

USO DE INTELIGÊNCIA ARTIFICIAL PARA VERIFICAR A CONFORMIDADE DE TERMOS DE PRIVACIDADE COM A LEI GERAL DE PROTEÇÃO DE DADOS

Deivid Valença de Araujo, Gustavo Luiz Silva Lima, Lucas Ricardo Ribeiro Souza, Roger William Belas Cruz, Vinicius de Oliveira Mendes Lima¹

Orientador: Professor Me. Gilberto Alves Pereira

Faculdade Impacta
São Paulo, SP, Brasil
Junho de 2022

RESUMO

Com a entrada da Lei Geral de Proteção de Dados (LGPD) em vigor desde 2020, as empresas precisaram adequar as políticas de privacidade. O objetivo do presente trabalho foi verificar a possibilidade de definir se os textos dos termos de privacidade estão aderentes à LGPD, utilizando um algoritmo de Aprendizado de Máquina. Foram selecionados 140 textos de termos de privacidade, sendo metade aderentes à LGPD; foram aplicados os algoritmos Naive Bayes e Regressão Logística. Nos resultados foi observado um desempenho de 86% de precisão do algoritmo Naive Bayes e 92% de precisão para a Regressão Logística. Em conclusão, foi observado que é possível utilizar Machine Learning para identificar se os termos estão ou não aderentes à LGPD, porém o trabalho precisa ser replicado com um maior volume de dados e, a princípio, essa técnica pode ser utilizada como forma de triagem.

Palavras-Chave: Lei Geral de Proteção dos Dados, Machine Learning, Privacidade.

ABSTRACT

With the beginning of the General Data Protection Law (GDPL) in 2020, companies must adapt their privacy policies. The objective of this work is to verify the possibility of defining if privacy terms are compliant to the GDPL, using a Machine Learning algorithm. 140 privacy terms were selected, half of which adhered to the GDPL, the Naive Bayes and Logistic Regression algorithms were applied. In the results, it was observed a performance of 86% in the precision of the Naive Bayes algorithm and 92% of precision for the Logistic Regression. In conclusion, it was observed that it is possible to use Machine Learning to identify whether or not the terms are compliant with the GDPL, but the work needs to be replicated with a greater volume of data.

Keywords: General Data Protection Law, Machine Learning, Privacy.

1. INTRODUÇÃO

A Lei Geral de Proteção de Dados Pessoais (LGPD) é a lei que regula as atividades de tratamento de dados pessoais na internet (ALECRIM, 2018). A

¹ Os autores podem ser contatados respectivamente pelos seus correios eletrônicos: deivid.araujo@aluno.faculdadeimpacta.com.br, gustavo.luiz@aluno.faculdadeimpacta.com.br, lucas.ribeiro@aluno.faculdadeimpacta.com.br, roger.cruz@aluno.faculdadeimpacta.com.br, vinicius.mendes@aluno.faculdadeimpacta.com.br

legislação criou novos conceitos jurídicos e definiu as condições nas quais os dados pessoais podem ser tratados. Estabelecendo papéis, condutas e responsabilidades para que haja maior cuidado com o tratamento e compartilhamento de dados pessoais (ALECRIM, 2018).

A LGPD foi publicada em 14 de agosto de 2018 e entrou em vigor em 18 de setembro de 2020. A princípio as empresas teriam 24 meses para se adequar, porém devido a pandemia de Covid-19 este tempo foi estendido e as sanções começaram em 1º de agosto de 2021. Portanto, a partir da entrada da lei em vigor as empresas precisaram iniciar o processo de adequação para evitar as sanções administrativas (PEREIRA; SILVA, 2022). Por meio de mapeamento dos dados é possível entender quais dados cada área manipula e onde trafegam dados sensíveis e pessoais, a partir deste ponto a empresa poderá traçar estratégias de como tratar os dados, isto envolve a revisão dos termos de privacidade (MALLMANN, 2020).

Políticas de privacidade podem ser definidas como uma descrição abrangente das práticas de coletas de dados de um site ou aplicação. Uma política de privacidade define quais informações são coletadas, a finalidade da coleta das informações e como os dados serão manipulados, armazenados e utilizados pela aplicação. Também é responsável por fornecer as informações a respeito das permissões de uso das informações pessoais coletadas e para resolver disputas relacionadas à privacidade com o site. Os termos de privacidade são a ferramenta de aplicação das políticas de privacidade (TERRA, 2021)

Para estar de acordo com a LGPD, a estratégia de coleta e tratamento dos dados devem estar explícitos nos termos de privacidade, o uso de Machine Learning pode auxiliar no processo de adequação das empresas à LGPD, pois o processo de revisão e ajustes dos termos de forma manual pode ser muito oneroso. O presente trabalho teve como objetivo conferir a possibilidade de verificar se Termos de Privacidade estão aderentes à LGPD, utilizando um algoritmo de Aprendizado de Máquina.

2. DEFINIÇÕES DA LGPD

Para melhor entendimento do trabalho é preciso conhecer as definições de dados, assim como os principais papéis e responsáveis que foram estabelecidas pela lei.

Dado pessoal é a informação referente a pessoa natural identificada ou identificável. Dado pessoal sensível é um dado sobre origem racial, étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político. Também são considerados dados sensíveis os dados referentes à saúde, vida sexual, genético e biométrico, quando estes estiverem vinculados a uma pessoa natural. Dado anonimizado é um dado relativo a um titular que não possa ser identificado. Banco de dados é definido como um conjunto estruturado de dados pessoais, estabelecido em um ou mais locais, podendo ter suporte eletrônico ou físico (BRASIL, 2018).

A lei descreve os seguintes papéis dos envolvidos e responsáveis:

- Titular é a pessoa natural a quem se referem os dados pessoais que são objeto de tratamento.
- Controlador é a pessoa responsável pelas decisões de tratamento dos dados.
- Operador é a pessoa que realiza o tratamento dos dados pessoais em nome do Controlador.
- Encarregado é a pessoa indicada pelo Controlador e Operador para ser o canal de comunicação entre o Controlador, o Titular e a Autoridade Nacional de Proteção de Dados (ANPD).
- A ANPD é o órgão da administração pública responsável por zelar,

implementar e fiscalizar o cumprimento da lei.

