

UNIVERSIDADE FEDERAL DE VIÇOSA  
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

GUILHERME VESCOVI NICCHIO

**UTILIZAÇÃO DE ALGORÍTIMO IMUNOLÓGICO ADAPTATIVO  
PARA ALOCAÇÃO DE BASES DE RESTABELECIMENTO DE  
ENERGIA ELÉTRICA**

VIÇOSA

2018

GUILHERME VESCOVI NICCHIO

**UTILIZAÇÃO DE ALGORÍTIMO IMUNOLÓGICO ADAPTATIVO  
PARA ALOCAÇÃO DE BASES DE RESTABELECIMENTO DE  
ENERGIA ELÉTRICA**

Monografia apresentada ao Departamento de Engenharia Elétrica do Centro de Ciências Exatas e Tecnológicas da Universidade Federal de Viçosa, para a obtenção dos créditos da disciplina ELT 490 – Monografia e Seminário e cumprimento do requisito parcial para obtenção do grau de Bacharel em Engenharia Elétrica.  
Orientador: Prof. Dr. José Tarcísio de Resende.

VIÇOSA  
2018



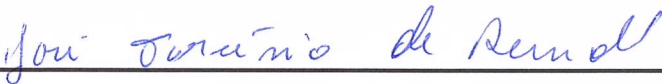
**GUILHERME VESCOVI NICCHIO**

**UTILIZAÇÃO DE ALGORÍTIMO IMUNOLÓGICO ADAPTATIVO  
PARA ALOCAÇÃO DE BASES DE RESTABELECIMENTO DE  
ENERGIA ELÉTRICA**

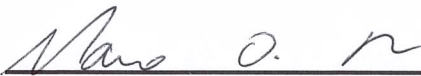
Monografia apresentada ao Departamento de Engenharia Elétrica do Centro de Ciências Exatas e Tecnológicas da Universidade Federal de Viçosa, para a obtenção dos créditos da disciplina ELT 490 – Monografia e Seminário e cumprimento do requisito parcial para obtenção do grau de Bacharel em Engenharia Elétrica.

Aprovada em 04 de Junho de 2018.

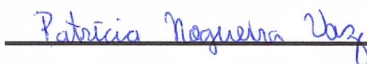
**COMISSÃO EXAMINADORA**



**Prof. Dr. José Tarcísio de Resende - Orientador**  
Universidade Federal de Viçosa



**Prof. Dr. Mauro de Oliveira Prates - Membro**  
Universidade Federal de Viçosa



**Engenheira Eletricista Patricia Nogueira Vaz - Membro**  
Universidade Federal de Viçosa

*“... from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.” – Charles Darwin*

## *Agradecimentos*

Agradeço a meus familiares pelo apoio inabalável que me deram todos estes anos e me prepararam para os desafios da vida. Sem vocês eu não seria nada.

Agradeço à minha namorada Joyce pelo companheirismo e motivação para encarar e superar os desafios.

Agradeço aos companheiros de trabalho Otávio Torres, Leonardo Ávila, Carlos Henrique Silva e demais colegas da Elektro pela ajuda para desenvolver este trabalho.

Obrigado Professor José Tarcísio de Resende pelo apoio e orientação todos estes anos de graduação.

Ao povo e governo brasileiro por me dar a oportunidade de realizar intercâmbio na Alemanha.

À Universidade Federal de Viçosa, por me ensinar minha profissão.

À República Interditados por ser eternamente a minha casa.

## *Resumo*

NICCHIO, Guilherme Vescovi. Monografia de Graduação, Universidade Federal de Viçosa, junho de 2018. **Utilização de algoritmo imunológico adaptativo para alocação de vases de restabelecimento de energia elétrica.** Orientador: José Tarcísio de Resende.

Atualmente o mundo vive uma revolução tecnológica em que os sistemas de informação integrados possibilitam a geração de um grande volume de dados. A análise inteligente de dados tornou-se importante para as empresas melhorarem seus processos, podendo assim enxergar correlações entre dados que até então estavam escondidas.

Para a análise deste grande volume de dados, foram desenvolvidas nas últimas décadas poderosas ferramentas computacionais envolvendo conceitos de estatística e programação para tratar estas informações. Dentre algumas destas técnicas destaca-se a clusterização, que consiste em agrupar dados de acordo com sua similaridade.

O setor elétrico brasileiro opera sobre um modelo de concessão que incentiva as empresas de distribuição de energia a buscarem eficiência em sua operação. Estas companhias são remuneradas pela melhoria de indicadores de qualidade, os quais são impactados pelo tempo levado para se reestabelecer o fornecimento de energia elétrica interrompido.

Objetiva-se através deste trabalho a avaliação do algoritmo de clusterização imuno-adaptativo ARIA para alocação otimizada de bases de restabelecimento de energia utilizando a localização das ocorrências emergenciais no estado de Pernambuco.

Este trabalho foi desenvolvido na empresa CELPE durante a revisão dos processos desta concessionária. Foi utilizado a ferramenta *Matlab*, desenvolvida pela empresa *Mathworks*, para desenvolver e executar o código de clusterização.

Como resultado, conclui-se que algoritmos de clusterização são úteis para abordar problemáticas de empresas com dispersão geográfica em sua operação. A utilização do algoritmo ARIA foi eficiente para a problemática de bases de restabelecimento de energia, no entanto ele não basta para definir a localização exata da base, sendo necessárias outras ferramentas para validação da localização proposta pelo algoritmo.

## *Abstract*

Currently the world is experiencing a technological revolution in which integrated information systems enable the generation of a large amount of data. Intelligent data analysis has become important for companies to improve their processes so they can see correlations between previously hidden data.

For the analysis of this large volume of data, in the last decades powerful computational tools were developed involving statistics and programming concepts to work with this information. Some of these techniques include clustering, which consists of grouping data according to their similarity.

The Brazilian electricity sector operates on a concession model that encourages energy distribution companies to seek efficiency in their operation. These companies are remunerated for the improvement of quality indicators, which are impacted by the time taken to reestablish the interrupted electricity supply.

The objective of this work is to evaluate the ARIA algorithm for optimum allocation of energy restoration bases using the location of emergency occurrences in the state of Pernambuco.

This work was developed in the company CELPE during the review of the processes of this concessionaire. The Matlab tool, developed by Mathworks, was used to develop and execute the clustering code.

As a result, it is concluded that clustering algorithms are useful to address problems of companies with geographic dispersion in their operation. However, it is not enough to define the exact location of the base, and other tools are necessary to validate the location proposed by the algorithm.



# *Sumário*

<b>1</b>	<b><i>Introdução</i></b> .....	<b>1</b>
1.1.	SETOR ELÉTRICO BRASILEIRO .....	1
1.2	AGENTES DO SETOR .....	1
1.3	A DISTRIBUIÇÃO DE ENERGIA ELÉTRICA .....	3
1.4	REMUNERAÇÃO DA DISTRIBUIÇÃO DE ENERGIA ELÉTRICA .....	3
1.4.1	A COMPOSIÇÃO TARIFÁRIA .....	4
1.5	QUALIDADE DO FORNECIMENTO DE ENERGIA .....	5
1.6	ASPECTOS OPERACIONAIS .....	6
1.6.1	ESTRUTURA DE ATENDIMENTO .....	7
1.7	MOTIVAÇÃO .....	8
1.8	OBJETIVO .....	9
<b>2</b>	<b><i>Metodologia</i></b> .....	<b>10</b>
2.1.	CLUSTERIZAÇÃO .....	10
2.2	ALGORITMOS IMUNO-INSPIRADOS .....	12
<b>3</b>	<b><i>Resultados</i></b> .....	<b>20</b>
<b>4</b>	<b><i>Conclusão</i></b> .....	<b>25</b>
	<b><i>Referências Bibliográficas</i></b> .....	<b>26</b>

## *Lista de Figuras*

Figura 1. Agentes do Setor Elétrico Brasileiro (ABRADEE, 2018). .....	2
Figura 2. Clusterização k-means (k=3).....	11
Figura 3. Clusterização K-means (k = 6).....	11
Figura 4. Árvore de clusters (modificado de Brian, et. al., 2011). .....	12
Figura 5. Exemplo de clusterização com ARIA. (Von Zuben, s. d.).....	14
Figura 6. Exemplo de Clusterização do ARIA. (a) Início das iterações; (b) Clusterização em andamento; (c) Final do processo. ....	18

## *Lista de Tabelas*

Tabela 1. Variáveis do pseudocódigo.....	16
Tabela 2. Exemplo da estrutura de planilha.....	20

# ***1 Introdução***

## **1.1. SETOR ELÉTRICO BRASILEIRO**

Na década de 1990 o setor elétrico brasileiro foi submetido a uma reestruturação chamada RESEB (Projeto de Reestruturação do Setor Elétrico Brasileiro). Neste novo modelo o Estado passou a atuar como “regulador”, estabelecendo normas para guiar o desenvolvimento do setor. Neste processo diversas empresas públicas do setor elétrico foram privatizadas e alguns órgãos públicos foram criados, como por exemplo, a agência reguladora ANEEL (Agência Nacional de Energia Elétrica). Apesar disso em 2001, após alguns problemas na oferta de energia que culminaram em grandes racionamentos, foram formulados novos ajustes com intuito de reduzir riscos de falta de energia (MME, Ministério de Minas e Energia. RESEB, 2011).

