



UNIVERSIDADE  
ESTADUAL DE LONDRINA

---

EDISON ANTONIO SAHD FILHO

**FORMAÇÃO DE CONHECIMENTO E AUTOMATIZAÇÃO  
DO PROCESSO DE TOMADA DE DECISÃO EM  
GOVERNANÇA DE TECNOLOGIA DA INFORMAÇÃO  
UTILIZANDO ALGORITOMO GENETICO**

---

Londrina  
2016



EDISON ANTONIO SAHD FILHO

**FORMAÇÃO DE CONHECIMENTO E AUTOMATIZAÇÃO  
DO PROCESSO DE TOMADA DE DECISÃO EM  
GOVERNANÇA DE TECNOLOGIA DA INFORMAÇÃO  
UTILIZANDO ALGORITOMO GENETICO**

Dissertação apresentada ao Programa de Mestrado em Ciências da Computação da Universidade Estadual de Londrina para obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Rodolfo Miranda de Barros  
Coorientador: Prof. Dr. Sylvio Barbon Junior

Londrina  
2016

---

Edison Antonio Sahd Filho

Formação de conhecimento e automatização do processo de tomada de decisão em governança de tecnologia da informação utilizando algoritmo genético/ Edison Antonio Sahd Filho. – Londrina-PR, 2016-

93 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof(a). Dr(a). Rodolfo Miranda de Barros

– Universidade Estadual de Londrina, 2016.

1. Palavra-chave1. 2. Palavra-chave2. I. Orientador. II. Universidade xxx. III. Faculdade de xxx. IV. Título

CDU 02:141:005.7

---

EDISON ANTONIO SAHD FILHO

**FORMAÇÃO DE CONHECIMENTO E AUTOMATIZAÇÃO DO  
PROCESSO DE TOMADA DE DECISÃO EM GOVERNANÇA DE  
TECNOLOGIA DA INFORMAÇÃO UTILIZANDO ALGORITOMO  
GENETICO**

Dissertação apresentada ao Programa de Mestrado em Ciências da Computação da Universidade Estadual de Londrina para obtenção do título de Mestre em Ciência da Computação.

**BANCA EXAMINADORA**

---

Orientador: Prof. Dr. Rodolfo Miranda de Barros  
Universidade Estadual de Londrina – UEL

---

Prof. Dr. Bruno Bogaz Zarbelão  
Universidade Estadual de Londrina – UEL

---

Prof. Dr. Elieser Botelho Manhas Junior  
Universidade Estadual de Londrina – UEL

---

Prof. Dr. Lisandro Modesto  
Serviço Nacional de Aprendizagem Industrial –  
SENAI Londrina

Londrina, 10 de maio de 2016.



## AGRADECIMENTOS

Primeiramente, agradeço ao professor Rodolfo Miranda de Barros pela nossa amizade, por o total apoio, compreensão e dedicação, não apenas em relação ao desenvolvimento dos projetos, e por todas as oportunidades que me foram concedidas.

Agradeço ao professor Sylvio Barbon Junior pela nossa amizade, por sua total dedicação e paciência durante as orientações, e por auxiliar com o meu crescimento, não só profissional.

A toda a minha família, em especial ao meu pai Edison Antonio Sahd, minha mãe Vera Lucia Stoeglehner Sahd e a minha irmã Claudia Stoeglehner Sahd.

Aos docentes, em especial ao professor Bruno Bogaz Zarpelão, que contribuíram com o meu aprendizado e desenvolvimento pessoal.

Ao Departamento de Computação da Universidade Estadual de Londrina.

Aos meus verdadeiros amigos, em especial Aline Nascimento, Bruno Gonçalves Dias, Cláudio Kawakani, Felipe Rodrigues Werner, Fernanda Martins de Oliveira, Karen Takarada, Luiz Fernando Carvalho, Mario Henrique Adaniya, Paulo Cavaleiro, Rafael Gomes Mantovani, Rafael Thiago da Silva, Rodrigo Ghiraldi, Rodrigo Igawa, Valter Bercini Junior, Vinicius Barros, Wilson Vinícius Teixeira e aos demais amigos, pelo total apoio e compreensão durante esta etapa da minha vida.



*“A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê.”*  
*(Arthur Schopenhauer)*



SAHD FILHO, E. A.. **Formação de conhecimento e automatização do processo de tomada de decisão em governança de tecnologia da informação utilizando algoritmo genético.** 93 p. Dissertação de Mestrado (Mestrado em Ciência da Computação) – Universidade Estadual de Londrina, Londrina–PR, 2016.

## RESUMO

A governança de TI é um conjunto de atividades complexas que possui o propósito de alinhar os objetivos da tecnologia da informação com a organização e entregar valor para a mesma. O objetivo deste trabalho é propor um modelo capaz de auxiliar o processo de tomada de decisão em governança de tecnologia da informação. Este modelo deve ser capaz de armazenar o conhecimento e recomendar pontos de melhoria relativo ao contexto de governança de TI. Ambos os objetivos foram alcançados no presente trabalho através de um modelo de repositório de conhecimento baseado em *data warehouse* e de algoritmo genético. Por se tratar de uma abordagem com características estocásticas, o algoritmo genético pode encontrar diferentes soluções para uma determinada situação. Entretanto, um padrão é estabelecido devido ao critério de seleção da população inicial. Desta forma, é necessário calcular a probabilidade de um determinado ponto de melhoria ser de fato o mais recomendado.

**Palavras-chave:** Governança de TI. Repositório de conhecimento. Automação de tomada de decisão. Algoritmo genético.



SAHD FILHO, E. A.. **Knowledge construction and decision making automation in information technology using genetic algorithm**. 93 p. Master's Thesis (Master in Science in Computer Science) – State University of Londrina, Londrina–PR, 2016.

## ABSTRACT

Information technology governance is a set of complex activities that aims to align the information technology goals to the organization, in order to deliver value. This work aims to propose a model able to help in the process of decision making in information technology governance. This model must store the knowledge and recommend improvements related to the IT governance context. Both the goals have been achieved in the present work through a knowledge warehouse model based on concepts of data warehouse and genetic algorithm. This approach may find different solutions to a given situation because of its stochastic features. However, a pattern is established due to the criterion of initial population selection. Thus, it is possible to deal with the probability of a certain improvement be the most recommended.

**Keywords:** IT governance. Knowledge warehouse. Decision making automation. Genetic algorithm.



## LISTA DE ILUSTRAÇÕES

Figura 1 – Processo de tomada de decisão em governança de TI. Adaptado de [1].	27
Figura 2 – Visão geral do <i>framework</i> . Adaptado de [2].	28
Figura 3 – Exemplo de questão de avaliação. Disponível em [2].	29
Figura 4 – Relacionamento entre questão e domínios de TI. Disponível em [2].	29
Figura 5 – Estrutura do modelo de repositório. Adaptado de [2].	31
Figura 6 – Modelo estrela. Adaptado de [3].	33
Figura 7 – Modelo floco de neve. Adaptado de [3].	33
Figura 8 – Modelo constelação de fatos. Adaptado de [3].	34
Figura 9 – Paradigmas da inteligência computacional. Disponível em [4].	35
Figura 10 – Exemplo de resultado de uma estratégia aglomerativa. Disponível em [3].	38
Figura 11 – Fluxograma do funcionamento do GA. Adaptado de [5].	40
Figura 12 – Esquematização do processo de <i>crossover</i> com corte simples e duplo.	42
Figura 13 – Processo de tomada de decisão através dos questionários de avaliação de governança de TI.	44
Figura 14 – Arquitetura do sistema de recomendação de pontos de melhoria em governança de TI.	45
Figura 15 – Estrutura geral do repositório de conhecimento em TI.	48
Figura 16 – Relacionamento entre os fatos do repositório.	59
Figura 17 – Mapeamento do questionário pelo repositório.	60
Figura 18 – Mapeamento das soluções pelo repositório.	60
Figura 19 – Exemplo de evolução do <i>fitness</i> por geração.	68



## LISTA DE TABELAS

Tabela 1 – Quantidade de alternativas por questão . . . . .	52
Tabela 2 – Resultados da análise de descritores de TI. . . . .	57
Tabela 3 – Descrição dos fatores e tipos de atributo. . . . .	58
Tabela 4 – Precisão dos modelos de agrupamento. . . . .	63
Tabela 5 – Matriz de confusão do <i>hierarchical clustering</i> . . . . .	63
Tabela 6 – Taxa de atendimento dos domínios de TI dos objetos do <i>cluster 5</i> . . . . .	64
Tabela 7 – Precisão dos algoritmos de geração de dados sintéticos. . . . .	65
Tabela 8 – Configurações de experimentos com GA. . . . .	67



## LISTA DE ABREVIATURAS E SIGLAS

AI	<i>Artificial Intelligence</i>
AIS	<i>Artificial Immune Systems</i>
CMM	<i>Capability Maturity Model</i>
EC	<i>Evolutionary Computation</i>
FS	<i>Fuzzy Systems</i>
GA	<i>Genetic Algorithm</i>
IC	<i>Inteligência Computacional</i>
ML	<i>Machine Learning</i>
NN	<i>Neural Network</i>
OLAP	<i>Online Analytical Processing</i>
OLTP	<i>Online Transaction Processing</i>
SI	<i>Swarm Intelligence</i>
TI	Tecnologia da Informação



# SUMÁRIO

1	INTRODUÇÃO . . . . .	21
2	FUNDAMENTAÇÃO TEÓRICA . . . . .	25
2.1	Governança de tecnologia da informação . . . . .	25
2.2	Framework de desenvolvimento de governança de TI . . . . .	27
2.2.1	Mecanismo de análise . . . . .	29
2.2.2	Modelo de repositório de conhecimento em governança de TI . . . . .	30
2.3	Data warehouse . . . . .	31
2.4	Inteligência computacional . . . . .	34
2.4.1	<i>Clustering</i> . . . . .	35
2.4.1.1	<i>K-means</i> . . . . .	36
2.4.1.2	<i>Agglomerative hierarchical clustering</i> . . . . .	37
2.4.2	Geração de dados sintéticos . . . . .	38
2.4.3	Algoritmo genético . . . . .	40
3	DESCRIÇÃO DO MODELO PROPOSTO . . . . .	43
3.1	Processo de tomada de decisão baseado no <i>framework</i> de desenvolvimento de governança de TI . . . . .	43
3.2	Arquitetura do modelo proposto . . . . .	44
3.2.1	Estrutura do repositório de conhecimento em TI . . . . .	47
3.2.2	Algoritmos de geração de dados sintéticos . . . . .	49
3.2.3	Características do algoritmo genético . . . . .	51
4	METODOLOGIA DE PESQUISA E RESULTADOS . . . . .	55
4.1	Repositório de conhecimento em TI . . . . .	55
4.1.1	Metodologia . . . . .	56
4.1.2	Resultados e discussão . . . . .	56
4.2	Agrupamento de dados . . . . .	61
4.2.1	Metodologia . . . . .	61
4.2.2	Resultados e discussão . . . . .	63
4.3	Geração de dados sintéticos . . . . .	64
4.3.1	Metodologia . . . . .	64
4.3.2	Resultados e discussão . . . . .	65
4.4	Busca pelo ponto de melhoria . . . . .	66
4.4.1	Metodologia . . . . .	66
4.4.2	Resultados e discussão . . . . .	67

5	CONCLUSÃO . . . . .	71
	REFERÊNCIAS . . . . .	75
	APÊNDICES	79
	APÊNDICE A – IMPLEMENTAÇÃO DA GERAÇÃO DE DADOS SINTÉTICOS DETERMINÍSTICO EM R . . . . .	81
	APÊNDICE B – IMPLEMENTAÇÃO DA GERAÇÃO DE DADOS SINTÉTICOS TABELA EM R . . .	83
	APÊNDICE C – QUESTIONÁRIO DE VALIDAÇÃO DOS DESCRITORES DE GOVERNANÇA DE TI . . . . .	85
	Trabalhos Publicados pelo Autor . . . . .	93

# 1 INTRODUÇÃO

O século XXI vem sendo marcado pelo grande avanço tecnológico. Atualmente, a importância da informação e da tecnologia é amplamente reconhecida por organizações, em diversos níveis, como um diferencial estratégico. Assim, a noção de sustentação de uma empresa através dos recursos de tecnologia da informação (TI) vem se tornando cada vez mais comum.

De forma geral, a governança de TI é uma série de atividades, implantadas em diversos níveis organizacionais, responsáveis por sustentar e estender os objetivos da organização. O objetivo de uma TI precisamente implantada é a entrega de valor ao negócio. Entretanto, para alcançar este propósito é necessário que os objetivos de TI sejam conhecidos, os riscos mitigados, os recursos gerenciados e as responsabilidades delegadas [6].

Contudo, a aplicação e desenvolvimento de governança de TI requer um alto nível de conhecimento por parte dos envolvidos. Este conhecimento está relacionado tanto às características de negócio quanto à gestão e recursos de TI. Além disso, a governança de TI é um processo dinâmico e iterativo, ocasionando mudanças constantes no conhecimento necessário. Ainda, cada etapa do seu desenvolvimento é um processo custoso e longo [1].

Em consequência das dificuldades encontrada em aplicar e desenvolver a governança de TI, existe uma grande taxa de falhas e desistências por parte das organizações. Devido a este fato, foi apresentado por Briganó [2] um *framework* para o desenvolvimento de governança de TI. Esta ferramenta busca avaliar o cenário organizacional através de um questionário e identificar o seu nível de maturidade, de forma a auxiliar a tomada de decisão por parte dos envolvidos.

Com base no mecanismo de análise apresentado em [2], a proposta geral do presente trabalho é demonstrar um modelo computacional capaz de automatizar o processo de tomada de decisão em governança de TI. A automatização deste processo visa reunir o conhecimento obtido em situações distintas e, a partir deste, recomendar os pontos de melhoria mais indicados para a atual situação da TI de uma organização, seguindo a tendência de gestões semelhantes e que obtiveram resultados positivos.

Desta forma, o primeiro objetivo deste trabalho está relacionado ao processo de armazenamento do conhecimento referente à tomada de decisão em governança de TI. Através de uma revisão literária e baseando-se em conceitos de *data warehouse* e do *framework* de Briganó [2], é apresentado um modelo de repositório capaz de armazenar o conhecimento relacionado a este processo.

Além de colaborar com a automatização do processo de tomada de decisão, o repositório de conhecimento em governança de TI contribui com o acesso e a análise da

informação disponível, conforme a seguir:

- **Centralização de soluções:** a centralização de atividades e soluções de TI em uma única base é capaz de facilitar a busca e análise de soluções para uma determinada situação;
- **Mapeamento do processo de tomada de decisão:** ao se estabelecer a relação entre a descrição do cenário organizacional, a solução aplicada para alcançar uma melhoria e os resultados obtidos, é possível constituir padrões e facilitar a análise e a compreensão sobre a natureza de tomada de decisão em governança de TI;
- **Adaptação do conhecimento:** a abordagem de *data warehouse* mantém uma sequência histórica do conhecimento. Desta forma, é possível verificar a variação do conhecimento com o passar do tempo e compreender o comportamento do processo de tomada de decisão;
- **Compatibilidade com ferramentas *Online Analytical Processing* (OLAP) e *data mining*:** a estrutura do repositório de conhecimento em governança de TI é compatível com a aplicação de ferramentas de consulta do tipo OLAP. Ferramentas OLAP facilitam a visualização e compreensão das informações contidas em uma base de dados. Além disso, é possível aplicar *data mining* para explorar o conhecimento disponível e identificar informações implícitas e de difícil detecção;

O segundo objetivo deste trabalho se refere ao processo de recomendação de pontos de melhoria por meio do conhecimento mantido. Assim, é demonstrada a capacidade de modelos de Inteligência Computacional (IC) em desempenhar esta função. Em uma primeira etapa, modelos de *clustering* são analisados para desempenhar a tarefa de agrupamento dos cenários organizacionais, baseando-se em seus níveis de maturidade. A fase de *clustering* garante a redução do espaço de busca, garantindo que os questionários utilizados nesse processo sejam semelhantes entre si e apresentem um padrão comportamental.

Em seguida, um modelo conhecido como Algoritmo Genético, ou *Genetic Algorithm* (GA), é investigado com o objetivo de identificar o ponto de melhoria mais apropriado para uma determinada situação. Através de uma população inicial composta por indivíduos com um alto grau de semelhança com um cenário que deseja-se aprimorar, ou em análise, este modelo é capaz de extrair os padrões presentes nos demais cenários e recomendar soluções adequadas às tendências da governança de TI.

O principal motivo pela escolha do GA é a sua representação cromossômica. Basicamente, GAs representam os dados através de um vetor de características conhecido como cromossomo. Os genes deste vetor representam as posições de cada característica e o alelo de um gene representa o valor atribuído ao mesmo. Desta forma, um cromossomo

representa o questionário como um todo, os genes simbolizam cada questão presente no questionário e os alelos descrevem os coeficientes atribuídos às alternativas selecionadas em cada questão.

Além disso, a sua habilidade de combinar as possíveis soluções disponíveis em um processo iterativo torna o algoritmo apto para reproduzir o modelo de tomada de decisão apresentado pelo *IT Governance Institute* (ITGI) [1]. Por meio deste modelo, é possível compreender o método iterativo de identificação de ações a serem aplicadas no cenário organizacional, mediante aos objetivos de TI detectados.

Este trabalho está estruturado da seguinte forma: no capítulo 2 são apresentados os principais conceitos referente à governança de TI, ao *framework* de desenvolvimento de governança de TI, aos *data warehouses* e à inteligência computacional. O capítulo 3 aborda a estrutura do modelo proposto. O capítulo 4 apresenta a modelagem do repositório de conhecimento, os métodos de IC utilizados e seus respectivos resultados. Por fim, o capítulo 5 apresenta a conclusão do trabalho e diretrizes para pesquisas futuras.



## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados os conceitos fundamentais para a compreensão da relação entre governança de TI, o repositório de conhecimento e os algoritmos utilizados para a automação do processo de tomada de decisão. O conteúdo apresentado neste capítulo se delimita a uma descrição básica dos conceitos, com a finalidade de prover um conhecimento prévio para o entendimento do presente trabalho.

Desta forma, o capítulo está organizado da seguinte forma: a seção 2.1 define a governança de TI, esclarece o processo de tomada de decisão e apresenta os domínios que uma boa TI deve atender. A seção 2.2 elucida o *framework* proposto por [2]. A seção 2.3 apresenta os principais conceitos sobre *data warehouse* e sua concepção. Por fim, a seção 2.4 apresenta a noção acerca dos algoritmos de inteligência computacional utilizados neste trabalho.

### 2.1 Governança de tecnologia da informação

De acordo com o ITGI [1], a governança de TI é um conjunto de atividades e disciplinas distribuídas, relacionadas à liderança e à estrutura e os processos organizacionais que objetivam o alinhamento e sustentação dos objetivos e estratégias da empresa por meio da tecnologia da informação. Desta forma, a governança de TI é de responsabilidade da alta administração de uma organização, mas ocorre em todos os níveis.

