Classificação de boletins de ocorrência de crimes digitais com Aprendizado de Máquina

Ana Paula Vieira Dias[1], **Nelcileno Virgilio De Souza Araujo**[1]

[1] Instituto de Computação – Universidade Federal de Mato Grosso (UFMT)
Cuiabá – MT – Brasil

apaulavieira927@gmail.com, nelcileno@ic.ufmt.br

Abstract. *The study investigated the growth of online fraud in Cuiabá in 2024. To address this crime, Natural Language Processing (NLP) techniques were applied, and four Machine Learning models were developed using TF-IDF and Word2Vec representations with the Random Forest and Naive Bayes algorithms. The best result was achieved with Random Forest and TF-IDF (accuracy of 0.95), while the lowest performance was obtained with Naive Bayes and Word2Vec (0.59). The findings demonstrate the potential of Machine Learning for classifying cybercrimes, contributing to advances in digital security.*

Resumo. *O estudo investigou o crescimento do estelionato virtual em Cuiabá no ano de 2024. Para auxiliar no enfrentamento desse crime, foram aplicadas técnicas de Processamento de Linguagem Natural (PLN) e desenvolvidos quatro modelos de Aprendizado de Máquina com as representações TF-IDF e Word2Vec, utilizando os algoritmos Random Forest e Naive Bayes. O melhor resultado foi obtido com Random Forest e TF-IDF (acurácia de 0.95), enquanto o pior foi com Naive Bayes e Word2Vec (0.59). Os resultados demonstram o potencial do uso de Aprendizado de Máquina na classificação de crimes cibernéticos, contribuindo para o avanço da segurança digital.*

1. Introdução

O advento da Internet, ocorrido na década de 1960, passou a exercer grande influência sobre a sociedade global. Com o surgimento da Rede Mundial de Computadores e a popularização dos computadores pessoais, um número crescente de pessoas passou a se conectar por meio de dispositivos eletrônicos. Esse novo cenário trouxe inúmeras oportunidades, mas também abriu espaço para que criminosos utilizassem esses meios para a prática de delitos, que gradualmente se tornaram parte do cotidiano digital. Com a pandemia de Covid-19 e a maior demanda dos usuários por serviços e conexões online, houve uma mudança significativa no modo de interação da sociedade com o meio digital, o mundo migrou para o virtual, e assim, também o fizeram os criminosos.

No contexto dos crimes digitais, mais especificamente dos estelionatos virtuais, observou-se um aumento significativo a partir de sua tipificação em 2021, quando o crime de fraude eletrônica foi adicionado ao Código Penal Brasileiro, no artigo 171 [Brasil 2021]. A mudança observada na vida dos brasileiros é evidenciada pela tendência de migração dos crimes presenciais para os crimes digitais [Alcadipani et al. 2024]. Esses delitos não são necessariamente complexos: muitos dos golpes aplicados utilizam

técnicas de Engenharia Social, que têm como objetivo manipular as emoções da vítima e enganá-la, levando-a a revelar credenciais ou a enviar valores financeiros ao golpista.

O enfrentamento desses crimes é uma tarefa desafiadora para as delegacias do Brasil. Há um descompasso entre a quantidade de crimes registrados e os recursos disponíveis, sejam eles financeiros, humanos ou técnicos [Lima e Bueno 2023]. Ou seja, o combate a esses delitos crescentes exige preparo das instituições e de seus agentes, que necessitam de treinamento para lidar com a alta demanda.

Na esfera estadual, os relatórios de segurança pública mato-grossenses não trazem informações detalhadas sobre os estelionatos digitais. A ausência de dados oficiais abertos para consulta pública pode gerar a sensação de que tais delitos não possuem grande relevância ou de que permanecem impunes aos olhos da lei. Diante desse cenário, o presente trabalho tem como objetivo propor uma solução para a classificação automática de golpes digitais em boletins de ocorrência, por meio da aplicação de modelos de Aprendizado de Máquina. A proposta busca agilizar um processo atualmente moroso, permitindo que os crimes registrados como estelionato em meio eletrônico sejam subdivididos em categorias mais específicas. Dessa forma, será possível mapear os delitos de maneira mais precisa e fornecer informações relevantes para a segurança pública do estado de Mato Grosso.

