Uma Abordagem de Aprendizado de Máquina para Avaliar Políticas de Privacidade em um Contexto de Desenvolvimento de Sistemas de Informação com Privacidade por design

Gabriel Cortizo Ferraz, Jéssyka Vilela, Carla Silva

[1]Centro de Informática, Universidade Federal de Pernambuco, Recife, Brasil

gabrielferrazcortizo@gmail.com,{jffv,ctlls}@cin.ufpe.br

Resumo. *A adoção da "Privacidade por Design"tornou-se essencial no desenvolvimento de sistemas de informação devido à LGPD, exigindo a incorporação da privacidade desde as fases iniciais do desenvolvimento. No entanto, as políticas de privacidade costumam ser extensas e complexas, dificultando sua compreensão e aplicação prática, o que pode comprometer a conformidade legal. Para solucionar esse problema, este estudo propõe uma abordagem baseada em aprendizado de máquina para avaliar e melhorar a transparência dessas políticas. A metodologia incluiu a análise de 23 políticas e a aplicação de diferentes algoritmos para avaliar sua clareza e adequação legal. O melhor algoritmo apresentou 81% de eficácia, enquanto a ferramenta desenvolvida atingiu 72,7% na avaliação das políticas em conformidade com "Privacidade por Design". Os resultados mostram que a ferramenta pode auxiliar desenvolvedores na implementação de requisitos de privacidade desde o início dos projetos. Assim, este estudo reforça o uso de aprendizado de máquina para integrar privacidade ao ciclo de vida do software.*

1. Introdução

No Brasil, a Lei Geral de Proteção de Dados Pessoais (LGPD) [3 2018] regula a coleta, uso e retenção de dados pessoais, garantindo ao titular o direito de acessar informações sobre o tratamento de seus dados. Essas informações são geralmente apresentadas em políticas de privacidade, que descrevem como os dados são coletados, armazenados e utilizados. No entanto, essas políticas costumam ser longas e complexas [Singh et al. 2011][Zimmerman et al. 2015], exigindo um nível de leitura universitário [Zimmerman et al. 2015], o que dificulta a compreensão dos usuários. Essa limitação evidencia a necessidade de maior transparência e clareza [Solove 2015] desde a fase de requisitos no desenvolvimento de software [Santana et al. 2022].

Atualmente, a privacidade é muitas vezes tratada de forma reativa, nas fases finais do desenvolvimento ou após o lançamento do software [Peixoto et al. 2023], o que pode comprometer a conformidade legal e a confiança dos usuários. Para abordar esse problema, este trabalho propõe uma ferramenta baseada em aprendizado de máquina e processamento de linguagem natural para avaliar a qualidade das políticas de privacidade [Terra et al. 2022]. Essa abordagem busca garantir que a privacidade seja integrada desde o início do desenvolvimento, promovendo maior conformidade com a legislação e transparência no uso dos dados pessoais.

Este trabalho se divide em 5 seções. Na seção 2 é apresentado o referencial teórico e são discutidos alguns trabalhos relacionados. A seção 3 descreve a metodologia utili-

zada na construção da ferramenta. A seção 4 apresenta a ferramenta desenvolvida. A seção 5 contempla a conclusão e discute trabalhos futuros.

2. Revisão da literatura e Trabalhos Relacionados

Políticas de privacidade descrevem como organizações coletam, utilizam, armazenam e protegem dados pessoais [Herold 2015], garantindo transparência e permitindo que usuários compreendam o tratamento de suas informações. Regulamentações como a GDPR e a LGPD [3 2018] exigem que essas informações sejam apresentadas de forma clara e acessível.

PLN combina linguística, ciência da computação e matemática para permitir que máquinas analisem e gerem linguagem natural [Manning and Schütze 1999]. Técnicas como análise sintática e extração de informação frequentemente utilizam aprendizagem de máquina para aprimorar tarefas como classificação de textos e tradução automática [Devlin et al. 2018].

Aprendizagem de máquina permite que sistemas automatizados aprendam a partir de dados, sendo aplicada em áreas como reconhecimento de fala e PLN. Modelos de classificação de texto são usados com algoritmos como Support Vector Machine, Árvores de Decisão, Random Forest, Regressão Logística e Naive Bayes.