Apesar das profundas mudanças ao longo da história do sistema elétrico brasileiro, estas duas últimas constituem a base do sistema vigente atualmente. De forma sintética o modelo é caracterizado pela possibilidade das concessionárias de distribuição públicas e privado coexistirem; planejamento da operação centralizado em uma sede; tarifas diferentes para diferentes concessões; possibilidade de clientes livres cativos; leilões de energia regulados para consolidação de contratos com as concessionárias que fornecem energia a clientes cativos; serviços de comercialização, transmissão, geração e distribuição devem ser operados separadamente (ABRADEE, 2018).

## **1.2 AGENTES DO SETOR**

A ANEEL é o mais importante órgão do setor, responsável por estabelecer regras e definir as interações entre as entidades que o constituem. A Câmara de Comercialização de Energia Elétrica (CCEE) é o órgão encarregado da administração dos contratos firmados entre as partes envolvidas na comercialização de energia, como também por definir o valor da energia no mercado livre. O Operador Nacional do Sistema (ONS) é incumbido pela operação do sistema elétrico Integrado (SIN), assegurando a estabilidade e eficiência na geração e transmissão de energia (ABRADEE, 2018).

Outros órgãos governamentais que constituem o setor são responsáveis por definir diretrizes e estratégias de longo prazo, citando o Ministério de Minas e Energia (MME), o Comitê de Monitoramento do Setor Elétrico (CMSE), o Conselho Nacional de Política Energética (CNPE) e a Empresa de Pesquisa Energética (EPE). Na base do sistema encontram-se os agentes de geração, transmissão, distribuição e comercialização, representados pelas companhias públicas e privadas que administram suas concessões através de diretrizes determinadas pelos órgãos reguladores, garantindo eficiência e qualidade no setor (ABRADEE, 2018).

Na Figura 1 é ilustrado o fluxograma dos órgãos que constituem o setor elétrico brasileiro.

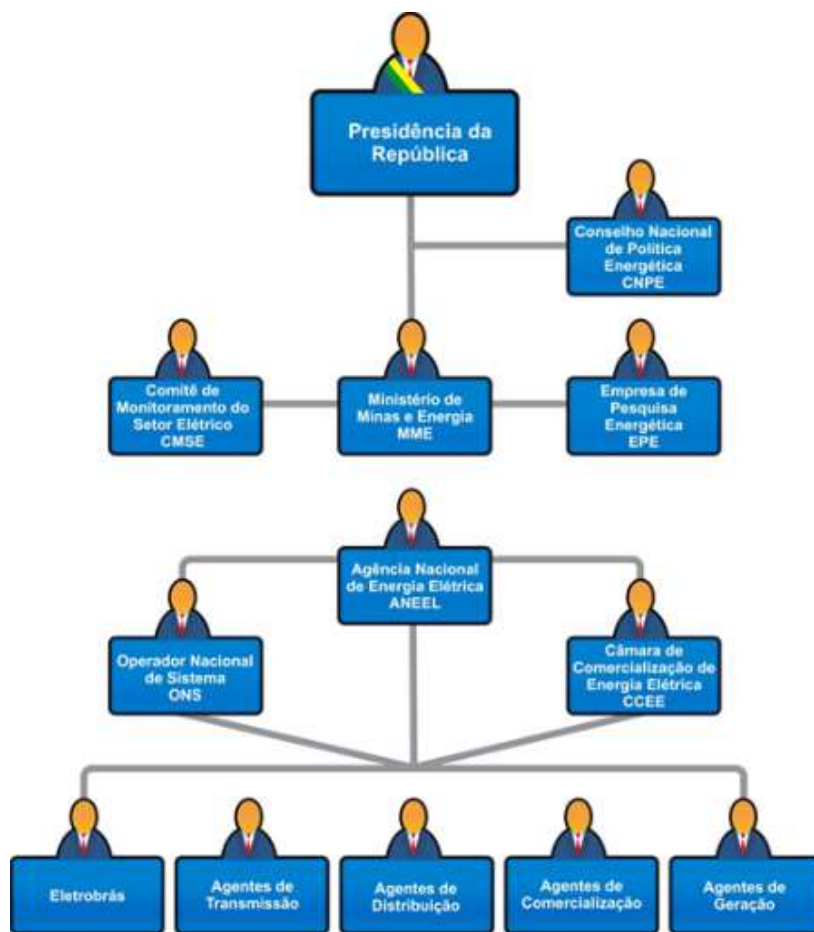


Figura 1. Agentes do Setor Elétrico Brasileiro (ABRADEE, 2018).

### 1.3 A DISTRIBUIÇÃO DE ENERGIA ELÉTRICA

A distribuição de energia no Brasil é realizada por empresas particulares e públicas, muitas delas compõem grupos que operam em diversos estados e áreas de concessão, buscando eficiência no processo e na operação (ABRADEE, 2018).

As áreas de concessão consistem em uma região delimitada geograficamente, em que os ativos presentes (subestações, redes de média e baixa tensão, etc.) são administrados pelas empresas, promovendo manutenção e expansão do sistema elétrico, como também promovendo atendimento dos serviços de acordo com as normas determinadas pela ANEEL (ABRADEE, 2018).

É de responsabilidade das concessionárias de energia realizarem investimentos para garantir que o crescimento de demanda seja acompanhado pela expansão do sistema elétrico, não sobrecarregando o sistema. Este investimento deve seguir a regra de “investimento prudente” da ANEEL, que visa garantir que as empresas façam investimentos condizentes com a modicidade tarifária e que sejam viáveis economicamente (Fontana, 2015).

Para administrar a sua área de concessão as empresas possuem uma estrutura de bases distribuídas geograficamente e comumente uma ou duas sedes administrativas nas quais se concentram os serviços corporativos. A estrutura operacional é composta por bases operacionais, responsáveis pelo suporte à equipe de campo, com veículos, ferramental e toda a estrutura necessária para o sucesso na execução dos serviços. É nestas bases também que se concentram a estrutura de pessoal, responsável pela gestão dos processos e estoque (planejadores, encarregados, almoxarifes, engenheiros e supervisores). Por outro lado, a sede administrativa é composta por áreas que dão suporte à operação de campo como, por exemplo, engenharia, recursos humanos, tecnologia da informação, finanças, *supply chain*, comercial e atendimento ao cliente. É nela que se encontram os cargos executivos responsáveis pela gerência estratégica da empresa (Fontana, 2015).

### 1.4 REMUNERAÇÃO DA DISTRIBUIÇÃO DE ENERGIA ELÉTRICA

Nas reformas realizados no RESEB o setor elétrico brasileiro teve seu primeiro arranjo regulatório, o qual estruturou um modelo em que a atividade de distribuição de energia

elétrica é remunerada através de uma tarifa predefinida pela ANEEL a ser paga pelos consumidores (MME, 2011).

As tarifas determinadas só podem ser modificadas caso atendam algumas condições especificadas no módulo 2 do Procedimento de Regulação Tarifária (PRORET, 2011):

- Para corrigir os efeitos da inflação, é previsto um reajuste tarifário anual, cujo método de cálculo é predefinido nos contratos de concessão;
- Em intervalos periódicos é realizada uma revisão tarifária, podendo variar de 3 a 5 anos dependendo da concessão, sendo que nesta revisão são respeitados os critérios e as metodologias presentes nas resoluções da ANEEL;
- Caso algum evento tenha um impacto significativo no equilíbrio econômico-financeiro da concessão, pode gerar uma justificativa para uma revisão tarifária extraordinária.