Após esta introdução, este trabalho está estruturado nas seguintes seções: Revisão Bibliográfica, na qual são apresentados os principais estudos relacionados ao tema; Metodologia, que descreve os métodos empregados no desenvolvimento da pesquisa; Desenvolvimento, onde são detalhadas as etapas de aplicação das técnicas e métodos propostos; Conclusão, que expõe os resultados obtidos e as considerações finais; e, por fim, Trabalhos Futuros, que apresenta perspectivas para a continuidade desta pesquisa.

2. Revisão Bibliográfica

A revisão bibliográfica foi separada de forma a abordar duas áreas: Estelionato digital e Aprendizado de Máquina para classificação de crimes. De cada uma delas foram escolhidos os trabalhos mais relevantes, começando pelo artigo de [Nivette et al. 2021], que tratou de como a pandemia de Covid-19 influenciou e impactou a forma com que os crimes são realizados, ligando a diminuição de circulação das pessoas nas ruas e a queda dos crimes presenciais. Na mesma linha, [Lima e Bueno 2023] falam sobre a situação brasileira, onde os números totais de roubos diminuem dando espaço para crimes digitais, ainda é levantada a importância das polícias civis no combate aos crimes digitais. Em [Alcadipani et al. 2024] o crime digital é tratado como uma tendência em crescimento, que fortalece o crime organizado e a sensação de proteção que os golpistas sentem no ambiente virtual.

[Harkin, Whelan e Chang 2018] apresentam uma pesquisa feita com agentes policiais sobre os principais desafios sobre o aumento acelerado dos crimes digitais, um estudo na mesma área foi feito por [Hadlington et al. 2018] sobre os desafios enfrentados por policiais no combate ao cibercrime, no trabalho é sugerido que haja melhores treinamentos para instruir os agentes. Em [Henriques e Gonçalves 2024] são abordadas as principais leis na área de crimes digitais, entre elas a Lei Carolina Dieckmann (Lei n.º 12.737/2012) de dezembro de 2012, o Marco Civil da Internet (Lei n.º 12.695/2014), Lei Geral da Proteção de Dados (LGPD - Lei 13.709/2018) e a atualização do código

penal em 2021. [Ferreira 2024] trabalha com a possibilidade do uso de Aprendizado de Máquina para a manipulação e análise dos dados policiais a fim de facilitar o serviço dos agentes.

Na área de Aprendizado de Máquina, em [Almeida et al. 2022] foi realizada a classificação de crimes ocorridos no estado de São Paulo, tendo como objetivo analisar boletins de ocorrência relacionados a crimes de homicídio, a fim de verificar se haveria possibilidade de intervenção policial no momento do fato. [Brandenburg 2017] analisa relatórios policiais de crimes cibernéticos com Aprendizado de Máquina e Mineração de Texto de forma a classificá-los em três categorias distintas. [Lal et al. 2020] identificam crimes reportados por meio da rede social *Twitter* utilizando de técnicas de pré-processamento e quatro algoritmos de Aprendizado de Máquina. O estudo de [Kumar e Bhalaji 2016] teve como objetivo classificar crimes nas categorias "violento" e "não violento", empregando dois algoritmos de Aprendizado de Máquina. [Mandalapu et al. 2023] fazem uma revisão de 51 artigos mais relevantes para classificação de crimes com Aprendizado de Máquina.