A dificuldade de interpretação das políticas de privacidade [Singh et al. 2011] motivou pesquisas sobre sua avaliação automatizada. Mineração de texto foi utilizada por [Li et al. 2021] para extrair informações e verificar conformidade com GDPR e CCPA, atingindo 94% de precisão. Paula [de Paula 2022] propôs uma ferramenta para avaliar políticas de privacidade brasileiras com base nos critérios de Terra, Vilela e Peixoto [Terra et al. 2022], utilizando PLN para identificar 14 critérios em 6 políticas. Este estudo compara a adoção de aprendizagem de máquina com o método de Paula [de Paula 2022], analisando sua eficácia na identificação de critérios de privacidade.

3. Metodologia

A metodologia deste trabalho foi dividida em seis etapas: (1) Seleção de critérios de avaliação de políticas de privacidade; (2) Criação da base de dados utilizada para o treinamento de modelos; (3) Definição do modelo de aprendizagem de máquina para classificar políticas de privacidade; (4) Definição de tecnologias utilizadas na elaboração da ferramenta; (5) Implementação da ferramenta; (6) Testes na ferramenta.

Foram adotados os 14 critérios utilizados no trabalho de Paula [de Paula 2022], derivados do estudo de Terra, Vilela e Peixoto [Terra et al. 2022]. A Tabela 1 apresenta os critérios, suas descrições e a distribuição das amostras na base de treinamento.

Criação da base de dados para treinamento de modelos. Foram coletados 375 títulos de políticas de privacidade de 23 empresas, abrangendo diversos setores, como redes sociais, e-commerce e mídia. Os dados foram organizados em um arquivo CSV contendo três colunas: *company* (empresa associada ao título), *text* (texto do título) e *class* (critério de avaliação correspondente). A base incluiu 15 categorias, sendo 14 representando critérios específicos e uma indicando a ausência de critérios.

Definição do modelo de aprendizagem de máquina. Cinco modelos supervisionados foram avaliados: Naive Bayes, Random Forest, Árvore de Decisão, Re-

Tabela 1. Lista de Critérios de Avaliação de Política de Privacidade utilizados na base de dados para o treinamento dos modelos.

Critério	Descrição	# de amostras
A política específica claramente quais dados são coletados?	É importante que a Política de Privacidade detalhe claramente quais dados serão coletados pela aplicação. Os dados coletados se dividem em categorias bem definidas e a Política deve indicar essas áreas.	28
A Política de Privacidade específica claramente como a empresa pode usar os dados coletados?	A Política deve indicar qual o propósito da coleta de informações dos usuários. É necessário afirmar, por exemplo, se os dados estão sendo coletados para contactar o usuário, melhorar os serviços fornecidos, análise e monitoramento durante o uso da aplicação, personalizar a experiência, publicidade direcionada, entre outras ações.	29
A política trata questões relacionadas à privacidade de crianças?	É necessário que a Política explique claramente como ocorre questões relacionadas à privacidade com crianças que acessam a aplicação.	11
A Política específica claramente como os dados são coletados?	A Política precisa expressar com clareza quais ferramentas a aplicação utiliza para coletar dados.	7
A Política de Privacidade claramente especifica se as informações podem ser compartilhadas ou vendidas para terceiros?	Caso envolva terceiros, é necessário descrever que tipo de informações são compartilhadas, quem são os terceiros e como os terceiros podem ser classificados, além de estar anexada a Política de Privacidade dessa empresa terceira. É necessário afirmar também caso não haja o compartilhamento com outras organizações.	25
Decisões Automatizadas	Aqui o critério verifica se a política discute se existem recursos tecnológicos que realizam decisões automatizadas para fim de melhorar o serviço prestado pela empresa	1
A Política de Privacidade claramente especifica quais são as medidas adotadas pela aplicação para garantir a confidencialidade, a integridade e a qualidade dos dados?	Este critério busca avaliar se a aplicação possui algum método para garantir a confidencialidade e integridade dos dados do usuário. Por exemplo, se o armazenamento dos dados é criptografado ou alguma máscara de IP é utilizada.	22
A política explica claramente o que acontece com os dados do usuário caso ele exclua a conta?	É importante que esteja descrito na política o que acontece caso o usuário se desvincule da aplicação.	0
A Política de Privacidade claramente especifica os direitos do usuário?	As leis de privacidade apresentam direitos que os usuários possuem. É uma boa prática que a política descreva esses direitos em relação a seus dados pessoais.	20
A Política de Privacidade fala sobre como ela utiliza cookie no seu site?	Este critério busca avaliar se o site fala sobre os tipos de cookies utilizando pelo website.	19
A Política de Privacidade claramente informa dados para contato com a empresa?	Idealmente deve haver o contato da área da empresa que trate de questões de privacidade dos dados de seus usuários.	16
A Política de Privacidade claramente específica como os dados são armazenados?	Ao informar como os dados são armazenados a empresa passa uma maior credibilidade para seus usuários.	15
A Política de Privacidade fala sobre transferir dados do usuário em nível internacional?	O critério idealmente deve falar sobre como transferir os dados do cliente para outras regiões fora do Brasil.	8
Como as alterações nas políticas são tratadas?	Após uma eventual alteração na Política de Privacidade, os usuários precisam ser informados e notificados sobre isso.	17

