

SÃO LUÍS - MA, BRASIL

2018

Roberto Arturo Quezada Sales

**PROPOSTA DE METODOLOGIA BASEADA EM
SISTEMA DE SUPORTE À DECISÃO ESPACIAL E
APRENDIZADO NÃO SUPERVISIONADO PARA
ANÁLISE E MITIGAÇÃO DE FALTAS EM REDES DE
DISTRIBUIÇÃO MODERNAS**

Dissertação apresentada ao Programa de Pós-Graduação em Energia e Ambiente da Universidade Federal do Maranhão, para obtenção do título de Mestre em Energia e Ambiente.

Orientador: Prof. Dr. Osvaldo Ronald Saavedra Mendez

São Luís - MA, Brasil

2018

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Núcleo Integrado de Bibliotecas/UFMA

Quezada Sales, Roberto Arturo.

PROPOSTA DE METODOLOGIA BASEADA EM SISTEMA DE SUPORTE À
DECISÃO ESPACIAL E APRENDIZADO NÃO SUPERVISIONADO PARA
ANÁLISE E MITIGAÇÃO DE FALTAS EM REDES DE DISTRIBUIÇÃO
MODERNAS / Roberto Arturo Quezada Sales. - 2018.

108 p.

Orientador(a): Osvaldo Ronald Saavedra Mendez.

Dissertação (Mestrado) - Programa de Pós-graduação em
Energia e Ambiente/ccet, Universidade Federal do Maranhão,
São Luís, 2018.

1. Aprendizado de máquina não supervisionado. 2.
Descoberta de conhecimento em base de dados. 3.
Otimização da confiabilidade. 4. Sistema de suporte à
decisão espacial. 5. Smart grid. I. Saavedra Mendez,
Osvaldo Ronald. II. Título.


ROBERTO ARTURO QUEZADA SALES

**PROPOSTA DE METODOLOGIA BASEADA EM
SISTEMA DE SUPORTE A DECISÃO ESPACIAL E
APRENDIZADO NÃO SUPERVISIONADO PARA
ANÁLISE E MITIGAÇÃO DE FALTAS EM REDES DE
DISTRIBUIÇÃO MODERNAS**


Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Energia e Ambiente da Universidade Federal do Maranhão como parte dos requisitos necessários para obtenção do título de Mestre em Energia e Ambiente.

Aprovada em: 30/11/2018


BANCA EXAMINADORA



Prof. Dr. Osvaldo Ronald Saavedra Mendez (Orientador)
Departamento de Engenharia Elétrica/Universidade Federal do Maranhão – UFMA



Prof. Dr. Francisco Sávio Mendes Sinfrônio (Membro interno)
Departamento de Engenharia Elétrica/Universidade Federal do Maranhão – UFMA



Prof. Dr. Mauro Sérgio Silva Pinto (Membro externo)
Departamento de Engenharia da Computação/Universidade Estadual do Maranhão – UEMA

"Não desista nas primeiras tentativas, a persistência é amiga da conquista. Se você quer chegar aonde a maioria não chega, faça o que a maioria não faz."

Bill Gates

Agradecimentos

Agradeço primeiramente a Deus, pela dádiva da vida, saúde e por ter me guiado nessa trajetória.

A minha esposa Vanessa pelo amor, paciência e cooperação.

Aos meus pais, José Roberto Quezada Peña e Verônica Soraia Sales, pelo amor carinho e a presença constante em minha vida.

Aos meus irmãos Rafael, Rodolfo e toda a minha família que me ajudaram e incentivaram nesta realização pessoal.

Ao meu orientador, Professor Dr Osvaldo Saavedra, que encaminhou e proporcionou a concretização desta realização.

À Universidade Federal do Maranhão por ter proporcionado esta oportunidade ímpar de aprendizado. Em especial aos professores do Programa de Pós-Graduação em Energia e Ambiente da UFMA, pelos valiosos conhecimentos repassados e à toda a turma 2016.

À Companhia Energética do Maranhão (CEMAR) pela contribuição na pesquisa sobre o tema no setor elétrico.

Aos colegas da CEMAR por permitirem e incentivarem o meu desenvolvimento contínuo enquanto profissional. Em especial ao Charles, João e Ronnie por terem investido e acreditado no meu trabalho enquanto foram meus gestores imediatos. Aos colegas da Automação: Bruno, Gleidson, Juliano e Sebastião pelas discussões a respeito do trabalho.

Por fim, todas as pessoas que estão comigo ou que de alguma forma colaboraram para o que eu sou hoje.

*A minha família, principalmente, a
meus pais e meus irmãos.*

Resumo

A disponibilidade e viabilidade de implantação de novas tecnologias (medidores inteligentes, recursos distribuídos de geração e armazenamento, estruturas de monitoramento, sensores, etc), bem como a ampliação das estruturas topológicas dos sistemas de potência (extensão de rede e equipamentos) e a crescente pressão do órgão regulador e da sociedade por um serviço de maior qualidade, resultaram em um aumento significativo da complexidade das tarefas de planejamento e operação das redes de distribuição de energia elétrica. Para permitir uma operação com qualidade, confiabilidade e segurança neste ambiente, é consenso da necessidade de integrar a camada elétrica da rede (*physical network*) com a camada de comunicação e informação (*cyber network*). A forma de executar esta integração é através das chamadas Redes de Energia Inteligentes ou *Smart Grids*. Este trabalho propõe o desenvolvimento de uma metodologia que permitirá avaliar o desempenho das tecnologias, bem como mitigar faltas em sistemas de distribuição de energia elétrica. O método proposto inclui a avaliação dos principais fatores que têm influência na qualidade do serviço. Topologia da Rede, disponibilidade de equipamentos de proteção, manobra e telecontrole e tráfego nas vias de acesso são avaliados em conjunto com a confiabilidade no fornecimento de energia, a fim de combinar os resultados destes critérios com indicadores de qualidade, que possibilitará mitigar os impactos de tais faltas e propor ações de investimentos a curto e médio prazo mais eficientes. O trabalho também abordará uma metodologia de análise baseada em Sistema de Suporte à Decisão Espacial, Técnica de Descoberta de Conhecimento em Base de Dados e Aprendizado de Máquina Não Supervisionado, proporcionando de maneira prática e fácil a comparação entre alimentadores, subestações ou mesmo regiões que compõem um sistema de distribuição.

Palavras-chave: *smart grid*, otimização da confiabilidade, descoberta de conhecimento em base de dados, aprendizado de máquina não supervisionado, sistema de suporte à decisão espacial.

Abstract

The availability and feasibility of deploying new technologies (smart meters, distributed generation and storage resources, monitoring structures, sensors, etc.), as well as the expansion of the topological structures of power systems (network extension and equipment) and the increasing pressure from the regulatory agency and society for a higher quality service resulted in a significant increasing in the complexity of the planning and operation tasks of electricity distribution networks. To allow an operation with quality, reliability and security in this environment, it is a consensus of the need to integrate the physical layer of the “physical network” network with the communication and information layer, “cyber network”. The way to perform this integration is through the so-called Smart Energy Networks or Smart Grids. This work proposes the development of a methodology that will allow to evaluate the performance of the technologies, as well as to mitigate faults in electric power distribution systems. The proposed method includes the evaluation of the main factors that influence the quality of the services. Network topology, availability of protective equipment, maneuvering, telecontrol and traffic in the access routes are evaluated together with the reliability of the power supply, in order to combine the results of these criteria with quality indicators, which will allow mitigating the impacts of the faults and propose more efficient short and medium term investments. The paper will also approach an analysis methodology based on Spatial Decision Support System, Database Knowledge Discovery Technique and Unsupervised Machine Learning, providing in a practical and easy way the comparison between feeders, substations or even regions that make up a distribution system. Keywords: smart grid, reliability optimization, fault mitigation, electrical distribution systems, database knowledge discovery, unsupervised machine learning, simulation of traffic on the access routes.

Keywords: smart grid, reliability optimization, knowledge discovery in database, untrained machine learning, spatial decision support systems.

Lista de ilustrações

Figura 1 – Caracterização do Sistema Elétrico.	25
Figura 2 – Sistema de distribuição primária com diferentes laterais.	27
Figura 3 – (a) Evolução DEC Brasil e (b) Evolução FEC Brasil	38
Figura 4 – (a) Evolução DEC CEMAR e (b) Evolução FEC CEMAR	39
Figura 5 – Compensações pagas no BRASIL, em milhões de R\$.	40
Figura 6 – Compensações pagas pela CEMAR, em milhões de R\$.	40
Figura 7 – Visão geral do sistema energético com o uso da REI.	46
Figura 8 – Arquitetura <i>Web-SDSS</i>	51
Figura 9 – Rede elétrica de uma região da cidade de São Luís do Maranhão (Camada 1) com o respectivo carregamento das vias (Camada 2).	52
Figura 10 – Etapas do Processo de <i>KDD</i>	57
Figura 11 – Arquitetura de um Sistema Típico de <i>Data Mining</i>	60
Figura 12 – Arquitetura de um sistema típico de <i>Data Mining</i>	66
Figura 13 – Agrupamento de um conjunto de objetos baseado no método <i>k-médias</i>	68
Figura 14 – Identificação do Ponto de <i>Elbow</i>	69
Figura 15 – Aba <i>Visualize</i> do <i>Weka</i>	71
Figura 16 – Integração dos Módulos no ambiente das Redes de Distribuição Modernas.	73
Figura 17 – Diagrama relacional com as tabelas e atributos utilizados no sistema.	74
Figura 18 – Filtragem dos dados.	77
Figura 19 – Transformação dos dados.	78
Figura 20 – Arquivo no Formato <i>ARFF</i>	78
Figura 21 – Simulação para identificar número ótimo de <i>clusters</i>	80
Figura 22 – Quantidade ideal de <i>clusters</i> a serem gerados utilizando os atributos considerados relevantes.	81
Figura 23 – Atributos utilizados na Mineração de Dados.	81
Figura 24 – Interface inicial da execução da tarefa de agrupamento na ferramenta <i>Weka</i>	82
Figura 25 – Diagrama tridimensional de correlação entre o tempo de acesso de cada bloco de rede considerando o tráfego (Acessibilidade), a distância elétrica de cada bloco e o impacto de uma possível falta de cada bloco em função do número de consumidores e a taxa de falha do bloco.	83
Figura 26 – Interface do resultado do <i>cluster</i> utilizando a opção <i>Use training set</i>	84
Figura 27 – Pagina inicial do SIMAUTEK - Módulo de gestão da automação e manutenção	87
Figura 28 – Camada que ilustra o diagrama do alimentador VLP-01C4.	88
Figura 29 – Camada que os marcadores do centro de despacho e o bloco em falha.	88

Figura 30 – (a) Tempo de deslocamento e (b) Tempo de inspeção	89
Figura 31 – Módulo de acompanhamento dos indicadores de qualidade do serviço. .	90
Figura 32 – Módulo de acompanhamento da taxa de falha anual para os blocos da rede de distribuição.	91
Figura 33 – (a) Tempo de Deslocamento Ocorrência SLF-01C8 e (b) Marcadores origem/destino Ocorrência SLF-01C8.	93
Figura 34 – (a) Tempo de Inspeção Ocorrência SLF-01C8 e (b) Camada Alimentador SLF-01C8.	94
Figura 35 – (a) Tempo de Deslocamento Ocorrência ARG-01C4 e (b) Marcadores origem/destino Ocorrência ARG-01C4.	94
Figura 36 – (a) Tempo de Inspeção Ocorrência ARG-01C4 e (b) Camada Alimenta- dor ARG-01C4	95
Figura 37 – Diagrama unifilar SLF-01C8.	96
Figura 38 – Diagrama unifilar ARG-01C4.	96
Figura 39 – Diagrama Unifilar SLF	97

Lista de tabelas

Tabela 1 – Tensões usuais em SEP	25
Tabela 2 – Indicadores individuais de continuidade do serviço.	30
Tabela 3 – Indicadores coletivos de continuidade do serviço.	31
Tabela 4 – Indicadores de tempo de atendimento às ocorrências emergenciais.	32
Tabela 5 – Indicadores de compensação pela ultrapassagem dos limites dos indicadores individuais.	33
Tabela 6 – Compensações de continuidade pagas pelas concessionárias no Brasil, no período de 2010 a 2013.	34
Tabela 7 – Compensações de continuidade pagas pelas concessionárias no Brasil, no período de 2014 a 2017.	34
Tabela 8 – Custo das Interrupções programadas por região e setor (R\$/kWh).	36
Tabela 9 – Custo das Interrupções não programadas por região e setor (R\$/kWh).	36
Tabela 10 – Indicador de continuidade DEC e FEC apurado e os limites estabelecidos para o Brasil e CEMAR, no período de 2005 a 2017.	38
Tabela 11 – Valor das compensações (R\$) e quantidade de compensações que foram pagas, no Brasil e CEMAR, no período de 2010 a 2017.	39
Tabela 12 – Consumo de energia elétrica (MWh) do setor residencial do Brasil e da CEMAR no período de 2010 a 2017.	41
Tabela 13 – Demanda do setor residencial do Brasil e das CEMAR, em kW, no período de 2010 a 2017.	41
Tabela 14 – Energia Interrompida para o setor residencial do Brasil e CEMAR, em kWh interrompido, para o período de 2010 a 2017.	42
Tabela 15 – Custo total das interrupções estimado para o setor industrial do Brasil e CEMAR, em R\$/kWh interrompido, para o período de 2010 a 2017.	43
Tabela 16 – Algoritmos de agrupamento conhecidos, divididos em técnicas e categorias, de acordo com a metodologia de construção dos <i>clusters</i>	63
Tabela 17 – Funções de distância entre dois padrões x e y.	66
Tabela 18 – Atributos primários utilizados na metodologia.	75
Tabela 19 – Principais parâmetros disponíveis para execução do algoritmo <i>SimpleKMeans</i> , suas descrições e valores utilizados durante o estudo de caso.	80
Tabela 20 – Centróides 0 a 3 resultantes da execução da tarefa de clusterização.	85
Tabela 21 – Centróides 4 a 7 resultantes da execução da tarefa de clusterização.	85
Tabela 22 – Comparação de Custos.	95

Lista de abreviaturas e siglas

AHC	– Algoritmo Heurístico Construtivo
ANEEL	– Agência Nacional de Energia Elétrica
API	– <i>Application Programming Interface</i>
ASP	– <i>Active Server Pages</i>
CEMAR	– Companhia Energética do Maranhão
CENS	– Custo da Energia Não Suprida
CSS	– <i>Cascading Style Sheets</i>
DEC	– Duração Equivalente de Interrupção por Unidade Consumidora
DIC	– Duração de Interrupção Individual por Unidade Consumidora
DICRI	– Duração da Interrupção Individual ocorrida em dia crítico por Unidade Consumidora
DMIC	– Duração Máxima de Interrupção Individual por Unidade Consumidora
DSS	– <i>Decision Support System</i>
FEC	– Frequência Equivalente de Interrupção por Unidade Consumidora
FIC	– Frequência de Interrupção Individual por Unidade Consumidora
GIS	– <i>Geographical Information System</i>
HTML	– <i>Hyper Markup Language</i>
KDD	– <i>Knowledge Discovery in Database</i>
LA	– <i>Loop Automation</i>
MLLB	– <i>Mixed Binary Linear Programming</i>
MME	– Ministério de Minas e Energia
MT	– Média Tensão
MVC	– <i>Model View Controller</i>
NIST	– <i>National Institute of Standards and Technology</i>

<i>NSGA-II</i>	– <i>Non-dominated Sorting Genetic Algorithm II</i>
<i>OMS</i>	– <i>Outage Management System</i>
PRODIST	– Procedimento de Distribuição de Energia Elétrica
QEE	– Qualidade de Energia Elétrica
REI	– Rede Elétrica Inteligente
<i>SCADA</i>	– <i>Supervisory Control and Data Acquisition</i>
SDMT	– Sistema de Distribuição de Média Tensão
<i>SDSS</i>	– <i>Spatial Decision Support Systems</i>
SEP	– Sistema Elétrico de Potência
<i>SH</i>	– <i>Self-healing</i>
SIMAUTEK	– Sistema Integrado de Manutenção e Automação de Equipamentos
<i>SQL</i>	– <i>Structured Query Language</i>
TMAE	– Tempo Médio de Atendimento a Emergências
TMD	– Tempo Médio de Deslocamento
TME	– Tempo Médio de Execução
TMP	– Tempo Médio de Preparação
<i>Web-SDSS</i>	– <i>Web-based Spatial Decision Support System</i>
<i>WEKA</i>	– <i>Waikato environment for knowledge analysis</i>

Sumário

1	INTRODUÇÃO	16
1.1	O Tema e sua Importância	16
1.2	Caracterização do Problema	17
1.3	Objetivos	19
1.3.1	Objetivo Geral	19
1.3.2	Objetivos Específicos	19
1.4	Revisão Bibliográfica	20
1.5	Estrutura do Trabalho	23
2	ASPECTOS RELACIONADOS AOS SISTEMAS DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA E ÀS REDES ELÉTRICAS INTELIGENTES	24
2.1	Aspectos Relacionados aos Sistemas de Distribuição	24
2.2	Qualidade de Energia Elétrica: Qualidade do Serviço	28
2.2.1	Indicadores de Continuidade e Tempo de Atendimento	29
2.2.2	Compensações pelas Transgressões dos Limites de Continuidade	33
2.2.3	Análise Econômica das Interrupções de Serviço	35
2.2.4	Análise BRASIL X CEMAR	37
2.3	Redes Elétricas Inteligentes	44
2.3.1	Características, Funcionalidades, Benefícios e Desafios da REI	46
3	SISTEMAS DE SUPORTE A DECISÃO ESPACIAL	49
3.1	Componentes do <i>Web-SDSS</i>	50
3.2	Acessibilidade da Rede de Distribuição	51
3.3	A Aplicação SIMAUTEK CEMAR	53
3.3.1	Requisitos e objetivos	53
3.3.2	Levantamento de Dados do Sistema Corporativo	54
4	EXTRAÇÃO DE CONHECIMENTO	56
4.1	Conceitos e Etapas Operacionais de <i>KDD</i>	56
4.2	Mineração de Dados	59
4.3	Análise de Agrupamentos	62
4.3.1	Tipos Básicos de Dados em Clusterização	64
4.3.2	Medidas de Distância	65
4.3.3	Algoritmo <i>K-means</i>	66
4.4	Ferramenta <i>Weka</i>	70

5	METODOLOGIA DESENVOLVIDA	72
5.1	Etapa de Seleção dos Dados	74
5.2	Etapa de Pré-processamento dos Dados	76
5.3	Etapa de Transformação	77
5.4	Etapa de Mineração de Dados	79
5.5	Etapa de Interpretação e Avaliação dos Resultados	84
5.5.1	Portal SIMAUTEK	86
5.5.1.1	Modulo de Gestão da Automação e Manutenção	86
5.5.2	Módulo de Simulação de Tempos de Deslocamento	87
5.5.3	Módulo de Acompanhamento do Desempenho e da Taxa de falha dos Blocos da Rede de Distribuição	89
6	RESULTADO E DISCUSSÕES	93
7	CONCLUSÃO	98
7.1	Trabalhos Futuros	99
	REFERÊNCIAS	101
	APÊNDICES	106
	APÊNDICE A – DIAGRAMA RELACIONAL - SIMAUTEK	107
	APÊNDICE B – TERMO DE CONSENTIMENTO	108

1 Introdução

1.1 O Tema e sua Importância

A constante evolução econômica e tecnológica atrelada ao aumento do consumo de energia elétrica vêm motivando a ampliação e modernização dos Sistemas Elétricos de Potência (SEP). A reestruturação do SEP conduziu a uma mudança nas relações entre as empresas concessionárias, forçando a busca por melhorias que atendam aos padrões e metas que definem com rigor a qualidade de serviço no fornecimento de energia (DUGAN et al., 1996; KAGAN et al., 2009).

Com o objetivo de adequar-se a esses padrões de qualidade, as empresas de energia elétrica precisam identificar corretamente os fatores que influenciam nos índices de confiabilidade dos seus sistemas, e desta forma, frente a estes desafios, as empresas distribuidoras de energia estão buscando o auxílio de tecnologias e técnicas de otimização e priorização com o intuito de obterem êxito na elaboração de um plano de ações que leve em consideração aspectos técnicos e econômicos. As variáveis a serem considerados para a construção de um modelo otimizado são de diversas ordens e podem ser alteradas de acordo com as características do sistema elétricos envolvido (FERRET et al., 2012).

Em meio às soluções em desenvolvimento a de maior notoriedade é a aplicação dos conceitos de tecnologias provenientes da implantação da Rede Elétrica Inteligente (REI), do anglicanismo *smart grid*, as quais preconizam fornecer energia com altos índices de confiabilidade, continuidade e segurança. Para isto, busca-se aplicar nas redes de distribuição de energia ferramentas de informação, monitoramento, comunicação, controle e medição, com o objetivo de integrar toda a rede para viabilizar serviços e informações em tempo real tanto para as concessionárias quanto para os consumidores, compondo os chamados sistemas de distribuição modernos (FANG et al., 2012; DUTRA et al., 2013; KAGAN et al., 2013).

Com o advento das redes modernas, nos últimos anos, a área de distribuição tem recebido investimentos no tocante a estudos em sistemas de proteção. Deste modo, novas técnicas e equipamentos têm sido explorados para implementação nesta área, tornando-os mais eficientes e robustos frente aos desafios correntes (KINDERMANN, 2010; FILHO; MAMEDE, 2011; SATO; FREITAS, 2015).

Dentre os desafios, as interrupções de energia elétrica causadas por falhas se destacam em virtude do potencial impacto que podem causar no SEP e seus elementos constituintes, às concessionárias de distribuição e principalmente aos consumidores, comprometendo inclusive, a qualidade da energia fornecida (KAGAN et al., 2010). No Brasil,

segundo as normas da Agência Nacional de Energia Elétrica (ANEEL), órgão regulador do sistema de energia, as interrupções se enquadram no item Qualidade do Serviço referente à área de Qualidade de Energia Elétrica (QEE) do Módulo 8 do documento normativo Procedimento de Distribuição de Energia Elétrica no Sistema Elétrico Nacional (PRO-DIST) (ANEEL, 2018b). A alínea Qualidade do Serviço refere-se aos indicadores para contabilização da duração e frequência de interrupções no fornecimento de energia a que os consumidores estão expostos, seja em virtude de interrupções por falhas no sistema ou atividades de manutenção programadas (ANEEL, 2018b).

A fim de aumentar a confiabilidade do sistema, as causas de interrupções não programadas merecem grande destaque na busca da sua identificação e análise de suas inter relações com as demais variáveis do Sistema de Distribuição de Media Tensão (SDMT). Assim, nesta dissertação, propõe-se uma metodologia para avaliar as tecnologias presentes na rede de distribuição e mitigar as faltas, que contemplará as inovações tecnológicas agregadas as redes modernas, considerando para este fim aspectos do SDMT, do tráfego nas vias de acesso e dos recursos de manobra e telecontrole disponíveis.

Após essa introdução, este primeiro capítulo trará na [seção 1.2](#) a caracterização do problema; na [seção 1.3](#) seus objetivos; na [seção 1.4](#) o referencial bibliográfico; e na [seção 1.5](#) a forma como a presente dissertação fora organizada. Em alguns momentos, para exemplificar as situações descritas neste trabalho, serão utilizados dados da Companhia Energética do Maranhão - CEMAR, empresa do grupo Equatorial que possui a concessão da distribuição de energia no estado do Maranhão, atendendo aos 217 Municípios, com aproximadamente 2 milhões de unidades consumidores.

1.2 Caracterização do Problema

Um dos grandes desafios do setor elétrico são os desligamentos não programados no fornecimento de energia provocado por defeitos transitórios ou permanentes. A ocorrência de faltas nos sistemas de distribuição deve-se a diferentes fatores, tais como problemas elétricos, mecânicos, de isolamento, de manutenção, e de natureza térmica, entre outros (KINDERMANN, 2010). Devido aos danos físicos para rede elétrica e seus equipamentos, a correção deve ser realizada o mais rápido possível por equipes de manutenção, que em muitas situações fazem uso de inspeção visual para localização dos pontos de defeito (CAMINHA, 2004; FILHO; MAMEDE, 2011).

Nestas situações, o reestabelecimento de energia pode ser extremamente oneroso e demorado, sobretudo em SDMT localizados em regiões de difícil acesso. Além disso, as interrupções causadas por faltas no sistema elétrico podem ocasionar condições de operação indesejáveis, riscos aos usuários, transtornos a consumidores residências e interrupção ou falha de processos industriais. Este conjunto de consequências negativas afeta diretamente os

indicadores de qualidade e continuidade do serviço de distribuição de energia, implicando em elevados custos operacionais às concessionárias devido às compensações aos consumidores afetados (KAGAN et al., 2010). Ademais, as concessionárias também têm custos financeiros devido a perdas de faturamento por energia não fornecida, e custos sociais associados à manutenção da imagem junto aos intervenientes (*stakeholders*). Por outro lado, os clientes se veem atingidos por perdas de faturamento de suas unidades de negócio, tanto em atividades comerciais quanto industriais, implicando também em custos financeiros (FILHO; MAMEDE, 2011).

O estudo realizado pela *SmartGrid Consumer Collaborative* (COLLABORATIVE, 2013) mostra que, a inclusão de tecnologias voltadas para a confiabilidade do sistema de distribuição tem o potencial de reduzir em até 22,3 minutos o tempo anual médio que cada consumidor fica sem energia. Esse mesmo estudo sugere que o custo estimado do impacto de uma desenergização sobre a produtividade da sociedade seria de aproximadamente U\$1,80/min para cada consumidor.

Nesse contexto, os investimentos em melhorias no sistema elétrico são uma necessidade constante e, portanto, uma realidade enfrentada pelas empresas distribuidoras de energia, logo, percebe-se que, apesar das atividades de expansão e melhoria já serem executadas de forma programada, estas constituem tarefa não trivial, requerendo o aperfeiçoamento contínuo e atualização dos procedimentos, de forma a atender às exigências impostas pela evolução dos sistemas elétricos.

Investimentos requerem disponibilidade de recursos, que nem sempre são suficientes. O que geralmente ocorre é que a quantidade de recursos disponíveis não é suficiente para cobrir o custo de todas as melhorias necessárias. Surge, então, a necessidade de priorizar as ações de melhoria do sistema elétrico. Para fazer esta priorização, deve-se de antemão conhecer a situação real das instalações. Quanto mais conhecimento disponível, melhor para uma correta quantificação das alternativas de investimento bem como das prioridades.

Nesse sentido, a união de variáveis utilizadas na área de engenharia de tráfego (tempo de deslocamento, velocidade médias, volume, etc.) com dados vinculados à operação da rede elétrica, obtidos através da base de registros mais recorrentes de desligamentos por meio da técnica de Descoberta de Conhecimento em Base de Dados (do anglicanismo, *Knowledge Discovery in Data Bases - KDD*), em algoritmos de aprendizado não supervisionado para o aprimoramento dos processos de tomada de decisão se apresenta como uma abordagem interessante para o problema de mitigação de faltas e priorização de investimentos em sistemas de distribuição.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo principal do presente trabalho é propor uma metodologia para mitigar faltas no SDMT, bem como avaliar o desempenho das soluções presentes na rede de distribuição a partir da aquisição de informações disponíveis em base de dados preexistentes, considerando aspectos do SEP, dos recursos de manobra disponíveis e do tráfego nas vias de acesso. O método proposto busca desenvolver um sistema capaz de analisar de forma rápida e eficiente, a agregação de várias variáveis e suas inter-relações no impacto de uma falta na rede, permitindo então a análise de ações de investimento capazes de mitigá-las de forma a otimizar os indicadores de qualidade do serviço e maximizar os benefícios em relação à confiabilidade da rede de distribuição de media tensão.

1.3.2 Objetivos Específicos

- a) Elencar variáveis significativas para os blocos de rede, formando um vetor multidimensional que leva em consideração características elétricas, lógicas e de acessibilidade (critério multivariável);
- b) Compor uma modelagem heterogênea de dados da rede de distribuição através de um processo de *KDD* para construção de uma base de dados a partir de um conjunto incompleto de informações;
- c) Identificar organizações não evidentes nos dados (similaridades físicas e diferenças tecnológicas) através da criação de agrupamentos consistentes, utilizando técnicas de aprendizado não supervisionado.
- d) Avaliar a importância dos Sistemas de Suporte a Decisão Espacial como provedor de dados em Sistemas de Distribuição de Energia;
- e) Propor uma metodologia que, a partir dos dados extraídos, esteja apta a estimar de forma rápida e precisa um conjunto de ações previstas nas redes de distribuição modernas e o peso das ações realizadas sobre a confiabilidade e seus custos associados;
- f) Realizar estudo de caso em rede real.

1.4 Revisão Bibliográfica

Uma revisão bibliográfica criteriosa sobre os temas abordados, e a busca pelo estado da arte da pesquisa, são fundamentais para estabelecer os limites da pesquisa e apresentar as contribuições do trabalho. Nesta seção é apresentada uma revisão bibliográfica sobre os principais tópicos envolvidos na dissertação: aspectos relacionados ao planejamento como forma de melhorar a confiabilidade de sistemas elétricos de distribuição, a utilização de sistemas de suporte a decisão espacial como provedor de dados e a utilização de técnicas de mineração de dados com ênfase para a aplicação de técnicas de clusterização como etapa do processo de Descoberta de Conhecimento em Banco de Dados, comumente referenciada pela sigla inglesa *KDD*.

Oliveira et al. (2010) propõe o planejamento estático de sistemas de distribuição como um problema de programação linear inteiro misto. São levados em consideração os objetivos de médio e longo prazo e ações de curto prazo como alocação de equipamentos. Desta forma, para avaliar estes fatores de maneira simultânea, implementam-se duas técnicas: um Algoritmo Construtivo Especialista (AHC) e um algoritmo *Branch-and-Bound* não linear. O modelo proposto tem por objetivo minimizar custos de operação e construção, sujeito às restrições de variação de potência, magnitude de tensão, capacidade dos circuitos e subestações, radialidade e controle de *TAP*.

Em Sousa (2013), para priorização de projetos em sistemas de distribuição, são apresentados dois diferentes modelos de programação linear binária mista (*MBLB*). O primeiro efetua o planejamento ótimo do sistema sem considerar a confiabilidade e o segundo visa obter a melhor solução para o problema do planejamento, considerando alocação de chaves para situações de contingência. Dessa forma, através de técnicas clássicas, são abordados os custos dos investimentos, aliado ao comportamento da operação do sistema, e os custos operacionais através do cálculo de energia não distribuída.

Mussoi e Teive (2013) sugere uma metodologia para maximização do valor de portfólio dos projetos como parte do processo de planejamento de expansão. Para tal, consideram-se os impactos financeiros, número de consumidores e aspectos relacionados com a confiabilidade do sistema e qualidade de energia. A otimização do modelo de decisão é realizada através de um algoritmo genético multiobjetivo (*NSGA-II*) e a solução é obtida por meio da fronteira ótima de Pareto, composta por um conjunto de compromissos com os portfólios e as restrições dos projetos. A metodologia reduz a subjetividade nas decisões e prioriza projetos sistematicamente com base nos dados disponíveis.

Na área de pesquisa sobre a utilização de sistemas *web* para simulação de tráfego nas vias de acesso as novas ferramentas de análise (tanto histórico, como tempo real), disponíveis através de *Application Programming Interface (API)* de acesso *web*, se apresentam como importantes provedores de dados e são amplamente vinculados com uma eficiente operação

de distribuição de energia. Apesar deste fato, pouco se explorou no setor elétrico além do simples despacho de equipe.

O uso de Sistemas de Suporte a Decisão (*DSS – Decision Support Systems*) que auxiliem a tomada de decisão, solucionando as diversas demandas e minimizando os erros é essencial (ARNOTT; PERVAN, 2005), principalmente quando o ambiente de decisão é dinâmico.

Segundo Brail e Klosterman (2001) o uso de Sistemas de Suporte a Decisão (*DSS – Decision Support Systems*) podem ser compostos por um *GIS (Geographical Information Systems)*. Assim, esse sistema pode incorporar, além de ferramentas tradicionais do *DSS*, também pode incorporar aquelas referentes aos *GIS*. Assim, a necessidade de se tomar decisões pelos gestores em diversos campos do conhecimento humano, com base na informação geográfica levou ao desenvolvimento dos Sistemas de Suporte a Decisão Espacial (*SDSS – Spacial Decision Support Systems*).

Em Armstrong e Densham (1990) o *SDSS* é definido como um sistema que incorpora os componentes essenciais de um *DSS*, como banco de dados (espaciais e não espaciais), modelos analíticos e de simulação. Dos *GIS*, os *SDSS* herdaram o georreferenciamento da informação (atrelar a localização geográfica a cada informação do banco de dados), a possibilidade de usar diferentes componentes geográficas de forma integrada ou em separado e a interface direcionada ao tratamento de informações geográficas. Assim, o *SDSS* dispõe de estrutura para integrar: (i) capacidades de modelagem analítica e espacial; (ii) gerenciamento de dados espaciais e não-espaciais; (iii) conhecimento de domínio; (iv) capacidades de exibição espacial; e (v) capacidades de gerar relatórios.