[Ahmed, Nafis e Biswas 2017] utilizaram Mineração de Texto e Naive Bayes, via RapidMiner, para analisar boletins de ocorrência da Índia (2012–2014), mostrando que o algoritmo é eficaz na classificação e previsão de crimes. No trabalho de [Padirayon et al. 2021] foi utilizado o Naive Bayes para classificar crimes em Sanchez Mira (2013–2019), identificando padrões como maior incidência às terças-feiras, às 14h, e no mês de maio. Em [Riego e Villarba 2023] usaram TF-IDF e Naive Bayes Multinomial para verificar a credibilidade de notícias no *Twitter* nas Filipinas, alcançando 99.46% de acurácia. [Passos et al. 2024] Implementaram TF-IDF com Naive Bayes e Random Forest para classificar dados sensíveis segundo a LGPD, obtendo 93% de acurácia com Random Forest e 89% com Naive Bayes. Por fim, no trabalho de [Chingmuankim e Jindal 2023] foram comparadas as variações do Naive Bayes (Gaussian, Bernoulli e Multinomial) com TF-IDF, demonstrando melhor desempenho em classificação textual do que o modelo Bag of Words.

3. Metodologia

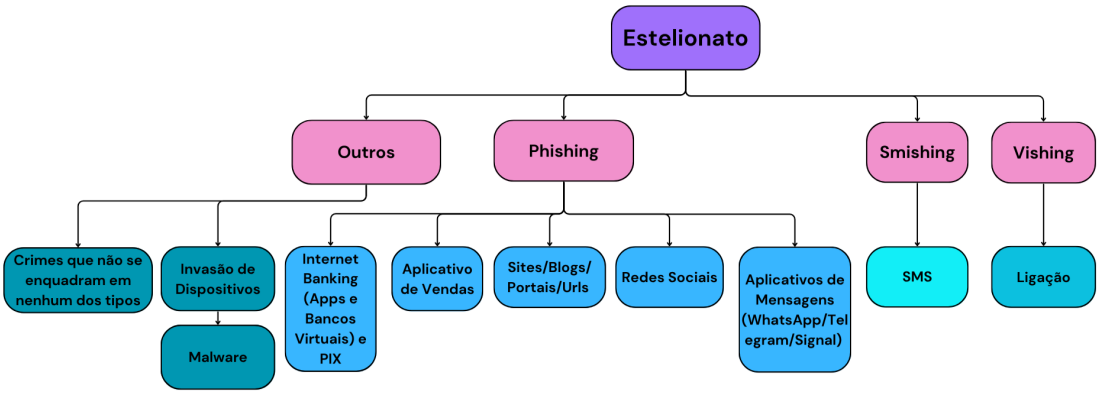
O presente trabalho teve por objetivo desenvolver e avaliar quatro modelos de Aprendizado de Máquina, a pesquisa é aplicada e adota uma abordagem quantitativa. A metodologia foi dividida nas seguintes fases: Levantamento Bibliográfico; Coleta dos dados; Classificação Manual; Pré-Processamento; Vetorização; Balanceamento das Classes; Criação do Modelo; Treino/Teste e Avaliação. Foi utilizada a técnica de Mineração de Texto para o *Knowledge Discovery from Text* (KDT), Descoberta de Conhecimento a partir de Texto, relacionada ao Processamento de Linguagem Natural (PLN), área da computação que busca extrair e interpretar significados presentes na linguagem humana.

O levantamento bibliográfico foi abordado na seção anterior e conforme o analisado indetificou-se a necessidade de trabalhos acadêmicos que abordem os golpes digitais na área de computação, isso demonstra a importância da presente pesquisa. Passando para a segunda fase, a Coleta de Dados foi feita a partir de um pedido formal à Diretoria da Polícia Judicial Civil (PJC) de Mato Grosso, a fim de obter as narrativas contidas em Boletins de Ocorrência (B.O) registrados em Cuiabá no ano de 2024. Com esses dados, passou-se para a terceira fase, Classificação Manual, onde foram lidas as 2.707 narrativas

obtidas. Esse processo foi muito importante para entender como as narrativas estavam distribuídas e qual o seu contexto, esses textos são informados pela vítima no momento do B.O.