gressão Logística e Support Vector Machine (SVM). A escolha dos modelos considerou a limitação da quantidade de amostras e sua aplicação em trabalhos anteriores [Manning and Schütze 1999][Géron 2019][Sinha et al. 2018]. Antes do treinamento, os textos foram pré-processados, incluindo conversão para minúsculas, lematização, remoção de stopwords e caracteres especiais. Para representação numérica, utilizou-se a técnica Term Frequency-Inverse Document Frequency (TF-IDF).

A avaliação dos modelos foi realizada por validação cruzada (k=5) [Santana et al. 2022][Peixoto et al. 2023], resultando em acurácias de: Decision Tree (64%), Naive Bayes (69%), Random Forest (76%), Regressão Logística (79%) e SVM (81%). Dado o melhor desempenho, o SVM foi adotado na ferramenta.

Definição de tecnologias utilizadas na elaboração da ferramenta. A ferramenta foi desenvolvida em Python, utilizando *scikit-learn* para aprendizado de máquina, *nltk* para pré-processamento e Flask para a interface web. O Bootstrap foi escolhido para o front-end, facilitando a prototipação.

Implementação da ferramenta. Com as tecnologias definidas, foi elaborado o fluxo da aplicação, seguido da implementação utilizando os componentes mencionados.

Testes na ferramenta. Três métricas foram analisadas: detecção de cabeçalhos, classificação de títulos e eficácia total. O SVM alcançou 81% de acurácia na classificação. A ferramenta foi testada em quatro políticas não incluídas no treinamento (Gov.br, Epic Games, Estadão e Casas Bahia). Detectou 37 critérios, mas cinco foram classificados erroneamente e sete não foram identificados. Com um total de 44 critérios, a ferramenta obteve uma eficácia final de 72,7%.

4. Resultados

Nesta seção, são apresentados os resultados obtidos na construção da ferramenta, os passos para a utilização da mesma e comparação com ferramentas e trabalhos já existentes.

4.1. Visão geral da ferramenta

Para avaliar uma política de privacidade, o usuário faz o upload de um arquivo em formato PDF, escolhido devido à falta de padronização na estrutura das políticas nos sites. Diferentes frameworks e tags HTML dificultam a extração de informações diretamente das páginas web. Após o upload, o texto é processado e os títulos são extraídos com base no tamanho da fonte. Em seguida, os títulos passam por pré-processamento e são classificados pelo modelo de aprendizagem de máquina nos critérios estabelecidos. Os resultados da predição (Figura 1a) e os títulos detectados (Figura 1b) são formatados e apresentados ao usuário.

A ferramenta também possui uma tela de informação sobre os critérios de ava-liação de políticas de privacidade (Figura 2a), explicando o conceito de critérios de avaliação de políticas de privacidade e enumerando os critérios utilizados na implementação da ferramenta. A última tela da ferramenta (Figura 2(b)) possui informações sobre a motivação para a criação da ferramenta, e algumas das tecnologias utilizadas no processo.

O código fonte da ferramenta, o modelo de aprendizagem e a base de dados utili-

Figura 1. (a) Tela de resultados da ferramenta para a política de privacidade selecionada e (b)Tela dos títulos da política de privacidade detectados.

Tabela 2. Comparação entre o trabalho proposto e trabalhos relacionados.