#### **1.4.1 A COMPOSIÇÃO TARIFÁRIA**

A composição da tarifa referente às distribuidoras de energia no Brasil é dividida em duas parcelas, A e B. A parcela A corresponde aos gastos não gerenciáveis, aqueles que a concessionária não consegue atuar como, por exemplo, os valores do contrato de compra de energia elétrica, às conexões ao sistema de transmissão, encargos setoriais definidos pela legislação, entre outros que não dependem da gestão da distribuidora (PRORET, 2011).

Já a parcela B está relacionada à remuneração da distribuidora pelos investimentos feitos em ativos. Os ativos são subdivididos em Base de Remuneração Regulatória (BRR) e Base de Ativos Regulatória (BAR), os quais representam o capital empenhado na expansão e manutenção do sistema e na respectiva infraestrutura, nesta ordem. Em cima destas bases é calculada a depreciação dos ativos e a taxa de remuneração. A parcela B da tarifa correspondem aos custos gerenciáveis, os quais a empresa tem amplo controle (PRORET, 2011).

## 1.5 QUALIDADE DO FORNECIMENTO DE ENERGIA

Para estimular a busca pela qualidade na distribuição de energia elétrica o fornecimento é avaliado com base em indicadores e metas estabelecidos pela ANEEL nos Procedimentos de Distribuição (PRODIST, 2011). O PRODIST define indicadores técnicos como também formas de cálculo para as perdas do sistema elétrico, sejam técnicas ou não técnicas.

Os principais indicadores técnicos de qualidade são: a Duração Equivalente de Interrupção por Unidade Consumidora (DEC) e a Frequência Equivalente de Interrupção por Unidade Consumidora (FEC). O DEC indica a média de horas que o consumidor fica sem energia elétrica em um período, já o FEC indica quantas vezes ele ficou sem energia, ou seja, a frequência das interrupções (PRODIST, 2011).

Outros indicadores que abrangem a qualidade no fornecimento são o DIC, FIC e DMIC. O DIC (Duração de Interrupção por Unidade Consumidora) e seu semelhante FIC (Frequência de Interrupção por unidade Consumidora) são os paralelos do DEC e FEC, mas analisados para uma unidade consumidora específica. Já o DMIC é a Duração Máxima de Interrupção por Unidade Consumidora, que indica o limite máximo de tempo que uma unidade consumidora pode ficar sem energia, impedindo que a distribuidora demore a reestabelecer o suprimento de energia (PRODIST).

Os indicadores DEC e FEC variam entre diferentes tipos de consumidores, embora a tarifa seja única dentro de uma área de concessão. Isso deve-se ao fato de que a distribuidora atua em regiões com diferentes aspectos geográficos e estruturais. Por exemplo, alguns clientes se encontram em zonas rurais, outros em centros urbanos próximos aos centros de carga e quanto mais longe uma ocorrência estiver dos centros de carga, mais tempo a distribuidora levará para atender a emergência, sendo assim, também um serviço mais dispendioso (ABRADEE, 2018).

Caso a empresa ultrapasse os limites dos indicadores estabelecidos pela ANEEL o consumidor deve ser recompensado de acordo com uma fórmula que analisa a gravidade da transgressão. A fórmula consiste em multiplicar o tempo em que o limite foi ultrapassado pelo valor equivalente da hora do custo de distribuição e também pelo fator de compensação estabelecido para o consumidor residencial (ABRADEE, 2018).



Visando não transgredir estes indicadores, as empresas devem realizar investimentos para melhorar estes indicadores e, conseqüentemente, aumentar a compensação tarifária (PRODIST).

## 1.6 ASPECTOS OPERACIONAIS

Conforme comentado, a qualidade e eficiência na operação do sistema elétrico é fundamental para a distribuidora, a qual pode ser punida ou remunerada pelos seus resultados. Para melhor entender como a empresa pode melhorar seus indicadores, é necessário o entendimento dos serviços demandados na rede de distribuição de energia.

Os serviços de rede de distribuição compreendem atividades executadas nas redes de média e baixa tensão, ou seja, nos alimentadores primários (redes de média tensão) provenientes de subestações e estruturados em ramais radiais por zonas rurais e urbanas onde a tensão eventualmente é rebaixada para entrega de energia às unidades consumidores em baixa tensão. Além de toda a infraestrutura amplamente conhecida como, cabos e postes, existe também a presença de equipamentos especiais, como reguladores de tensão, religadores, indicadores de falta, entre outros, que demandam mão de obra especializada. Os serviços operacionais de rede podem ser divididos em três grandes grupos: os serviços técnicos comerciais, expansão e preservação de rede e a demanda gerada por acidentes e ocorrências em geral que levam à interrupção no fornecimento de energia (Fontana, 2015).

Os chamados Serviços Técnicos Comerciais (comumente referido como STC) abrangem atividades em clientes em âmbito comercial como, por exemplo, ligações de novos clientes, corte no fornecimento de energia, inspeção em medidores, troca de padrão, poda de árvores, entre outras atividades que são comumente realizadas em rede secundária. A maioria destas atividades são de baixa complexidade e reguladas, ou seja, seguem protocolo de operação definido pela ANEEL, tendo prazo de atendimento pré-definido, não podendo ser negligenciado com risco de prejuízo regulatório-financeiro impactando indicadores de satisfação do cliente, recuperação de energia e inadimplência (Fontana, 2015).

A preservação e expansão de redes representam atividades que visam incrementar a capacidade do sistema através de construções, promovendo o fornecimento de energia elétrica em áreas isoladas e melhoria da robustez da rede contra falhas. Estas atividades são

---

previamente planejadas para execução com a rede energizada ou desenergizada de acordo com o prazo estipulado pelo regulador (Fontana, 2015).

As atividades de atendimento de emergências abrangem as ocorrências que acarretam interrupção no fornecimento de energia e são o foco principal deste trabalho. Estas ocorrências têm grande impacto nos indicadores de qualidade apresentados anteriormente e podem impactar desde clientes isolados até grandes blocos de clientes como cidades inteiras ou até grandes regiões. Para contabilizá-las as distribuidoras contam com o auxílio de sistemas de informação capazes de levantar a quantidade e duração destes eventos, sendo assim uma valiosa base de dados para análises do desempenho e eficiência da operação, como também, para desenvolvimento de estudos estratégicos, prevendo futuras ocorrências e mapeando zonas de risco (Fontana, 2015).

### **1.6.1 ESTRUTURA DE ATENDIMENTO**

A estrutura de atendimento consiste em uma infraestrutura predial comumente chamada de base, composta pelo escritório, pátio, almoxarifados, depósito de materiais entre outras estruturas de suporte para as atividades realizadas em uma área geográfica.

No escritório se encontram os profissionais responsáveis pelo apoio à operação e administrativo, como técnicos, estagiários, engenheiros, supervisores. No entanto, estudos complexos de engenharia são localizados na sede corporativa citada anteriormente. O pátio abriga os veículos à disposição do serviço a ser realizado em campo, é a área utilizada para carga e descarga de materiais, montagem e teste de equipamentos, estoque de postes, transformadores, entre outras estruturas (Fontana, 2015).

A área de atuação geográfica da base é limitada à uma região específica abrangendo um número de municípios próximos. Para um atendimento otimizado estas bases devem estar localizadas próximos a maior concentração de ocorrências, garantindo assim agilidade no atendimento e redução nos custos associados ao deslocamento que vão desde a consumo de gasolina, desgaste de veículos, remuneração por tempo de trabalho, etc.

---

Visto isso, é notório que a revisão dos processos associados à realização dos serviços é de fundamental importância para a melhoria do desempenho, podendo uma alteração na localização destas bases trazer um grande impacto nos resultados da empresa.

## 1.7 MOTIVAÇÃO

As atividades citadas anteriormente podem ser realizadas por profissionais próprios da distribuidora ou por empresas terceirizadas. As concessionárias do setor elétrico brasileiro, em sua maioria, trabalham com um quadro misto destes dois tipos. Atualmente, motivados por problemas trabalhistas judiciais, baixa eficiência e insegurança dos serviços realizados por terceiros, as empresas estão revisando os processos e avaliando a primarização de algumas de suas atividades.

Em 2017 as concessionárias brasileiras pertencentes ao grupo Neoenergia COELBA, CELPE e COSERN, as quais atuam na totalidade dos estados da Bahia, Pernambuco e Rio Grande do Norte iniciaram um processo de unificação com a Elektro, distribuidora que atua em parte do estado de São Paulo. Este movimento é uma estratégia do grupo espanhol Iberdrola, acionista majoritário de todas estas, num processo que busca consolidação da empresa no setor elétrico brasileiro (Neoenergia, 2018). Visando compartilhamento de boas práticas entre as distribuidoras envolvidas, muitos dos processos destas empresas foram revisados, dentre eles, a estrutura operacional para atendimento das ocorrências de emergência.