A governança de TI é um elemento da governança corporativa. De acordo com o Instituto Brasileiro de Governança Corporativa [7], a governança corporativa é um sistema responsável por guiar, monitorar e incentivar as organizações. Esse processo é realizado por meio de práticas como: os relacionamentos entre sócios, o conselho de administração, a diretoria, os órgãos de fiscalização e controle e as demais partes interessadas.

Para garantir uma governança corporativa de qualidade, é necessário assegurar o foco nos seguintes princípios:

- **Transparência:** disponibilizar todas as informações relevantes para as partes interessadas, sem uma pré seleção das mesmas;
- **Equidade:** tratamento justo e equivalente para todas as partes interessadas;
- **Prestação de contas:** prestação de contas da atuação dos agentes de governança de modo claro, conciso, compreensível e tempestivo;
- **Responsabilidade corporativa:** zelar pela viabilidade econômico-financeira e pela qualidade das organizações;

Rodrigues *et al.* [8] afirmam que a TI pode assumir diferentes papéis dentro de uma organização, de acordo com o seu nível de maturidade. O primeiro papel é a TI como fornecedora de serviços, onde são ofertados serviços que objetivam a informação, mas sem influenciar nos processos da organização e em seu desempenho final.

O segundo papel é a TI alinhada ao negócio, onde a TI auxilia a tomada de decisão, envolvendo uma taxa de riscos médios. Este papel é menos comum do que o primeiro. Por último, a TI como diferencial competitivo é um papel muito incomum e que participa ativamente na formulação de estratégias para a organização.

Todas as abordagens citadas por Rodrigues estão relacionadas ao alinhamento da TI com os objetivos da organização e às tomadas de decisão para atingir e evoluir cada objetivo de TI. De acordo com o ITGI [1], o modelo de tomada de decisão em governança de TI é sintetizado através da Figura 1.

O processo se inicia com a definição dos objetivos da TI. Esses objetivos, relacionados aos objetivos da organização, estabelecem um ponto de partida para o processo de aprimoramento da TI. Em seguida, é estabelecido um laço de repetição contínuo. A primeira etapa deste laço é a comparação da atual situação da TI com situações ideais. Esta comparação é realizada através da comparação do cenário da organização com cenários que já obtiveram sucesso.

A partir das comparações realizadas, as direções para o aprimoramento dos objetivos de TI são identificadas. Estas direções estão relacionadas à um conjunto de atividades responsáveis por garantir que os objetivos sejam atingidos. Tais atividades são aplicadas no cenário organizacional de uma corporação, de forma a gerar resultados. Esses resultados podem ser tanto positivos quanto negativos para uma determinada situação. A última etapa do laço de repetição é responsável por avaliar os resultados obtidos e mensurar a performance da TI.

De acordo com Lunardi et al. [9], uma governança de TI eficiente deve focar uma alta performance nos seguintes domínios de TI:

- **Alinhamento estratégico:** alinhamento das estratégias e planos de TI com os objetivos da organização;
- **Valor de TI:** mensura o valor que a TI atribui a uma organização;
- **Gerenciamento de riscos:** gerenciamento e estratégias para contornar eventos prejudiciais à TI e à organização como um todo;
- **Gerenciamento de recursos:** gerenciamento do uso apropriado dos recursos de TI;

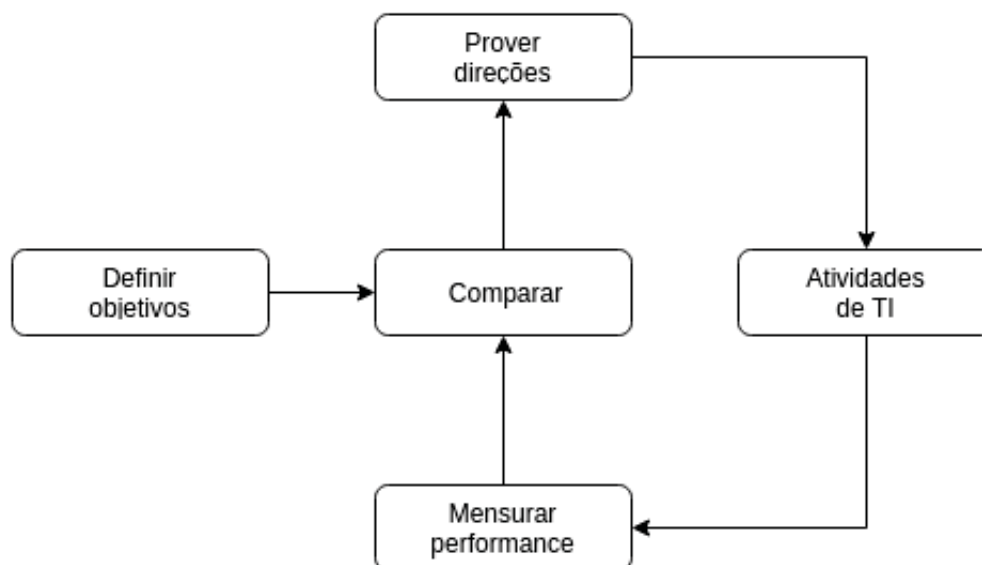


Figura 1 – Processo de tomada de decisão em governança de TI. Adaptado de [1].

- **Medidas de performance:** monitoramento da performance operacional e valor do TI;

Além dos domínios citados anteriormente, o domínio de responsabilidades é de grande importância para a eficiência de uma governança de TI. O domínio de responsabilidades possui o objetivo de definir os papéis e responsabilidades acerca da tomada de decisão de TI e assegurar a sua divulgação para a ciência de toda a organização [10].

A taxa de atendimento da governança de TI nos domínios citados acima define o nível de maturidade da mesma. Para compreender a sua evolução, é necessário adotar um modelo de maturidade responsável por estabelecer o patamar de evolução do processo. Um modelo muito comum é o *Capability Maturity Model* (CMM), onde os níveis de maturidade variam entre: inicial, gerenciado, definido, quantitativamente gerenciado e otimizado [11].

## 2.2 Framework de desenvolvimento de governança de TI

As ações relacionadas à governança de TI possuem um grande nível de complexidade, devido ao grande número de fatores correspondentes à organização como um todo. Assim, a complexidade em aplicar governança de TI é influenciada pelas características e particularidades da organização, o que, em diversas situações, ocasiona falhas e desistências por parte dos envolvidos.

Com base nesta situação, foi proposto por Brigano [2] um *framework* capaz de auxiliar a identificação de áreas com necessidade de melhorias, ou pontos de melhoria, e dos meios para implantar a governança de TI. Neste sentido, o *framework* divide-se em duas partes complementares que, embora correlacionadas, podem ser utilizadas isoladamente.

Essas partes são: (i) o mecanismo de análise e (ii) o modelo de repositório de soluções de governança de TI.

A primeira parte compreende o mecanismo de análise, e seu objetivo é, através de um questionário, analisar o cenário organizacional e posicionar as suas características como uma taxa de atendimento dos domínios de governança de TI. A segunda parte é uma estrutura para o desenvolvimento de um repositório de conhecimento em governança de TI.

A Figura 2 apresenta a visão geral do *framework*. O processo inicia-se com o ambiente organizacional, que fornece as informações sobre a sua TI. Estas informações são extraídas através do questionário de diagnóstico de TI, que destaca os pontos fortes e fracos da governança de TI. As características extraídas alimentam o repositório e um conjunto de soluções são identificadas para o atual cenário. Ao aplicá-las no ambiente organizacional, as soluções geram um conjunto de resultados, que devem ser armazenados também no repositório.

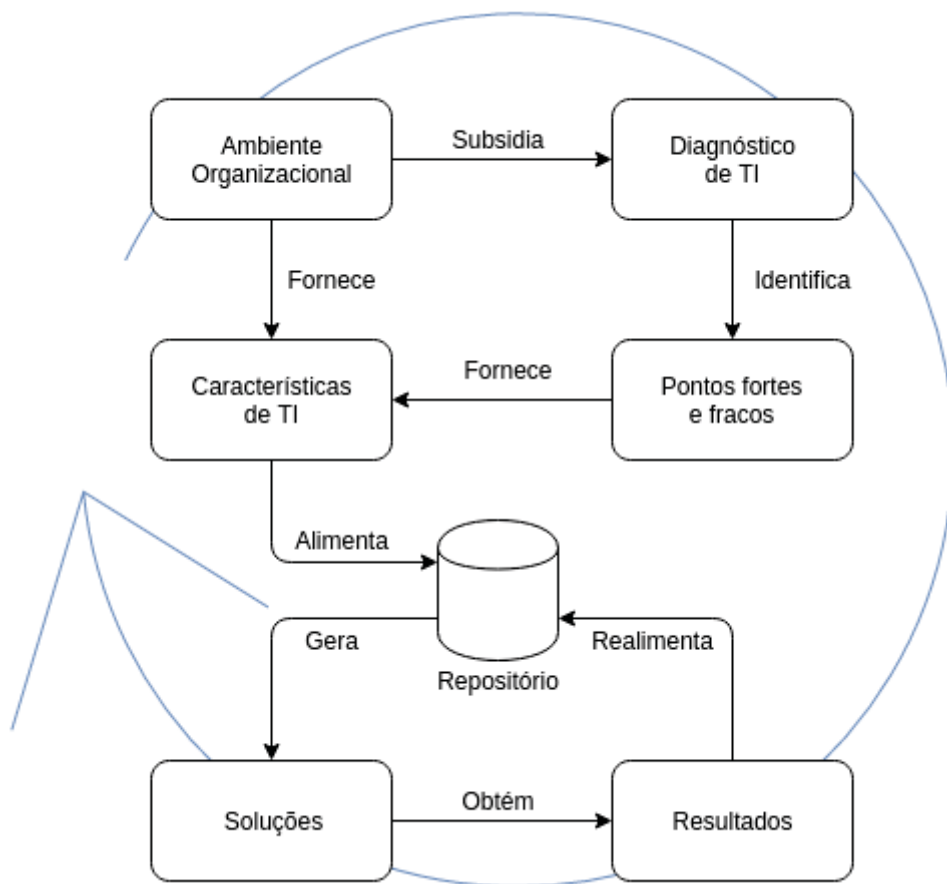


Figura 2 – Visão geral do *framework*. Adaptado de [2].

### 2.2.1 Mecanismo de análise

O objetivo do mecanismo de análise é extrair as características do ambiente organizacional e posicionar a sua taxa de atendimento nos domínios de governança de TI. A extração das características do ambiente organizacional ocorre através de um questionário de avaliação de governança de TI. Este questionário possui 2 questões objetivas de posicionamento do nível econômico da organização e 51 questões de múltipla escolha para avaliação do cenário.

A Figura 3 apresenta um exemplo de uma questão do *framework*. Através desta figura, é possível observar que a questão possui os seguintes elementos: (i) um enunciado, (ii) um conjunto de alternativas de respostas e (iii) o coeficiente relacionado a cada alternativa. O coeficiente representa o grau de aplicação de um determinado item e seu valor varia entre +3 e -3.

Além disso, cada questão possui um relacionamento com os domínios de governança de TI. A Figura 4 apresenta os elementos do relacionamento. Para cada eixo, ou domínio, existem uma justificativa e um peso relacionado. O peso representa o grau de importância do item avaliado para o determinado domínio. Os valores dos pesos variam entre 0 e +4. Maiores detalhes sobre o questionário podem ser encontrados em [2].

53. O suporte aos usuários da organização é eficaz?	
A	O suporte aos usuários é excelente, com alta eficácia na resolução dos problemas e dentro dos prazos determinados.
B	O nível técnico do suporte é excelente, mas, por algumas vezes, os prazos são excedidos.
C	O nível técnico do suporte é bom e os prazos são frequentemente excedidos.
D	O suporte aos usuários é insatisfatório, com dificuldades técnicas para a resolução dos problemas.

Figura 3 – Exemplo de questão de avaliação. Disponível em [2].

Após a obtenção do questionário respondido, é necessário posicionar a sua taxa de atendimento nos domínios de governança de TI. Este processo inicia-se com a equação 2.1, onde o resultado do questionário  $RQ(e)$  é computado. Para isto, é realizada uma soma ponderada entre o coeficiente das alternativas selecionadas em cada questão  $f_i$  e os pesos de cada questão  $p_{i,e}$  relacionados ao domínio  $e$ .

Eixo	Justificativa	Peso
Alinhamento estratégico	O suporte de TI não é um mecanismo que contribui com o alinhamento estratégico da governança de TI.	0
Valor de TI	Por meio de um suporte bem implementado e eficiente pode-se extrair mais valor da TI.	3
Gerenciamento de Riscos	O conhecimento proporcionado pelo suporte pode auxiliar no gerenciamento de riscos do mesmo.	1
Recursos de TI	O bom aproveitamento dos recursos de TI pela organização é diretamente proporcional à qualidade do suporte.	4
Medidas de Performance	Embora não relacionado diretamente, o suporte pode fornecer algumas medidas com relação à qualidade e tempo da TI.	2
Responsabilidades	As responsabilidades da TI não são influenciadas pelo suporte. Estas devem ser definidas em um nível gerencial.	0

Figura 4 – Relacionamento entre questão e domínios de TI. Disponível em [2].

$$RQ(e) = \sum_{i=1}^Q f_i \times p_{i,e} \quad (2.1)$$

Para determinar os limites da faixa de pontuação em relação ao resultado bruto do questionário, é necessário calcular a faixa de pontuação  $FP(e)$  do questionário. A faixa de pontuação é calculada através da diferença entre o melhor e o pior caso possível do questionário, conforme a equação 2.2. Nesta equação,  $max(.)$  é um operador que retorna o maior coeficiente e  $min(.)$  retorna o menor coeficiente de uma questão.

$$FP(e) = \sum_{i=1}^Q max(f) \times p_{i,e} - \sum_{i=1}^Q min(f) \times p_{i,e} \quad (2.2)$$

Em seguida, o fator de ajuste  $FA(e)$  é aplicado ao questionário com a finalidade de estabelecer a relação entre o valor do resultado bruto do questionário com a sua faixa de pontuação em um determinado eixo. Esta operação regula o resultado do questionário para baixo, com base no pior caso do questionário, conforme a equação 2.3.

$$FA(e) = 0 - \sum_{i=1}^Q min(f) \times p_{i,e} \quad (2.3)$$

Por último, a taxa de atendimento do questionário  $TA(e)$  para o domínio  $e$  é calculada através da equação 2.4. Como resultado, a equação retorna um valor entre 0 e 100, representando o percentual de atendimento da TI de uma organização em um determinado domínio.

$$TA(e) = \frac{RQ(e) + FA(e)}{FP(e)} \quad (2.4)$$

## 2.2.2 Modelo de repositório de conhecimento em governança de TI

De acordo com Briganó [2], para facilitar a tarefa de identificação de ações de TI, é necessário centralizar tais ações em uma única base de dados. Desta forma, o repositório de conhecimento em TI centraliza diversas ações de governança de TI e os resultados obtidos ao aplicar uma ação. A Figura 5 apresenta a estrutura do modelo de repositório.

A estrutura do repositório possui um alto nível de abstração e disponibiliza o conhecimento através de itens de repositório, conforme a seguir:

- **Mecanismos de TI:** fornecem a base para a gestão de TI, produzindo informações e visão sobre as funções de TI [12];
- **Melhores práticas:** são práticas isoladas que possuem seus resultados reconhecidos por diversas organizações [13];

- **Tutoriais:** é qualquer tipo de tutorial relacionado com governança de TI [2];
- **Processos:** compreende uma sequência de atividades bem definidas e inter-relacionadas com objetivos específicos [14];
- **Ferramentas:** soluções de *software* utilizadas para executar tarefas complexas e com uma taxa de aplicação e execução lenta [2];
- **Normas e modelos:** são quaisquer regras ou diretrizes escritas para motivar a harmonia, otimização ou segurança de um objeto, materiais ou não materiais, quando se trata de repetir um processo de negócio [15];
- **Treinamentos:** são métodos de capacitação que promovem a renovação do conhecimento dos profissionais diante ao cenário dinâmico da TI [2];
- **Estudo de caso:** responsável por armazenar os relatos e o conhecimento obtido por uma organização ao aplicar uma solução de algum outro item de repositório [2];

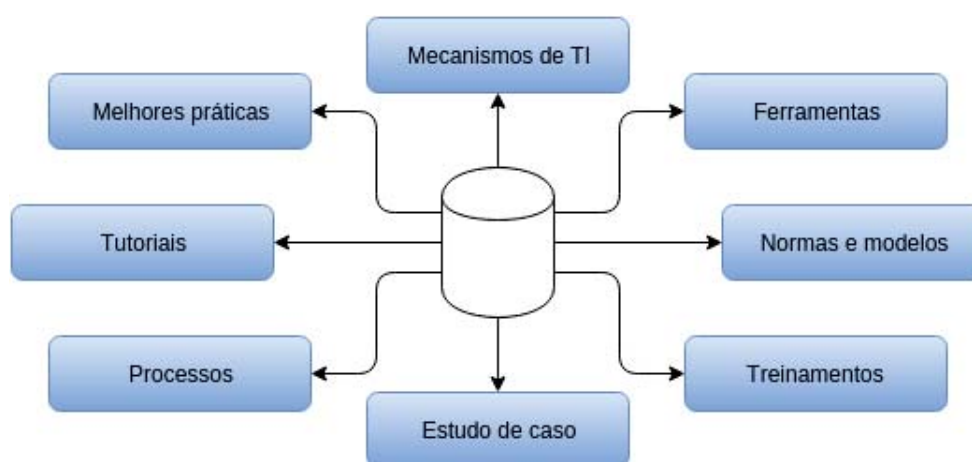


Figura 5 – Estrutura do modelo de repositório. Adaptado de [2].

## 2.3 Data warehouse

*Data warehouse* é um modelo de base de dados responsável por centralizar e sintetizar a informação de fontes remotas e heterogêneas, com a finalidade de suportar a tomada de decisão em um contexto e a análise e compreensão do conhecimento. Desta forma, um *data warehouse* pode ser compreendido como um repositório de conhecimento que visa centralizar as informações de um contexto [16] [17].

Por tratar-se de um modelo multidimensional de dados, um *data warehouse* organiza o seu conhecimento através do conceito de *data cube*. Um *data cube* permite a modelagem e visualização dos dados através de uma estrutura com múltiplas dimensões.

Na prática, esta estrutura é concebida por meio de elementos conhecidos como fatos e dimensões [3].

Um *data cube* é representado por um tema central. Um fato é um objeto mensurável que representa um tema sobre um determinado contexto. As dimensões são entidades que armazenam fragmentos de informações referentes a um tema. Esses fragmentos, em conjunto com um fato central, constituem o conhecimento sobre um determinado contexto. Fatos e dimensões são termos que representam a característica e a funcionalidade que uma Tabela deve assumir.

