XXIX CONGRESSO NACIONAL DO CONPEDI BALNEÁRIO CAMBORIU - SC

DIREITO, GOVERNANÇA E NOVAS TECNOLOGIAS IV

MAIQUEL ÂNGELO DEZORDI WERMUTH

LEONEL SEVERO ROCHA

D597

Direito, governança e novas tecnologias IV [Recurso eletrônico on-line] organização CONPEDI

Coordenadores: Leonel Severo Rocha; Maiquel Ângelo Dezordi Wermuth.

– Florianópolis: CONPEDI, 2022.

Inclui bibliografia
ISBN: 978-65-5648-626-0
Modo de acesso: www.conpedi.org.br em publicações
Tema: Constitucionalismo, Desenvolvimento, Sustentabilidade e Smart Cities

1. Direito – Estudo e ensino (Pós-graduação) – Encontros Nacionais. 2. Direito. 3. Governança e novas tecnologias.

XXIX Congresso Nacional do CONPEDI Balneário Camboriu - SC (3: 2022: Florianópolis, Brasil).

CDU: 34

Conselho Nacional de Pesquisa
e Pós-Graduação em Direito Florianópolis
Santa Catarina – Brasil
www.conpedi.org.br

XXIX CONGRESSO NACIONAL DO CONPEDI BALNEÁRIO CAMBORIU - SC

DIREITO, GOVERNANÇA E NOVAS TECNOLOGIAS IV

Apresentação

Apresentação

Apresentam-se os trabalhos exibidos, no dia 07 de dezembro de 2022, no Grupo de Trabalho (GT) "Direito, Governança e Novas Tecnologias IV", no âmbito do XXIX Congresso do Conselho Nacional de Pesquisa e Pós-Graduação em Direito – CONPEDI – "Constitucionalismo, Desenvolvimento, Sustentabilidade e Smart Cities" – realizado no campus da UNIVALI em Balneário Camboriú/SC.

O GT, de coordenação dos trabalhos dos Professores Doutores Leonel Severo Rocha e Maiquel Ângelo Dezordi Wermuth, envolveu 20 artigos que, entre perspectivas teóricas e práticas, nos impulsionam à imprescindibilidade da observação dos dilemas da atualidade a partir da ótica do direito, da governança e das novas tecnologias. Os trabalhos apresentados abriram caminho para uma importante discussão, a partir da qual os pesquisadores do Direito puderam interagir, levando-se em consideração o momento político, social e econômico vivido pela sociedade brasileira.

O primeiro trabalho é "DISTÚRBIO DE INFORMAÇÃO: FAKE NEWS E PSICOLOGIA" desenvolvido por Lilian Novakoski e Adriane Nogueira Fauth de Freitas. No referido estudo, os autores analisam o fenômeno das fake news desde a criação da informação falsa até a recepção da notícia pelo leitor. A pesquisa trata da epidemia de informação, traçando comentários voltados a uma economia comportamental e a própria relação do direito com a psicologia.

"EFICÁCIA E APLICABILIDADE DA INTELIGÊNCIA ARTIFICIAL COMO MECANISMO REDUCIONAL DO CUSTO DO PROCESSO JURÍDICO", desenvolvido por Ricardo da Silveira e Silva e Rodrigo Valente Giublin Teixeira trata da aplicação da Inteligência Artificial como instrumento eficaz na redução dos custos processuais e consequente facilitação do acesso à justiça.

Letícia Feliciana dos Santos Cruz, Stephanny Resende De Melo, Victor Ribeiro Barreto são autores do artigo “O DILEMA DAS REDES” E AS TECNOLOGIAS DE VIGILÂNCIA NAS CIDADES GLOBALIZADAS: COMO SE PROTEGER?”, cujo estudo tem como objetivo central a discussão da segurança de dados pessoais pelas empresas.

O tema “SMART CITIES E O USO DE CÂMERAS DE VIGILÂNCIA COM INTELIGÊNCIA ARTIFICIAL E RECONHECIMENTO FACIAL” desenvolvido por Emerson Gabardo e Juliana Horn Machado Philippi tem como objetivo analisar as consequências do uso de câmeras de monitoramento com inteligência artificial e reconhecimento facial no contexto das smart cities, bem como propor regulação para evitar violações a direitos fundamentais.

O artigo de autoria de Pedro Augusto Gregorini e Maria Paula Costa Bertran Munoz, intitulado como “JURIMETRIA APLICADA ÀS DEMANDAS BANCÁRIAS: ESTATÍSTICA DOS TIPOS DE PROCEDIMENTO E ASSUNTOS MAIS FREQUENTES NAS AÇÕES AJUIZADAS PELOS BANCOS NO TRIBUNAL DE JUSTIÇA DE SÃO PAULO”, investiga a proporção de ações em que os bancos são autores no estado de São Paulo e dos tipos de procedimento e assuntos mais frequentes.

De autoria de Lourenço de Miranda Freire Neto, Larissa Dias Puerta de Miranda Freire e Thomaz Matheus Pereira Magalhães, é o artigo “PROTEÇÃO DE DADOS E GOVERNANÇA CORPORATIVA SOCIAL E AMBIENTAL COMO INSTRUMENTOS DE DEFESA DOS TRABALHADORES E CONSUMIDORES”, que parte dos avanços tecnológicos e dos novos meios de comunicação para analisar as dinâmicas das relações de emprego que vem se alterando rapidamente nos últimos anos.

“POSSIBILIDADES PARA UMA GOVERNANÇA GLOBAL: A EDUCAÇÃO COMO INSTRUMENTO DE GOVERNANÇA TRANSNACIONAL”, desenvolvido por Ornella Cristine Amaya e Clovis Demarchi, cuja pesquisa discute o conceito de educação para a era das acelerações.

“OS INFLUENCIADORES DIGITAIS NAS RELAÇÕES DE CONSUMO: CONTRIBUIÇÕES DOUTRINÁRIAS E JURISDICIONAIS SOBRE O TEMA”, é o trabalho de Isadora Balestrin Guterres, Luiz Henrique Silveira Dos Santos e Rosane Leal Da Silva. Os autores analisam como as plataformas digitais são utilizadas por influenciadores – pessoas que exploram sua imagem para divulgar produtos e serviços em seus canais – o que suscita que se questione qual a natureza jurídica de sua atuação e suas responsabilidades em relação ao consumidor.

O artigo “GOVERNO DIGITAL E NOVAS TECNOLOGIAS: ANALISE DA ADOÇÃO DA BLOCKCHAIN NA ADMINISTRAÇÃO PÚBLICA”, desenvolvido por Caroline Vicente Moi, Alexandre Barbosa da Silva e Rahiza Karaziaki Merquides, cujo estudo contextualiza a adoção da BLOCKCHAIN na administração pública, suscitando um aumento da eficiência e na redução de custos quando adotadas pelos entes públicos.

Pedro Henrique Freire Vazatta e Marcos Vinícius Viana da Silva são autores do artigo “DADOS OBTIDOS DAS ESTAÇÕES DE RÁDIO BASE NA CONTRIBUIÇÃO DA INVESTIGAÇÃO CRIMINAL E O DIREITO FUNDAMENTAL À PRIVACIDADE”, que dispõe sobre a proteção da intimidade e da vida privada e a sua respectiva relação com a coleta de dados das estações de rádio base.

“COMPLIANCE NA SOCIEDADE DE RISCO” é o trabalho de Renato Campos Andrade, em que o autor parte da análise dos desafios do compliance na sociedade de risco de Ulrich Beck.

Cibele Andréa de Godoy Fonseca, Emerson Wendt e Ismar Frango Silveira desenvolveram o trabalho “CRIMES CIBERNÉTICOS E SUA PREVISÃO COM USO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA E DE DADOS HETEROGÊNEOS: UM MAPEAMENTO SISTEMÁTICO DE TÉCNICAS DE ANÁLISE E PREDITIVIDADE DE DELITOS”, em que o referido estudo trata do avanço da prática de crimes cibernéticos, suscitando o anonimato de criminosos pelas falhas na persecução criminal na esfera cibernética.

Matheus Adriano Paulo e Márcio Ricardo Staffen explanaram em seu artigo “CONSIDERAÇÕES SOBRE A PROTEÇÃO DE DADOS PESSOAIS COMO UM MECANISMO DE DIREITO TRANSNACIONAL”, acerca da proteção de dados pessoais como um mecanismo de direito transnacional, mencionando o case envolvendo França e Google na política de cookies e no rastreamento/compartilhamento de dados.

