

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE INFORMÁTICA
ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS E SUAS APLICAÇÕES**

RAFAELA SOMAVILA LIMA

**CRIAÇÃO DE PROJETO DE CIÊNCIA DE DADOS UTILIZANDO A
METODOLOGIA CRISP-DM EM CONFORMIDADE COM A LGPD**

**CURITIBA
2021**

RAFAELA SOMAVILA LIMA

**CRIAÇÃO DE PROJETO DE CIÊNCIA DE DADOS UTILIZANDO A
METODOLOGIA CRISP-DM EM CONFORMIDADE COM A LGPD**

Trabalho de Conclusão de Curso de Especialização apresentado como requisito parcial à obtenção do título de Especialista em Ciência de Dados e Suas Aplicações, do Departamento Acadêmico de Informática, da Universidade Tecnológica Federal do Paraná.

Orientadora: Profa. Dra. Rita Cristina Galarraga Berardi

**CURITIBA
2021**



Ministério da Educação
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
UTFPR - CAMPUS CURITIBA
DIRETORIA-GERAL - CAMPUS CURITIBA
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO - CAMPUS CURITIBA
DEPARTAMENTO DE APOIO DAS ESPECIALIZAÇÕES LATO-SENSU DOS
CURSOS DE INFORMÁTICA - CAMPUS CURITIBA
CURSO DE ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS E SUAS APLICAÇÕES



TERMO DE APROVAÇÃO

CRIAÇÃO DE PROJETO DE CIÊNCIA DE DADOS UTILIZANDO A METODOLOGIA CRISP-DM EM CONFORMIDADE COM A LGPD.

por

Rafaela Somavila Lima

Este Trabalho de Conclusão de Curso foi apresentado às 20h30min do dia 28 de julho de 2021 por videoconferência como requisito parcial à obtenção do grau de Especialista em Ciência de Dados e suas Aplicações na Universidade Tecnológica Federal do Paraná - UTFPR - Campus Curitiba. A aluna foi arguida pela Banca de Avaliação abaixo assinados. Após deliberação, a Banca de Avaliação considerou o trabalho aprovado.

Profa. Dra. Rita Cristina Galarraga Berardi (Presidente/Orientadora – DAINF-CT/ UTFPR-CT)

Prof. Dr. Alexandre Reis Graeml (Avaliador 1– DAINF-CT/ UTFPR-CT)

Prof. Dr. Prof. Dr. Luiz Celso Gomes Junior (Avaliador 2 – DAINF-CT/ UTFPR-CT)

O Termo de Aprovação assinado encontra-se no sistema SEI- nº Processo nº 23064.029980/2021-16

AGRADECIMENTOS

Agradeço especialmente à Prof.^a Dr.^a Rita Berardi por todo apoio e paciência durante a orientação deste trabalho e também durante toda a duração do curso, por incentivar a não desistir da Especialização nem do trabalho de conclusão. Obrigada por ouvir minha vaga ideia de tema e ajudar a transformar em algo interessante de se escrito. Sem o seu incentivo, esse agradecimento (e trabalho) nunca seria escrito.

Agradeço à minha amiga Andressa Francieli Prawucki de Paiva que me ajudou e incentivou durante todo o período da Especialização e que também não me deixou desistir, sem as piadas e memes este tema talvez nem seria escolhido.

Gostaria de agradecer à oito pessoas que fizeram e fazem parte do processo de escrita deste trabalho e em todos os outros dias: Park Seonghwa, Kim Hongjoong, Jeong Yunho, Kang Yeosang, Choi San, Song Mingi, Jung Wooyoung e Choi Jongo.

Agradeço aos meus pais que me incentivaram a continuar até o fim.

Just keep it up
모두가 볼 수 있도록
더 just keep it up

Apenas continue
Para que todos possam ver
Apenas continue

(ATEEZ – Sunrise)

RESUMO

LIMA, Somavila Rafaela. Criação de Projeto de Ciência de Dados utilizando a metodologia CRISP-DM em conformidade com a LGPD. 2021. 34 f. Trabalho de Conclusão de Curso de Especialização, Departamento Acadêmico de Informática, Universidade Tecnológica Federal do Paraná. Curitiba, 2021.

A Ética é uma parte da Filosofia que discute o que pode ser considerado bom ou mau, tratando-se de comportamentos dos indivíduos dentro de uma sociedade. A rapidez com que a computação foi evoluindo, o grande volume de dados gerados diariamente, a importância dos mesmos para gerar informações valiosas e sua aplicabilidade no contexto atual de Big Data e popularização do uso de inteligência artificial e automação, impulsionou a discussão sobre a ética na manipulação de dados e também as implicações legais para uso de dados de terceiros. Desde 2020 está em vigor a Lei Geral de Proteção de Dados Pessoais (LGPD), considerando-se diversos critérios pensados para proteger os dados da população brasileira. Este trabalho apresenta um mapeamento de como um projeto de Ciência de Dados pode estar em conformidade com a LGPD desde o momento da sua concepção. Propõe-se, assim, analisar a metodologia CRISP-DM, que é uma metodologia para desenvolver projetos de Ciência de Dados e relacionar as etapas aos artigos aplicáveis da Lei Geral de Proteção de Dados Pessoais.

Palavras-chave: Ética em computação. LGPD. CRISP-DM. Proteção de dados pessoais.

ABSTRACT

LIMA, Somavila Rafaela. Creation of a Data Science Project using the CRISP-DM methodology in compliance with LGPD. 2021. 34 f. Trabalho de Conclusão de Curso de Especialização, Departamento Acadêmico de Informática, Universidade Tecnológica Federal do Paraná. Curitiba, 2021.

Ethics is a field of Philosophy that discusses what can be considered "good" or "bad", in regards to the behavior of individuals within a society. The quickness with which computing has been evolving, the large amount of daily generated data, its importance to generate valuable information and its applicability in the current Big Data context and the popularization of artificial intelligence and automation use, boosted the discussion on ethics in data manipulation and also the legal implications for using personal third-party data. Since 2020, the Brazilian General Data Protection Law (LGPD) has been in force, covering several criteria designed to protect the data of Brazilian population. This work presents a mapping of how a Data Science project can comply with the LGPD from the moment of its conception. It is proposed to analyze the CRISP-DM methodology, a methodology to develop Data Science projects and relate the steps to the applicable articles of the Brazilian General Data Protection Law.

Keywords: Computer ethics. LGPD. CRISP-DM. Personal data protection.

LISTA DE ILUSTRAÇÕES

Figura 1 - Etapas do modelo CRISP-DM	17
--	----

LISTA DE SIGLAS

IA	Inteligência Artificial
KDD	Knowledge Discovery Database
MIT	Massachusetts Institute of Technology
LGPD	Lei Geral de Proteção de Dados Pessoais

LISTA DE ACRÔNIMOS

CRISP-DM	Cross Industry Standard Process for Data Mining
SEMMA	Sample, Explore, Modify, Model e Assess

SUMÁRIO

1 INTRODUÇÃO	10
2 REFERENCIAL TEÓRICO	11
2.1 ÉTICA.....	11
2.2 ÉTICA E DOMÍNIOS DE APLICAÇÃO	13
2.2.1 Computação	13
2.2.2 Medicina	13
2.2.3 Jogos Eletrônicos	14
2.2.4 Inteligência Artificial.....	14
2.2.5 Discussões e Preocupações	14
2.3 REGULAMENTAÇÃO DA ÉTICA NA COMPUTAÇÃO	15
2.4 MODELOS DE PROJETO DE CIÊNCIA DE DADOS	17
3 PROPOSTA DE MAPEAMENTO CRISP-DM E LGPD	19
3.1 MAPEAMENTO	19
3.1.1 Etapa 1 – Entendimento do Negócio	19
3.1.2 Etapa 2 – Entendimento dos Dados	21
3.1.3 Etapa 3 – Preparação dos Dados	22
3.1.4 Etapa 4 – Modelagem	23
3.1.5 Etapa 5 – Avaliação	23
3.1.6 Etapa 6 – Implantação.....	24
3.2 EXEMPLO DE USO	24
3.2.1 Etapa 1 – Entendimento do negócio	25
3.2.2 Etapa 2 – Entendimento dos dados	26
3.2.3 Etapa 3 – Preparação dos dados	27
3.2.4 Etapa 4 – Modelagem	28
3.2.5 Etapa 5 – Avaliação	28
3.2.6 Etapa 6 – Implantação.....	29
4 CONCLUSÃO.....	30
REFERÊNCIAS.....	31

1 INTRODUÇÃO

A filosofia existe há mais de vinte séculos (CHAUÍ, 2000), e nesse extenso período, permanece sendo um dos campos mais presentes. A ética, juntamente com a moral, é entendida como um conjunto de práticas e deveres dos membros de uma sociedade.