*Data warehouses* são ferramentas do tipo OLAP. Bases de dados tradicionais, ou *online transaction processing* (OLTP), mapeiam operações cotidianas de um contexto. Diferentemente do OLTP, o OLAP organiza e apresenta o conhecimento de forma a auxiliar a tomada de decisão. Este apoio é possível devido à forma com que os dados são estruturados e às operações OLAP. Exemplos dessas operações são a *drill-down* e a *roll-up* que, respectivamente, aumenta e diminui o nível de abstração de um *data cube* [18].

Han [3] destaca quatro características acerca de *data warehouse*:

- **Orientado a assuntos:** organizar o conhecimento baseado em assuntos;
- **Integrado:** constituído através de múltiplas fontes heterogêneas de dados;
- **Variação com o tempo:** armazenar o conhecimento de forma a prover uma perspectiva histórica da variação dos dados;
- **Não volátil:** o conhecimento retido é apenas acumulado, não sofrendo alterações;

No geral, a arquitetura adotada para um sistema de *data warehouse* possui três camadas: (i) camada de bases de dados, (ii) camada OLAP e (iii) camada *front-end*. A primeira camada consiste em um conjunto de fontes externas de dados. A segunda camada apresenta a estrutura do *data warehouse* e as funções OLAP disponíveis. Por fim, a terceira camada disponibiliza ferramentas de análise e *data mining*.

Ainda do ponto de vista de sua arquitetura, Hajmoosaei, Kashfi e Kailasam e [19] destacam quatro modelos básicos de *data warehouses*:

- **Centralizado:** reúne toda a informação sobre assuntos, abrangendo o contexto como um todo;
- **Data mart:** distribui a informação em *marts* distintos, cada um com a responsabilidade de mapear um determinado assunto sobre o contexto;
- **Distribuído:** integra logicamente uma coleção de dados distribuída em diversas fontes heterogêneas, utilizando o conceito de *views*;

- **Híbrido:** combina as abordagens citadas anteriormente, de acordo com a necessidade;

A fase de modelagem de um *data warehouse* difere-se da modelagem de uma base de dados relacional. Enquanto a modelagem de bases de dados relacionais visa a definição das entidades e o relacionamento entre as mesmas, *data warehouse* requer uma modelagem orientada a assuntos, de forma a facilitar a análise dos dados [3].

No geral, existem três formas de modelagem: (i) estrela, (ii) floco de neve e (iii) constelação de fatos. A Figura 6 apresenta o modelo estrela. Nesta abordagem, o *data warehouse* conta com um fato central relacionado a um contexto e diversas dimensões, que complementam o conhecimento. A normalização das Tabelas não é um fator levado em consideração nesta abordagem.

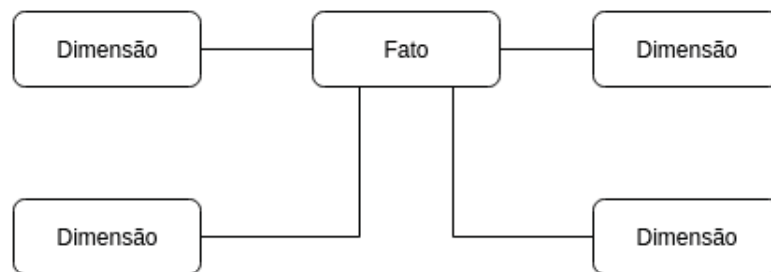


Figura 6 – Modelo estrela. Adaptado de [3].

Na abordagem de modelagem floco de neve, o *data warehouse* conta com um fato central relacionado a um contexto e diversas dimensões, complementando o conhecimento, semelhante à abordagem estrela. A diferença entre as duas abordagens está na possibilidade de normalização das Tabelas de dimensão, conforme a Figura 7.

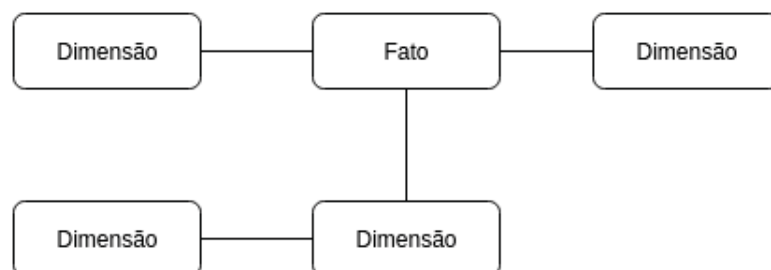


Figura 7 – Modelo floco de neve. Adaptado de [3].

Por último, a Figura 8 apresenta o modelo constelação de fatos. Esta abordagem possui um nível de complexidade maior por permitir que um *data warehouse* possua mais de uma Tabela de fato. Além disso, a normalização das Tabelas de dimensão é um fator permitido nesta abordagem.

## 2.4 Inteligência computacional

De acordo com [4], a Inteligência Computacional (IC) é um campo responsável por estudar mecanismos adaptativos e capazes de simular o comportamento inteligente em ambientes complexos e dinâmicos. Esses mecanismos são providos da inteligência artificial (IA) e sua capacidade adaptativa está relacionada à propensão de generalizar o conhecimento envolvido no contexto. A área responsável por estudar os elementos de aprendizado é chamada de aprendizado de máquina, ou *machine learning* (ML).

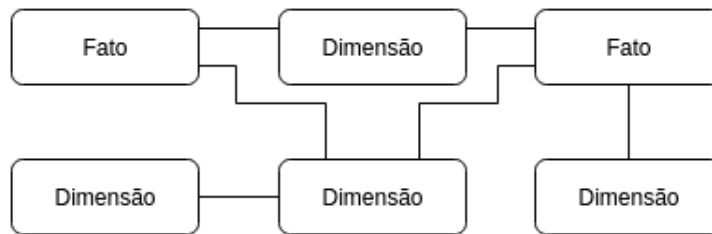


Figura 8 – Modelo constelação de fatos. Adaptado de [3].

Um atributo, ou dimensão, é a unidade básica de dados relacionada à IC. Esta unidade é responsável por representar as características de um contexto. Os atributos podem ser categorizados como [3]:

- **Nominal:** características relacionadas com nomes, representando categorias, estados e rótulos;
- **Binário:** característica nominal que possui apenas duas categorias (0 e 1);
- **Ordinal:** características cujos valores representam uma escala, ordem ou *ranking*;
- **Numérico:** características quantitativas representadas por valores inteiros ou reais;

A Figura 9 apresenta a relação entre os paradigmas relacionados à IC. De forma geral, redes neurais artificiais, ou *neural networks* (NN), são abordagens não lineares inspiradas no processamento paralelo dos neurônios biológicos. Computação evolucionária, ou *evolutionary computation* (EC), mimetiza o processo de evolução natural, seguindo os conceitos de adaptação e sobrevivência. Inteligência de enxame, ou *swarm intelligence* (SI), realiza otimização e agrupamento de dados com base no estudo de organização de colônias.

Sistema imunológico artificial, ou *artificial immune system* (AIS), é uma abordagem que baseia-se no sistema imunológico para identificar células pertencentes a um corpo e células externas. Sistemas *fuzzy* (FS) trabalham com incerteza, não limitando-se apenas às classes de intervalo unitário real  $[0, 1]$ . Por fim, outras abordagens da IA são utilizadas

em conjunto com as abordagens citadas anteriormente para auxiliá-las na resolução de problemas [4].

No geral, as abordagens citadas anteriormente estão associadas, mas não restritas apenas, à resolução das seguintes tarefas:

- **Classificação:** dado um conjunto  $N$  de objetos e  $L$  de classes que representam as características do contexto, a classificação visa identificar a qual classe um dado objeto  $o$  pertence [20];
- **Agrupamento:** dado um conjunto  $N$  de objetos, o agrupamento possui o objetivo de particionar os objetos em  $C$  grupos, de acordo com as suas semelhanças [21];
- **Busca:** dado um espaço de busca, visa localizar a configuração de parâmetros, ou solução ótima para um problema [4];
- **Outros:** demais tarefas, como regressão, análise de *outlier* (anomalias), associação, correlação, caracterização, discriminação, etc;

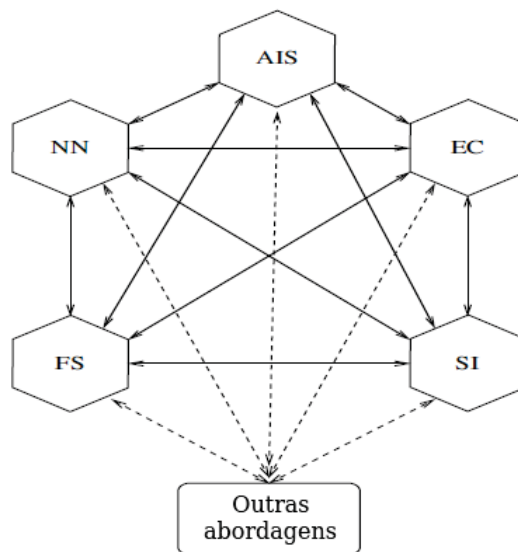


Figura 9 – Paradigmas da inteligência computacional. Disponível em [4].

Nas próximas seções, serão aprofundadas as definições dos campos da IC relacionados ao desenvolvimento do presente trabalho.

#### 2.4.1 *Clustering*

Agrupamento de dados, ou *clustering*, é uma abordagem de aprendizado não supervisionado aplicada em contextos onde as classes dos objetos de um *dataset* não são

conhecidas. Desta forma, seu objetivo é particionar estes objetos e posicioná-los em sub-grupos, chamados *clusters*. Através das suas características, os objetos de um *cluster* (*intracluster*) são similares entre si e dissimilares dos demais objetos (*extracuster*) [3].

De acordo com Paredes e Vargas [22], existem cinco categorias de algoritmos de *clustering*: (i) método de particionamento (ii) método hierárquico, (iii) método baseado em densidade, (iv) método baseado em grade e (v) método baseado em modelos. O método de particionamento procura agrupar um conjunto de dados não rotulados em  $k$  grupos. Um grupo deve possuir, no mínimo, um objeto e os objetos são agrupados de acordo com a sua similaridade.

O método hierárquico realiza o agrupamento de dados através da distribuição dos dados em uma estrutura de árvore, onde cada nível representa a granularidade de agrupamento. No geral, existem dois tipos de estratégias hierárquicas: (i) aglomerativa e (ii) divisiva. Na estratégia divisiva, os objetos são distribuídos no nó raiz e divididos de acordo com a sua similaridade, até que cada objeto seja distribuído em seu próprio *cluster*. A estratégia aglomerativa será apresentada na seção 2.4.1.2.

O método baseado em densidade expande um *cluster* com forma arbitrária até que um limiar de densidade seja alcançado. Para realizar esta tarefa, é necessário definir uma área de vizinhança dos objetos e calcular a densidade de acordo com a quantidade de objetos presentes em cada vizinhança. O método baseado em grade realiza as operações de agrupamento em uma quantidade finita de subdivisões do espaço, conhecida como células. Por último, o método baseado em modelos utiliza métodos estatísticos ou de aprendizado de máquina para generalizar um modelo que descreva cada *cluster*.

#### 2.4.1.1 *K-means*

O *k-means* é um algoritmo de agrupamento de dados por particionamento que possui como função objetivo o propósito de manter a semelhança *intracluster* e a diversidade *intercluster*. *Intracluster* e *intercluster* significam, respectivamente, objetos de um mesmo *cluster* e objetos de *clusters* distintos. Este particionamento é realizado através da distribuição dos objetos em  $k$  *clusters* [3].

As equações 2.5, 2.6 e 2.7 formalizam a definição do *k-means*. O processo de agrupamento mapeia o *dataset*  $X$  para o conjunto de classes  $m$ . O conjunto de *clusters*  $C$  agrupa os objetos com características semelhantes. Cada *cluster* é denotado por  $i$  e este não pode ser um conjunto vazio. A união de todos os *clusters* deve compor o *dataset* original e a intersecção de dois *clusters* deve resultar em um conjunto vazio [23].

$$C_i \neq \emptyset, i = 1, 2, \dots, m \quad (2.5)$$

$$\bigcup_{i=1}^m C_i = X \quad (2.6)$$

$$C_i \cap C_j = \emptyset, i, j = 1, 2, \dots, m; i \neq j \quad (2.7)$$

Além disso, o *k-means* é uma técnica baseada em *centroid*. Um *centroid* é a média dos dados de um *cluster* e descreve as características do mesmo. A similaridade entre os objetos é calculada através de uma distância. A distância habitualmente utilizada é a distância Euclidiana, apresentada na equação 2.8. Nesta equação,  $p$  e  $q$  representam dois vetores de características distintos [3].

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (2.8)$$

O algoritmo 1 apresenta uma visão geral do *k-means*, onde  $k$  é a quantidade de *clusters* e  $D$  é o *dataset*. Primeiramente, o algoritmo atribui objetos aleatórios como centros de *clusters*. Em seguida, ele posiciona cada objeto no *cluster* com o centro mais semelhante, através do cálculo da distância euclidiana entre o objeto e o *centroid*, e atualiza os valores dos *centroids*. Esse processo repete-se até que não ocorra nenhuma mudança nos *clusters*.

**Data:**  $k, D$

**Result:** Um conjunto de  $k$  *clusters*

Escolher arbitrariamente  $k$  objetos de  $D$  como *centroids* iniciais;

**repeat**

    (Re)atribuir cada objeto para o *cluster* mais semelhante, baseado na distância euclidiana;

    Recalcular o *centroid*;

**until** Não haja mudanças;

**Algoritmo 1:** K-Means. Disponível em [3].

#### 2.4.1.2 Agglomerative hierarchical clustering

A estratégia de método hierárquico aglomerativo busca distribuir os objetos em uma estrutura de árvore utilizando o método *bottom-up*. Isso significa que os objetos são primeiramente posicionados nos nós de folha, representando um *cluster* único para cada objeto. Iterativamente, os *clusters* são combinados de acordo com a sua semelhança até que todos os objetos estejam presentes em um único *cluster*, que é o nó raiz. O nó raiz representa o *dataset* como um todo [3].

A Figura 10 apresenta o dendrograma de uma estratégia aglomerativa. O dendrograma é uma forma de representação da distribuição dos objetos na árvore. É possível

observar que cada nível da árvore representa uma granularidade diferente da quantidade de *clusters*. A similaridade entre dois objetos é calculada através de uma equação de distância como, por exemplo, a distância Euclidiana descrita na seção 2.4.1.1.

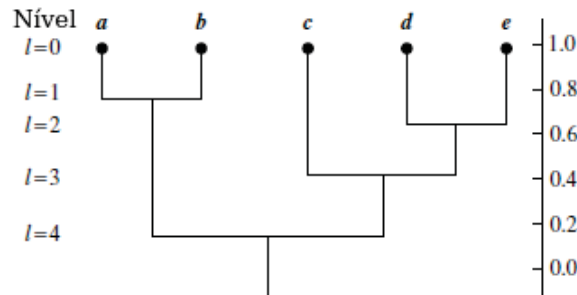


Figura 10 – Exemplo de resultado de uma estratégia aglomerativa. Disponível em [3].

O algoritmo 2 apresenta de forma abstrata o funcionamento desta estratégia, onde  $D$  é um *dataset* a ser agrupado. Primeiramente, inicializa-se um vetor ativo  $A$  que representa os objetos que ainda faltam ser processados. A árvore de agrupamento  $T$  é representada como um vetor que armazena as sequências de combinações entre pares de objetos. O próximo passo é atribuir todos os objetos iniciais de  $A$  em  $T$ , representando os nós folhas e os dados em seus próprios *clusters*.

**Data:**  $D$

**Result:** Uma árvore com os objetos agrupados

Inicializar um conjunto ativo  $A$  com os objetos de  $D$ ;

Inicializar uma árvore  $T$  com  $A$ ;

**while**  $|A| > 1$  **do**

    Escolher o par de objetos de  $A$  com a melhor distância;

    Remover o par de  $A$ ;

    Adicionar a união do par em  $A$ ;

    Adicionar a união do par em  $T$ ;

**end**

**Algoritmo 2:** *Hierarchical agglomerative clustering*. Disponível em [24].

Em seguida, são calculadas as distâncias entre todos os objetos disponíveis em  $A$  (matriz de distância) e é selecionado o par de objetos mais semelhante, ou seja, com menor distância. Este par é removido de  $A$ , e uma representação de união entre os objetos é adicionada em  $A$  e em  $T$ . Este processo repete-se até que  $A$  possua um único elemento representando a união de todos os objetos de  $D$  (nó raiz).

## 2.4.2 Geração de dados sintéticos

A aplicação de algoritmos de IC e ML para a resolução de problemas complexos vem aumentando nos últimos anos. Entretanto, é comum deparar-se com situações onde

as bases de dados disponíveis possuem pouca quantidade de dados ou classes desbalanceadas. Além disso, dados de natureza fraudulenta como, por exemplo, fraude em transações bancárias, ocorrem em pequenas quantidades em diversos *datasets*, dificultando a generalização do conhecimento [25].

A geração de dados sintéticos é um recurso de IC, ML e *Data mining* responsável por reproduzir dados artificiais e possui o propósito de reforçar a quantidade de dados disponíveis e aprimorar a generalização do conhecimento. Os dados artificiais devem ser realistas e fiéis aos dados originais, de forma a intensificar os padrões presentes em uma base de dados [26] [27].

Em problemas de classificação, uma base de dados desbalanceada é aquela onde as classes não possuem uma representação aproximadamente igual. Ou seja, a quantidade de objetos da classe majoritária é extremamente maior em relação à quantidade de objetos da classe minoritária. Desta forma, um modelo de ML não é capaz de extrair o conhecimento de um conjunto de dados com essas características, ocorrendo um "*under-fitting*" [28].

Em geração de dados sintéticos, existem duas técnicas básicas para a produção de objetos artificiais: (i) *under-sampling* e *over-sampling*. Ambos os métodos são baseados na distribuição de amostragens. *Under-sampling* reduz a quantidade de dados disponíveis na classe majoritária, aproximando-a do tamanho da classe minoritária. *Over-sampling* duplica os dados disponíveis na classe minoritária, aproximando-a do tamanho da classe majoritária.

As duas abordagens solucionam o problema de balanceamento de classes. Entretanto, ambas causam desvantagens na generalização do conhecimento. *Under-sampling* pode negligenciar objetos com alta relevância, reduzindo a performance do classificador. No caso do *over-sampling*, o tempo de convergência do modelo pode aumentar, além de causar *over-fitting*, onde o modelo "memoriza" os padrões apresentados na fase de treinamento.

Em algumas situações, é necessário solucionar problemas diferentes da classificação. Como exemplo, pode-se ressaltar a busca por um objeto ótimo, como no caso do presente trabalho. Além disso, diversas situações disponibilizam bases de dados com uma quantidade de objetos disponíveis escassa. Nessa condição, é impossível determinar *a priori* a quantidade de classes relacionadas ao problema.