Os papéis de Controlador e Operador podem ser assumidos por pessoa natural ou jurídica de direito público ou privado (BRASIL, 2018).

2.1 TERMOS DE PRIVACIDADE

O Termo de Privacidade é o documento que faz parte da política de privacidade de uma empresa. Ele contém as práticas e medidas de Privacidade e Segurança adotadas pela organização. Seu objetivo é fornecer informações sobre como a empresa obtém, utiliza, armazena e protege os dados pessoais que coleta, a fim de levar transparência aos donos dos dados pessoais que possui (NONES, 2021).

Como um dos pilares da LGPD é oferecer transparência ao titular dos dados, os termos de privacidade devem abordar os seguintes tópicos: como os dados pessoais são recolhidos, quais dados estão sendo recolhidos, qual a finalidade do uso dos dados, por quanto tempo os dados ficarão armazenados, o que será utilizado para garantir a segurança dos dados coletados, se os dados serão ou não compartilhados ou transferidos a terceiros, orientação sobre atualizações da política e informações de como o titular deve entrar em contato com a empresa (BRASIL, 2018).

3. APRENDIZADO DE MÁQUINA

Aprendizado de Máquina ou *Machine Learning* (ML) é uma área da inteligência artificial que tem como objetivo o desenvolvimento de técnicas computacionais sobre o aprendizado e a construção de sistemas capazes de adquirir conhecimento de forma automática. Um sistema de ML toma decisões baseado em experiências acumuladas por meio de resoluções bem-sucedidas de problemas anteriores (MONARD; BARANAUSKAS, 2003).

Existem paradigmas para serem usados no processo de aprendizado do algoritmo. O Simbólico, em que o sistema aprende a partir da construção de representações simbólicas de um conceito por meio da análise de exemplos e contra-exemplos. O método Estatístico, como o próprio nome sugere, usa modelos estatísticos para encontrar uma aproximação do conceito induzido; dentro deste método há diversas técnicas, que são utilizadas de acordo com o objetivo do algoritmo. Há o método chamado de Baseado em Exemplos que consiste em relacionar um exemplo com outro similar apresentado anteriormente, assumindo dessa forma que eles terão a mesma classe. O método Conexionista recebe este nome pois remete a redes neurais, que envolve unidades altamente interconectadas; a comunidade científica acredita que as redes neurais possuem um grande potencial na resolução de problemas que requerem um intenso processamento sensorial humano (MONARD ; BARANAUSKAS, 2003).

4. USO DE MACHINE LEARNING NO PROCESSAMENTO DE TEXTOS

O presente estudo teve como base alguns trabalhos realizados na área de processamento textual.

Há métodos para análise discursiva automática, por meio da teoria da estrutura retórica desenvolvida nos anos de 1980 por William Mann, Sandra Thompson e Christian Matthiessen. A partir da anotação retórica e da análise de diversos textos científicos da área da computação, o autor desenvolveu o primeiro analisador retórico da língua portuguesa, chamado de *Discourse Analyzer* (DiZer), que é capaz de reconhecer relações discursivas baseados em unidades complexas

de conteúdos (PARDO, 2005).

Pesquisas apresentaram a aplicação de uma nova abordagem de ML baseada em argumentação aplicada à área jurídica. O experimento usou um conjunto de dados já utilizado em experimentos anteriores. Este método demonstrou robustez a ruídos em dados de aprendizado (MOŽINA et al. 2006).

Teodorescu (2017) utilizou o método de processamento de linguagem natural para a análise de textos. Com foco em textos da área da administração, o autor focou na análise de textos e métodos de aprendizado com o intuito de pesquisa estratégica na área de gestão.

Durante o congresso *Developers Conference* sediado em Porto Alegre-RS no ano de 2019, foi apresentada a aplicação de ML para a classificação de um conjunto de textos jurídicos com o intuito de categorizá-los como sendo do âmbito cível ou criminal. O processo de teste do algoritmo foi realizado com 2400 amostras de texto (PACHECO, 2019).

Estudiosos descreveram as etapas do processo de mineração de texto e a análise de diferentes métodos de classificação, utilizando algoritmos de aprendizado de máquina supervisionados a partir de jurisprudência. Os autores concluíram que por meio da utilização da técnica de aprendizado de máquina supervisionado é possível apresentar informações relevantes de um texto sem a necessidade de leitura integral (MEDEIROS ; SILVA 2020).

Scikit-learn é uma biblioteca de código aberto para aplicar ML, ela inclui diversos algoritmos para classificação e regressão (PEDREGOSA et al., 2011).

Um dos principais algoritmos do Scikit-learn é o Naive Bayes (NB), um classificador multinomial utilizado para classificação com características discretas. Este algoritmo é utilizado para categorizar texto baseado na frequência das palavras usadas. Sua principal característica é desconsiderar a correlação entre as variáveis (RISH, 2001).

O kit de linguagem natural (NLTK) é um conjunto de bibliotecas e programas utilizados no processamento de linguagem natural (BIRD; KLEIN; LOPES, 2019).

A regressão logística é uma técnica estatística que busca estimar a probabilidade de a variável dependente assumir um determinado valor em função dos outros valores. Uma das vantagens dessa técnica está no fato de os resultados serem mostrados em probabilidade, precisa apenas de um pequeno número de suposições para realizar as previsões e possui um alto grau de confiabilidade (SANTOS, 2019).

Com base na publicação de CHEN et al. (2020) as métricas de ML são: acurácia, que serve como uma boa indicação geral do desempenho do modelo. Ela mensura quanto dos exemplos foram de fato classificados corretamente, independente da classe. Essa métrica é definida pela razão entre o que o modelo acertou e todos os exemplos. E a precisão, que é a métrica mais comum para avaliar modelos de classificação, há uma ênfase maior para os erros por falso positivo. Essa métrica é definida pela razão entre a quantidade de exemplos classificados corretamente como positivos e o total de exemplos classificados como positivos.