Por se referir ao estudo de valores morais de ações e sociedade, a ética é um assunto presente em vários âmbitos e também um tema atual, especialmente na computação.

Com a evolução da computação e o acesso a grandes volumes de dados disponíveis para utilização, a discussão sobre questões éticas na utilização dessas informações tornou-se um tema central.

O impacto do mau uso de informações pessoais de indivíduos pode ser grande e em diversos âmbitos, tais quais privacidade e comportamento (MÜLLER, 2021), discriminação e também na política, como foi o caso da descoberta em 2018 do uso de dados pela empresa *Cambridge Analytica* através do *Facebook* (MA e GILBERT, 2019).

As legislações não conseguem acompanhar a velocidade das atualizações tecnológicas. Porém, é possível tentar prever situações futuras e pensar em eventuais soluções que sejam necessárias. Em 2020 entrou em vigor a Lei Geral de Proteção de Dados Pessoais (BRASIL, 2018), seguindo o movimento de outros países ao pensar no tratamento e proteção de dados, tornando o tema ética ainda mais importante.

O objetivo deste trabalho é realizar um mapeamento entre os artigos da LGPD com as etapas de desenvolvimento de um projeto de Ciência de Dados seguindo a metodologia CRISP-DM para que sirva de apoio na construção de projetos de Ciência de Dados éticos desde o seu planejamento até a entrega. Para isso foi montado um exemplo com situações de aplicação dos artigos da LGPD dentro de cada etapa da metodologia CRISP-DM, do início à implantação de um projeto, sugerindo questionamentos a serem considerados no desenvolvimento do projeto.

2 REFERENCIAL TEÓRICO

2.1 ÉTICA

A ética é uma parte da filosofia que diz respeito ao que pode ser considerado moralmente bom ou mau. A palavra “ética” deriva de duas outras palavras, *ethos* (hábito, costume, caráter), do grego e *mores* (costumes), do latim. Moral e ética são conceitos que andam juntos. Moral é um conjunto de regras, hábitos e costumes de uma sociedade e que são aceitos pelos integrantes dessa sociedade, (GERT e GERT, 2020). Já a ética reflete o comportamento de um indivíduo e como esse indivíduo escolhe se relacionar com o próximo (WEX, 2021). O campo da ética também é chamado de filosofia da moral, pois envolve ações de recomendações e sistematização de conceitos do que é um comportamento certo ou errado (FIESER, 2021).

Para Orts e Navarro (2001), a ética não pode nem deve ser neutra, pois parte-se do princípio de que a ética é crítica dos costumes morais, portanto não possui neutralidade.

Ética é subdividida em outras três grandes áreas: metaética, ética normativa e ética aplicada. A metaética estuda e investiga de onde surgiram os princípios éticos e o que eles significam e se possuem de fato algum significado. Ela está ligada à metafísica e à psicologia. A ética normativa possui objetivo mais prático: entender os padrões morais que definem e regulam o que torna uma ação boa ou má. Ela também envolve hábitos adquiridos e consequência do comportamento para com terceiros. Por fim, a ética aplicada está relacionada com o estudo de assuntos morais considerados controversos, tais quais direitos dos animais, guerras, aborto, questões ambientais, entre outros temas (FIESER, 2021). A ética aplicada combina os conceitos da metaética e as ferramentas da ética normativa e tenta aplicar a teoria em situações reais e cotidianas.

Por ser uma dimensão da ética que é empregada em situações reais, a ética aplicada pode ser abordada em diversos campos, como biologia, ambiente corporativo, setores públicos, política, computação e ciência de dados, entre outras variadas áreas. Para James H. Moor (1985), a ética em computação é analisar a natureza e impacto social da tecnologia computacional. A tecnologia computacional

abrange não apenas computadores, mas também *software*, redes e uma tecnologia mais atual, a nuvem. Não existe um roteiro determinado para que uma ação seja considerada ética ou não na computação. Não existem políticas que abranjam todos os cenários possíveis de problemas. Para cada caso, é necessário avaliar uma política que guie na resolução da questão. Esse é um campo dinâmico e complexo, pois além de considerar valores que mudam para cada indivíduo, é preciso considerar também o constante avanço na tecnologia.

Moor (1985) comenta sobre o que chama de “fator da invisibilidade” no uso da tecnologia computacional, e o divide em três tipos: abuso invisível, usar intencionalmente operações de um computador para agir com conduta antiética. São exemplos disso, a possibilidade de espionar outros indivíduos acessando seus dados privados sem deixar registros e alterar dados de terceiros. Valores invisíveis de programas, desenvolver um software incluindo, mesmo que não intencionalmente valores pessoais da pessoa programadora e vieses. O terceiro tipo é chamado de cálculos complexos invisíveis, por serem capazes de realizar cálculos muito complexos. Como confiar que esses cálculos estão corretos realmente se um ser humano não consegue replicar e revisar? Um exemplo desse tipo de invisibilidade é o uso dos computadores em ambiente militar, onde as consequências de uma eventual falha possuem grande peso ético.

O acesso à internet, o aumento da capacidade de armazenamento de dados através de tecnologias como a nuvem e a facilidade com que é possível criar uma conta em redes sociais, fez com que o volume de informações pessoais inseridos na web por indivíduos aumentasse muito. Entre esses dados inseridos em plataformas, constam informações pessoais, incluindo fotos, nomes e localização. Em redes sociais, apesar de ser possível alterar configurações de privacidade, muitos dos dispositivos utilizados para fazer upload têm a capacidade de incluir metadados nessas postagens (SMITH et al., 2012).

Dados pessoais possuem grande valor. Se usados de forma antiética podem causar danos e grandes impactos, como foi o caso do escândalo envolvendo o *Facebook* e a empresa *Cambridge Analytica*. Os dados coletados a partir de uma pesquisa via *Facebook* - na qual os usuários não consentiram que suas informações fossem repassadas à *Cambridge Analytica* (CHEN, 2018), foram utilizados pela empresa para gerar perfis comportamentais de eleitores e através disso, enviar anúncios direcionados de campanha, de acordo com o alinhamento político de cada

perfil. A intenção era manipular eleitores durante a votação presidencial dos Estados Unidos da América em 2018 para favorecer o candidato do Partido Republicano.

2.2 ÉTICA E DOMÍNIOS DE APLICAÇÃO

Por ser um campo atemporal e aplicável em diversas disciplinas, a ética é aplicada também à computação e a outros campos relacionados à computação, citados a seguir.

2.2.1 Computação

A ética em computação é um braço da ética aplicada. O conceito foi primeiramente usado por um professor do MIT, Norbert Wiener durante os anos 1940 enquanto participava de um projeto durante a Segunda Guerra Mundial (BYNUM, 2020). Em 1954 foi publicado, também por Wiener, o livro “The Human Use of Human Beings: cybernetics and society” que dá base para a análise e pesquisa da ética em computação.