Atualmente o mundo vive uma revolução tecnológica em que os sistemas de informação integrados possibilitam uma grande geração de dados. Diariamente são deixadas pegadas digitais seja ao pesquisar no Google, navegar pelo Facebook, assistindo um vídeo no Youtube, ou até mesmo ao utilizar um GPS para chegar a determinado local. Isso tudo é devidamente armazenado em um banco de dados, chamado de *Big Data* (McAfee et. al., 2012).

Para a análise deste grande volume de dados, foram desenvolvidas nas últimas décadas poderosas ferramentas computacionais. Envolvendo conceitos de estatística e programação para tratar estas informações, a análise inteligente de dados tornou-se parte das empresas,

utilizada para melhorar seus processos, podendo agora enxergar correlações que até então estavam escondidas (McAfee et. al., 2012).

Conforme comentado, os dados extraídos das ocorrências na rede elétrica através dos sistemas informatizados, formam uma valiosa base de dados para a empresa. Utilizando-se das ferramentas computacionais é possível mapear as ocorrências geograficamente e com base nisso definir onde deveriam se localizar as bases de atendimento para que o serviço seja o mais rápido e eficiente possível, o que vai de encontro às políticas estabelecidas pelo sistema de concessão brasileiro.

### **1.8 OBJETIVO**

Em face à revisão dos processos da distribuidora CELPE, objetiva-se através deste trabalho a avaliar o algoritmo de clusterização imuno-adaptativo ARIA para propor uma localização otimizada das bases de restabelecimento utilizando a base de dados referentes às ocorrências emergenciais no estado de Pernambuco.

## 2 Metodologia

A base dos dados relativos às ocorrências emergenciais de uma empresa ao longo de anos é um grande desafio para um cientista de dados, a partir dela são extraídas inúmeras informações como, por exemplo, zonas carentes de manutenção, regiões de crescimento de carga, sazonalidades de ocorrências geradas por eventos humanos, tais como copa do mundo e eleições ou mesmo não humanos como eventos climáticos.

### 2.1. CLUSTERIZAÇÃO

O primeiro passo para se analisar o grande volume de dados provenientes do sistema é encontrar similaridades entre as amostras para então agrupá-las através de um processo denominado clusterização (do inglês *clustering* – que significa a criação de grupos). Este processo utiliza de ferramentas estatísticas e computacionais para reconhecimento de padrões entre dados, até então escondidos aos olhos humanos (Estivill-Castro, 2002).

Existem inúmeras formas de se realizar a clusterização, todas elas com suas vantagens e desvantagens, não existindo uma melhor que a outra, apenas sendo uma mais apropriada para um problema específico. Algoritmos de clusterização podem ser classificados de acordo com sua abordagem na definição da similaridade entre as amostras, variando em centenas de tipos diferentes. A seguir são descritos métodos importantes.

Um dos mais importantes e populares tipos de clusterização é uma abordagem baseada no centroide dos dados no espaço, também conhecida como *k-means*. A ideia principal do algoritmo é definir os grupos de uma amostra baseado em uma quantidade predefinida de grupos ( $k$ ) que se deseja obter. (Berkhin, 2006). As figuras 2 e 3 mostram exemplos da clusterização por centroide com 3 e 6 grupos, respectivamente.

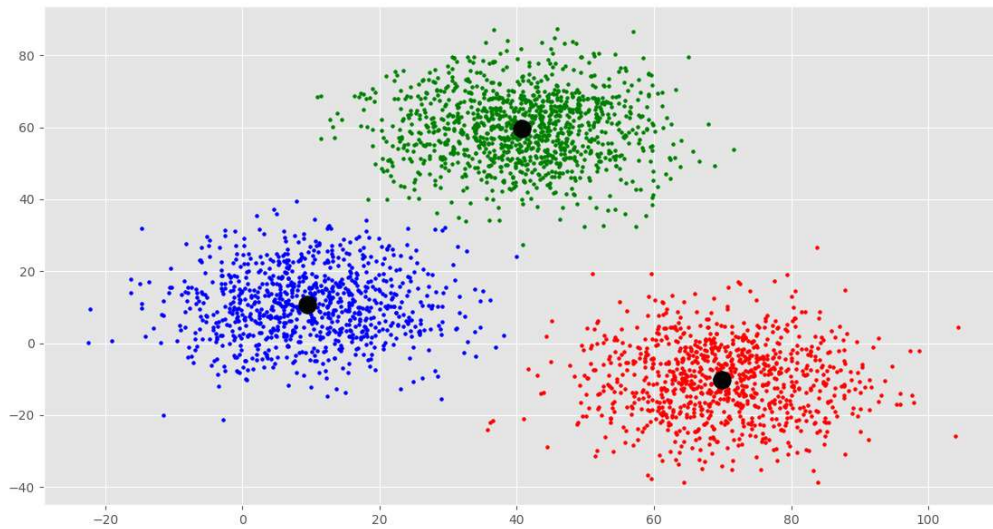


Figura 2. Clusterização k-means ( $k=3$ )

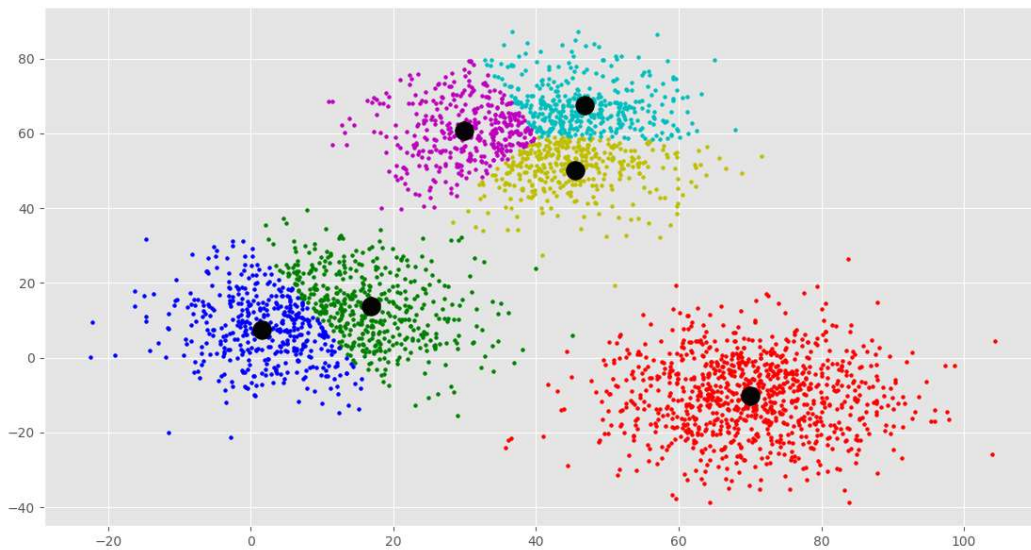


Figura 3. Clusterização K-means ( $k=6$ )

De forma resumida, os centroides, em preto, são posicionados inicialmente em coordenadas aleatórias e ao decorrer das iterações, estes se aproximam dos dados mais próximo até se posicionarem no centro dos dados e formarem a quantidade  $k$  de clusters (Berkhin, 2006).

Outro modelo muito utilizado é a clusterização hierárquica, que consiste na ideia básica de que as amostras mais próximas entre si são mais similares do que as que se

encontram distantes. Assim, os clusters são gerados baseados na distância entre as amostras formando uma “árvore de clusters”. Neste tipo a quantidade de clusters é definida quando o usuário define um limitante para a divisão de distâncias. Na figura 4 é apresentada a árvore de clusters (Brian, et. al., 2011).

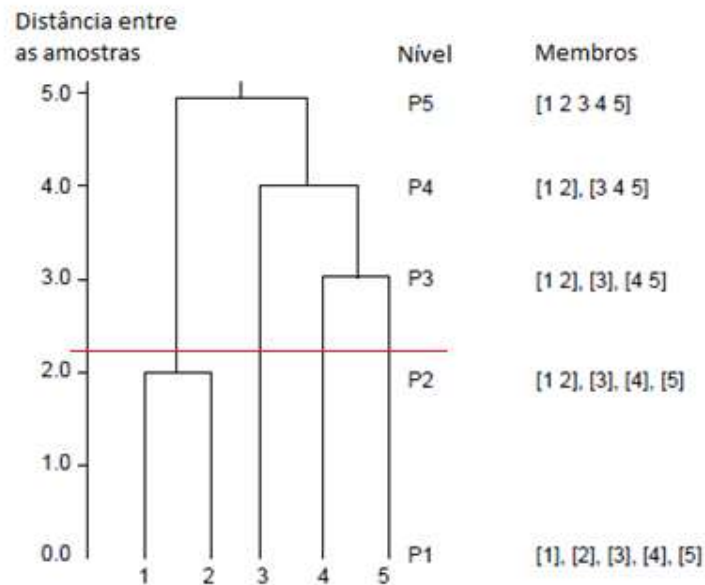


Figura 4. Árvore de clusters (modificado de Brian, et. al., 2011).