Mediante a leitura dos dados, foi entendido que essas narrativas deveriam ser separadas nas quatro seguintes classes: “Phishing”, ‘’Smishing”, “Vishing”, e “Outros”. Como pode ser observado na Figura 1, a classificação foi feita de modo a entender o meio onde os crimes ocorriam. Dessa forma, com as narrativas devidamente classificadas e retiradas as que não faziam sentido para a composição do *dataset*, o resultado final teve um conjunto de 2673 narrativas. Sendo a distribuição final: “Phishing”: 1560, “Outros”: 655, “Vishing”: 411, “Smishing”: 47.

Figura 1. Classificação dos golpes



Fonte: Adaptado de [Chiew, Yong e Tan 2018]

Além disso, conforme apresentado na Tabela 1, também a partir da classificação manual, foi possível separar e identificar os principais tipos de golpes ocorridos. Na classe “Phishing” o principal foi o golpe do Intermediário, utilizando os aplicativos OLX ou Facebook Marketplace. Sites falsos que se passavam por empresas e instituições confiáveis também ocorreram como no exemplo de um site que imitava o do Detran-MT. No “Vishing” e no “Smishing” houve golpes parecidos, com uma falsa central de banco entrando em contato e alegando uma compra suspeita.

Tabela 1. Principais golpes ocorridos em Cuiabá no ano de 2024

Classificação	Golpe	Descrição
Phishing	Falso contato no WhatsApp	O golpista se passa por um parente ou amigo da vítima e solicita transferências com urgência.
	Golpe da OLX – Intermediário	Envolve comprador, vendedor e o golpista que se apresenta como intermediário. O comprador envia o dinheiro ao golpista, acreditando tratar-se do vendedor.

Continua

Classificação	Golpe	Descrição
	Golpe do Marketplace – Intermediário	Variante do golpe da OLX, aplicado no Marketplace do Facebook. O golpista intermedia falsamente a negociação e recebe o pagamento.
	Golpe das tarefas diárias	A vítima acredita estar realizando trabalho remoto com promessas de lucros crescentes mediante investimentos, mas nunca recebe retorno financeiro.
	Sites de compras falsos	Sites fraudulentos imitam lojas virtuais legítimas, enganando a vítima a realizar compras falsas e fornecer dados sensíveis.
	Redes sociais – anúncios de investimento falso	A partir de contas hackeadas, golpistas divulgam falsos investimentos. Vítimas acreditam na veracidade e acabam transferindo dinheiro ou tendo contas roubadas.
	Golpe do advogado	A vítima recebe mensagens de supostamente seu advogado, informando um ganho de causa, mas precisa pagar taxas para liberá-lo. Trata-se de um golpe com número clonado.
	Site falso do Detran	Sites falsos imitam o site oficial do Detran. A vítima gera boletos falsos acreditando estar quitando débitos reais.
Vishing	Ligação da falsa central do banco	Golpista liga fingindo ser da central de segurança do banco, alegando compras suspeitas e induzindo a vítima a fornecer dados bancários ou senhas.
	Ligação de falsa facção criminosa	A vítima recebe uma ligação com ameaças e extorsão, acreditando estar sendo vigiada por uma facção criminosa.
Smishing	Aviso de falsa compra (falsa central do banco)	A vítima recebe um SMS com alerta de compra suspeita. Ao entrar em contato, é induzida a fornecer dados bancários.

Continua

Classificação	Golpe	Descrição
	Atualização de dados de programas sociais	O golpista envia SMS solicitando atualização de dados em programas sociais. O link leva a um site falso que rouba informações pessoais.

Conclusão

Exemplos comuns na classe "Outros" incluíam: Contas em redes sociais hackeadas, não é possível identificar qual meio foi utilizado para a ação; Empréstimos realizados sem consentimento; Débitos em contas bancárias de Instituto Nacional do Seguro Social (INSS) sem consentimento; Golpes amorosos que não utilizam de meios eletrônicos claros.

Passadas as três primeiras etapas da Mineração de Texto, na quarta, Pré-Processamento, foi realizada a limpeza e padronização dos dados. Os passos incluídos nessa parte foram a remoção de frases com Expressões Regulares (REGEX), remoção de *stopwords*, utilização de dicionários criados para remover palavras específicas e a aplicação do *stemming* para reduzir as palavras ao seu radical. A vetorização foi a etapa seguinte, onde as palavras são transformadas em números, para isso foram utilizados dois métodos: TF-IDF e Word2Vec.