Em Santos et al. (2011) é apresentado um sistema de sistema de suporte a decisão espacial baseado na *web* com o objetivo de gerar rotas de veículos otimizadas para vários problemas de roteamento de veículos que visam atender à demandas localizadas ao longo dos arcos de uma rede de transporte. O *Web-SDSS* incorpora o *GoogleMapsTM*, um banco de dados, uma heurística e uma meta-heurística baseada no algoritmo de Colônia de Formigas para gerar rotas e mapas de rotas de veículos individuais detalhadas. O sistema foi testado para coleta de lixo urbano em Coimbra, Portugal.

Na área de pesquisa sobre a utilização de métodos inteligentes para manipulação de dados, existem vários trabalhos no ramo computacional chamado Descoberta de Conhecimento em Base de Dados. Dedicar-se a pesquisa bibliográfica à etapa operacional da mineração de dados com foco às técnicas de clusterização ou agrupamento de dados.

Assim como em muitas outras áreas, em sistemas de potência existe um grande crescimento na quantidade de dados armazenada pelos sistemas de informação. Segundo Olaru et al. (1999), as principais fontes de desses dados são: (i) dados de campo, que podem ser coletados através de diversos dispositivos ao longo do sistema, (ii) sistema *SCADA*,

(iii) dados obtidos através de ambientes de simulação de planejamento ou operação. Esse enorme conjunto de dados torna difícil para um profissional a tarefa de entendê-lo como um todo. Nesse contexto é que a utilização de algoritmos capazes de sintetizar estruturas a partir de dados se torna uma necessidade. Implementar e aplicar esses algoritmos a problemas reais é o que se propõe o campo de mineração de dados. Esse artigo descreve diversas ferramentas utilizadas na mineração de dados, exemplificando através de um software desenvolvido para estudos de avaliação da segurança dinâmica.

No trabalho realizado em [Peng et al. \(2004\)](#) equipes de manutenção coletaram cada falha ocorrida no sistema numa tabela que incluía hora, data, mês, ano, endereço, equipamento em que ocorreu a falha, causa ou acidente, e etc. O banco de dados acumulou uma grande base de informação durante muitos anos. Nesse foram utilizados dados brutos e mineração de dados para derivar padrões e regras para as o diagnóstico de falhas ocorridas nos equipamentos da rede de distribuição e localização de falhas.

No artigo [Nassar et al. \(2004\)](#) é feita a modelagem de uma Rede *Bayesiana* desenvolvida a partir do conhecimento de *clusters* (ou agrupamentos) climáticos e sua relação com o número de paradas não programadas, utilizando-se uma base de dados considerando as observações diárias num período de 12 meses. Primeiramente, foram identificados os padrões climáticos mensais através de técnicas estatísticas e redes neurais. Posteriormente, investigou-se a relação dos *clusters* climáticos com o número de paradas não programadas. Finalmente, foi desenvolvida uma Rede *Bayesiana* que se constitui num cenário que propaga a informação climática diária de forma a oferecer uma previsão do número de paradas emergenciais. Dessa forma, a Rede *Bayesiana* pode ser útil como apoio à gestão de equipes de atendimentos a falhas no sistema elétrico.

O artigo [Mori \(2006\)](#) apresenta um estudo sobre o estado da arte em mineração de dados para sistemas de potência. Devido à não linearidade existente nos sistemas de potência, o autor ressalta a importância da utilização de técnicas de extração de conhecimento de banco de dados para suporte da operação e no planejamento. A mineração de dados é uma etapa de todo o processo de extração de conhecimento. Alguns métodos utilizados em mineração de dados são citados e sucintamente descritos. Ainda são apresentadas 42 referências sobre trabalhos que dão uma visão geral sobre o estado da arte em aplicações de mineração de dados para sistemas de potência.

Em [Madan et al. \(1997\)](#), é descrito uma série de aplicações de mineração de dados para sistemas de potência. O autor diz que técnicas de inteligência artificial, principalmente sob a forma de sistemas especialistas, são utilizadas como ferramentas de apoio para os operadores. No entanto, ele lembra que essas técnicas possuem algumas limitações. O desempenho do sistema especialista irá depender da quantidade e da qualidade da informação utilizada com base de conhecimento. É difícil construir sistemas especialistas que possam capturar totalmente o conhecimento de especialistas humanos. As técnicas

de mineração de dados são utilizadas para contornar esse problema. Conhecida também como extração de conhecimento em banco de dados, esse processo permite organizar a informação de uma forma mais adequada para a utilização em algoritmos de aprendizado de sistemas especialistas.

1.5 Estrutura do Trabalho

Para facilitar a compreensão, o trabalho foi dividido em 7 Capítulos. No [Capítulo 1](#) desenvolve-se uma breve introdução ao tema estudado, apresentam-se as motivações e objetivos da pesquisa.

No [Capítulo 2](#) apresenta-se brevemente a constituição do Sistema Elétrico de Potência (SEP) e a forma com que seus elementos se interconectam, são apresentados os principais indicadores de qualidade de energia elétrica abordados no procedimento de distribuição e nas normas vigentes, como funciona o pagamento das compensações pelas distribuidoras e o custo da interrupção no Brasil, além de aspectos referentes a conceituação, caracterização, desenvolvimento e implantação das Redes Elétricas Inteligentes (REIs).

No [Capítulo 3](#) contextualiza-se a evolução dos Sistemas de Suporte a Decisão (*SSD*), são definidos os principais componentes e a arquitetura de um Sistema de Suporte a Decisão Espacial (*SSDS*) e apresenta-se o software SIMAUTEK, desenvolvido na CEMAR, para suporte a esta tecnologia.

No [Capítulo 4](#) são apresentados os conceitos e etapas do processo de descoberta de conhecimento em base de dados, discorre-se sobre a técnica de agrupamento e a heurística utilizada como base para solução do problema proposto.

No [Capítulo 5](#) se encontra a metodologia, com detalhamento de todas etapas realizadas, definição dos atributos relevantes para o sistema de distribuição, simulação dos tempos de deslocamento a partir do *SSDS* e implementação do módulo de aprendizado não supervisionado.

No [Capítulo 6](#) são descritas as simulações realizadas e os resultados obtidos são apresentados e discutidos.

Finalmente, no [Capítulo 7](#) as conclusões e sugestões para trabalhos futuros são apresentados.

2 Aspectos Relacionados aos Sistemas de Distribuição de Energia Elétrica e às Redes Elétricas Inteligentes

Neste Capítulo é apresentada, de forma sucinta, a composição básica dos sistemas elétricos de potência e descrita a forma com que seus elementos se interconectam desde a geração até a distribuição de energia aos consumidores finais. A seguir, são delineados os métodos de avaliação da qualidade de serviço e o impacto econômico que estes têm nos consumidores e no próprio operador da rede de distribuição. São apresentadas características, funcionalidades, benefícios e desafios para o desenvolvimento e implantação da Rede Elétrica Inteligente (REI) como uma prática para a melhoria da qualidade.

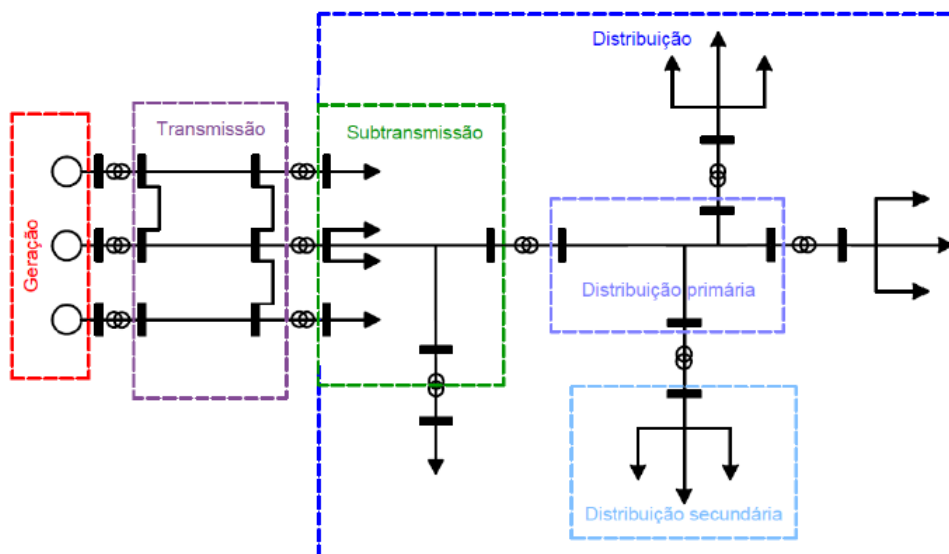
2.1 Aspectos Relacionados aos Sistemas de Distribuição

Com o objetivo de contextualizar o problema, nesta seção são abordadas as características de um sistema de distribuição a fim de identificar os elementos a serem considerados nos estudos. Os SEP são definidos como o conjunto de todas as instalações e equipamentos destinados à geração, transmissão e distribuição de energia elétrica. Os sistemas elétricos de potência podem ser subdivididos em três grandes blocos (KAGAN et al., 2010):

- **Geração** – destinado para o papel de conversão de alguma forma de energia (mecânica, térmica, entre outros) em energia elétrica.
- **Transmissão** – responsável pelo transporte de energia elétrica dos centros de produção aos centros consumidores.
- **Distribuição** – encarregado de distribuir a energia elétrica recebida do sistema de transmissão aos grandes, médios e pequenos consumidores.

Na [Figura 1](#), tem-se um diagrama unifilar, uma maneira bastante usual de se representar o SEP, por meio da qual se observa de forma gráfica como os blocos de um sistema elétrico típico se interconectam. Nesta figura, nota-se a presença de um bloco de geração formado por três usinas, um conjunto de linhas de transmissão, uma rede de subtransmissão, uma rede de distribuição primária (média tensão) e três redes de distribuição secundárias (baixa tensão) (KAGAN et al., 2010).

Figura 1 – Caracterização do Sistema Elétrico.



Fonte: Adaptação de KAGAN et al.

No Brasil, os valores eficazes das tensões, com frequência de 60 Hz são fixados por decreto do Ministro de Minas e Energia. Na tabela [Tabela 1](#) estas tensões de atendimento aos consumidores, anteriormente mencionadas, são apresentadas de acordo com o campo de aplicação e a área do sistema de potência a que se referem. Ressalta-se que o objetivo da pesquisa desenvolvida tem por foco a análise dos sistemas de distribuição, mais especificamente no campo de aplicação do SEP aqui nomeado como Distribuição Primária, o qual será trabalhado no desenvolvimento deste trabalho (KAGAN et al., 2010).

Tabela 1 – Tensões usuais em SEP

Tensão Padronizada (kV)	Campo de Aplicação	Área do Sistema de Potência
220 / 127 (V)	Distribuição Secundária	
380 / 220 (V)		
13.8	Distribuição Primária	Distribuição
34.5		
34.5	Subtransmissão	
69		
138		
138	Transmissão	Transmissão
230		
345		
500		

Fonte: Kagan et al. (2010)

Seguindo a [Figura 1](#), observa-se que para a ligação entre o bloco de Geração e o bloco de Transmissão inicialmente é realizada por uma Subestação Elevadora de Transmissão, está elevação da tensão de geração para a tensão de transmissão é devido às distâncias a serem percorridas e o montante de energia a ser transportado. A seguir, o Sistema de Transmissão se encarrega de transportar a energia dos centros de produção até os centros consumidores.

Caso existam consumidores que necessitem de energia nos níveis de tensão de transmissão, os mesmos são alimentados diretamente a partir destes ramais do sistema. Caso contrário, chegando aos centros de consumo, na Subestação Abaixadora de Subtransmissão, ocorre o primeiro abaixamento de tensão. Esta, então, é distribuída pelo Sistema de Subtransmissão para grandes consumidores atendidos por níveis de tensão de subtransmissão e para as Subestações de Distribuição, nas quais se sucede o segundo rebaixamento dos níveis de tensão, reduzindo-se a tensão de subtransmissão para tensão de distribuição primária, ponto este que é o foco dos estudos da presente pesquisa.

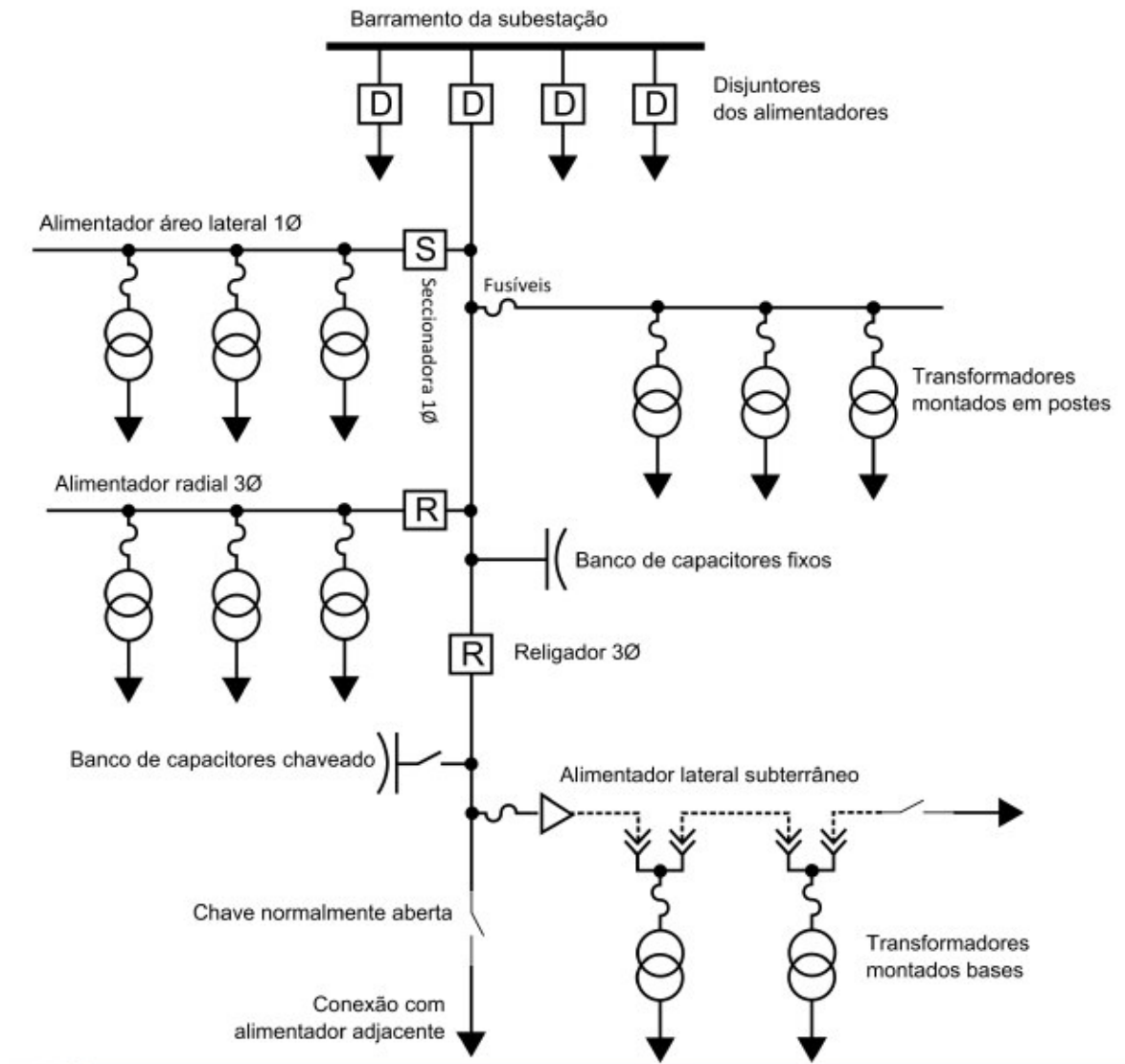
A seguir, o Sistema de Distribuição Primária atende os consumidores de médio porte em tensão de distribuição primária (média tensão), levando energia até os Transformadores de Distribuição. Os Transformadores de Distribuição reduzem a tensão de distribuição primária e assim o Sistema de Distribuição Secundária distribui energia aos pequenos consumidores em tensão de distribuição (baixa tensão).

Existem diversas configurações de sistemas de distribuição, porém a mais usual é a radial (primária e secundária). A configuração radial possui muitas vantagens. Podemos destacar as seguintes ([SHORT, 2014](#)):

- Facilidade no controle de tensão;
- Facilidade na implementação de proteção de sobrecorrente;
- Facilidade na previsão e controle do fluxo de potência;
- Baixas magnitudes de corrente comparando-se com outros circuitos;
- Baixo custo.

A Figura 2 ilustra um sistema de distribuição primária, contemplando diferentes tipos de alimentadores laterais e componentes do sistema.

Figura 2 – Sistema de distribuição primária com diferentes laterais.



Fonte: Adaptação de KAGAN et al.

Conforme ilustrado na Figura 2, na configuração radial os alimentadores partem de um barramento comum na subestação, após o qual são instalados disjuntores, comandados por relés de sobrecorrente e relés de religamento, que efetuam a proteção individual de cada alimentador. Ao longo dos alimentadores podem ser instalados outros dispositivos de proteção e controle, tais como: religadores, chaves telecomandadas, seccionizadores, chaves fusíveis repetidoras, chaves religadora e dispositivos para facilitar a localização de faltas.

Os alimentadores são compostos de um circuito principal, denominado tronco do alimentador, geralmente trifásico, com condutores de bitolas superiores aos demais circuitos. Os ramais laterais são circuitos também primários que derivam diretamente do tronco do alimentador. Os circuitos que derivam de ramais laterais são denominados sub-ramais. Os ramais laterais e sub-ramais podem ser trifásicos, bifásicos ou monofásicos (GÖNEN, 2015; FERREIRA, 2009).

No momento em que se implanta um sistema de distribuição de energia elétrica, a população local imediatamente passa a ter inúmeros benefícios nas atividades produtivas e de lazer, criando-se uma dependência natural. A percepção da dependência de energia aparece quando os clientes constatarem interrupções no fornecimento que interfere no desenvolvimento de suas atividades e, nesse sentido, é natural que se estabeleça um processo de discussão quanto a qualidade do serviço prestado. Desta forma, cabe ao Poder Público concedente, através de seus organismos de regulação e fiscalização, zelar pela qualidade da energia elétrica que será entregue aos clientes, através da emissão de leis e regulamentos que estabeleçam os critérios mínimos de qualidade a cumprir por parte das concessionárias, bem como os deveres de todos os clientes.

Em termos gerais, os indicadores de continuidade do fornecimento de energia monitoram a duração e frequência das interrupções. A partir dos indicadores se estabelecem padrões mínimos de qualidade tanto para clientes (valores individuais) quanto para agrupamentos (conjuntos, concessionárias, etc.). O prejuízo causado pela violação dos padrões mínimos de qualidade é avaliado para cada cliente em forma de compensação que se reflete diretamente na tarifa de energia.

Para melhor definir estes indicadores que serão utilizados na metodologia do trabalho, a próxima seção vai tratar dos aspectos relacionados à qualidade de energia elétrica com foco para qualidade do serviço.

2.2 Qualidade de Energia Elétrica: Qualidade do Serviço

De modo geral o marco regulatório da qualidade da energia elétrica está segregado em regulamentos que tratam separadamente a qualidade do produto, a qualidade comercial e a qualidade do serviço (ANEEL, 2018b).

- A qualidade comercial trata os prazos de execução dos serviços estabelecidos e as penalidades em caso de violação de atividades que envolvem, dentre outros: obras para atendimento inicial, formalização dos contratos, atividades regulares do ciclo de faturamento como: leitura, faturamento, entrega da fatura, cobrança e o pagamento; atendimento telefônico e reclamações.

- A qualidade do produto avalia a conformidade de tensão em regime permanente e as perturbações na forma de onda de tensão;
- A qualidade do serviço estuda os impactos no fornecimento de energia elétrica.

Para o âmbito deste estudo a qualidade comercial e a qualidade do produto não serão alvos da análise, sendo apenas relevante a qualidade de serviço. Na qualidade do serviço, o órgão regulador deve garantir o equilíbrio entre os clientes e os agentes prestadores de serviço, a fim de reduzir o prejuízo causado pelas interrupções (diz respeito aos clientes) e as incertezas associadas à tomada de decisão sobre os investimentos que serão remunerados (diz respeito às empresas) (ANEEL, 2018a).

2.2.1 Indicadores de Continuidade e Tempo de Atendimento

A qualidade do serviço prestado pelas distribuidoras, assim como o desempenho geral do sistema elétrico são avaliados pela ANEEL - Agência Nacional de Energia Elétrica (ANEEL, 2018a), com base em indicadores coletivos e individuais. Em termos coletivos, a ANEEL analisa o desempenho das empresas não apenas através dos valores apurados de cada indicador em relação aos padrões regulatórios de cada empresa, mas também através da comparação para gerar o *Ranking* de Qualidade das distribuidoras. Em termos de indicadores individuais, as distribuidoras de energia elétrica devem apurar a qualidade do serviço prestado a cada cliente ou unidade consumidora através dos indicadores apresentados na Tabela 2. De acordo com o módulo 8 do PRODIST (ANEEL, 2018b), esses indicadores são a base de cálculo dos indicadores coletivos.

Os indicadores DIC (Duração de Interrupção Individual por Unidade Consumidora) e FIC (Frequência de Interrupção Individual por Unidade Consumidora) indicam por quanto tempo e o número de vezes, respectivamente, que uma unidade consumidora ficou sem energia elétrica durante os períodos de apuração mensais, trimestrais e anuais. O DMIC (Duração Máxima de Interrupção Individual por Unidade Consumidora) é um indicador que monitora o tempo máximo das interrupções mensais. O DICRI (Duração da Interrupção Individual ocorrida em dia crítico por Unidade Consumidora) é um indicador específico para interrupções ocorridas em dias críticos, onde a quantidade de ocorrências emergenciais, em um determinado conjunto de unidades consumidoras, supera a média acrescida de três desvios padrões dos valores diários dos 24 meses anteriores ao ano em curso (ANEEL, 2018b).

Os indicadores individuais são a base para o cálculo dos indicadores coletivos que representam parte do sistema (conjunto) ou o sistema como um todo (empresa). Assim a qualidade do fornecimento, a nível de sistema, é monitorada através dos indicadores DEC (Duração equivalente da Interrupção por Unidade Consumidora) e FEC (Frequência Equivalente de Interrupção por Unidade Consumidora).

Tabela 2 – Indicadores individuais de continuidade do serviço.

INDICADORES INDIVIDUAIS	EXPRESSÃO	DESCRIÇÃO
DIC		
(Duração de Interrupção Individual por Unidade Consumidora)	$DIC = \sum_{i=1}^n t(i)$	Indica o intervalo de tempo acumulado que, no período de apuração, em cada unidade consumidora ocorreu interrupção da distribuição de energia elétrica [horas e centésimos de horas].
FIC		
(Frequência de Interrupção Individual por Unidade Consumidora)	$FIC = n$	Indica o número de interrupções ocorridas, no período de apuração, em cada unidade consumidora.
DMIC		
(Duração Máxima de Interrupção Individual por Unidade Consumidora)	$DIC = t(i)_{max}$	Indica o intervalo de tempo máximo e contínuo em que ocorreu interrupção da prestação de serviço em uma unidade consumidora, no período de apuração [horas e centésimos de horas]
DICRI		
(Duração da Interrupção Individual ocorrida em dia crítico por Unidade Consumidora (horas))	$DICRI = t_{critico}$	Indica o intervalo de tempo que, no dia crítico, em cada unidade consumidora ocorreu interrupção da distribuição de energia elétrica, [horas e centésimos de horas].

i = Índice de interrupção da unidade consumidora , no período de apuração, varia de 1 a n .
 n = Número de interrupções da unidade consumidora considerada, no período de apuração.
 $t(i)$ = Duração de cada interrupção (i) da unidade consumidora considerada, no período de apuração.
 $t(i)_{max}$ = Tempo da máxima duração de interrupção contínua (i), no período de apuração, na unidade consumidora considerada, expresso em horas e centésimos de horas.
 $t_{critico}$ = Duração da interrupção ocorrida em dia crítico.

Fonte: Adaptação de ANEEL (2018b)

O DEC indica o número de horas, em média, que uma unidade consumidora fica sem energia elétrica durante períodos de tempo mensais, trimestrais e anuais. Já o FEC indica quantas vezes, em média, houve interrupções nas unidades consumidoras, no período considerado. Em um primeiro nível de controle coletivo, estes indicadores são circunscritos aos conjuntos de unidades consumidoras, conforme apresentado na Tabela 3.

Tabela 3 – Indicadores coletivos de continuidade do serviço.

INDICADORES COLETIVOS	EXPRESSÃO	DESCRIÇÃO
DEC (Duração Equivalente de Interrupção por Unidade Consumidora)	$DEC = \frac{\sum_{i=1}^{C_c} DIC(i)}{C_c}$	Indica o intervalo de tempo, em média, em que ocorreu descontinuidade da prestação de serviço em cada unidade consumidora do conjunto considerado, no período de apuração, em horas.
FEC (Frequência Equivalente de Interrupção por Unidade Consumidora)	$FEC = \frac{\sum_{i=1}^{C_c} FIC(i)}{C_c}$	Indica o número de vezes, em média, em que ocorreu descontinuidade da prestação de serviço em cada unidade consumidora do conjunto considerado, no período de apuração.

i = Índice de unidades consumidoras atendidas em BT ou MT faturadas do conjunto.
 C_c = Número total de unidades consumidoras faturadas do conjunto no período de apuração (mensal, trimestral ou anual), atendidas em BT e MT.

Fonte: Adaptação de ANEEL (2018b).

Ainda em ANEEL (2018b), o atendimento às ocorrências emergenciais deverá ser supervisionado, avaliado e controlado por meio de indicadores que expressam os valores vinculados a conjuntos de unidade consumidoras, são definidos os indicadores de tempo de atendimento referente às ocorrências emergenciais, conforme descrito na Tabela 4.

Tabela 4 – Indicadores de tempo de atendimento às ocorrências emergenciais.

INDICADORES DE TEMPO DE ATENDIMENTO	EXPRESSÃO	DESCRIÇÃO
TMP (Tempo Médio de Preparação)	$TMP = \frac{\sum_{i=1}^n TP(I)}{n}$	Indica o tempo médio de preparação da equipe de atendimento de emergências, expresso em minutos.
TMD (Tempo Médio de Deslocamento)	$TMD = \frac{\sum_{i=1}^n TD(I)}{n}$	Indica o tempo médio de deslocamento da equipe de atendimento de emergência, expresso em minutos.
TME (Tempo Médio de Execução)	$TME = \frac{\sum_{i=1}^n TE(I)}{n}$	Indica o tempo de execução do serviço até seu restabelecimento pela equipe de atendimento de emergência, expresso em minutos.
TMAE (Tempo Médio de Atendimento a Emergências)	$TMAE = TMP + TMD + TME$	Indica o tempo médio de atendimento a ocorrências emergenciais, representando o tempo médio para atendimento de emergência, expresso em minutos.

n = Número de ocorrências emergenciais verificadas no conjunto de unidades consumidoras
TP = Tempo de preparação da equipe de atendimento de emergência para cada ocorrência emergencial, expresso em minutos.
TD = Tempo de deslocamento da equipe de atendimento de emergência para cada ocorrência emergencial, expresso em minutos.
TE = Tempo de execução do serviço até seu restabelecimento pela equipe de atendimento de emergência para cada ocorrência emergencial.

Fonte: Adaptação de [ANEEL \(2018b\)](#).

Todos os aspectos regulatórios do assunto estão descritos no Módulo 8 dos Procedimentos de Distribuição - PRODIST ([ANEEL, 2018b](#)).

2.2.2 Compensações pelas Transgressões dos Limites de Continuidade

A Tabela 5 apresenta as fórmulas para cálculos das compensações para ultrapassagem dos limites estabelecidos dos indicadores de continuidade individuais, no período de apuração (ANEEL, 2018b).

Tabela 5 – Indicadores de compensação pela ultrapassagem dos limites dos indicadores individuais.

INDICADORES	EXPRESSÃO PARA CÁLCULO DAS COMPENSAÇÕES
DIC	$ValorDIC = \left(\frac{DICv}{DICp} - 1 \right) \times DICp \times \frac{EUSDmédio}{730} \times kei$
FIC	$ValorFIC = \left(\frac{FICv}{FICp} - 1 \right) \times DICp \times \frac{EUSDmédio}{730} \times kei$
DMIC	$ValorDMIC = \left(\frac{DMICv}{DMICp} - 1 \right) \times DMICp \times \frac{EUSDmédio}{730} \times kei$
DICRI	$ValorDICRI = \left(\frac{DICRIv}{DICRIp} - 1 \right) \times DICRIp \times \frac{EUSDmédio}{730} \times kei$
<i>DICv</i>	= Duração de interrupção por unidade consumidora ou por ponto de conexão, conforme cada caso, verificada no período considerado, expressa em horas e centésimos de hora.
<i>DICp</i>	= Limite de continuidade estabelecido no período considerado para o DIC por unidade consumidora ou por ponto de conexão, expresso em horas e centésimos de hora.
<i>FICv</i>	= Frequência de interrupção por unidade consumidora ou por ponto de conexão, conforme cada caso, verificada no período considerado, expressa em número de interrupções.
<i>FICp</i>	= Limite de continuidade estabelecido no período considerado para o FIC por unidade consumidora ou por ponto de conexão, expresso em número de interrupções e centésimo do número de interrupções.
<i>DMICv</i>	= Duração máxima de interrupção contínua por unidade consumidora ou por ponto de conexão, conforme cada caso, verificada no período considerado, expressa em horas e centésimos de hora.
<i>DMICp</i>	= Limite de continuidade estabelecido no período considerado para o DMIC por unidade consumidora ou por ponto de conexão, expresso em horas.
<i>EUSDmédio</i>	= Média dos encargos de uso do sistema de distribuição correspondente aos meses do período de apuração do indicador.
730	= Número de hora no mês
<i>kei</i>	= 15 para unidade em BT, 20 para unidades em MT, 27 para unidade em AT.

Fonte: Adaptação de ANEEL (2018b).

Desde a aprovação do PRODIST, em 2008, foi definido que a partir de 2009, a violação dos limites DEC e FEC nos períodos pré-definidos de apuração não mais implica pagamento de multa pela distribuidora à ANEEL. Atualmente, quando houver violação dos limites de continuidade DIC, FIC e DMIC no período de apuração mensal, ou dos limites DIC e FIC nos períodos de apuração trimestral ou anual, ou dos limites de DICRI (somente mensal) para interrupções em dia crítico, a empresa prestadora do serviço deverá calcular a compensação ao consumidor e efetuar o crédito na fatura, em até dois meses após o período de apuração.

As [Tabela 6](#) e [Tabela 7](#) apresentam as compensações de continuidade pagas pelas concessionárias no Brasil, no período de 2010 a 2017, considerando os dados mensais, trimestrais e anuais. Para o DIC, FIC e DMIC, as informações de compensação estão disponíveis a partir de 2010. Para o DICRI, as informações de compensações estão disponíveis a partir de 2012 ([ANEEL, 2018a](#)).

Tabela 6 – Compensações de continuidade pagas pelas concessionárias no Brasil, no período de 2010 a 2013.

Compensações de Continuidade das Concessionárias			2010	2011	2012	2013
DIC, FIC e DMIC	Mensal	QTD	64.528.029	68.376.650	66.031.346	67.586.765
		VLR	259.409.387,88	290.941.585,92	305.792.525,44	265.045.891,13
	Trimestral	QTD	18.311.034	21.500.368	19.139.127	20.071.661
		VLR	42.844.877,84	40.513.635,13	41.990.777,27	37.144.197,46
	Anual	QTD	12.252.627	15.260.861	12.882.181	13.716.452
		VLR	58.543.287,88	68.836.266,39	69.891.361,86	61.779.107,02
DICRI	QTD	0	0	9.170.700	3.378.948	
	VLR	0,00	0,00	19.796.625,84	14.584.068,41	
TOTAL GERAL			95.091.690	105.137.879	107.223.354	104.753.826
			360.797.553,60	400.291.487,44	437.471.290,41	378.553.264,02

Fonte: Adaptação de [ANEEL \(2018a\)](#).

Tabela 7 – Compensações de continuidade pagas pelas concessionárias no Brasil, no período de 2014 a 2017.

Compensações de Continuidade das Concessionárias			2014	2015	2016	2017
DIC, FIC e DMIC	Mensal	QTD	69.433.683	78.455.430	73.450.080	67.515.365
		VLR	279.043.522,19	450.176.623,42	405.084.576,21	346.651.524,11
	Trimestral	QTD	20.546.783	23.656.339	20.694.645	19.100.488
		VLR	38.215.040,48	64.540.627,96	59.469.423,39	56.291.118,10
	Anual	QTD	13.032.543	16.064.721	13.597.798	11.827.829
		VLR	54.988.606,81	102.218.280,49	84.604.859,24	70.926.779,44
DICRI	QTD	7.195.740	7.024.954	4.293.725	1.908.786	
	VLR	18.513.244,11	39.957.952,19	26.429.211,32	17.000.875,95	
TOTAL GERAL			110.208.749	125.201.444	112.036.248	100.352.468
			390.760.413,59	656.893.484,06	575.588.070,16	490.870.297,60

Fonte: Adaptação de [ANEEL \(2018a\)](#).

Observa-se nas [Tabela 6](#) e [Tabela 7](#), que a partir de 2010, quando tem início o pagamento das compensações, nota-se que houve aumento anual dos valores, considerando o total geral apresentado, o que pode ser entendido como uma dificuldade das concessionárias em se adequar aos limites de continuidade estabelecidos pelo órgão regulador.