Para estes problemas, é necessário incrementar a quantidade de objetos para toda a base de dados, sem levar em conta uma determinada classe. O método de *over-sampling* pode não ser uma opção viável, uma vez que apenas as características do dado duplicado são reforçadas. A seção 3.2.2 descreve duas abordagens propostas neste trabalho que procuram reforçar o padrão dos objetos, ao invés de suas características.

### 2.4.3 Algoritmo genético

Algoritmo Genético (GA) é um modelo evolucionário inspirado no processo de seleção natural proposto por Charles Darwin. Neste processo, o indivíduo mais adaptado sobrevive ao meio e é capaz de se reproduzir, mantendo as suas características nas gerações posteriores. Esta abordagem soluciona problemas de busca e otimização [4].

A Figura 11 apresenta o fluxo de execução de um GA. O processo se inicia com a inicialização da população inicial, podendo ser aleatório ou seguindo um processo pre-estabelecido. Em seguida, o algoritmo entra em um estado de *loop*, onde é utilizada uma função *fitness* para calcular a aptidão dos indivíduos. O critério de parada estabelece o limiar do *fitness* de uma solução ótima e verifica se algum indivíduo atingiu este limiar. Caso não haja soluções ótimas em uma determinada geração, ou iteração, os indivíduos da população são submetidos aos operadores genéticos [5].

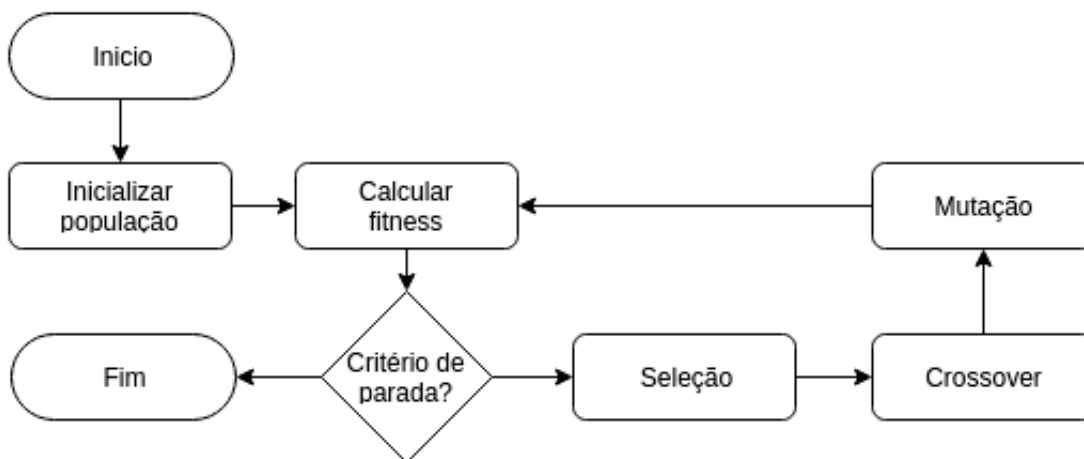


Figura 11 – Fluxograma do funcionamento do GA. Adaptado de [5].

O primeiro operador é a seleção, onde os indivíduos mais aptos são selecionados para a troca de material genético. Em seguida, o operador *crossover* utiliza uma política de corte de cromossomos para realizar uma recombinação, onde é realizada a troca de material genético entre indivíduos. Por último, o operador de mutação é responsável por realizar uma modificação genética aleatória, sutil e com uma baixa taxa de ocorrência. Após o passo de mutação, os processos do *loop* repetem-se até que o critério de parada seja satisfeito e um ou mais indivíduos ótimos sejam identificados.

Cada indivíduo de uma população representa uma potencial solução para o contexto onde aplica-se o GA. Cada solução é representada através de um vetor de características, responsável por descrever as propriedades da mesma. Este vetor de características é conhecido como cromossomo e deve ser devidamente representado através da natureza do problema [4].

O processo de representação do vetor de características é conhecido como representação cromossômica e pode ser definido através das seguintes estruturas: binária, valor real, valor inteiro, permutação ou outros. Além disso, um cromossomo é composto por genes e alelos. Um gene é uma estrutura hereditária e corresponde à posição que uma determinada característica assume no cromossomo. Relacionado ao gene, um alelo representa o valor atribuído à esta característica.

A função *fitness* é responsável por mensurar a qualidade, ou aptidão, de indivíduo em relação ao meio de aplicação. Se o problema em questão for de maximização de função, quanto maior o *fitness* de um cromossomo, mais qualificado este é como solução do problema. Um dos maiores desafios do uso de GA é a definição da função *fitness*, pois cada problema a ser solucionado possui uma formulação matemática exclusiva e, muitas vezes, não conhecida.

O critério de parada deve definir uma regra responsável por garantir que uma solução ótima seja encontrada. Esta regra pode ser definida de acordo com duas vertentes: (i) através da avaliação do *fitness* dos indivíduos ou (ii) por tempo de execução, conhecido também como quantidade de gerações[29].

Em algumas situações, existe mais de um critério de *fitness* a ser avaliado. Uma função *fitness* com múltiplos objetivos  $f(x)$  avalia um vetor de objetivos  $X = x_1, x_2, \dots, x_n$ , onde  $x$  representa cada objetivo relacionado ao problema. Uma forma de lidar com funções com múltiplos objetivos é o método baseado em pesos  $w_i$ . Através da equação 2.9, esta abordagem é capaz de transformar o vetor de objetivos em um único escalar, ao aplicar uma soma ponderada entre os valores de *fitness* e um peso atribuído a cada objetivo. A principal vantagem desta abordagem é a capacidade de destacar a importância de cada objetivo por meio do valor do peso [5].

$$g(x) = \sum_{i=1}^k w_i \times f_i(x) \quad (2.9)$$

De acordo com Guo *et al.* [30], o operador de seleção possui a finalidade de identificar os indivíduos mais aptos para a reprodução. Entretanto, a seleção exclusiva de indivíduos com os maiores valores de *fitness* não é adequada. Um critério capaz de atribuir chances de seleção para os indivíduos menos aptos, proporcional ao seu valor de *fitness*, evita máximos locais. Os métodos de seleção mais comuns na literatura sobre GA são:

- **Seleção por roleta:** a seleção do indivíduo baseia-se na distância entre os valores de *fitness*, onde indivíduos com mais *fitness* possuem limites maiores e, consequentemente, maiores chances de serem selecionados [31];

- **Seleção por torneio:** dois indivíduos são selecionados ao acaso e ambos disputam pelo direito de seleção. O indivíduo mais apto possui uma maior chance de vencer a disputa [32];

O operador de *crossover* utiliza uma política de corte para separar partes dos cromossomos e realizar a troca de material genético de dois indivíduos [30]. Existem dois critérios de corte mais difundidos na literatura, conforme a Figura 12 [5]:

- **Corte simples:** dividir dois cromossomos em duas partes, em posições equivalentes, e alternar os materiais genéticos de ambos;
- **Corte duplo:** dividir dois cromossomos em três partes, em posições equivalentes, e realizar a troca do material genético correspondente à segunda parte do corte;

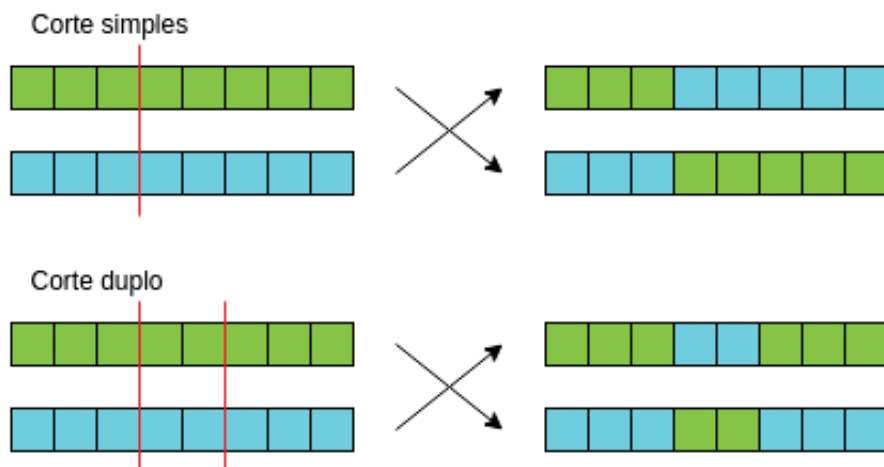


Figura 12 – Esquematização do processo de *crossover* com corte simples e duplo.

Por fim, o operador de mutação aplica modificações aleatórias nos alelos de um cromossomo. O benefício da mutação é a garantia de fuga de um máximo local. Ou seja, caso todos os indivíduos de uma população, em uma determinada geração, sejam representados pelos mesmos alelos, a mutação garante a chance do GA encontrar uma solução ótima. A taxa de ocorrência de mutação deve ser um valor pequeno, geralmente menor que 1% de chance [30].

### 3 DESCRIÇÃO DO MODELO PROPOSTO

Este capítulo é responsável por elucidar a arquitetura do modelo avaliado neste trabalho. Desta forma, a seção 3.1 apresenta uma forma manual de aplicação do *framework* de desenvolvimento de governança de TI de Briganó [2]. Em seguida, é apresentada a arquitetura do modelo proposto na seção 3.2. Esta arquitetura é capaz de automatizar o processo apresentado na seção 3.1.

#### 3.1 Processo de tomada de decisão baseado no *framework* de desenvolvimento de governança de TI

Por meio do processo de tomada de decisão em governança de TI, apresentado na seção 2.1, e do *framework* de desenvolvimento de governança de TI, apresentado na seção 2.2, foi especificada uma forma plausível para a aplicação manual do mecanismo de avaliação de governança de TI para o processo de tomada de decisão. Nesta abordagem, é necessário identificar os pontos de melhoria de um questionário analisando os resultados obtidos previamente em situações semelhantes.

A Figura 13 apresenta o processo de tomada de decisão através do mecanismo de análise de governança de TI [2]. O primeiro passo corresponde à aplicação do questionário em um cenário organizacional, a fim de obter o atual estado da TI de uma empresa e seu respectivo nível de maturidade. O conjunto de taxas de atendimento que uma organização possui em cada domínio de TI representa o seu respectivo nível de maturidade, conforme descrito na seção 2.1.

Em posse de um conjunto de questionários as respostas semelhantes ao obtido, o próximo passo compreende a combinação entre as questões dos questionários selecionados com as do questionário em análise. A finalidade desta tarefa é realizar uma combinação sucessiva entre pares e gerar novas situações, a partir dos padrões contidos nos questionários. Estes padrões são garantidos através dos consenso e das práticas definidas no contexto de governança de TI.

Como resultados do processo de combinação sucessiva, são gerados novos pares a partir dos questionários combinados na iteração anterior. Os novos questionários mantêm as características dos originais, garantindo assim uma variação suave de suas configurações e a geração de cenários semelhantes aos reais. Em uma determinada iteração, o processo de combinações sucessivas pode ou não encontrar o ponto de melhoria para o cenário em análise.

Assim que um ponto de melhoria é identificado, é necessário determinar uma ação

de TI apropriada e aplicá-la no ambiente organizacional, com a finalidade de aprimorar o seu atual nível de maturidade. A identificação de uma ação de TI é um trabalho complexo e exige um alto nível de conhecimento e experiência por parte do especialista responsável. De acordo com Brinanó [2], o desenvolvimento de um repositório de soluções de TI é capaz de auxiliar a execução desta tarefa, uma vez que este elemento é responsável por centralizar diversas soluções em uma única base de dados.

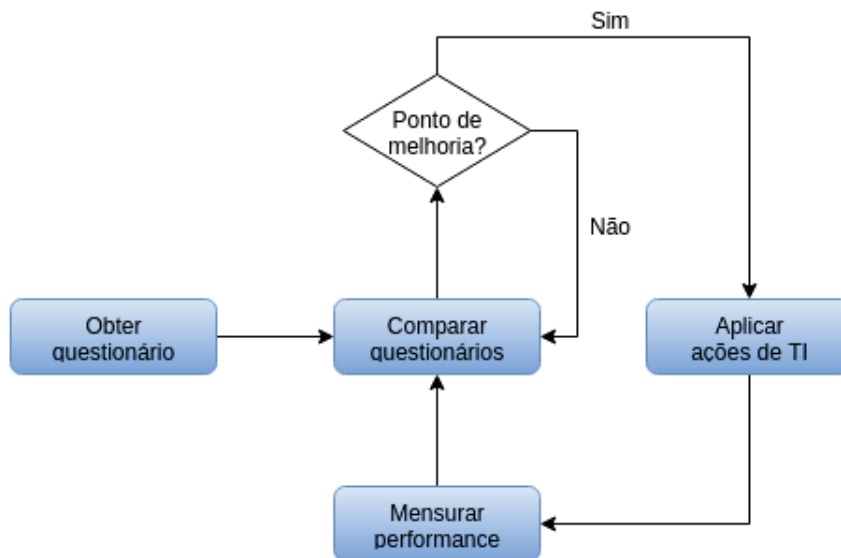


Figura 13 – Processo de tomada de decisão através dos questionários de avaliação de governança de TI.

A aplicação de uma ação de TI no contexto organizacional gera alguns resultados, tanto positivos quanto negativos. Por meio do estado "mensurar performance", os resultados são coletados e analisados, visando verificar se os objetivos foram alcançados. Sempre que um resultado é obtido, o conhecimento sobre o comportamento da tomada de decisão em governança de TI é aprimorado e documentado. Este processo se repete até que todos os objetivos sejam alcançados, ou seja, o nível mais alto de maturidade seja conquistado.

### 3.2 Arquitetura do modelo proposto

De uma forma geral, o objetivo deste trabalho está relacionado à automatização do processo de tomada de decisão apresentado na seção 3.1. A Figura 14 apresenta o modelo proposto neste trabalho. O processo se inicia com a atividade "questionário", que consiste em: (i) aplicar o questionário de avaliação de governança de TI em um cenário organizacional e (ii) computar o nível de maturidade ao questionário obtido.

A atividade "questionário" é executada através do mecanismo de análise, descrito na seção 2.2. O nível de maturidade representa a taxa de atendimento da TI de uma organização em relação aos domínios de TI, citados na seção 2.1. Assim, o nível de maturidade

é uma representação quantitativa do resultado questionário e o questionário é capaz de extrair e representar o conhecimento referente ao atual cenário da TI de uma organização.

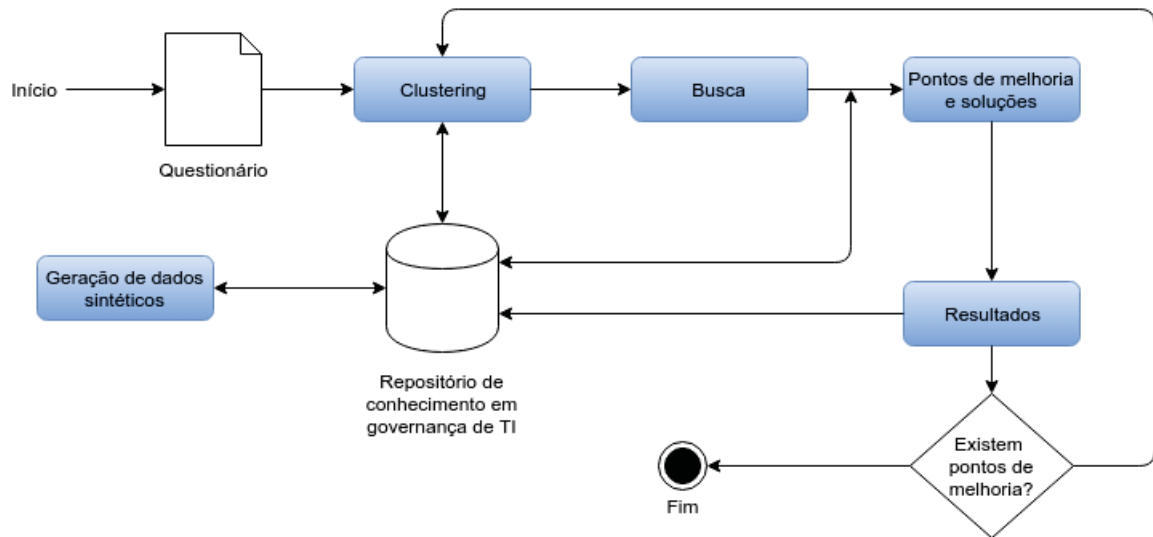


Figura 14 – Arquitetura do sistema de recomendação de pontos de melhoria em governança de TI.

A partir da obtenção de um questionário em análise para a análise, o próximo passo do processo é selecionar respostas semelhantes às obtidas. A atividade "*clustering*" é responsável por realizar esta tarefa em dois passos: (i) selecionar os questionários disponíveis no repositório de conhecimento em governança de TI e (ii) agrupar os questionários, através de um algoritmo de *clustering*, com base em seus níveis de maturidade. Este processo é necessário para levar em consideração o mapeamento entre a variação média das taxas de atendimento dos domínios de TI e a extensão de cada nível de maturidade, conforme será definido a diante.

O objetivo da atividade "*clustering*" é restringir a diversidade dos questionários selecionados. Para que o ponto de melhoria seja encontrado, é necessário restringir o espaço de busca e analisar os padrões disponíveis para a faixa de maturidade do questionário em análise. Isso garante que organizações não sejam otimizadas com base em cenários muito distantes de suas realidades, e assegura uma convergência suave e dentro de suas capacidades.

Através do padrão comportamental presente nos questionários selecionados na atividade de *clustering*, é possível encontrar um ou mais pontos de melhoria para uma determinada situação de TI. O padrão comportamental representa a cultura de desenvolvimento de TI dentro das organizações. A atividade "busca" é responsável por buscar, no espaço de busca disponível, os pontos de melhoria mais propícios para a determinada situação.

O algoritmo selecionado para realizar o processo de busca por pontos de melhoria foi o GA. De forma geral, esta escolha foi realizada devido à sua característica de realizar combinações sucessivas para encontrar uma possível solução de um problema e a sua compatibilidade com o processo de tomada de decisão em governança de TI. Além disso, o GA é um modelo amplamente explorado na literatura, especificamente em problemas de busca e otimização.

Após a identificação dos pontos de melhoria, é necessário definir qual ação, ou solução, de TI deverá ser aplicada ao ambiente organizacional. Desta forma, a atividade "ponto de melhoria e soluções" consiste em um estado no qual o usuário seleciona o ponto de melhoria eleito para ser aprimorado e a solução que será utilizada para esta tarefa. Em seguida, o modelo entra em um estado de espera e aguarda a aplicação da solução no cenário organizacional e, conseqüentemente, a aquisição dos resultados.