O TF-IDF é utilizado na vetorização para o cálculo da relevância da palavra no contexto do texto. Palavras frequentes em um único documento tendem a ter valores mais altos no TF-IDF, enquanto termos muito comuns, como preposições, recebem valores baixos por terem pouca relevância na busca [Ramos 2003]. O Word2Vec é uma técnica de *embedding*, que representa a palavra de forma a manter seu contexto, uma palavras e outras ao seu redor possuem representações baseadas em vetores de N dimensões. No caso do Word2Vec uma de suas arquitetura é a *Skip-Gram*, em que a palavra central é a entrada e a saída são as palavras ao seu redor [Mikolov et al. 2013]. Essa arquitetura foi escolhida após alguns testes, verificando que modelos que se utilizavam do *Skip-Gram* obtinham melhor desempenho.

Além disso, foram escolhidos dois algoritmos para a criação dos modelos, o primeiro sendo o Random Forest, e o segundo o Naive Bayes Multinomial, ambos muito utilizados na área de classificação de texto. O Random Forest trabalha com a criação de uma "floresta" de árvores de decisão e o Naive Bayes utiliza o Teorema de Bayes, um cálculo estatístico para determinar a classe.

4. Desenvolvimento

A próxima etapa foi o balanceamento das classes, para isso foi utilizado o SMOTE (*Synthetic Minority Oversampling TEchnique*). Essa técnica utiliza a sobreamostragem da classe minoritária, ou seja, cria exemplos sintéticos [Chawla et al. 2002]. A seguir, será discutido o desenvolvimento dos modelos, e é importante frisar que foi utilizado o *Cross Validation* em 5 *folds*. Ou seja, o conjunto foi dividido em cinco partes, onde quatro delas são para treinamento e uma para teste, esse processo foi repetido cinco vezes, o resultado obtido é a média dos valores.

No modelo Random Forest com TF-IDF, inicialmente foram ajustados os hiperparâmetros do Random Forest, `n_estimators` e `max_depth`. O modelo foi trei-

nado algumas vezes com diferentes combinações desses parâmetros para identificar a configuração que proporcionasse a maior acurácia, cujos resultados foram visualizados em um mapa de calor. Observou-se que a combinação `n_estimators` = 150 e `max_depth` = 18 alcançou a acurácia máxima de 0.95, um resultado muito bom, considerando a alta profundidade das árvores e a ausência de indícios de *overfitting*.

Utilizando o Random Forest com Word2Vec, foi feito um mapa de calor com os valores de acurácia para os hiperparâmetros do Random Forest. Observou-se que as curvas de aprendizado, que foram geradas para análise, apresentaram valores de treino muito próximos de 1.0, indicando um possível *overfitting*. Para investigar, o conjunto de dados foi dividido em 75% para treino, 20% para teste e 5% para validação. Com o novo treinamento, o modelo alcançou acurácia de 0.92, mantendo bons valores de *F1-score*. Na validação, feita com 5% de dados não vistos e sem balanceamento, a acurácia foi de 0.71, resultado considerado aceitável devido ao desbalanceamento das classes.

Passando para o Naive Bayes com TF-IDF, não foram necessárias alterações de parâmetros, assim o modelo foi treinado e obteve um resultado de 0.89 na acurácia, um bom resultado. Ao montar um relatório com as métricas, a análise por classe mostrou que "Outros" teve *precision* de 0.93 e *recall* de 0.78, indicando bom desempenho, mas com alguns casos não reconhecidos. "Phishing" apresentou equilíbrio entre as métricas, enquanto "Smishing" obteve valores altos de *precision* e *recall*, possivelmente devido às entradas sintéticas criadas. O "Vishing" teve *recall* de 0.96, mostrando ótima capacidade de identificação. Na matriz de confusão, a classe "Smishing" foi classificada corretamente em todos os casos, enquanto "Outros" apresentou menor desempenho.