2.2.3 Análise Econômica das Interrupções de Serviço

Diante da importância da qualidade do serviço de distribuição de energia elétrica, a ANEEL apresenta através da Nota Técnica nº 0173/2016-SRD/ANEEL, uma análise da regulamentação da continuidade do fornecimento de energia elétrica, com enfoque sobre a avaliação dos custos relacionados à confiabilidade do serviço de distribuição ([ANEEL, 2016](#)), com objetivo de incorporar o custo da interrupção na regulação da qualidade.

Os custos da interrupção representam o prejuízo do consumidor diante de uma interrupção no fornecimento de energia elétrica, sendo diferente para cada tipo de consumidor. O custo da energia não suprida (CENS) é o custo decorrente da energia que deixou de ser consumida por causa de uma interrupção e é contabilizado através dos custos diretos e custos indiretos. Os custos diretos estão ligados aos custos referentes à perda de produção, perda de matéria prima, retomada dos processos de produção. Já os custos indiretos estão associados aos investimentos em equipamentos ou instalações para adequar o nível de confiabilidade desejado ([ANEEL, 2016](#)).

Para estimar o CENS para o Brasil foram consideradas algumas questões tais como o momento do dia, classe do dia (útil ou final de semana) e a probabilidade de ocorrência de interrupções. Particularmente, foram desenvolvidas curvas de valor de ócio horário e diário, com base nos padrões de consumo residencial, disponibilidade de equipamentos elétricos, além da inclusão das curvas de carga dos setores analisados. No que diz respeito ao método do Valor Adicionado, também foram construídas curvas de carga com base na quantidade de empresas em operação em cada hora do dia, diferenciando assim o dia útil do final de semana. Em ambas as categorias (residencial e não residencial) foram alocadas probabilidades de ocorrências que possibilitaram calcular valores esperados das interrupções.

Os valores estimados para o CENS são consistentes com a realidade do país e são apresentados na [Tabela 8](#) e [Tabela 9](#)

Tabela 8 – Custo das Interrupções programadas por região e setor (R\$/kWh).

Setor	Nacional	Norte	Nordeste	Centro-Oeste	Sudeste	Sul
Industrial	1,8	3,1	1,5	1,8	2,3	0,9
Comercial e Serviços	2,4	1,2	1,1	4,0	3,1	1,1
Poder Público	1,6	1,2	1,5	2,0	1,6	1,8
Serviço Público	1,0	1,5	0,5	1,5	1,0	1,5
Rural e Rural Irrigante	1,6	1,6	1,0	2,9	1,1	2,4
Residencial	2,6	2,5	2,4	2,1	2,7	2,5
Baixa Renda	0,6	0,6	0,6	0,5	0,4	0,6
Média Região	2,0	2,0	1,5	2,5	2,4	1,5

Fonte: Adaptação de ANEEL (2016).

Tabela 9 – Custo das Interrupções não programadas por região e setor (R\$/kWh).

Setor	Nacional	Norte	Nordeste	Centro-Oeste	Sudeste	Sul
Industrial	13,9	23,7	11,7	13,8	17,3	6,6
Comercial e Serviços	18,1	9,6	8,3	30,6	24,1	8,7
Poder Público	12,3	9,2	11,3	15,5	12,4	13,9
Serviço Público	7,5	11,6	3,9	11,5	7,6	11,4
Rural e Rural Irrigante	12,4	12,4	8,1	22,0	8,2	18,4
Residencial	19,7	18,9	18,4	16,5	20,8	19,0
Baixa Renda	4,4	4,9	4,9	4,1	3,2	4,5
Média Região	15,7	15,7	11,3	19,2	18,8	11,6

Fonte: Adaptação de ANEEL (2016).

Os valores apresentados são uma estimativa conforme a metodologia aplicada, sendo assim algumas questões precisam ser melhor avaliadas, se os valores de custo de energia não fornecida estão de acordo com as realidades nacional e regional, qual o período para atualizar os valores, se há possibilidade de separar os valores por áreas de concessão, para assim conseguir apresentar valores com maior nível de certeza.

2.2.4 Análise BRASIL X CEMAR

Esta seção visa apresentar os indicadores DEC e FEC apurados e as compensações que foram pagas aos consumidores pelas interrupções no fornecimento de energia, considerando os indicadores apresentados para CEMAR e Brasil. Considerando que a classe de consumo com maior representação na CEMAR é a classe residencial (ANEEL, 2018a), este estudo vai focar no custo da interrupção total estimada para o consumidor residencial em relação às compensações pagas.

A Companhia Energética do Maranhão - CEMAR, sociedade anônima de capital aberto, com sede na cidade São Luís, no Estado do Maranhão, controlada pela Equatorial Energia S.A., é a concessionária do serviço público de distribuição de energia elétrica e atividades associadas ao serviço de energia elétrica, podendo prestar serviços técnicos de sua especialidade na área de concessão que abrange todo o Estado do Maranhão com 331.937 km², atendendo 2.468.119 consumidores em 217 municípios, suas atividades regulamentadas e fiscalizadas pela Agência Nacional de Energia Elétrica - ANEEL, vinculada ao Ministério de Minas e Energia - MME. A Companhia possui suas ações negociadas unicamente no Mercado de Balcão Organizado da B3 (GERAIS-CEMAR, 2017).

O sistema de distribuição de energia da CEMAR atualmente é composto por 144 subestações com capacidade instalada de aproximadamente 3.005 MVA, 103.741 km de linhas de distribuição aérea, além de 5.114 km de linhas de subtransmissão. Para realizar a supervisão e a operação de uma rede tão extensa, a empresa investiu constantemente nos últimos anos em equipamentos para realizar a automação na rede, sistemas de comunicação para enviar as informações coletadas às subestações e ao centro de operação, além da digitalização das informações de supervisão, controle e proteção presentes na subestação (GERAIS-CEMAR, 2017).

Para o cálculo da demanda (kW), levantaram-se os dados de consumo de energia elétrica (MWh) do setor residencial e por fim para estimar o custo da interrupção de energia para os consumidores residenciais (R\$/kWh), especificado para cada região do Brasil, conforme apresentado pela Nota Técnica nº 0173/2016-SRD/ANEEL (ANEEL, 2016).

Levantamento dos indicadores de continuidade DEC e FEC e compensações pagas

Os indicadores de continuidade DEC e FEC, apurado (AP) e limite (LI), são utilizados para se ter uma visão geral da qualidade do serviço prestado pelas distribuidoras aos clientes, além de possibilitar a comparação entre os indicadores praticados. Os valores de DEC e FEC foram obtidos entre 2005 e 2017, para que se possa acompanhar o desempenho ao longo do período, na busca pela melhoria da qualidade de energia elétrica através dos indicadores de continuidade, conforme apresentado na Tabela 10.

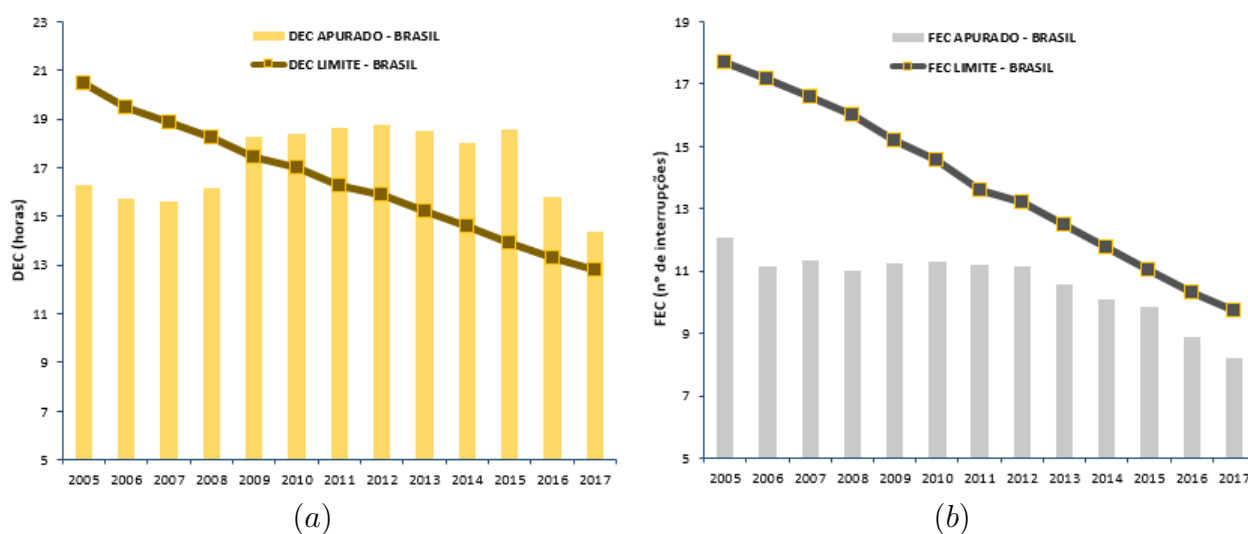
Tabela 10 – Indicador de continuidade DEC e FEC apurado e os limites estabelecidos para o Brasil e CEMAR, no período de 2005 a 2017.

	BRASIL				CEMAR			
	DEC AP	DEC LI	FEC AP	FEC LI	DEC AP	DEC LI	FEC AP	FEC LI
2005	16,26	20,44	12,09	17,71	54,56	79,48	32,90	47,60
2006	15,73	19,48	11,17	17,17	42,40	63,03	24,55	45,34
2007	15,63	18,85	11,36	16,57	28,59	58,55	19,78	39,20
2008	16,17	18,26	10,99	15,99	27,19	53,95	16,75	34,35
2009	18,26	17,44	11,27	15,19	23,45	46,24	15,10	29,70
2010	18,42	17,01	11,31	14,54	21,41	40,10	13,79	26,01
2011	18,61	16,23	11,21	13,61	21,44	34,18	11,60	23,52
2012	18,78	15,87	11,17	13,19	21,64	29,38	10,91	21,02
2013	18,49	15,19	10,60	12,47	18,85	25,19	10,88	18,42
2014	18,03	14,58	10,08	11,77	17,00	23,37	11,01	16,71
2015	18,60	13,94	9,86	11,03	15,28	21,68	8,95	15,03
2016	15,81	13,31	8,87	10,33	14,22	19,92	7,51	13,35
2017	14,35	12,77	8,20	9,72	13,28	18,58	7,14	11,95

Fonte: Adaptação de ANEEL (2018a).

O perfil de DEC e FEC para o Brasil, conforme apresentado na Figura 3, mostra que os limites do FEC não foram ultrapassados, no período em questão. Ao contrário para o DEC, que a partir de 2009, houve transgressões conforme os limites foram diminuindo. O que pode ser explicado como uma dificuldade na adequação das concessionárias a limites mais rigorosos.

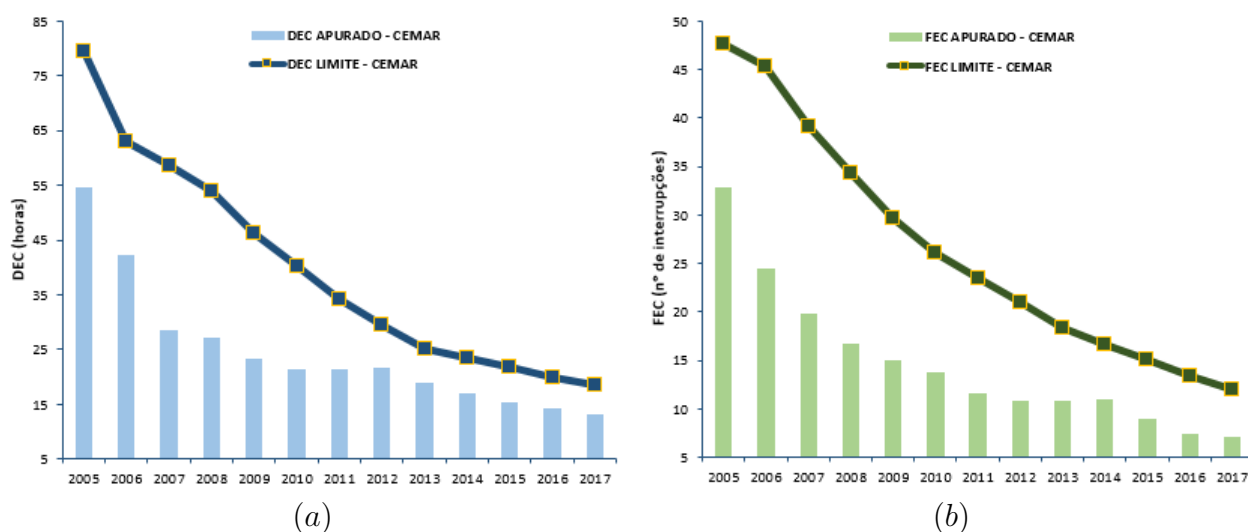
Figura 3 – (a) Evolução DEC Brasil e (b) Evolução FEC Brasil



Fonte: Adaptação de ANEEL (2018a).

A Figura 4 apresenta o perfil de DEC e FEC para a concessionária CEMAR, visto que não há ultrapassagem dos limites para os indicadores DEC e FEC durante o período analisado. A evolução apresentada ao longo dos anos para a CEMAR pode ser associada ao aumento da eficiência operacional. Observa-se também que os valores praticados são bem baixos se comparados com os valores praticados no Brasil.

Figura 4 – (a) Evolução DEC CEMAR e (b) Evolução FEC CEMAR



Fonte: Adaptação de ANEEL (2018a).

Os valores de compensações pagas em reais aos consumidores, devido às transgressões dos indicadores estabelecidos também foram levantados, porém considerando o início em 2010 até 2017, período quando começam o pagamento das compensações, conforme apresentado na Tabela 11.

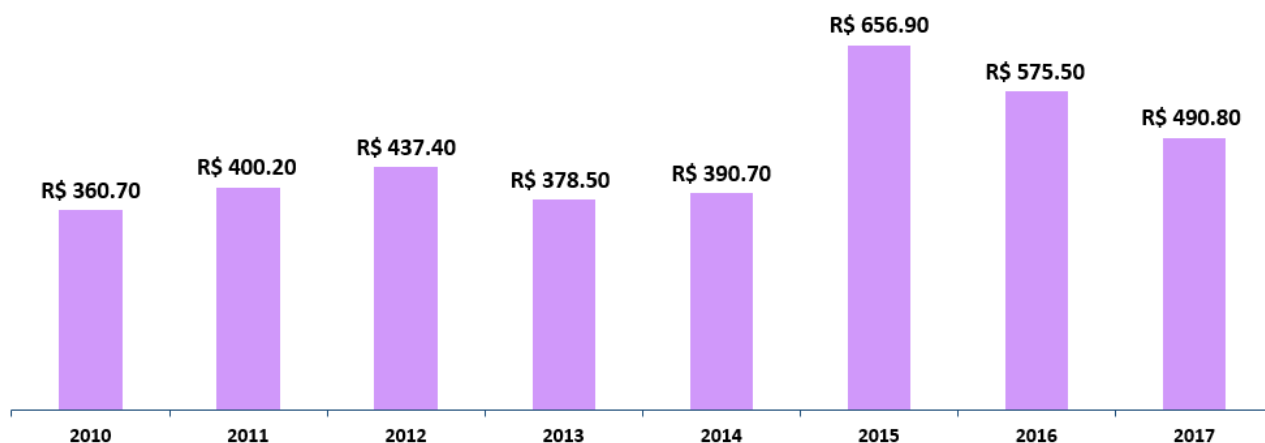
Tabela 11 – Valor das compensações (R\$) e quantidade de compensações que foram pagas, no Brasil e CEMAR, no período de 2010 a 2017.

PERIODO	BRASIL		CEMAR	
	QTD (unidade)	VALOR (R\$)	QTD (unidade)	VALOR (R\$)
2010	95.091.690	360.797.553,60	1.905.948	4.921.628,47
2011	105.137.879	400.291.487,44	3.918,364	13.664.140,73
2012	107.223.354	437.471.290,41	3.049.186	12.511.681,45
2013	104.753.826	378.553.264,02	2.435.823	7.434.708,95
2014	110.208.749	390.760.413,59	2.517.815	6.953.591,04
2015	125.201.444	656.893.484,06	2.116.251	6.578.062,15
2016	112.036.248	575.588.070,16	2.266.867	6.912.409,03
2017	100.352.468	490.870.297,60	2.288.491	5.682.568,56

Fonte: Adaptação de ANEEL (2018a).

No Brasil, o pagamento das compensações se mantém constante, com um pico em 2015, ano com o maior DEC apurado, conforme [Figura 5](#).

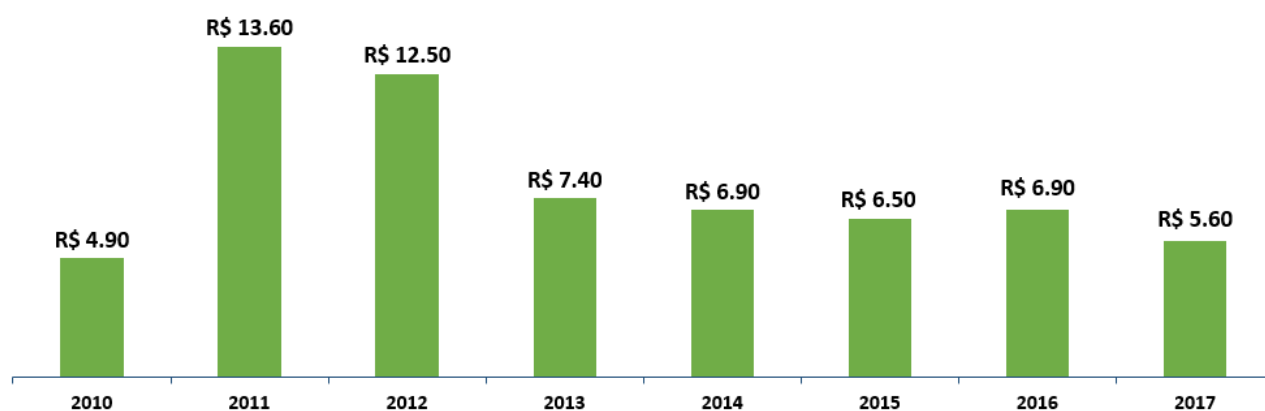
Figura 5 – Compensações pagas no BRASIL, em milhões de R\$.



Fonte: Adaptação de [ANEEL](#).

A [Figura 6](#) apresenta as compensações pagas no período de 2010 a 2017 pela CEMAR, onde se verifica uma diminuição significativa dos valores pagos entre 2012 e 2013, período em que se observa maior evolução do indicador DEC, conforme [Figura 4](#). Observa-se também que de 2014 a 2016 os valores pagos por compensações não tiveram evolução expressiva, o que pode ser entendido como uma dificuldade em melhorar os indicadores DEC e FEC.

Figura 6 – Compensações pagas pela CEMAR, em milhões de R\$.



Fonte: Adaptação de [ANEEL](#).

Cálculo da Energia Elétrica Interrompida

Os consumos de energia elétrica (MWh) pelo setor residencial foram retirados da ANEEL, que os disponibiliza anualmente, de acordo com a região e as distribuidoras de energia (ANEEL, 2018a). Para esse trabalho foram utilizados os dados de consumo de energia no período de 2010 a 2017, do Brasil e CEMAR, conforme disponibilizado pela Tabela 12.

Tabela 12 – Consumo de energia elétrica (MWh) do setor residencial do Brasil e da CEMAR no período de 2010 a 2017.

Ano	Consumo (MWh)	
	Brasil	CEMAR
2010	108.531.246,75	3.788.071,25
2011	113.226.334,34	2.041.283,58
2012	119.026.208,28	2.257.538,81
2013	126.487.397,52	2.563.489,83
2014	133.746.825,09	2.785.695,25
2015	132.846.391,21	2.916.575,43
2016	134.470.491,75	3.124.254,64
2017	136.760.142,40	3.177.519,47

Fonte: Adaptação de ANEEL (2018a).

Para o cálculo da demanda residencial (kW), dividiu-se o consumo total de energia (MWh) pelo número de horas do ano (8760 horas = 365 dias x 24 horas), já que os dados de consumo apresentados são anuais, conforme Equação 2.1.

$$Demanda = \left(\frac{\text{Consumo de Energia Elétrica}}{8760} \right) * 1000 \quad (2.1)$$

na qual:

1000 = fator de conversão de MWh para kWh

Os valores da demanda residencial estão apresentados na Tabela 13 para o período de 2010 a 2017.

Tabela 13 – Demanda do setor residencial do Brasil e das CEMAR, em kW, no período de 2010 a 2017.

Ano	Demanda kW	
	Brasil	CEMAR
2010	12.389.411,73	432.428,22
2011	12.925.380,63	233.023,24
2012	23.587.466,70	257.709,91
2013	14.439.200,63	292.634,83
2014	15.267.902,41	318.001,74
2015	15.165.113,15	332.942,40
2016	15.350.512,76	356.650,07
2017	15.611.888,40	362.730,53

Fonte: Autoria própria.

A energia interrompida (kWh) foi calculada pelo produto da demanda residencial (kW) com o indicador DEC apurado (horas) do Brasil e da CEMAR. O cálculo da energia interrompida foi feito conforme [Equação 2.2](#).

$$EnergiaInterrompida = Demanda * DEC \text{ apurado} \quad (2.2)$$

O resultado da energia interrompida para o setor residencial do Brasil e CEMAR é apresentado pela [Tabela 14](#)

Tabela 14 – Energia Interrompida para o setor residencial do Brasil e CEMAR, em kWh interrompido, para o período de 2010 a 2017.

Ano	Energia Interrompida (kWh)	
	Brasil	CEMAR
2010	228.212.964,06	9.258.288,29
2011	240.541.333,57	4.996.018,26
2012	255.172.624,60	5.576.842,45
2013	266.980.819,65	5.516.185,31
2014	275.280.280,41	5.046.029,59
2015	282.071.104,62	5.087.359,88
2016	242.691.606,69	5.071.564,04
2017	224.030.598,57	4.817.061,48

Fonte: Autoria própria.

Cálculo do custo da interrupção de energia elétrica estimado

Para a obtenção da estimativa da interrupção de energia (R\$/kWh) para o consumidor residencial foi considerado o custo unitário de cada interrupção não programada e o custo unitário de cada interrupção programada, de acordo com a região e o setor, conforme [Tabela 8](#) e [Tabela 9](#), apresentado na [subseção 2.2.3](#) e o DEC apurado do Brasil e CEMAR.

Os custos das interrupções de energia foram obtido pelo produto da energia interrompida e o custo unitário das interrupções, conforme equações abaixo:

$$CI_{np} = Energia \text{ Interrompida} * CustoUnitário \text{ Interrupção } np \quad (2.3)$$

$$CI_p = Energia \text{ Interrompida} * CustoUnitário \text{ Interrupção } p \quad (2.4)$$

na qual:

CI_{np} = Custo Interrupção não programada (R\$);

CI_p = Custo programada (R\$);

De acordo com a Nota Técnica nº 0173/2016-SRD/ANEEL, foi estabelecido que o custo unitário de cada interrupção programada é 13% do custo unitário de cada interrupção não programada. Assim, a estimativa do custo total das interrupções foi obtida a partir da soma do custo das interrupções não programadas com o custo das interrupções programadas, considerando o setor residencial e a região, de acordo com a equação abaixo (ANEEL, 2016):

$$\text{Custo Total das Interrupções} = 0,87 * CI_{np} + 0,13 * CI_p \quad (2.5)$$

na qual:

0,87 = Fator de proporcionalidade para o custo das interrupções não programadas.

0,13 = Fator de proporcionalidade para o custo das interrupções programadas.

Assim, a Tabela 15 apresenta o custo total das interrupções estimado para o setor residencial do Brasil (em Bilhões de R\$/kWh) e CEMAR (em Milhões de R\$/kWh), para os anos de 2010 a 2017.

Tabela 15 – Custo total das interrupções estimado para o setor industrial do Brasil e CEMAR, em R\$/kWh interrompido, para o período de 2010 a 2017.

Ano	Custo (R\$/kWh)	
	Brasil	CEMAR
2010	3.988.477.972,82	151.095.264,97
2011	4.203.940.886,79	81.535.018,00
2012	4.459.651.960,14	91.014.068,76
2013	4.666.023.785,04	90.024.144,22
2014	4.811.073.460,69	88.226.402,99
2015	4.929.756.695,51	83.025.713,28
2016	4.241.521.210,05	82.767.925,11
2017	3.915.382.711,14	78.614.443,35

Fonte: Autoria própria.

Esta seção teve por objetivo explorar os aspectos importantes da regulação no que tange o SEP. As recomendações são principalmente abordadas no PRODITS e dão uma ideia dos critérios que merecem ser avaliados frente da proposta deste trabalho. De acordo com o que foi apresentado, os principais critérios que devem ser levados em consideração estão relacionados à qualidade do serviço. Com base nestas informações, obtém-se um conjunto de possíveis critérios a serem estudados e levados em consideração no método de decisão proposto.

Ainda nesta conjuntura, visando contornar as adversidades apresentadas, tem ganhado grande destaque na implementação dos conceitos e técnicas de Redes Elétricas Inteligentes (REI), também conhecidas como *Smart Grids*. A REI, de forma geral, busca agregar as novas tecnologias de medição, tecnologia da informação, comunicação, automação, geração e infraestrutura às redes elétricas a fim de promover a aplicação e utilização de técnicas avançadas de monitoramento, controle, automação, operação, gerenciamento e manutenção.

Posto isto, dentro do escopo da proposta desenvolvida no presente trabalho, a seguir são expostos os aspectos referentes à REI mais direcionados aos sistemas de distribuição, bem como características e funcionalidades focadas na infraestrutura da automação da distribuição.

2.3 Redes Elétricas Inteligentes

A definição de REI apresenta-se aberta e tem diferentes conceituações, já em 2003, o *DOE (United States Department of Energy)* previa como uma rede de distribuição de energia totalmente automatizada que monitora e controla cada cliente e nó, garantindo fluxo bidirecional de informações e de energia elétrica entre usina e o aparelho, e todos os pontos entre eles ([DOE, 2003](#)).

Em 2006, a Comissão Europeia destinada ao estudo e avaliação das perspectivas tecnológicas das redes elétricas e de fontes de geração sustentáveis, conceituava REI como uma rede elétrica capaz de integrar de forma inteligente as ações de todos os usuários conectados a ela, sejam eles geradores, consumidores ou aqueles que assumem ambos os papéis, com a finalidade de prover energia elétrica de forma eficiente, sustentável, econômica e segura ([SMARTGRIDS, 2006](#)).

Posteriormente, em [Siddiqui et al. \(2008\)](#), apresenta-se uma definição segundo a qual REI pode ser entendida como sobreposição dos sistemas unificados de comunicação e controle à infraestrutura de energia elétrica existente, para fornecer as informações certas para a entidade correta (por exemplo, equipamentos de uso final e consumidores) no instante adequado para se tomar a decisão correta. É um sistema que otimiza o suprimento e a entrega de energia, minimizando perdas, é auto recuperável, e permite a próxima geração eficiência energética e resposta à demanda de aplicação.

Em ([DOLLEN et al., 2009](#)), a pedido do *National Institute of Standards and Technology (NIST)* dos Estados Unidos, é exposta uma das conceituações mais citadas na literatura, a qual aponta que REI refere-se a uma modernização do sistema de distribuição de eletricidade, para que o mesmo seja capaz de monitorar, proteger e otimizar automaticamente a operação dos seus próprios elementos interconectados, a partir da gestão central e distribuída, passando pela rede de transmissão de alta tensão e do sistema de

distribuição, até os usuários industriais e sistemas de automação predial, as instalações de armazenamento de energia e consumidores finais. Além disso, reporta-se que a REI será caracterizada por um fluxo bidirecional de eletricidade e informações para criar uma automatizada e amplamente distribuída rede de distribuição de energia, e que ela incorpora na rede os benefícios da computação e comunicação para fornecer informações em tempo real.

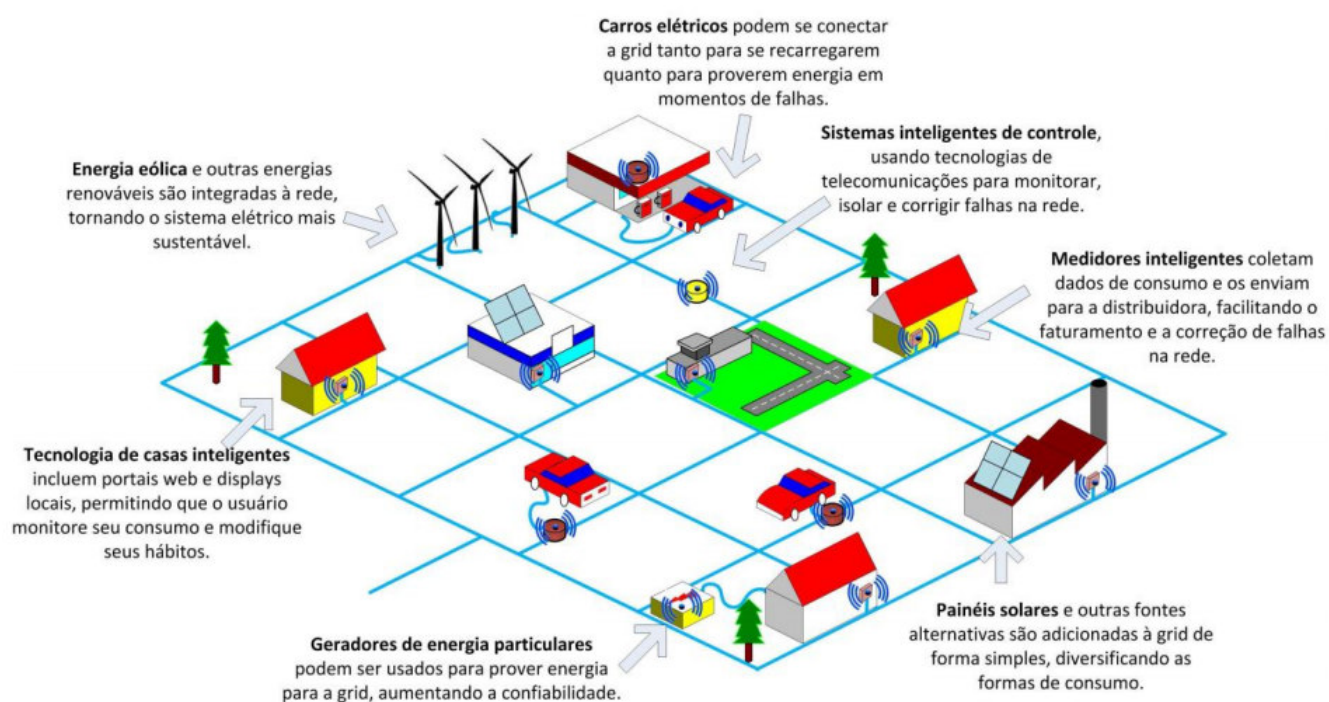
Ainda em 2009, em [Gellings \(2009\)](#) REI é caracterizada pela utilização de sensores, comunicação, capacidade computacional e controle para melhorar a funcionalidade total do sistema de fornecimento de energia elétrica. Além disso, diz-se que um sistema convencional se torna inteligente a partir do sensoriamento, comunicação e inteligência aplicada. Para o SEP, isto possibilita várias funções que permitem otimização, em combinação, do uso de geração e armazenamento massivo, transmissão, distribuição, fontes distribuídas e uso final do consumidor com objetivo de garantir confiabilidade, e otimizar ou minimizar o uso de energia, mitigar o impacto ambiental, gerenciar ativos e conter gastos.

Já em [Borlase \(2016\)](#) REI é definida como a integração de infraestruturas elétricas e de comunicações com avançados processos de automação e tecnologia da informação no âmbito da rede elétrica existente. REI representa uma mudança completa na forma como concessionárias, políticos, consumidores e outros participantes da indústria pensam a entrega de eletricidade e os serviços e ela relacionados. Este novo processo de pensamento provavelmente trará diferentes tecnologias e soluções de redes inteligentes, mas proverá benefícios com maior integração da engenharia e operações das concessionárias e a partir de novos modelos de negócios.

Atualmente, o [ENERGY \(2018\)](#) sinaliza que de forma resumida a tecnologia digital que permite a comunicação de duas vias entre a concessionária e os seus clientes, e o sensoriamento ao longo das linhas de transmissão é o que torna a rede inteligente. Ressalta-se ainda que a REI representa uma oportunidade sem precedentes para mover a indústria de energia a uma nova era de confiabilidade, disponibilidade e eficiência que contribuirá para a saúde econômica e ambiental. Durante o período de transição, será fundamental realizar testes, melhorias tecnológicas, educação do consumidor, desenvolvimento de normas e regulamentos, e compartilhamento de informações entre projetos para garantir que os benefícios previstos a partir da REI se tornem uma realidade.

Neste contexto, observa-se que a definição de REI se apresenta ampla e tem diferentes conceituações, as quais compartilham a opinião de que as mesmas trazem consigo conceitos bastante abrangentes relativos a introdução de novas tecnologias nos SEP com perspectiva de mudanças regulatórias, estruturais, técnicas e comerciais. Conceitos estes aplicáveis desde os sistemas de geração, transmissão e distribuição até os sistemas de geração distribuída e microgeração, industriais, edifícios e casas inteligentes, veículos elétricos e armazenamento de energia, como ilustrado na [Figura 7](#).