As áreas de pesquisa da ética computacional são vastas, incluem crimes cibernéticos, privacidade, uso dos dados, vigilância e monitoramento em cidades inteligentes. Atualmente com a popularização da automatização de ambientes e objetos, aumenta a discussão sobre a Internet das Coisas e veículos inteligentes.

2.2.2 Medicina

A computação tem sido aliada da prática da medicina e da assistência médica, seja com modelos para estudos, partes e modelos robóticos para prática de cirurgias, o auxílio em diagnósticos de doenças, e também para tratamento de pacientes. A informática também auxilia no armazenamento, transmissão dos dados dos pacientes para que seja possível ter sistemas integrados e facilitar o tratamento e controle do histórico do paciente. Esse armazenamento de informações levanta questões éticas (MARCKMANN e GOODMAN, 2006). Algumas delas: de quem é o dono dessas informações sobre a saúde do paciente? Como essas informações devem ser utilizadas? Quais benefícios e riscos para os pacientes?

2.2.3 Jogos Eletrônicos

Jogos eletrônicos, sejam eles *online*, *offline*, jogados individualmente ou em comunidade, têm se tornado alvo de debates sobre sua influência, considerando o impacto econômico e social dos novos jogos (BUCHANAN e ESS, 2005). Os jogos eletrônicos são parte relevante da computação e atraem interesse para a área da ética computacional, por terem interação com outras esferas como política e sociedade, segundo Buchanan e Ess (2005). Para Consalvo (2005), a ética em jogos eletrônicos existe em várias camadas. Não é algo que pode ser pensado apenas de modo binário, considerando um jogo “bom” ou “mau”. É preciso considerar que além da forma como foi desenvolvido o jogo, também há o jogador, que é um agente ativo que faz suas próprias escolhas quando está jogando.

2.2.4 Inteligência Artificial

A ética aplicada à inteligência artificial é uma vertente da ética computacional que foca em sistemas inteligentes (MÜLLER, 2021), considerando que esses sistemas são desenvolvidos e utilizados por humanos que possuem visões próprias do que consideram certo ou errado. Outra discussão ética sobre sistemas inteligentes envolve o comportamento desses sistemas.

Existem diversos campos de estudo e também desafios da ética dentro da inteligência artificial, sendo alguns deles a robótica, onde tem-se a situação hipotética de os sistemas inteligentes se tornarem inteligentes demais, a chamada “singularidade”, na qual a IA evolui tanto que foge ao controle da humanidade (KURZWEIL, 1999).

2.2.5 Discussões e Preocupações

A tecnologia não é imparcial, pois apenas por sua existência já implica em efeitos no mundo. Ela é desenvolvida por pessoas, que possuem suas próprias visões e que podem acabar influenciando na visão dos usuários, causando vieses de

interpretação, pois pode deixar dados importantes de fora (WHELCHER, 1986). Por não ser livre de valores, algumas questões e preocupações são pensadas na aplicação de tecnologias inteligentes.

Uma das questões é a utilização de tecnologias de reconhecimento facial, que podem ser usadas para marginalizar, discriminar, perseguir e vigiar ainda mais alguns grupos, como nos casos dos protestos no conflito entre China e Hong Kong que os líderes dos protestos são identificados e punidos (MOZUR, 2019). No Brasil em 2019, segundo estudo da Rede de Observatórios de Segurança (NUNES, 2019), 90% dos presos através de sistema de reconhecimento facial, no período de março a outubro eram negros. Nos Estados Unidos o sistema de reconhecimento também é utilizado sob um viés racial (GARVIE e FRANKLE, 2016)

O conceito da inteligência artificial é pensado e desenhado sob uma visão de pessoas brancas (CAVE e DIHAL, 2020). Sistemas desenvolvidos que necessitam de algum tipo de reconhecimento funcionam com poucos erros em peles claras, mas não em peles negras (LOHR, 2018).

2.3 REGULAMENTAÇÃO DA ÉTICA NA COMPUTAÇÃO

A maioria das aplicações citadas possui algo em comum: o uso ou armazenamento de dados de terceiros. Pensando na questão de que dados são informações sensíveis e pessoais, existem leis já escritas ou em processo de escrita sobre proteção de dados em diversos países.

Um exemplo de regulamentação para o processamento de dados pessoais é a GDPR (*General Data Protection Regulation*), formulada em 2016 e em vigor desde maio de 2018, aplicada para países participantes da União Europeia. Apesar de ser uma regulamentação sobre moradores europeus, a GDPR tem impacto mundial, pois trata da proteção dos dados de pessoas (físicas e não falecidas), residentes em países da União Europeia, independentemente da localização ou origem dos terceiros interessados nesses dados, de acordo com o art. 3º da GDPR (2016).

A GDPR define princípios, regras e papéis para utilização de dados, que contemplam desde a definição de indivíduo, razões para obtenção dos dados, consentimento para uso dos dados, anonimização de dados para processamento, direito de arrependimento do consentimento, segurança dos dados, especifica prazos

para reportar vazamentos de informações, no art. 33 GDPR (2016) e eventuais punições em caso do não cumprimento das normas. O tempo para adequação às normas de proteção de dados especificadas pela GDPR foi de dois anos, entre a proposta e a implementação.

No Brasil, temos a LGPD (Lei Geral de Proteção de Dados Pessoais) (BRASIL, 2018), sancionada em agosto de 2018, e vigorada em agosto de 2020. Discussões sobre proteção de dados existem no Brasil desde 2010 (SERPRO). Algumas dessas propostas e discussões se tornaram leis: em 2011 a Lei nº 12.527 de 18/11/2011, chamada “Lei de Acesso à Informação” (BRASIL, 2011). No ano de 2012 a Lei nº 12.737 de 30/11/2012, também chamada “Lei Carolina Dieckmann” (BRASIL, 2012) sobre a tipificação de crimes cibernéticos. Em 2014 é sancionada a Lei nº 12.965, de 23/4/2014, o “Marco Civil da Internet”, que estabelece princípios e deveres para uso da internet (BRASIL, 2014), até chegar à mais recente lei, a LGPD.