“CIBERESPAÇO E O ASSÉDIO A DEMOCRACIA: A CONSTRUÇÃO DE UMA PONTE ENTRE A REGULAÇÃO E A LIBERDADE DE ESCOLHA” é o trabalho de Gustavo Marshal Fell Terra, Marco Antonio Zimermann Simão e Willian Amboni Scheffer, oriundo de pesquisa em que os autores tratam de estudos ligados aos assédios sofridos pela democracia frente às novas práticas virtuais. A análise parte do pressuposto existente entre as regulações atuais e as que surgirão e de que modo esse arcabouço técnico pode influenciar a liberdade na Constituição Federal.

Ranivia Maria Albuquerque Araújo e Lara Jessica Viana Severiano são autores do artigo “A RESPONSABILIDADE CIVIL DECORRENTE DOS ATOS JURÍDICOS PRATICADOS PELOS SISTEMAS DE INTELIGÊNCIA ARTIFICIAL”, em que se busca analisar a possibilidade de responsabilização da inteligência artificial.

“INTELIGÊNCIA ARTIFICIAL E VIOLÊNCIA DOMÉSTICA: A GARANTIA À INTEGRIDADE FÍSICA POR MEIO DA RELATIVIZAÇÃO DA PRIVACIDADE” de

Isabelle Brito Bezerra Mendes trata da relativização da proteção de dados diante de situações de violência doméstica e da possibilidade legal de utilização da inteligência artificial como prova nesses tipos de delitos.

“A INTELIGÊNCIA ARTIFICIAL COMO FERRAMENTA PARA TRAZER EFETIVIDADE AO PROCESSO JUDICIAL” de Marcus Jardim da Silva, cujo trabalho trata a inteligência artificial como meio de efetivação da justiça, citando o caso do robô pesquisador.

O artigo “A LEI GERAL DE PROTEÇÃO DE DADOS: UMA ANÁLISE DO PAPEL DO PODER PÚBLICO NA PROTEÇÃO DOS DADOS PESSOAIS DO CIDADÃO” escrito por Camila Barreto Pinto Silva e Cristina Barbosa Rodrigues, tem por objetivo esclarecer a forma como a administração deverá tratar os dados pessoais diante da LGPD.

“ORGANIZAÇÕES, RISCO E PROTEÇÃO DE DADOS PESSOAIS NA CULTURA DAS REDES: OBSERVANDO O PAPEL DO RELATÓRIO DE IMPACTO À PROTEÇÃO DE DADOS PESSOAIS (RIPD)” de Ariel Augusto Lira de Moura, Bernardo Leandro Carvalho Costa e Leonel Severo Rocha objetiva analisar o Relatório de Impacto a Proteção de Dados na cultura das redes a partir do questionamento sobre que de pontos pode-se observar de modo a conectá-lo à um contexto maior de transformações da sociedade contemporânea.

O artigo “A AUTORREGULAÇÃO REGULADA DAS PLATAFORMAS DIGITAIS: UMA APROXIMAÇÃO AO COMPARTILHAMENTO DE INFORMAÇÕES” escrito por Fabio Luis Celli, Alfredo Copetti e Sylvia Cristina Gonçalves da Silva analisa a necessidade de regulação das plataformas digitais relacionadas às redes sociais e aos aplicativos de serviços de mensageria privada para o compartilhamento de informações por parte dos usuários.

Agradecemos a todos os pesquisadores da presente obra pela sua inestimável colaboração. Desejamos uma ótima e proveitosa leitura!

Coordenadores:

Prof. Dr. Leonel Severo Rocha – UNISINOS

Prof. Dr. Maiquel Ângelo Dezordi Wermuth - UNIJUÍ

CRIMES CIBERNÉTICOS E SUA PREVISÃO COM USO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA E DE DADOS HETEROGÊNEOS: UM MAPEAMENTO SISTEMÁTICO DE TÉCNICAS DE ANÁLISE E PREDITIVIDADE DE DELITOS

CYBER CRIMES AND THEIR PREDICTION USING MACHINE LEARNING ALGORITHMS AND HETEROGENEOUS DATA: A SYSTEMATIC MAPPING OF CRIME ANALYSIS AND PREDICTABILITY TECHNIQUES

Cibele Andréa de Godoy Fonseca [1]
Emerson Wendt [2]
Ismar Frango Silveira [3]

Resumo

Devido ao avanço das tecnologias e o crescimento do uso da internet, principalmente no período de pandemia global que iniciou em 2020, observar-se o avanço do crime cibernético, acreditando os criminosos no anonimato, no não monitoramento, havendo falhas na persecução criminal na esfera cibernética, consequentemente sensação de impunidade e de insegurança. Outros avanços são vistos, como (a) quantidade de dados que estão disponíveis nas mídias digitais, (b) o avanço da computação em nuvem, (c) o uso de algoritmos e (d) a inteligência artificial. Reforça-se a importância e necessidade de analisar e prever a ocorrência de crimes cibernéticos com o objetivo de ajudar a polícia a planejar e executar ações preventivas. No Brasil, o sistema de segurança pública não otimizou, não padronizou e não uniformizou seus processos na mesma velocidade do avanço da tecnologia e dos crimes cibernéticos. Por esse motivo, os dados quali-quantitativos dos boletins de ocorrência de crimes cibernéticos não estão organizadamente disponíveis aos cidadãos/pesquisadores, o que não possibilita a manipulação deles por meio do uso de inteligência artificial, especificamente aprendizado de máquina. Nesta pesquisa, verifica-se a falta de dados sobre crimes cibernéticos no Brasil e o quanto isso impossibilita a análise e possíveis prevenções desses crimes. Assim, realizou-se um mapeamento sistemático que mostra o que está sendo feito em relação à análise e previsão de outros tipos de crimes em todo o mundo, além dos cibernéticos, cujos dados heterogêneos são abertos, ou seja, estão disponíveis, bem como suas análises e previsões através do uso do aprendizado de máquina.

---

[1] Doutoranda em Engenharia Elétrica e de Computação pela Universidade Presbiteriana Mackenzie, Mestre em Engenharia Elétrica e de Computação, Bacharel em Ciências com especialização em Matemática, Advogada. ORCID: https://orcid.org/my-orcid?orcid=0000-0001-5270-1480.

[2] Mestre e Doutorando em Direito e Sociedade (PPGD Universidade La Salle Canoas-RS). Delegado de Polícia Civil no RS. Membro do Conselho Superior da Polícia Civil do RS. Lattes: http://lattes.cnpq.br /9475388941521093.

[3] Doutor em Engenharia Elétrica pela POLI-USP, mestre em Ciências - Informática pelo ITA. Professor Adjunto da Universidade Presbiteriana Mackenzie e Professor Titular da Universidade Cruzeiro do Sul. ORCID: https://orcid.org/0000-0001-8029-072X.

Palavras-chave: Aprendizado de máquina, Crimes cibernéticos, Dados heterogêneos, Inteligência artificial, Prevenção

Abstract/Resumen/Résumé

Due to the advancement of technologies and the growth of internet use, especially in the period of the global pandemic that started in 2020, the advancement of cybercrime can be observed, believing criminals in anonymity, in non-monitoring, with failures in criminal prosecution in the sphere cybernetics, consequently a feeling of impunity and insecurity. Other advances are seen, such as (a) the amount of data that is available in digital media, (b) the advancement of cloud computing, (c) the use of algorithms, and (d) artificial intelligence. It reinforces the importance and needs to analyze and predict the occurrence of cybercrimes to help the police to plan and execute preventive actions. In Brazil, the public security system did not optimize, and standardize its processes at the same speed as the advancement of technology and cyber crimes. For this reason, the quality-quantitative data of cybercrime bulletins are not organized and available to citizens/researchers, which does not allow their manipulation through the use of artificial intelligence, specifically machine learning. In this research, there is a lack of data on cyber crimes in Brazil, and how much this makes the analysis and possible prevention of these crimes impossible. Thus, a systematic mapping was carried out that shows what is being done concerning the analysis and prediction of other types of crimes around the world, in addition to cybernetic ones, whose heterogeneous data are open, that is, they are available, as well as their analysis and predictions through the use of machine learning.