Figura 7 – Visão geral do sistema energético com o uso da REI.



Fonte: Adaptação de LOPES et al.

2.3.1 Características, Funcionalidades, Benefícios e Desafios da REI

A partir da aplicação dos conceitos de Redes Elétricas Inteligentes (REI) e do desenvolvimento das mesmas espera-se prover modernização ao SEP pela promoção de ações, tais como: gestão eficiente de ativos, gerenciamento do fluxo bidirecional de energia, fornecimento de informações em tempo real, inserção de fontes renováveis de geração, aumento do desempenho, eficiência, segurança e confiabilidade da operação, redução de perdas técnicas e comerciais, e viabilidade da participação do usuário no controle de demanda/produção de energia, dentre outros.

Para tanto, faz-se necessário às redes inteligentes possuir características e funcionalidade específicas, tais como (DOLLEN et al., 2009; GELLINGS, 2009; BORLASE, 2016; BAYINDIR et al., 2016):

- **Infraestrutura de comunicação:** dispositivos para envio e recebimento de dados e comandos operacionais entre a rede e o centro de operações;
- **Infraestrutura avançada de medição:** disponibilidade de medidores com capacidade para medição multivariáveis em tempo real e comunicação bidirecional
- **Armazenamentos de dados:** capacidade para armazenamento de uma grande

quantidade de dados trocados entre os pontos da rede e o centro de operações;

- **Automação e controle:** emprego de ferramentas ativas de apoio à operação na tomada de decisão;
- **Monitoramento:** disponibilidade de equipamentos para supervisão de demanda e fornecimento tanto por consumidores quanto pela concessionária;
- **Auto recuperação:** capacidade de automaticamente detectar, identificar, localizar, analisar e corrigir faltas na rede elétrica;
- **Inclusão dos consumidores:** fornecimento de meios para participação ativa dos consumidores no planejamento, gerenciamento e operação da própria demanda;
- **Segurança:** tolerância a ataques cibernéticos;
- **Adaptabilidade:** habilidade de gerir fluxo bidirecional de energia a partir da inserção e exclusão de fontes de geração, saída de partes do sistema devido a faltas, aumento e diminuição da demanda, entre outros;
- **Desempenho:** respostas a comandos em tempo real e de forma adequada fornecendo energia e qualidade;
- **Predição:** capacidade de analisar dados operacionais a fim de prevenir potenciais falhas de equipamentos;
- **Sustentabilidade:** adoção e inclusão prioritária de fontes renováveis e de baixo impacto ambiental;
- **Economia:** operação capitalizada por regras de oferta e demanda definidas pelos órgãos reguladores tendo em vista geração, consumo e armazenamento de energia;
- **Eficiência:** redução da perdas técnicas e aumento no rendimento da geração.

O desenvolvimento de redes elétricas com tais características e funcionalidades poderá trazer inúmeros benefícios, dentre os quais destaca-se a vasta e complexa rede de monitoramento, controle, automação e comunicação, que implica em melhoria da qualidade, confiabilidade e continuidade da energia fornecida, uma vez que com o uso de tais tecnologias as mesmas tornar-se-ão capazes de identificar antecipadamente perturbações e prover a auto recuperação e reparação do sistema isolando circuitos e/ou redirecionando fluxos energéticos de fontes próximas ao local das falhas, minimizando distúrbios e reduzindo a ocorrência de interrupções (DOLLEN et al., 2009; BORLASE, 2016; COLAK et al., 2016).

Outro benefício com o qual se espera contar é com alto grau de privacidade de usuários e segurança da rede inteligente por meio do monitoramento constante e troca de

informações em tempo real com o centro de operações, provendo segurança virtual frente a ataque aos sistemas de comunicação e informação, protegendo-se do roubo de dados do próprio sistema elétrico, pessoais e inclusive financeiros, referentes ao comércio de energia pelos usuários (DOLLEN et al., 2009; BORLASE, 2016; COLAK et al., 2016).

Também se objetiva maior sustentabilidade e eficiência energética pelo emprego distribuído de fontes renováveis de energia, otimização de recursos, aumento da produtividade da geração, armazenamento do excedente de energia e gerenciamento operacional, visando menos perdas nos processos de geração, transmissão e distribuição, bem como redução de picos de demanda e estímulo ao uso de equipamentos mais eficientes e ao consumo racional de energia (DOLLEN et al., 2009; BORLASE, 2016; COLAK et al., 2016).

Em contrapartida, para implementação da REI surge desafios e obstáculos, os quais para serem superados necessitarão do desenvolvimento de novos softwares dedicados ao comportamento inteligente da rede, projeto de sistemas de medição, controle, automação e comunicação adaptáveis e robustos, concepção de estratégias para o gerenciamento do grande volume de dados das redes inteligentes, atualização constante dos sistemas de segurança virtuais frente as investidas de ataques cibernéticos cada vez mais evoluídos, entre outros (DOLLEN et al., 2009; BORLASE, 2016; COLAK et al., 2016).

Além dos desafios tecnológicos de implementação da REI, a regulação desse novo sistema é outro ponto extremamente importante. A aproximação com o consumidor que a *Smart Grid* propõe, provocará modificações significativas nos sistemas atuais de regulação, já que o consumidor deixará de ser um elemento passivo para obter uma posição importante na nova rede que é totalmente dinâmica.

Observa-se que as principais áreas envolvidas na aplicação dos conceitos de REI são medição, monitoramento, controle, automação, segurança, geração distribuída, comercialização de energia, e sustentabilidade, em meio as quais dentro do escopo das pesquisas da presente tese algumas destas soluções serão avaliadas e associadas à outras variáveis do sistema elétrico com base em informações históricas de faltas.

3 Sistemas de Suporte a Decisão Espacial

Um Sistemas de Suporte a Decisão (*do anglicanismo, Decision Support System - DSS*) pode ser definido, como sendo um sistema computacional que auxilia a tomadores de decisão nas análises e proposição de soluções para os problemas em estudo, através de simulação de cenários possíveis. O *DSS* utiliza a combinação de modelos (físicos, abstratos, simbólicos e matemáticos), técnicas analíticas e recuperação de informações para desenvolver e avaliar problemas complexos (LIANG et al., 2005).

O desenvolvimento de interfaces gráficas, da *Internet*, e a necessidade de sistemas que levem em consideração a componente geográfica deu origem ao *GIS* (*Geographical Information System*). O *GIS* integra ferramentas que possibilitam ao gestor ampla visão do problema em estudo, permitindo a visualização, de informações espaciais em diversos formatos, através da interpretação de dados oferecidos pelo próprio sistema (ARONOFF, 1989; BRAIL; KLOSTERMAN, 2001; HUXHOLD et al., 1991a). Sua estrutura de dados é baseada em coordenadas geográficas que permitem a localização espacial dos elementos envolvidos no estudo (HUXHOLD et al., 1991b). Ele permite o mapeamento da região que envolve o problema em estudo e possibilita a codificação, o armazenamento e a análise de dados espaciais e alfanuméricos e associa atributos gráficos e não-gráficos (ARONOFF, 1989).

Segundo Brail e Klosterman (2001), os *DSS* podem ser compostos por um *GIS*. Assim, esse sistema pode incorporar, além de ferramentas tradicionais do *DSS*, também pode incorporar aqueles referentes aos *GIS*. Assim, a necessidade de se tomar decisões pelos gestores em diversos campos do conhecimento humano, com base na informação geográfica levou ao desenvolvimento dos *SDSS*.

Atualmente, os *SDSS* evoluíram para possibilitar trabalhos colaborativos em ambientes baseados na *Word Wide Web*, originando os *Web-SDSS* (*Web-based Spatial Decision Support System*) (SUGUMARAN; SUGUMARAN, 2007). O processo participativo no *Web-SDSS* pode ocorrer dos seguintes modos: entre participantes que se encontram no mesmo lugar ao mesmo tempo; entre colaboradores que apesar de estarem em uma mesma localidade podem acessar o sistema em diferentes períodos; entre colaboradores localizados em lugares diferenciados, mas conectados ao mesmo tempo; e entre colaboradores localizados em locais e horários diferentes (JANKOWSKI, 1995).

3.1 Componentes do *Web-SDSS*

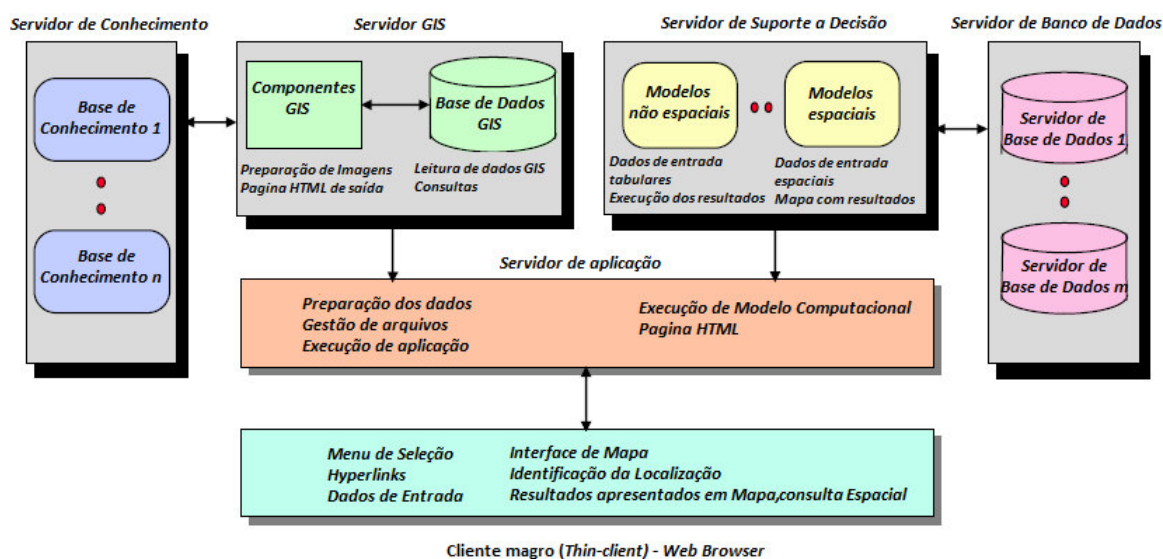
O *Web-SDSS* inclui um sistema de informações geográficas baseado na *Web* como um solucionador de problemas e facilita a recuperação, exibição e análise de dados geográficos. Nesta arquitetura combina-se vários componentes diferentes incluindo interfaces de usuário *HTML*, interface de programas *internet*, modelos computacionais e base de dados geográfica.

Existem duas maneiras de configurar o *Web-SDSS*: (1) processamento no lado do servidor; e (2) processamento no lado do cliente. Para nosso trabalho vamos utilizar a abordagem do lado do servidor que é uma abordagem computacional denominada de cliente-servidor do inglês *thin client* (cliente magro) e a maior parte do processamento, incluindo e manipulação de dados espaciais acontece do lado do servidor. As informações e os objetos de imagem resultantes são enviadas para o cliente para serem renderizadas e, ou, processadas por outras aplicações.

O *WebSDSS* do lado servidor requer somente um navegador instalado na máquina do cliente para realizar as tarefas de *SDSS*. No entanto, toda ação de usuário requer comunicação entre cliente e servidor. Os componentes típicos desta arquitetura são apresentados na [Figura 8](#) e incluem: Servidor *GIS*, Servidor de Suporte a Decisão, Servidor de Banco de Dados, Servidor de Conhecimento e Servidor *Web*. Estes componentes são brevemente descritos abaixo (SUGUMARAN; SUGUMARAN, 2007).

- **Servidor *GIS***: permite o acesso a mapas, modelos e ferramentas dentro ou fora de uma organização. Os servidores *GIS* geralmente gerenciam grandes bancos de dados *GIS*, entrega de informações geográficas e fornecem funcionalidade *GIS* abrangente, incluindo consulta e modelagem;
- **Servidor de Suporte à Decisão**: fornece acesso a um grande número de modelos necessários para analisar e resolver problemas não estruturados. Suporta modelos espaciais e não-espaciais e também facilita o desenvolvimento e teste de novos modelos. O servidor de suporte à decisão pode incorporar diferentes tipos de modelos, como modelos univariável e multivariável.
- **Servidor de Banco de Dados**: fornece acesso a dados não espaciais armazenados no sistema. A tomada de decisão requer fácil acesso a grandes volumes de diferentes tipos de dados (quantitativos, qualitativos, espaciais, temporais, etc.). O componente de interface de base de dados fornece acesso aos dados necessários para a aplicação.
- **Servidor de Conhecimento**: pode conter regras que permitem ao usuário selecionar o tipo apropriado de modelo a ser usado para uma tarefa específica e realizar uma análise de sensibilidade. Também pode conter políticas, procedimentos, regras de negócios e restrições organizacionais que podem ser relevantes para o problema em questão.

- **Servidor de Aplicação:** atua como *front-end* que conecta todos os serviços do *Web-SDSS*. Também é responsável por gerar páginas *HTML* dinâmicas retornados dos servidores.

Figura 8 – Arquitetura *Web-SDSS*

Fonte: Adaptação de [SUGUMARAN; SUGUMARAN](#).

A interface do usuário do *Web-SDSS* inclui menus, mapas gráficos, botões de controle e entrada de formulário. Esses utilitários de interface executam seleções, dados de entrada e exibições e consultas de mapas, usando *tags HTML*, *Javascript* e outros protocolos da *Internet* ([SUGUMARAN; SUGUMARAN, 2007](#)).

Os componentes descritos anteriormente podem ser implementados usando várias tecnologias para criar ambientes *Web-SDSS* sofisticados, flexíveis e fáceis de usar.

3.2 Acessibilidade da Rede de Distribuição

Algumas empresas fornecem junto a um software ou a um serviço um conjunto de rotinas e padrões estabelecidos para a utilização das suas funcionalidades por aplicativos que não pretendem envolver-se em detalhes da implementação do *software*, mas apenas usar seus serviços. A este conjunto de funções dá-se o nome de Interface de Programação de Aplicação, ou *API* na sigla em inglês. Neste trabalho, a *API* que fundamentou o propósito da pesquisa foi o *Google Maps API* que é responsável por fornecer o suporte à composição de mapas e às marcações com referência em geolocalização ([GOOGLE, 2018](#)).

A partir desta ferramenta é possível obter um histórico de velocidades médias para cada arco da rede viária. Esses dados são obtidos de forma automática pelos *smartphones* e fornecem uma fonte de dados coletiva, que agregadas em um servidor, garantem uma amostragem significativa para futuras inferências.

Através desses dados de tráfego, é possível definir, para cada par origem / destino desejado, um tempo máximo, um tempo mínimo e um tempo médio de acesso. Ou seja, uma vez que os centros de despachos são fixos (origem), é possível identificar o tempo provável de acesso para cada segmento de rede associado a um arco viário (destino). A Figura 9 apresenta um exemplo de uma rede elétrica associada ao carregamento do tráfego em um momento qualquer do dia.

Figura 9 – Rede elétrica de uma região da cidade de São Luís do Maranhão (Camada 1) com o respectivo carregamento das vias (Camada 2).



Fonte: Autoria própria.

A associação de um trecho de rede a um arco de via é parte crucial deste trabalho. Além disso, a análise da posição de cada equipamento, sua capacidade de manobra bem como a acessibilidade das equipes em caso de falha compõe o núcleo fundamental do sistema, que deve ter peso importante durante o processo de agrupamento dos blocos. Através desta metodologia, será possível compor uma modelagem heterogênea da rede de distribuição, considerando não somente suas características construtivas, mas também sua interação com as equipes de campo e o impacto financeiro de sua confiabilidade.

3.3 A Aplicação SIMAUTEK CEMAR

Esta seção descreve o Sistema Integrado de Manutenção e Automação de Equipamentos - SIMAUTEK, uma plataforma *Web* utilizada para gestão da operação, automação e manutenção de equipamentos presentes na rede de distribuição. Os tópicos serão abordados em ordem semelhante a como foram desenvolvidos, dando prioridade para cobrir os principais elementos de um *Web-SDSS*.

3.3.1 Requisitos e objetivos

Este trabalho foi desenvolvido seguindo a metodologia ágil de produção de *software*, que consiste em um conjunto de técnicas de desenvolvimento de *software* que têm o objetivo de acelerar os processos, visando uma melhoria contínua e incremental, de modo a entregar parcelas funcionais do projeto aos usuários, possibilitando uma rápida resposta quando da necessidade de alterações (ALIANCE, 2018).

Um atributo marcante desta metodologia é a subdivisão dos objetivos em metas menores com janelas de tempo, tipicamente mensais, para sua conclusão. Estas janelas de tempo recebem o nome de *scripts* e tem o objetivo de, além de facilitar o acompanhamento das metas, propiciar a entrega particionada das etapas do projeto.

Para tanto, seguiu-se uma agenda de execução de tarefas por meio de iterações periódicas. A cada iteração eram definidas funcionalidades que deveriam ser adicionadas na aplicação, focando sempre na entrega de versões utilizáveis do programa ao fim do período.

Para o desenvolvimento desta aplicação, utiliza-se tecnologias *web*, como: *HTML* (Linguagem de Marcação de Hipertexto, em português), *JavaScript* e *CSS* (Folha de Estilo em Cascata, em português) para a apresentação das informações, *front-end*. Para a comunicação com os servidores, *back-end*, utiliza-se a metodologia *MVC* (modelo, visualização e controlador, em português) na linguagem *C#* pela *framework ASP.NET*.

O SIMAUTEK é uma aplicação que além de permitir a visualização dos dados que estão espalhados em diferentes bancos de dados, possibilita o processamento e relacionamento destes dados com o intuito de gerar conhecimento, auxiliando assim, os usuários na tomada de decisões mais rápidas e assertivas. Além do uso dos dados internos da companhia, ele também possui acesso à *API* que auxilia na visualização dos dados, como por exemplo o *Google Maps*. (MICROSOFT, 2018; MACORATTI, 2018; GOOGLE, 2018)

Contudo, a fim de esclarecer e identificar os objetivos do *software*, foi realizado o levantamento dos requisitos junto às áreas de manutenção e operação.

Definiu-se, portanto, que a aplicação SIMAUTEK deve possuir os seguintes módulos:

- Módulo para cálculo da taxa de falta nos blocos: Este módulo será responsável por buscar valores do histórico de atuações em equipamentos telecontrolados (Disjuntores, Religadores e Seccionalizadoras) pelo *SCADA* (*Supervisory Control and Data Acquisition*) e ocorrências de eventos de faltas na rede de média tensão registrados pelo *OMS* (*Outage Management System*), de forma a se obter a taxa anual de faltas em cada bloco de rede (Trechos de rede delimitados por elementos de manobra). A taxa de falta anual de cada bloco será utilizada pelo módulo de aprendizado não supervisionado de agrupamento para a avaliação do custo anual de faltas no bloco.
- Módulo de simulação de tempos de deslocamento: Este módulo, através de informações provenientes do *Google Maps API*, será capaz de avaliar os possíveis tempos de acesso e procura a defeitos em cada bloco da rede de distribuição. Com base neste dado, é possível avaliar o impacto de possíveis faltas e servir como base para a locação de detectores de falta na rede de distribuição.
- Módulo de acompanhamento do desempenho e análise de investimento: Este módulo terá como objetivo avaliar o impacto das alterações realizadas na rede sobre os indicadores de continuidade/compensações financeiras relacionadas à confiabilidade da rede de distribuição. Para isto, serão utilizadas informações provenientes do módulo de aprendizado não supervisionado. Também deverá fornecer ferramentas de análises gráficas, que permitirão detectar com maior facilidade as oportunidades de investimentos e o peso das ações realizadas sobre a confiabilidade e seus custos associados.
- Módulo de gestão da automação e manutenção: Este módulo terá como objetivo avaliar as informações de disponibilidade das soluções de comunicação e efetividade operacional dos ativos de automação, os indicadores de taxa de sucesso de telecomando e disponibilidade da automação são avaliados com base nestas informações. Também vai permitir a análise de equipamentos em operação e fora de operação.

3.3.2 Levantamento de Dados do Sistema Corporativo

Esta secção apresenta os sistemas corporativos utilizados para prover informações importantes para o SIMAUTEK. As consultas aos dados foram realizadas através da Linguagem *SQL* (*Structured Query Language*). Os dados importados de cada fonte externa estão descritos a seguir:

- **GIS (Geographic Information System)**: O sistema responsável pelo gerenciamento do cadastro dos equipamentos existentes na rede de distribuição da CEMAR é o GeoRede. Nesse sistema são armazenados e atualizados os cadastros de todos componentes da rede de distribuição, seus parâmetros elétricos, a conectividade entre eles, posição geográfica e seus identificadores únicos.
- **OMS (Outage Management System)**: Utilizado atualmente pelo centro de controle da CEMAR é o OPER, através dessa ferramenta, os operadores registram toda e qualquer manobra realizada na rede de distribuição, armazenando as informações de estado de chaves sem supervisão. Essas informações são consideradas essenciais para a correta interpretação da topologia da rede, impactando na segurança e na tomada de decisão por parte do operador. Esta aplicação fornece também metas de indicadores de continuidade, registros das ocorrências emergenciais na rede de média tensão, causa das faltas, ações de reparo, tempo de reparo, tempo de preparação;
- **SCADA (Supervisory Control and Data Acquisition)**: Utilizado atualmente pelo centro de controle da CEMAR é o *Elipse Power*. Este *software* é responsável pela supervisão e controle de todos os equipamentos de subestação e aproximadamente 350 equipamentos da rede de MT (Média Tensão) de distribuição, entre chaves, seccionadoras, religadores e reguladores de tensão. Esta aplicação fornece o Histórico de atuações permanentes e temporárias do sistema de proteção (religadores, disjuntores e seccionadoras), a disponibilidade dos equipamentos, sinalizações de correntes/localização de faltas, atuações do sistema de automação, *Self-healing* e *Loop Automation*.

Por possuir muitas entidades definidas, o diagrama de relacionamento entre as bases de dados acaba sendo bastante complexo, contudo facilita a compreensão o funcionamento do *software*. O [Apêndice A](#), detalha o diagrama de relacionamento entre cada entidade e os demais componentes do sistema.

Os resultados obtidos no desenvolvimento da aplicação *web* SIMAUTEK são apresentados no [Capítulo 5](#) que trata da Metodologia juntamente com as etapas da Extração de Conhecimento, que serão descritas no próximo Capítulo.

4 Extração de Conhecimento

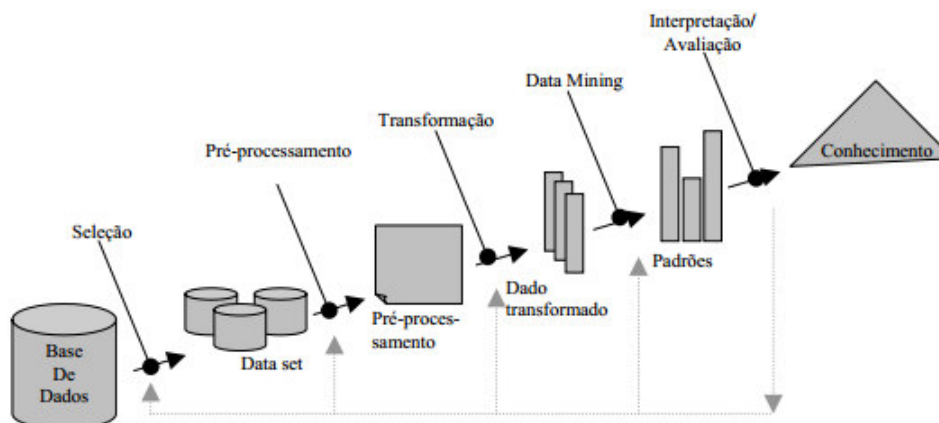
A utilização de informações armazenadas em bases de dados para compor padrões úteis de informação tem sido um constante desafio nos dias atuais. Técnicas como mineração de dados, extração de conhecimento, descoberta de informação, processamento de padrões de dados são alguns termos utilizados para este tipo de tarefa. Uma forma de fazer uso adequado de dados existentes, guardando as características necessárias para o fim que se propõe, é a utilização das técnicas de Extração de Conhecimento em Banco de Dados, comumente referenciado pela sigla inglesa *KDD* (FAYYAD et al., 1996).

Neste Capítulo discorre-se sobre as particularidades e etapas operacionais de *KDD*, com ênfase na mineração de dados e seus métodos de grande potencial à aplicação no problema de identificação de padrões. Dedicar-se uma seção especial às técnicas de clusterização ou agrupamento de dados.

4.1 Conceitos e Etapas Operacionais de *KDD*

O processo de *KDD* é composto por um conjunto de etapas com a finalidade de obter novos conhecimentos a respeito de um determinado domínio, a partir de uma base de dados em estado bruto (FAYYAD; UTHURUSAMY, 2002). O termo processo utilizado no *KDD* implica que este compreende diversos passos ou etapas, as quais envolvem a preparação dos dados, a busca por padrões, a avaliação do resultado e as etapas adicionais intermediárias que têm o objetivo de preparar a transição das etapas (FAYYAD et al., 1996).

O *KDD* é um processo que envolve a utilização de um banco de dados e de procedimentos adicionais que selecionam, pré-processam, realizam amostragem ou transformam os dados. O processo ainda conta com uma etapa denominada *data mining*, que é a aplicação de métodos (algoritmos) que extraem padrões a partir dos dados. Por fim, a avaliação dos produtos obtidos da *data mining* é realizada a fim de identificar se um subconjunto de padrões obtidos podem ser julgados como novos conhecimentos. As principais etapas do processo são apresentadas, na forma de uma visão geral, Figura 10.

Figura 10 – Etapas do Processo de *KDD*

Fonte: adaptação de [FAYYAD et al.](#)

Geralmente uma etapa está relacionada diretamente com sua etapa antecedente, de forma que o processo é considerado um processo iterativo. Esta relação possibilita melhores resultados em cada etapa, uma vez que, trabalhar com o resultado da etapa anterior é mais produtivo e gera mais qualidade nos resultados das etapas subsequentes. Pode-se dizer que o processo de *KDD* compreende, em outras palavras, todo o ciclo que o dado percorre até que este se transforme em um conhecimento ([CEREDA; JOSE, 2014](#)).

Embora as etapas do processo *KDD* devam ser executadas na ordem apresentada na [Figura 10](#), o processo é iterativo e interativo ([CIOS et al., 2007](#)). Iterativo, por ser uma sequência finita de operações em que o resultado de cada uma depende do resultado das que a precedem. Interativo, pois o usuário pode, e às vezes necessita, intervir e controlar o curso das atividades. Destaca-se, na mesma figura, a possibilidade da execução de “laços” entre quaisquer duas etapas, dados pelas setas tracejadas apresentadas abaixo de cada transição entre etapas ([FAYYAD et al., 1996](#)). Isso significa que o usuário pode, quando necessário e em qualquer transição, retornar a qualquer etapa anterior para execução com quaisquer alterações desejadas.

As etapas principais de seleção de dados, pré-processamento, transformação, *data mining* e interpretação e avaliação dos dados do processo de *KDD* são descritas a seguir:

1. **Seleção dos dados:** nesta etapa é criado um subconjunto de dados selecionados a partir da(s) base(s) de dados disponíveis. Este subconjunto conterá apenas aqueles dados relevantes para a solução do problema. O sucesso do processo depende da escolha correta dos dados que formam o conjunto de dados alvo, pois é neste subconjunto que, mais adiante no processo, serão aplicados os algoritmos para descoberta de conhecimento.

2. **Pré-processamento:** Após etapa de seleção, inicia a limpeza e pré-processamento dos dados. Os dados quando armazenados num banco de dados, muitas vezes aparecem com alguns problemas aparentes, tais como: informações incompletas, dados redundantes, ruído e incerteza. Nesta etapa, devem ser estudadas e aplicadas as estratégias para tratamento desses dados.
3. **Transformação dos dados:** A execução da etapa de pré-processamento produz um conjunto de dados muitas vezes complexos para a execução posterior das etapas do *KDD*. Assim, como procedimento subsequente, executa-se a etapa de transformação dos dados, cujo objetivo é consolidar o conjunto de dados para possibilitar o trabalho com informações relevantes. Além disso, esta etapa permite otimizar o tempo de processamento do algoritmo de *data mining*, uma vez que reduz o número e a complexidade dos atributos disponíveis para análise.
4. **Data mining:** Etapa caracterizada pela busca de padrões nos dados. Nesta etapa, é escolhido o método e são definidos os algoritmos que realizarão a busca pelo conhecimento implícito e útil do banco de dados. É a fase mais importante do processo de *KDD* onde dados são transformados em informação. Por isso, é importante que seja realizada quando os dados estiverem corretos e a tarefa seja adequada para alcançar o objetivo.
5. **Interpretação/Avaliação:** Onde é realizada a interpretação dos resultados obtidos após a aplicação do algoritmo minerador. A principal meta dessa fase é melhorar a compreensão do conhecimento obtido, em forma de relatórios demonstrativos, com a documentação e explicação das informações relevantes descobertas no processo de *KDD*. Os resultados do processo de descoberta do conhecimento podem ser mostrados de forma que possibilite uma análise criteriosa para identificar a necessidade de retornar a qualquer uma das etapas anteriores do processo de *KDD*, caso os resultados não sejam satisfatórios.

Devido à importância deste assunto no desenvolvimento do trabalho, a seguir são descritas com mais detalhes as tarefas de *data mining*, com ênfase para a tarefa de agrupamento sendo que esta é utilizada neste trabalho durante a aplicação do estudo de caso, pelas soluções oferecidas na tarefa e à adequação aos problemas deste trabalho.

4.2 Mineração de Dados

Grande parte das teorias, métodos e algoritmos utilizados no processo de mineração de dados para a extração de conhecimento em grande base de dados, são provenientes das áreas de aprendizado de máquina e reconhecimento de padrões.

Do reconhecimento de padrões, por exemplo, provêm duas abordagens principais denominadas *aprendizado supervisionado* e *aprendizado não-supervisionado* (COSTA; MONTAGNOLI, 2001):

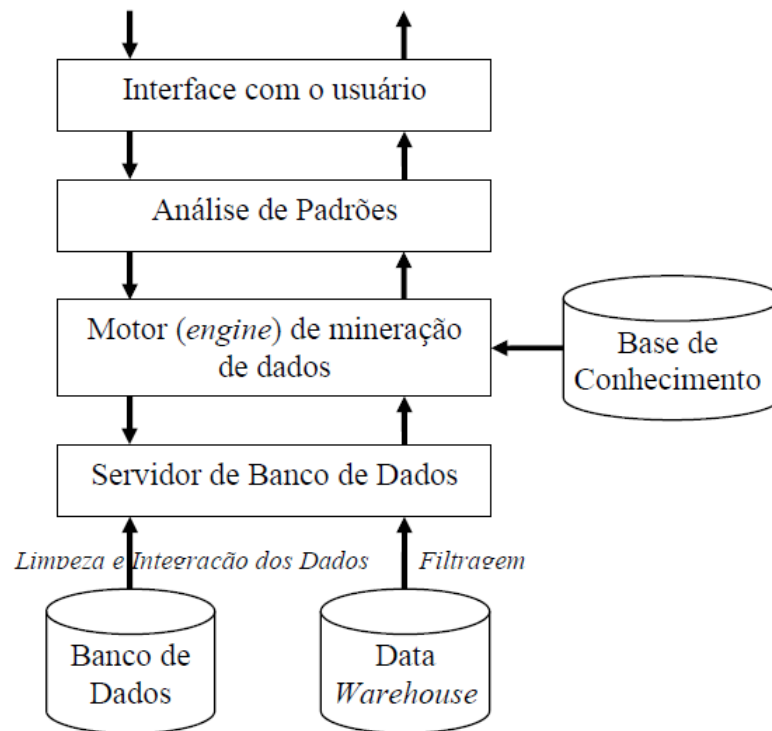
- *aprendizado supervisionado* é aquele que requer algum tipo de supervisão do processo que informe sobre exemplos típicos de cada classe, os quais, dentro de um processo de treinamento, permitem identificar a classe de novos objetos. Um exemplo é ensinar alguém que nunca viu maçãs ou laranjas a reconhecê-las, mostrando-lhes várias frutas e identificando cada uma; depois disso, esse alguém provavelmente será capaz de classificar novas amostras.
- *aprendizado não-supervisionado* não possui exemplos e o número típico de classes é desconhecido. Nesse caso, é preciso identificar como os objetos podem ser agrupados em classes, sempre com base em atributos dos mesmos. O sistema de reconhecimento deverá deduzir o número de classes e quais objetos pertencem a quais classes. No exemplo acima, as frutas seriam mostradas sem identificação, devendo a pessoa descobrir os dois tipos de frutas (classes).

O processo não-supervisionado é mais complexo que o supervisionado, já que nestes sistemas é necessário definir o número de classes e os atributos típicos para formação dos padrões. Os métodos desse processo, também chamado *clustering*, tendem a apresentar bons resultados quando as classes (nuvem de pontos) estão bem separadas no espaço de atributos, e apresentam dificuldades se elas estiverem sobrepostas (quando os objetos são semelhantes). A aplicação de reconhecimento de padrões de padrões juntamente com bases de dados constitui a mineração de dados (COSTA; MONTAGNOLI, 2001).

Uma definição que adota uma visão ampla sobre a funcionalidade de mineração de dados é a seguinte (HAN et al., 2011):

"Data Mining é o processo de descoberta de conhecimento interessante de grande quantidade de dados armazenados tanto em banco de dados, como em data warehouses ou outros repositórios de informações".