As soluções apresentadas ao usuário são ações aplicáveis no contexto de governança de TI, com a finalidade de aprimorar o nível de maturidade da organização, e podem classificadas em quatro grupos:

1. **Soluções recomendadas:** são as soluções utilizadas nos questionários que de fato otimizaram a empresa em análise durante a atividade "busca";
2. **Soluções próximas:** são as soluções empregadas nos demais questionários utilizados na atividade de "busca";
3. **Novas soluções:** são novas soluções cadastradas e, conseqüentemente, utilizadas pelo usuário em seu atual nível de maturidade;
4. **Outras soluções:** demais soluções presentes do repositório;

Depois de aplicar a solução no cenário organizacional, o usuário deve informar os resultados obtidos para o sistema. A atividade "resultados" é encarregada de capturar os resultados quantitativos e qualitativos obtidos, além de permitir que o usuário altere as alternativas das questões influenciadas pela ação aplicada. Isso garante que o novo conhecimento seja coletado e que casualidades nas demais questões sejam identificadas.

Por último, o modelo verifica se o nível de maturidade máximo foi alcançado pela organização. Caso ainda existam melhorias a serem adotadas pela organização, o modelo retorna para o início do processo, na atividade de "*clustering*", e todo o processo descrito se repete até que este critério seja satisfeito.

Devido às características da governança TI e à natureza do *framework* de Brigano [2], a aplicação de uma ação e a avaliação de seu resultado e conseqüências em um cenário organizacional são processos custosos e podem demorar, geralmente, de 6 meses ou mais

para serem realizados. Em consequência deste fato, existe uma certa dificuldade em adquirir dados para popular o repositório de conhecimento em governança de TI e otimizar questionários em análise através dos padrões disponíveis.

Para contornar esta característica, o modelo conta com a atividade "geração de dados sintéticos". Esta atividade é responsável por expandir a quantidade de dados disponíveis, respeitando a similaridade entre os dados originais e sintéticos. Desta forma, este mecanismo é capaz de reforçar os padrões comportamentais presentes no repositório e garantir a qualidade dos dados disponíveis.

A atividade "geração de dados sintéticos" não influencia as características reais do problema em questão. Sempre que um novo padrão comportamental surgir, o modelo levará em conta este novo conhecimento através da atividade de "*clustering*", que utiliza um algoritmo de aprendizado não supervisionado. Ainda, o GA é um modelo autoadaptável, adequando seu conhecimento através da inicialização do algoritmo. Esta característica o difere dos modelos de aprendizado de máquina.

Além disso, caso o mecanismo de geração de dados sintéticos produzir uma grande quantidade de dados que não condizem com a realidade, e um padrão sintético seja criado, o usuário pode analisar e ignorar a recomendação do modelo naquela iteração. Este ato não compromete o funcionamento do modelo, muito menos o seu uso futuro pelo usuário em questão, uma vez que o usuário pode selecionar o ponto de melhoria que deseja aprimorar e a solução que será aplicada. Desta forma, o usuário contribui com o desenvolvimento do conhecimento disponível.

O elemento "repositório de conhecimento em governança de TI" representa uma base de dados responsável por armazenar o conhecimento necessário para a compreensão e tomada de decisão no contexto de governança de TI. A estrutura desta base de dados fundamenta-se nos conceitos de *data warehouse* devido às suas características de centralização e organização do conhecimento, além da possibilidade de aplicação de ferramenta de análise dos dados.

### 3.2.1 Estrutura do repositório de conhecimento em TI

O principal objetivo do repositório é representar de forma coerente o conhecimento referente à governança de TI e mapear o seu processo de tomada de decisão. Desta forma, o repositório deve ser estruturado para auxiliar não somente o modelo de recomendação de pontos de melhoria proposto neste trabalho, mas também a compreensão do comportamento de TI por especialistas e pesquisadores.

A Figura 15 apresenta a estrutura geral do repositório de conhecimento de governança de TI. Para abstrair esta estrutura, foram analisados: (i) os elementos de conhecimento disponíveis em fontes remotas, (ii) elementos não disponíveis, mas que são

necessários para representar o conhecimento e (iii) o relacionamento entre os elementos de conhecimento.

Os elementos de conhecimento são subconjuntos de informações do conhecimento referente à governança de TI. Esses subconjuntos representam informações específicas e são equivalentes aos *data marts*. As informações de um *data mart* podem se situar em fontes de dados remotas e recuperadas através de um processo de aquisição de dados, ou no próprio repositório.

O *data mart* "soluções de TI" organiza-se por meio do modelo de itens de repositório, apresentados na seção 2.2, e é capaz de centralizar e disponibilizar diversos tipos de ações referentes à governança de TI. Estas ações são armazenadas seguindo uma estrutura hierárquica, conforme apresentado no capítulo 4, possibilitando uma maior liberdade para a seleção de soluções.

A "avaliação de TI" possibilita o armazenamento de informação capaz de mensurar e representar o atual estado em que a TI de uma organização se encontra, respeitando uma metodologia de avaliação preestabelecida. O questionário de avaliação de governança de TI, apresentado na seção 2.2, foi selecionado como metodologia de avaliação devido à sua capacidade de representar o nível de maturidade na forma de taxa de atendimento dos domínios de TI.

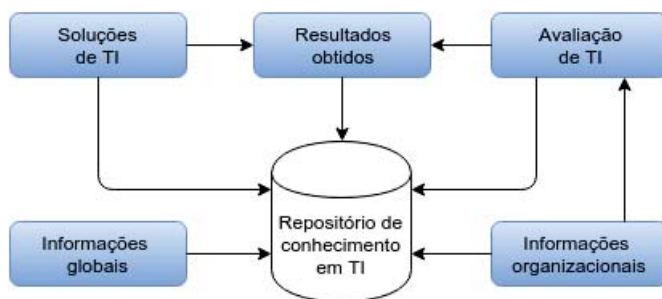


Figura 15 – Estrutura geral do repositório de conhecimento em TI.

De acordo com o modelo de tomada de decisão explicado nas seções 3.1 e 3.2, após a identificação do ponto de melhoria, uma ação de TI é aplicada no cenário organizacional. Esta ação é responsável por gerar resultados de natureza quantitativa e qualitativa. Desta forma, o item "resultados obtidos" é responsável por armazenar tais resultados e mapear a relação entre uma situação de TI e as ações presentes no repositório.

As informações relacionadas à organização, seu faturamento anual e a quantidade de funcionários são mantidas pelo item "informações organizacionais". Além de descrever as características gerais de uma organização, este item é responsável por auxiliar a identificação das avaliações de uma determinada organização.

Por fim, para aumentar o nível de compreensão do comportamento de TI, é neces-

sário avaliar as características econômicas, sociais e políticas em um determinado período. Assim, o item "informações globais" dedica-se em descrever a variação do estado da economia global e relacionar com o restante do conhecimento.

### 3.2.2 Algoritmos de geração de dados sintéticos

Conforme citado na seção 3.2, a aplicação de uma ação e a avaliação de seus resultados em um cenário organizacional são processos custosos e podem demorar, aproximadamente, 6 meses ou mais. Para contornar esta característica, foram propostas duas abordagens de geração de dados sintéticos. No presente trabalho, o problema não se trata de balanceamento de classes minoritárias, mas sim uma solução para expandir a quantidade de dados disponíveis, respeitando a similaridade entre os dados originais e sintéticos e reforçando os padrões presentes no repositório.

A primeira abordagem, apresentada no algoritmo 3, foi intitulada "geração de dados sintéticos determinístico", devido à presença de um padrão de mudanças preestabelecido. O algoritmo é inicializado com os seguintes parâmetros:

- **Dados originais:** estrutura com os dados originais que serão utilizados para a geração de dados sintéticos;
- **Quantidade de dados sintéticos:** a quantidade necessária de dados sintéticos a gerar;
- **Taxa de mudança:** um valor externo entre 0 e 100 para determinar a chance de ocorrer uma mudança;
- **Alfabeto:** estrutura com os valores que cada posição dos dados sintéticos pode possuir;
- **Padrão:** arranjo com a regra de mudança que um conjunto de dados sintéticos devem seguir;

O primeiro passo do algoritmo 3 é verificar a existência de um padrão de mudanças. Caso não exista, o algoritmo deve gerar um padrão aleatório e fixo para um determinado conjunto de dados originais. Este padrão é um arranjo constituído pelos seguintes valores:

- **Valor 1:** o algoritmo deve aumentar o valor da posição atual, de acordo com o alfabeto;
- **Valor 0:** o algoritmo deve manter o valor da posição atual;
- **Valor -1:** o algoritmo deve diminuir o valor da posição atual, de acordo com o alfabeto

Em seguida, o algoritmo seleciona e replica aleatoriamente  $i$  dados originais, onde  $i$  é um contador que itera de 1 até a quantidade de dados sintéticos desejados. Seguindo o padrão e a taxa de mudanças, este processo garante a alta taxa de similaridade entre os dados sintéticos e originais, uma vez que, para a maior parte de um dado sintético (arranjo), serão mantidos os valores iguais aos originais.

Por último, o algoritmo percorre os  $i$  dados replicados. Para cada posição  $j$  de um dado, é gerado um valor aleatório entre 0 e 100, com a finalidade de compará-lo com a taxa de mudança. Caso este valor seja menor ou igual à taxa de mudança, a posição é alterada, seguindo o padrão preestabelecido. Esta característica garante que os dados sintéticos sejam alterados o mínimo possível, mantendo a semelhança com os dados originais.

A segunda abordagem, intitulada "geração de dados sintéticos Tabela", é um modelo semelhante ao algoritmo 3, diferindo apenas na forma com que as mudanças ocorrem. De acordo com o algoritmo 4, determinam-se as alterações em um dado de acordo com o acaso. Assim, os parâmetros de inicialização do algoritmo permanecem os mesmos, com exceção do parâmetro de padrão de mudanças, que é ausente nesta abordagem.

```

Data: dados originais, quantidade de dados sintéticos, taxa de mudança,
         alfabeto, padrão
Result: conjunto de dados sintéticos
if padrão == NULL then
  | gerar um padrão aleatório
end
sintéticos = NULL
for i em 1 até quantidade de dados sintéticos do
  | replicar aleatoriamente os dados originais em sintéticos
end
for i em 1 até quantidade de linhas de sintéticos do
  | for j em 1 até quantidade de colunas de sintéticos do
  | | if aleatório entre 0 e 100 <= taxa de mudança then
  | | | if padrão[j] == 1 then
  | | | | sintético[i,j] desloca uma alternativa para cima (de acordo com
  | | | | o alfabeto)
  | | | end
  | | | if padrão[j] == -1 then
  | | | | sintético[i,j] desloca uma alternativa para baixo (de acordo com
  | | | | o alfabeto)
  | | | end
  | | end
  | end
end

```

**Algoritmo 3:** Algoritmo de geração de dados sintéticos determinístico

O primeiro passo da geração de dados sintéticos Tabela é a replicação aleatória de  $i$  dados originais, onde  $i$  é um contador que itera de 1 até a quantidade de dados sintéticos

desejados. Assim como no processo anterior, este passo garante a similaridade entre os dados sintéticos e originais.

Em seguida, para cada posição  $j$  de um dado, é gerado um valor aleatório entre 0 e 100, com a finalidade de compará-lo com a taxa de mudança. Caso este valor seja menor ou igual à taxa de mudança, a posição é alterada de forma aleatória, estabelecendo uma maior diversidade entre os dados gerados. Os detalhes sobre a implementação dos algoritmos 3 e 4 são apresentados nos apêndices A e B.

### 3.2.3 Características do algoritmo genético

A aplicação do GA no contexto de governança de TI e do *framework* de desenvolvimento de governança de TI possui o objetivo de buscar um ponto de melhoria para aprimorar um questionário em análise. Um ponto de melhoria compreende a alteração de uma alternativa em uma determinada questão. Esta alternativa é capaz de aprimorar a taxa de atendimento de uma organização em um ou mais domínios de TI, sem que haja um decremento nos demais.

```

Data: dados originais, quantidade de dados sintéticos, taxa de mudança,
         alfabeto
Result: conjunto de dados sintéticos
sintéticos = NULL
for  $i$  em 1 até quantidade de dados sintéticos do
  | replicar aleatoriamente os dados originais em sintéticos
end
for  $i$  em 1 até quantidade de linhas de sintéticos do
  | for  $j$  em 1 até quantidade de colunas de sintéticos do
  | | if aleatório entre 0 e 100  $\leq$  taxa de mudança then
  | | | if aleatório entre 0 e 1  $==$  0 then
  | | | | sintético[ $i,j$ ] desloca uma alternativa para cima (de acordo com
  | | | | o alfabeto)
  | | | else
  | | | | sintético[ $i,j$ ] desloca uma alternativa para baixo (de acordo com
  | | | | o alfabeto)
  | | | end
  | | end
  | end
end

```

**Algoritmo 4:** Algoritmo de geração de dados sintéticos Tabela

Logo, o algoritmo deve encontrar as tendências (padrões) de pontos de melhoria presentes em questionários semelhantes ao em análise e testar a sua aplicabilidade. Uma das vantagens em aplicar o GA é a sua capacidade de gerar possíveis soluções, testá-las e, caso uma solução seja encontrada, encerrar a execução do algoritmo através de um critério de parada.

Outra vantagem do GA é a sua capacidade em encontrar soluções em problemas com o espaço de busca extenso e que não possuem algoritmos eficientes para encontrar tal solução. O espaço de busca do problema levantado no presente trabalho são as configurações do questionário de avaliação de governança de TI, ou seja, as possíveis combinações entre suas alternativas. A Tabela 1 apresenta a quantidade de alternativas por questão. Através desses dados é possível calcular o espaço de busca do problema. O total de combinações entre as alternativas do questionário é  $7^1 \times 6^0 \times 5^3 \times 4^{33} \times 3^{12} \times 2^2 \times 1^0 = 1,372469856 \times 10^{29}$ .

A aptidão de um questionário é definida pelo cálculo da taxa de atendimento, apresentada na seção 2.2. Para encontrar um ponto de melhoria, a formalização do problema define dois objetivos: (i) aprimorar um ou mais domínios de TI e (ii) evitar que os demais domínios sofram um decremento. Desta forma, trata-se de um problema multiobjetivo e optou-se por utilizar a função *fitness* baseada em pesos.

Tabela 1 – Quantidade de alternativas por questão

Quantidade de alternativas	Quantidade de questões
7	1
6	0
5	3
4	33
3	12
2	2
1	0

A função *fitness* baseada em pesos possibilita o cenário onde o usuário é capaz de selecionar o domínio que o mesmo deseja aprimorar. Entretanto, devido ao tempo disponível e à complexidade envolvida neste problema, foi decidido tratar os diferentes objetivos com o mesmo peso neste trabalho, desconsiderando a aplicação de maior importância para um determinado domínio.

Para garantir que uma solução com a menor quantidade de alterações seja encontrada, o critério de parada do GA segue a seguinte regra: "encerrar a execução quando a população apresentar um indivíduo com o *fitness* maior que o *fitness* do questionário em análise". De outra forma, o algoritmo poderia encontrar soluções muito distantes ao cenário do questionário em análise.

Além disso, a população inicial é composta por indivíduos com o *fitness* menor ou igual à do questionário em análise. Isso, em conjunto com a fase de *clustering*, garante a semelhança entre os indivíduos, e a convergência do algoritmo com base nas tendências e padrões apresentados pelos demais questionários selecionados. Ainda, este processo assegura o funcionamento do critério de parada, uma vez que nenhum indivíduo da primeira geração possui o *fitness* maior que o *fitness* do questionário em análise.

A representação cromossômica é a forma de representar um problema para o GA. Neste contexto, uma população é um conjunto de questionários selecionados através de um critério preestabelecido. Para representar os questionários, optou-se por representar um cromossomo através de um arranjo de inteiros. Cada gene é a alternativa selecionada em uma determinada questão. Por fim, o alelo é o valor do coeficiente da alternativa selecionada.

Nessa seção, foram apresentadas as características gerais do GA. Entretanto, as características relacionadas aos experimentos realizados, como os operadores genéticos de seleção, *crossover* e mutação, serão abordadas no capítulo 4.



## 4 METODOLOGIA DE PESQUISA E RESULTADOS

Este capítulo descreve a metodologia de pesquisa empregada para a concepção e validação dos elementos do modelo apresentado no capítulo 3, assim como os resultados obtidos. Desta forma, o capítulo se organiza da seguinte maneira: a seção 4.1 aborda a concepção do repositório de conhecimento em governança de TI; a seção 4.3 debate sobre a validação dos algoritmos de geração de dados sintéticos; a seção 4.2 trata a validação dos métodos de agrupamento de dados; e a seção 4.4 discute a validação do algoritmo de busca pelo ponto de melhoria em governança de TI.

### 4.1 Repositório de conhecimento em TI

Em seu trabalho, Briganó [2] levantou a importância de uma estrutura de armazenamento de conhecimento fortemente relacionada com o *framework* de desenvolvimento de governança de TI. Estes elementos, embora correlacionados, podem ser utilizados isoladamente. Esta afirmação é comprovada devido à descrição abstrata e, conseqüentemente, na ausência de um detalhamento sobre a estrutura do repositório de conhecimento por Briganó, sem que seus resultados fossem comprometidos.

No presente trabalho, o repositório de conhecimento em governança de TI possui o propósito de centralizar e representar o conhecimento referente à governança de TI e ao seu processo de tomada de decisão. Isso garante a formalização do conhecimento, auxilia a compreensão do contexto e possibilita a automação do processo através de mecanismos de inteligência computacional.

Para idealizar o repositório de conhecimento em governança de TI, optou-se por fundamentar esta base de dados em conceitos relacionados à *data warehouse*. Este fato garante as seguintes vantagens:

- Centralizar informações de fontes remotas em uma única base de dados;
- Manter séries históricas com a finalidade de possibilitar a análise da variação do conhecimento com o passar do tempo;
- Estruturar e organizar o conhecimento relevante em uma forma consistente e coerente (qualidade dos dados);
- Facilitar a análise do conhecimento e a aplicação de ferramentas de OLAP e de data mining;

### 4.1.1 Metodologia

A elaboração do repositório de conhecimento em governança de TI requer duas etapas fundamentais: (i) análise dos descritores de governança de TI e (ii) análise e modelagem das entidades (dimensões e fatos) e relacionamentos. A análise dos descritores possui o objetivo de identificar as informações que caracterizam uma governança de TI e suas soluções, e que o armazenamento no repositório possua relevância ao contexto.

A análise dos descritores de governança de TI é crucial devido ao fato de que essas informações não são oriundas de uma base de dados remota. Caso estas informações fossem obtidas de uma fonte remota, seria necessário apenas mapear as entidades e relacionamentos relevantes. Além disso, esta atividade define as métricas de TI. Essas métricas são capazes de mensurar a qualidade de uma governança.

