

Risk assessment of revictimization in cases of domestic violence against women

Ricardo Vasconcelos¹, Carlos Caminha², Vasco Furtado³

¹ Universidade de Fortaleza, Tribunal de Justiça do Estado do Ceará, Brazil
ricardo.vasconcelos@tjce.jus.br

² Universidade Federal do Ceará - UFC, Brazil
caminha@ufc.br

³ Universidade de Fortaleza, Cientista-chefe do Tribunal de Justiça do Estado do Ceará, Brazil
vasco@unifor.br

Abstract. Risk assessment tools for domestic violence against women show limited predictive accuracy when grounded only in structured professional judgment. In Brazil, the National Risk Assessment Form (FONAR) lacks statistically grounded weighting, constraining its effectiveness. This study evaluates FONAR's predictive capacity and tests whether judicial-data enrichment improves revictimization risk estimation. Administrative and judicial records were integrated through probabilistic linkage, yielding 8,694 cases (1,429 revictimizations) from 10,069 FONAR forms collected between March 2022 and September 2024. Propensity score matching (1:1, nearest-neighbour with caliper on the logit) estimates an effect of protective measures on revictimization of 0.1246 (ATC; 95% CI [0.1089, 0.1412]) and 0.1226 (ATT; 95% CI [0.1064, 0.1399]), consistent with selection of higher-risk cases for treatment. Supervised models were tested under two configurations: FONAR-only and enriched. XGBoost performed best (AUC = 0.6139, recall = 0.8027), outperforming FONAR-only (AUC = 0.5709). Protective-measure indicators and prior judicial proceedings dominate predictions, acting as proxies for case severity rather than causal drivers. The study contributes an empirical assessment of FONAR, evidence that data integration improves risk modelling, and a pipeline to support judicial decision-making.

CCS Concepts: • **Applied computing** → **Law**.

Keywords: criminal justice, domestic violence, machine learning, revictimization, risk assessment

1. INTRODUÇÃO

O Painel da Violência Doméstica Contra Mulheres registra, de 2020 a 2025, aumento de 85% nos casos de violência doméstica e de 185% nos registros de feminicídio ¹. Em 2025, medidas de proteção de urgência responderam por 85% das decisões, com crescimento de 184% no período. Esses indicadores expõem deficiências das proteções institucionais e tornam urgente o desenvolvimento de políticas de mitigação. Em 2021, a Lei nº 14.149, instituiu o Formulário Nacional de Avaliação de Riscos (FONAR), questionário estruturado destinado a identificar fatores de recorrência e escalada da violência [Lisboa et al. 2019].

Este estudo investiga a vitimização em casos de violência doméstica contra mulheres a partir de dados retrospectivos, integrando o FONAR a registros do sistema judicial para caracterizar vítimas, agressores, históricos processuais e decisões de proteção. A pesquisa aborda duas questões: o efeito das medidas de proteção sobre a revitimização e o potencial das variáveis do FONAR, combinadas a dados contextuais, para apoiar estimativas de risco.

Metodologicamente, o estudo compreende três estágios: engenharia de dados, análise retrospectiva

¹Disponível em: <https://www.cnj.jus.br/programas-e-acoes/estatistica/>

e aprendizado de máquina. Identificam-se os casos iniciais de vitimização e comparam-se as taxas de revitimização com e sem medidas de proteção, empregando pareamento por escore de propensão (*propensity score matching*) para mitigar o viés de seleção. Em seguida, propõem-se algoritmos de aprendizado de máquina para estimar o risco de revitimização a partir do FONAR combinado a dados processuais.

Os algoritmos foram aplicados para apoiar a tomada de decisão judicial quanto à concessão de medidas protetivas, auxiliando magistrados na identificação de casos vulneráveis. Duas configurações foram testadas: uma com variáveis do FONAR isoladamente e outra com dados enriquecidos com histórico judicial. Os modelos enriquecidos superaram aqueles baseados exclusivamente no FONAR, indicando que o ganho preditivo decorre da incorporação de variáveis históricas que ampliam a capacidade discriminativa. Como contribuições, o trabalho oferece: (i) uma avaliação empírica da capacidade preditiva do FONAR com base integrada a registros judiciais; (ii) evidência quantitativa do ganho preditivo associado à integração de dados; e (iii) um pipeline reprodutível de aprendizado de máquina para apoio à decisão judicial.

2. ESTADO DA ARTE

O uso de ferramentas de avaliação de risco no sistema de justiça criminal tornou-se uma estratégia padrão para prevenir a violência doméstica e proteger as vítimas; essas ferramentas são amplamente utilizadas em países desenvolvidos para identificar vulnerabilidades e orientar medidas de proteção [Grogger et al. 2021] [da Silva 2021] [Lisboa et al. 2019].

Esses instrumentos operam com base em escalas derivadas de formulários, nos quais a classificação de risco resulta de um julgamento profissional estruturado. Embora promovam padronização, transparência e integração, apresentam limitações: interpretação humana, qualidade dos dados e capacidade preditiva limitada, particularmente em casos de alto risco. Também simplificam fenômenos complexos e podem perpetuar desigualdades estruturais. Como ferramentas de apoio, têm alcance limitado quando utilizadas como mecanismos centrais de tomada de decisão [Berk et al. 2021] [Turner et al. 2021] [Myhill et al. 2023].