Com base nessa ideia, a arquitetura de um sistema típico de mineração de dados com seus componentes principais é apresentada na Figura 11 (HAN et al., 2011):

Figura 11 – Arquitetura de um Sistema Típico de *Data Mining*

Fonte: adaptação de HAN et al.

1. **Banco de dados, *data warehouse***: pode ser somente um ou um conjunto de banco de dados, planilhas, ou outro tipo de banco de dados informações, sobre os quais podem ser realizadas técnicas de limpeza e integração de dados.
2. **Servidor de Banco de Dados ou servidor de *data warehouse***: o servidor de banco de dados é responsável por localizar e carregar os dados relevantes, baseado nas consultas de usuário.
3. **Base de conhecimento**: é o domínio de conhecimento utilizado para guiar a busca, ou avaliar o interesse em algum padrão resultante. Esses dados são a entrada do motor de inferência e dessa base será extraído o conhecimento de interesse para o sistema.
4. **Motor (*engine*) de mineração de dados**: é o responsável pela inferência com base nas informações obtidas na base de conhecimento. Busca padrões de interesse através de técnicas de mineração de dados tais como regras de associação, regras de classificação, padrões de sequências, agrupamento *clustering*, séries temporais, regras de produção, entre outras.
5. **Avaliação de padrões**: interage com o motor de mineração de dados com o objetivo de avaliar se os padrões ou regras geradas são interessantes à aplicação.

6. **Interface com o usuário:** é o módulo de comunicação entre o sistema de mineração de dados, permitindo que o usuário interaja com o sistema através de consultas e visualização da informação extraída.

As técnicas de mineração de dados devem ser vistas sob o foco da escalabilidade, de forma a satisfazer os requisitos para o relacionamento com grandes conjuntos de dados. Para essa finalidade, novos métodos foram desenvolvidos e métodos estatísticos e matemáticos antigos e bem conhecidos, bem como métodos de redes neurais foram atualizados e retrabalhados para aplicação de mineração de dados (GRABMEIER; RUDOLPH, 2002).

A utilização de mineração de dados permite a aquisição de conhecimento interessante ou informação de alto nível, que pode ser explorada de ângulos diferentes, e aplicada em tomadas de decisões, controle de processos, gerenciamento de informação e processamento de consultas. Assim, é considerada como uma das fronteiras mais importantes em sistemas de bancos de dados e um dos mais promissores desenvolvimentos interdisciplinares na indústria da informação (HAN et al., 2011).

Os objetivos da mineração de dados podem ser alcançados através de diferentes métodos, também denominados de tarefas. As tarefas são variadas e distintas devido à variedade de padrões existentes em grandes volumes de dados e, para cada tipo de padrão são necessários diferentes métodos e abordagens. As diferentes tarefas de mineração de dados podem ser definidas como sumarização, classificação, agrupamento, associação e regressão (FAYYAD et al., 1996; GHEWARE et al., 2014).

A tarefa de sumarização realiza a abstração ou generalização dos dados, cuja aplicação resulta em um conjunto de dados menor, que fornece uma visão geral desses dados e que possuem, normalmente, algum conhecimento agregado. A classificação é a aprendizagem de regras ou funções que determinam a classe de um objeto de acordo com seus atributos.

A tarefa de agrupamento realiza a identificação de grupos similares, ou *clusters*, a partir dos dados. O objetivo do agrupamento é agrupar conjuntos de objetos de forma que os objetos pertencentes ao mesmo grupo sejam mais semelhantes entre si, em relação àqueles em outros grupos.

A tarefa de associação envolve a descoberta de uniões ou conexões entre objetos, denominadas regras de associação, que revelam relacionamentos entre objetos no conjunto de dados.

A tarefa de regressão é uma metodologia estatística que visa determinar funções com erros mínimos que modelem os dados. E, por fim, a regressão pode fornecer informações sobre o relacionamento e as interdependências entre os objetos do conjunto de dados (GHEWARE et al., 2014; CHEN et al., 1996).

4.3 Análise de Agrupamentos

O agrupamento ou clusterização é a identificação de classes, também denominadas de grupos ou *clusters*, a partir um conjunto de dados cujos objetos possuem classes desconhecidas. Em outras palavras, a tarefa de agrupamento realiza a identificação de grupos similares a partir de dados não estruturados. A tarefa se caracteriza por agrupar um conjunto de dados de entrada em grupos, de tal forma que os dados incluídos em um mesmo grupo são os mais similares entre si e menos similares em relação aos dados que estão em outros grupos (GIUDICI, 2005).

Uma vez que os grupos tenham sido definidos a partir do conjunto de dados, os objetos são rotulados de acordo com seu grupo correspondente. A rotulação dos dados em seus *clusters* permite distinguir características comuns de forma sumarizada para compor a descrição de cada *cluster*.

Um bom agrupamento produz segmentos de alta qualidade, em que a similaridade intraclasse é alta e a interclasse é baixa. A qualidade desse resultado também depende da métrica de similaridade usada pelo método e de sua implementação, além de sua habilidade em descobrir algum ou todos os padrões escondidos (LOPES, 1999).

A tarefa de agrupamento pode ser realizada utilizando diferentes técnicas. Tradicionalmente, as técnicas de agrupamento são divididas em técnicas hierárquicas e técnicas de particionamento. Os algoritmos que utilizam as técnicas hierárquicas constroem *clusters* gradativamente, formando uma árvore de *clusters*.

Para as técnicas hierárquicas são categorizados algoritmos aglomerativos e divisivos, de acordo com a forma em que os *clusters* são construídos. As técnicas de particionamento, por outro lado, geram os conjuntos diretamente, tentando descobrir *clusters* com a realocação de pontos entre os subconjuntos, ou ainda identificando *clusters* em uma área com alta densidade de dados.

As técnicas de particionamento podem ser categorizadas em métodos probabilísticos, métodos *k-medoids*, métodos *k-means* e métodos baseados em densidade (BERKHIN, 2006). Na Tabela 16 apresenta-se a listagem de algoritmos de agrupamento de acordo com suas categorias e subcategorias.

Tabela 16 – Algoritmos de agrupamento conhecidos, divididos em técnicas e categorias, de acordo com a metodologia de construção dos *clusters*.

Técnica	Categoria	Algoritmo	
Hierárquica	Aglomerativo	Cure Chameleon	
	Divisivo	SLINK COBWEB	
Particionamento	Probabilístico	EM (Expectation Maximization) SNOB AutoClass MCLUST	
		K-medoids	PAM CLARA CLARANS
			K-means
	Baseado em densidade	DBSCAN OPTICS DBCLASD DENCLUE	

Fonte: [Berkhin \(2006\)](#).

Diante de uma tarefa de mineração de dados, a definição de algoritmos de agrupamento para um determinado problema é uma tarefa complexa. Suas abordagens são provenientes de uma variedade de disciplinas e, em função disso, o uso de diferentes terminologias e definições básicas, dificultam a escolha de uma abordagem relevante para um determinado problema ([FASULO, 1999](#)).

Para a escolha de um algoritmo, no entanto, é fundamental observar se este atende a diversos requisitos especiais próprios de *clusterização*, desejáveis para a formação de bons agrupamentos, alguns dos quais serão descritos a seguir ([HAN et al., 2011](#); [FASULO, 1999](#)):

- *Escalabilidade e usabilidade*: as técnicas devem apresentar rapidez e escalabilidade para agrupar dados, de forma a suportar todo o conjunto de dados. Além disso, devem ser insensíveis à ordem de entrada dos registros de dados apresentados ao sistema. Devem, ainda, ser capazes de detectar agrupamentos de formas arbitrárias, visto que a maioria desses algoritmos determina agrupamentos baseados em medidas de distância euclidiana, que tendem a encontrar agrupamentos esféricos com tamanhos e densidades similares.

- *Habilidade de trabalhar com diferentes tipos de atributos*: os algoritmos devem ser capazes de agrupar diversos tipos de dados, tais como aqueles baseados em intervalos, numéricos, binários, categóricos, nominais, ordinais ou ainda, combinar vários tipos de dados.
- *Tratamento efetivo e alta disponibilidade*: os bancos de dados, normalmente, apresentam várias dimensões ou atributos, enquanto que muitos algoritmos de agrupamentos conseguem apenas tratar dados envolvendo duas ou três dimensões. Portanto, é um desafio agrupar objetos de dados em um espaço de alta dimensionalidade, especialmente considerando que tais dados podem estar bem esparsos e altamente distorcidos. Além disso, funções de distância aplicadas sobre dados com ruídos, dados excedentes, ausentes, desconhecidos ou errôneos, serão ineficientes, e esses algoritmos devem ser capazes de tratar esses tipos de problemas.
- *Interoperabilidade dos resultados*: os usuários esperam resultados de agrupamento que sejam interpretáveis, compreensíveis e usáveis. Devem estar ligados a uma semântica específica para interpretações e aplicações. É importante estudar com o objetivo de uma aplicação pode influenciar a seleção de métodos de agrupamentos.

4.3.1 Tipos Básicos de Dados em Clusterização

Os principais algoritmos de clusterização operam sobre as seguintes estruturas de dados:

- *Matriz de Dados Linear (ou estrutura objeto-por-variável)*: representa n objetos com p variáveis (ou atributos) na forma de uma tabela relacional ou uma matriz e dimensão $n \times p$ (n por p variáveis). O vetor de n objetos (ou *tuplas*) é representado da seguinte forma (HAN et al., 2011):

$$\begin{bmatrix} X_{11} & \cdots & X_{1f} & \cdots & X_{1p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ X_{il} & \cdots & X_{if} & \cdots & X_{ip} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ X_{nl} & \cdots & X_{nf} & \cdots & X_{np} \end{bmatrix} \quad (4.1)$$

- *Matriz de Dissimilaridade (ou estrutura objeto-por-objeto)*: armazena medidas de proximidade relacionada a qualquer par dentre n objetos de um conjunto de dados, na forma de uma matriz de dimensão $n \times n$, como mostrado a seguir:

$$\begin{bmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ \vdots & \vdots & \vdots & 0 & \\ d(n,1) & d(n,2) & \cdots & \cdots & 0 \end{bmatrix}, \quad (4.2)$$

Em que $d(x, y)$ é a diferença ou dissimilaridade entre os objetos x e y . Em geral, $d(x, y)$ é um número não-negativo que está próximo de 0 quanto objetos x e y são altamente similares ou "próximos" uns dos outros, e se tornam maiores à medida que diferem. Desde que $d(x, y) = d(y, x)$ e $d(x, x) = 0$, tem-se a matriz (4.2).

As linhas e colunas da matriz de dados representam entidades diferentes, enquanto que as linhas e colunas da matriz de dissimilaridade representa a mesma entidade.

4.3.2 Medidas de Distância

A qualidade de um agrupamento é dada em função da métrica de similaridade dos objetos desse agrupamento. A definição de similaridade ou dissimilaridade entre objetos depende do tipo de dado considerado e que tipo de similaridade se está procurando e é geralmente expressa em termos de uma função de distância $d(x, y)$, cuja métrica deve satisfazer às seguintes condições (HAN et al., 2011):

$$\begin{aligned} d(x, y) &\geq 0, \forall x \text{ e } y \\ d(x, x) &\geq 0, \forall x \\ d(x, y) &= d(y, x) \end{aligned} \quad (4.3)$$

Já a métrica distância é um conceito mais restritivo, requerendo a satisfação do teorema da desigualdade triangular; ou seja, para qualquer padrão \mathbf{X} , \mathbf{Y} e \mathbf{Z} , tem-se:

$$d(x, y) \leq d(x, z) + d(z, y) \quad (4.4)$$

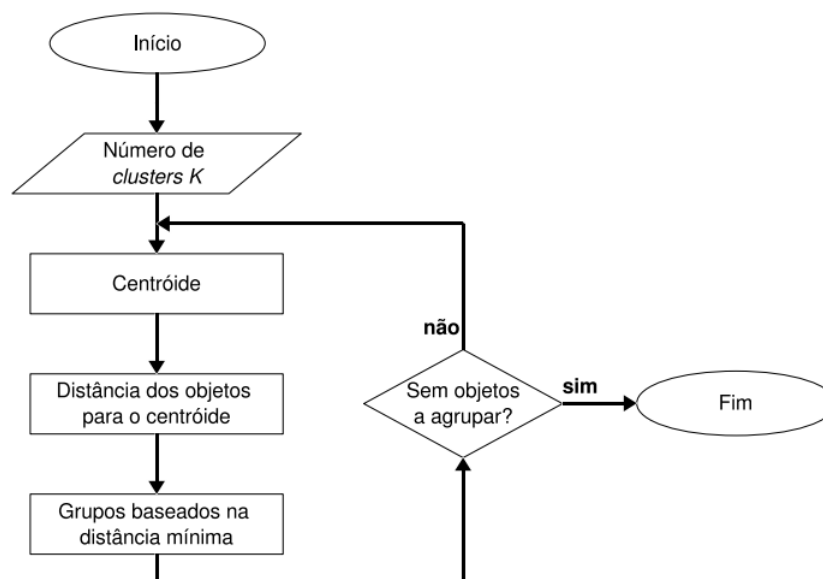
No caso de atributos contínuos, uma variedade de funções pode ser usada (vide Tabela 17), cada qual apropriada a um determinado tipo de problema.

Tabela 17 – Funções de distância entre dois padrões x e y .

Função de Distância	Expressão
Euclidiana	$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
Hamming (city block)	$d(x, y) = \sum_{i=1}^n x_i - y_i $
Tchebyshev	$d(x, y) = \max_{i=1,2,\dots,n} x_i - y_i $
Minkowski	$d(x, y) = \sqrt[p]{\sum_{i=1}^n (x_i - y_i)^p}$, $p > 0$
Canberra	$d(x, y) = \sum_{i=1}^n \frac{ x_i - y_i }{x_i + y_i}$, x_i e y_i positivos
Separação Angular	$d(x, y) = \frac{\sum_{i=1}^n x_i y_i}{[\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2]^{1/2}}$

4.3.3 Algoritmo *K-means*

O algoritmo *K-Means* (MACQUEEN et al., 1967) é um dos mais simples algoritmos de aprendizado não-supervisionado que resolve o problema de agrupamento de dados. O procedimento segue uma maneira relativamente simples e fácil de classificar um determinado conjunto de dados através de um certo número de *clusters* (assumindo *k-clusters*), fixado a priori. O fluxograma do algoritmo é mostrado na Figura 12

Figura 12 – Arquitetura de um sistema típico de *Data Mining*.

Fonte: Adaptação de HAN et al..

O algoritmo funciona basicamente enquanto houver elementos na base sem pertencer a nenhum agrupamento, faz-se a comparação entre elementos (nova informação comparada aos centroides dos *clusters* existentes) para incluir os novos.

O algoritmo segue basicamente os seguintes passos (HAN et al., 2011):

Passo 1: Particionar os objetos, de forma aleatória, em k grupos não vazios, sendo que cada partição será o centroide desses grupos;

Passo 2: Assinalar cada objeto restante como pertencente ao *cluster* ao qual apresenta-se mais próximo do seu centroide de acordo com a função distância;

Passo 3: Atualizar os centroides de cada *cluster*. As novas posições dos centroides será a média dos elementos que pertencem ao *cluster*, conforme Equação (4.5):

$$c_j = \frac{\sum_{i=1}^{n^{(j)}} x_i^{(j)}}{n^{(j)}} \quad (4.5)$$

na qual: $x_i^{(j)}$ é um elemento associado ao *cluster* j e $n^{(j)}$ é o número de elementos associados ao *cluster* j ;

Passo 4: Se elementos mudarem de *cluster* no *Passo 2*, repetir 2-3. Se não, o algoritmo para.

Se o número de dados é menor ou igual ao número de *clusters* então atribuir cada dado como o centroide de um *cluster*. Cada centroide terá associado um dos k *clusters*. Se o número de dados é maior do que o número de *cluster*, para cada um dos dados, calcular a distância a todos os centroides até obter a distância mínima. Cada dado pertencerá ao *cluster* para o qual possui a menor distância ao seu centroide.

Uma vez que não tenha certeza sobre a localização do centroide, precisa-se ajustar a localização do centroide com base nos dados atuais. Então, atribui-se todos os dados para esse novo centroide. Este processo é repetido até que não haja mais dados que se deslocam para outros *clusters*. A convergência irá sempre ocorrer quando a soma das distâncias de cada amostra de treinamento para o centroide for reduzida.

Finalmente este algoritmo visa minimizar uma função objetivo, neste caso uma função de erro quadrático. A função objetivo é mostrada na Equação (4.6):

$$J = \sum_{i=1}^k \sum_{j=1}^n d(X_i, C_j)^2, \quad (4.6)$$

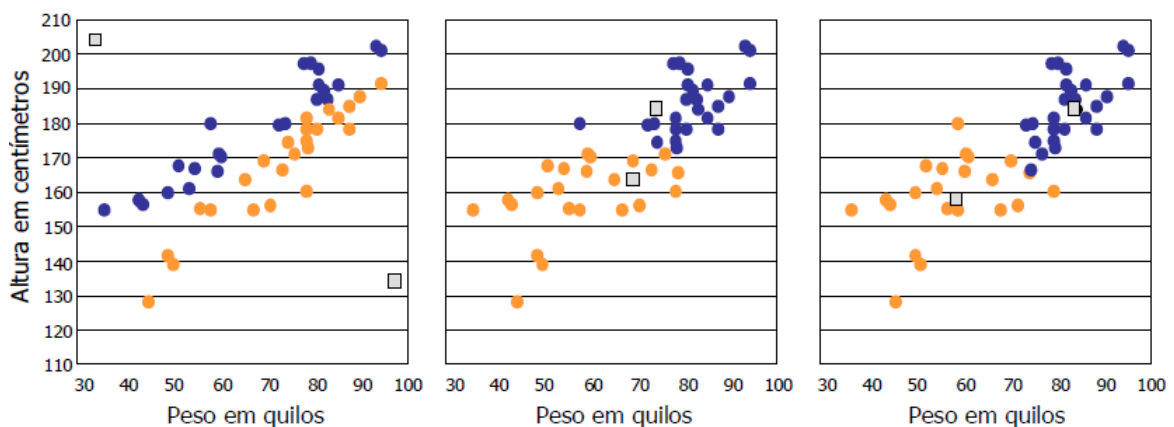
na qual:

- n é o número de instâncias

- k é o número de *clusters* considerando no algoritmo, o qual deve ser decidido antes da execução;
- X_i é o i -ésimo dado;
- C_j é o j -ésimo *cluster*; e
- $d(X_i, C_j)$ é a distância entre X_i e C_j , usualmente a distância é a Euclidiana (d_E).

Um exemplo deste algoritmo pode ser visualizado na [Figura 13](#), na qual se deseja agrupar um conjunto de objetos em 2 grupos em função dos atributos peso e altura.

Figura 13 – Agrupamento de um conjunto de objetos baseado no método *k-médias*.



Fonte: Adaptação de [COSTA; MONTAGNOLI](#).

O *k-means* foi escolhido por ser um algoritmo simples, escalonável e suportar uma grande quantidade de dados.

O problema da definição de grupos

Existem vários critérios para a determinação do número de grupos e quase todos funcionam da seguinte maneira: realizar o agrupamento dos dados considerando 2 grupos e calcular o valor de uma função proposta que tenha o número de grupos como um de seus parâmetros, realizar o agrupamento dos dados considerando 3 grupos e calcular o valor da mesma função, repetir este procedimento até atingir um número máximo de grupos estabelecido. O agrupamento que ocasionar o valor máximo (ou, em alguns casos, mínimo) da função, deve ser considerado como o melhor agrupamento possível para a base de dados ([KODINARIYA; MAKWANA, 2013](#)).

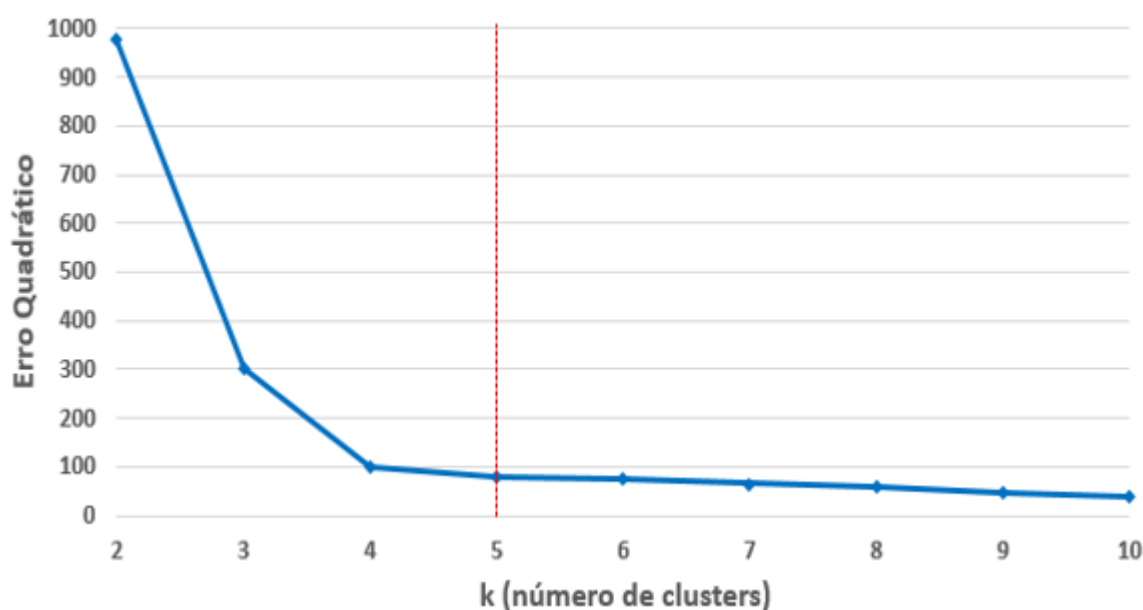
Como será explicado mais adiante, foi medida a soma de erros quadráticos quando para a variação do valor de k . Pelo método de *elbow*, também chamado de regra do cotovelo, é possível avaliar qual valor aproximado de k a ser dado de entrada ao algoritmo *k-means*, utilizando o conjunto de dados deste trabalho (KODINARIYA; MAKWANA, 2013).

A ideia da regra do cotovelo é executar o *k-means* no conjunto de dados n vezes numa faixa de valores, e para cada valor de n calcular a soma dos erros quadrados. Ao final da execução é gerado um gráfico mostrando erro versus o número de clusters, ou seja, a soma dos erros quadráticos de acordo com a quantidade de clusters gerada (KODINARIYA; MAKWANA, 2013).

Se a linha do gráfico é semelhante ao desenho de um braço, o ponto mais baixo representa onde o erro começa a se estabilizar, ou seja, a diferença de erros entre determinado valor k de *clusters* e o seu sucessor $k + 1$ é abaixo de um *threshold*, sendo este o valor ideal (KODINARIYA; MAKWANA, 2013).

Um exemplo do método de *elbow* para definição do valor de k é apresentado na Figura 14, onde o valor ideal de k é 5.

Figura 14 – Identificação do Ponto de *Elbow*



Fonte: adaptação de: KODINARIYA; MAKWANA

4.4 Ferramenta *Weka*

Nos últimos anos, o grande aumento do uso de processos de *data mining* pelas empresas motivou o surgimento de muitos softwares proprietários e livres para tal finalidade, por exemplo, o *RapidMiner*, o *Apache Spark*, o *Weka*, entre outros. Para o presente trabalho serão utilizadas as ferramentas de aprendizado não supervisionado do *Weka*.

Segundo [Witten et al. \(2011\)](#), o *Weka* é um software muito popular de aprendizagem de máquina (*machine learning*) e *data mining*. É totalmente livre, do tipo *open source*. O sistema foi desenvolvido em *Java* e pode ser executado em praticamente qualquer plataforma, tendo sido testado em sistemas operacionais *Linux*, *Windows* e *Macintosh*.

O *Weka* se consolidou como o sistema de data mining mais utilizado em ambientes acadêmicos ([WITTEN et al., 2011](#)). Este sistema é utilizado não apenas em pesquisas científicas, mas principalmente para fins didáticos. Foi exatamente a sua adequação para este último tipo de aplicação que o tornou popular.

O sistema *Weka* foi criado em uma universidade da Nova Zelândia (*The University of Waikato*). *Weka* significa *Waikato environment for knowledge analysis*. O sistema provê uma interface uniforme para diferentes algoritmos de aprendizagem, além de métodos para pré e pós-processamento e avaliação do resultado da aprendizagem em um certo conjunto de dados ([HOLMES et al., 1994](#)).

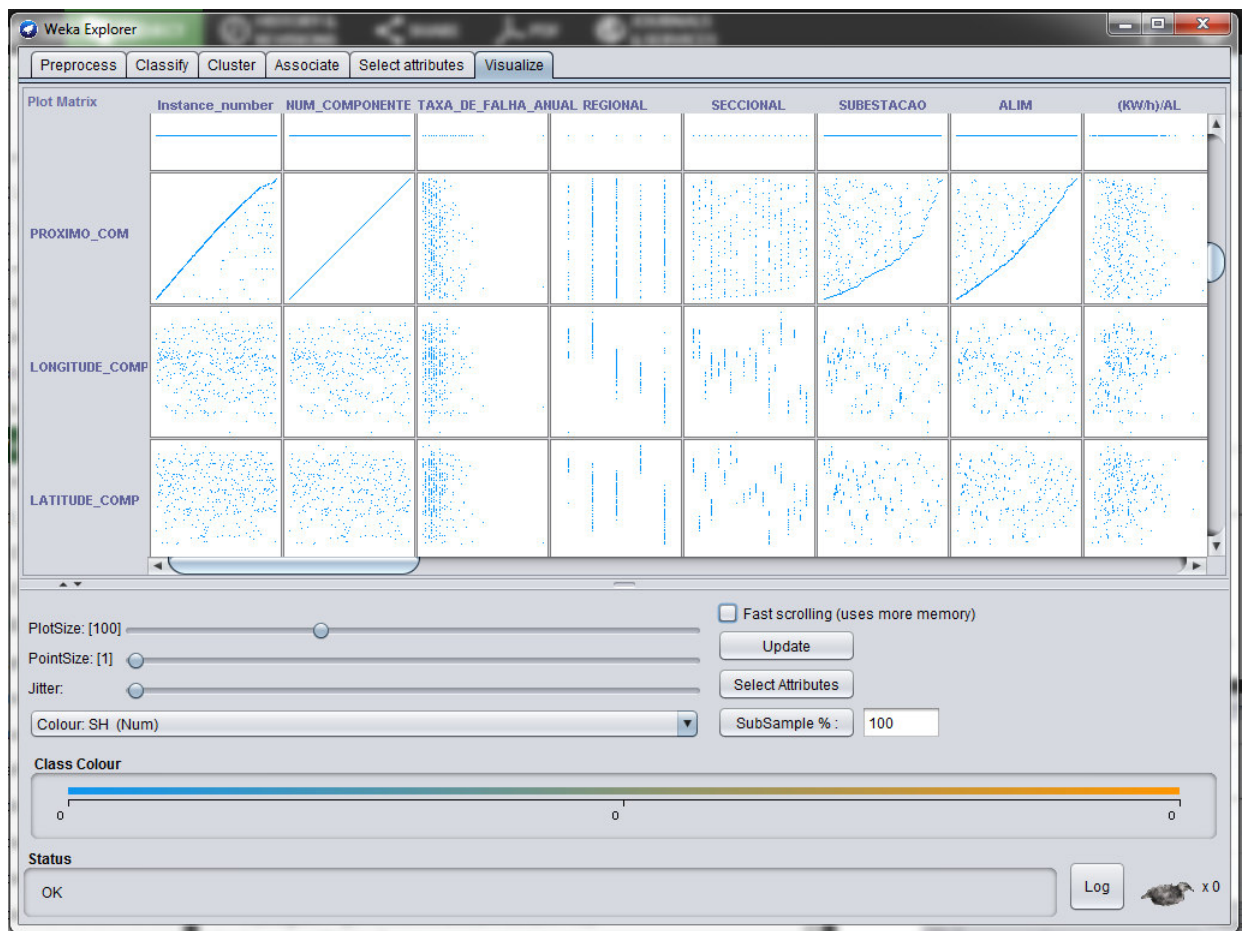
O sistema oferece implementações de algoritmos de aprendizagem que podem ser facilmente aplicados em conjunto de dados, além de incluir uma variedade de ferramentas para transformação de dados, tais como seleção de atributo e limpeza de dados. É possível pré-processar um conjunto de dados, alimentar um algoritmo de aprendizagem e analisar o resultado e seu desempenho, sem que seja necessário escrever sequer uma linha de código ([HOLMES et al., 1994](#)).

O sistema *Weka* inclui métodos para os principais problemas de *data mining*: Regressão, Classificação, Agrupamento e Associação. O sistema disponibiliza também muitas ferramentas para visualização e pré-processamento de dados. O recurso mais valioso que o *Weka* oferece é a implementação de algoritmos de aprendizagens reais. Mas, as ferramentas de pré-processamento, chamadas de filtros, também são importantes ([WITTEN et al., 2011](#)).

Para [Witten et al. \(2011\)](#), o *Weka* pode ser considerado uma coleção de classes da linguagem *Java*, que compreendem, entre outras, classes de árvore de decisão, filtros de pré-processamento ou classes que implementam algoritmos específicos de aprendizagem. No *Weka* cada implementação de um algoritmo está encapsulada em uma classe e pode depender de outras classes para suas funcionalidades. A vantagem desta estrutura é a possibilidade de se criar códigos próprios em *Java* e integrar a eles classes de algoritmos de aprendizagem disponibilizados pelo *Weka*.

A Figura 15 apresenta uma das telas do *Weka*, a aba de visualização, utilizada para visualizar os atributos do conjunto de dados de entrada. Sua interface exibe uma matriz de duas dimensões de plotagem espalhada para cada par de atributos. Além disso, é possível alterar alguns parâmetros desta interface, como cor e tamanho, para melhorar a visualização dos dados no espaço bidimensional.

Figura 15 – Aba *Visualize* do *Weka*.



Fonte: Autoria própria.

No decorrer da metodologia será descrita de forma mais detalha algumas das funcionalidades do software *Weka*.

5 Metodologia Desenvolvida

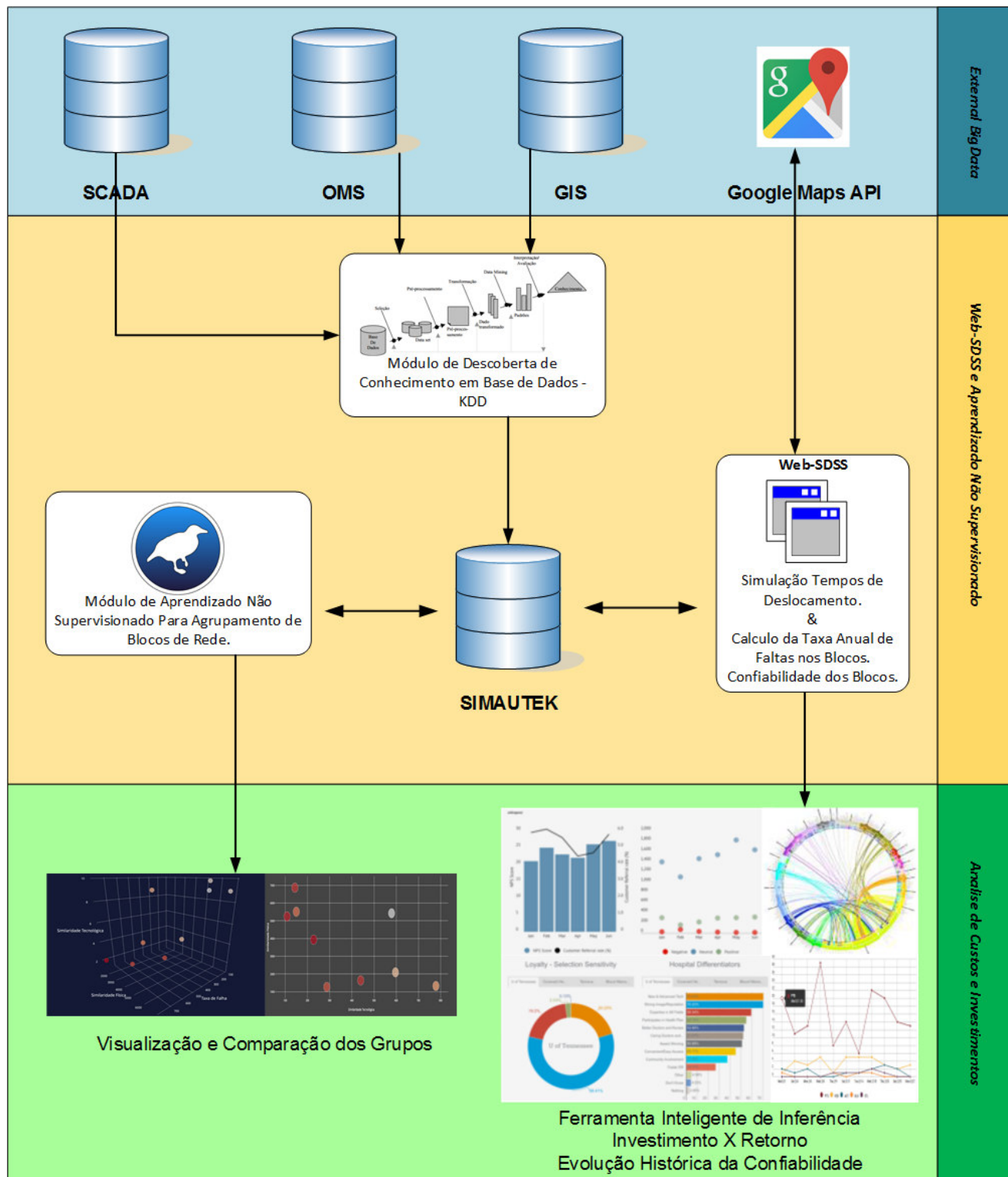
De modo geral, a metodologia desenvolvida deverá utilizar técnicas de *KDD* para filtrar, contextualizar e extrair informações úteis dos diferentes sistemas pré-existentes na empresa, isto é, bases de dados que não foram projetadas com um fim específico e orientados para algum aplicativo computacional. A metodologia concentra a execução de algoritmos de *data mining* através da ferramenta *Weka*, em que as etapas anteriores são executadas de forma independente, através da linguagem *SQL (Structured Query Language)*. Como a metodologia proposta é baseada no *KDD*, após a execução dos algoritmos de *data mining*, uma etapa final é necessária para avaliação e compreensão dos resultados obtidos.