O levantamento dos descritores foi realizado através de uma revisão literária. Nesta revisão, foram analisados 15 artigos, relacionados ao estudo de aplicação de ações e soluções em governança de TI, e identificados os fatores que sofreram influências, tanto positivas quanto negativas, a partir de tal aplicação. A quantidade de artigos analisados foi definida com base na complexidade da investigação.

Em seguida, foi formulado um questionário com a finalidade de validar os fatores encontrados com base no conhecimento especialista. Desta forma, o questionário possui 28 questões relacionadas a cada fator. Cada questão procura avaliar o grau de relevância do fator em mensurar as características de TI de forma genérica. Todas as questões possuem cinco alternativas e avaliam a influência do fator de muito baixo até muito alto. Para a validação, foram selecionados 10 especialistas na área para responder o questionário. Os detalhes sobre o questionário estão presentes no apêndice C.

Por último, a análise e modelagem das entidades e relacionamentos compreende em verificar como o conhecimento será representado no repositório, quais informações são pertinentes e como as informações se relacionam entre si. Como resultado, foi esquematizado um diagrama de entidade-relacionamento que apresenta a estrutura do repositório. Este modelo baseia-se em modelagem multidimensional constelação de fatos, oriunda dos conceitos de *data warehouse*.

### 4.1.2 Resultados e discussão

A Tabela 2 apresenta os resultados do levantamento de descritores de governança de TI e sua validação através do conhecimento especialista. Desta forma, a primeira coluna da Tabela apresenta os descritores identificados, a segunda coluna apresenta os artigos que citaram a importância dos fatores e a terceira coluna apresenta o grau de relevância médio, obtido através da aplicação do questionário para os especialistas.

É possível constatar que alguns fatores são amplamente reconhecidos na literatura,

como o caso dos fatores "performance" e "riscos". Porém, existe uma escassez de análise para outros fatores, como o caso da "demanda" e da "usabilidade". Entretanto, a relevância de cada fator é comprovada através da média dos resultados do questionário aplicado aos especialistas em governança de TI.

Tabela 2 – Resultados da análise de descritores de TI.

<b>Fator</b>	<b>Artigos encontrados</b>	<b>Relevância</b>
Alinhamento estratégico	[33], [34]	Alto
Aprimoramento de produtividade	[35]	Médio
Aprimoramento de serviço	[36], [35]	Médio
Benefícios	[37]	Médio
Conhecimento	[35], [33], [38], [39], [40], [41]	Alto
Custo	[42], [43], [33], [41], [37]	Alto
Demanda	[44]	Alto
Economia	[45], [44], [41]	Alto
Estrutura de TI	[42], [36], [43], [33], [38], [34], [44], [37]	Médio
Forma de capital	[45], [44], [38], [40]	Alto
Investimento	[35]	Médio
Inovação	[33]	Alto
Nicho de aplicação	[45]	Alto
Performance	[42], [35], [33], [38], [34], [39], [44], [40], [46], [41], [37]	Alto
Prioridades de TI	[33]	Alto
Recursos de TI	[33], [38], [34], [41], [44]	Alto
Requisitos	[44]	Alto
Restrições	[44]	Alto
Riscos	[42], [33], [38], [34], [44], [40], [46], [41], [37]	Alto
Satisfação	[35]	Médio
Segurança	[42], [41]	Alto
Tamanho da organização	[44]	Alto
Tempo	[45], [42], [36], [35], [43], [38], [39], [33], [41], [44]	Alto
Tempo da organização no presente nível	[40]	Alto
Tempo da organização	[40]	Alto
Usabilidade	[35]	Alto
Valor de TI	[41], [33], [37]	Alto

O cálculo da média das relevâncias relacionadas aos fatores revelam que, no geral, os descritores identificados possuem uma alta relevância no contexto. Desta forma, optou-se por utilizar todos os fatores para descrever o contexto de governança TI e suas soluções no repositório de conhecimento.

A Tabela 3 apresenta o relacionamento entre os fatores identificados e seus respectivos significados dentro do contexto. Além disso, são apresentados os tipos de atributo associado à cada descritor. Os tipos de atributos variam entre: (i) atributos nominais, (ii) atributos ordinais e (iii) atributos numéricos.

A Figura 16 apresenta o relacionamento entre os fatos do repositório. Através

Tabela 3 – Descrição dos fatores e tipos de atributo.

<b>Fator</b>	<b>Descrição</b>	<b>Tipo de atributo</b>
Alinhamento estratégico	Fatores que tornam os objetivos da TI correspondentes aos objetivos da organização	Nominal
Aprimoramento de produtividade	Fatores da solução que aprimoram a produtividade dos envolvidos na TI e na organização como um todo	Nominal
Aprimoramento de serviço	Fatores que auxiliam o desenvolvimento de novos serviços e a preservação de serviços já existentes de forma eficaz	Nominal
Benefícios	Conjunto de fatores positivos que a ação proporciona	Nominal
Conhecimento	Conhecimento necessário, por parte dos envolvidos, para implantar e manter a ação	Nominal
Custo	Custo necessário para implantar e manter a solução	Numérico
Demanda	Descrição do cenário aonde a ação se encaixa	Nominal
Economia	Fatores que descrevem a economia da organização, do país e a economia mundial	Nominal
Estrutura de TI	Forma na qual a tomada de decisão da TI é executada (centralizada, descentralizada, federal e outros)	Nominal
Forma de capital	Formas de bens econômicos que a solução suporta	Nominal
Investimento	Quantidade de verba aplicada para o desenvolvimento da TI da organização	Numérico
Inovação	Fatores que diferenciam uma ação das demais	Nominal
Nicho de aplicação	Restrição de área de aplicação de uma solução	Nominal
Performance	Fatores de desempenho da ação	Nominal
Prioridades de TI	Prioridades da TI de uma organização suportadas pela solução	Nominal
Recursos de TI	Recursos de TIC necessários para a aplicação da solução	Nominal
Requisitos	Requisitos gerais para aplicar a solução	Nominal
Restrições	Restrições de aplicação da ação	Nominal
Riscos	Riscos envolvidos na aplicação de uma ação	Nominal
Satisfação	Grau de satisfação de uma organização ao aplicar uma determinada solução	Numérico
Segurança	Fatores de segurança da informação relevantes para a aplicação da ação	Nominal

Tamanho da organização	Quantidade de funcionários da organização	Numérico
Tempo	Tempo necessário para aplicar a ação	Numérico
Tempo da organização no presente nível	Tempo em que a organização se encontra em um determinado nível de maturidade	Numérico
Tempo da organização	Quantidade de tempo em que a organização existe	Numérico
Usabilidade	Fatores de usabilidade da solução	Nominal
Valor de TI	Fatores que agregam valor ao TIC de uma organização	Nominal

deste relacionamento, é possível notar que os fatos do repositório referem-se aos principais assuntos do contexto, sendo eles: (i) questionário de avaliação de governança de TI, (ii) soluções de TI, (iii) informações sobre a organização e (iv) descrição da economia global. Além disso, é possível observar a presença de uma dimensão de tempo. Esta dimensão é capaz de armazenar todos os fatores relacionados ao tempo.

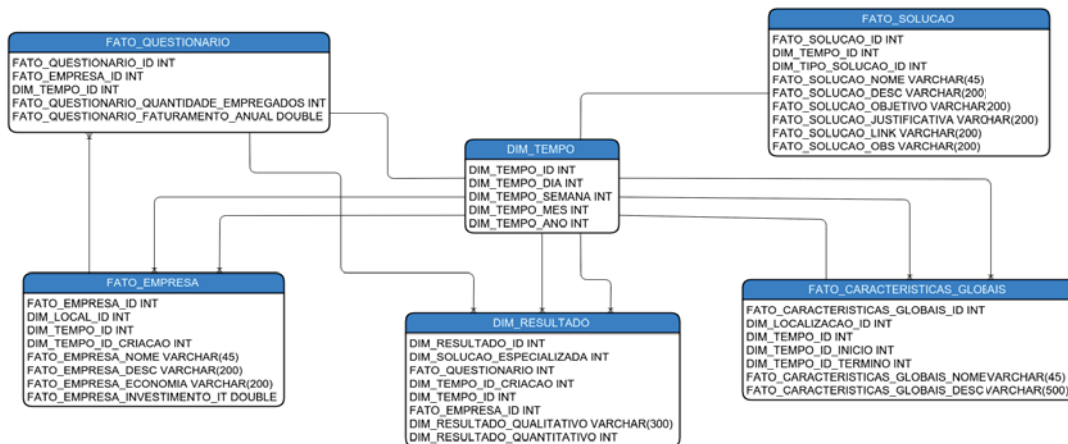


Figura 16 – Relacionamento entre os fatos do repositório.

A dimensão de resultados é responsável por constituir o conhecimento disponível sobre a tomada de decisão no contexto de governança de TI. Através desta dimensão, são associados a descrição do atual cenário organizacional (questionário e nível de maturidade), as informações sobre a organização, com a finalidade de identificar a mesma, e a solução aplicada para garantir a melhoria do cenário organizacional. Ademais, a dimensão de resultados armazena os resultados qualitativos e quantitativos da aplicação da solução em um determinado cenário, relacionando-se com os fatores de performance e satisfação.

O fato "organização" armazena a descrição de sua economia e o total de investimentos para o setor de TI, respectivamente referente aos fatores "economia" e "investimento". Além disso, este fato possui um relacionamento adicional com a dimensão de tempo, responsável por identificar a data de criação da organização.

Além de mapear toda a estrutura do questionário de avaliação de governança de TI e o nível de maturidade associado, o fato "questionário", Figura 17, possui dois atributos especiais: (i) faturamento anual e (ii) quantidade de funcionários. Esses atributos estão diretamente associados, respectivamente, com os descritores "investimento" e "tamanho da organização".

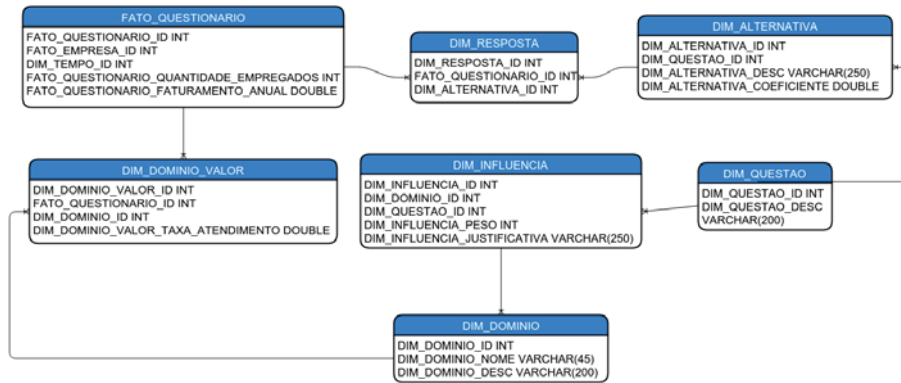


Figura 17 – Mapeamento do questionário pelo repositório.

A estrutura de mapeamento do questionário é capaz de armazenar um conjunto de questões e suas respectivas alternativas. Os domínios de governança de TI também são mantidos nessa estrutura. Cada domínio é associado ao valor alcançado no mesmo por um determinado questionário, representando a taxa de atendimento para aquele domínio. Além disso, cada questão está associada aos domínios disponíveis, mapeando a influência do objeto de análise da questão com o domínio. As respostas de um questionário é armazenada através de uma Tabela associativa, responsável por estabelecer uma relação N-N entre os questionários e as alternativas.

Por último, a Figura 18 apresenta a abstração da estrutura de soluções do repositório. O nível mais alto desta abstração é a dimensão "tipo de solução", responsável por mapear o itens de repositório, apresentados na Figura 5. Em seguida, a dimensão "solução" apresenta ações de governança de TI de forma genérica, sem detalhar uma solução específica. Por fim, a dimensão "solução especializada" disponibiliza implementações particulares para escolha do usuário.

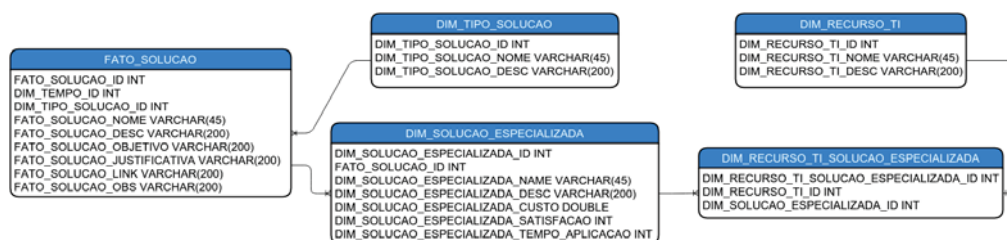


Figura 18 – Mapeamento das soluções pelo repositório.

A dimensão recursos de TI mantém os diferentes tipos de recursos necessários para a aplicação de uma determinada solução. Esta dimensão, em conjunto com a sua Tabela associativa, está relacionada ao fator "recursos de TI". Os demais fatores não citados foram mapeados seguindo a mesma relação N-N da dimensão "recursos de TI".

## 4.2 Agrupamento de dados

Basicamente, *clustering* é uma técnica de agrupamento de dados contidos em um *dataset*, de forma que objetos organizados em um mesmo *cluster* tendem a ser similares entre si e diferente dos objetos organizados em outros *clusters* [47]. A aplicação de técnicas de *clustering* no presente trabalho objetiva o agrupamento de questionários através dos níveis de maturidade, conforme elucidado na seção 2.1.

Para alcançar este objetivo, foram analisados dois modelos clássicos de algoritmos de agrupamento de dados: (i) *k-means* e (ii) *hierarchical clustering*. O agrupamento de dados no *k-means* ocorre através da aproximação dos objetos de pontos centrais de um *cluster*, conhecidos como *centroids*. Esses *centroids* são gerados aleatoriamente e adaptados ao centro dos dados do grupo, conforme os objetos são posicionados no *cluster*.

No caso do *hierarchical clustering*, os objetos são distribuídos em forma de uma árvore de acordo com a sua similaridade. Cada nível da árvore representa uma quantidade de *clusters*, e cada subárvore de um elemento representa os objetos do *clusters*. Nessa seção, será apresentada a metodologia aplicada para validar o agrupamento de questionários e seus respectivos resultados.

### 4.2.1 Metodologia

A validação do processo de agrupamento de questionários visa identificar o algoritmo com melhor capacidade de agrupamento, seguindo o modelo de maturidade estipulado pelo CMM [11]. Através deste modelo, os algoritmos devem agrupar os dados em 5 *clusters*, conforme a seguir:

- **Nível inicial:** média das taxas de atendimento dos domínios de TI entre 0% e 19.999%;
- **Nível gerenciado:** média das taxas de atendimento dos domínios de TI entre 20% e 39.999%;
- **Nível definido:** média das taxas de atendimento dos domínios de TI entre 40% e 59.999%;
- **Nível gerenciado quantitativamente:** média das taxas de atendimento dos domínios de TI entre 60% e 79.999%;

- **Nível otimizado:** média das taxas de atendimento dos domínios de TI entre 80% e 100%;

A validação do processo de agrupamentos de questionário consiste em verificar se os objetos de um determinado *cluster* possuem a mesma faixa de nível de maturidade. Desta forma, foram agrupados apenas os questionários da base de dados original, que possui apenas 25 objetos. A baixa quantidade de objetos disponíveis facilita a análise dos resultados sem que haja algum tipo de influência nos mesmos, uma vez que os questionários estão bem distribuídos nas faixas de níveis de maturidade.

O algoritmo *hierarchical clustering* não baseia-se na inicialização aleatória de centros, mas sim na distribuição dos dados em uma estrutura de dados, de acordo com a similaridade dos dados. Desta forma, este modelo sempre forma *clusters* com os mesmos objetos. Portanto, foi necessário executar esta abordagem uma única vez e analisar os resultados de agrupamento obtidos através de uma matriz de confusão.

Por outro lado, o algoritmo *k-means* é baseado na inicialização aleatória dos centros e na adaptação e deslocamento dos mesmos, de acordo com os objetos do *cluster*. Assim, o algoritmo pode gerar grupos com objetos diferentes a cada iteração. Desta forma, a validação deste modelo consiste em executar 10 iterações do *k-means* e calcular as respectivas precisões através de uma matriz de confusão. Através da média das precisões obtidas, é possível computar a precisão geral do modelo.

A matriz de confusão, ou matriz de erro, é um modelo de validação responsável por apresentar a quantidade de acertos e erros do modelo. Esta matriz é  $n \times n$ , onde  $n$  é o número de classes, ou agrupamentos, envolvidos [48]. Além disso, as seguintes variáveis da matriz de confusão são relevantes para calcular a precisão dos algoritmos de agrupamento:

- **Verdadeiro positivo (*true positive* - TP):** se a instância pertence a uma classe e é classificada como tal;
- **Falso positivo (*false positive* - FP):** se a instância pertence a uma classe e é classificada com outra classe;

As linhas da matriz representam a classe em análise, e as colunas representam os resultados obtidos através de um modelo. Assim, é possível concluir que a diagonal principal da matriz de confusão representa os acertos do modelo, ou verdadeiros positivos, e os demais elementos representam os erros do modelo, ou falsos positivos. A equação 4.1 representa a porcentagem de dados corretamente classificados em relação à todos os dados [49].

$$precisão = \frac{TP}{TP + FP} \quad (4.1)$$

### 4.2.2 Resultados e discussão

Através do experimento citado na seção anterior, foi possível identificar que a melhor opção para o agrupamento de questionários é o algoritmo *hierarchical clustering*. A Tabela 4 apresenta a precisão obtida por cada uma das abordagens. De acordo com a Tabela, o algoritmo *K-means* obteve uma precisão de 86.6% e o algoritmo *hierarchical clustering* obteve uma precisão de 88%.

Tabela 4 – Resultados dos modelos de agrupamento.

	<i>K-means</i>	<i>Hierarchical Clustering</i>
Iteração 1	98%	88%
Iteração 2	76%	—
Iteração 3	88%	—
Iteração 4	84%	—
Iteração 5	84%	—
Iteração 6	84%	—
Iteração 7	84%	—
Iteração 8	84%	—
Iteração 9	76%	—
Iteração 10	84%	—
Média	86.6%	88%

A Tabela 5 apresenta a matriz de confusão obtida através do algoritmo *hierarchical clustering*. A partir desta Tabela, é possível observar que 22 objetos do *dataset* foram agrupados obedecendo o critério da média das taxas de atendimento por nível de maturidade, estabelecidos na seção 4.2.1. Apenas três objetos foram agrupados de forma incorreta.

Por meio de uma análise dos objetos presentes no *cluster* 5, equivalente ao nível otimizado, a Tabela 6 apresenta as taxas de atendimento dos seus objetos em cada domínio de TI. Através da média das taxas de atendimento, observa-se que apenas o primeiro objeto atende ao critério do *cluster*. Os demais objetos atendem ao critério do *cluster* de nível gerenciado quantitativamente.