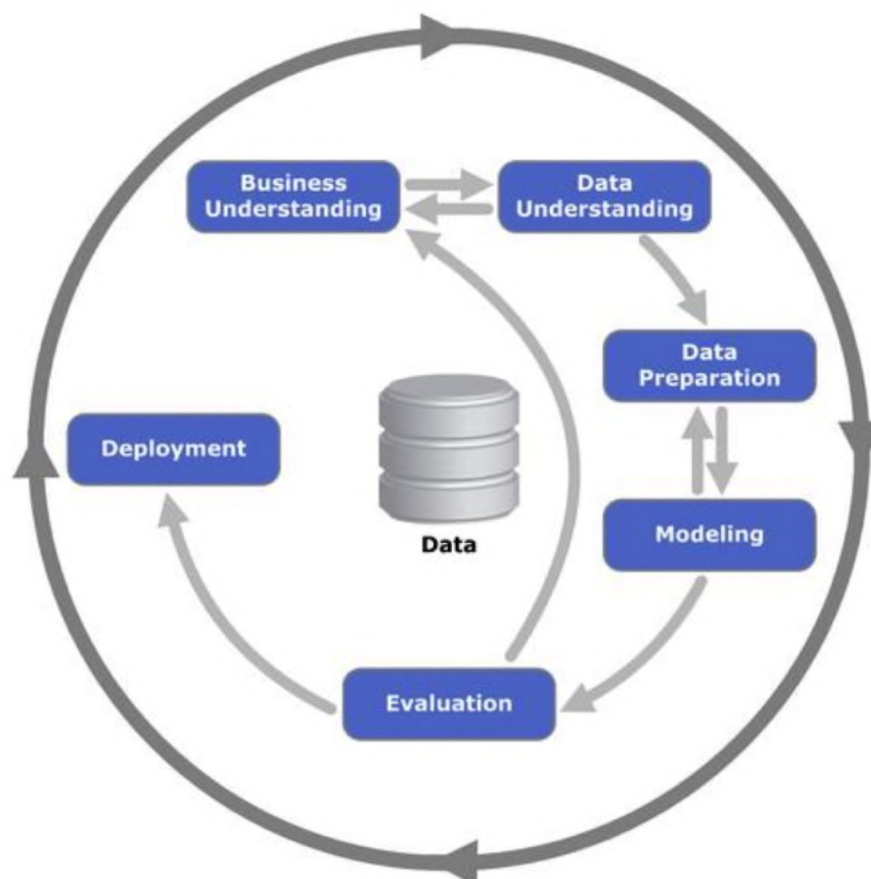
A LGPD, assim como a GDPR, tem por objetivo regulamentar o tratamento de dados pessoais, definindo requisitos para que esses dados sejam processados por pessoas físicas ou jurídicas, fundamentando-se em valores como o direito à liberdade e privacidade (BRASIL, 2020). Para a LGPD, existem conjuntos diferentes de dados, sendo eles os “dados pessoais sensíveis”, descritos no art. 5º, II (BRASIL, 2018) que incluem informações raciais, religiosas, políticas, informações médicas e genéticas e “dados pessoais”, estes descritos no art. 5º, I, que são informações relacionadas à pessoa física de forma que seja possível identificá-la, como por exemplo nome e CPF. Também são definidos papéis, direitos e diretrizes, contemplando o direito de confirmação de processamento dos dados, anonimização, consentimento e revogação do consentimento, correção de dados incorretos, conceitos definidos no Artigo 5º, V a X (BRASIL, 2018) e eventuais punições em caso de descumprimento da LGPD, definidas nos arts. 52 a 54 (BRASIL, 2018).

Devido ao contexto atual de pandemia da Covid-19, as sanções em casos de descumprimento, passarão a ser aplicadas a partir de agosto de 2021, de acordo com o Artigo 65, inciso I-A da LGPD, seguindo a Lei nº 14.010 (BRASIL, 2020), dando assim um tempo maior de adequação sem punições.

2.4 MODELOS DE PROJETO DE CIÊNCIA DE DADOS

O grande volume de dados gerados diariamente por usuários de aplicativos e empresas demanda o uso de técnicas e ferramentas de mineração de dados para que as informações puras se tornem algo mais interpretável (SHAFIQUE e QAISER, 2014) que possa auxiliar nos negócios, seja identificando padrões ou modos de consumidores. Existem diversos modelos utilizados para Ciência de Dados, sendo os mais comuns o modelo KDD (*Knowledge Discovery Databases*), o modelo SEMMA (*Sample, Explore, Modify, Model e Assess*) e o modelo CRISP-DM (*Cross-Industry Standard Process for Data Mining*).

Os 3 modelos são muito similares, sendo o CRISP-DM o mais amplamente utilizado por ser mais completo em comparação aos outros. O modelo CRISP-DM consiste em seis passos, conforme indicado na Figura 1 (CHAPMAN et al., 2000).



Fonte: Jensen (2012)

Figura 1 - Etapas do modelo CRISP-DM

As etapas do modelo não são rígidas, ou seja, é possível retornar a alguma etapa anterior caso seja necessário alterar alguma coisa. As fases são as seguintes:

Entendimento do Negócio (*Business Understanding*) - É a etapa inicial e também a mais importante do projeto, é a fase na qual são pensados os objetivos do projeto, identificadas quais as necessidades do cliente, e um plano inicial é pensado para tentar alcançar esses objetivos. Também devem ser considerados nessa fase recursos, requisitos, critérios de sucesso e avaliados os riscos (CHAPMAN et al., 2000).

Entendimento dos Dados (*Data Understanding*) - Fase ligada à anterior, onde ocorrem outros passos para familiarização com os dados que serão utilizados, tais quais a coleta inicial, análise exploratória, validação e descrição desses dados e eventual formação de hipóteses para resolução do problema (CHAPMAN et al., 2000).

Preparação dos Dados (*Data Preparation*) - Consiste de atividades ligadas ao tratamento dos dados, incluindo a seleção do que será de fato usado como *dataset*, limpeza e transformação caso seja necessário (CHAPMAN et al., 2000).

Modelagem (*Modeling*) - Etapa na qual são selecionados os modelos que serão aplicados e testados. É possível testar vários modelos e avaliá-los posteriormente em outra fase, de acordo com o guia CRISP-DM 1.0 de Chapman et al. (2000), na fase de modelagem pode ser necessário retornar com frequência à fase 3 dependendo do modelo escolhido.

Avaliação (*Evaluation*) - Nesta fase, após criados um ou mais modelos, é necessário avaliá-los antes da implantação. A avaliação consiste em identificar se o modelo escolhido está apto a cumprir os objetivos que foram definidos na primeira fase, caso não esteja, é necessário voltar à primeira etapa e rever o escopo e/ou objetivos. Em caso de resultado positivo na avaliação dos critérios de sucesso, pode-se seguir para a sexta etapa (CHAPMAN et al., 2000). Também deve-se revisar e validar se nenhum passo que foi definido anteriormente foi deixado de lado.

Implantação (*Deployment*) - Chapman et al. (2000), cita que a implantação do modelo não é o fim do projeto. É necessário definir um plano de monitoramento e adaptação, pois o cliente muitas vezes não tem entendimento dos dados raw. É preciso transformar os dados em algum relatório mais compreensível em caso de necessidade.

3 PROPOSTA DE MAPEAMENTO CRISP-DM E LGPD

Seguindo os passos do Modelo CRISP-DM para desenvolvimento de um projeto de Ciência de Dados, propõe-se nesta seção indicar e relacionar a LGPD (BRASIL, 2018) às etapas do modelo, para que o projeto já seja pensado em acordo com a Lei de Proteção de Dados Pessoais.

3.1 MAPEAMENTO

3.1.1 Etapa 1 – Entendimento do Negócio

Por ser a etapa que rege as outras, é preciso considerar muitos cenários possíveis de aplicação da LGPD e cuidar de algumas questões, avaliando as necessidades do escopo do projeto e pensando como se adequar. O entendimento do negócio é a base para as outras atividades, portanto é esperado que se relacione com a maioria dos itens da Lei nº 13.709 de 14/8/2018 (BRASIL, 2018).

A começar pelo art. 3º I, II e III, pois o projeto terá os dados tratados em território brasileiro (I), caso a finalidade do projeto seja o fornecimento de bens ou serviços/fins comerciais (II) e os dados forem coletados em solo brasileiro (III). É necessário atentar-se às definições dos papéis e atividades relacionadas aos dados. Essas definições são importantes para que se compreenda o texto (GONZÁLEZ, 2019) e torne possível adequar um projeto corretamente. O art. 5º define os conceitos, papéis e atividades que serão aplicados aos dados, nos incisos I e II diferencia os tipos de dados; os incisos V a XIX definem papéis sobre a utilização dos dados e termos importantes de atividades executadas durante o tratamento dos mesmos.