Tal etapa é realizada de forma detalhada, analisando cada caso individualmente, e os resultados obtidos são apresentados para cada um dos estudos de caso através do *Web-SDSS SIMAUTEK*. O sistema proposto deverá reduzir significativamente o tempo de planejamento, tornando mais eficiente a análise do grande volume de variáveis envolvidas no processo de decisão e permitirá maximizar o retorno dos investimentos para a empresa. O fluxograma proposto pelo trabalho está apresentado na [Figura 16](#) e é composto por 3 níveis: *External Big Data*; *Web-SDSS* e *Aprendizado Não Supervisionado e Análise de Custos e Investimento*.

Para os desenvolvimentos, utiliza-se uma base de dados real, na qual são armazenados milhares de ocorrências vinculadas a desligamentos na rede de distribuição de energia. Após analisar cuidadosamente os campos e o conteúdo da base disponível, verificou-se que a mesma apresentava campos incompletos, informações contraditórias, campos sem dados e muitos outros problemas. Esses fatores colaboraram para concluir que a base, no estado inicial, era inapropriada para utilização direta. Dessa forma, o processo de *KDD* foi escolhido para regularizar os problemas existentes.

As próximas seções descrevem as etapas que envolvem o processo de *KDD* utilizado na metodologia. Algumas etapas são mais pormenorizadas devido a sua importância no processo geral. Um exemplo pode ser observado nas etapas de mineração de dados, a qual será apresentada na [seção 5.4](#) e o portal *Web SDSS SIMAUTEK*, para análise dos dados, descrito na [seção 5.5](#).

Figura 16 – Integração dos Módulos no ambiente das Redes de Distribuição Modernas.



Fonte: Autoria própria.

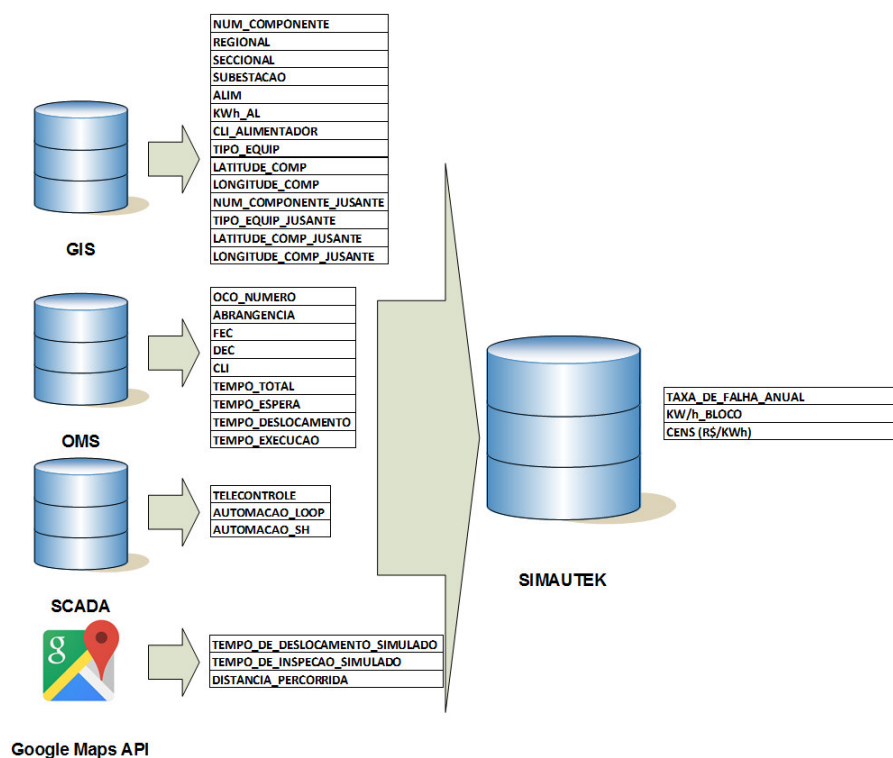
5.1 Etapa de Seleção dos Dados

O desafio desta metodologia é desenvolver uma técnica capaz de agrupar atributos do sistema elétrico e do sistema viário, com o objetivo de proporcionar subsídios importantes à análise do custo das faltas e as oportunidades de investimentos.

Os atributos devem ser adequadamente selecionados de forma a codificar a maior quantidade possível de informações relacionada à tarefa de interesse, além de apresentarem a mínima redundância entre eles. Nesta proposta, pretende-se utilizar atributos referentes às características elétricas, características lógicas e de acessibilidade, que formarão um espaço multidimensional.

Conforme apresentado na Figura 17, o conjunto de atributos utilizados originou-se de quatro tabelas principais: *GIS*, *OMS*, *SCADA*, e *API GOOGLE*. Da tabela cadastro, extraíram-se variáveis relacionadas à identificação dos equipamentos de proteção dos blocos impactados pelas faltas. A segunda tabela possui informações referentes à operação e ao impacto das faltas nos indicadores de continuidade. Na terceira tabela são extraídas informações referentes à disponibilidade para telecontrole e esquemas especiais de automação. Já o *Google Maps API* disponibiliza simulações dos tempos de deslocamento, inspeção e e distancias percorridas para mitigar as faltas.

Figura 17 – Diagrama relacional com as tabelas e atributos utilizados no sistema.



Fonte: Autoria própria.

A descrição dos atributos utilizados no sistema consta na [Tabela 19](#). A partir dos atributos primários se derivou o CENS, definido na [subseção 2.2.4](#), a taxa de falha anual, que consiste na soma das ocorrências anuais de falta para determinado componente e o consumo para determinado bloco através da proporcionalidade com o número de clientes do alimentador.

Tabela 18 – Atributos primários utilizados na metodologia.

Tabela	Atributo	Descrição
GIS	NUM_COMPONENTE	Identificador único do componente no banco de dados.
GIS	REGIONAL	Regional elétrica do componente
GIS	SECCIONAL	Seccional elétrica do componente
GIS	SUBESTACAO	Subestação ao qual o componente pertence
GIS	ALIM	Alimentador ao qual o componente pertence
GIS	kWh_AL	Consumo de energia estimado do circuito alimentador
GIS	CLI_ALIMENTADOR	Quantidade de clientes do alimentador ao qual o componente pertence
GIS	TIPO_EQUIP	Tipo do equipamento
GIS	LATITUDE_COMP	Latitude do componente, no formato WGS84
GIS	LONGITUDE_COMP	Longitude do componente, no formato WGS84
GIS	NUM_COMPONENTE_JUSANTE	Identificador único do componente, a jusante, no banco de dados.
GIS	TIPO_EQUIP_JUSANTE	Tipo do equipamento a jusante
GIS	LATITUDE_COMPONENTE_JUSANTE	Latitude do componente a jusante, no formato WGS84
GIS	LONGITUDE_COMPONENTE_JUSANTE	Longitude do componente a jusante, no formato WGS84
OMS	OCO_NUM	Número da ocorrência
OMS	ABRANGENCIA	Segmento ao qual pertence a ocorrência (subestação - SE, alimentador - AL, trafo - TF, chave - CH)
OMS	DEC	Duração Equivalente de Interrupção por Unidade Consumidora, apurado para cada ocorrência de falta
OMS	FEC	Frequência Equivalente de Interrupção por Unidade Consumidora, apurado para cada ocorrência de falta;
OMS	CLI	Quantidade de clientes interrompidos
OMS	TEMPO_TOTAL	Tempo médio de atendimento a emergências
OMS	TEMPO_ESPERA	tempo médio de espera
OMS	TEMPO_DESLOCAMENTO	tempo médio de deslocamento
OMS	TEMPO_EXECUCAO	tempo médio de execução
SCADA	TELECONTROLE	Flag que indica a presença ou não de telecontrole
SCADA	AUTOMAÇÃO_LOOP	Flag que indica se o componente pertence ao <i>loop automation</i>
SCADA	AUTOMAÇÃO_SH	Flag que indica se o componente pertence ao <i>self-healing</i>
API_OOGLE	TEMPO_DESLOCAMENTO_SIMULADO	tempo médio simulado de deslocamento
API_OOGLE	TEMPO_INSPEÇÃO_SIMULADO	tempo médio simulado de inspeção
API_OOGLE	DISTANCIA_PERCORRIDA	distância percorrida simulada para mitigar a fata

Fonte: Autoria própria.

A concessionária de energia cedeu para análise 17.920 registros de desligamentos ocorridos, referente ao segmento chave (CH), de janeiro a setembro de 2018. As tabelas de dados foram fornecidas no *Oracle*, dessa forma foi possível valer-se de uma ferramenta mais poderosa para tratamento dos dados. A consulta ao banco *Oracle* foi realizada utilizando uma versão de desenvolvimento do software *PL/SQL Developer*.

Cada atributo deverá compor uma dimensão no espaço ocupado pelos blocos. Os algoritmos de aprendizado não supervisionados *K-means*, não estão sujeitos à restrição dimensional (HAN et al., 2011).

5.2 Etapa de Pré-processamento dos Dados

Depois de selecionados os dados a serem trabalhados, passou-se para a etapa de pré-processamento dos dados. Utilizando o histórico de eventos fornecido pela empresa, identificou-se a estrutura relacional aplicada ao banco de dados através do seu diagrama de entidades e relacionamentos, e verificaram-se os tipos de ferramentas necessárias para proceder à etapa em questão.

As consultas ao banco de dados foram realizadas utilizando a linguagem *Structured Query Language (SQL)*. Dessa forma, foi possível extrair da estrutura original as variáveis definidas na seção anterior, abandonando a estrutura original e adotando uma nova estrutura relacional, adequada ao armazenamento das variáveis de interesse.

Na [Figura 18](#) apresenta-se um exemplo descritivo de uma sequência de filtros criados com o objetivo de limpar e integrar as variáveis de interesse na nova estrutura utilizada. Assim, do conjunto original de eventos de desligamentos armazenados no banco de dados, constatou-se que apenas metade desses dados se adequava ao domínio de estudo. Portanto, grande parte desses registros foi descartada da nova base de dados.

Conforme a [Figura 18](#), os primeiros registros a serem eliminados da base de eventos foram os que apontavam como causa um desligamento que havia sido programado pela companhia distribuidora de energia. Cita-se como exemplo, desligamentos para manutenção preventiva, alteração para melhorias, desligamento a pedido do cliente, entre outros. Em seguida, foram retirados os registros contendo os valores '000', por ser um valor inexistente na lista de códigos de causa da companhia e valores do tipo *NULL*, indicando que o campo não foi preenchido pelo eletricitista na hora do atendimento da ordem de serviço e também registros de componentes que não existem mais na base de cadastro, devido à dificuldade em se conseguir informações para a análise, este último caso refere-se por exemplo a chaves que deixaram de existir na rede ou mudaram de alimentador devido à dinâmica do sistema de distribuição de energia.

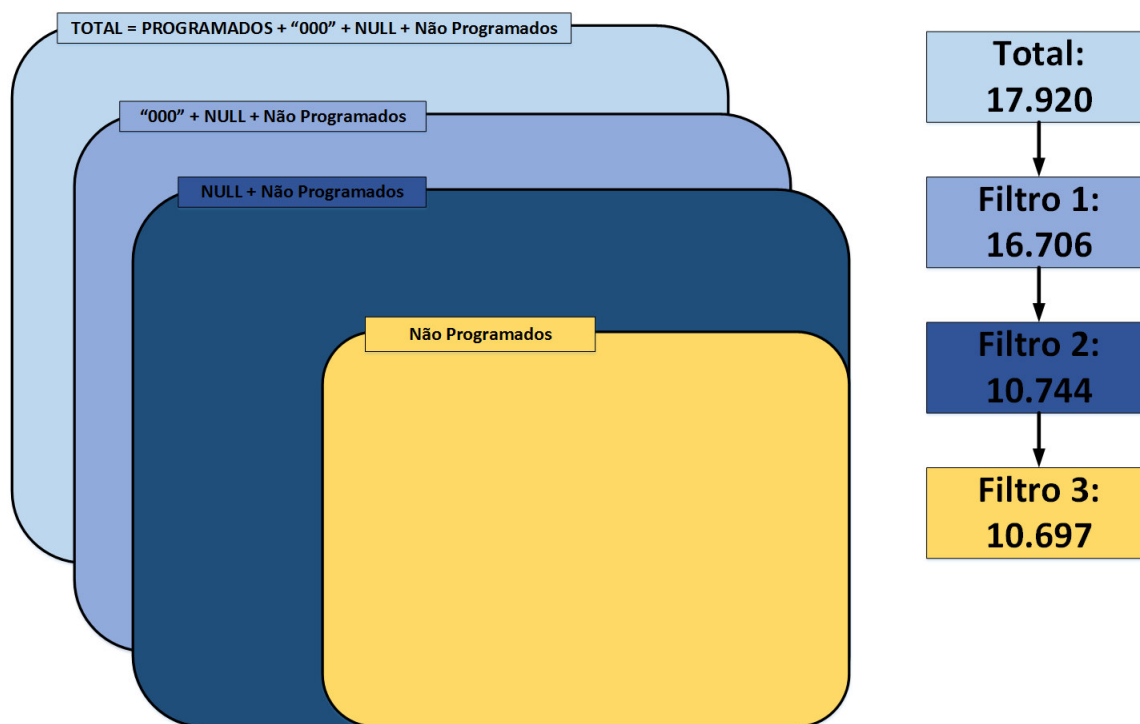
Desta forma, foi possível separar apenas registros de desligamentos não programados, os quais interessam ao estudo em questão. A [Figura 18](#) apresenta uma visão da ordem de grandeza (quantificação) de dados eliminados do conjunto inicial utilizando os critérios descritos acima.

Após essa limpeza e integração, dos 17.920 registros fornecidos, apenas 59% dos dados foram considerados aderentes aos eventos registrados para serem utilizados no sistema, totalizando 10.697 registros. Neste ponto, é importante elaborar uma reflexão acerca dos registros coletados em campo.

Uma das maiores preocupações em ([PRETTO et al., 2006](#)) foi retirar a responsabilidade de apontamento de uma causa de interrupção do eletricitista (coletor), pois observou-se que durante a recomposição da rede, muitas vezes não é possível registrar adequadamente

os eventos, e a grande maioria das vezes, os eletricitas são direcionados para recompor rapidamente a interrupção, com o objetivo de mitigar indicadores de duração e frequência.

Figura 18 – Filtragem dos dados.



Fonte: Autoria própria.

Dessa forma, a redução proposta para a base de dados traduz uma forte aderência aos aspectos de identificação das variáveis definidas para o problema. A nova base foi acomodada em uma estrutura com as seguintes quantificações: 10.697 eventos registrados. É importante observar que a base iniciou com 17.920 registros e passou para 10.697 eventos de interrupções com condições de utilização.

5.3 Etapa de Transformação

A etapa de transformação consiste em aplicar procedimentos sobre os dados de forma que eles sejam modificados para execução das etapas posteriores. As modificações sobre os dados são realizadas em cada atributo de acordo com seu tipo de valor e de acordo com alguma simplificação que se deseja obter.

Neste trabalho, a etapa de transformação teve por objetivo simplificar a forma como os valores são apresentados, de forma que, as informações que antes eram distribuídas por ocorrência a partir desta etapa passaram a ser identificadas por blocos impactados por faltas (componente), conforme descrito na [Figura 19](#).

Figura 19 – Transformação dos dados.

OCO_NUMERO	NUM_COMPONENTE	ALIM	KWh_AL	CLI_ALIMENTADOR	TIPO_EQUIP	ABRANGENCIA	FEC	DEC	CLI	KWh_BLOCO	TEMPO_TOTAL
555702	0152404	CDO01C1	979.7852359	869	RELIGADOR	CH	0.0017001	0.00066293	869	979.7852359	14.64
60456	0152404	CDO01C1	979.7852359	869	RELIGADOR	CH	0.00169848	0.002109341	869	979.7852359	13.68
251170	0152404	CDO01C1	979.7852359	869	RELIGADOR	CH	7.6232E-05	0.000234773	869	979.7852359	4.08
110727	0152404	CDO01C1	979.7852359	869	RELIGADOR	CH	0.00170453	0.00109831	869	979.7852359	3.12

NUM_COMPONENTE	TAXA_DE_FALHA_ANUAL	ALIM	KWh_AL	CLI_ALIMENTADOR	TIPO_EQUIP	ABRANGENCIA	FEC	DEC	CLI	KWh_BLOCO	CENS (R\$/KWh)	TEMPO_TOTAL
152404	4	CDO01C1	979.7852359	869	RELIGADOR	CH	0.00517934	0.004105355	3476	3919.140944	181.8109485	35.52

Fonte: Autoria própria.

O passo final da etapa de transformação é a construção do arquivo que compreende o conjunto de dados preparado para posterior execução do algoritmo de *data mining* aceito pelo *WEKA*, o *ARFF* (*Attribute-Relation File Format*). Um exemplo contendo um trecho de um conjunto de dados no formato *ARFF* é apresentado na [Figura 20](#).

Figura 20 – Arquivo no Formato *ARFF*.

```

Scheme:      weka.clusterers.SimpleKMeans -
Relation:    ANALISE_CLUSTER-weka.filters.u
Instances:   385
Attributes:  31
             TAXA_DE_FALHA_ANUAL
             TIPO_EQUIP
             FEC
             DEC
             CLI
             KWh_BLOCO
             CENS (R$/Kwh)
             TEMPO_TOTAL
             TEMPO_DE_DESLOCAMENTO_SIMULADO
             TEMPO_DE_INSPECAO_SIMULADO
             DISTANCIA
             AUTOMACAO
             LOOP
             SH

```

Fonte: Autoria própria.

No cabeçalho são definidos o nome do conjunto de dados (*@relation*) e os atributos e seus respectivos tipos (*@attribute*). A seguir, a região identificada por *@data* contém uma sequência de registros em cada linha e os valores dos seus atributos separados por vírgula. No final desta etapa foram considerados 6.468 registros.

5.4 Etapa de Mineração de Dados

A quarta etapa da metodologia para descoberta de conhecimento em bancos de eventos de falta em rede de distribuição de energia é a etapa de *data mining*, onde é executada a técnica e o algoritmo definido. Esta é a etapa na qual os dados são, de fato, minerados. Em outras palavras, e como descrito na [seção 4.3](#), é onde a informação é esculpida para que se possa obter um novo conhecimento, uma informação nova e útil.

Na aplicação da tarefa de agrupamento deste estudo de caso, optou-se por trabalhar com a técnica de particionamento. Esta técnica divide o conjunto de dados de forma que os dados mais semelhantes entre si são agrupados e, ao mesmo tempo, estes ficam separados dos demais.

O objetivo é que, através dos agrupamentos, seja possível reduzir significativamente a quantidade de dados a serem tratados e o esforço computacional de avaliação, pois cada classe deverá agrupar blocos de redes muito similares do ponto de vista de contribuição no custo de falta, facilitando a análise de seus custos e do impacto de diferentes ações de investimento. Desta forma, encontrando o grupo com melhor retorno de investimento, o mesmo tipo de ação poderá ser aplicado para cada bloco do mesmo grupo, facilitando a priorização e o acompanhamento dos investimentos já realizados.

Com o objetivo da etapa de *data mining* definido e os dados para análise selecionados e transformados, deve-se definir o algoritmo que melhor se adequa nesta situação. Há várias opções de algoritmos para executar a técnica de particionamento. O *k-means* foi o método escolhido, sendo denominada de *SimpleKMeans* a implementação em *Java* do algoritmo *k-means* na ferramenta *Weka*. É importante observar que o algoritmo *SimpleKMeans* utiliza, como medida de distância entre os atributos numéricos, a medida de distância euclidiana. Essa medida, por sua vez, aplica automaticamente a normalização dos atributos numéricos ao realizar o agrupamento.

A escolha do método *k-means* se dá pela sua simplicidade e eficiência, sendo que, em comparação com diversos outros algoritmos de agrupamento, o *k-means* é o mais eficiente em termos de tempo de execução ([MACQUEEN et al., 1967](#)).

Este algoritmo possui vários parâmetros de entrada, como pode ser visto na [Tabela 19](#). Os parâmetros estão disponíveis na ferramenta *Weka* com valores pré-definidos (valores *default*), e, com exceção do parâmetro *numClusters*, os demais parâmetros tiveram seus valores mantidos. Em *displayStdDevs*, o valor foi definido como verdadeiro, pois se deseja visualizar dados adicionais estatísticos de desvio padrão dos atributos; o parâmetro *distanceFunction* foi definido como a distância euclidiana para comparação entre os dados, que é o cálculo de distância mais simples baseado no plano cartesiano; *dontReplaceMissingValues* foi utilizado como falso, pois é o valor pré-definido, mas como não há valores ausentes nos conjuntos de dados, esse parâmetro não é relevante para

o estudo de caso; o parâmetro *initializationMethod* foi definido como aleatório para a escolha randômica inicial dos centroides; em *maxIterations*, o valor de iterações máximo é definido como 500, pois é um limite superior para o número de iterações no processo de recalcular os centroides; em *preserveInstancesOrder*, o valor falso foi utilizado para que não seja necessário analisar as instâncias na ordem, de entrada; o parâmetro *seed* foi mantido em 10, pois é o valor pré-definido para a geração de números aleatórios no algoritmo.

Tabela 19 – Principais parâmetros disponíveis para execução do algoritmo *SimpleKMeans*, suas descrições e valores utilizados durante o estudo de caso.

Parâmetro	Descrição	Valor Utilizado
displayStdDev	Exibe desvios padrão de atributos numéricos e a contagem de atributos nominais	Verdadeiro
distanceFunction	A função de distância a ser usada para comparar instâncias	Distância Euclidiana
dontReplaceMissingValues	Substituir globalmente valores ausentes com a média/moda	Falso
initializationMethod	Define a forma de inicialização dos centros de <i>cluster</i>	Aleatório
maxIterations	Define o número máximo de interações	500
preserveInstancesOrder	Preserva a ordem das instâncias.	Falso
seed	Valor aleatório de <i>seed</i> a ser usado	10

Fonte: Autoria própria.

Como já dito anteriormente, a regra do cotovelo executa o *k-means* no conjunto de dados, com o intuito de identificar a quantidade ideal de *clusters*. A partir das simulações do *Weka* foi possível obter os valores de erro quadrático, conforme ilustrado na [Figura 21](#) e a partir de então construir o gráfico gerado por este método apresentando o erro quadrático de acordo com o número de *clusters*, apresentado na [Figura 22](#).

É considerada a quantidade ideal, o ponto onde os erros começam a se estabilizar, observe que este ponto estável é identificado pelo valor 8. Portanto, o valor de *numClusters* é 8. Para a execução da regra do cotovelo, foram utilizados apenas os atributos escolhidos na etapa de seleção.

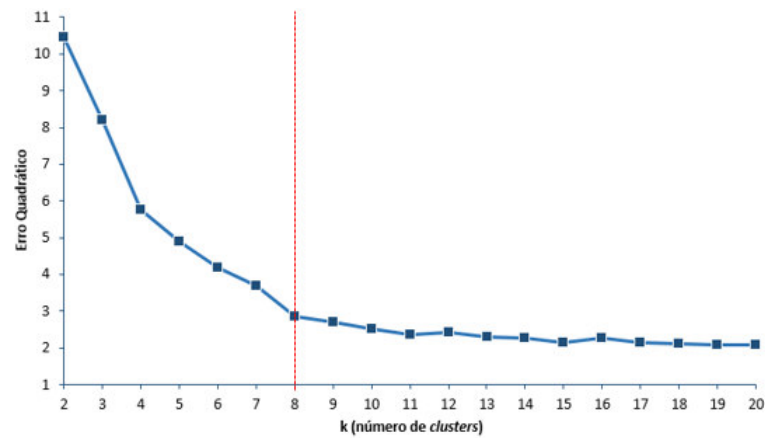
Figura 21 – Simulação para identificar número ótimo de *clusters*.

```
kMeans
=====

Number of iterations: 33
Within cluster sum of squared errors: 2.863209138583003E8
```

Fonte: Autoria própria.

Figura 22 – Quantidade ideal de *clusters* a serem gerados utilizando os atributos considerados relevantes.



Fonte: Autoria própria.

Todos os testes foram executados utilizando apenas quatorze dos trinta e um atributos selecionados na etapa de seleção dos dados. A grande maioria dos atributos são cadastrais e foram eliminados, visto que não influenciaram nos resultados. Manteve-se, então, os campos referentes aos indicadores de qualidade do serviço, tempos de atendimento, distâncias e as informações de automação, conforme apresentado na [Figura 23](#).

Figura 23 – Atributos utilizados na Mineração de Dados.

```

Attributes:  31
TAXA_DE_FALHA_ANUAL
TIPO_EQUIP
FEC
DEC
CLI
Kwh_BLOCO
CENS (R$/kwh)
TEMPO_TOTAL
TEMPO_DE_DESLOCAMENTO_SIMULADO
TEMPO_DE_INSPECAO_SIMULADO
DISTANCIA
AUTOMACAO
LOOP
SH

Ignored:
NUM_COMPONENTE
REGIONAL
SECCIONAL
SUBESTACAO
ALIM
Kwh_AL
CLI_ALIMENTADOR
ABRANGENCIA
TEMPO_ESPERA
TEMPO_DESLOCAMENTO
TEMPO_EXECUCAO
LATITUDE_COMP
LONGITUDE_COMP
PROXIMO_COMP
TIPO_PROX_COMP
LATITUDE_PROX_COMP
LONGITUDE_PROX_COMP

```

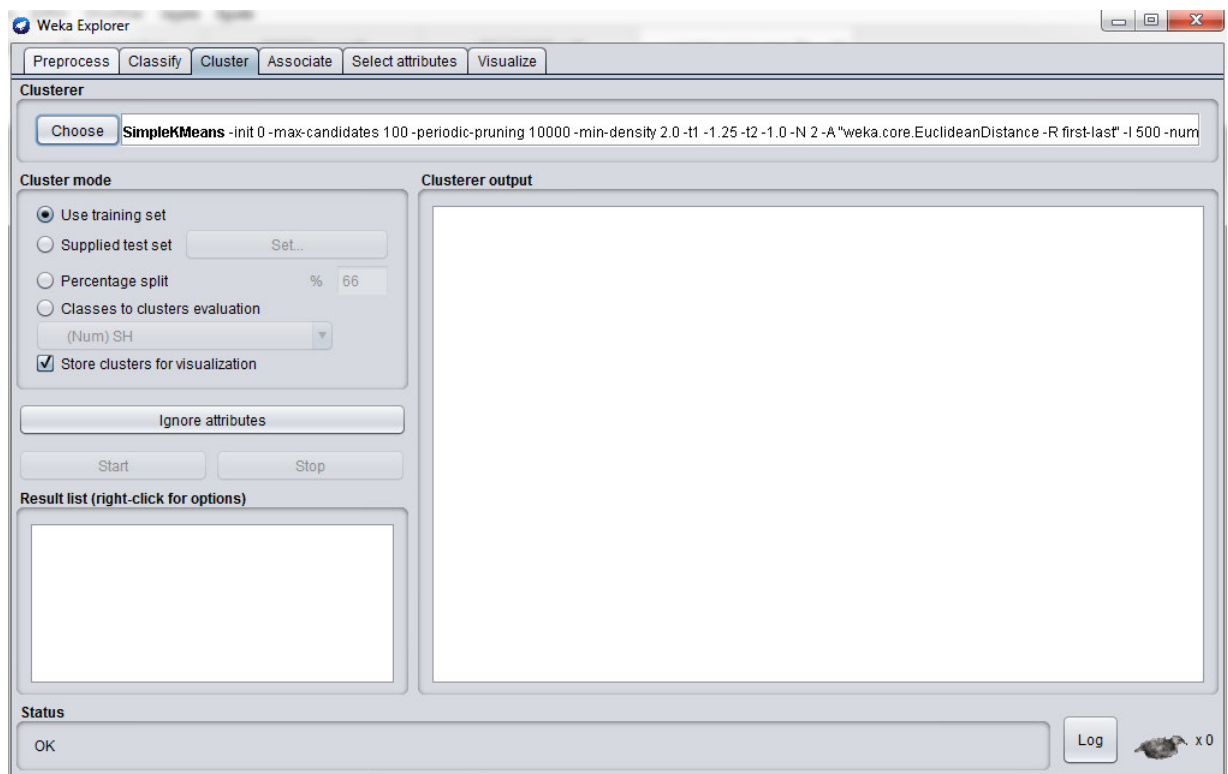
Fonte: Autoria própria.

A tela da ferramenta *Weka* para a geração de *cluster* é apresentada na [Figura 24](#).

Nesta figura pode ser observado que a definição do algoritmo é realizada através do botão *Choose* e também que é necessário escolher um modo de agrupamento dentre os disponíveis no painel *Cluster mode*. Cada modo de agrupamento é avaliado de maneira distinta e, para cada modo, é realizada a avaliação dos modelos de *clusters* gerados.

A ferramenta *Weka* executa o algoritmo de *cluster* inicialmente para a geração dos *clusters* desejados. Cada instância do conjunto de dados é associada a um *cluster* de acordo com a proximidade de seus atributos com o centroide do *cluster*. Este processo é considerado o treinamento do algoritmo de *cluster*. Além disso, a ferramenta *Weka* também contabiliza a porcentagem de instâncias associadas a cada *cluster*. Este processo é considerado como o teste do agrupamento e é realizado de maneiras diferentes em cada modo de execução, descritos nesta seção.

Figura 24 – Interface inicial da execução da tarefa de agrupamento na ferramenta *Weka*.

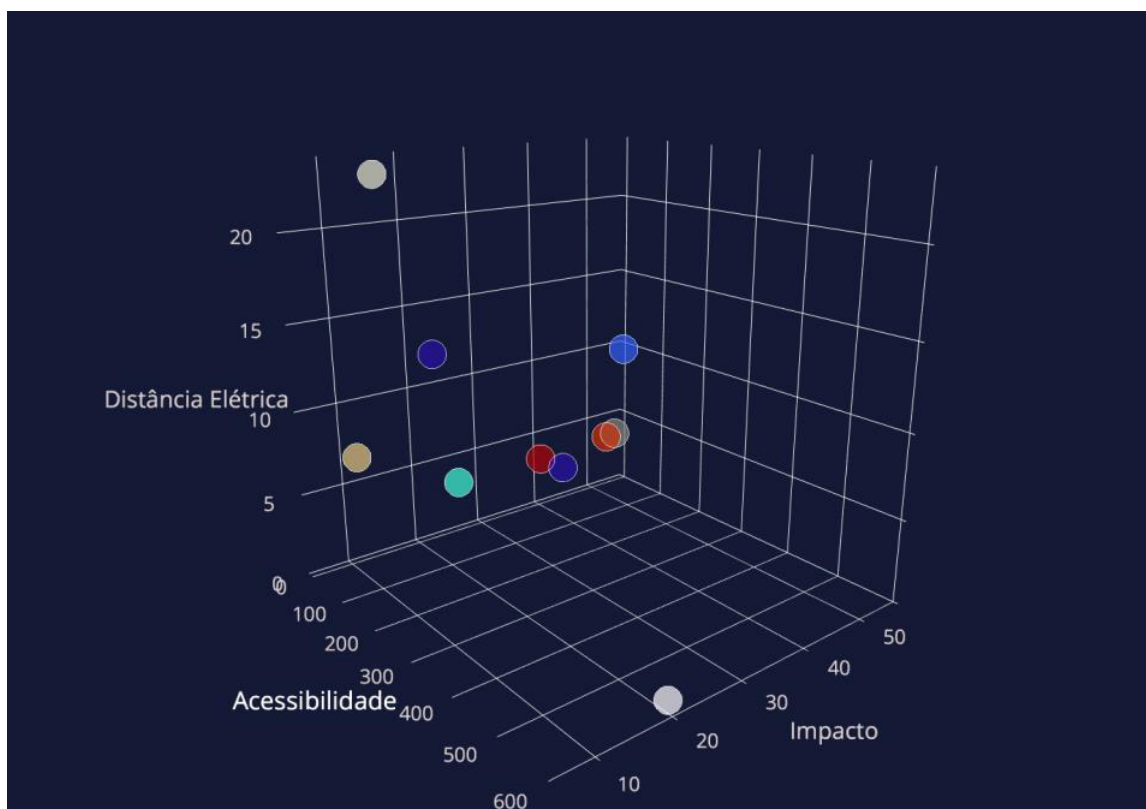


Fonte: Autoria própria.