Tabela 5 – Matriz de confusão do *hierarchical clustering*.

	<i>Cluster</i> 1	<i>Cluster</i> 2	<i>Cluster</i> 3	<i>Cluster</i> 4	<i>Cluster</i> 5
<i>Cluster</i> 1	2	0	0	0	0
<i>Cluster</i> 2	0	7	0	0	0
<i>Cluster</i> 3	0	0	6	0	0
<i>Cluster</i> 4	0	0	1	6	0
<i>Cluster</i> 5	0	0	0	2	1

Isso acontece devido ao objeto 1 ser o questionário com nível mais elevado da base. Mesmo sendo o mais elevado, a média de suas taxas de atendimento é de 80.16%,

o que corresponde ao limite inferior da característica do grupo 5. Desta forma, os dados disponíveis, apesar de bem distribuídos, não descrevem corretamente o último grupo. Entretanto, a aquisição de mais questionários que descrevam as características do *cluster* 5 reformam este grupo, e o algoritmo *hierarchical clustering* é capaz de representá-lo corretamente.

Tabela 6 – Taxa de atendimento dos domínios de TI dos objetos do *cluster* 5.

	Objeto 1	Objeto 2	Objeto 3
Alinhamento estratégico	79.63%	73.76%	70.45%
Valor de TI	82.46%	70.43%	68.17%
Gerenciamento de riscos	80.97%	73.45%	69.91%
Recursos de TI	81.74%	67.63%	75.30%
Medidas de performance	71.42%	78.57%	75.23%
Responsabilidades	84.71%	72.42%	68.77%
Média dos domínios	80.16%	72.71%	71.31%

### 4.3 Geração de dados sintéticos

Devido à dificuldade em obter avaliações de cenários de governança de TI, conforme citado no capítulo 3, é necessário recorrer ao dados sintéticos. Basicamente, um gerador de dados sintéticos é responsável por estender a quantidade de dados disponíveis, de forma a reforçar os padrões presentes na base de dados. Os dados sintéticos devem possuir uma grande taxa de semelhança com os dados originais, diferentemente do processo de replicação de dados.

No geral, a geração de dados sintéticos é aplicada em problemas de balanceamento de classes para classificadores. Nesta abordagem, o modelo é aplicado para equilibrar a quantidade de dados disponíveis para as classes, garantindo que o classificador generalize com maior precisão o conhecimento.

Entretanto, o problema do presente trabalho está relacionado à uma base de dados com uma pequena quantidade de dados disponíveis. Neste caso, é necessário reforçar todo e qualquer tipo de padrão presente na base de dados. Esta seção descreve a metodologia utilizada para avaliar os algoritmos apresentados na seção 3.2.2. Além disso, os resultados obtidos também serão abordados a seguir.

#### 4.3.1 Metodologia

A validação dos algoritmos de geração de dados sintéticos propostos neste trabalho baseia-se na relação entre os dados gerados sinteticamente e os dados originais. Basicamente, um modelo é apropriado se o dados gerados a partir deste possuem alta

semelhança com os dados originais. Para ratificar a semelhança dos dados, é necessário analisar as características dos conjuntos de dados.

Desta forma, a análise inicia-se com a aplicação de um algoritmo de *clustering* nos dados originais. Em seguida, os dados sintéticos são gerados e uma comparação é realizada através de outro passo de *clustering*. Esta comparação visa verificar se o conjunto de dados sintéticos possui as mesmas características dos dados originais. Assim, o conjunto de dados sintéticos gerados a partir de dados presentes em um determinado *cluster* devem ser agrupados em conjunto.

Foram estabelecidos dois valores distintos para a taxa de mudança, sendo eles: 5% e 50%. O valor mais baixo induz a geração de dados com poucas alterações. Desta forma, o valor mais alto complementa a validação, levantando a capacidade dos modelos em situações distintas. Além disso, cada abordagem foi executada 10 vezes e, em cada iteração, foram gerados 100 dados sintéticos a partir dos dados originais.

O melhor algoritmo é eleito com base na precisão do modelo em gerar dados semelhantes, porém não idênticos, aos dados originais. Para isso, optou-se em utilizar a média de precisão alcançada através da matriz de confusão. Para cada iteração, um valor de precisão é calculado e, ao término das 10 iterações, a média entre os valores é computada, definindo a precisão geral do modelo.

### 4.3.2 Resultados e discussão

Por meio do experimento citado na seção 4.3.1, a Tabela 7 apresenta os resultados obtidos. Na primeira comparação, utilizando uma taxa de mudança de 5%, ambos os modelos obtiveram um bom percentual de acertos. Entretanto, o modelo determinístico obteve um melhor percentual.

Considerando a aleatoriedade da abordagem estocástica, foi realizada a segunda etapa do experimento. Esta etapa consiste em validar se o modelo possui de fato a capacidade de gerar dados semelhantes aos originais ou se a alta taxa de precisão está relacionada apenas à baixa taxa de mudança. Desta forma, a taxa de mudança foi alterada para 50%, o que permite uma maior alteração dos dados pelo modelo.

Tabela 7 – Precisão dos algoritmos de geração de dados sintéticos.

Taxa de mudança	Determinístico	Tabela
5%	99.8%	97%
50%	97%	58.7%

Por meio da segunda etapa, foi possível verificar que o modelo determinístico manteve o alto percentual de acerto, de 99.8% para 97%, perdendo apenas 2.8% de precisão.

No caso da abordagem estocástica, a perda de precisão foi mais drástica, de 97% para 58.7%, perdendo 38.3 pontos percentuais de precisão.

Desta forma, conclui-se que a alta precisão do modelo Tabela está relacionada à baixa taxa de mudança. Quando esta taxa é aumentada, a aleatoriedade do modelo torna os dados gerados mais dispersos, alterando de forma evidente as características dos questionários. Por outro lado, a abordagem determinística, mesmo perdendo precisão, é capaz de manter as características dos questionários em um nível aceitável.

## 4.4 Busca pelo ponto de melhoria

Encontrar o ponto de melhoria de um determinado cenário de uma organização é uma tarefa complexa, pois exige um alto nível de conhecimento dos envolvidos, demanda uma grande quantidade de tempo e envolve uma alta taxa de riscos. Devido a este fato, o principal objetivo deste trabalho é validar um modelo para encontrar o ponto de melhoria de forma automatizada e com base em situações semelhantes, já vivenciado por outras organizações.

Para alcançar este objetivo, o GA foi eleito para automatizar esta tarefa. Através de um processo de busca em questionários representando cenários semelhantes, o modelo é capaz de identificar qual a questão mais apropriada para aplicar uma medida de melhoria. A identificação desta questão ocorre com base nos padrões comportamentais e tendências contidos nos questionários.

### 4.4.1 Metodologia

Para ratificar o processo de busca pelo ponto de melhoria, foi necessário identificar a configuração do GA que melhor se aplica à resolução do problema. Através da Tabela 8 é possível observar as 12 propostas de configurações para a resolução do problema. A coluna "quantidade de dados" se refere às bases de dados utilizadas. Foram geradas duas bases, uma com 75 dados sintéticos e 25 dados originais e a outra com 975 dados sintéticos e 25 dados originais, com o objetivo de verificar a influência da quantidade de dados nos resultados do GA.

Além disso, foram investigadas algumas variações de operadores genéticos. Como estratégia de seleção, foram examinados os métodos de seleção por roleta e seleção por torneio. Para a estratégia de *crossover*, foram apurados os métodos de corte simples, corte duplo e uma nova abordagem, proposta neste trabalho, intitulada como corte por gene. Nesta nova abordagem, a troca de material genético ocorre apenas entre dois genes, mantendo grande parte das características dos indivíduos.

Para realizar os experimentos, foram selecionados aleatoriamente 10 indivíduos em cada base de dados, com a finalidade de otimizá-los. Para cada indivíduo, foram

Tabela 8 – Configurações de experimentos.

	Quantidade de dados	Estratégia de seleção	Estratégia de <i>crossover</i>
exp1	100	roleta	corte simples
exp2	100	roleta	corte duplo
exp3	100	torneio	corte simples
exp4	100	torneio	corte duplo
exp5	1000	roleta	corte simples
exp6	1000	roleta	corte duplo
exp7	1000	torneio	corte simples
exp8	1000	torneio	corte duplo

executados 10 vezes o GA para autenticar os pontos de melhoria identificados. Desta forma, cada experimento foi executado 100 vezes, totalizando 1200 execuções do GA.

Para identificar a configuração mais apropriada, foram analisadas duas perspectivas dos resultados: (i) a ocorrência do ponto de melhoria em mais de uma solução e (ii) a quantidade de gerações necessárias para encontrar os pontos de melhoria. A ocorrência do ponto de melhoria em mais de uma solução garante que o GA foi capaz de extrair o padrão presente nos dados utilizados. A quantidade de gerações necessárias para encontrar os pontos de melhoria não deve ser muito alta nem muito baixa, a ponto de ocorrer uma convergência prematura.

#### 4.4.2 Resultados e discussão

Ao analisar os resultados dos experimentos, foi possível identificar que o experimento 2 possui os melhores resultados. Por ter sido executado mais de uma vez para otimizar um mesmo questionário e pela sua capacidade de generalizar mais de uma solução, o GA encontrou diversos pontos de melhoria para um mesmo questionário.

Desta forma, é necessário realizar uma contagem de vezes que um ponto de melhoria foi recomendado e calcular a probabilidade deste ser a solução para o aprimoramento de um determinado cenário. Através deste processo, a segunda configuração encontrou, no máximo, dois pontos de melhoria como melhores opções. Isso é, sem dúvida, uma característica plausível, uma vez que isto elimina a incerteza na tomada de decisão do modelo.

Além disso, esta abordagem seleciona os pontos de melhoria de forma coerente, sem que haja conflitos com outros pontos. Ou seja, as questões levantadas para sofrer melhoria possuem baixa dependência com as demais questões. Isso garante que um ponto de melhoria que depende de outra questão ainda não otimizada não seja recomendado.

Por fim, a Figura 19 apresenta a evolução das soluções. Nesta figura, o eixo x representa cada geração de uma determinada execução, e o eixo y representa a média dos

*fitness* dos indivíduos da geração. Desta forma, a população da primeira geração, onde os indivíduos não estão otimizados, possui um *fitness* médio de 1.23. Na geração 88, o *fitness* médio aumentou para 1.32, e a execução foi encerrada, pois o *fitness* de um ou mais indivíduos ultrapassaram o valor de *fitness* do questionário em análise.

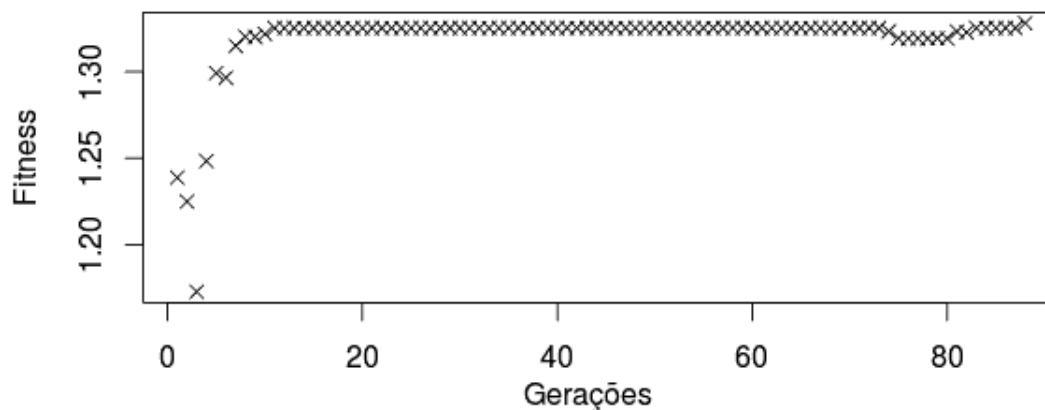


Figura 19 – Exemplo de evolução do *fitness* por geração.

Neste exemplo, os pontos de melhoria identificados foram as questões 34 e 36, que possuem, respectivamente, os coeficientes -1 e -3. A questão 31 avalia o incentivo do uso da TI como diferencial estratégico pela alta administração. Neste caso, o coeficiente -1 representa que todos os projetos de TI devem se comportar como todos os demais projetos. O objetivo desta questão é posicionar a TI de forma que a alta gerência proporciona capital de modo facilitado para todos (ou seguindo algum critério) os projetos de TIC e os projetos que não vão bem são barrados logo no início.

A questão 36 avalia o incentivo da capacitação e treinamento sobre o uso da TI. O coeficiente -3 representa a não existência de preocupação com treinamento e capacitação quanto ao uso da TIC. O objetivo desta questão é posicionar a TI de forma que existe uma política de treinamento de uso da TI estabelecida pela alta gerência. Dessa forma, em ambas as questões, o resultado do questionário apresenta uma certa deficiência diante à situação ideal de governança de TI. Assim, o algoritmo genético foi capaz de identificar corretamente os pontos de maior necessidades em um determinado cenário de TI.

Em relação aos outros experimentos com base de dados com 100 objetos, os resultados não foram satisfatórios. No geral, a convergência prematura e a grande quantidade de pontos de melhoria mais recomendados tornaram os resultados insatisfatório. Além disso, a falta de coerência nos pontos de melhoria e a alteração de grande parte da configuração do questionário, como no caso dos experimentos que utilizaram corte simples como estratégia de *crossover*, também foram fatores decisivos para a eleição do melhor modelo.

Para os experimentos com base de dados com 1000 objetos, a qualidade dos resultados foi baixa devido à quantidade de dados disponíveis para a busca. Como a base possui uma quantidade maior de dados, o GA é, conseqüentemente, inicializado com uma maior quantidade de indivíduos. Isso aumenta as chances e a quantidade de ocorrências de *crossover*. Devido à esta característica e ao critério de parada utilizado, o algoritmo pode ser forçado a atingir um máximo local, selecionando soluções não ideais. Por fim, por disponibilizar de uma grande quantidade de indivíduos, a execução o GA torna-se mais lenta.



## 5 CONCLUSÃO

O desenvolvimento de governança de TI é, de fato, de grande importância para o progresso de uma organização. Além disso, a dificuldade em compreender e tomar decisões acerca do aprimoramento deste contexto é eminente e explorada neste trabalho. Desta forma, a importância do desenvolvimento de uma modelo capaz de auxiliar esses processos é perceptível.

A análise do problema levantado no presente trabalho estabeleceu o modelo discutido no capítulo 3. Através deste paradigma, determinou-se que um modelo capaz de constituir o conhecimento de um contexto dinâmico e apoiar a tomada de decisão através de recomendação de pontos de melhoria deve possuir quatro elementos principais: (i) um repositório de conhecimento, (ii) aquisição do conhecimento, (iii) agrupamento dos dados e (iv) busca pelo ponto de melhoria.

O repositório de conhecimento deve ser capaz de mapear todos o conhecimento necessário para a compreensão e tomada de decisão no contexto. Desta forma, optou-se em basear o repositório em conceitos de *data warehouse*. Além de garantir a qualidade dos e a aquisição de dados de fontes remotas, estes conceitos estruturam a base de dados de forma a auxiliar a análise do conhecimento contido na mesma.

Por meio de um levantamento bibliográfico e consultas com especialistas, foi possível desenvolver um repositório de conhecimento em governança de TI capaz de mapear os principais pontos levantados sobre o contexto. Além disso, a estrutura definida neste trabalho é compatível com os descritores de governança de TI e com o mecanismo de análise proposto por Briganó [2].

Os elementos do *framework* de desenvolvimento de governança de TI, apresentado por Briganó [2] foi crucial para o desenvolvimento deste estudo. O mecanismo de análise é uma excelente ferramenta de aquisição, uma vez que ela é capaz de extrair o conhecimento empírico e posicioná-lo através de valores numéricos. Estes valores podem ser compreendidos tanto como a configuração do questionário através dos coeficientes das alternativas selecionadas, quanto a taxa de atendimento do cenário organizacional nos domínios de governança de TI.

Apesar da má formação do *cluster* relacionado ao nível de maturidade "otimizado", a fase de análise dos algoritmos de *clustering* elegeu o *hierarchical clustering* como a abordagem que melhor agrupou os questionários por nível de maturidade. A eficiência deste modelo se comprova através dos demais grupos, que realizam o agrupamento bem definido e com os dados característicos de cada grupo.

O problema relacionado ao último grupo de níveis de maturidade não é crítico para

o *hierarchical clustering*, uma vez que este se trata de uma abordagem de aprendizado de máquina não supervisionado. Por meio de um incremento da quantidade de dados relacionados ao último grupo, o algoritmo é capaz de generalizar corretamente o conhecimento, agrupar os dados precisamente e contornar a deficiência relacionada com a formação do *cluster* de nível otimizado.

Os resultados relacionados ao processo de busca pelo ponto de melhoria foram satisfatório para um estudo inicial neste campo. Por meio do GA, foi possível identificar o ponto de melhoria mais apropriado para um determinado cenário, com base no conhecimento obtido a partir de outras situações semelhantes à atual. A configuração que obteve o melhor resultado foi a seleção por roleta e *crossover* por corte duplo.

Através da execução consecutiva do GA para o aprimoramento de de um mesmo questionário e de suas características, foi possível verificar que esta abordagem pode identificar mais de um ponto de melhoria para um cenário exclusivo. Entretanto, através dos padrões contidos no conhecimento utilizado no processo de otimização, os pontos de melhoria mais recomendados aparecem com solução com maior frequência.

Outra característica vinculada ao problema abordado neste trabalho é a quantidade de dados disponíveis para a análise. A natureza da governança de TI disponibiliza as avaliações dos cenários organizacionais em um período de tempo relativamente extenso. Desta forma, verificou-se a necessidade de geração de dados sintéticos com base nos dados disponíveis. Desta forma, foram propostos dois modelos de geração de dados sintéticos, o determinístico e o Tabela. A abordagem determinística apresentou os melhores resultados. A avaliação de ambas as abordagens baseou-se na qualidade dos dados gerados em relação com a taxa de mudança atribuída para cada modelo.

A análise dos resultados obtidos a partir deste trabalho permite a predição de como uma ferramenta de recomendação de pontos de melhoria para o contexto de governança de TI deve funcionar. Logo, esta ferramenta deve utilizar as abordagens que obtiveram os melhores resultados. Uma observação final é que a fase de busca deve ser um processo iterativo, que apresenta diversos pontos de melhoria. Através de uma contagem dos pontos identificados, é possível recomendar todos os pontos de melhoria. Em conjunto com o ponto de melhoria, é apresentado o percentual de chance de determinado ponto ser o mais apropriado.