5. METODOLOGIA

Para o presente estudo foi desenvolvido um experimento, que teve como base utilizar um algoritmo de Machine Learning treinado com termos que estão de acordo com a LGPD.

5.1 MATERIAIS

Para o desenvolvimento do estudo foi utilizada a plataforma *Google Colab*, a linguagem de programação *Python*, os algoritmos Naive Bayes e Regressão Logística da biblioteca *Scikit-learn*, além do NLTK para processamento de linguagem natural.

5.2 PROCEDIMENTOS

Foram selecionados 70 (setenta) termos de privacidade que já estão aderentes à LGPD e 70 (setenta) termos que não estão aderentes à LGPD, totalizando uma amostra de 140 (cento e quarenta) textos. Para definir se um termo estava aderente à lei foram utilizados os principais tópicos que a lei exige serem abordados nos termos. Os tópicos são: informações gerais sobre o termo, como os dados são recolhidos, quais os dados pessoais são recolhidos, qual a finalidade da coleta dos dados, por quanto tempo os dados ficarão armazenados, quais as medidas de segurança serão tomadas, se os dados serão compartilhados e com quem, se os dados serão transferidos para terceiros, definição de cookies, pedido de consentimento, aviso sobre alterações da política de privacidade, jurisdição para resolução de conflitos e além desses o termo deve ter data igual ou superior à 2018 (BRASIL, 2018). O Anexo A deste artigo exemplifica um termo de privacidade que está de acordo com a LGPD.

Todos os termos que possuíam todos os tópicos foram entendidos como aderentes. Os termos que não apresentaram os tópicos exigidos pela lei ou com data anterior à 2018 foram considerados não aderentes, exemplo em Anexo B.

A busca dos termos de privacidade usados se deu por meio do acesso à página de privacidade dos sites. Os textos foram então copiados e tratados para eliminar espaços e parágrafos e, em seguida, foram colocados em duas planilhas distintas (do Microsoft Excel): aderentes e não aderentes.

A busca de textos aderentes foi realizada por meio do acesso à versão atual dos sites. A busca dos textos não aderentes foi realizada com a utilização da ferramenta *web Wayback Machine* que permite acesso a versões antigas de conteúdos de websites, foram coletados termos anteriores ao ano de 2018 (KAHLE et al., 1996).

Quadro 1: Resumo dos processos realizados em ordem de execução

ORDEM DE EXECUÇÃO	PROCESSO	DESCRIÇÃO
1	Importar dependências	Instalar os programas Python necessários para realizar o trabalho
2	Montar Dataframe	Carregar planilhas com os textos LGPD e não LGPD.
3	Classificação das amostras	Amostra aderente à LGPD recebe valor de 1. Amostra não aderente à LGPD recebe valor de 0.
4	Conversão de campos em	Código lê a coluna “Termo”

ORDEM DE EXECUÇÃO	PROCESSO	DESCRIÇÃO
	variáveis	e coloca os dados na variável “termos” E a coluna “classe” passa para a variável “y”
5	Pré processamento	Remoção do <i>caput</i> , que se trata da remoção de formatação dos cabeçalhos e títulos dos termos e tokenização.
6	Geração de <i>bag of words</i>	Geração de um multiconjunto de palavras desconsiderando a estrutura gramatical e até mesmo a ordenação, mas mantendo a frequência com a qual elas aparecem (Pedregosa, 2011).
7	Quebrando o conjunto de treino e teste	Gera valor numérico pseudo-aleatório para determinar a aleatoriedade da amostra. Configuração que determina que do total do conjunto 80% será para treino e 20% para teste.
8	Modelos	Importação e configuração dos modelos Naive Bayes e Regressão Logística
9	Classificação	Execução dos algoritmos com a saída dos resultados.

Fonte: Elaborado pelos autores

O quadro 1 apresenta os processos realizados no estudo. Na ferramenta *Google Colab* foi realizado o *upload* de ambos os arquivos para montar o *dataframe*, foi designado a cada um deles um valor binário denominado “Classe” (1 para os termos aderentes e 0 para os termos não aderentes), em seguida cada campo da planilha foi convertido em variável, ordenado de forma aleatória e inseridos no *dataframe*.

Em seguida foi realizado o pré-processamento dos dados para remoção do título do texto. Na etapa seguinte foi utilizada a biblioteca NLTK para fazer a tokenização do texto, que é o processo de dividir frases em palavras e na sequência a quebra em dois conjuntos, teste e treino, sendo 80% do conjunto para treinar o modelo e 20% para testar, os resultados apresentados neste trabalho usam como base o desempenho dos modelos na fase de teste. O trabalho utilizou como

referência o experimento e algoritmo de Pacheco (2019), o código desenvolvido para o estudo pode ser acessado por meio do link².

6. RESULTADOS

Quadro 2: Resultado estatístico da análise de Naive Bayes por classificação, utilizando 20% da amostra de 140 termos de privacidade.

Naive Bayes	Precisão	Amostra
Aderente	0.91	13
Não Aderente	0.82	15

Fonte: Elaborado pelos autores

No quadro 2 é possível observar que o algoritmo Naive Bayes apresentou para a classificação de termos aderentes à LGPD precisão de 0.91, ou seja, o algoritmo classificou corretamente 91% dos termos de privacidade que estavam classificados como aderentes à LGPD.

Para os termos não aderentes o modelo apresentou 0.82 de precisão, indicando que 82% dos termos classificados como não aderentes estavam realmente não aderentes à LGPD.