Em *Use training set*, o algoritmo de *cluster* é executado sobre o conjunto de dados carregado anteriormente e, então, as instâncias são associadas aos *clusters*, de acordo com o centroide de cada *cluster*, em um processo considerado treinamento do agrupamento. Posteriormente, a porcentagem de instâncias associadas a cada *cluster* é calculada como parte do teste do agrupamento. Após realizadas todas as definições para a execução do algoritmo de *cluster*, pode-se dar início ao processo através do botão *Start* (Figura 24) e, então, é finalizada a etapa de *data mining*.

Como forma de visualizar a ideia da etapa de mineração de dados, a Figura 25 apresenta uma correlação espacial entre três atributos: (a) o número de consumidores atingidos em uma possível falta, (b) o tempo de acesso ao bloco considerando a estrutura viária disponível e o carregamento médio das vias e (c) a distância elétrica do bloco. É importante salientar que esta é apenas uma simplificação para apresentar a ideia do uso da metodologia, pois na realidade são consideradas múltiplas dimensões.

Figura 25 – Diagrama tridimensional de correlação entre o tempo de acesso de cada bloco de rede considerando o tráfego (Acessibilidade), a distância elétrica de cada bloco e o impacto de uma possível falta de cada bloco em função do número de consumidores e a taxa de falha do bloco.



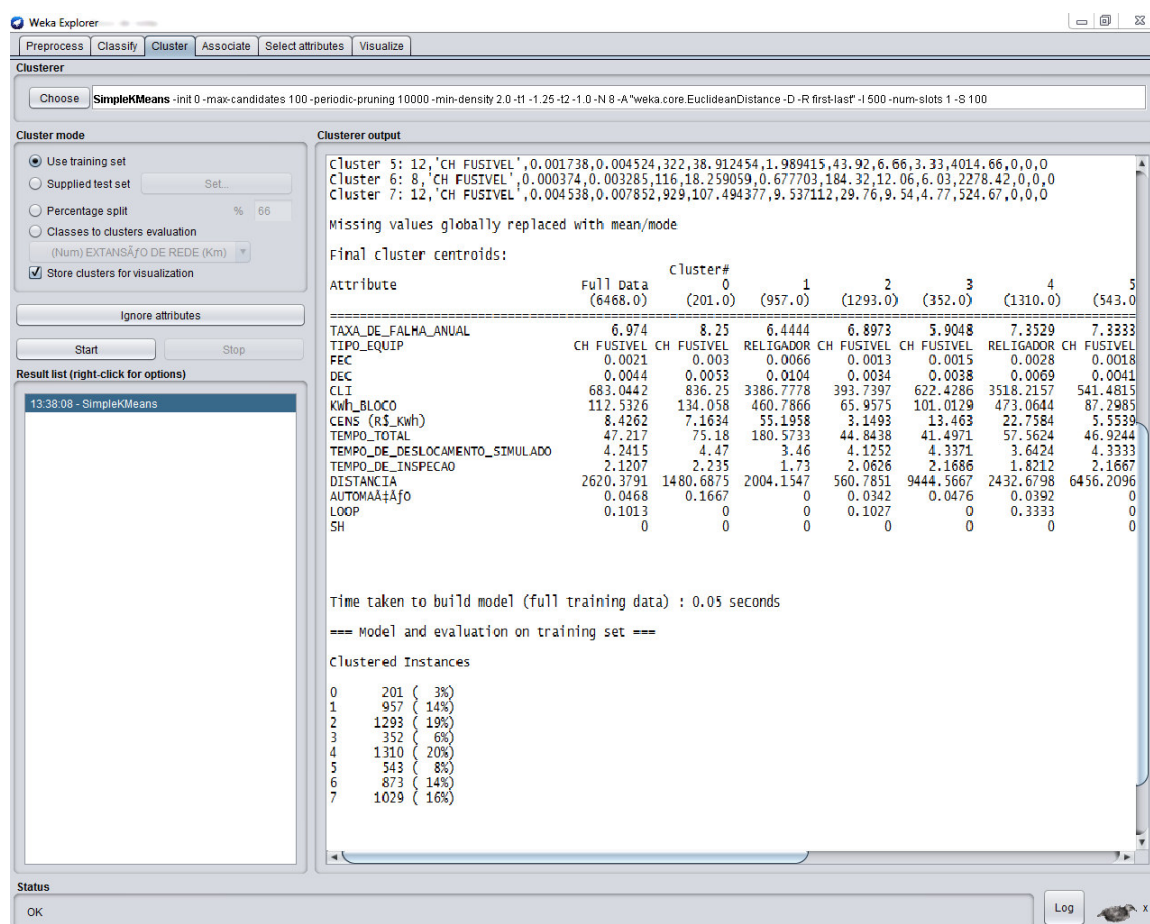
Fonte: Autoria própria.

5.5 Etapa de Interpretação e Avaliação dos Resultados

Na última etapa da metodologia verifica-se a possibilidade de interpretar o resultado obtido. Em caso positivo, é feita uma avaliação dos dados a fim de validar se o resultado é uma informação nova e se essa informação pode ser útil. Caso não seja possível interpretar o resultado ou o resultado não tenha gerado um conhecimento novo e útil, é necessário refazer a etapa de *data mining*, ou, então, refazer todas as etapas desde o início, até que seja encontrado um resultado satisfatório.

Os resultados foram obtidos através da opção *Use training set*, que pode ser visto na Figura 26, Tabela 20 e Tabela 21. Pode-se observar que foram gerados oito *clusters*. Os centroides de um *cluster* representam a média gerada pelo *k-means* em cada atributo pertencente a ele. Através dessas médias, foi possível analisar as características dos perfis, de acordo com variáveis definidas anteriormente. Cada centroide corresponde a um perfil de solução presente na rede.

Figura 26 – Interface do resultado do *cluster* utilizando a opção *Use training set*.



Fonte: Autoria própria.

Tabela 20 – Centróides 0 a 3 resultantes da execução da tarefa de clusterização.

Attribute	Full Data (6468.0)	0 (201.0)	1 (957.0)	2 (1293.0)	3 (352.0)
TAXA_DE_FALHA_ANUAL	6.794	8.25	6.4444	6.8973	5.9048
TIPO_EQUIP	CH FUSIVEL	CH FUSIVEL	RELIGADOR	CH FUSIVEL	CH FUSIVEL
FEC	0.0021	0.003	0.0066	0.0013	0.0015
DEC	0.0044	0.0053	0.0104	0.0034	0.0038
CLI	683.0442	836.25	3386.7778	393.739	622.4286
KWh_BLOCO	112.5326	134.058	460.7866	65.9575	101.0129
CENS (R\$ kWh)	8.4262	7.1634	55.1958	3.1493	13.463
TEMPO_TOTAL	47.217	75.18	180.5733	44.8438	41.4971
TEMPO_DE_DESLOCAMENTO_SIMULADO	4.2415	4.47	3.46	4.1252	4.3371
TEMPO_DE_INSPECAO	2.1207	2.235	1.73	2.0626	2.1686
DISTANCIA	2620.3791	1480.6875	2004.1547	560.7851	9444.5667
AUTOMACAO	0.0468	0.1667	0	0.0342	0.0476
LOOP	0.1013	0	0	0.1027	0
SH	0	0	0	0	0

Fonte: Autoria própria.

Tabela 21 – Centróides 4 a 7 resultantes da execução da tarefa de clusterização.

Attribute	4 (1310.0)	5 (543.0)	6 (873.0)	7 (1029.0)
TAXA_DE_FALHA_ANUAL	7.3529	7.3333	6	6.1538
TIPO_EQUIP	RELIGADOR	CH FUSIVEL	DISJUNTOR	RELIGADOR
FEC	0.0028	0.0018	0.0061	0.0022
DEC	0.0069	0.0041	0.0138	0.0051
CLI	3518.2157	541.4815	2561.9091	904.7778
KWh_BLOCO	473.0644	87.2985	413.0695	153.6022
CENS (R\$ kWh)	22.7584	5.5539	71.3743	8.5902
TEMPO_TOTAL	57.5624	46.9244	47.76	37.1911
TEMPO_DE_DESLOCAMENTO_SIMULADO	3.6424	4.3333	3.2564	3.3867
TEMPO_DE_INSPECAO	1.8212	2.1667	1.6282	1.6933
DISTANCIA	2432.6794	6456.2096	5314.6845	870.2241
AUTOMACAO	0.0392	0	1	0
LOOP	0.3333	0	0	0
SH	0	0	0	0

Fonte: Autoria própria.

Analisando os centroides é possível observar, de modo geral, que o *cluster* 4 é o mais populoso, com 20% dos registros. Observa-se também que os *clusters* que possuem maior impacto sobre os indicadores de qualidade são também os que apresentam o maior Custo da Energia Não Fornecida (CENS), e ainda, a partir desta análise, observa-se que os *clusters* 1 e 4 apresentam semelhança em relação ao tipo de equipamento e às características físicas e de acessibilidade da rede, porém, se observa também uma diferença com relação ao CENS destes *clusters*. Outra observação importante é que no *cluster* 03 o custo do bloco é intermediário com grandes distâncias e uma baixa quantidade de clientes. Esta análise não se esgota neste trabalho e pode ter várias interpretações. Para complementar as informações geradas pelo módulo de aprendizado não supervisionado através da ferramenta *Weka* utilizou-se o sistema de suporte a decisão espacial desenvolvido na empresa, o SIMAUTEk, onde é possível analisar através de uma ferramenta integrada as informações de acessibilidade da rede, indicadores de qualidade do serviço e disponibilidade de tecnologias. As análises aqui realizadas devem gerar propostas de ações de investimento com melhor o custo-benefício para cada bloco do sistema de distribuição.

5.5.1 Portal SIMAUTEK

Nesta subsecção serão apresentados os resultados obtidos no desenvolvimento do portal *web* do SIMAUTEK e como esta ferramenta se integra ao objetivo do trabalho. De maneira geral, a aplicação consiste em consultas, filtros e telas que alteram sua aparência e conteúdo conforme recebe solicitações de usuário e permitem as funcionalidades de acompanhamento e gestão da manutenção e automação.

5.5.1.1 Modulo de Gestão da Automação e Manutenção

Após realizar o login, é inicializada a tela que permite, através de *dashboards*, e gráficos de colunas avaliar as informações de disponibilidade das soluções de comunicação e efetividade operacional dos ativos de automação.

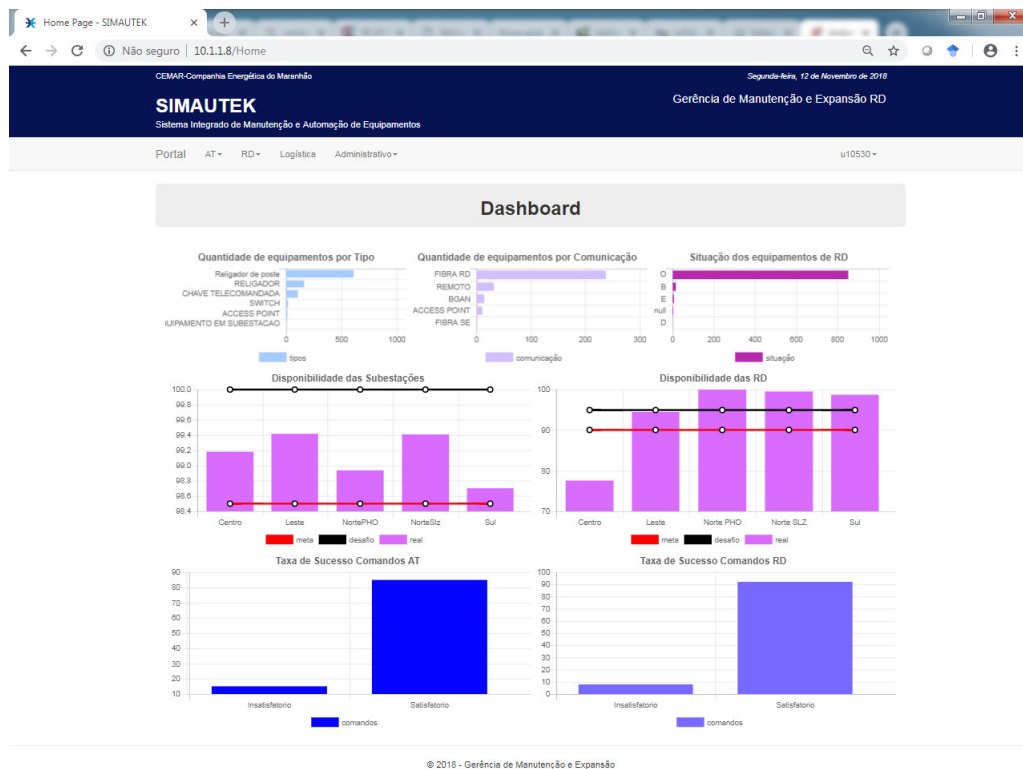
Na parte superior da página inicial é possível fazer o acompanhamento quantitativo dos equipamentos de proteção/seccionamento da rede MT, entre chaves, seccionadoras, religadores e reguladores de tensão, também permite acompanhar o status de operação destes, que pode ser em operação e fora de operação. Ainda na parte superior da tela é possível realizar uma análise quantitativa dos dispositivos de comunicação, seja fibra ótica, satélite ou rádio.

Além do acompanhamento quantitativo, estratificado por tipo de equipamento de proteção/seccionamento e comunicação, outro recurso importante implementado na tela inicial do SIMAUTEK foi o *dashboard* de disponibilidade da comunicação, esta análise permite que ações gerências sejam tomadas reduzindo os tempos de indisponibilidade das soluções de comunicação e o impacto nos indicadores de disponibilidade da automação.

Outro acompanhamento importante, ainda na tela inicial é a Taxa de Sucesso de Comando. Este *dashboard* reflete o percentual de sucesso dos telecomandos e é calculado a partir da relação dos comandos executados com sucesso e o total de comandos enviado pela operação.

As funcionalidades da página inicial do SIMAUTEK são apresentadas na [Figura 27](#).

Figura 27 – Pagina inicial do SIMAUTEK - Módulo de gestão da automação e manutenção



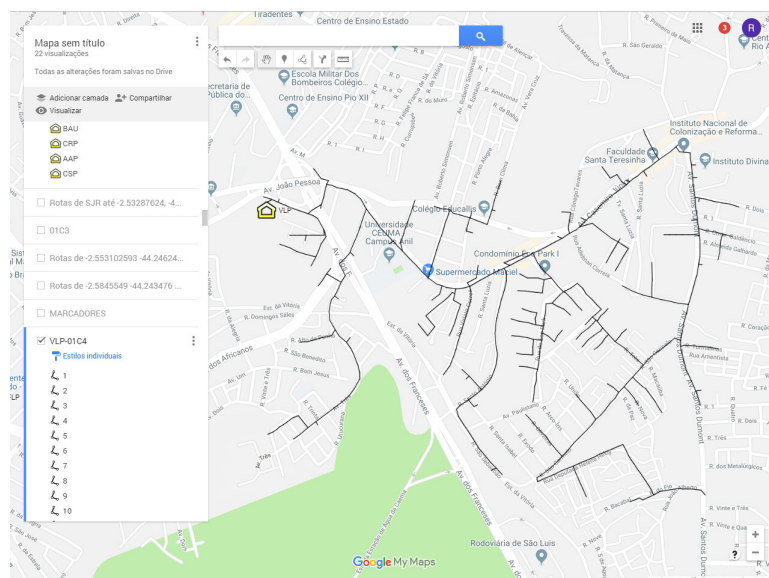
Fonte: Autoria própria.

5.5.2 Módulo de Simulação de Tempos de Deslocamento

Este módulo permite associar um trecho de rede elétrica a um arco de via e é parte crucial deste trabalho. Através deste módulo é possível criar camadas com marcadores e rotas que permitem a análise da posição de cada equipamento, sua capacidade de manobra bem como a acessibilidade das equipes em caso de falha.

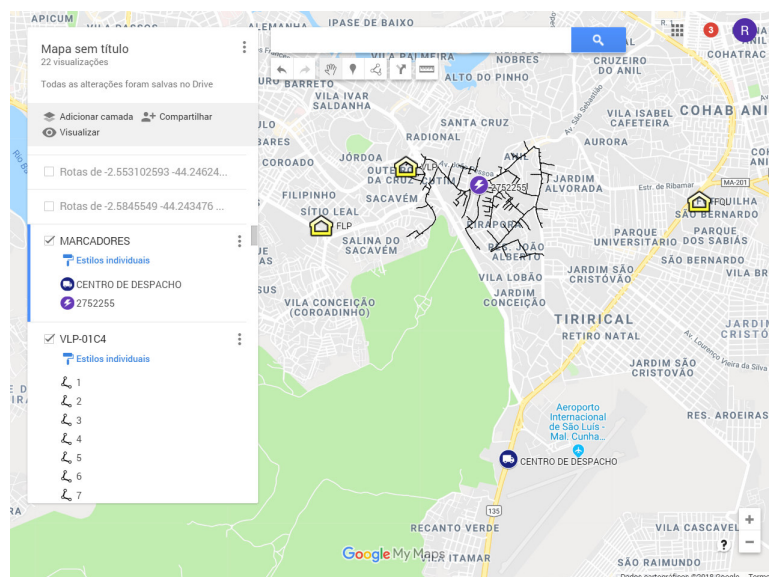
A Figura 28 apresenta um exemplo de uma rede elétrica, em seguida, na Figura 29 são apresentados os marcadores que definem os pontos de origem e destino e por último na Figura 30 são apresentadas as simulações de tempo de deslocamento e inspeção com relação a o componente/bloco 2752255 pertencente ao alimentador VLP-01C4.

Figura 28 – Camada que ilustra o diagrama do alimentador VLP-01C4.



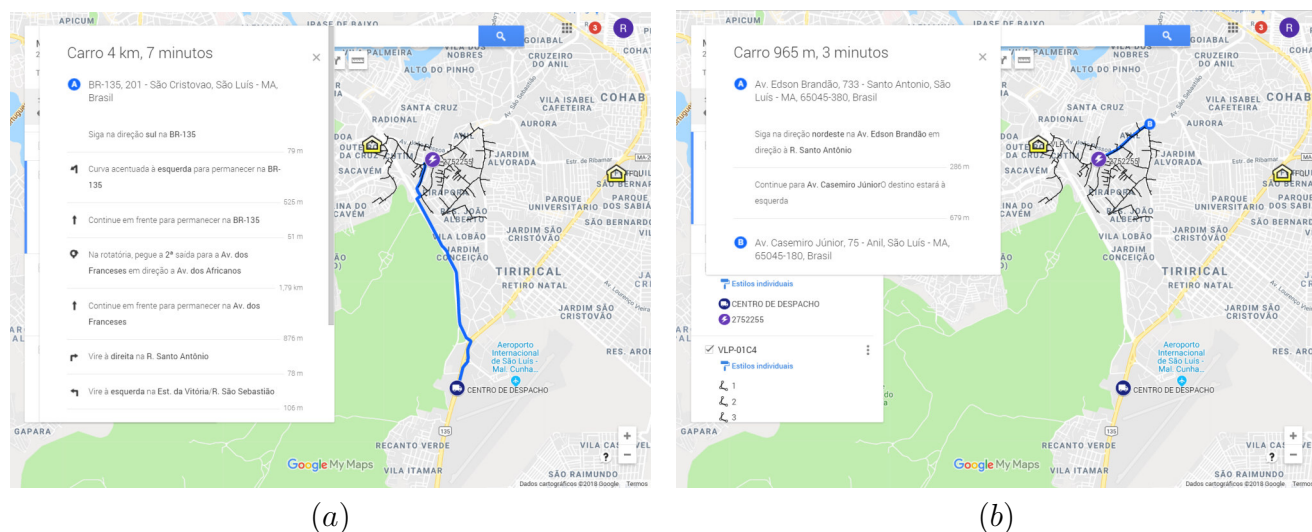
Fonte: Autoria própria.

Figura 29 – Camada que os marcadores do centro de despacho e o bloco em falha.



Fonte: Autoria própria.

Figura 30 – (a) Tempo de deslocamento e (b) Tempo de inspeção



Fonte: Autoria própria.

5.5.3 Módulo de Acompanhamento do Desempenho e da Taxa de falha dos Blocos da Rede de Distribuição

Este módulo é o responsável por buscar valores do histórico de atuações em equipamentos telecontrolados pelo *SCADA* e ocorrências de eventos de faltas na rede de média tensão registrados pelo *OMS*, de forma a se obter a taxa anual de faltas em cada bloco de rede (Trechos de rede delimitados por elementos de manobra). Este módulo também é responsável por avaliar, no decorrer do tempo, o impacto das alterações realizadas na rede sobre os indicadores de qualidade do serviço.

A partir deste módulo é possível analisar, através de gráficos e tabelas, os blocos com maior taxa de falha e com maior impacto e então realizar ações de para mitigar as falhas, conforme ilustrado na [Figura 31](#) e [Figura 32](#).

Figura 31 – Módulo de acompanhamento dos indicadores de qualidade do serviço.

Ocorrências DEC FEC

25 resultados por página

01/04/2018 15:48:29 12/11/2018 15:48:29

Data	Componente	SE	Alimentador	Regional	Causa
03/05/2018 17:37:49	0565644	BRD	01C4	Leste	DP - Falha em Co
26/05/2018 22:45:04	2731673	SJR	01C2	NorteSiz	DETERIORAÇÃO MATERIAL COMI
04/06/2018 13:43:03	2758555	BCP	09N2	Sul	DETERIORAÇÃO MATERIAL COMI
04/07/2018 08:19:25	2355337	MAR	01C2	NorteSiz	ARVORE NA REG COMINTER
09/08/2018 21:54:50	2813807	IPA	01C6	Sul	Transitorio
17/08/2018 15:00:36	2813807	IPA	01C6	Sul	FENÔMENOS NA COMI INTER
22/08/2018 21:13:43	2813807	IPA	01C6	Sul	Transitorio
26/08/2018 07:52:18	2813807	IPA	01C6	Sul	FENÔMENOS NA COMI INTER
09/09/2018 13:53:37	2704765	RCO	09F3	Sul	FENÔMENOS NA COMI INTER
10/09/2018 18:02:45	2704765	RCO	09F3	Sul	FENÔMENOS NA COMI INTER
12/09/2018 20:16:26	0229539	ACL	09F1	Sul	FENÔMENOS NA COMI INTER
16/10/2018 06:17:15	275214X	CET	01W2	NorteSiz	DETERIORAÇÃO MATERIAL COMI

Mostrando de 1 até 25 de 1.389 registros

Anterior 1 2 3 4 5 ... 55 Próximo

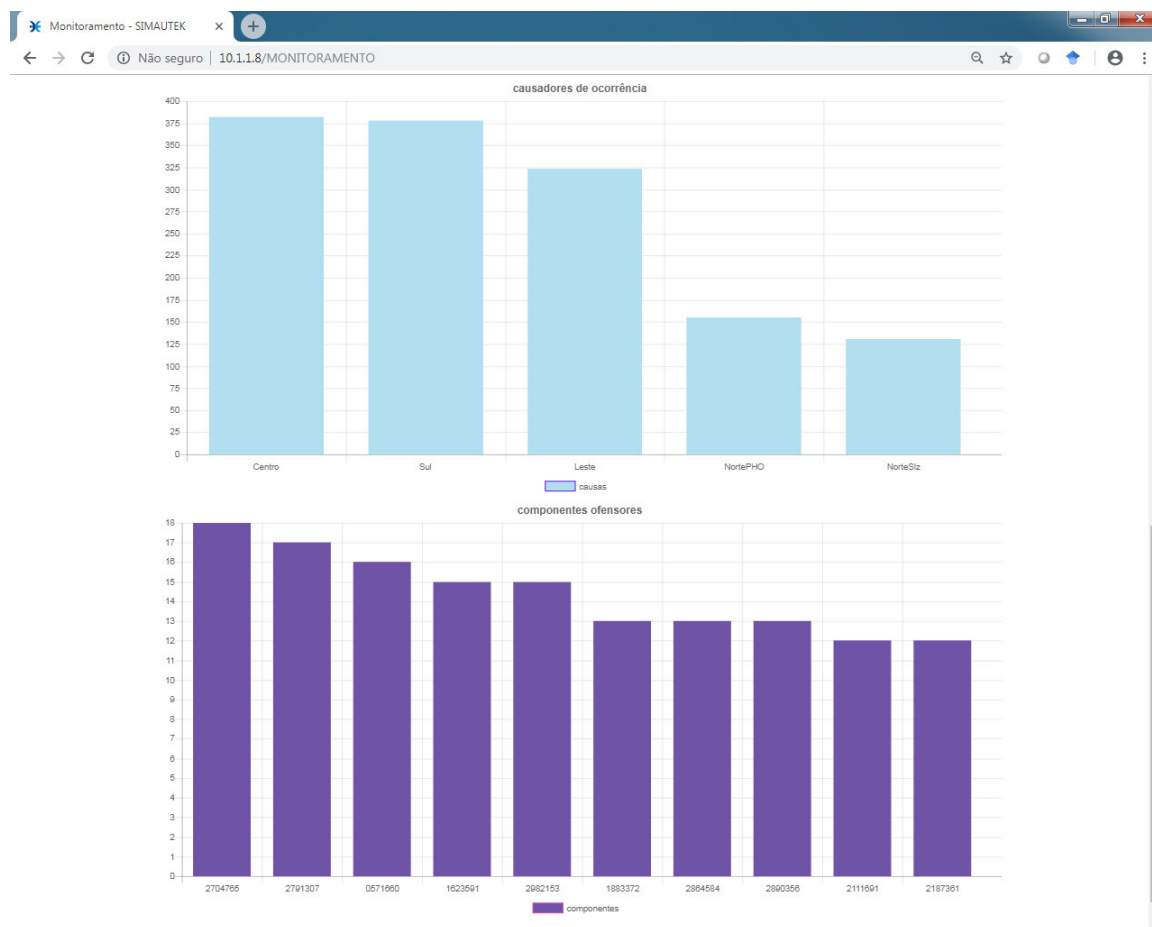
DEC: 0.7685484699941533 **FEC:** 1.0764631374667457 **Tempo:** 8173.679999999985 **Clientes:** 1914030

Dim: 2946.6300000000002

somatório dec somatório fec

Fonte: Autoria própria.

Figura 32 – Módulo de acompanhamento da taxa de falha anual para os blocos da rede de distribuição.



Fonte: Autoria própria.

Ainda na análise dos resultados, ao final desta metodologia será possível propor ações de investimentos sobre os blocos em tecnologias como:

- **Religadores:** Redução do impacto de faltas temporárias no sistema, tendo ação na melhoria do índice de confiabilidade do bloco;
- **Indicadores de Falta:** Redução do tempo de deslocamento das equipes de campo, tendo ação nas compensações DIC e DMIC;
- **Inclusão de Comunicação em Religadores:** Redução do tempo de preparação e possibilidade de redução do impacto de faltas em caso de existência de trecho adjacente redundante com transferência simples (Isolamento e Recomposição Remota);
- **Loop Automation:** Redução do impacto de faltas em caso de existência de trecho adjacente redundante, simplicidade de isolamento/recomposição e dificuldade de comunicação com o centro de controle;

- ***Self-healing***: Redução do tempo de execução de manobras complexas, tendo impacto em faltas de grandes proporções em caso de existência de trechos adjacentes, comunicação com religadores e manobras de transferência complexas (mais de uma alternativa ou trechos com restrições de transferência);
- **Medidores Eletrônicos com indicação *Last Gasp***: Redução do tempo de identificação do impacto de faltas, possibilitando a redução do tempo de preparação ao atendimento.

Considerando que a atual alocação de tecnologias voltadas para o incremento de confiabilidade foi realizada com base no uso de ferramentas de escopo limitado, ou até mesmo com base no conhecimento empírico de técnicos experientes, e que existe pouca reavaliação do desempenho de tais componentes e suas posições topológicas, a metodologia do presente trabalho se mostrou interessante pois apresentou uma proposta que permite definir ações de investimento baseada em dados utilizando critérios da rede de distribuição e das vias de acesso.

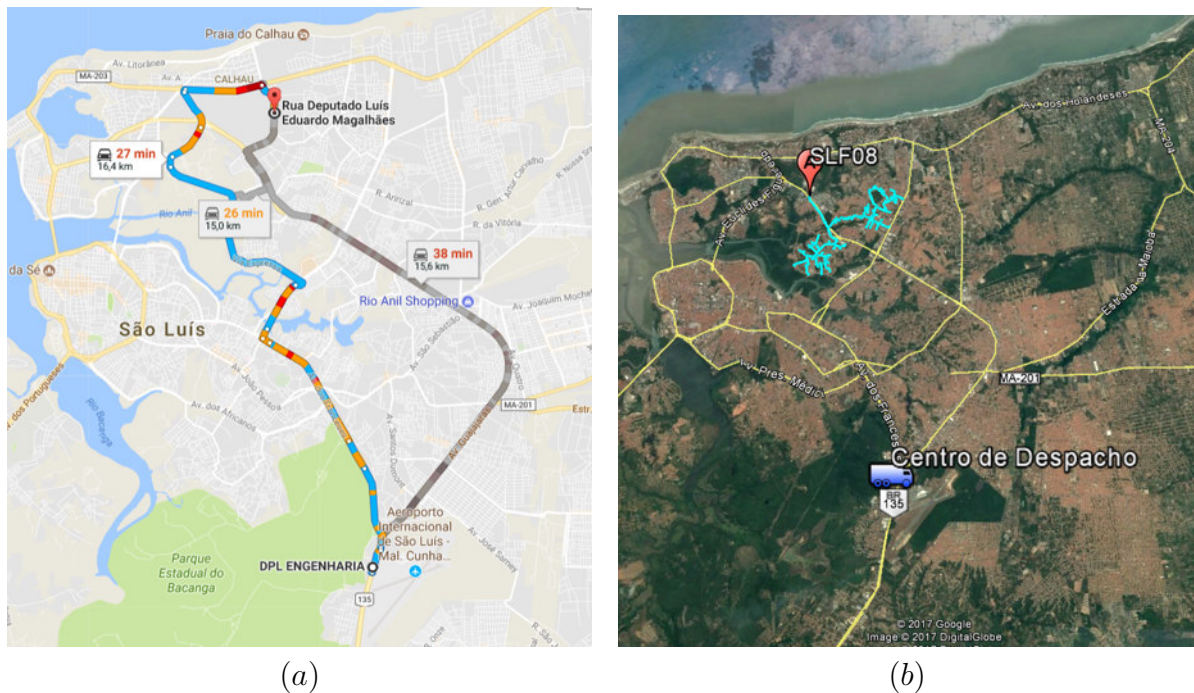
6 Resultado e Discussões

Como forma de avaliar os resultados obtidos através da metodologia proposta, foi analisado o grupo de *clusters* discutidos na [seção 5.5](#) deste trabalho. Os perfis comparados foram os *cluster 1* e *cluster 4*, que apresentaram custos diferentes apesar de características tecnológicas e de acessibilidade semelhantes.

Para cada *cluster* comparado foi selecionado um bloco para avaliação dos custos da falta e suas características. Para representar o *cluster 1* foi selecionado o bloco ARG-01C4 (Alimentador Araçagy 01C4) e para representar o *cluster 4* foi selecionado o bloco SLF-01C8 (Alimentador São Francisco 01C8).

O bloco SLF-01C8 teve seu tempo de deslocamento simulado de 30 minutos, conforme ilustrado na [Figura 33](#).

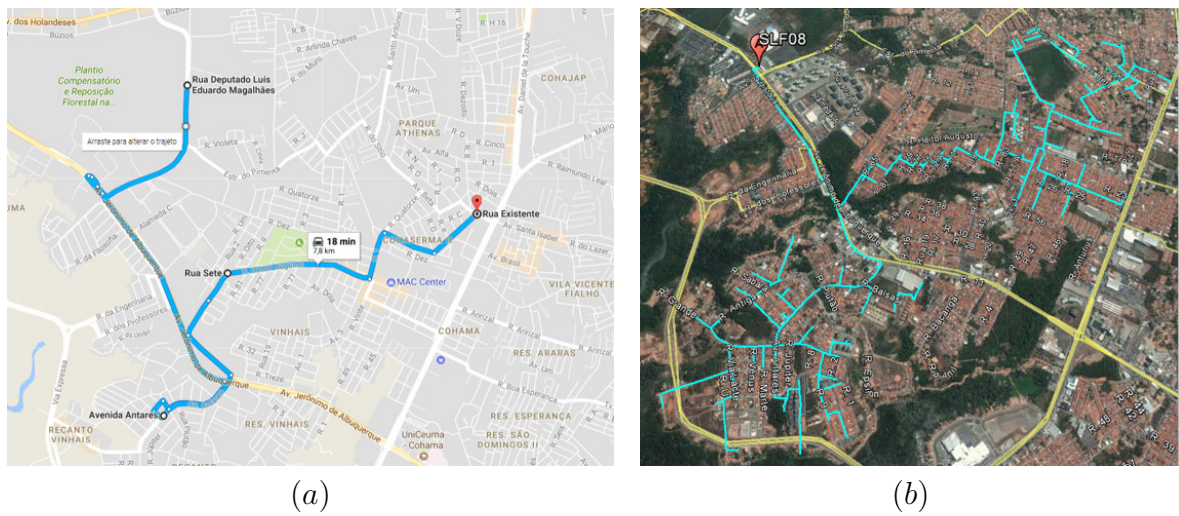
Figura 33 – (a) Tempo de Deslocamento Ocorrência SLF-01C8 e (b) Marcadores origem/destino Ocorrência SLF-01C8.