Keywords/Palabras-claves/Mots-clés: Machine learning, Cyber crimes, Heterogeneous data, Artificial intelligence, Prevention

1 INTRODUÇÃO

De acordo com o artigo publicado na CNN Brasil em agosto de 2022, no contexto da América Latina o Brasil é o segundo país no ranking dos ataques cibernéticos, atrás apenas do México, que teve 85 bilhões de tentativas. Essa publicação também informa que durante o primeiro semestre de 2022, o Brasil registrou 31,5 bilhões de tentativas de ataques de crimes cibernéticos, sendo que esse número é 94% maior em relação ao primeiro semestre de 2021, quando foram feitos 16,2 bilhões de registros (OLIVEIRA, 2022). Para Palmieri, Shortland e McGarry (2021, n.p.) "A Internet proporcionou às pessoas uma infinidade de oportunidades para se envolver em crimes *online*".

A pandemia da Covid-19 proporcionou o incremento da utilização da Internet e, consequentemente, a exploração pelos criminosos. Dados revelados pelo Fórum Brasileiro de Segurança Pública indicam o aumento de 497,5%, na variação de 2018 para 2021, no estelionato por meio eletrônico e no caso do estelionato sem especificação do meio, um aumento de 179,8%, no mesmo período. Importa observar que nem todos os Estados fazem a diferenciação, ou seja, a categorização do meio pelo qual o crime foi cometido, na hora do registro de ocorrência (BUENO; LIMA, 2022, 110-1, 120-1), prejudicando uma análise mais ampliada e correta.

Nesse panorama, o sistema de persecução criminal brasileiro não se encontra preparado para atuar em relação aos crimes que ocorrem na rede de computadores, mundial ou privada, tanto antes quanto durante o processo de investigação, além de não ter uma atuação proativa efetiva, atuando basicamente de forma reativa. Esse cenário faz com que, na população, prolifere uma sensação de insegurança na utilização das mídias digitais.

Analisar e prever crimes cibernéticos, o desafio da preditividade, poderá fazer com que a população se sinta mais segura. Neste contexto, justifica-se a análise realizada neste artigo, pois pretende-se mostrar o que é feito atualmente em relação ao uso de dados heterogêneos de crimes e como a inteligência artificial (AI) pode ser útil para análise e previsão de crimes, num sentido geral. A proposta, então, é usar os resultados do mapeamento como guia para quem deseja analisar e prever crimes cibernéticos.

O conceito de crimes cibernéticos ou de criminalidade cibernética poderá variar no contexto de outras pesquisas, pois que podem ser considerados crimes cibernéticos puros ou próprios (WENDT; JORGE, 2021) aqueles em que a questão tecnológica, a inserção de conceitos da Tecnologia da Informação e Comunicação e/ou da Internet, já

foi feita pelo legislador, a exemplo dos arts. 154-A[1] (BRASIL, 2012; 2021), 218-C[2] (BRASIL, 2018) e 171, §2-A[3] (BRASIL, 2021) do Código Penal. Já os crimes cibernéticos impuros ou impróprios são aqueles em que meio tecnológico e/ou a Internet é apenas utilizado para a prática de delitos, sem prever essa circunstância especificamente em seu texto normativo.

Nesse passo, o artigo parte de uma análise da fundamentação teórica, baseada no método de Bailey et al. (2007), utilizada para conduzir o mapeamento sistemático, que propicia o uso de aprendizado de máquina e de dados heterogêneos como mecanismo para a predição de crimes no âmbito cibernético, entendido este, especialmente para este estudo, como a rede mundial de computadores, a Internet.

---

[1] Invasão de dispositivo informático

Art. 154-A. Invadir dispositivo informático de uso alheio, conectado ou não à rede de computadores, com o fim de obter, adulterar ou destruir dados ou informações sem autorização expressa ou tácita do usuário do dispositivo ou de instalar vulnerabilidades para obter vantagem ilícita:

Pena – reclusão, de 1 (um) a 4 (quatro) anos, e multa.

§ 1º Na mesma pena incorre quem produz, oferece, distribui, vende ou difunde dispositivo ou programa de computador com o intuito de permitir a prática da conduta definida no caput.

§ 2º Aumenta-se a pena de 1/3 (um terço) a 2/3 (dois terços) se da invasão resulta prejuízo econômico.

§ 3º Se da invasão resultar a obtenção de conteúdo de comunicações eletrônicas privadas, segredos comerciais ou industriais, informações sigilosas, assim definidas em lei, ou o controle remoto não autorizado do dispositivo invadido:

Pena – reclusão, de 2 (dois) a 5 (cinco) anos, e multa.

§ 4º Na hipótese do § 3o, aumenta-se a pena de um a dois terços se houver divulgação, comercialização ou transmissão a terceiro, a qualquer título, dos dados ou informações obtidas.

§ 5º Aumenta-se a pena de um terço à metade se o crime for praticado contra:

I - Presidente da República, governadores e prefeitos;

II - Presidente do Supremo Tribunal Federal;

III - Presidente da Câmara dos Deputados, do Senado Federal, de Assembleia Legislativa de Estado, da Câmara Legislativa do Distrito Federal ou de Câmara Municipal; ou

IV - dirigente máximo da administração direta e indireta federal, estadual, municipal ou do Distrito Federal.

[2] Divulgação de cena de estupro ou de cena de estupro de vulnerável, de cena de sexo ou de pornografia

Art. 218-C. Oferecer, trocar, disponibilizar, transmitir, vender ou expor à venda, distribuir, publicar ou divulgar, por qualquer meio - inclusive por meio de comunicação de massa ou sistema de informática ou telemática -, fotografia, vídeo ou outro registro audiovisual que contenha cena de estupro ou de estupro de vulnerável ou que faça apologia ou induza a sua prática, ou, sem o consentimento da vítima, cena de sexo, nudez ou pornografia:

Pena - reclusão, de 1 (um) a 5 (cinco) anos, se o fato não constitui crime mais grave.

Aumento de pena

§ 1º A pena é aumentada de 1/3 (um terço) a 2/3 (dois terços) se o crime é praticado por agente que mantém ou tenha mantido relação íntima de afeto com a vítima ou com o fim de vingança ou humilhação.

Exclusão de ilicitude

§ 2º Não há crime quando o agente pratica as condutas descritas no caput deste artigo em publicação de natureza jornalística, científica, cultural ou acadêmica com a adoção de recurso que impossibilite a identificação da vítima, ressalvada sua prévia autorização, caso seja maior de 18 (dezoito) anos.

[3] Fraude eletrônica

§ 2º-A. A pena é de reclusão, de 4 (quatro) a 8 (oito) anos, e multa, se a fraude é cometida com a utilização de informações fornecidas pela vítima ou por terceiro induzido a erro por meio de redes sociais, contatos telefônicos ou envio de correio eletrônico fraudulento, ou por qualquer outro meio fraudulento análogo.

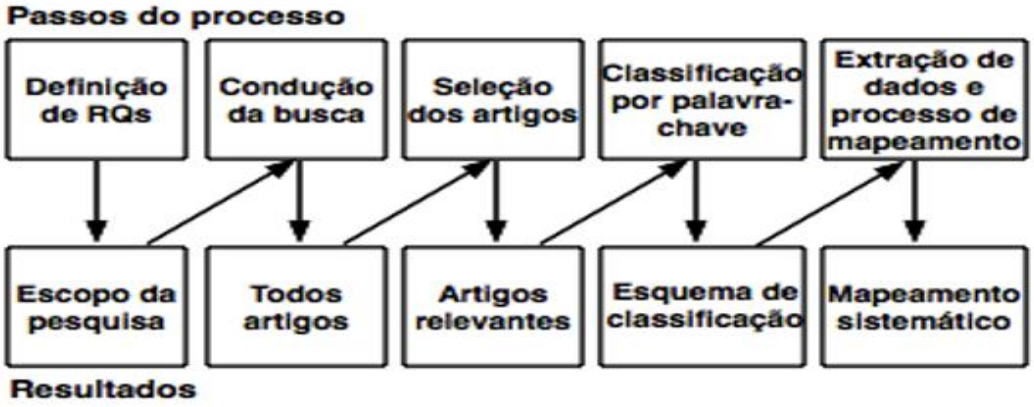
§ 2º-B. A pena prevista no § 2º-A deste artigo, considerada a relevância do resultado gravoso, aumenta-se de 1/3 (um terço) a 2/3 (dois terços), se o crime é praticado mediante a utilização de servidor mantido fora do território nacional.

No segundo momento, partir-se-á, com a metodologia de pesquisa empírica, para seleção de 51 (cinquenta e um) trabalhos e observação de 14 (quatorze) trabalhos selecionados e relacionados à análise e previsão de crimes com o uso de aprendizado de máquina e de dados heterogêneos.