O modelo Naive Bayes com Word2Vec obteve acurácia geral de 0.59, indicando dificuldade em generalizar bem o problema. A classe "Phishing" apresentou o melhor desempenho (F1-Score de 0.65 e recall de 0.72), seguida por "Vishing" (F1-Score de 0,68). O "Smishing" teve desempenho moderado (F1-Score de 0.59), enquanto "Outros" obteve o pior resultado (F1-Score de 0.48). A análise da matriz de confusão mostrou confusões frequentes entre as classes "Smishing" e "Vishing", e entre "Outros" e "Phishing", indicando dificuldade do modelo em distinguir padrões específicos. No entanto, o desempenho geral sugere que o Naive Bayes tem limitações para capturar relações complexas em *embeddings* Word2Vec.

5. Conclusão

Partindo para os resultados, conforme evidenciado na Tabela 2, o melhor modelo foi o Random Forest combinado com o TF-IDF, demonstrando que o uso de uma representação vetorial simples, aliado a um classificador robusto, pode ser muito eficaz para a solução desse problema. O segundo melhor modelo foi o que combinou o Random Forest com o Word2Vec, com acurácia de 0.92. Esses resultados indicam que, para este conjunto de dados, a análise baseada na frequência de termos foi mais eficaz do que as representações semânticas obtidas pelos *embeddings*.

Por outro lado, os modelos baseados em Naive Bayes apresentaram desempenho inferior. O modelo Naive Bayes com TF-IDF obteve acurácia de 0.89, resultado satisfatório e coerente com a natureza probabilística do classificador. Contudo, ao ser combinado com Word2Vec, o desempenho caiu significativamente para 0.59, indicando que o Naive Bayes não se adapta bem a representações vetoriais contínuas, especialmente

Tabela 2. Resultados de acurácia entre os modelos

	Random Forest	**Naive Bayes**
TF-IDF	0.95	0.89
Word2Vec	0.92	0.59

quando é necessário modificar seu funcionamento original para evitar valores negativos.

6. Trabalhos Futuros

Os resultados desta pesquisa demonstram que a combinação entre representações textuais e algoritmos de classificação impacta diretamente o desempenho dos modelos. A partir disso, algumas direções podem ser exploradas em trabalhos futuros para aprimorar a abordagem.

Primeiramente, recomenda-se o uso de *embeddings* mais avançados, como *fastText* e modelos baseados em *transformers* (por exemplo, BERT e suas variantes), capazes de capturar relações contextuais mais complexas do que aquelas baseadas apenas na frequência de termos. Além disso, sugere-se a investigação de outros modelos de aprendizado, como o *XGBoost*, que aprimora o desempenho das árvores de decisão, ou ainda o uso de redes neurais profundas, que podem oferecer maior capacidade de generalização e representação dos dados textuais.

7. Agradecimentos

Um agradecimento especial a PJC-MT por ter cedido os dados necessários para a realização desta pesquisa.

Referências

AHMED, W.; NAFIS, M. T.; BISWAS, S. S. Performance analysis of naïve bayes algorithm on crime data using rapid miner. *International Journal of Advanced Research in Computer Science*, v. 8, n. 5, May–June 2017. ISSN 0976-5697.

ALCADIPANI et al. *Anuário Brasileiro de Segurança Pública 2024*. São Paulo: Fórum Brasileiro de Segurança Pública, 2024. ISSN 1983-7364.

ALMEIDA, M. G. de et al. Utilização de machine learning para classificaç ao de crimes de morte no estado de são paulo. 2022.

BRANDENBURG, M. *Text Classification of Dutch police records*. Dissertação (Mestrado) — Utrecht University, 2017.

BRASIL. Lei nº 14.155, de 27 de maio de 2021. Brasília, DF, 2021. Disponível em: <https://www.planalto.gov.br/ccivil$_03/_{a}to2019-2022/2021/lei/l14155.htm$¿.