Para o art. 6º, deve-se cuidar com os incisos I, delimitar para qual finalidade os dados serão tratados; II, os dados deverão ser tratados apenas para a finalidade definida; III, limitar o tratamento dos dados ao mínimo necessário para cumprir a finalidade definida e informada ao titular dos dados; IV, garantia de livre acesso à consulta dos dados caso solicitado pelos titulares; IX, não utilização dos dados para fins abusivos e/ou discriminatórios. O titular dos dados tem o direito de ter acesso a informações sobre o tratamento dos mesmos, algumas das informações que devem ser cedidas caso sejam solicitadas são indicadas no art. 9º, II, o responsável pelas

decisões que serão tomadas no tratamento dos dados. Para utilizar dados pessoais em um projeto, é importante atentar-se ao art. 7º, I, que declara que os dados apenas poderão ser utilizados caso o titular dê seu consentimento. O consentimento deve ser fornecido por escrito ou de algum modo que mostre a vontade do titular em compartilhar seus dados, em acordo com o art. 8º, §§ 1º e 2º. O parágrafo 5º do art. 8º estabelece que a revogação do consentimento de uso dos dados pode ser solicitada a qualquer momento pelo titular, portanto também é um item a ser cuidado quando se pensa nos riscos do projeto na fase de entendimento.

Em caso da necessidade de utilização de dados pessoais sensíveis para o projeto, como informações médicas, étnicas ou religiosas (definidos pelo art. 5º, II), é importante preferir anonimizar esses dados, seguindo o art. 12, § 1º dados sensíveis que estejam anonimizados, serão considerados apenas dados pessoais, desde que não seja possível reverter esse processo. Para fins de planejamento, é importante considerar o art. 15, I a III, que especificam que o término do tratamento dos dados pode acontecer assim que a finalidade for completada (I), quando o período para tratamento acabou (II), se o titular dos dados solicitar a revogação do consentimento (III). Após o término do tratamento dos dados, os mesmos devem ser eliminados, considerando no caso de projetos, o art. 16, IV apresenta uma das possibilidades para as quais os dados podem ser preservados, que é para uso exclusivo do controlador (pessoa responsável pela decisão de tratamento dos dados), desde que essas informações permaneçam anonimizadas e não sejam compartilhadas com terceiros.

Em projetos que utilizem inteligência artificial e/ou aprendizado de máquina para tomar decisões de forma automatizada que possam influenciar a vida do titular, utilizando informações do mesmo, para criar perfis de consumo, ou pontuação de crédito, o titular tem o direito de solicitar revisão dessas decisões, solicitando informações sobre procedimentos utilizados para essa automatização, caso assim deseje, conforme prevê o art. 20, § 1º. Durante o entendimento do negócio, é importante saber se em algum momento do processo será necessária a transferência de dados para fora do Brasil e para qual país, a LGPD prevê no art. 33, I que a transferência é somente permitida para países que possuam a proteção de dados adequada ao que está previsto na Lei.

Durante o processo de entendimento, uma atividade importante é o controlador dos dados indicar um encarregado pelos mesmos e divulgar as informações sobre o escolhido publicamente. A pessoa que será o canal de

comunicação entre os envolvidos no tratamento dos dados, em acordo com o art. 41, § 1º. Outro ponto de atenção é a questão da responsabilização pelos dados, caso aconteça algo durante o tratamento dos dados que prejudique o titular de alguma forma, o mesmo deverá ser indenizado. Os responsáveis podem responder individual ou coletivamente (art. 42, § 1º, I e II e § 3º).

Todas as medidas de segurança dos dados devem ser consideradas. Os sistemas devem ser estruturados, atendendo aos requisitos de cibersegurança, conforme o art. 49. Essas medidas de segurança devem ser asseguradas desde o planejamento do projeto até sua entrega, conforme art. 46, § 2º. É imprescindível que não haja incidente de vazamento de dados. Caso ocorra, é necessário comunicar o incidente à autoridade nacional e também notificar os titulares, informando o tipo de dados afetados, informações sobre os titulares, indicação das medidas utilizadas para proteger esses dados, quais os riscos ocasionados por esse incidente e quais medidas serão tomadas para remediar ou reverter os efeitos do vazamento, conforme descrito no art. 48, § 1º, I a IV e VI, respectivamente.

A etapa de entendimento do negócio é a que irá ditar e ser utilizada como base para pensar em todo o projeto e executar as atividades necessárias, portanto, mesmo que não chegue a ser aplicado tudo que a LGPD estabelece, por estar fora do escopo, é importante considerar e conhecer os artigos.

Algumas perguntas que podem ser feitas como guias para auxiliar na etapa de entendimento de negócio: qual o objetivo a ser alcançado? Quais os requerimentos? Quais os riscos? O custo e os benefícios? Quais as ferramentas a serem utilizadas? Qual será o período necessário para retenção dos dados?

3.1.2 Etapa 2 – Entendimento dos Dados

Nesta etapa é importante pensar que, apesar de a fase anterior ter mapeado vários itens da LGPD (BRASIL, 2018), a Lei se aplica a todas as etapas, não apenas à de entendimento do negócio. Mesmo esta não sendo a etapa de seleção e limpeza dos dados, os mesmos ainda são manipulados para validação, descrição e análise exploratória, portanto é necessário se atentar a alguns aspectos da Lei. Em caso de tratamento de dados pessoais sensíveis, estes devem ser mapeados para posterior processo de anonimização na etapa de preparação dos dados.

É preciso se atentar para a origem dos dados, se foram coletados em território brasileiro, segundo o art. 3º, III. Caso os dados sejam provenientes de outro país, eles devem ser tratados em acordo com as regulamentações do país de origem, de acordo com o especificado no art. 4º, IV.

Mesmo na fase de entendimento, é importante atentar-se se há consentimento por parte do titular para a utilização dos dados para a finalidade que foi decidida durante o entendimento do negócio, em acordo com o art. 7º, I. Também é importante manter registro das operações feitas para tratamento dos dados (art. 37).

As medidas de segurança para proteção dos dados devem ser tomadas desde a etapa de entendimento, incluindo garantia de infraestrutura adequada para evitar acessos que não foram autorizados, situações de destruição, invasão de sistema, perda de dados, etc., conforme descrito no art. 46, caput do artigo.

Perguntas que podem servir de apoio à etapa de entendimento dos dados: os dados foram coletados em território nacional? Há consentimento do titular para uso dos dados? A finalidade de utilização foi explicitada? O período de retenção dos dados foi informado ao titular? Os direitos do titular sobre seus dados foram informados? Quais dados precisariam ser anonimizados para posterior análise?

3.1.3 Etapa 3 – Preparação dos Dados

Esta etapa de preparação é a que de fato tem manipulação dos dados para as atividades de tratamento, incluindo a seleção do que será utilizado no *dataset* para a modelagem. Nesta etapa, assim como nas anteriores, é preciso que haja consentimento por escrito pelo titular dos dados, em acordo com o art. 7º, II. Todo titular tem o direito assegurado pelo art. 18, caput do artigo da Lei 13.709 (BRASIL, 2018) de solicitar informações sobre a utilização e tratamento de seus dados. Essas informações devem ser disponibilizadas de forma clara para o titular, conforme art. 9º caput do artigo. Em caso de uso de dados pessoais sensíveis, mapeados na etapa de entendimento dos dados, os mesmos devem passar pelo processo de anonimização, evitando assim a identificação do titular.

Assegurar a transparência do tratamento dos dados, como prevê o § 2º do art. 10 é importante para provar que os dados estão sendo utilizados apenas para os fins definidos na etapa de entendimento do negócio.

Medidas de segurança para proteção dos dados também devem ser levadas em conta nesta etapa. Deve-se pensar na infraestrutura, medidas técnicas e administrativas adequadas para evitar incidentes, e estas medidas devem ser cumpridas desde o planejamento do projeto, conforme § 2º do art. 46 e caput deste artigo.

Para esta etapa, alguns questionamentos que podem apoiar a atividade: foram documentadas as atividades de exploração e descrição dos dados? Os dados disponíveis estão completos? Os dados disponíveis podem ser utilizados para se alcançar o objetivo do projeto? Existem dados sensíveis a serem anonimizados para utilização?