2 SELEÇÃO DO CONTEÚDO DE PESQUISA, CATEGORIZAÇÃO E DEFINIÇÃO DO OBJETO DE ANÁLISE: OS ESTUDOS SOBRE A CIBERCRIMINALIDADE

Para fins de realização desta pesquisa, utiliza-se do método de Bailey *et al.* (2007), visando a conduzir o mapeamento sistemático, cujos passos estão apresentados em seguida.

Figura 1: Processo do mapeamento sistemático



Fonte: Bailey *et al.* (2007, n.p.).

É importante explicar a etapa de classificação por palavras-chave porque é durante essa etapa que o mapeamento é realizado. Ela, de acordo com Bailey *et al.* (2007), consiste em outras duas etapas: a primeira, que contempla atividades referentes a análise do resumo dos artigos escolhidos para extrair daí um esquema de classificação dos trabalhos, permitindo o início dos agrupamentos. Em seguida, todos os artigos são classificados dentro desse esquema, que é atualizado à medida que o mapa é construído.

O resultado após conduzir todas as etapas do método de mapeamento sistemático proposto por Bailey *et al.* (2007) é composto por 51 trabalhos, eis que atendem ao objetivo desta pesquisa, que inclui o levantamento bibliográfico a fim de obter um panorama de quais técnicas e dados estão sendo utilizados para realizar a análise e previsão de crimes

cibernéticos. Com base neste objetivo, foi definida a questão principal [QP][4] deste mapeamento, que contempla a pergunta referente à quais técnicas de aprendizado de máquina estão sendo usadas para realizar análise e previsão de crimes cibernéticos.

Para ajudar a responder a esse [QP], foi definida a seguinte questão específica [QE], que contempla a pergunta referente a quais dados são usados para análise e previsão de crimes cibernéticos.

Foram pesquisados, em julho de 2022, trabalhos elaborados em português, inglês, francês, espanhol e italiano, referente ao período de junho de 2020 até julho de 2022 nas bases de dados que abrangem artigos acadêmicos, teses de doutorado, dissertações de mestrado e trabalhos de conclusão de curso. Sendo utilizadas nesse processo as plataformas: "Capes"[5], "IEEE"[6], "Mackenzie"[7], "Google Scholar"[8], "arXiv.org"[9] e "Science Direct"[10].

Para realizar a busca automática nas bases de dados selecionadas, foram empregadas as seguintes palavras-chave nas buscas (tanto em inglês quanto em português): "cyber crimes", "crimes cibernéticos", "crimes informáticos", "crimes tecnológicos", "inteligência artificial", “artificial inteligence”, "machine learning", “aprendizado de máquina”, "deep learning" e "aprendizado profundo".

Quanto às *strings*, foram utilizadas para conduzir as buscas as compostas pelas palavras “(cyber crimes OU crimes cibernéticos OU crimes informáticos OU crimes tecnológicos) E (inteligência artificial OU artificial inteligence OU machine learning OU aprendizagem de máquina OU deep learning OU aprendizagem profunda)”.

Quanto à metodologia utilizada para realizar a pesquisa, ela seguiu 4 etapas:

1. Exclusão por título,
2. Exclusão por resumo,
3. Eliminação por diagonal, e,
4. Exclusão de leitura completa.

Quanto à exclusão, foram adotadas as seguintes premissas:

1. Trabalhos que não estão em português, inglês, francês, espanhol ou italiano,

[4] Na imagem, RQ significa Research Question.
[5] Capes: https://www-periodicos-capes-gov-br.ezl.periodicos.capes.gov.br/index.php/buscador-primo.html.
[6] IEEE: https://www.ieee.org/publications/index.html.
[7] Mackenzie: https://www.mackenzie.br/noticias/artigo/n/a/i/bibliotecas-do-mackenzie-10-perguntas-para-voce-saber-tudo-que-precisa.
[8] Google Acadêmico: https://scholar.google.com.br/.
[9] ArXiv: https://arxiv.org/.
[10] Science Direct: https://www.sciencedirect.com/.

2. Trabalhos que não informam a base de dados utilizada, e

3. Trabalhos que não usam banco de dados real.

Durante a realização da etapa 4, com a leitura completa dos trabalhos, foram organizados critérios de qualidade para definir se o trabalho encontrado atende ao objetivo do mapeamento. Cada critério foi avaliado como “SIM”, “PARCIALMENTE” ou “NÃO” e pontuado, respectivamente, com os valores 1, 0,5 ou 0. Foram eliminados os trabalhos que atingiram nota inferior a 60%. Os critérios de qualidade criados para este trabalho foram:

1. O trabalho está bem escrito e contém uma quantidade considerável de informações?

2. O trabalho possui uma amostra de dados significativa/relevante para a pesquisa?

3. Foi usado um método estatístico para avaliar o modelo? e,

4. Os resultados apontam para informações relevantes para o estudo?

A Tabela 1 contém o número de trabalhos inicialmente encontrados e quantos foram mantidos em cada uma das fases de seleção.

Tabela 1: Trabalhos encontrados por fase

Plataformas	**Seleção**	**Fase 1**	**Fase 2**	**Fase 3**	**Fase 4**
Google Scholar	23	11	11	11	11
Capes *Journal*	14	0	0	0	0
IEEE	9	1	1	1	1
Mackenzie	0	0	0	0	0
arXiv.org	5	2	2	2	2
Science Direct	0	0	0	0	0

Fonte: Produzido pelos autores a partir da pesquisa realizada (2022)

Assim, passando por todas as etapas, restaram selecionados 14 (quatorze) trabalhos acadêmicos, úteis à análise pretendida neste estudo.

3 METODOLOGIAS DE ANÁLISE DOS DADOS CRIMINAIS: APRENDIZAGEM ALGORÍTMICA E PREDIÇÃO DE CRIMES CIBERNÉTICOS

Durante o mapeamento sistemático, foram selecionados 14 (quatorze) trabalhos relacionados à análise e previsão de crimes com o uso de aprendizado de máquina e de dados heterogêneos.

Costa (2020) apresentou uma proposta de utilização da abordagem de aprendizagem não supervisionada com o uso do algoritmo *K-means* a fim de estabelecer agrupamentos para apoiar os formuladores de políticas de segurança pública nos municípios do Estado de Pernambuco no Brasil.

A decisão de utilizar o algoritmo de agrupamento deve-se ao fato dele trabalhar com a formação de agrupamentos, obtendo homogeneidade dentro dos grupos formados, de modo que os componentes (no caso, os municípios) do grupo compartilhem características criminosas comuns que os diferencia dos municípios de outros agrupamentos.

Com base nesses fatos, ao agrupamento como ferramenta de auxílio à decisão de direcionamento das ações de segurança pública, proporciona uma melhor visão das regiões que apresentam maior necessidade de atenção no cenário do Estado de Pernambuco. Os dados utilizados são os disponibilizados pela Secretaria Nacional de Segurança Pública – SENASP[11], por meio da Lei de Acesso à Informação – LAI[12], e eles incluem as ocorrências criminais de 2018. O resultado apresentou a existência de práticas difusas de crimes entre as cidades e diante disso, a partir da obtenção dos agrupamentos, foi possível chegar a grupos com ocorrências basicamente semelhantes. Costa (2020) destacou que esses dados apresentam lacunas quando observados em um horizonte de tempo mais longo.

Joner (2020) apresenta o uso de técnicas de aprendizado de máquina com o uso dos algoritmos de Redes Neurais com modelos de Memória de Curto Prazo Longo (em inglês LSTM[13] - *Long Short Term Memory*) e redes neurais profundas[14] (em inglês DNN – Deep Neural Networks), Regressão e Classificação[15] para o problema de previsão de crimes a fim de servir de subsídio para a solução do problema de alocação de força

[11] Secretaria Nacional de Segurança Pública - SENASP, ver mais em: https://www.gov.br/mj/pt-br/assuntos/sua-seguranca/seguranca-publica/a-senasp.
[12] Lei de Acesso à Informação Estado de Pernambuco, ver mais em: https://www.lai.pe.gov.br/sds/.
[13] De acordo com Passos (2021), as redes neurais do tipo Long Short Term Memory (memória de curto e longo prazo) são um tipo especial de rede neural recorrente e capazes de aprender conexões em **sequências de informação**. Dentre as principais aplicações das redes neurais Long Short Term Memory destacam-se: processamento de tarefas de **linguagem natural**; processamento de **áudio**; e processamento de uma sequência de frames de **vídeo**.
[14] Pina *et. al* (2020) abordam o tema das redes neurais profundas.
[15] Para ampliar o tema da regressão e classificação, vide Escovedo e Koshiyama (2020).

policial. A base de dados utilizada pelo autor é a fornecida pelo Observatório de Segurança Pública (OSP) pertencente ao Gabinete Estadual de Gestão Integrada (GGI-E) do Governo do Estado do Rio Grande do Sul - Brasil. Ressalta-se que o banco de dados é atualizado a partir de 2017 e é fornecido por órgãos de segurança pública, como Polícia Civil (PC), Brigada Militar (BM), Instituto Geral de Perícia (IGP) e Superintendência de Serviços Penitenciários (SUSEPE). (JONER, 2020, p. 11).