Fonte: Autoria própria.

Analisando a ocorrência de falta neste bloco, observa-se que tem um impacto médio em 2889 consumidores sem energia e que resulta em um CENS de 390 R\$/kWh. O tempo de inspeção para este bloco é de aproximadamente 18 minutos, conforme simulação da [Figura 34](#).

Figura 34 – (a) Tempo de Inspeção Ocorrência SLF-01C8 e (b) Camada Alimentador SLF-01C8.



Fonte: Autoria própria.

De forma análoga à análise realizada no bloco SLF-01C8, foi realizada a mesma análise para o bloco selecionado do *cluster 4*, o componente ARG-01C4. Através da simulação, este bloco teve um tempo estimado de deslocamento de 35 minutos, conforme pode ser observado na [Figura 35](#).

Figura 35 – (a) Tempo de Deslocamento Ocorrência ARG-01C4 e (b) Marcadores origem/destino Ocorrência ARG-01C4.

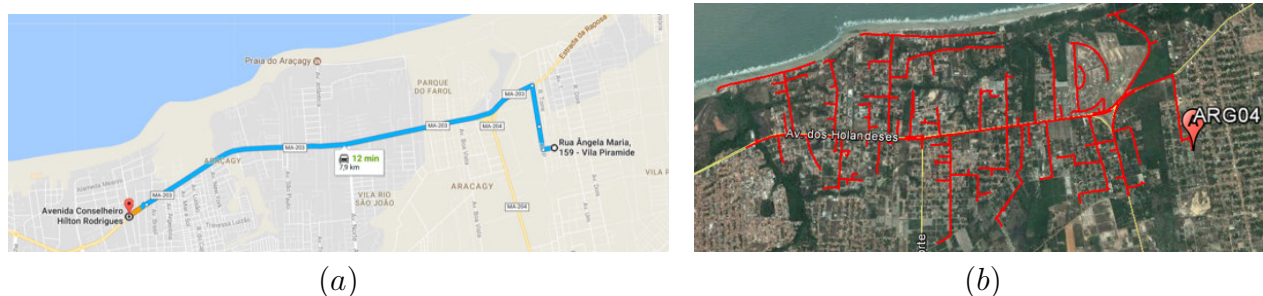


Fonte: Autoria própria.

Analisando a ocorrência de falta do bloco ARG-01C4, observa-se que possui um

impacto médio em 3889 consumidores e que o CENS é de 87 R\$/kWh. O tempo de inspeção médio simulado é de de aproximadamente 12 minutos, conforme simulação da [Figura 36](#).

Figura 36 – (a) Tempo de Inspeção Ocorrência ARG-01C4 e (b) Camada Alimentador ARG-01C4



Fonte: Autoria própria.

A [Tabela 22](#) apresenta de forma resumida uma comparação entre os blocos ARG-01C4 e SLF-01C8 onde é constatado a semelhança física dos blocos que apresentam o mesmo dispositivo de proteção, os tempos de acesso são semelhantes, porém cabe ainda entender melhor o motivo de o bloco ARG01C4, mesmo tendo impactado um número maior de consumidores e levando mais tempo para eliminar o defeito, teve um Custo de Energia Não Suprida menor que o bloco SLF01C8.

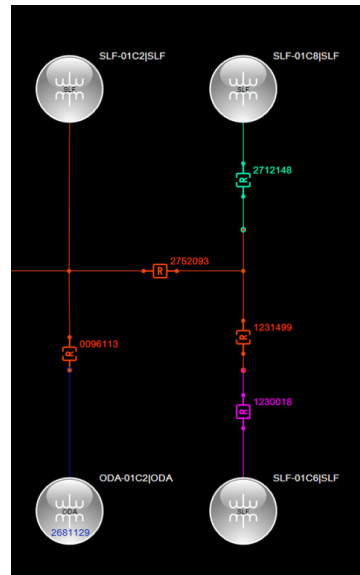
Tabela 22 – Comparação de Custos.

NUM_ COMPONENTE	TIPO_ EQUIP	AUTOMAÇÃO	TEMPO_ DESLOCAMENTO_ SIMULADO	TEMPO_ INSPEÇÃO_ SIMULADO	CLI	CENS_\$_Wh	TEMPO_ TOTAL
SLF-01C8	Disjuntor	SIM	30 minutos	18 minutos	2889	389.5166955	01:00:37
ARG-01C4	Disjuntor	SIM	35 minutos	12 minuto	3889	86.99801687	03:10:39

Fonte: Autoria própria.

Analisando os diagramas unifilares dos circuitos observa-se que o bloco SLF-01C8 não possui recursos para transferência de carga por telecontrole, o que acarreta em tempos de maiores para realização das manobras de isolamento e recomposição dos blocos saudáveis. Este bloco possui apenas uma Chave telecontrolada normalmente aberta com o bloco SLF-01C2, conforme apresentado na [Figura 37](#).

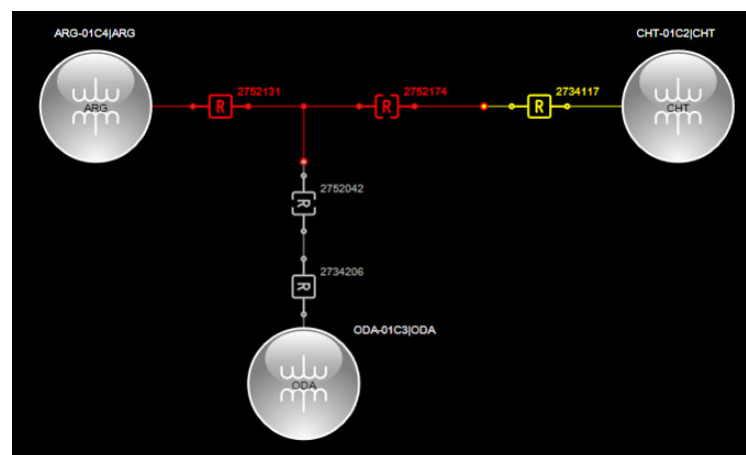
Figura 37 – Diagrama unifilar SLF-01C8.



Fonte: Autoria própria.

Aplicando a mesma análise para o bloco ARG-01C4 observa-se que este possui recurso de transferência de carga por telecontrole através de um Religador normalmente fechado com capacidade de isolar falhas e duas chaves normalmente abertas que permitem a transferência de cargas para blocos CHT-01C2 E ODA-01C3, adjacentes. A [Figura 38](#) ilustra o diagrama unifilar do bloco ARG-01C4.

Figura 38 – Diagrama unifilar ARG-01C4.



Fonte: Autoria própria.

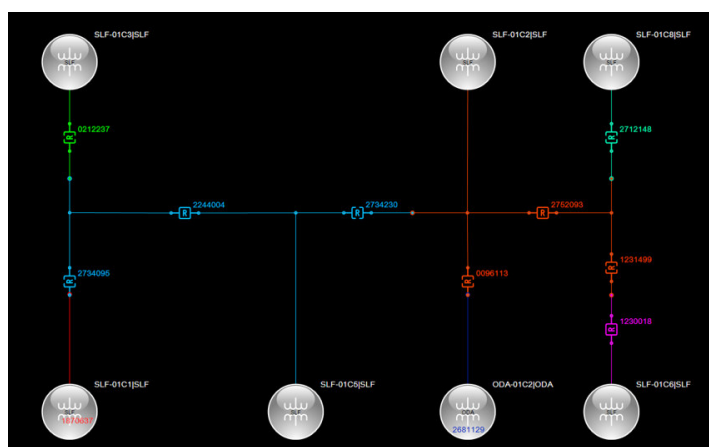
Logo, podemos concluir que para o perfil do *cluster 1* uma possível ação para mitigar o custo das faltas seria a aplicação de religadores com capacidade de telecomando

para Isolamento e Recomposição Remota, reduzindo assim os tempos de atendimento.

Ainda seguido a análise e como forma de obter uma solução de menor custo para o problema identificado no *cluster 1*, é possível agrupar blocos com equipamentos subutilizados na rede de distribuição e então priorizar ações de realocação de equipamentos em detrimento da aquisição de novos equipamentos.

Como exemplo, o diagrama esquemático da [Figura 39](#) apresenta uma sobra de equipamentos telecontrolados em SLF01C2, com provável subutilização, pois não permitem transferência de carga de forma remota (Falta equipamento normalmente fechado entre a subestação e a chave normalmente aberta). Algum desses equipamentos seria mais útil sendo utilizado como normalmente fechado, por exemplo, no alimentador SLF01C8.

Figura 39 – Diagrama Unifilar SLF



Fonte: autoria própria

Observa-se também que este tipo de situação é comum pois a rede de distribuição está em constante mudança, e com o passar do tempo, o planejamento inicial deixa de ser verdadeiro. A quantidade de dados a serem analisados e as inúmeras possibilidades (Topologia da rede x Equipamentos x Taxa de falhas x Condições de tráfego) tornam esta tarefa extremamente complexa e consumidora de homem-hora de engenharia.

As principais contribuições desse trabalho podem ser sintetizadas como: apresentar conceitos de *smart grid* associados ao desenvolvimento de futuras aplicações computacionais de planejamento, operação e manutenção; apresentar conceitos e procedimentos para realizar integração de bases de dados heterogêneas para criar aplicações utilizando os conceitos *KDD*, aprendizado não supervisionado e sistemas de suporte a decisão; discutir as correlações que aplicações gerenciais baseadas em bases de dados e aplicações de gestão de operação da rede através de ferramentas de suporte a decisão espacial podem fornecer para o apoio na solução de processos de análise e mitigação de faltas na rede de distribuição de energia elétrica.

7 Conclusão

Este trabalho apresentou uma proposta de metodologia capaz de analisar de maneira simples, a agregação de várias variáveis e suas inter-relações no impacto financeiro de uma falta na rede, permitindo então a análise de ações de investimento capazes de mitigá-las de forma viável. A metodologia desenvolvida utiliza as tarefas de *KDD* para filtrar, contextualizar e extrair informações úteis dos diferentes sistemas pré-existentes na empresa, isto é, bases de dados que não foram projetadas com um fim específico e orientado para algum aplicativo computacional.

O conjunto de atributos utilizado originou-se de quatro tabelas principais: *GIS* (*Geographic Information System*), *OMS* (*Outage Management System*), *SCADA* (*Supervisory Control and Data Acquisition*) e *GOOGLE MAPS API*.

As empresas concessionárias de energia possuem enormes quantidades de informação armazenadas em seus bancos de dados, porém, muitas vezes, com dados faltosos, inconsistentes e com elevado grau de incerteza. Ainda assim, existe uma informação valiosa contida nesses bancos de dados, que pode ser melhor utilizada para estudos e análises após etapas de qualificação dos dados.

Dessa forma, a aplicação do processo de *KDD* pode ser útil nesta tarefa, possibilitando a qualificação das informações utilizadas para planejar e operar os sistemas. Neste trabalho, buscou-se a construção de uma base de dados que representasse os registros de desligamentos não-programados em sistemas de distribuição, a partir de um universo de aproximadamente 17.920 eventos. Para alcançar esse objetivo, foram aplicadas técnicas que envolvem o tratamento e transformação de dados. No final do processo de *KDD* aplicado nesta dissertação restaram 6.468 eventos de interrupção não programados com condições de utilização. Isso tornou a base de dados mais confiável e adequada para a posterior etapa de análise dos dados.

Concluiu-se também, neste estudo, que a *Weka* é uma ferramenta importante para a realização da etapa de *data mining*. A *Weka* realiza a importação dos conjuntos de dados a serem trabalhados, implementa e executa os algoritmos de *data mining* utilizando tarefa de agrupamento, de forma simples e flexível, mostrando-se uma ferramenta adequada para a metodologia de descoberta de conhecimento proposta.

Para avaliar os resultados obtidos a partir da metodologia foi utilizado o sistema de suporte a decisão espacial SIMAUTEK, uma plataforma *Web* utilizada para gestão da operação, automação e manutenção de equipamentos presentes na rede de distribuição e também com integração às ferramentas do *Google Maps API*.

Como forma de avaliar a pertinência da metodologia estudada, analisou-se de forma comparativa dois *clusters*. Nesta análise concluiu-se que os dois blocos analisados possuíam características de acessibilidade dos blocos semelhantes, porém divergências nos impactos sobre os custos da energia não distribuída. Foi constatado que o bloco com maior impacto no CENS não possui recurso de transferência de carga por telecontrole o que torna as manobras para isolar o trecho com defeito e recompor o trecho saudável mais demorada e conseqüentemente mais cara.

Constatou-se também que existem oportunidades de melhoria na rede de distribuição de energia apenas com a realocação otimizada dos equipamentos existentes sem necessariamente adquirir novos equipamentos.

Considerando que atualmente o estudo para alocação de dispositivos voltados para o incremento de confiabilidade é realizado com base no uso de ferramentas de escopo limitado, ou até mesmo com base no conhecimento empírico de técnicos experientes, e que existe pouca reavaliação do desempenho de tais componentes e suas posições topológicas, a implementação deste sistema em uma distribuidora de energia elétrica proporciona a possibilidade de criação de uma base de dados consistente sobre desligamentos não programados e, ao mesmo tempo, tornando possível a criação de ferramentas para análise e mitigação das causas, para auxiliar no direcionamento dos investimentos em manutenção e a gestão de recursos financeiros, bem como o acompanhamento do desempenho da operação do sistema.

7.1 Trabalhos Futuros

Esta dissertação explorou um tema cada vez mais importante para as empresas concessionárias de energia. Esta área de pesquisa relacionada a análise e mitigação de causas de desligamentos não programados em redes de distribuição tem um caráter multidisciplinar, envolvendo aspectos computacionais, como por exemplo banco de dados, aspectos estatísticos e probabilísticos, confiabilidade de sistemas elétricos, inteligência computacional, metaheurísticas e aspectos de planejamento, operação e manutenção de sistemas de distribuição.

Apresentam-se a seguir alguns assuntos de trabalhos futuros para pesquisa e melhoria da metodologia desenvolvida:

- Criação de um novo modelo para classificação de causas, estratificando e organizando faltas permanentes e faltas transitórias, já que todas têm impacto sobre os clientes;
- Aprofundamento da pesquisa sobre identificação de causas baseada em informações incompletas e de múltiplas bases de dados, utilizando novas técnicas de descoberta de conhecimento em base de dados, aprendizagem automática e mineração de dados;

-
- Integrar ao SIMAUTEK um módulo para simulação elétrica da rede, permitindo análise de fluxo de potência que permitirá identificar, para cada bloco de rede, os possíveis pontos de manobra de isolamento e transferência parcial de carga e os pontos de inspeção para caso ocorra uma falta no bloco;
 - Utilização de ferramentas de inteligência artificial para retroalimentar a plataforma e gerar conhecimento;
 - Desenvolvimento de metodologias que permitam estabelecer uma correlação entre os desligamentos e outras rotinas relacionadas ao processo da empresa, não somente manutenção e operação, como também suprimentos, TI e comercial.

Referências

- ALIANCE, A. *Sobre - Metodologia Ágil para Desenvolvimento de Software*. 2018. Disponível em: <<https://www.agilealliance.org/agile101/the-agile-manifesto>>. Citado na página 53.
- ANEEL. *Agência Nacional de Energia Elétrica*. 2018. Disponível em: <<http://www.aneel.org.br>>. Citado 7 vezes nas páginas 29, 34, 37, 38, 39, 40 e 41.
- ANEEL. *Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional, Módulo 8-Qualidade da Energia Elétrica*. 2018. Disponível em: <<http://www.aneel.gov.br/modulo-8>>. Citado 7 vezes nas páginas 17, 28, 29, 30, 31, 32 e 33.
- ANEEL, N. T. *Agência Nacional de Energia Elétrica*. 2016. Disponível em: <http://http://www.aneel.gov.br/aplicacoes/consulta_publica/documentos/Nota%20T%C3%A9cnica_0173_SRD.pdf>. Citado 4 vezes nas páginas 35, 36, 37 e 43.
- ARMSTRONG, M. P.; DENSHAM, P. J. Database organization strategies for spatial decision support systems. *International Journal of Geographical Information Systems*, Taylor & Francis, v. 4, n. 1, p. 3–20, 1990. Citado na página 21.
- ARNOTT, D.; PERVAN, G. A critical analysis of decision support systems research. *Journal of information technology*, Springer, v. 20, n. 2, p. 67–87, 2005. Citado na página 21.
- ARONOFF, S. *Geographic information systems: a management perspective*. Taylor & Francis, 1989. Citado na página 49.
- BAYINDIR, R. et al. Smart grid technologies and applications. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 66, p. 499–516, 2016. Citado na página 46.
- BERKHIN, P. A survey of clustering data mining techniques. In: *Grouping multidimensional data*. [S.l.]: Springer, 2006. p. 25–71. Citado 2 vezes nas páginas 62 e 63.
- BORLASE, S. *Smart grids: infrastructure, technology, and solutions*. [S.l.]: CRC press, 2016. Citado 4 vezes nas páginas 45, 46, 47 e 48.
- BRAIL, R. K.; KLOSTERMAN, R. E. *Planning support systems: Integrating geographic information systems, models, and visualization tools*. [S.l.]: ESRI, Inc., 2001. Citado 2 vezes nas páginas 21 e 49.
- CAMINHA, A. C. *Introdução à proteção dos sistemas elétricos*. [S.l.]: E. Blucher, 2004. Citado na página 17.
- CEREDA, P. R. M.; JOSE, J. Adaptive data mining: Preliminary studies. *IEEE Latin America Transactions*, IEEE, v. 12, n. 7, p. 1258–1270, 2014. Citado na página 57.
- CHEN, M.-S.; HAN, J.; YU, P. S. Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and data Engineering*, IEEE, v. 8, n. 6, p. 866–883, 1996. Citado na página 61.

- CIOS, K. J. et al. *Data mining: a knowledge discovery approach*. [S.l.]: Springer Science & Business Media, 2007. Citado na página 57.
- COLAK, I. et al. A survey on the critical issues in smart grid technologies. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 54, p. 396–405, 2016. Citado 2 vezes nas páginas 47 e 48.
- COLLABORATIVE, S. G. C. Smart grid economic and environmental benefits: a review and synthesis of research on smart grid benefits and costs. *Atlanta, GA*, 2013. Citado na página 18.
- COSTA, L. d. F.; MONTAGNOLI, C. Máquinas tomam decisões: Reconhecimento de padrões e mineração de dados. *Ciência Hoje. Rio de Janeiro*, v. 30, n. 176, p. 22–29, 2001. Citado 2 vezes nas páginas 59 e 68.
- DOE, U. Grid 2030: A national vision for electricity’s second 100 years. *US DOE Report*, 2003. Citado na página 44.
- DOLLEN, D. V. et al. Report to nist on the smart grid interoperability standards roadmap. *Electric Power Research Institute (EPRI) and National Institute of Standards and Technology*, 2009. Citado 4 vezes nas páginas 44, 46, 47 e 48.
- DUGAN, R. C. et al. *Electrical power systems quality*. [S.l.]: mcgraw-Hill New York, 1996. v. 2. Citado na página 16.
- DUTRA, J. et al. Redes elétricas inteligentes no brasil: Subsídios para um plano nacional de implantação. *Rio de Janeiro: Synergia Editora*, v. 2, 2013. Citado na página 16.
- ENERGY, D. U. D. O. *What is the Smart Grid? Office of Electricity Delivery and Energy Reliability*. 2018. Disponível em: <http://smartgrid.gov/the_smart_grid>. Citado na página 45.
- FANG, X. et al. Smart grid—the new and improved power grid: A survey. *IEEE communications surveys & tutorials*, IEEE, v. 14, n. 4, p. 944–980, 2012. Citado na página 16.
- FASULO, D. *An analysis of recent work on clustering algorithms*. [S.l.], 1999. Citado na página 63.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI magazine*, v. 17, n. 3, p. 37, 1996. Citado 3 vezes nas páginas 56, 57 e 61.
- FAYYAD, U.; UTHURUSAMY, R. Evolving data into mining solutions for insights. *Communications of the ACM*, ACM, v. 45, n. 8, p. 28–31, 2002. Citado na página 56.
- FERREIRA, G. D. *OTIMIZAÇÃO DA CONFIABILIDADE DE SISTEMAS DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA: Uma*. 2009. Citado na página 28.
- FERRET, R. et al. *Hierarquização de alimentadores para fins de manutenção utilizando análise multicriterial*. Tese (Doutorado) — Universidade Federal de Santa Maria, 2012. Citado na página 16.

- FILHO, J. M.; MAMEDE, D. R. *Proteção de sistemas elétricos de potência*. [S.l.]: Grupo Gen-LTC, 2011. Citado 3 vezes nas páginas 16, 17 e 18.
- GELLINGS, C. W. *The smart grid: enabling energy efficiency and demand response*. [S.l.]: The Fairmont Press, Inc., 2009. Citado 2 vezes nas páginas 45 e 46.
- GERAIS-CEMAR, C. E. D. M. Relatório de sustentabilidade 2017. Disponível em: <http://www.equatorialenergia.com.br>>. Acesso em novembro de 2018, 2017. Citado na página 37.
- GHEWARE, S.; KEJKAR, A.; TONDARE, S. Data mining: task, tools, techniques and applications. *International Journal of Advanced Research in Computer and Communication Engineering*, v. 3, n. 10, 2014. Citado na página 61.
- GIUDICI, P. *Applied data mining: statistical methods for business and industry*. [S.l.]: John Wiley & Sons, 2005. Citado na página 62.
- GÖNEN, T. *Electric power distribution system engineering*. [S.l.]: CRC Press LCC, 2015. Citado na página 28.
- GOOGLE. *Sobre - Google Maps*. 2018. Disponível em: <<https://www.google.com/maps/about/>>. Citado 2 vezes nas páginas 51 e 53.
- GRABMEIER, J.; RUDOLPH, A. Techniques of cluster algorithms in data mining. *Data Mining and knowledge discovery*, Springer, v. 6, n. 4, p. 303–360, 2002. Citado na página 61.
- HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.l.]: Elsevier, 2011. Citado 9 vezes nas páginas 59, 60, 61, 63, 64, 65, 66, 67 e 75.
- HOLMES, G.; DONKIN, A.; WITTEN, I. H. Weka: A machine learning workbench. In: IEEE. *Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on*. [S.l.], 1994. p. 357–361. Citado na página 70.
- HUXHOLD, W. E. et al. An introduction to urban geographic information systems. *OUP Catalogue*, Oxford University Press, 1991. Citado na página 49.
- HUXHOLD, W. E. et al. An introduction to urban geographic information systems. *OUP Catalogue*, Oxford University Press, 1991. Citado na página 49.
- JANKOWSKI, P. Integrating geographical information systems and multiple criteria decision-making methods. *International journal of geographical information systems*, Taylor & Francis, v. 9, n. 3, p. 251–273, 1995. Citado na página 49.
- KAGAN, N. et al. Redes elétricas inteligentes no brasil: análise de custos e benefícios de um plano nacional de implantação. *Rio de Janeiro: Sinergia: Abradee*, 2013. Citado na página 16.
- KAGAN, N.; OLIVEIRA, C. C. B. D.; ROBBA, E. J. *Introdução aos sistemas de distribuição de energia elétrica*. [S.l.]: Edgard Blücher, 2010. Citado 5 vezes nas páginas 16, 18, 24, 25 e 27.
- KAGAN, N.; ROBBA, E. J.; SCHMIDT, H. P. Estimacão de indicadores de qualidade da energia elétrica. *Editora Blucher*, 2009. Citado na página 16.

KINDERMANN, G. *Curto-Circuito*. [S.l.]: Do Autor Fpolis, 2010. Citado 2 vezes nas páginas 16 e 17.

KODINARIYA, T. M.; MAKWANA, P. R. Review on determining number of cluster in k-means clustering. *International Journal*, v. 1, n. 6, p. 90–95, 2013. Citado 2 vezes nas páginas 68 e 69.

LIANG, T. P.; TURBAN, E.; ARONSON, J. E. Decision support systems and intelligent systems. *Yogyakarta: Penerbit Andi*, 2005. Citado na página 49.

LOPES, C. H. P. Classificação de registros em banco de dados por evolução de regras de associação utilizando algoritmos genéticos. *Dissertação de Mestrado, Departamento de Engenharia Eletrica, PUC-Rio*, 1999. Citado na página 62.

LOPES, Y. et al. Smart grid e iec 61850: Novos desafios em redes e telecomunicações para o sistema elétrico. *XXX Simpósio Brasileiro de Telecomunicações*, 2012. Citado na página 46.

MACORATTI. *Sobre - MVC*. 2018. Disponível em: <http://www.macoratti.net/08/06/asp_mvc1.htm>. Citado na página 53.

MACQUEEN, J. et al. Some methods for classification and analysis of multivariate observations. In: OAKLAND, CA, USA. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. [S.l.], 1967. v. 1, n. 14, p. 281–297. Citado 2 vezes nas páginas 66 e 79.

MADAN, S.; SON, W.-K.; BOLLINGER, K. Applications of data mining for power systems. In: IEEE. *Electrical and Computer Engineering, 1997. Engineering Innovation: Voyage of Discovery. IEEE 1997 Canadian Conference on*. [S.l.], 1997. v. 2, p. 403–406. Citado na página 22.

MICROSOFT. *Sobre - ASP.NET*. 2018. Disponível em: <<https://www.asp.net/>>. Citado na página 53.

MORI, H. State-of-the-art overview on data mining in power systems. In: IEEE. *Power Systems Conference and Exposition, 2006. PSCE'06. 2006 IEEE PES*. [S.l.], 2006. p. 33–34. Citado na página 22.

MUSSOI, F. L.; TEIVE, R. C. A multistage value-based model for prioritization of distribution projects using a multiobjective genetic algorithm. *Journal of Control, Automation and Electrical Systems*, Springer, v. 24, n. 5, p. 623–637, 2013. Citado na página 20.

NASSAR, S. et al. Identificação de condições climáticas adversas através de redes bayesianas. In: *IX SEPOPE-Symposium of Specialists in Electrical Operational and Expansion Planning*. [S.l.: s.n.], 2004. p. 1–5. Citado na página 22.

OLARU, C.; GEURTS, P.; WEHENKEL, L. Data mining tools and application in power system engineering. In: TRONDHEIM, NORWAY. *Proceedings of the 13th Power System Computation Conference, PSCC99*. [S.l.], 1999. p. 324–330. Citado na página 21.

OLIVEIRA, M. L. d. et al. *Planejamento integrado da expansão de sistemas de distribuição de energia elétrica*. Tese (Doutorado), 2010. Citado na página 20.

- PENG, J.-T.; CHIEN, C.; TSENG, T. Rough set theory for data mining for fault diagnosis on distribution feeder. *IEE Proceedings-Generation, Transmission and Distribution*, IET, v. 151, n. 6, p. 689–697, 2004. Citado na página 22.
- PRETTO, C. O. et al. Utilização de computação móvel para qualificação de rotinas de operação e manutenção de redes de distribuição. *Sba: Controle & Automação Sociedade Brasileira de Automatica*, SciELO Brasil, v. 17, n. 4, p. 446–458, 2006. Citado na página 76.
- SANTOS, L.; COUTINHO-RODRIGUES, J.; ANTUNES, C. H. A web spatial decision support system for vehicle routing using google maps. *Decision Support Systems*, Elsevier, v. 51, n. 1, p. 1–9, 2011. Citado na página 21.
- SATO, F.; FREITAS, W. Análise de curto-circuito e princípios de proteção em sistemas de energia elétrica-fundamentos e prática. *Rio de Janeiro: Campus*, 2015. Citado na página 16.
- SHORT, T. A. *Electric power distribution handbook*. [S.l.]: CRC press, 2014. Citado na página 26.
- SIDDIQUI, O.; HURTADO, P.; PARMENTER, K. *The Green Grid Energy Savings and Carbon Emissions Reductions Enabled by a Smart Grid*. [S.l.], 2008. Citado na página 44.
- SMARTGRIDS, E. *European technology platform smart grids: vision and strategy for Europe's electricity networks of the future*. [S.l.], 2006. Citado na página 44.
- SOUSA, J. d. Planejamento de sistemas de distribuição de energia elétrica através de um modelo de programação linear inteiro misto (plim). Universidade Estadual Paulista (UNESP), 2013. Citado na página 20.
- SUGUMARAN, V.; SUGUMARAN, R. Web-based spatial decision support systems (websdss): evolution, architecture, examples and challenges. *Communications of the Association for Information Systems*, v. 19, n. 1, p. 40, 2007. Citado 3 vezes nas páginas 49, 50 e 51.
- WITTEN, I. H.; FRANK, E.; HALL, M. Introduction to weka. *Data mining: practical machine learning tools and techniques*, v. 2, p. 365–368, 2011. Citado na página 70.

Apêndices

APÊNDICE B – Termo de consentimento

CEMAR		REQUISIÇÃO PARA TRABALHO ACADÊMICO	
DADOS DO SOLICITANTE			
NOME:	Roberto Arturo Quezada Sales		
CARGO:	Engenheiro de Distribuição PL	MATRÍCULA:	u10530
DIRETORIA:	Diretoria de Distribuição	GERÊNCIA:	GMAN
		RAMAL:	2305
DADOS ACADÊMICOS			
INSTITUIÇÃO DE ENSINO:	Universidade Federal do Maranhão		
ENDEREÇO:	Universidade Federal do Maranhão, CCET, bloco 2, apto 104		
COORDENADOR DO CURSO:	Shigeaki Leite de Lima		
FONE:	3272-9270		
TRABALHO ACADÊMICO			
TEMA:	METODOLOGIA METODOLOGIA BASEADA EM TÉCNICA DE DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS PARA ANÁLISE E MITIGAÇÃO DE FALTAS EM SISTEMAS DE DISTRIBUIÇÃO MODERNOS		
OBJETIVO:	<p>GERAL</p> <p><input type="checkbox"/> O objetivo geral da pesquisa conduzida neste trabalho é elaborar uma metodologia para análise e mitigação de faltas em um sistema de distribuição de energia elétrica a partir da aquisição de informações disponíveis em base de dados (Big Data), considerando aspectos do sistema de distribuição e dos recursos de manobra e telecontrole disponíveis. O método proposto busca desenvolver um sistema capaz de analisar de forma rápida e eficiente, a agregação de várias variáveis e suas inter-relações no impacto financeiro de uma falta na rede, permitindo então a análise de ações de investimento capazes de mitigá-las de forma a otimizar os indicadores de continuidade e maximizar os benefícios em relação à confiabilidade do sistema.</p> <p>ESPECÍFICOS</p> <p><input type="checkbox"/> Elencar variáveis significativas para os blocos de rede, formando um vetor multidimensional que leva em consideração características elétricas, lógicas e de acessibilidade (critério multivariável).</p> <p><input type="checkbox"/> Compor uma modelagem heterogênea de dados da rede de distribuição baseada em técnica de descoberta de conhecimento em base de dados.</p> <p><input type="checkbox"/> Identificar organizações não evidentes nos dados (similaridades físicas/diferenças tecnológicas) através da criação de agrupamentos consistentes.</p> <p><input type="checkbox"/> Proposição e implementação de um sistema especialista que, a partir dos dados extraídos, esteja apto a estimar de forma rápida e precisa um conjunto de ações de investimento e o peso das ações realizadas sobre a confiabilidade e seus custos associados</p> <p><input type="checkbox"/> Realizar estudo de caso em rede real.</p>		
INFORMAÇÕES NECESSÁRIAS:	<p>1) Topologia da Rede;</p> <p>2) Informações cadastrais: quantidade de clientes/demanda</p> <p>3) Indicadores de Qualidade Individuais/Coletivos (Disponível: www.aneel.gov.br;)</p> <p>4) Taxas de falha permanente e temporária;</p> <p>5) Tempo médio de reparo</p> <p>6) Tempo médio de chaveamento</p> <p>*OBS.: O estudo de caso utilizará nome fantasia para circuitos como forma de preservar as informações da empresa.</p>		

João Marcos Fernandes Mendonça
 Eng. Automação TIT
 Gerência de Expansão e Automação
 CREA: 10384 DMA

Ronnie Santiago Loureir
 Eng. Automação
 CREA: PE030827D

Antônio Simões Filho
 Executivo de Manutenção de Redes
 Mat. 7422

Felipe Vianelli
 Gerência de Operações do
 Sistema Elétrico - CEMAR
 Mat. 10718

Solicito a direção da CEMAR, as informações acima, como objetivo de divulgação e uso exclusivo na composição e construção do meu trabalho acadêmico. Estou ciente que não podem ser informados/utilizados dados confidenciais, apenas dados públicos e concordo em entregar uma cópia do trabalho concluído a CEMAR- Gente e Gestão e, ainda, autorizo a CEMAR a utilizar o conteúdo como fonte de referência e/ou divulgação, no todo ou em parte.

São Luís, 24/07/18

Roberto Sales
 Colaborador/Estagiário

Diretoria de Gente, Gestão e Finanças

Sérgio Souto Mello
 Diretor de Operações Técnicas
 e Comerciais

Ytaquirãe Soeiro
 Gerência de Gente e Gestão
 Companhia Energética do Maranhão-CEMAR