3.1.4 Etapa 4 – Modelagem

A etapa de modelagem abrange os testes de modelos utilizando os *datasets* preparados na etapa anterior, de preparação dos dados. Assim como nas outras etapas, é muito importante atentar-se ao art. 49 da Lei Geral de Proteção de Dados Pessoais (BRASIL, 2018). Os sistemas utilizados para tratamento de dados devem ser estruturados dentro dos requisitos de segurança e boas práticas, a fim de evitar incidentes de vazamentos de informações e também evitar danos para os titulares dos dados.

Como os titulares dos dados podem solicitar informações sobre o uso e tratamento de seus dados, é importante manter registradas as operações executadas durante o tratamento dos dados, seguindo o art. 37. Nesta etapa também é necessário atentar-se às medidas de transparência que foram adotadas e definidas nas etapas anteriores.

Na etapa de modelagem, pode-se pensar nas seguintes questões: os dados estão adequados para os modelos? A documentação permanece sendo feita? Protocolos de segurança dos dados continuam funcionais?

3.1.5 Etapa 5 – Avaliação

A etapa de avaliação consiste em validar se os modelos escolhidos na fase quatro estão aptos a cumprir os objetivos definidos durante o entendimento do negócio e, caso seja verificado que não cumpre, é preciso retornar à primeira etapa para rever o escopo ou os objetivos, portanto, a finalidade do projeto se altera. A cada alteração de escopo ou objetivos, é preciso seguir o § 6º do art. 8º, e informar os titulares sobre a alteração, solicitando novamente o consentimento para o uso dos dados. O titular dos dados pode revogar o consentimento após ser informado sobre a mudança de finalidade, caso não concorde com as alterações, em acordo com o art. 9º, § 2º.

Caso o modelo escolhido para atender o escopo do projeto possua alguma forma automatizada para tomada de decisões, incluindo inteligência artificial ou *machine learning*, para gerar informações de perfil de consumo, por exemplo, o titular dos dados pode solicitar a revisão das decisões que foram tomadas. Além disso, essas informações devem ser fornecidas sempre que solicitadas em acordo com o § 1º do art. 20.

Na avaliação do modelo, quando observado que o mesmo cumpre os objetivos definidos na primeira etapa de entendimento, considera-se, segundo o art. 15, I que o tratamento dos dados chegou a seu fim.

As seguintes perguntas guia podem ser pensadas na etapa de avaliação: o resultado do modelo foi satisfatório, alcançando o objetivo? O que será alterado caso não tenha alcançado o objetivo? A possibilidade de alteração de escopo havia sido explicitada para o titular?

3.1.6 Etapa 6 – Implantação

A etapa de implantação não significa o final do projeto, pois ainda é preciso definir um plano de adaptação e monitoramento. Mesmo na implantação o titular tem o direito de solicitar a revogação do consentimento, partindo do princípio do art. 8º, § 5º que define que a revogação do consentimento pode ser solicitada a qualquer momento e os dados devem ser eliminados quando solicitado.

As seguintes perguntas podem ser feitas nesta etapa: o tratamento dos dados chegou ao fim? A documentação do tratamento foi feita até a etapa de implantação?

3.2 EXEMPLO DE USO

Um canal de denúncias, também denominado *hotline*, é um mecanismo para receber informações sobre possíveis delitos, interligando assim os diversos atores (internos e externos de uma organização) de um incidente ou delito. A iniciativa de canais de denúncia é aplicada e desenvolvida desde meados de 1996 como resposta à existência de materiais ilegais na Internet (WILLIAMS, 1999). Ainda segundo Williams (1999), não existe um modelo padronizado de *hotline*, especificando o que deve ou não fazer parte deste mecanismo, pois a necessidade é algo particular de cada segmento que usará desse modelo para receber denúncias.

Canais para denúncias em ambiente laboral são muito importantes no auxílio à prevenção e tratativas de delitos. Existem diversos meios de se receber uma denúncia. Alguns dos mais comuns são através de página na internet, ligação telefônica ou e-mails. Permitir o acesso a esse mecanismo incentiva o reporte de situações antiéticas.

Por ser um produto que recebe diretamente informações de pessoas naturais, o processo de planejamento de um Canal de Denúncias será utilizado como exemplo de adequação à Lei 13.709 de 14/8/2018 (BRASIL, 2018).

A fim de fazer o mapeamento de conformidade à LGPD, seguindo a metodologia CRISP-DM para desenvolvimento de projetos, neste exemplo será utilizada uma empresa fictícia chamada “ATZ”, que possui como meio de contato para denúncias uma página *web* com formulários divididos em temas pré-definidos com espaço para adicionar outros temas caso o relatante sinta que as opções não representam sua necessidade. Além da averiguação dos relatos de forma interna, feito pela área de *compliance*, a intenção ao se analisar os dados é obter um relatório contendo informações estatísticas sobre os relatos recebidos, para que posteriormente este relatório possa ser utilizado como objeto para métricas de utilização. O exemplo irá considerar para fins de objetivo a exploração dos dados com o intuito de analisar as denúncias recebidas, verificar quais os temas mais recorrentes e, através de análise do histórico de denúncia/tratativa aplicada, sugerir formas de prevenção a fim de evitar ainda mais recorrências.

3.2.1 Etapa 1 – Entendimento do negócio

Para auxiliar no entendimento, pode-se imaginar perguntas para guiar o processo, pensando na solução e nos pontos da LGPD a serem considerados.

Nesta etapa é preciso definir quais os objetivos do projeto, avaliar riscos, requerimentos, avaliar quais ferramentas precisarão ser utilizadas inicialmente e também definir o critério de sucesso esperado.

Perguntas guias: Qual o objetivo a ser alcançado? Quais os requerimentos? Quais os riscos? O custo e os benefícios? Quais as ferramentas a serem utilizadas? (CHAPMAN et al., 2000).

O objetivo da empresa ATZ é receber denúncias de pessoas internas ou externas através de uma página web, encaminhar estas denúncias para a área de *compliance* para tratativa e utilizar as informações das denúncias e sobre as tratativas aplicadas para identificar os temas mais recorrentes e chegar a um modelo que auxilie na prevenção desses delitos mais comuns. Serão necessários recursos de infraestrutura (servidores, banco de dados) e segurança (virtual e física), definir um encarregado pelos dados, garantir que a finalidade do uso dos dados esteja explícita, permitindo que o denunciante consinta ou não com essa utilização.

Para que esteja em conformidade com a LGPD e que a Lei se aplique aos dados coletados, é preciso que o tratamento seja feito em território brasileiro ou que a coleta tenha sido feita em território nacional (Lei 13.709, Art. 3º, I e III). Para estar em acordo com o art. 46, § 2º, a ATZ precisa assegurar o tratamento com segurança dos dados, desde a concepção do produto até sua entrega, uma boa prática é descrever publicamente quais os esforços tomados para garantir o máximo de segurança possível.

3.2.2 Etapa 2 – Entendimento dos dados

Esta etapa contempla a coleta, descrição, exploração e verificação da qualidade dos dados.

Perguntas guias: Os dados foram coletados em território nacional? Há consentimento do titular para uso dos dados? A finalidade de utilização foi explicitada? O período de retenção dos dados foi informado ao titular? Os direitos do titular sobre seus dados foram informados? Quais dados precisam ser anonimizados para posterior análise?

É imprescindível que a ATZ obtenha o consentimento do titular para a utilização dos dados e que informe o prazo de retenção, finalidade do tratamento e os direitos do titular, seguindo os princípios de transparência no tratamento dos dados (art. 10, § 2º), e em acordo com o art. 18, I a IX, que prevê o direito de o titular fazer solicitações sobre os dados, incluindo informações, acesso aos dados, correções de informações erradas, anonimização e revogação do consentimento, entre outros. É necessário documentar todos os passos de manipulações feitas nesta fase durante a análise exploratória, seguindo o art. 37.