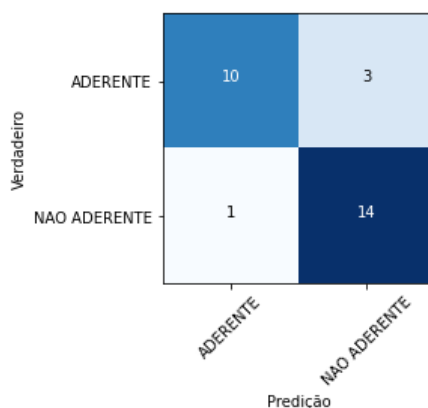
Quadro 3: Resultado estatístico de desempenho do modelo Naive Bayes, utilizando 20% da amostra de 140 termos de privacidade.

Modelo	Acurácia	Precisão
Naive Bayes	0.85	0.86

Fonte: Elaborado pelos autores.

O quadro 3 apresenta o desempenho geral do modelo, que atingiu 0.85 de acurácia indicando que 85% dos termos foram classificados corretamente. A precisão do modelo foi de 0.86, o que significa que 86% das classificações realizadas pelo modelo estão corretas.

Figura 1: Matriz de Confusão do Naive Bayes



Fonte: Gerado pelo algoritmo utilizado no experimento.

A figura 1 representa a distribuição dos registros em termos de suas classes verdadeiras e de suas classes preditas para o algoritmo Naive Bayes. Em que dos 13 registros aderentes 10 foram corretamente classificados e 1 foi classificado de forma errada. Para os 15 registros não aderentes 14 foram classificados

² <https://github.com/ViniciusOML/tetris>

corretamente e 3 foram classificados de forma errada.

Quadro 4: Resultado estatístico da análise de Regressão Logística por classificação, utilizando 20% da amostra de 140 termos de privacidade.

Regressão Logística	Precisão	Amostra
Aderente	0.92	13
Não Aderente	0.93	15

Fonte: Elaborado pelos autores.

No quadro 4 é possível observar que o algoritmo de Regressão Logística apresentou para a classificação de termos aderentes à LGPD precisão de 0.92, ou seja, há uma probabilidade de 92% dos termos classificados como aderentes estarem corretos.

Para os termos não aderentes o modelo apresentou precisão de 0.93, ou seja, há uma probabilidade de 93% dos termos classificados como não aderentes serem realmente não aderentes.

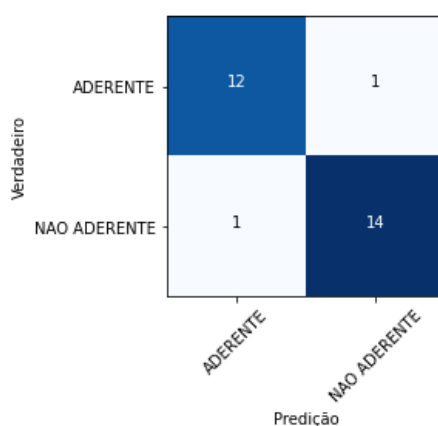
Quadro 5: Resultado estatístico de desempenho do modelo de Regressão Logística, utilizando 20% da amostra de 140 termos de privacidade.

Modelo	Acurácia	Precisão
Regressão Logística	0.92	0.92

Fonte: Elaborado pelos autores.

O quadro 5 apresenta o desempenho geral do modelo de Regressão Logística, que atingiu 0.92 de acurácia indicando que é possível classificar corretamente os termos com uma exatidão de 92%. A precisão do modelo foi de 0.92, o que significa que há uma probabilidade de 92% das classificações realizadas pelo modelo estarem corretas.

Figura 2: Matriz de confusão Regressão Logística



Fonte: Gerado pelo algoritmo utilizado no experimento.

A figura 2 representa a distribuição dos registros em termos de suas classes verdadeiras e de suas classes preditas para o algoritmo de Regressão Linear. Em que dos 13 registros aderentes 12 foram corretamente classificados e 1 foi classificado de forma errada. Para os 15 registros não aderentes 14 foram classificados corretamente e 1 foi classificado de forma errada.

7. ANÁLISE

De acordo com Juliano (2019) a métrica de precisão busca responder qual a proporção de identificações positivas foi realmente correta. Em relação a essa métrica, o modelo Naive Bayes teve um desempenho 91% para os termos aderentes, sobre a amostra usada na fase de teste do algoritmo. O modelo de Regressão Logística apresentou desempenho de 92%. Ambos os desempenhos foram inferiores aos achados de Juliano (2019). Em seu trabalho, em que utilizou ambos os algoritmos para diferenciar textos jurídicos, Pacheco (2019) obteve em suas classificações uma precisão de 98% e 99% com Naive Bayes e 99% e 100% com a Regressão Logística. Apesar dos desempenhos serem inferiores comparados ao estudo de Juliano (2019) os resultados podem ser entendidos pelo baixo volume amostral dos dados deste estudo, 140, em comparação ao trabalho de Pacheco (2019), 2400.

Os termos não aderentes apresentaram precisão de 82% com Naive Bayes e 93% com Regressão Logística. Para essa classificação, ambos os modelos também ficaram abaixo dos resultados do estudo de Pacheco (2019). Para este tipo de classificação é possível entender que este resultado se deu pois os termos não aderentes não possuíam nenhuma espécie de padrão, como os tópicos obrigatórios definidos na LGPD. Alguns dos termos, apesar de serem todos anteriores à lei, são bem específicos em relação aos tratamentos dos dados, outros abordam apenas os dados coletados via *cookies*, trechos de códigos utilizados para identificar os usuários, controlar o tráfego nos sites e identificar as sessões de navegação (HORMOZI, 2005) por ser uma amostra pouco homogênea a precisão foi baixa em relação aos termos aderentes, para reverter este cenário é preciso que haja uma amostra maior de termos não aderentes.

A precisão do desempenho geral do modelo Naive Bayes foi de 86% e para o modelo de Regressão Logística foi de 92%. No estudo de Pacheco (2019) essa métrica foi de 98% com Naive Bayes e 99% com Regressão Logística. Em relação à acurácia dos modelos utilizados no estudo, o resultado com Naive Bayes foi de 85% e 92% com Regressão Logística. Ficando abaixo do esperado em relação ao estudo de Juliano (2019) que foi de 98% com Naive Bayes e 99% com Regressão Logística, porém novamente este resultado pode ser explicado pelo baixo volume de dados e heterogeneidade da amostra dos termos não aderentes.