Na Figura 4, por exemplo, tem-se o limite definido pela linha vermelha na distância aproximada de 2, onde há o nível P2, com 4 clusters, neste caso membros dos clusters seriam 1 e 2 em conjunto e os outros individuais.

## 2.2 ALGORITMOS IMUNO-INSPIRADOS

O estudo de sistemas imunológicos na ciência da computação visa imitar alguns processos do sistema imunológico de seres vertebrados no âmbito computacional para solucionar problemáticas. Estudos nessa área não são novidade. O primeiro trabalho a relacionar estes dois campos da ciência foi realizado em 1846, por Farmer et al. Nas décadas seguintes, diversos estudos aprofundaram-se nesta área, e com a revolução dos dados e o *Big Data*, estas técnicas vêm sendo constantemente revisadas (de França et. al., 2010). Neste trabalho é utilizada uma parte desta família de algoritmos para realizar a clusterização, tendo alguns motivos para sua utilização que serão explicados adiante.

Na década de 1970 cientistas desenvolveram teorias para explicar o funcionamento da resposta imunológica adaptativa do sistema imunológico dos vertebrados, onde a Teoria da Seleção Clonal (Burnet, 1978) e a Teoria da Rede Imunológica (Jerne, 1974) inspiraram o desenvolvimento da família de algoritmos imuno-adaptativos. Apesar destas teorias terem perdido espaço em detrimento de outras, por falta de evidências empíricas no ramo da medicina, as técnicas provenientes da ideia da Rede Imunológica ganharam força no âmbito computacional (de França et. al., 2010). Os conceitos do sistema serão sucintamente explicados a seguir para uma melhor compreensão do algoritmo.

O sistema imunológico dos vertebrados utiliza de mecanismos para identificação de um agente específico causador de doença (patógeno) que são de extrema importância para manter um organismo saudável. De acordo com a Teoria da Seleção Clonal a resposta imunológica ocorre da seguinte maneira (de Castro & Timmis, 2002a):

1. Patógeno entra no organismo;
2. Antígeno (uma parte do patógeno que é reconhecida pelo sistema imunológico) se liga ao anticorpo ou ao linfócito B (célula capaz de produzir anticorpos) que desencadeiam a resposta imunológica para eliminar o invasor;
3. Linfócitos B iniciam um processo de proliferação através de clonagem, para aumentar a geração de anticorpos capazes de identificar antígeno em questão;
4. Na clonagem os linfócitos B sofrem uma mutação controlada, para modificar a afinidade de seus anticorpos ao antígeno;
5. Se os anticorpos destas novas células tiverem maior afinidade aos antígenos, elas são selecionadas para permanecerem no sistema enquanto que a original é removida, podendo também o contrário acontecer.

O mecanismo de clonagem, mutação e supressão foi a fonte de inspiração para os algoritmos de clusterização. Desta forma é possível fazer um paralelismo com um pseudocódigo que clona e modifica um cluster para melhor se adaptar aos dados em questão, e de maneira análoga, se o cluster não for melhor que o original ele é suprimido (de França et. al., 2010).

Neste trabalho é utilizado o algoritmo denominado Algoritmo Imunológico com Raio Adaptativo ou ARIA (do inglês *Adaptive Radius Immune Algorithm*). Este algoritmo utiliza-se de mecanismos de expansão clonal e supressão da rede de acordo com os parâmetros de



densidade apresentados nos dados, para assim melhor abranger as amostras. O algoritmo é rápido em termos computacionais e de simples implantação (Bezerra et. al., 2005).

Na figura 5 são exemplificados pequenos antígenos espalhados, a localização e o raio dos anticorpos são atualizados para melhor abranger os antígenos.

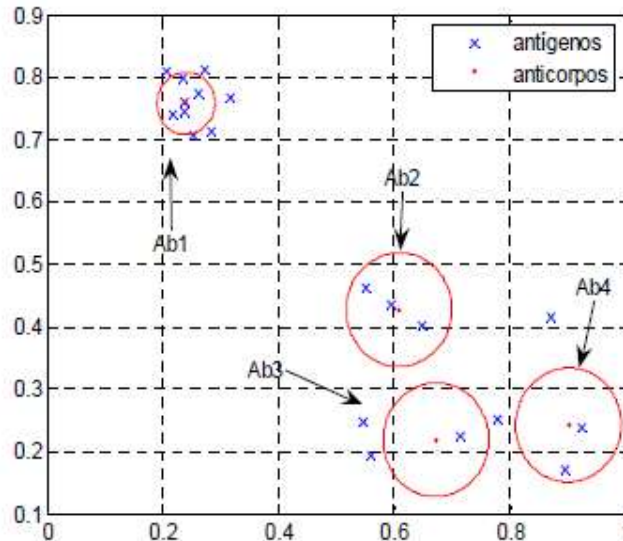


Figura 5. Exemplo de clusterização com ARIA. (Von Zuben, s. d.)

O mecanismo do ARIA permite a supressão de anticorpos e adaptação de seu raio de acordo com a densidade local de amostras (antígenos). Desta forma, em áreas com grande densidade de dados os anticorpos diminuem seu raio e podem se agrupar próximos uns aos outros, e de maneira oposta, assumem grandes raios em áreas com amostras dispersas (Bezerra et. al., 2005).

Este algoritmo foi escolhido por ser um dos mais conhecidos algoritmos imunológicos, tendo sua eficiência já sido testada em grandes grupos de dados e por conseguir preservar as informações de densidade de informação após a clusterização, gerando dados mais assertivos (Bezerra et. al., 2005).

O algoritmo pode ser simplificado em três etapas:

- 1 Maturação da afinidade: os anticorpos sofrem mutação para melhor abranger os antígenos (amostras) apresentados;

- 2 Expansão clonal: Os anticorpos que melhor se adaptam são clonados e a rede de anticorpos se expande;
- 3 Supressão da rede: Os anticorpos que reconhecem outro anticorpo dentro de seu raio são eliminados.

Um pseudocódigo e suas respectivas variáveis de entrada são mostrados de forma resumida no algoritmo a seguir (Bezerra et. al., 2005):

*1 Inicialização de variáveis*

*2 For it = 1 to max\_it do:*

*2.1 For I = 1 to N do:*

*2.1.1 Seleciona o anticorpo (Ac) que melhor se adapta a um antígeno (Ag);*

*2.1.2 Aproxima o Anticorpo (Ac) com taxa  $\mu$  na direção do antígeno (Ag);*

*end*

*2.2 Elimina os Anticorpos (Ac) que não foram estimulados;*

*2.3 Clona os Anticorpos (Ac) que reconhecem antígenos (Ag) localizados a uma distância maior que seu raio (R);*

*2.4 Calcula a densidade local para cada anticorpo (Ac);*

*2.5 Calcula o raio limiar para a supressão dos anticorpos através da fórmula:*

$$R_{\text{anticorpo}} = \text{raio mínimo } (r) * \left( \frac{\text{den}_{\text{max}}}{\text{den}_{\text{iteração}}} \right)^{1/\text{dim}}$$

*2.6 Suprime anticorpos redundantes dando prioridade aos que tem menor raio (R);*

*2.7 Atualiza o raio médio ( $E = \text{média}(R)$ );*

*2.8 Se a iteração atual é maior que a metade número total de iterações.*

*2.8.1 Reduz a taxa de mutação ( $\mu = \mu * \text{decaimento}$ );*

*end*

*end*

Na tabela 1 são descritas as variáveis do pseudocódigo.

Tabela 1. Variáveis do pseudocódigo.