A finalidade é a análise de dados históricos preenchidos à medida que denúncias são recebidas e tratadas, essa informação precisa estar explícita juntamente com o período de retenção, que pode ser em algum dos casos: revogação de consentimento pelo titular ou enquanto for necessário para a atividade.

A segurança dos dados deve ser levada em conta tanto na etapa anterior quanto nesta (art. 46, caput), é preciso atentar-se à localização dos servidores, quem está autorizado a acessar, quais medidas de segurança físicas e tecnológicas (*firewall*, criptografia, etc.) estão sendo tomadas para proteger os dados

Quanto ao formato dos dados, os mesmos já serão inseridos no banco de dados com formatação padrão, sendo necessário apenas conferir na etapa de preparação qual o formato necessário para os modelos que serão testados.

3.2.3 Etapa 3 – Preparação dos dados

Etapa contempla o tratamento dos dados, limpeza dos mesmos e seleção de quais dados farão parte do *dataset* utilizado para modelagem.

Perguntas guias: Foram documentadas as atividades de exploração e descrição dos dados? Os dados disponíveis estão completos? Os dados disponíveis podem ser utilizados para se alcançar o objetivo do projeto? Existem dados sensíveis a serem anonimizados para utilização?

A ATZ deve manter a documentação de todo o tratamento dos dados, incluindo os processos executados, controles de versão, entre outros, para poder entregar ao titular caso seja solicitado, conforme art. 19, § 3º. Por serem etapas complementares, a de entendimento e preparação dos dados possuem os mesmos quesitos de adequação que devem ser seguidos pela ATZ para estar em conformidade com a Lei 13.709 (BRASIL, 2018).

Como o objetivo é utilizar uma base de dados que contém informações pessoais como nome completo e local de trabalho, para utilizar os dados para modelagem, a ATZ precisa tratar os dados para que não seja possível identificar o titular durante os testes de modelagem.

3.2.4 Etapa 4 – Modelagem

Etapa de testes de modelos para encontrar o que melhor atinge o objetivo do projeto.

Perguntas guia: Os dados estão adequados para os modelos? A documentação permanece sendo feita? Protocolos de segurança dos dados continuam funcionais?

A empresa ATZ precisa manter a transparência dos processos (art. 10, § 2º) documentando o que foi feito durante as atividades dessa etapa (art. 37). Em acordo com o art. 49, é necessário que os sistemas utilizados sejam estruturados para atender os requisitos de segurança e boas práticas previstos na Lei.

Os *datasets* selecionados pela ATZ na etapa de preparação dos dados para o projeto serão testados nos modelos devem estar anonimizados, para que não seja possível identificação de titulares durante os testes e tratamento dos dados.

3.2.5 Etapa 5 – Avaliação

A fase de avaliação verifica se o modelo cumpre os objetivos especificados na etapa de entendimento do negócio, em caso positivo, pode-se seguir para a etapa de implementação. Caso o modelo não cumpra com os objetivos, é preciso retornar à primeira fase e rever os objetivos.

Perguntas guia: O resultado do modelo foi satisfatório, alcançando o objetivo? O que será alterado caso não tenha alcançado o objetivo? A possibilidade de alteração de escopo havia sido explicitada para o titular?

O modelo selecionado cumpre o objetivo inicial definido na etapa de entendimento do negócio, que é auxiliar a sugerir modos de prevenção para delitos denunciados. Posteriormente, se for necessário incluir ou desconsiderar alguma

informação, mudando assim a finalidade de uso dos dados, a ATZ precisa solicitar novamente o consentimento dos titulares, informando quais foram as alterações feitas.

3.2.6 Etapa 6 – Implantação

Nesta fase é realizada a entrega do projeto desenvolvido pela ATZ.

Perguntas guia: O tratamento dos dados chegou ao fim? A documentação do tratamento foi feita até a etapa de implementação?

É importante considerar que a Lei Geral de Proteção de Dados Pessoais garante o direito de o titular solicitar a revogação do consentimento do compartilhamento de seus dados a qualquer momento, e pode solicitar a exclusão dos mesmos conforme o art. 8º, § 5º mesmo depois da implementação.

Como a ATZ informou que o período de retenção se estende até o titular solicitar a revogação do consentimento ou até a finalização da atividade, os dados ainda permanecem em tratamento.

A implementação não significa o final do projeto, é preciso ter um plano de acompanhamento e monitoramento posterior. A ATZ permanece com dados sendo inseridos no banco para utilização no modelo e continuidade do projeto.

4 CONCLUSÃO

A discussão sobre ética, apesar de antiga, permanece um tema atual devido às suas várias aplicações em diversos campos, incluindo a computação. Considerando-se o contexto atual do grande volume de dados gerados por máquinas e humanos diariamente e a possibilidade de utilização dessas informações valiosas, surgiu o conceito de *Big Data*.

A disponibilidade para exploração de dados trouxe à tona discussões que já existiam, mas que se tornaram ainda mais relevantes sobre a proteção de dados pessoais. Em 2016 na Europa foi criado um regulamento para o tratamento e uso de informações pessoais, a *General Data Protection Regulation* (GDPR), que entrou em vigor no ano de 2018. Após a regulamentação da GDPR, outros países seguiram a União Europeia e fizeram suas próprias legislações, entre esses países o Brasil, que em 2018 aprovou a Lei Geral de Proteção de Dados Pessoais (LGPD), inspirada na GDPR. A LGPD entrou em vigor efetivamente em 2020, com exceção de alguns artigos sobre as sanções que entrarão em vigor a partir de 1º de agosto de 2021.

O presente trabalho buscou mapear artigos da LGPD seguindo a metodologia CRISP-DM, para que um projeto de Ciência de Dados já seja projetado pensando-se em conformidade com a LGPD. Durante o processo, o mapeamento mostrou-se difícil, pois a própria LGPD possui alguns artigos que não especificam diretamente a utilização, apenas aplicação geral, algo bom por poder abranger diversos usos, mas também que dificulta o entendimento.

A Lei Geral de Proteção de Dados Pessoais deve ser aplicada para qualquer tipo de dado pessoal manipulado, por isso a intenção de guiar o planejamento de um projeto em acordo com a mesma. Por ser uma Lei, a LGPD pode ser atualizada, vetando artigos ou incluindo novos, portanto o objetivo não é definir um guia definitivo, mas sim guiar para pontos de atenção que precisam ser considerados quando se está desenvolvendo um projeto de Ciência de Dados.