Por fim, o presente trabalho possibilita o desenvolvimento de algumas pesquisas futuras, como por exemplo:

1. A análise de aplicabilidade de outros modelos de inteligência computacional para o processo de recomendação de pontos de melhoria (ex.: mapas auto-organizáveis de Kohonen);

2. A análise de busca por pontos de melhoria através de permutação de questionários;
3. A aplicação de *data mining* no repositório para estudar e compreender o comportamento relativo à governança de TI.



## REFERÊNCIAS

- [1] ITGI. *Board briefing on IT governance*. [S.l.]: IT Governance Institute, 2003.
- [2] BRIGANÓ, G. U. *Um framework para desenvolvimento de governança de TIC*. 155 f. Dissertação (Mestrado) — Universidade Estadual de Londrina, 2012.
- [3] HAN, J.; KAMBER, M.; PEI, J. *Data mining: concepts and techniques*. 3. ed. Waltham: Morgan kaufmann, 2012. 703 p.
- [4] ENGELBRECHT, A. P. *Computational intelligence: an introduction*. 2. ed. [S.l.]: Wiley, 2007.
- [5] LINDER, R. *Algoritmos genéticos*. 3. ed. [S.l.]: Editora ciência moderna, 2012.
- [6] RADOVANOVIC, D. et al. Information technology governance x2014; cobit model. In: *MIPRO, 2011 Proceedings of the 34th International Convention*. [S.l.: s.n.], 2011. p. 1426–1429.
- [7] IBGC. <<http://www.ibgc.org.br/inter.php?id=18163>>. Acesso em: 12 mar, 2016.
- [8] RODRIGUES, L. C.; MACCARI, E. A.; A, S. A. S. O desenho da gestão da tecnologia das 100 maiores empresas na visão dos executivos de TI. *JISTEM - Journal of Information Systems and Technology Management*, scielo, v. 6, p. 483 – 506, 00 2009. ISSN 1807-1775. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1807-17752009000300006&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1807-17752009000300006&nrm=iso)>.
- [9] LUNARDI, G. L.; MAÇADA, A. C. G.; BECKER, J. L. It governance effectiveness and its antecedents: An empirical examination in brazilian firms. In: *2014 47th Hawaii International Conference on System Sciences*. [S.l.: s.n.], 2014. p. 4376–4385. ISSN 1530-1605.
- [10] LUNARDI, G. L.; BECKER, J. L.; MAÇADA, A. C. G. Governança de ti e suas implicações para a gestão da ti: um estudo acerca da percepção dos executivos. *XXXIV Encontro da ANPAD*. Rio de Janeiro, 2010.
- [11] TACONI, L. H. *GAIA catálogo de serviços de TI: um framework para construção de catálogos de serviços de tecnologia da informação*. 137 p. Dissertação (Mestrado) — Universidade Estadual de Londrina, 2014.
- [12] LUNARDI, G. L.; BECKER, J. L.; MACADA, A. C. G. The financial impact of it governance mechanisms' adoption: An empirical analysis with brazilian firms. In: *System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on*. [S.l.: s.n.], 2009. p. 1–10. ISSN 1530-1605.
- [13] KOZLOVA, E.; HASENKAMP, U.; KOPANAKIS, E. Use of it best practices for non-it services. In: *2012 Annual SRII Global Conference*. [S.l.: s.n.], 2012. p. 725–734. ISSN 2166-0778.
- [14] PMI. *Um guia do conhecimento em gerenciamento de projetos (PMBOK)*. 4. ed. [S.l.: s.n.], 2008.

- [15] MUELLER, T. et al. Because everybody is different: Towards understanding the acceptance of organizational it standards. In: *System Sciences (HICSS), 2015 48th Hawaii International Conference on*. [S.l.: s.n.], 2015. p. 4050–4058. ISSN 1530-1605.
- [16] SEAH, B. K.; SELAN, N. E. Design and implementation of data warehouse with data model using survey-based services data. In: *Innovative Computing Technology (INTECH), 2014 Fourth International Conference on*. [S.l.: s.n.], 2014. p. 58–64.
- [17] HUANG, T. et al. Digital urban planning oriented data warehouse constructing supported by gis: Taking greater beijing regional planning as a case. In: *2010 18th International Conference on Geoinformatics*. [S.l.: s.n.], 2010. p. 1–5. ISSN 2161-024X.
- [18] DEHNE, F. et al. A distributed tree data structure for real-time olap on cloud architectures. In: *Big Data, 2013 IEEE International Conference on*. [S.l.: s.n.], 2013. p. 499–505.
- [19] HAJMOOSAEI, A.; KASHFI, M.; KAILASAM, P. Comparison plan for data warehouse system architectures. In: *Data Mining and Intelligent Information Technology Applications (ICMiA), 2011 3rd International Conference on*. [S.l.: s.n.], 2011. p. 290–293.
- [20] KLEINBERG, J.; TARDOS, E. Approximation algorithms for classification problems with pairwise relationships: metric labeling and markov random fields. In: *Foundations of Computer Science, 1999. 40th Annual Symposium on*. [S.l.: s.n.], 1999. p. 14–23. ISSN 0272-5428.
- [21] ZHANG, Z. et al. Clustering aggregation based on genetic algorithm for documents clustering. In: *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*. [S.l.: s.n.], 2008. p. 3156–3161. ISSN 1089-778X.
- [22] PAREDES, G. E.; VARGAS, L. S. Circle-clustering: A new heuristic partitioning method for the clustering problem. In: *The 2012 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2012. p. 1–8. ISSN 2161-4393.
- [23] ZHANG, C.; XIA, Q.; YANG, G. Reconsideration about clustering analysis. In: *Industrial Electronics and Applications (ICIEA), 2015 IEEE 10th Conference on*. [S.l.: s.n.], 2015. p. 1517–1524.
- [24] ADAMS, R. P. *Hierarchical agglomerative clustering*. <[https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwiwwOLZoKzMAhWIEpAKHe--CVgQFgghMAA&url=https%3A%2F%2Fcanvas.harvard.edu%2Fcourses%2F1862%2Ffiles%2F260018%2Fdownload%3Fverifier%3D65N5RCbwWZppOxVmOoGLYJfYlgFTUGrb2gfainaH%26wrap%3D1&usg=AFQjCNGixfODaS5yqBy3OB8ogtI03fhY1Q&sig2=BmDzNH\\_\\_oqsLLtByxywv-Q&cad=rja](https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwiwwOLZoKzMAhWIEpAKHe--CVgQFgghMAA&url=https%3A%2F%2Fcanvas.harvard.edu%2Fcourses%2F1862%2Ffiles%2F260018%2Fdownload%3Fverifier%3D65N5RCbwWZppOxVmOoGLYJfYlgFTUGrb2gfainaH%26wrap%3D1&usg=AFQjCNGixfODaS5yqBy3OB8ogtI03fhY1Q&sig2=BmDzNH__oqsLLtByxywv-Q&cad=rja)>. Acesso em: 25 mar, 2016.
- [25] CHAWLA, N. V. et al. Smote: Synthetic minority over-sampling technique. *J. Artif. Int. Res.*, AI Access Foundation, USA, v. 16, n. 1, p. 321–357, jun. 2002. ISSN 1076-9757. Disponível em: <<http://dl.acm.org/citation.cfm?id=1622407.1622416>>.

- [26] ANDERSON, J. W. et al. Synthetic data generation for the internet of things. In: *Big Data (Big Data), 2014 IEEE International Conference on*. [S.l.: s.n.], 2014. p. 171–176.
- [27] CANO, I.; TORRA, V. Generation of synthetic data by means of fuzzy c-regression. In: *Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference on*. [S.l.: s.n.], 2009. p. 1145–1150. ISSN 1098-7584.
- [28] HU, S. et al. Msmote: Improving classification performance when training data is imbalanced. In: *Computer Science and Engineering, 2009. WCSE '09. Second International Workshop on*. [S.l.: s.n.], 2009. v. 2, p. 13–17.
- [29] LIU, C. Y.; ZOU, C. M.; WU, P. A task scheduling algorithm based on genetic algorithm and ant colony optimization in cloud computing. In: *Distributed Computing and Applications to Business, Engineering and Science (DCABES), 2014 13th International Symposium on*. [S.l.: s.n.], 2014. p. 68–72.
- [30] GUO, P.; WANG, X.; HAN, Y. The enhanced genetic algorithms for the optimization design. In: *2010 3rd International Conference on Biomedical Engineering and Informatics*. [S.l.: s.n.], 2010. v. 7, p. 2990–2994. ISSN 1948-2914.
- [31] ZHANG, L.; CHANG, H.; XU, R. Equal-width partitioning roulette wheel selection in genetic algorithm. In: *2012 Conference on Technologies and Applications of Artificial Intelligence*. [S.l.: s.n.], 2012. p. 62–67. ISSN 2376-6816.
- [32] HONG, Y. et al. A comprehensive comparison between real population based tournament selection and virtual population based tournament selection. In: *2007 IEEE Congress on Evolutionary Computation*. [S.l.: s.n.], 2007. p. 445–452. ISSN 1089-778X.
- [33] KREY, M. Information technology governance, risk and compliance in health care - a management approach. In: *Developments in E-systems Engineering (DESE), 2010*. [S.l.: s.n.], 2010. p. 7–11.
- [34] MENDES, F. F. A method for diagnosing information technology governance processes. In: *System Sciences (HICSS), 2013 46th Hawaii International Conference on*. [S.l.: s.n.], 2013. p. 4395–4404. ISSN 1530-1605.
- [35] DAN, Y. The utility evaluation analysis of information technology based on it's usage frequency. In: *Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on*. [S.l.: s.n.], 2011. p. 278–282.
- [36] ALVES, V.; RIBEIRO, J.; CASTRO, P. Information technology governance x2014; a case study of the applicability of itil and cobit in a portuguese private school. In: *Information Systems and Technologies (CISTI), 2012 7th Iberian Conference on*. [S.l.: s.n.], 2012. p. 1–6. ISSN 2166-0727.
- [37] RIBBERS, P. M. A.; PETERSON, R. R.; PARKER, M. M. Designing information technology governance processes: diagnosing contemporary practices and competing theories. In: *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on*. [S.l.: s.n.], 2002. p. 3143–3154.

- [38] AJAYI, B. A.; HUSSIN, H. Exploring information technology governance in a malaysian public university: Providers' perspectives. In: *Information and Communication Technology for The Muslim World (ICT4M), 2014 The 5th International Conference on*. [S.l.: s.n.], 2014. p. 1–6.
- [39] KILIC, N.; METIN, B. Importance of education in information technology governance. In: *2012 4th IEEE International Symposium on Logistics and Industrial Informatics*. [S.l.: s.n.], 2012. p. 65–68. ISSN 2156-8790.
- [40] KAUR, J.; MOHAMED, N.; AHLAN, A. R. Modeling the impact of information technology governance effectiveness using partial least square. In: *Statistics in Science, Business, and Engineering (ICSSBE), 2012 International Conference on*. [S.l.: s.n.], 2012. p. 1–5.
- [41] PEREIRA, G. V. et al. Information technology governance practices adoption through an institutional perspective: The perception of brazilian and american cios. In: *System Sciences (HICSS), 2013 46th Hawaii International Conference on*. [S.l.: s.n.], 2013. p. 4446–4455. ISSN 1530-1605.
- [42] CHENG, X.; GONG, Y. A study of the application of information technology in corporate governance of li ning co., ltd. In: *Computational and Information Sciences (ICCIS), 2012 Fourth International Conference on*. [S.l.: s.n.], 2012. p. 750–753.
- [43] FAN, L.; WU, Y. Information technology, decision-making mechanism and selection of decision-making mode. In: *Computing, Control and Industrial Engineering (CCIE), 2010 International Conference on*. [S.l.: s.n.], 2010. v. 2, p. 81–86.
- [44] KREY, M. et al. Approach to the evaluation of a method for the adoption of information technology governance, risk management and compliance in the swiss hospital environment. In: *System Science (HICSS), 2012 45th Hawaii International Conference on*. [S.l.: s.n.], 2012. p. 2810–2819. ISSN 1530-1605.
- [45] SU, H. The information technology spillover effect and substitution effect analyses on economics in china. In: *Management and Service Science (MASS), 2010 International Conference on*. [S.l.: s.n.], 2010. p. 1–4.
- [46] NUGROHO, B.; SURENDRO, K. Using organizational culture approach and cobit framework in designing of information technology governance on non ministerial government institute (lpnk), case study: Center for scientific documentation and information - indonesian institute of sciences. In: *Electrical Engineering and Informatics (ICEEI), 2011 International Conference on*. [S.l.: s.n.], 2011. p. 1–5. ISSN 2155-6822.
- [47] NISHA; KAUR, P. J. Cluster quality based performance evaluation of hierarchical clustering method. In: *Next Generation Computing Technologies (NGCT), 2015 1st International Conference on*. [S.l.: s.n.], 2015. p. 649–653.
- [48] ORACLE. *Data Mining Concepts*. <[https://docs.oracle.com/cd/E11882\\_01/datamine.112/e16808/classify.htm](https://docs.oracle.com/cd/E11882_01/datamine.112/e16808/classify.htm)>. Acesso em: 9 abr, 2016.
- [49] ADANIYA, M. H. A. C. *Detecção de anomalias utilizando firefly harmonic clustering algorithm e assinatura digital de segmento de rede*. 81 f. Dissertação (Mestrado) — Universidade Estadual de Londrina, 2011.

## Apêndices



## APÊNDICE A – IMPLEMENTAÇÃO DA GERAÇÃO DE DADOS SINTÉTICOS DETERMINÍSTICO EM R

```
deterministic_synthetic_data_generation <- function(data = NULL, quantity = 1,
                                                    change_rate = 0.05, alphabet = NULL,
                                                    mask = NULL){
  if(is.null(mask)){
    n = floor(dim(data)[2] * (10/100));
    p = floor(dim(data)[2] * (5/100));
    mask = rep(0, dim(data)[2]);

    for(i in 1:sample(1:n, 1)){
      r = sample(1:dim(data)[2], 1);
      while(mask[r] == 1 || mask[r] == -1){
        r = sample(1:dim(data)[2], 1);
      }
      mask[r] <- -1;
    }
    for(i in 1:sample(1:p, 1)){
      r = sample(1:dim(data)[2], 1);
      while(mask[r] == 1 || mask[r] == -1){
        r = sample(1:dim(data)[2], 1);
      }
      mask[r] <- 1;
    }
  }

  synthetic_data = NULL;
  parent = NULL;

  for(i in 1:quantity){
    index = sample(1:dim(data)[1], 1);
    parent = c(parent, index);
    synthetic_data = rbind(synthetic_data, data[index, ]);
  }

  for(i in 1:dim(synthetic_data)[1]){
    for(j in 1:dim(synthetic_data)[2]){
      if(runif(1, 0, 1) <= change_rate){
        if(mask[j] == 1){
          if(!is.na(alphabet[j, which(alphabet[j, ] == synthetic_data[i, j]) + 1]) &&
              alphabet[j, which(alphabet[j, ] == synthetic_data[i, j])] < dim(alphabet)[2]){
            synthetic_data[i, j] = alphabet[j, which(alphabet[j, ] == synthetic_data[i, j])+1];
          }
        }
        if(mask[j] == -1){
          if(which(alphabet[j, ] == synthetic_data[i, j]) > 1){
            synthetic_data[i, j] = alphabet[j, which(alphabet[j, ] == synthetic_data[i, j])-1];
          }
        }
      }
    }
  }
  return(list(data = synthetic_data, index = parent));
}
```



## APÊNDICE B – IMPLEMENTAÇÃO DA GERAÇÃO DE DADOS SINTÉTICOS TABELA EM R

```

stochastic_synthetic_data_generation <- function(data = NULL, quantity = 1,
                                                change_rate = 0.05, alphabet = NULL){
  synthetic_data = NULL;
  parent = NULL;

  for(i in 1:quantity){
    index = sample(1:dim(data)[1], 1);
    parent = c(parent, index);
    synthetic_data = rbind(synthetic_data, data[index, ]);
  }

  for(i in 1:dim(synthetic_data)[1]){
    for(j in 1:dim(synthetic_data)[2]){
      if(runif(1, 0, 1) <= change_rate){
        if(sample(0:1, 1) == 0){
          if(which(alphabet[j, ] == synthetic_data[i,j]) > 1){
            synthetic_data[i,j] = alphabet[j, which(alphabet[j, ] == synthetic_data[i,j])-1];
          }else{
            synthetic_data[i,j] = alphabet[j, which(alphabet[j, ] == synthetic_data[i,j])+1];
          }
        }else{
          if(!is.null(alphabet[j, which(alphabet[j, ] == synthetic_data[i,j]) + 1])
            && !is.na(alphabet[j, which(alphabet[j, ] == synthetic_data[i,j]) + 1])){
            synthetic_data[i,j] = alphabet[j, which(alphabet[j, ] == synthetic_data[i,j])+1];
          }else{
            synthetic_data[i,j] = alphabet[j, which(alphabet[j, ] == synthetic_data[i,j])-1];
          }
        }
      }
    }
  }
  return(list(data = synthetic_data, index = parent));
}

```



## APÊNDICE C – QUESTIONÁRIO DE VALIDAÇÃO DOS DESCRITORES DE GOVERNANÇA DE TI

Qual é o nível de influência que o "alinhamento estratégico" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?	
A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que o "aprimoramento da produtividade" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?	
A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que o "aprimoramento de serviços" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?	
A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que os "benefícios da ação/solução" possuem/sofrem quando aplicada uma ação/solução na TIC de uma organização?	
A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que o "conhecimento" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que o "custo" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que a "demanda de ações/soluções" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que a "economia" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que a "estrutura de TIC" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que o "alinhamento estratégico" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que os "investimentos" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que a "inovação" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que o "nicho de mercado" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que a "performance" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que as "prioridades de TIC" possuem/sofrem quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que os "recursos de TIC" possuem/sofrem quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que os "requisitos da ação/solução" possuem/sofrem quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que as "restrições" possuem/sofrem quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que os "riscos" possuem/sofrem quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que a "grau de satisfação da ação/solução" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que a "segurança da informação" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que o "tamanho da organização" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que o "tempo" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que o "tempo da organização" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que o "tempo da organização na presente posição" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que a "usabilidade da ação/solução" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que as "vantagens da ação/solução" possuem/sofrem quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo

Qual é o nível de influência que o "valor de TIC" possui/sofre quando aplicada uma ação/solução na TIC de uma organização?

A	Muito alto
B	Alto
C	Médio
D	Baixo
E	Muito baixo



## TRABALHOS PUBLICADOS PELO AUTOR

1. SAHD FILHO, E. A.; SENEFONTE, H. C. M. ; BARROS, R. M. . Proposta de Sistema Baseado em Regras Aplicado ao Desenvolvimento de TIC. In: 8ª Conferência Ibérica de Sistemas e Tecnologias de Informação (CISTI'2013), 2013, Lisboa. 8ª Conferência Ibérica de Sistemas e Tecnologias de Informação (CISTI'2013), 2013. v. II. p. 224-227.