Castro (2020) contemplou em seu trabalho o uso do aprendizado supervisionado por meio da exploração de fontes heterogêneas de dados para entender padrões de comportamento criminoso e prever a ocorrência de crimes por tipo e por região geográfica. Evidências de fontes de dados foram selecionadas e combinadas com o objetivo de prever a tendência e o número de ocorrências de tipos de crimes por regiões geográficas.

Para realizar as previsões, foram utilizadas cinco técnicas de aprendizado de máquina, são elas: k-Nearest Neighbor (k-NN)[16], Support Vector Machine (SVM)[17], Random Forest (RF)[18], eXtreme Gradient Boosting (XGBoost)[19] e a Rede Neural Long Short Term Memory (LSTM)[20]. Conforme menciona o autor, os dados utilizados para a realização da pesquisa são provenientes de fontes heterogêneas, contendo registros criminais oficiais e não oficiais e eles incluem um conjunto de antecedentes criminais oficiais coletados junto à Secretaria de Segurança do Estado de Minas de Gerais[21] e um conjunto de dados não oficiais coletados do site "Onde fui roubado"[22] (DE CASTRO, 2020, p. 55).

Prado (2020), por sua vez, em sua obra analisou o uso do aprendizado não supervisionado através do uso do algoritmo Apriori[23] . Para realizar as análises, foram utilizadas 3.443.750 transações para gerar regras de associação. As transações foram separadas por ano, resultando em cinco subgrupos (2015, 2016, 2017, 2018 e 2019). Para ajudar a identificar os pontos fortes das regras de associação geradas pelo experimento, foram utilizadas as medidas de interesse: Suporte, Confiança, Elevação, r (Coeficiente de Correlação de Pearson) e Qui-Quadrado, com seu nível de significância (p-valor). O autor

[16] Ver mais sobre a técnica em Ileoh (2018).
[17] Ver mais sobre modelo de predição SVM em Coutinho (2019).
[18] Ver mais sobre *random forest* em Oshiro (2013).
[19] Ver mais sobre XGBoost em Gomes (2019).
[20] Ver Passos (2021).
[21] Sobre Secretaria de Segurança Pública de Minas Gerais, ver: http://www.seguranca.mg.gov.br/.
[22] Ver: http://www.ondefuiroubado.com.br.
[23] Vide Santos (2018).

(PRADO, 2020, p. 121) destaca que o conjunto de dados utilizado na sua pesquisa inclui as ocorrências criminais que ocorreram em todos os estados do Brasil, podendo seus dados serem consultados no site do Projeto Transparência Traduzida[24].

Já os autores Assad e Chagas (2019), contemplaram, em seus estudos, o estilo do aprendizado supervisionado através do uso de árvore de decisão[25] e algoritmos de regressão logística[26]. Os modelos foram gerados utilizando a ferramenta *Open Source Orange Canvas*, que permite a criação de todo o *workflow* de um projeto de aprendizado de máquina por meio de uma interface gráfica.

Conforme explicam os autores (ASSAD; CHAGAS, 2019, p. 22), os dados utilizados para a realização de sua pesquisa, foram extraídos do site da Secretaria de Segurança Pública (SSP) do Estado de São Paulo[27] em março de 2019 para a ano de 2018. Entre os 54 atributos associados a cada crime, destacam-se: número do boletim, data de emissão, período do crime, flagrante, cidade, estado, latitude, longitude, descrição do local, nome da delegacia, sexo, idade, número do estado, cor do veículo, cidade do veículo, ano de fabricação do veículo, ano do modelo do veículo, tipo de veículo, número de telefones celulares, marca do celular, bairro e rua.

Os demais atributos são chaves, diretamente relacionados à identificação do registro criminal, ou pessoa, ou são o atributo com grande falta de informação. Portanto, poucos atributos foram utilizados para realizar os experimentos, aqueles mais relevantes para os bons resultados dos modelos, foram definidos empiricamente. Assad e Chagas (2019) destacam que o modelo apresentava algumas limitações, sendo a principal a disponibilidade dos dados originais e a ausência de algumas informações coletadas nos boletins criminais, como, por exemplo, a idade e o sexo da vítima - estas são importantes para previsões mais precisas e permite a caracterização de novos padrões criminais. Uma das discrepâncias destacadas pelo autor refere-se ao fato de o furto de celular estar sempre presente no mesmo boletim de ocorrência de furto de carro ou furto de carro, quando esses crimes acontecem simultaneamente, impossibilitam a distinção entre um e outro e geram confusão na previsão de modelos.

[24] Ver: www.transparenciatraduzida.ufs.br.
[25] Ver mais sobre árvore de decisão em Casali (2021).
[26] Ver mais sobre regressão logística em Fernandes *et. al* (2021)
[27] Ver: https://www.ssp.sp.gov.br/.

Souza (2018) utiliza modelos de aprendizado supervisionado através do uso de algoritmos de Árvore de Decisão[28], Classificação Gaussiana Naive Bayes[29] e K-NN K-Nearest Neighbor[30] . Ele acredita que eles podem ser usados para implementar uma solução de análise e previsão de fatos usando dados do cenário de segurança pública do estado do Rio de Janeiro - Brasil. O autor (2018, p. 42) informa que o conjunto de dados utilizado para a realização do trabalho inclui os dados criminais fornecidos pela página de Dados Abertos do Instituto de Segurança Pública do Estado do Rio de Janeiro[31]. De acordo com o site[32], é possível acessar as bases de dados de antecedentes criminais e atividade policial no estado do Rio de Janeiro. (SOUZA, p. 42).

As estatísticas publicadas são baseadas nos Registros de Ocorrência (RO) coletados pelas Delegacias da Polícia Civil de todo o estado, esses registros de ocorrência são complementados por informações de órgãos específicos da Polícia Militar do estado do Rio de Janeiro. Antes de serem consolidadas no Instituto de Segurança Pública (ISP), as informações e registros de ocorrências passam por controle de qualidade pela Corregedoria da Polícia Civil (COINPOL).

As estatísticas produzidas são baseadas na data em que o Registro de Ocorrências foi criado. Souza (2018) destacou o fato de que é muito difícil encontrar um conjunto de dados de acesso livre com informações atualizadas no Brasil.

Ratul e Rab (2020) em seu trabalho, utilizam dados (478.578 incidentes) de crimes e acidentes na cidade de Denver - EUA, ocorridos de janeiro de 2014 a maio de 2019. Também foram utilizados algoritmos de classificação, como *random forest*, árvore de decisão, classificador AdaBoost, classificador de árvore extra, análise discriminante linear, classificadores K-Neighbors e modelos ensemble[33] para classificar diferentes classes de crimes.

Wang *et al.* (2017) utilizaram dados de crimes cometidos em Los Angeles – EUA de um determinado período de 6 meses e a adaptação do preditor espaço-temporal de *deep learning*, e ST-ResNet [34], para prever coletivamente a distribuição de crimes na área de Los Angeles Angeles – EUA.

[28] Ver mais sobre árvore de decisão em Casali (2021).
[29] Ver mais sobre a metodologia Naive Bayes em Gruendemann *et. al* (2019).
[30] Ver mais sobre a técnica em Ileoh (2018).
[31] Disponível em: http://dados.gov.br/.
[32] Disponível em: http://www.ispdados.rj.gov.br/.
[33] Ver mais sobre modelos assemble em De Castro, De Pádua e Andrade (2005).
[34] Ver sobre em Zhang, Zheng e Qi (2017).