Critério	Ferreira [de Paula 2022]	Este trabalho
Ano	2022	2023
Paradigmas utilizados	Processamento de linguagem natural.	Processamento de linguagem natural e modelo de aprendizado supervisionado Support Vector Machine.
Número de políticas utilizadas na construção da ferramenta	7	23
Número de critérios avaliados	14	14
Eficácia em um conjunto de políticas não utilizados na construção da ferramenta (%)	48%	72.7%

zada no treinamento dos modelos estão disponibilizados no repositório [1].

4.2. Comparação com ferramentas existentes

A Tabela 2 faz o comparativo entre a ferramenta desenvolvida nesse trabalho e a ferramenta desenvolvida no trabalho de Paula [de Paula 2022], tendo em vista que ambas as ferramentas propõem a detecção de critérios de avaliação de políticas de privacidade de maneira automatizada.

Tendo em vista que o desempenho de um algoritmo de aprendizado de máquina está diretamente relacionado ao número de exemplos e a qualidade de dados utilizados para treinar o modelo [Khalid and Mehmood 2019], um número maior de políticas foi utilizado neste trabalho. Quanto à avaliação da eficácia, tanto o trabalho de Paula [de Paula 2022] como este trabalho utilizaram o cálculo da proporção de previsões corretas em relação ao número total de previsões, divergindo quanto ao número de políticas utilizadas no teste que foram respectivamente de duas e quatro políticas.

4.3. Limitações

O suporte da ferramenta limitado à políticas de privacidade em formato PDF afeta diretamente a experiência do usuário, tendo em vista que as políticas de priva-cidade salvo exceções são disponibilizadas em páginas web, sendo necessário que o usuário faça o

[1] https://github.com/GabrielCortizo/avaliador-de-politicas-de-privacidade

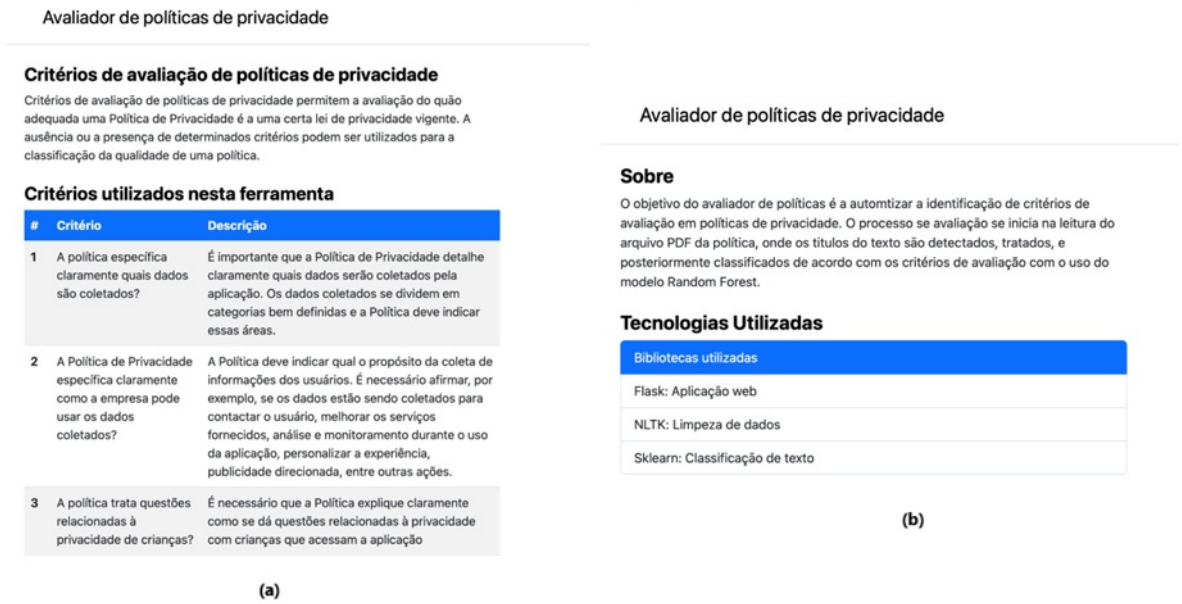

Figura 2. (a) Tela de Critérios de avaliação de políticas de privacidade; e, (b) Tela sobre a ferramenta.

download do conteúdo da página web para cada política e que o converta em formato PDF para utilizar a ferramenta.