8. CONSIDERAÇÕES FINAIS

Com o presente estudo foi possível observar que é possível, com o uso de algoritmos de Machine Learning, verificar se Termos de Privacidade estão ou não aderentes à LGPD. Apesar disso, devido à falta de estudos publicados com o mesmo objetivo deste trabalho e por conta da complexidade na coleta de dados, principalmente nos termos considerados como aderentes à lei pois eles necessitam de leitura prévia, consideramos que os resultados obtidos não devem ser utilizados como uma forma única de avaliar se um Termo de privacidade está aderente à LGPD, mas pode ser usado como um método de triagem.

Sugere-se que novos estudos sejam feitos a partir deste, mantendo o objetivo central porém melhorando a etapa coleta de dados, seja via raspagem de dados ou chamando especialistas na área de proteção de dados para sugerir melhorias na metodologia.

REFERÊNCIAS

ALECRIM, Emerson. **O que você deve saber sobre a lei de proteção de dados pessoais do Brasil**. 2018. Disponível em: <https://tecnoblog.net/responde/lei-geral-protecao-dados-brasil/>. Acesso em: 13 fev. 2022.

BIRD, Steven; KLEIN, Ewan; LOPER, Edward. **UPDATED FOR NLTK 3.0**, 2019. Disponível em: <https://www.nltk.org/book/ch00.html>. Acesso em: 20 mar. 2022.

BRASIL. Lei nº 13709, de 14 de agosto de 2018. **Lei Geral de Proteção dos Dados Pessoais**. Diário Oficial da União, 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 18 out. 2021.

CHEN, Dehua *et al.* **Métricas de Avaliação em Machine Learning: Classificação**. 2020. Disponível em: <https://medium.com/kunumi/m%C3%A9tricas-de-avalia%C3%A7%C3%A3o-em-machine-learning-classifica%C3%A7%C3%A3o-49340dcdb198>. Acesso em: 24 jan. 2022.

HORMOZI, Amir. **Cookies and Privacy. PRO QUEST**, v. 13, ed. 6, p. 51-59, 2005. Disponível em: <https://www.proquest.com/openview/8b6509a9d2f04d9e456849d4349185f2/1?pq-origsite=gscholar&cbl=52433>. Acesso em: 11 abr. 2022.

KAHLE, Brewster *et al.* **Wayback Machine**. 1996. Disponível em: <https://archive.org/about/>. Acesso em: 7 abr. 2022.

MALLMANN, Júlia. **LGD vem aí: um checklist para saber se sua empresa está em conformidade**. 2020. Disponível em: <https://cio.com.br/tendencias/lgd-vem-ai-um-checklist-para-saber-se-sua-empresa-esta-em-conformidade/>. Acesso em: 13 dez. 2021.

MEDEIROS, Bruno A; SILVA, Eric C. Marcelino. **Comparação de métodos de mineração de texto para classificação de documentos jurídicos**. 2020. Disponível em: <https://repositorio.animaeducacao.com.br/bitstream/ANIMA/8438/1/Artigo%20-%20Bruno%20Angelo%20Medeiros%20-%20Eric%20Marcelino.pdf>. Acesso em: 25 jan. 2022.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. **Conceitos sobre Aprendizado de Máquina**. In: MONARD, Maria Carolina; BARANAUSKAS, José Augusto. *Sistemas Inteligentes Fundamentos e Aplicações: Conceitos sobre Aprendizado de Máquina*. 1. ed. Barueri: Manole LTDA, 2003. v. 1, cap. 4, p. 89-114. Disponível em: <https://dcm.ffclrp.usp.br/~augusto/publications/2003-sistemas-inteligentes-cap4.pdf>. Acesso em: 12 jan. 2022.

MOŽINA, Martin. **Argument Based Machine Learning Applied to Law**. 2006. Disponível em: <https://link.springer.com/article/10.1007/s10506-006-9002-4>. Acesso em: 14 dez. 2021.

NONES, Fernanda. **Política de Privacidade: o que é, sua relação com**

estratégias de Marketing e Vendas e como fazer de acordo com a LGPD. 2021. Disponível em: <https://resultadosdigitais.com.br/marketing/politicas-de-privacidade/>. Acesso em: 14 maio 2022.

PACHECO, Juliano. **THE DEVELOPERS CONFERENCE**, 2019, Porto Alegre. **ML - Aplicações de Machine Learning na área Jurídica, um exemplo com classificação de textos.** 2019. Disponível em: https://s3-sa-east-1.amazonaws.com/thedevconf/presentations/TDC2019POA/machine/ZXJ-8873_2019-12-09T105647_20191128%20-%20TDC%20-%20Aplicacoes%20de%20Machine%20Learning%20na%20area%20Juridica,%20um%20exemplo%20com%20classificacao%20de%20textos.pdf. Acesso em: 8 nov. 2021.

PARDO, Thiago Alexandre Salgueiro. **Métodos para Análise Discursiva Automática. Orientador: Profa. Dra. Maria das Graças Volpe Nunes.** 2005. Tese (Doutorado) - Instituto de Ciências Matemáticas e de Computação da USP, 2005. Disponível em: <http://www.nilc.icmc.usp.br/dizer2/PDF/THESIS-Pardo.pdf>. Acesso em: 1 dez. 2021.

PEDREGOSA, Fabian; VAROQUAUX, Gaël; MICHEL, Vincent; THIRION, Bertrand. **Scikit-learn: Machine Learning in Python - Countvectorizer.** Journal of Machine Learning Research, 2011. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html. Acesso em: 23 maio 2022.