Wang (2021) apresentou a relação entre crimes de ódio e alguns fatores, incluindo desigualdade de renda e renda familiar média, usando aprendizado de máquina (regressão linear simples, regressão linear múltipla, algoritmos K-means[35] e K-Nearest Neighbors) e os dados de um conjunto de dados de taxas de crimes de ódio nos EUA em 2016 que ocorreram antes e depois da eleição presidencial, bem como taxas de crimes de ódio em todos os estados dos EUA de 2010 a 2015.

Safat, Asghar e Gillani (2021) usaram dados criminais de Chicago e Los Angeles-EUA e como algoritmos regressão logística, SVM, Naïve Bayes, KNN, arvore de decisão, MLP, Random Forest, XGBoost e LSTM para classificar, processar e prever tempestades em série com intervalos de tempo de duração desconhecida. De acordo com os autores (2021, p. 70092), "a análise de dados exploratórios mostrou visualizações extensas de detalhes do crime, incluindo taxas de crimes em diferentes períodos de tendências diários a anuais, tipos de crimes e áreas de alta intensidade com base em padrões históricos".

Stalidis, Semertzidis e Daras (2018) apresentaram a ideia de prever tipos de crimes mais prováveis de serem cometidos em áreas abertas e em ambientes urbanos. Para que isso fosse possível, eles alimentaram os métodos de *deep learning* com informações históricas espaciais, temporais e do tipo de crime. Usaram dados abertos oriundos de relatórios policiais dos departamentos de polícia das cidades de Seattle, Minneapolis, Philadelphia, San Francisco, e da Metropolitan DC - EUA. Trata-se de cinco *datasets* que podem ser encontrados no *Kaggle*. Foram comparados os resultados dos algoritmos arvore de decisão, Naive Bayes, Logit Boost, SVM, Random Forests, KNN–MLP, ST-ResNet, etc.

Já o objetivo dos autores Lopes e Felix (2019) é identificar os determinantes econômicos, sociais e demográficos dos crimes de homicídios no Brasil, além de realizar a predição do nível de crimes em território nacional. O trabalho contempla o estudo quantitativo por meio da adoção de métodos de árvore de regressão, *random forest*, e K-Nearest Neighbors (KNN), e dos dados sobre categorias de renda, educação e demográficas obtidos por meio do IPEADATA e IBGE (LOPES; FELIX, 2019, p. 274).

Lopes e Felix (2019) acreditam que tal abordagem ajuda a produzir respostas mais robustas sobre os efeitos dos fatores sociais, econômicos e demográficos sobre o crime. Seria, portanto, uma nova ferramenta para orientar os formuladores de políticas públicas.

[35] Ver mais sobre agrupamento k-means em Miguel (20??).

Stec e Klabjan (2018) utilizaram redes neurais profundas para fazer previsões de contagem de crimes no dia seguinte de suas ocorrências. O *dataset* utilizado é composto dos dados de crimes de Chicago e Portland - EUA, somados aos dados referentes ao clima, dados do censo e transporte público.

Eles organizaram as contagens dos crimes dividindo-as em 10 partes e o modelo por eles utilizado prevê diariamente a parte mais provável de suas ocorrências. Para que isso fosse possível, eles treinaram os dados usando estruturas de rede neural, incluindo variações que são adequadas aos aspectos espaciais e temporais do problema de previsão de crime.

Os autores Braz *et al.* (2009) argumentam que, na área de segurança pública, a mineração de dados pode ser utilizada para (a) determinar os locais com maior criminalidade, (b) definir os perfis de vítimas e criminosos, (c) identificar a existência de quadrilhas e *serial killers,* (d) detectar quais dias da semana ocorrem mais delitos bem como (e) as suas causas etc. Eles citam um dos exemplos de utilização de mineração de dados no setor de segurança pública: o apoio ao planejamento estratégico adotado pelo Departamento de Polícia de Richmond – EUA. Neste caso foi utilizado o emprego da tarefa de agrupamento, utilizando o algoritmo Simple K-means[36] em relação as condutas dos criminosos e das vítimas, procurando potencializar indicações de diretrizes para policiamento em termos de distribuição e das razões da ocorrência policial. Foram obtidos resultados relevantes, gerando justamente grupos de características comuns de condutas criminosas e de situações de risco das vítimas Braz *et al*. (2009).

Vital *et al* (2020) explicam no trabalho que desenvolveram que a violência é um dos maiores desafios da sociedade brasileira. O número de homicídios no país vinha crescendo de forma heterogênea em todo o seu território desde a década de 1980, e, recentes estudos analisaram os impactos das políticas públicas sobre o crime e as condições econômicas e sociais podem ser muito mais importantes para explicá-lo.

Baseados nessas informações, os autores (VITAL *et al*, 2020) elaboraram este trabalho, cujo objetivo foi investigar as características que mais contribuem para caracterizar o crime no Brasil por meio da adoção de abordagem utilizando árvores de decisão. Com a sua aplicação, é possível analisar uma quantidade significativa de variáveis que podem impactar o crime. Como resultados do trabalho, destacam-se os principais achados, que mostram que as condições de moradia, a densidade demográfica,

[36] Ver mais sobre *simple k-means* em Xie e Jiang (2010).

a falta de religião e o número de homicídios nas regiões vizinhas são significativos para explicar a taxa de homicídios em 2016 nas cidades brasileiras.

Rayhan e Hashem (2020), por sua vez, enfatizam que devido aos impactos adversos que os crimes podem ter na vida humana, na economia e na segurança, é necessário utilizar um modelo que possa prever a ocorrência futura de crimes com a maior precisão possível, para que seja possível tomar medidas a fim de evitá-los. Destaca-se também a necessidade de utilizar um modelo interpretável, que revele a razão por trás da previsão de um modelo, garantindo sua transparência e permitindo planejar as etapas de prevenção ao crime. Nesse trabalho foram utilizados dados referentes aos crimes que ocorrem em Chicago – EUA e o deep learning para conduzir os estudos baseados em dados históricos.

Kshatri *et al.* (2021), por meio da utilização do aprendizado de máquina e do comitê de máquinas, conduziram previsão de crimes na Índia. Referindo ainda, que nesse país, é um problema desafiador identificar a natureza dinâmica dos crimes (KSHATRI *et al.*, 2021). O *dataset* foi elaborado baseado nos dados criminais de 2001 a 2015, oriundos dos registros de crimes do *National Crime Record Bureau* (NCRB) de todos os estados da Índia referentes aos relatórios factuais sobre assassinato, estupro e roubo (crimes violentos). Cerca de 60.000 crimes ocorreram nesse período.

O autor Vaquero Barnadas (2016) mostrou como resolver um problema real de classificação de dados usando diferentes algoritmos, como o *K-Nearest Neighbours*, e *neural networks*. O utilizado *dataset* contém todos os crimes classificados em categorias de diferentes tipologias de crime de San Francisco – EUA. O principal objetivo do desafio, segundo Vaquero Barnadas (2016), foi prever as categorias de crimes ocorridos. O resultado apresentado pelo autor mostra que, com o problema de classificação de crimes, o algoritmo mais preciso era a rede neural artificial.

Embora seja verdade que a ANN foi melhor do que o KNN, ela não o superou. Isso pode ter sido um problema de design de modelo, seleção de recursos inadequada, treinamento ou ajuste inadequado ou uma combinação de todos eles. Provavelmente, adicionando-se mais camadas à rede ou estendendo o processo de treinamento, os resultados poderiam ter sido melhores.

Segundo O'Neil (2021), para que fosse possível manter ou melhorar o policiamento com força reduzida, o comandante de polícia de Reading – EUA investiu,

em 2013, em um software da empresa Predpol[37] (*startup* de *big data* cuja sede era em Santa Cruz na Califórnia), cuja finalidade era prever crimes. O *modus operandi* do Predpol se baseia no processamento de dados históricos de crimes e com a frequência de hora em hora ele calculava a probabilidade da ocorrência do local do crime (O'NEIL, 2021, p.134). Por meio da utilização do Predpol, os policiais podiam ver em quadrantes (cujas medidas equivaliam a dois campos de futebol) as conclusões do programa. Quanto mais tempo passassem patrulhando esses quadrantes, aumentavam as chances de desencorajar os criminosos. Após um ano, isso resultou na queda do percentual de assaltos em 23% explica O'Neil (2021).

Informa, ainda, O'Neil (2021, p. 135), que Nova York - EUA utiliza um sistema similar ao Predpol intitulado de *CompStat*[38]; e a Filadélfia – EUA utiliza o sistema *HunchLAb*[39]. Explica, a autora que os *inputs* principais do Predpol são: tipo e local do crime e quando ele ocorreu. Portanto, é possível destacar que a vantagem quanto à utilização de um software como o Predpol é que ele não vai à cata dos criminosos, ou seja, ele não se concentra no indivíduo, mas, sim, privilegia a localização geográfica (O'NEIL, 2021, p. 135).