Variável	Descrição
Ac	Anticorpo
Ag	Antígeno
R	Raio de cada anticorpo
r	Menor raio admitido
$\mu$	Taxa de mutação
decaimento	Taxa de decaimento da mutação
E	Raio médio para o cálculo da densidade de amostras
dim	Dimensão das amostras

Os passos a seguir descrevem as etapas do algoritmo (Bezerra et. al., 2005):

1. Inicialização de variáveis. O tamanho inicial da população de anticorpos e seu respectivo raio é indiferente, pois este número será adaptado após as iterações para melhor abranger as amostras. A taxa de mutação inicial pode ser igual a 1 e sua taxa de diminuição igual a 0,95. É escolhido também o raio mínimo permitido para os anticorpos que deve ser escolhido cuidadosamente de acordo com a densidade das amostras, mas isso será discutido mais adiante;
  - 2.1 Nesta etapa ocorre a maturação da afinidade entre anticorpo e antígeno;
    - 2.1.1 O anticorpo que melhor se adapta a um antígeno, isto é, aquele com a menor distância euclidiana a um antígeno específico, sofre mutação;
    - 2.1.2 Este antígeno é “mutado”. Em outras palavras, movimentado em direção ao antígeno. O processo então repete-se para um antígeno diferente, até que todos os antígenos tenham sido avaliados;
 
$$\text{Nova posição do anticorpo} = (\text{Posição do anticorpo original}) + \text{decaimento} * (\text{número aleatório entre 0 e 1}) * (\text{distância entre antígeno e anticorpo})$$
  - 2.2 Se algum anticorpo não for mutado nenhuma vez, é porque ele não foi o anticorpo mais próximo a um antígeno nenhuma vez, então ele é excluído;
  - 2.3 Nesta etapa ocorre a Expansão Clonal. Se ainda existem antígenos fora do raio de atuação dos anticorpos, duplica-se o anticorpo que está mais próximo do antígeno isolado;

- 2.4 Calcula-se o raio médio de todos os anticorpos e com isso a densidade das amostras de cada anticorpo. O valor é o número de amostras dentro do raio médio (E) para cada anticorpo;
- 2.5 Em seguida é atualizado o raio dos anticorpos existentes com base na densidade de antígenos dentro do raio de acordo com a fórmula mencionada no pseudocódigo;
- 2.6 Nesta etapa ocorre a supressão da rede. Após a atualização, se dois anticorpos estão muito próximos de modo que seus raios englobem o vizinho, então o de maior raio deve ser excluído para eliminar redundâncias;
- 2.7 A média dos raios dos anticorpos é atualizada;
- 2.8 O decaimento é atualizado.

Estas iterações são repetidas várias vezes de acordo com a definição do próprio usuário, mas empiricamente foi constatado que após 150 iterações temos poucas mudanças na configuração dos clusters. O único parâmetro crucial que precisa de atenção ao executar o código é a constante “r”. Ela representa o valor do menor raio possível em uma base, o qual ocorrerá na região com maior densidade de dados. O valor escolhido varia e está relacionado a forma de distribuição dos dados, uma má escolha pode ter grande impacto no resultado (Bezerra et. al., 2005).

A figura 6 mostra um exemplo da clusterização através do ARIA, onde o azul são as amostras distribuídas no espaço e o ponto preto com raio vermelho os anticorpos com seu raio de atuação. No exemplo temos o início do processo com poucos anticorpos distribuídos em (a) e pode-se observar a adaptação dos anticorpos a medida que aumenta-se o número de iterações em (b) e (c).

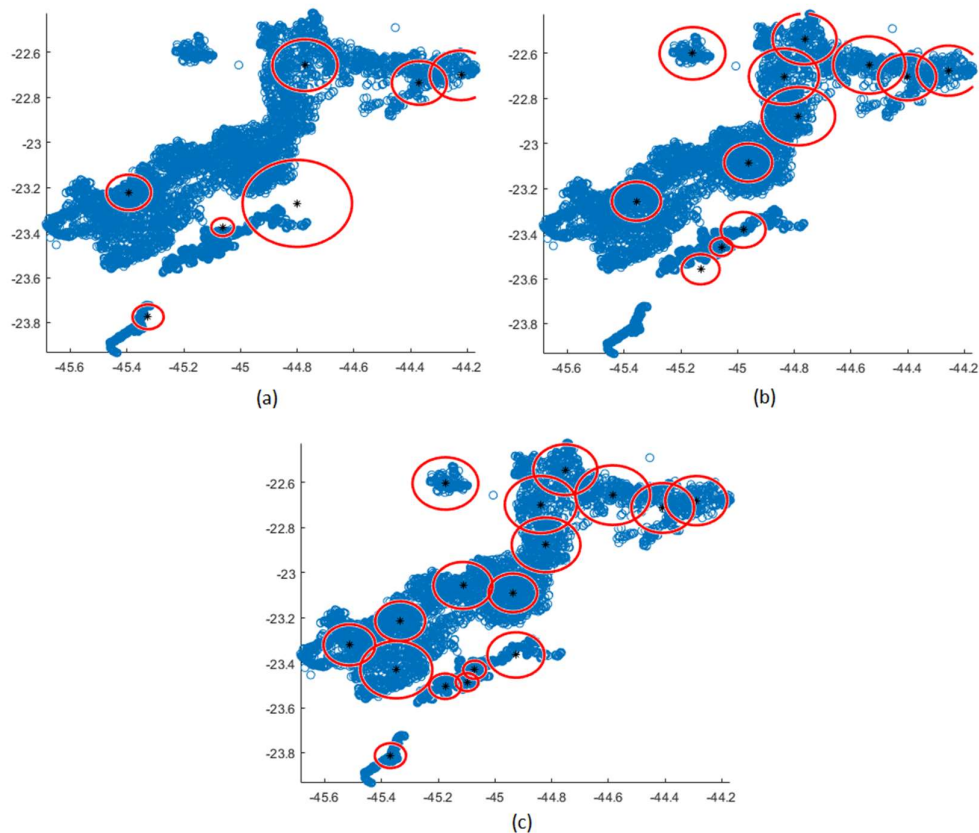


Figura 6. Exemplo de Clusterização do ARIA. (a) Início das iterações; (b) Clusterização em andamento; (c) Final do processo.

A grande vantagem deste método é que, ao contrário de outras propostas de clusterização, não é necessário definir previamente em quantos grupos será dividido o banco de dados, o próprio algoritmo fornece esta resposta (de França et. al., 2010).

Neste trabalho, os dados analisados consistem em ocorrências emergenciais, que se comportam como os antígenos. O objetivo é posicionar bases de atendimento, que funcionam como os anticorpos. No cotidiano das empresas as ocorrências podem ser distantes ou próximas das bases o que diminui ou aumenta a densidade de ocorrências, alterando assim o raio de atuação da base. Conforme visto, o ARIA se comporta da mesma maneira, tornando este algoritmo muito útil para esta problemática específica.

Para formular o estudo, utilizou-se a ferramenta Matlab para processar o algoritmo com os dados de ocorrências emergenciais no estado de Pernambuco. Estes dados foram coletados nos últimos 4 anos pela distribuidora de energia CELPE e constam na base de dados

da empresa. Os dados foram organizados em uma planilha do Excel. O período de 4 anos foi escolhido devido ao fato de que este período abrange a maioria dos eventos sazonais no Brasil, sejam eles causados pelo homem ou naturais. Para validar os deslocamentos entre as bases sugeridas e as ocorrências foi utilizado o website Google Maps (Google Maps) e para gerar o mapa de calor das ocorrências o website Google Fusion Tables (Google Fusion Tables).

### 3 *Resultados*

Devido às características do sistema de coleta de dados vigente na CELPE os dados das ocorrências da distribuidora não são adquiridos na sua localização exata. Desta forma, a base é organizada pela quantidade de eventos que ocorreram em uma base já existente.

Atualmente temos no estado de Pernambuco 50 bases de restabelecimento. Sendo assim, a base de dados em questão é organizada conforme a estrutura exemplificada na tabela 2, representando: Nome da cidade em que se encontra a base (Localidade); Quantidade de ocorrências registradas (OC); Quantidade de ocorrências normalizada (OC normalizadas); Latitude da cidade (Latitude); Longitude da cidade (Longitude). Sendo que as coordenadas de latitude e longitude foram obtidas através do site oficial do IBGE (Instituto Brasileiro de Geografia e Estatística). A tabela 2 mostra um exemplo do formato da base de dados.

Tabela 2. Exemplo da estrutura de planilha

Localidade	OC	OC normalizadas	Latitude	Longitude
CAMARAGIBE	27696	2770	-8,02	-34,98

A coluna descrita como “OC normalizadas” representa a quantidade de ocorrências divididas por 10 e arredondadas para cima para que tenhamos um número inteiro maior ou igual a 1, uma forma de simplificar a base e tornar mais ágil o processamento dos dados. Esta normalização tem um efeito pouco significativo nas características dos dados, visto que a maioria das localidades possui mais de mil ocorrências, totalizando 656.392 eventos.