PEDREGOSA, Fabian; VAROQUAUX, Gaël; MICHEL, Vincent; THIRION, Bertrand. **Scikit-learn: Machine Learning in Python - Naive Bayes.** Journal of Machine Learning Research, 2011. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html. Acesso em: 20 mar. 2022.

PEREIRA, Fábio; SILVA, Cecília. **Ano de vigência da LGPD: o que esperar para 2022?** 2022. Disponível em: <https://techcompliance.org/ano-de-vigencia-da-lgpd/>. Acesso em: 22 de maio de 2022.

RISH, Irina. **An empirical study of the naive Bayes classifier.** JCA 2001 workshop on empirical methods in artificial intelligence, v. 3, ed. 22, p. 41-46, 2001. Disponível em: <https://www.cc.gatech.edu/home/isbell/classes/reading/papers/Rish.pdf>. Acesso em: 22 maio 2022.

SANTOS, Vinicius. **O que é regressão logística e como aplicá-la usando Python.** 2019. Disponível em: <https://www.computersciencemaster.com.br/o-que-e-regressao-logistica-e-como-aplicar-la-usando-python/>. Acesso em: 1 maio 2022.

TEODORESCU, Mike Horia. **Machine Learning Methods for Strategy Research.** 2017. Disponível em: https://www.hbs.edu/ris/Publication%20Files/18-011_579b0249-df04-4755-b0f6-3fe17c0b26cd.pdf. Acesso em: 10 jan. 2022.

TERRA, Augusto Henriques. **Catálogo de critérios para avaliação de políticas de privacidade.** Orientador: Jéssyka Flavianne Ferreira Vilela. 2021. TCC (GRADUAÇÃO) - CENTRO DE INFORMÁTICA GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO, 2021. Disponível em: https://www.cin.ufpe.br/~jffv/docs/TG_Augusto.pdf. Acesso em: 7 abr. 2022.

ANEXOS

ANEXO A - TERMO ADERENTE À LGPD

Políticas de Privacidade

Prezado Cliente, Quando você compra no KaBuM! você nos confia seus dados pessoais. É pensando na manutenção desta confiança que criamos nossa Política de Privacidade e Proteção de Dados Pessoais: nela é possível entender as nossas práticas sobre o tratamento dos seus dados. O KaBuM! entende que é fundamental explicar de modo transparente nossos processos de tratamento de dados, bem como esclarecer seus direitos como titular destes dados. Esta política se aplica à todas as pessoas que interagem com nossos serviços de Comércio Eletrônico na plataforma web e no aplicativo móvel. Lembrando que alguns dados são necessários para a prestação de nossos serviços. Desta forma, caso você não os queira fornecer, poderá optar por não utilizar nossa plataforma/aplicativo. Alguns outros dados não terão o caráter obrigatório, mas a ausência de seu consentimento poderá comprometer a prestação dos nossos serviços a você, como é o caso de recursos terem sua formatação e funcionalidades limitadas em nossa plataforma/aplicativo. **O Kabum Comércio Eletrônico S.A.** é o controlador dos dados pessoais tratados em nossa plataforma/aplicativo.

Como coletamos seus dados

Nós coletamos seus dados a partir das seguintes fontes:

- Plataforma E-Commerce KaBuM!: através de seu cadastro na plataforma, coletamos informações como: nome, CPF, RG, data de nascimento, telefone, e-mail, endereço, CEP. A totalidade dos dados coletados e seu uso estão descritos na seção **Dados que coletamos**;
- Aplicativo E-Commerce KaBuM!: oferecido nas lojas de terceiros como o Google Play e Apple Store, nele também coletamos as informações referentes ao cadastro conforme mencionado no item acima;
- Canais de Atendimento: poderemos coletar dados por meio dos canais de atendimento KaBuM!
- Blog KaBuM!
- Através de plataformas de terceiros, como, Facebook, Google. Estas plataformas possuem suas próprias políticas de privacidade e os dados que coletamos são aqueles cadastrados e autorizados naquela plataforma, posteriormente compartilhados com o KaBuM!. Dentro de nossa empresa esses dados são cuidados da mesma forma como os dados coletados diretamente por nós. Em relação ao tratamento de dados feitos pelas plataformas de terceiros, recomendamos que sejam lidas as políticas de privacidade das respectivas plataformas.

Dados que coletamos

Durante sua interação com a nossa plataforma e aplicativo, coletamos dados pessoais para poder prestar serviços a você e melhorar sua interação com o nosso e-commerce. Para que possamos vender nossos produtos, alguns dados pessoais são utilizados para cumprimento legal ou regulatória, promoção do nosso negócio através do legítimo interesse, prevenção contra ações ilícitas e fraudes e campanhas de Marketing, estas com base em seu consentimento. Abaixo é possível ver os dados coletados e o propósito do tratamento:

Dados		Finalidade
Cadastrais	Nome; e-mail; sexo; número de CPF; número de RG; data de nascimento; número de telefone e endereço	São dados coletados para realização de cadastro ou para alteração de cadastro na plataforma e aplicativo; Estes são utilizados para iniciar a prestação dos nossos serviços e viabilizar a compra e venda de produtos; Campanhas de Marketing; Procedimentos relacionados à emissão de nota fiscal; Atendimento de requisitos regulatórios; Armazenamento de registros de acesso à plataforma em conformidade com o Marco Civil da Internet; Prevenção à fraude na plataforma;
Financeiros	Dados bancários para processamento de pagamentos, informações do cartão de crédito (nome do titular, número do cartão, data de validade, código verificador, etc.)	Procedimentos relacionados ao pagamento de produtos vendidos na plataforma e aplicativo do KaBuM!.
Dados Computacionais e de navegação	Podemos coletar endereço IP utilizado na conexão, tipo de conexão à internet, localização geográfica, data e hora de acesso, tipo de dispositivo, identificação de navegador, tempo de uso no e-commerce, informações estatísticas sobre as suas interações, cliques na plataforma, produtos e páginas visualizadas, inclusive através de tecnologias como cookies, pixels, tags e tecnologias similares, rastreadores de terceiros. Podemos ainda coletar modelos de hardware, sistemas operacionais e versões, software, idiomas preferidos, identificadores exclusivos de aparelhos, identificadores de publicidade, localização aproximada do dispositivo por meio do endereço IP.	Melhoria da plataforma, como, performance e conteúdo de preferência dos clientes; Armazenamento de registro de acesso; atendimento ao Marco Civil da Internet para possível investigação; prevenção à fraude e outros atos ilícitos à plataforma e-commerce e aplicativo;
Outros dados pessoais	Podemos coletar informações de plataformas de terceiros, como Facebook e Google, no momento que você acessa o e-commerce KaBuM! por meio destas plataformas. Atualmente as informações coletadas são: nome, login (e-mail), idioma preferido e foto do perfil, desde que você permita isso nestas plataformas. Também coletamos dados de fontes publicamente disponíveis.	Identificação dos nossos clientes; oferecer serviços aprimorados e facilitados aos clientes;