CHAWLA, N. V. et al. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, AI Access Foundation and Morgan Kaufmann Publishers, v. 16, p. 321–357, 2002.

CHIEW, K. L.; YONG, K. S. C.; TAN, C. L. A survey of phishing attacks: Their types, vectors and technical approaches. *Expert Systems with Applications*, v. 106, p. 1–20, 2018. ISSN 0957-4174. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417418302070>.

CHINGMUANKIM; JINDAL, R. Classification and analysis of textual data using naive bayes with tf-idf. New Delhi, India, 2023.

FERREIRA, M. F. G. *Estelionato e ambiente virtual: Desafios para agências policiais em Minas Gerais a partir do olhar da complexidade e das Ciências Policiais*. Dissertação (Mestrado) — Universidade do Estado de Minas Gerais, 2024.

HADLINGTON, L. et al. A qualitative exploration of police officers' experiences, challenges, and perceptions of cybercrime. *Policing: A Journal of Policy and Practice*, v. 15, n. 1, p. 34–43, 12 2018.

HARKIN, D.; WHELAN, C.; CHANG, L. The challenges facing specialist police cybercrime units: an empirical analysis. *Police Practice and Research*, Routledge, v. 19, n. 6, p. 519–536, 2018. Disponível em: <https://doi.org/10.1080/15614263.2018.1507889>.

HENRIQUES, T. A.; GONçALVES, S. M. Crimes digitais: análise sobre o estelionato virtual. *Revista Eletrônica de Ciências Jurídicas*, v. 14, n. 1, out. 2024. Disponível em: <https://revista.fadipa.br/index.php/cjuridicas/article/view/576>.

KUMAR, K. B. S.; BHALAJI, N. A study on classification algorithms for crime records. In: UNAL, A. et al. (Ed.). *Smart Trends in Information Technology and Computer Communications*. Singapore: Springer Nature Singapore, 2016. p. 873–880. ISBN 978-981-10-3433-6.

LAL, S. et al. Analysis and classification of crime tweets. *Procedia Computer Science*, v. 167, p. 1911–1919, 2020. ISSN 1877-0509. International Conference on Computational Intelligence and Data Science. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050920306761>.

LIMA, R. S. d.; BUENO, S. *Anuário Brasileiro de Segurança Pública 2023*. São Paulo: Fórum Brasileiro de Segurança Pública, 2023. ISSN 1983-7364.

MANDALAPU, V. et al. Crime prediction using machine learning and deep learning: A systematic review and future directions. *IEEE Access*, v. 11, p. 60153–60170, 2023.

MIKOLOV, T. et al. *Efficient Estimation of Word Representations in Vector Space*. 2013. Disponível em: <https://arxiv.org/abs/1301.3781>.

NIVETTE, A. E. et al. A global analysis of the impact of covid-19 stay-at-home restrictions on crime. *Nature Human Behaviour*, v. 5, n. 7, p. 868–877, 2021. ISSN 2397-3374. Disponível em: <https://doi.org/10.1038/s41562-021-01139-z>.

PADIRAYON, L. M. et al. Mining the crime data using naïve bayes model. *Indonesian Journal of Electrical Engineering and Computer Science*, v. 23, n. 2, p. 1084–1092, August 2021. ISSN 2502-4752.

PASSOS, E. H. d. S. et al. Identificação e classificação de dados sensíveis usando técnicas de processamento de linguagem natural (pln). *Revista Direitos Democráticos & Estado Moderno*, v. 3, n. 12, 2024.

RAMOS, J. Using tf-idf to determine word relevance in document queries. In: CITESEER. *Proceedings of the first instructional conference on machine learning*. [S.l.], 2003. v. 242, n. 1, p. 29–48.

RIEGO, N. C. R.; VILLARBA, D. B. Utilization of multinomial naive bayes algorithm and term frequency-inverse document frequency (tf-idf vectorizer) in checking the cre-

dibility of news tweet in the philippines. General Luna, corner Muralla St, Intramuros, Manila, 1002 Metro Manila, 2023.