A partir da localização geográfica e das informações que estiverem mais tempo nos quadrantes indicados pelo Predpol, será possível repelir ladrões e assaltantes. Portanto, a comunidade será beneficiada, conclui (O'NEIL, 2021, p.135).

Quanto à Europa, é possível mencionar que a Itália atua com a Polícia Preditiva para Segurança Urbana por meio da utilização do software *X-Law* (L'INTELLIGENZA, 2020). Com o software, é possível prever, a partir de inteligência artificial, os seguintes delitos: roubo, furto, furto de carteira, golpes dentre outros crimes predatórios. Segundo *X-Law.it*, o software *X-Law* é baseado na ideia de que

> crimes são cíclicos, possuem permanência e podem ser previstos se for possível definir uma lógica de previsão adequada e transferi-la para modelos de aprendizado de máquina que com base em determinadas informações sobre os fenômenos criminais e sobre as dinâmicas sociais nos contextos em que ocorrem, têm a tarefa de descodificar o desenho criminoso na origem dos crimes individuais, proporcionando a sua reconfiguração no tempo e no espaço, com o objetivo de criar a condição ideal para o reaproveitamento evita-os de forma mais eficaz do que o método tradicional. (L'INTELLIGENZA, 2020, n.p.)

37 Saiba mais em: https://www.predpol.com/.
38 Saiba mais em: https://www.compstat360.org/.
39 Saiba mais: https://teamupturn.gitbooks.io/predictive-policing/content/systems/hunchlab.html.

É possível, após a condução do mapeamento, confirmar que seus resultados respondem às questões principal (QP) e específica (QE) estruturadas nesse trabalho, ou seja, foram descritas as principais técnicas de aprendizado de máquina estão sendo usadas para realizar análise e previsão de crimes, que também podem ser aplicadas aos crimes cibernéticos. Aliás, também pontuaram os trabalhos quais os dados são usados para análise e previsão de crimes, o que merece um aprimoramento, dado a multiplicidade de bancos de dados, oficiais ou não oficiais, estruturados ou não estruturados.

4 CONSIDERAÇÕES FINAIS

Existe um amplo estudo e corpo de literatura que inclui análise e previsão de crimes com dados heterogêneos e inteligência artificial, com os mais variados métodos de análise algorítmica. O resultado do mapeamento mostra que os dados de crimes são oriundos dos meios digitais, utilizados pelos pesquisadores e encontram-se nos relatórios criminais, em boletins de ocorrência e em pesquisas feitas com a população. Não foram encontrados trabalhos que utilizem dados oriundos de decisões que transitaram e julgado (*res judicata*), tampouco estão disponíveis no Brasil dados organizados sobre os crimes cibernéticos, sendo o dado mais próximo de consolidação o mencionado pelo Fórum Brasileiro de Segurança Pública sobre o estelionato, mesmo assim, são dados parciais.

Quanto aos algoritmos usados, os trabalhos utilizam os de aprendizado de máquina, especificamente aprendizado supervisionado e não supervisionado e poucos com redes neurais. São as mais variadas metodologias utilizada nos 14 trabalhos analisados.

É possível acrescentar o fato de que uma ação é considerada crime quando ela for prevista como crime em lei, sendo então a conduta investigada e julgada, finalizando a persecução criminal quando o processo transita em julgado. Porém, ainda carece no Brasil sistema de segurança pública otimizado, padronizado e uniformizado em relação aos crimes cometidos no âmbito cibernético, ou seja, há que se acelerar o processo de análise desses dados na mesma velocidade com que a tecnologia avança, na mesma velocidade com que os crimes cibernéticos estão sendo registrados.

Verificou-se, então, que em razão de os dados quali-quantitativos contidos nos boletins de ocorrência relacionados a crimes cibernéticos não estarem organizados e disponíveis aos cidadãos e pesquisadores, não foram ainda objeto de estudos e, uma vez

tabulados, podem possibilitar a manipulação, estudo e análise deles por meio do uso de inteligência artificial, especificamente aprendizado de máquina.

Pontua-se, assim, a falta de dados sobre crimes cibernéticos no Brasil e, consequentemente, o quanto isso impossibilita a análise e possíveis prevenções desses crimes. Por isso, como trabalho futuro, propõe-se a utilização de dados reais de crimes cibernéticos punidos por juízes em todas as instâncias do Poder Judiciário brasileiro - sendo estes chamados de coisa julgada, ou seja, processos que já foram julgados (res judicata) e que foram punidos como crimes cibernéticos. A categorização ainda precisa ser definida, mas há a possibilidade de se trabalhar a pesquisa vindoura com crimes cibernéticos próprios e/ou impróprios.

Com esses dados em mãos, a proposta é fazer previsões e análises por meio do uso de aprendizado de máquina envolvendo algoritmos de aprendizado supervisionado e não supervisionado. Os resultados produzidos tendem a ajudar os especialistas a descobrir padrões de crimes cibernéticos e agir proativamente para reduzi-los.

REFERÊNCIAS

ASSAD, Felipe José Perpétuo; CHAGAS, Jorge Felipe Campos. **Análise preditiva de manchas criminais no estado de São Paulo**. Ph.D. thesis. Universidade Federal Fluminense 2019.

BAILEY, John *et al*. Evidências relacionadas ao design de software orientado a objetos: uma pesquisa. *In*: **Primeiro Simpósio Internacional de Engenharia e Medição de Software Empírico (ESEM 2007)**. IEEE, 2007. p. 482-484.

BRASIL. **Lei nº 12.737, de 30 de novembro de 2012**. Dispõe sobre a tipificação criminal de delitos informáticos; altera o Decreto-Lei nº 2.848, de 7 de dezembro de 1940 - Código Penal; e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12737.htm. Acesso em: 10 out. 2022.

BRASIL. **Lei nº 13.718, de 24 de setembro de 2018**. Altera o Decreto-Lei nº 2.848, de 7 de dezembro de 1940 (Código Penal), para tipificar os crimes de importunação sexual e de divulgação de cena de estupro, tornar pública incondicionada a natureza da ação penal dos crimes contra a liberdade sexual e dos crimes sexuais contra vulnerável, estabelecer causas de aumento de pena para esses crimes e definir como causas de aumento de pena o estupro coletivo e o estupro corretivo; e revoga dispositivo do Decreto-Lei nº 3.688, de 3 de outubro de 1941 (Lei das Contravenções Penais). Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13718.htm. Acesso em: 10 out. 2022.

BRASIL. **Lei nº 14.155, de 27 de maio de 2021b**. Altera o Decreto-Lei nº 2.848, de 7 de dezembro de 1940 (Código Penal), para tornar mais graves os crimes de violação de dispositivo informático, furto e estelionato cometidos de forma eletrônica ou pela internet; e o Decreto-Lei nº 3.689, de 3 de outubro de 1941 (Código de Processo Penal), para definir a competência em modalidades de estelionato. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2021/Lei/L14155.htm. Acesso em: 03 jun. 2021.

BRAZ, Lucas M.; FERREIRA, Rafael; DERMEVAL, Diego; VÉRAS, Douglas; LIMA, Marcilio; TIENGO, Willy. Aplicando mineração de dados para apoiar a tomada de decisão na segurança pública do estado de Alagoas. *In* **Workshop de Computação Aplicada em Governo Eletrônico**, Bento Goncalves, Rio Grande do Sul, Brasil. SBC Brazilian Computer Society, 2009. Pág. 1475-1488.

BUENO, Samira; LIMA, Renato Sérgio de Lima. Anuário Brasileiro de Segurança Pública 2022. **Fórum Brasileiro de Segurança Pública**, Ano 16, 2022. ISSN 1983-7364. Disponível em: https://forumseguranca.org.br/wp-content/uploads/2022/06/anuario-2022.pdf?v=5. Acesso em: 2 set. 2022.

CASALI, Rudiney. Árvore de decisão: como se aplica no aprendizado da máquina. **Digital House**, 19 mar. 2021. Disponível em: https://www.digitalhouse.com/br/blog/arvore-de-decisao/. Acesso em: 11 out. 2022.

COSTA, Jefferson Carlos de Oliveira Ribeiro. **Identificação de grupos de municípios Pernambucanos para recomendação de políticas de segurança pública utilizando uma técnica de clusterização.** Dissertação de Mestrado. Universidade Federal de Pernambuco. 2020.