Na figura 7 é exibido o resultado da clusterização na base de dados em 4 etapas, na progressão de (a) até (d).

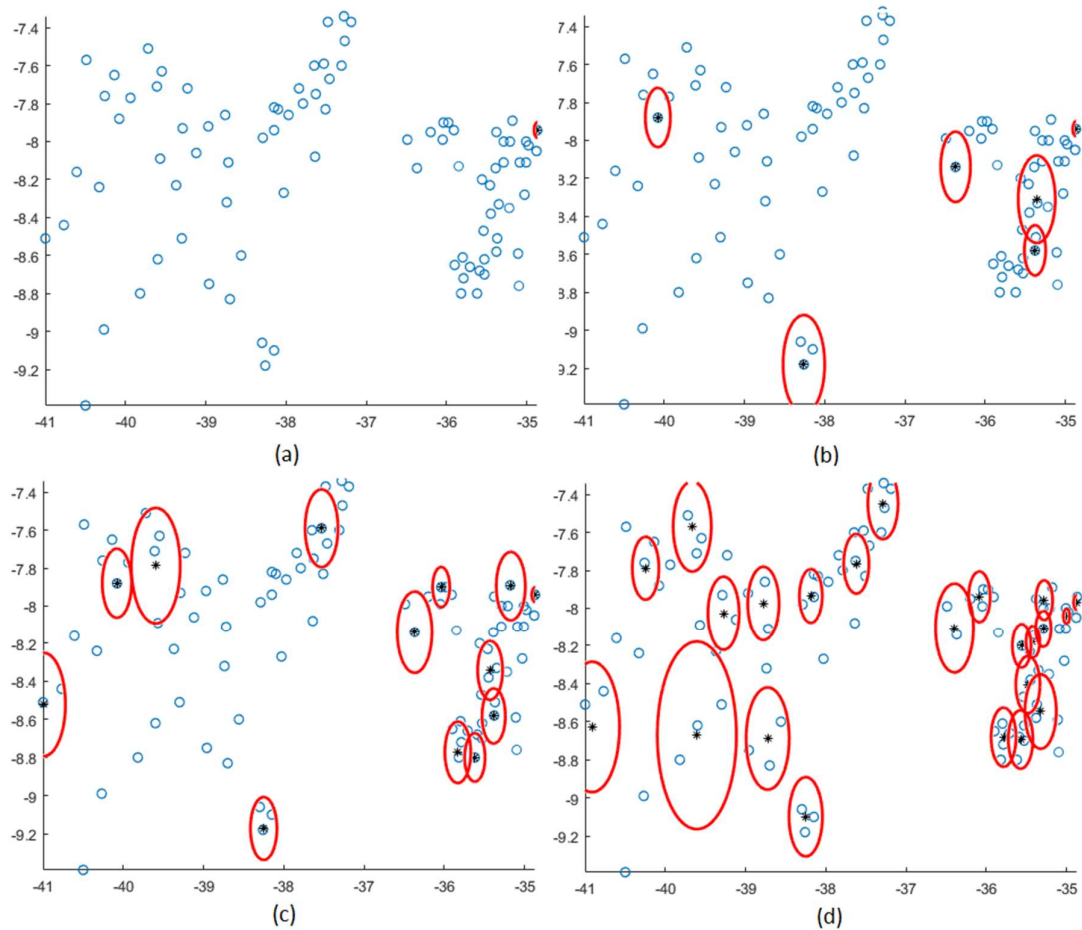


Figura 7. Processo de clusterização na base de dados. (a) 1 Iteração; (b) 50 Iterações; (c) 100 Iterações; (d) 150 Iterações e final do processo.

Em (a) tem-se a primeira iteração do código a base de dados das 50 cidades, cada uma representada por um ponto azul e com seu respectivo “peso” de ocorrências geograficamente distribuídas. Pode-se observar o ponto preto com raio vermelho como a primeira base (ou anticorpo) posicionada. Em (b) e em (c), o código encontra-se em sua 50<sup>a</sup> e 100<sup>a</sup> iteração, respectivamente. Observa-se a progressão na adaptação das bases e seus respectivos raios. O código segue fazendo otimizações e se adaptando até que se chega ao resultado final em (d), após 150 iterações predefinidas. Em (d) observa-se os diferentes raios das bases que se adaptaram à densidade de ocorrências na região, visando manter uniforme a densidade os raios aumentaram ou diminuíram respeitando o raio mínimo predeterminado no código. Os



raios na Figura 7 aparecem em forma elíptica embora sejam circulares, pois, os eixos verticais e horizontais estão em diferentes escalas. Para gerar os dados em questão foi escolhido uma taxa de mutação inicial igual a 1 e a constante de decaimento igual a 0,95, esta escolha permitiu que por volta da 150ª iteração já não houvesse grandes mudanças no sistema.

Sendo assim, observa-se na Figura 8 os clusters resultantes. Foram obtidos 23 clusters representados pelas diferentes cores e suas respectivas posições ideais para as bases representadas pelo asterisco preto.

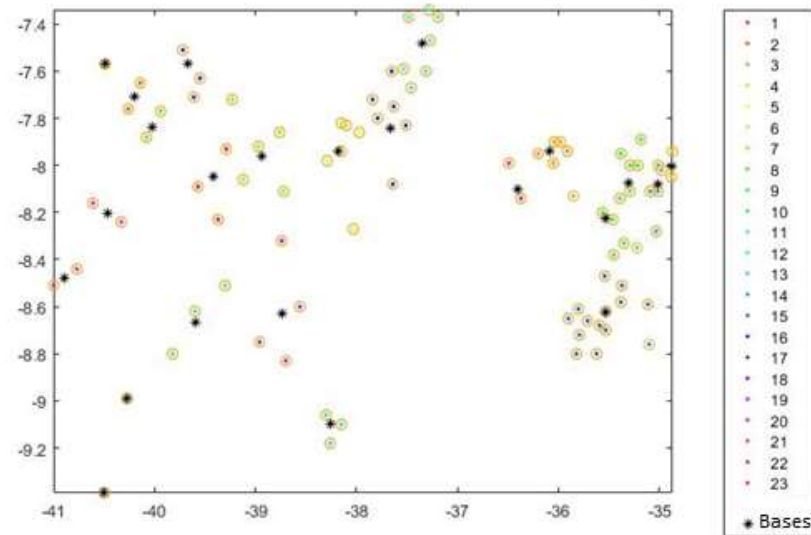


Figura 8. Clusters resultantes.

O código oferece a localização ideal para as bases, retornando para o usuário a latitude e longitude ideais destas. Sendo assim, o resultado são pontos no eixo cartesiano, após analisar estas localizações no Google Maps, observa-se que as características geográficas da região nem sempre comportam uma base como, por exemplo, zonas rurais ou localidades remotas.

Na figura 9 são exemplificadas bases sugeridas pela clusterização que não resultam em alocações plausíveis.

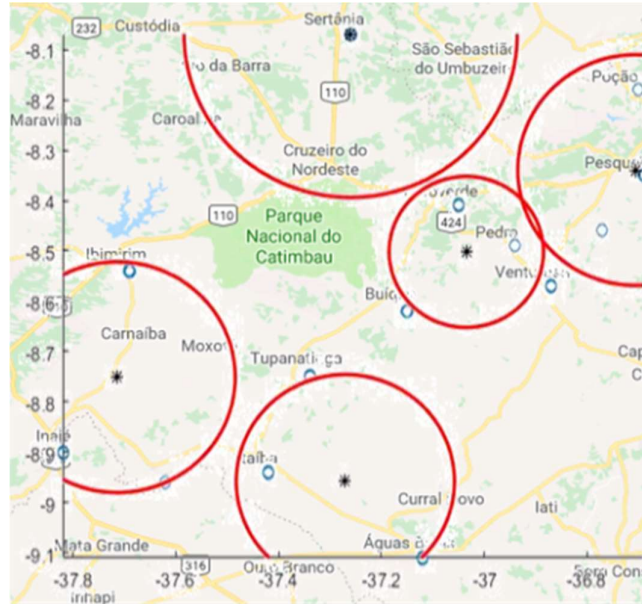


Figura 9. Exemplo de bases sugeridas pela clusterização que não resultaram em alocações plausíveis por serem coordenadas de região rural.

Algumas bases sugeridas, como as exibidas na Figura 10, encontram-se em zona rural ou em lugares remotos, cabendo ao usuário realizar uma análise mais subjetiva a fim de encontrar uma cidade estratégica para posicioná-la.