Atendimento	Coletamos informações que os clientes enviam quando entram em contato com o atendimento, quando fazem reclamações, avaliações ou elogios, ou entram em contato com o KaBuM! de outra forma, podendo inclusive gravar e ouvir as ligações, Isso pode incluir comentários, fotos ou outros registros, inclusive de terceiros coletados pelos próprios clientes.	Atendimento à requisito regulatório e para proteção do KaBuM! em alguma disputa judicial e análise interna caso haja necessidade, incluindo compartilhamento com seguradoras e empresas de transporte
-------------	---	---

De uma forma geral, utilizaremos seus dados para melhorar a segurança dos nossos clientes e serviços, principalmente para: prevenção à fraude; prestar suporte ao clientes; para enviar comunicações (sobre mudanças em nossos termos, políticas ou serviços, por exemplo), bem como para comunicações de marketing (para divulgar nossos serviços, promoções, sorteios, notícias, eventos e também para divulgar produtos e serviços de nossos parceiros, por exemplo); para utilizá-los em procedimentos jurídicos; para fornecer serviços e recursos; oferecer, processar ou facilitar pagamentos de nossos produtos, bem como oferecer garantia estendida e serviços de instalação.

Decisões automatizadas

Utilizamos recursos tecnológicos que realizam decisões automatizadas para aumentar a qualidade e proteção dos nossos serviços, como, prevenção do ambiente tecnológico contra ações danosas, gerenciamento de riscos de crédito, combate à fraude ao KaBuM! e aos nossos clientes.

Dados de Crianças e Adolescentes

A plataforma E-Commerce KaBuM! e Aplicativo são destinados à qualquer pessoa que gosta de tecnologia, contudo o cadastro e a compra são condicionados às exigências legais. Caso seja menor de 18 anos, seu cadastro e compra deverão ser aprovados pelo representante legal. Ao se cadastrar, o cliente da plataforma deve informar sua idade real para que cadastros sejam criados com informações verdadeiras. Se tivermos conhecimento de cadastro realizado por criança ou adolescentes de forma indevida, tomaremos as medidas necessárias para eliminação dos dados pessoais, incluindo, mas não se limitando à retirada da conta da plataforma/aplicativo.

Compartilhamento de Dados Pessoais

Para que possamos prestar nossos serviços com qualidade e atender a legislação vigente, precisamos compartilhar alguns dados pessoais. Compartilhamos seus dados quando aplicável com:

- Instituições financeiras, administradoras de cartão de crédito e de pagamentos;
- Empresas de prevenção à fraude;
- Lojistas que utilizam a nossa plataforma de Marketplace;
- Empresas de transportes;
- Instituições Governamentais/autoridades policiais e judiciárias, para atendimento de legislação vigente; atendimento às demandas judiciais/administrativas e controle fiscal;
- Parceiros de marketing ou comerciais;
- Empresas de serviço de garantia estendida;
- Empresas de serviço de instalação e/ou assistência técnica;
- Com as empresas do mesmo grupo econômico.
- Provedores de armazenamento;
- Consultores, advogados, contadores e outros prestadores de serviços;
- Mediante consentimento, de maneiras diferentes das listadas nesta política.

Medidas de Segurança

O KaBuM! investe continuamente e dispõe de medidas técnicas e administrativas de segurança para proteger seus dados, alinhadas as melhores práticas de segurança da informação. O KaBuM! preza pela Confidencialidade, Integridade e Disponibilidade das informações e dos dados pessoais dos clientes, colaboradores e parceiros de negócios. Há a busca contínua de métodos de proteção contra ações danosas que visam utilizar, acessar, alterar e/ou destruir dados indevidamente, dentre elas, citamos:

- Controle de acesso físico aos ambientes corporativos;
- Acesso aos dados somente por profissionais que exercem funções necessárias para atendimento das finalidades específicas;
- Monitoramento contínuo dos sistemas de informações, de modo a prevenir, detectar e bloquear ações danosas;
- Armazenamento e transferência de dados utilizando métodos seguros.

Sobre a retenção e eliminação dos dados pessoais

De acordo com a legislação vigente, O KaBuM! faz uso dados por tempo necessário para atender as finalidades para as quais foram coletados, para cumprir com os requerimentos legais aplicáveis, para atendimento do legítimo interesse do KaBuM!, bem como para atendimento dos interesses dos titulares dos dados. A pedido do titular de dados, eliminaremos todos aqueles dados pessoais que forem permitidos pela legislação vigente, como, sua subscrição às campanhas e notificações de Marketing, eliminação de contas e demais dados pessoais. Apenas manteremos dados pessoais para cumprimento legal ou regulatórios, cumprimento de exigências fiscais, seguros, auditorias, contábeis e/ou para proteção do KaBuM!, como, prevenção a fraudes e subsídio para eventuais demandas judiciais/administrativas.