REFERÊNCIAS

- BRASIL. **Lei nº 12.527 Lei de Acesso à Informação**, de 18 de novembro de 2011. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>. Acesso em: 10 maio 2021.
- _____. **Lei nº 12.737 Lei Carolina Dieckmann**, de 30 de novembro de 2012. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12737.htm>. Acesso em: 10 maio 2021.
- _____. **Lei nº 12.965 Marco Civil da Internet**, de 23 de abril de 2014. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm>. Acesso em: 10 maio 2021.
- _____. **Lei nº 13.709 Lei Geral de Proteção de Dados Pessoais (LGPD)**, de 14 de agosto de 2018. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>. Acesso em: 10 maio 2021.
- _____. **Lei nº 14.010**, de 10 de junho de 2020. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2020/Lei/L14010.htm>. Acesso em: 10 maio 2021.
- _____. Ministério da Defesa. **Proteção de Dados Pessoais - LGPD**. 2020. Disponível em: <<https://www.gov.br/defesa/pt-br/acesso-a-informacao/lei-geral-de-protecao-de-dados-pessoais-lgpd>>. Acesso em 10 maio 2021.
- BUCHANAN, Elizabeth. A.; ESS, Charles. **Introduction: The Ethics of E-Games**. The International Review of Information Ethics, Edmonton, Canada, v. 4, p. 2–6, 2005. Disponível em: <<http://informationethics.ca/index.php/irie/article/view/16>>. Acesso em: 05 abr. 2021.
- BYNUM, Terrell Ward. 2000. **The foundation of computer ethics**. ACM SIGCAS Computers and Society. v. 30, n. 2, p.6-13, 2000. Disponível em: <<https://doi.org/10.1145/572230.572231>>. Acesso em: 09 abr. 2021.
- CAVE, Stephen; DIHAL, Kanta. **The Whiteness of AI**. Philos. Technol. 2020. v. 33 p. 685–703. Disponível em: <<https://doi.org/10.1007/s13347-020-00415-6>>. Acesso em: 30 abr. 2021.
- CHAPMAN, Pete; CLINTON, Julian; KERBER, Randy; KHABAZA, Thomas; REINARTZ, Thomas; SHEARER, Colin e WIRTH, Rüdiger. 2000. **CRISP-DM 1.0 - Step-by-step data mining guide**, CRISP-DM Consortium. Disponível em: <<https://www.the-modeling-agency.com/crisp-dm.pdf>>. Acesso em: 30 abr. 2021.
- CHAUÍ, Marilena. **Convite à Filosofia**. São Paulo: Ática, 2000. Disponível em: <https://home.ufam.edu.br/andersonlfc/Economia_Etica/Convite%20Filosofia%20-%20Marilena%20Chaui.pdf>. Acesso em: 01 maio 2021.

CHEN, Adrian. **Cambridge Analytica and Our Lives Inside the Surveillance Machine**. 2018. The New Yorker. Disponível em: <<https://www.newyorker.com/tech/annals-of-technology/cambridge-analytica-and-our-lives-inside-the-surveillance-machine>>. Acesso em: 11 jun 2021.

CONSALVO, Mia. **Rule Sets, Cheating, and Magic Circles: Studying Games and Ethics**. The International Review of Information Ethics, Edmonton, Canada, v. 4, p. 7–12, 2005. Disponível em: <<http://informationethics.ca/index.php/irrie/article/view/162>>. Acesso em: 05 abr. 2021.

FIESER, James. **Ethics**. The Internet Encyclopedia of Philosophy. 2021. Disponível em: <<https://iep.utm.edu/ethics/>>. Acesso em: 05 abr. 2021.

GARVIE, Clare; FRANKLE, Jonathan. **Facial-recognition software might have a racial bias problem**. The Atlantic. 2016. Disponível em: <<https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/>>. Acesso em: 22 jun. 2021.

GERT, Bernard; GERT, Joshua. **The Definition of Morality**. The Stanford Encyclopedia of Philosophy (Fall 2020 Edition), Edward N. Zalta (ed.). Disponível em: <<https://plato.stanford.edu/archives/fall2020/entries/morality-definition>>. Acesso em: 12 abr. 2021.

GONZÁLEZ, Mariana. **LGPD Comentada**. 2019. Disponível em: <<https://guialgpd.com.br/lgpd-comentada/>>. Acesso em: 14 abr. 2021.

KURZWEIL, Ray. **The Age of Spiritual Machines: when computers exceed human intelligence**. Nova Iorque: Penguin, 1999. 289 p. Disponível em: <https://jimdo-storage.global.ssl.fastly.net/file/aff560e-b5bf-43df-8a7b-061dfd4e08d3/THE_AGE_OF_SPIRITUAL_MACHINES_WHEN_COMPU.pdf>. Acesso em: 09 maio 2021.

LOHR, Steve. **Facial Recognition Is Accurate, if You're a White Guy**. The New York Times. 2018. Disponível em: <<https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html>>. Acesso em: 14 jun. 2021.

MA, Alexandra; GILBERT, Ben. **What Is Cambridge Analytica? a Timeline of the Facebook Controversy**. 2019. Business Insider. Disponível em: <<https://www.businessinsider.com/cambridge-analytica-a-guide-to-the-trump-linked-data-firm-that-harvested-50-million-facebook-profiles-2018-3>>. Acesso em: 31 maio 2021.

MARCKMANN, Georg; GOODMAN, Kenneth. W. **Introduction: Ethics of Information Technology in Health Care**. The International Review of Information Ethics, Edmonton, Canada, v. 5, p. 2–5, 2006. Disponível em: <<http://informationethics.ca/index.php/irrie/article/view/188>>. Acesso em: 05 abr. 2021.

MOOR, James H. **What Is Computer Ethics?** *Metaphilosophy* v. 16, n. 4, 1985. p. 266-271. Disponível em: <https://www.cs.ucdavis.edu/~koehl/Teaching/ECS188/PDF_files/MOOR-1985-Metaphilosophy.pdf>. Acesso em: 12 abr. 2021.

MOZUR, Paul. **In Hong Kong Protests, Faces Become Weapons**. 2019. *The New York Times*. Disponível em: <<https://www.nytimes.com/2019/07/26/technology/hong-kong-protests-facial-recognition-surveillance.html>>. Acesso em: 15 jun. 2021.

MÜLLER, Vincent C. **Ethics of Artificial Intelligence and Robotics**, *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.). Disponível em: <<https://plato.stanford.edu/entries/ethics-ai/>>. Acesso em: 06 abr. 2021.

NUNES, Pablo. **Maioria dos presos por reconhecimento facial são negros**. *The Intercept*. 2019. Disponível em: <<https://theintercept.com/2019/11/21/presos-monitoramento-facial-brasil-negros/>>. Acesso em: 14 jun. 2021.

ORTS, Adela Cortina; NAVARRO, Emilio Martínez. **Ética**. 2001. Resumo Prof. Dr. Roque Junges (UNISINOS). Disponível em: <http://www.projeto.unisinos.br/humanismo/etica/etica_adela_cortina.pdf>. Acesso em: 27 abr. 2021.

SERPRO. **Linha do tempo da proteção de dados pessoais e da Lei Geral de Proteção de Dados Pessoais, no Brasil**. Disponível em: <<https://www.serpro.gov.br/lgpd/menu/arquivos/linha-do-tempo-1/view>>. Acesso em: 10 maio 2021.

SHAFIQUE, Umair; QASIR, Haseeb. **A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA)**. 2014. *International Journal of Innovation and Scientific Research*. v. 12, n. 1, p. 217-222. Disponível em: <<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.686.840&rep=rep1&type=pdf>>. Acesso em: 15 jun. 2021.

SMITH, Matthew; SZONGOTT, Christian; HENNE, Benjamin; VON VOIGT, Gabriele. **Big data privacy issues in public social media**. 2012. 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST). Disponível em: <<https://ieeexplore.ieee.org/abstract/document/6227909>>. Acesso em: 11 jun. 2021.

UNIÃO EUROPEIA. **Regulamento 2016/679 General Data Protection Regulation**. 2016. Disponível em: <<https://gdpr-info.eu/>>. Acesso em: 09 maio 2021.

WEX | LII / Legal Information Institute. **Ethics**. Disponível em: <<https://www.law.cornell.edu/wex/Ethics>>. Acesso em: 05 abr. 2021.

WHELCHER, Robert. J. **Is Technology Neutral?** *IEEE Technology and Society Magazine*, v. 5, n. 4, p. 3–8. 1986. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/5010049>>. Acesso em: 11 jun. 2021.

WIENER, Norbert. **The Human Use of Human Beings: cybernetics and society.** Londres: Eyre And Spottiswoode, 1954. 199 p. Disponível em: <http://asounder.org/resources/weiner_humanuse.pdf>. Acesso em: 20 abr. 2021.

WILLIAMS, Nigel. **The Contribution Of Hotlines To Combating Child Pornography On The Internet.** 1999. Disponível em: <<https://www.childnet.com/ufiles/combating-child-pornography.pdf>>. Acesso em: 22 jun. 2021.