Para escolher a cidade deve-se considerar as condições de infraestrutura das cidades próximas ao ponto sugerido, e também avaliar o tempo de deslocamento da base escolhida até as demais cidades dentro do cluster em questão.

Muitas vezes a cidade escolhida para base e uma cidade vizinha, apesar de próximas no plano cartesiano, são divididas por barreiras naturais como rios ou parques ecológicos, que tornam o deslocamento entre elas mais demorado. Nesta situação, deve-se verificar se o deslocamento entre esta cidade isolada e a base do cluster vizinho não é mais ágil, afim de se obter o melhor desempenho na atividade de restabelecimento.

Outra problemática não prevista no código acontece em regiões metropolitanas. Estas regiões têm como característica alta concentração de ocorrências e um tempo lento de deslocamento causado muitas vezes por congestionamento de veículos.

Para solucionar este problema na região metropolitana de Recife, foram separadas as cidades que compõem esta região e o código foi executado sobre esta região específica. Para a

maior parte do estado foi utilizado o raio mínimo de 0,1, sendo este satisfatório para a densidade dos dados apresentados na figura 9, garantindo o máximo de uma hora no deslocamento quando verificado no Google Maps. No entanto, para personalizar a alocação das bases da região metropolitana o raio mínimo para as bases foi diminuído para 0,04 e o tempo de deslocamento de 30 minutos, podendo então o algoritmo se adaptar até formar clusters com pequenos raios de atuação. Isto resulta em um número maior de clusters, com menores raios de atuação do que o limite estabelecido para o resto do estado. Isto significa que se tem mais bases distribuídas para evitar os grandes tempos de deslocamento em zonas com congestionamento de veículos. Na figura 10 é ilustrado o mapa de calor das ocorrências da região metropolitana e as bases resultantes em preto.

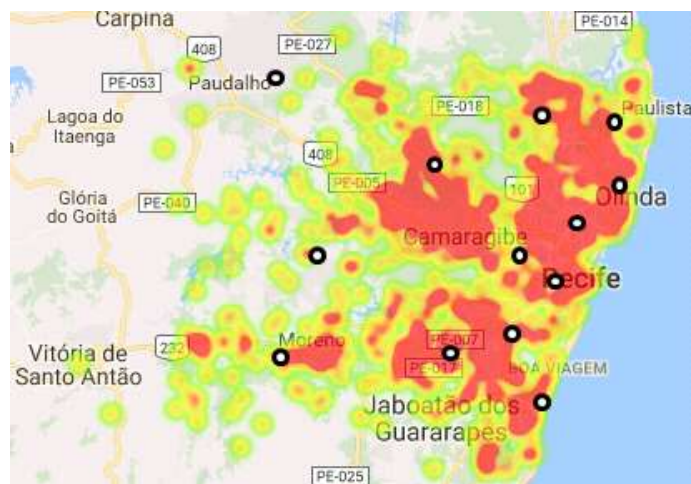


Figura 10. Mapa de calor das ocorrências da região metropolitana e as bases resultantes em preto.

A maior concentração de bases nesta região é também muito justificada quando se considera os indicadores de qualidade no fornecimento de energia DEC e FEC. Sabendo-se que a maior parte das ocorrências do estado de Pernambuco se concentra no litoral metropolitano, um sucesso na operação da região metropolitana traz um impacto significativo nos indicadores da companhia e contribuem significativamente para a saúde do negócio.

## ***4 Conclusão***

Através do trabalho realizado, pôde-se notar que técnicas de análise inteligente de dados, são ferramentas poderosas para melhoria de processo e tomada de decisões em empresas. A revolução de dados é uma realidade e empresas que melhor aproveitarem as oportunidades desta nova realidade terão vantagem competitiva.

Através da utilização de técnicas de clusterização é possível abordar de várias maneiras diferentes problemáticas para dispersão geográfica de dados. A utilização do algoritmo ARIA foi bastante eficiente para a problemática de bases de restabelecimento de energia, podendo futuramente serem incorporadas novas ferramentas computacionais ao código para sanar falhas na metodologia. É possível, por exemplo, através de um código mais aprofundado validar automaticamente os tempos de deslocamento no Google Maps e propor a base ótima com este parâmetro.

Baseado neste estudo é possível aprimorar as técnicas de clusterização para propor soluções concretas de posicionamento geográfico em diversos setores da indústria. Esta técnica pode ser aplicada à centros de distribuição de logística ou outras atividades que requeiram uma base de atuação para um determinado raio de abrangência geográfica. Este tipo de configuração está presente em muitas empresas, sendo suas aplicações infinitas.

Por fim, a mudança na operação da empresa para as bases propostas pelo algoritmo depende de diversos fatores que garantem a viabilidade econômico financeira do negócio. Fatores econômicos como, por exemplo, o investimento necessário, taxa de retorno, modificação no número de pessoas e veículos necessários, além de aquisição de infraestrutura predial tem grande papel na decisão da implantação ou não da estrutura. No entanto, é evidente que a estrutura proposta melhora a saúde operacional da atividade de restabelecimento do fornecimento de energia elétrica, além de ir ao encontro com a política de incentivo a investimentos em ativos promovidos pela agência reguladora ANEEL.

## ***Referências Bibliográficas***

- [1] ABRADDEE, Associação Brasileira de Distribuidores de Energia Elétrica. Website <http://www.abradee.com.br>, 2018.
- [2] ANEEL, Agência Nacional de Energia Elétrica. Condições Gerais de Fornecimento de Energia Elétrica – REN no. 499/2012, 2012 Revisão Tarifária – Módulos 2.1 e 2.2, 2011.
- [3] ANEEL, Agência Nacional de Energia Elétrica. PRODIST, Procedimentos de Distribuição – Módulo 8, 2011.
- [4] ANEEL, Agência Nacional de Energia Elétrica. PRORET, Procedimentos de Revisão Tarifária – Módulos 2.1 e 2.2, 2011.
- [6] Berkhin, P. (2006). A Survey of Clustering Data Mining Techniques. In: Kogan J., Nicholas C., Teboulle M. (eds) Grouping Multidimensional Data. Springer, Berlin, Heidelberg.
- [7] Bezerra G.B., Barra T.V., de Castro L.N., Von Zuben F.J. (2005) Adaptive Radius Immune Algorithm for Data Clustering. In: Jacob C., Pilat M.L., Bentley P.J., Timmis J.I. (eds) Artificial Immune Systems. ICARIS 2005. Lecture Notes in Computer Science, vol 3627. Springer, Berlin, Heidelberg.
- [8] Brian S. Everitt . Sabine Landau Morven Leese . Daniel Stahl (2011). Cluster Analysis 5th Edition. King's College London, UK.
- [9] Burnet, F. M. (1978). Clonal Selection and After . In Bell, G. I., Perelson, A. S., & Pimgley, G. H. Jr., (Eds.), Theoretical Immunology (pp. 63–85). New York: Marcel Dekker.
- [10] de França, F. O., Coelho, G. P., Castro, P. A., & Von Zuben, F. J. (2010). Conceptual and Practical Aspects of the aiNet Family of Algorithms. International Journal of Natural Computing Research (IJNCR), 1(1), 1-35.
- [11] de Castro, L. N., & Von Zuben, F. J. (2002b). aiNet: An artificial immune network for data analysis. In H. A. Abbass, R. A. Sarker, & C. S. Newton (Eds.), Data Mining: A Heuristic Approach (pp. 231-259). Hershey, PA: IGI Global.
- [12] Estivill-Castro, Vladimir (2002). Why so many clustering algorithms – A Position Paper.
- [13] Fontana, Heron (2015). Logística Operacional – Alocação de bases operacionais em distribuição de energia elétrica. Universidade de São Paulo, Brasil.
- [14] Google Maps, Website <https://www.google.com.br/maps>
- [15] Google Fusion Tables, Website <https://sites.google.com/site/fusiontablestalks/>
- [16] Jerne, N. K. (1974). Towards a network theory of the immune system. Ann. Immunol. Inst. Pasteur, 125C, 373–389.

- [17] McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, 90(10), 60-68.
- [18] MME, Ministério de Minas e Energia. RESEB – Reestruturação do Setor Elétrico Brasileiro – Sumário Executivo, 2011.
- [19] Neoenergia, Website <http://www.neoenergia.com>, 2018.
- [20] Von Zuben F.J. Tutorial do Aria (Adaptive Radius Immune Algorithm). Universidade de Campinas. Universidade de Campinas, Brasil.