Seus Direitos

O Kabum visa atender todos os direitos de nossos clientes, conforme definido na Lei, no que tange tratamento de dados pessoais:

- Confirmar a existência de seus dados;
- Acessar os dados;
- Obter informação adicional sobre o uso e compartilhamento de dados;
- Corrigir os dados;
- Solicitar a portabilidade dos dados;
- Revogar o consentimento de algum tratamento de dados;
- Solicitar a eliminação de dados pessoais, quando possível, conforme previsto nessa política.

Política de Cookies

O que são Cookies?

Os cookies são pequenos arquivos enviados pelos sites durante a navegação. Eles ficam armazenados nos dispositivos de computadores e dispositivos móveis, que permitem melhorar o desempenho e a experiência de navegação dos nossos clientes, deixando as plataformas do KaBuM! personalizadas de acordo com as suas preferências.

A inserção de cookies auxilia no reconhecimento do dispositivo para a próxima visita, guardando apenas informações genéricas como a forma e/ou local de acesso e o modo como o cliente utiliza nossas plataformas (e-commerce, hotspots etc).

Nós utilizamos Cookies primários, os quais são adicionados diretamente nos dispositivos pelas nossas plataformas, bem como Cookies de terceiros, que por sua vez, são adicionados por empresas terceiras de nossa confiança através das nossas plataformas.

Os cookies possuem tempo de vida, sendo os temporários expiram logo após o fechamento do navegador e os persistentes permanecem no seu dispositivo por tempo determinado ou até que seja feita a exclusão manual.

Tipos de Cookies que utilizamos

Cookies necessários:

Como o próprio nome diz, são os necessários para o funcionamento das nossas plataformas. Portanto, os clientes que desabilitarem os cookies em seus navegadores poderão não ter a melhor experiência durante a navegação, incluindo o comprometimento de funcionalidades essenciais.

Cookies de desempenho/performance:

Este tipo de cookie coleta informações anônimas sobre as páginas acessadas, possibilitando validar, por exemplo, se houve erros, quanto tempo o usuário permaneceu na página, como o usuário interage entre as páginas e como chegou até a página.

Cookies de funcionalidade:

Eles memorizam as preferências e escolhas dos clientes, evitando a necessidade de configuração e/ou preenchimento das informações novamente, toda vez que há o acesso às nossas plataformas.

Cookies de publicidade/propaganda:

Eles direcionam anúncios com base nos interesses do cliente, auxiliando na busca, por exemplo, de produtos. Estes anúncios podem ser exibidos em forma de banner. Quando este tipo de cookie não estiver habilitado, as publicidades continuarão a ser exibidas, porém não propriamente direcionadas aos interesses do usuário.

Como você pode gerenciar os cookies em seu navegador?

É possível bloquear tipos específicos de cookies por meio das configurações do navegador utilizado. A grande maioria dos navegadores aceita cookies automaticamente, portanto cabe ao usuário personalizar o seu próprio navegador. Para mais detalhes sobre como apagar ou rejeitar cookies, há tutoriais que tratam do assunto:

A inserção de cookies auxilia no reconhecimento do dispositivo para a próxima visita, guardando apenas informações genéricas como a forma e/ou local de acesso e o modo como o cliente utiliza nossas plataformas (e-commerce, hotspots etc).

- Google Chrome
- Mozilla Firefox
- Internet Explorer
- Microsoft Edge
- Safari
- Opera

Atualização desta Política de Privacidade

Quando mudarmos a forma de tratamento de dados pessoais esta Política será atualizada, por esse motivo recomendamos que acesse esta política com frequência e em caso de dúvidas, procure-nos em nossa [central de Atendimento](#).

Dúvidas, comentários e reclamações sobre as práticas de tratamento de dados do KaBuM! podem ser enviados para privacidade@kabum.com.br.

Exerça seus Direitos de Privacidade [clicando aqui](#).

Última alteração: fevereiro de 2022.

Fonte: O KABUM COMÉRCIO ELETRÔNICO S.A. **Políticas de Privacidade**. 2022. Disponível em: <https://www.kabum.com.br/privacidade>. Acesso em: 12 abr. 2022.

ANEXO B - TERMO NÃO ADERENTE À LGPD

Política de Privacidade

1) Obtenção e Armazenamento dos dados

1.1. O KaBuM! zela pela segurança e privacidade de seus clientes, portanto todas as informações fornecidas durante a navegação e compra no site são armazenadas de forma automática no banco de dados, não sendo vendidas, trocadas ou divulgadas a terceiros - exceto quando estas informações forem necessárias para processo de entrega, cobrança, participação de promoções solicitadas pelos clientes, confirmação de dados objetivando a segurança na finalização da compra, ordem judicial e administrativa.

1.1.1. Para a preservação de seus dados, o KaBuM! não aconselha a divulgação de senha a terceiros, mesmo que familiares ou amigos.

1.2. O KaBuM! também poderá utilizar outras tecnologias padrões as quais coletam informações para melhorar a experiência de navegação do usuário, baseada nos hábitos e preferências deste, como os Cookies, que são os dados enviados para o navegador do usuário após acesso ao site KaBuM!. Recordando que, para bloqueios dos Cookies, caberá ao próprio visitante/usuário configurar o seu navegador.

1.2.1. O usuário consente que o KaBuM! utilize as informações para realização de marketing direto.

2) Disposições Finais

2.1. O conteúdo constante nesta Política de Privacidade poderá ser atualizado ou modificado a qualquer momento, cabendo aos usuários verificá-lo sempre que efetuar o acesso ao site.

2.2. O usuário, ao colocar informações em uma área pública do KaBuM!, está ciente de que elas estarão disponíveis a outros visitantes e usuários do site.

2.3. Caso persistir dúvida em relação ao conteúdo dessa Política, o usuário deverá entrar em contato com a [Central de Atendimento](#).

Atualizado em 25 de julho de 2016.

Fonte: KAHLE, Brewster et al. **Wayback Machine - KaBuM!**. 1996. Disponível em: https://web.archive.org/web/20170218114616/http://www.kabum.com.br/cgi-local/site/institucional/politica_privacidade.cgi . Acesso em: 7 maio 2022.