COUTINHO, Bernardo. Modelos de Predição | SVM: Aprenda a criar seu primeiro algoritmo de classificação com SVM. **Turing Talks**, 28 jul. 2019. Disponível em: https://medium.com/turing-talks/turing-talks-12-classifica%C3%A7%C3%A3o-por-svm-f4598094a3f1. Acesso em: 11 out. 2022.

DE CASTRO, Ursula Rosa Monteiro. **Explorando aprendizagem supervisionada em dados heterogêneos para predição de crimes**. Ph.D. thesis. Pontifícia Universidade Católica de Minas Gerais. 2020.

DE CASTRO, Cristiano Leite; DE PÁDUA BRAGA, Antônio; ANDRADE, Alessandro Vivas. Aplicaçao de um Modelo Ensemble de Redes Neurais Artificiais para Previsão de Séries Temporais não Estacionárias. In: **XXV Congresso da Sociedade Brasileira de Computação**, São Leopoldo, RS. 2005.

ESCOVEDO, Tatiana; KOSHIYAMA, Adriano. **Introdução a Data Science: Algoritmos de Machine Learning e métodos de análise**. Casa do Código, 2020.

FERNANDES, Antônio Alves Tôrres et al. Leia este artigo se você quiser aprender regressão logística. **Revista de Sociologia e Política**, v. 28, 2021.

GOMES, Pedro César Tebaldi. Conheça o algoritmo XGBoost. **DataGeeks**, 3 jun. 2019. Disponível em:

https://www.datageeks.com.br/xgboost/#:~:text=O%20XGBoost%20%C3%A9%20um%20algoritmo,os%20outros%20algoritmos%20ou%20frameworks. Acesso em: 11 out. 2022.

GRUENDEMANN, Felipe Camargo et al. Equipamento eletromédico intravenoso: uma proposta para qualificação da geração de alertas utilizando Naive Bayes. In: **Anais da XIX Escola Regional de Alto Desempenho da Região Sul**. SBC, 2019.

ILEOH. O Algorítmo K-Nearest Neighbors (KNN) em Machine Learning. **Portal Data Science**, 12 dez. 2018. Disponível em: https://portaldatascience.com/o-algoritmo-k-nearest-neighbors-knn-em-machine-learning/. Acesso em: 10 out. 2022.

JONER, Henrique. **Inferência preditiva geoespacial da criminalidade em Porto Alegre**: uma abordagem de aprendizado de máquina. Ph.D. thesis. Faculdade de Ciências Econômicas da UFRGS. 2020.

KSHATRI, Sapna Singh, *et al.* An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach. **IEEE Access** v. 9, 2021. p. 67488–67500.

L'INTELLIGENZA Artificiale che ha rivoluzionato il paradigma della sicurezza urbana. **X LAW**. 2020. Disponível em: https://www.xlaw.it/presentazione/. Acesso em: 12 set. 2022.

LOPES, Lucas Pereira; FELIX, Sabrina Vieira. Determinantes e predição de crimes de homicídios no Brasil: uma abordagem de aprendizado de máquina. **Revista Brasileira de Biometria**, v. 37, n. 2, 2019. p. 272-289.

MIGUEL, Tussevana. K-Means Clusteering (Agrupamento k-means). Aprender Data Science, sem data. Disponível em: https://aprenderdatascience.com/k-means-clustering-agrupamento-k-means/. Acesso em: 11 out. 2022.

OLIVEIRA, Ingrid. Levantamento mostra que ataques cibernéticos no Brasil cresceram 94% 2022. 22/08/2022, 11h24min. **CNN Brasil**. Disponível em: https://www.cnnbrasil.com.br/tecnologia/levantamento-mostra-que-ataques-ciberneticos-no-brasil-cresceram-94/. Acesso em: 13 set. 2022.

OSHIRO, Thais Mayumi. **Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica**. 2013. Tese de Doutorado. Universidade de São Paulo.

O'NEIL, Cathy. **Algoritmos de destruição em massa**: como o big data aumenta a desigualdade e ameaça a democracia. Tradução Rafael Abraham. 1ª ed. Santo André, São Paulo: Editora Rua do Sabão, 2021.

PALMIERI, Michael; SHORTLAND, Neil; MCGARRY, Presley. Personality and online deviance: The role of reinforcement sensitivity theory in cybercrime. **Computers in Human Behavior.** 120, 106745. 2021. Disponível em: https: //www.sciencedirect.com/science/article/pii/S0747563221000674. Acesso em: 13 set. 2022. doi:https://doi.org/10.1016/j.chb.2021.106745.

PASSOS, Bianka Tallita. Long Short Term Memory: redes neurais artificiais que são capazes de ler, ouvir e ver. **Ateliware Blog**. 09 nov. 2021. Disponível em: https://ateliware.com/blog/long-short-term-memory. Acesso em: 10 out. 2022.

PINA, Débora; KUNSTMANN, Liliane; OLIVEIRA, Daniel de; VALDURIEZ, Patrick; MATTOS, Marta. Uma abordagem para coleta e análise de dados de configurações em redes neurais profundas. In: **SBBD 2020-35ª Simpósio Brasileiro de Banco de Dados**. 2020. p. 1-6.

PRADO, Kleber Henrique de Jesus. **Data Science Aplicada à Análise Criminal Baseada nos Dados Abertos Governamentais do Brasil**. Ph.D. thesis. Universidade Federal de Sergipe. 2020.

RATUL, Md; RAB, Aminur. A comparative study on crime in denver city based on machine learning and data mining. **ArXiv abs/2001.02802**. 2020.

RAYHAN, Yeasir; HASHEM, Tanzima. AIST: An interpretable attention-based deep learning model for crime prediction. **arXiv preprint arXiv:2012.08713**. 2020.

SAFAT, Wajiha; ASGHAR, Sohail; GILLANI, Saira Andleeb. Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques. **IEEE Access**, v. 9, p. 70080–70094. doi:10.1109/ACCESS.2021.3078117. 2021.

SANTOS, Vinicius dos. Como funciona o algoritmo Apriori. **Computer Science Master**, 10 out. 2018. Disponível em: https://www.computersciencemaster.com.br/como-funciona-o-algoritmo-apriori/. Acesso em: 10 out. 2022.

SOUZA, José Renato Mendes de. **Utilização de aprendizagem de máquina na predição de crime**. Ph.D. thesis. Universidade Federal Fluminense. 2018.

STALIDIS, Panagiotis; SEMERTZIDIS, Theodoros; DARAS, Petros. **Examining deep learning architectures for crime classification and prediction**. Available in: https://arxiv.org/abs/1812.00602, access at:13 set. 2022. doi:10.48550/ARXIV.1812.00602. 2018.

STEC, Alexandre; KLABJAN, Diego. **Forecasting crime with deep learning**. Available in: https://arxiv.org/abs/1806.01486, access at: 13 set. 2022. doi:10.48550/ARXIV.1806.01486. 2018.

VAQUERO BARNADAS, Miquel. **Machine learning applied to crime prediction**. B.S. thesis. Universitat Politècnica de Catalunya. 2016.

VITAL, Tauã Magalhães *et al*. **Features of crime in brazil**: an approach based on decision trees algorithms. 2020.

WANG, Bao et al. Deep learning for real-time crime forecasting and its ternarization. **Chinese Annals of Mathematics**, Series B, v. 40, n. 6, p. 949-966, 2017. doi:10.1007/s11401-019-0168-y.

WANG, Shaoxuan. Hate crime analysis based on artificial intelligence methods, *in*: **E3S Web of Conferences**, EDP Sciences. 2021. p. 01062.

WENDT, Emerson; JORGE, Higor Vinicius Nogueira. **Crimes Cibernéticos**: Ameaças e procedimentos de investigação. 3ª Edição. Rio de Janeiro: Brasport, 2021.

XIE, Juanying; JIANG, Shuai. A simple and fast algorithm for global K-means clustering. In: **2010 Second International Workshop on Education Technology and Computer Science**. IEEE, 2010. p. 36-40.

ZHANG, Junbo; ZHENG, Yu; QI, Dekang. Deep spatio-temporal residual networks for citywide crowd flows prediction. In: **Thirty-first AAAI conference on artificial intelligence**. 2017. Disponível em: https://dl.acm.org/doi/10.5555/3298239.3298479. Acesso em: 11 out. 2022.