A ferramenta também assume que o texto da política de privacidade em avaliação seja dividido em títulos, e que os títulos possuam uma formatação de texto diferente dos parágrafos, o que inviabilizaria a identificação de critérios caso documento de política não segue essas duas limitações.

5. Conclusões e Trabalhos Futuros

A avaliação automatizada de políticas de privacidade é essencial devido à complexidade e extensão desses documentos [Singh et al. 2011]. Este trabalho implementou uma ferramenta baseada em aprendizado de máquina para avaliar a qualidade das políticas utilizando critérios previamente definidos na literatura. A ferramenta obteve 72,7% de eficácia na identificação de critérios em políticas não utilizadas no treinamento, superando a ferramenta de Paula [de Paula 2022]. Disponibilizada online junto com seu código-fonte e base de dados, a ferramenta enfrenta desafios como a falta de padronização nos documentos, presença de imagens e tabelas, e a escassez de amostras para alguns critérios, o que impacta a precisão da classificação.

Algumas sugestões de trabalhos futuros envolvem: (i) Aumento do número de critérios de avaliação de políticas de privacidade utilizados na ferramenta a fim de melhorar a eficácia da ferramenta; (ii) Teste de usabilidade da ferramenta; (iii) Testes de eficácia da ferramenta; (iv) A possibilidade de leitura de textos de políticas a partir de links URL e a habilitação da ferramenta na utilização não apenas dos títulos das políticas, mas também do texto completo; (v) Evoluir a ferramenta para verificação de conformidade da política de privacidade com a LGPD, tal como realizado em [Cejas et al. 2024] para a GDPR.

Referências

(2018). Lei geral de proteção de dados pessoais. http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm.

Cejas, O. A., Abualhaija, S., and Briand, L. C. (2024). Compai: A tool for gdpr completeness checking of privacy policies using artificial intelligence. In *Proc. of the 39th IEEE/ACM International Conference on Automated Software Engineering*, ASE '24, page 2366–2369, New York, NY, USA. Association for Computing Machinery.

de Paula, R. F. O. (2022). Utilizando processamento de linguagem natural para avaliar políticas de privacidade. Trabalho de Graduação. Centro de Informática, UFPE.

Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)*.

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2nd edition.

Herold, R. (2015). *The Privacy Papers: Managing Technology, Consumer, Employee, and Legislative Actions*. Auerbach Publications, 1st edition.

Khalid, H. and Mehmood, A. (2019). Sentiment analysis of product reviews using machine learning techniques. In *2019 2nd International Conference on Intelligent Sustainable Systems (ICISS)*, pages 324–327. IEEE.

Li, F., Li, Q., Yang, W., and Zhang, S. (2021). Automatic privacy policy analysis and its legal implications. *IEEE Transactions on Information Forensics and Security*, 16:1267–1278.

Manning, C. D. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.

Peixoto, M., Ferreira, D., Cavalcanti, M., Silva, C., Vilela, J., Araújo, J., and Gorschek, T. (2023). The perspective of brazilian software developers on data privacy. *Journal of Systems and Software*, 195:111523.

Santana, E., Vilela, J., and Peixoto, M. (2022). Diretrizes para apresentação de políticas de privacidade voltadas à experiência do usuário. In *WER 2022*.

Singh, R. I., Sumeeth, M., and Miller, J. (2011). Evaluating the readability of privacy policies in mobile environments. *Intl. Journal of Mobile HCI (IJMHCI)*, 3(1):55–78.

Sinha, A., Sharma, A., and Jha, S. (2018). Identifying ptsd symptoms using twitter data: A microblogging platform-based machine learning approach. In *2018 2nd Intl. Conference on Trends in Electronics and Informatics (ICOEI)*, pages 1097–1101. IEEE.

Solove, D. J. (2015). *Understanding Privacy*. Harvard University Press, 1st edition.

Terra, A., Vilela, J., and Peixoto, M. (2022). A catalog of quality criteria to guide the assessment of applications' privacy policies. In *WER 2022*.

Zimmerman, J., Toubia, O., and Schwartz, H. A. (2015). The readability of privacy policies. *Computers in Human Behavior*, 52:479–487.