Em contrapartida, modelos estatísticos e de aprendizado de máquina reduzem a subjetividade ao identificar padrões complexos e apresentam melhor desempenho em contextos não lineares ou de interatividade [Turner et al. 2021] [Berk et al. 2021]. Estudos mostram que a incorporação de variáveis criminais históricas a esses formulários gera ganhos substanciais de desempenho, destacando a importância da integração de dados [Turner et al. 2019] [Turner et al. 2021] [Grogger et al. 2021]. [Allard et al. 2024] analisaram o SARA (V2 e V3), constatando forte validade convergente, uma AUC de 0,65, consistência interna moderada e boa confiabilidade entre avaliadores. Uma meta-análise de 29 estudos revelou que instrumentos atuariais superam as ferramentas de julgamento estruturado, embora ambos apresentem precisão baixa a moderada (AUC = 0,64 para violência doméstica) [Ryland et al. 2025]. De forma semelhante, [Viljoen et al. 2025] revisou 31 estudos, demonstrando que instrumentos estruturados superam julgamentos não estruturados, com precisão variando de 67,8% a 71,4%. [Weber et al. 2025] analisaram 102 estudos, mostrando que a validade depende das taxas basais de reincidência e que confiar exclusivamente na AUC superestima o risco em cenários de declínio.

Quanto às evidências empíricas, [Caulfield et al. 2025] avaliou aproximadamente 25.000 casos do DASH (AUC=0,67 para letalidade; 0,62 para reincidência), com melhor desempenho na versão abreviada do protocolo. [Gray et al. 2025] encontrou AUC entre 0,69 e 0,71 para três instrumentos aplicados a 94 homens canadenses, com taxa de reincidência de 12,8%. A seleção de variáveis é crucial: [Júnior and Ribeiro 2019] destaca ganhos com dados históricos, enquanto [Pham et al. 2026] mostra que apenas um subconjunto de fatores é preditivo.

No Brasil, identificou-se uma lacuna na literatura quanto à integração de dados de questionários e registros judiciais com aprendizado de máquina no contexto da violência contra a mulher. Difer-

entamente de instrumentos clássicos como SARA, DASH e outros, centrados em itens estáticos, este estudo mostra que o risco de revitimização está fortemente associado ao histórico processual e à gravidade acumulada do agressor, evidenciando que agregar dados judiciais ao formulário amplia o poder preditivo. Este estudo avança sobre essa base ao combinar predição da revitimização, integração de bases e apoio à decisão judicial em um mesmo desenho.

3. METODOLOGIA

3.1 Fontes e caracterização de dados

O questionário estruturado FONAR captura informações sobre violência contra a mulher em três dimensões: vítima, agressor e contexto [Lisboa et al. 2019]. O FONAR é preenchido em distritos policiais e enviado ao Poder Judiciário com o processo, que pode conter ou não um pedido de medida protetiva. Nesta pesquisa, o conjunto de dados do FONAR foi enriquecido com registros judiciais. A vinculação probabilística por nome, filiação, data de nascimento e documento permitiu recuperar registros criminais, execuções e ocorrências relacionadas a medidas protetivas, feminicídio e violência doméstica envolvendo vítima ou agressor. A qualidade do processo foi avaliada em amostra de 119 formulários revisados manualmente (precisão de 100%). Adotou-se limiar conservador de 1,741 — o menor escore entre pares validados —, priorizando precisão e evitando associações indevidas entre históricos de indivíduos distintos, o que recuperou 121 processos na amostra. Foi também identificada a concessão (ou não) das medidas protetivas solicitadas, em janela de até seis meses após a ocorrência e antes da baixa processual. Ressalta-se que essa janela é utilizada exclusivamente para a construção da variável-resposta (revitimização).

Após as etapas de seleção, integração e preparo, o conjunto final, totalmente anonimizado, integra variáveis de seis dimensões sintetizadas na Tabela I.

Dimensão	Descrição	Nº	Tipo
Formulário FONAR	Q01–Q27 (1 = sim; 0 = não/desconhecida)	27	Binária
Vínculo vítima-agressor	Casamento, união estável, namoro, etc.	1	Categórica
Histórico do agressor	Procedimentos, processos, execuções, MP, violência doméstica, feminicídio	6	Binária
Histórico da vítima	Procedimentos, processos e concessões anteriores de MP	4	Binária
Idade	Vítima, agressor e diferença etária	3	Númerica
Total de covariáveis		41	
Variável-resposta	Revitimização (1 = novo episódio; 0 = ausência)	1	Binária

Table I. Composição do conjunto de variáveis.

A pesquisa seguiu diretrizes éticas nacionais para dados secundários, com aprovação do Comitê de Ética (parecer nº 7.023.476, 22/08/2024) e autorização do TJCE (Processo no 8514389-46.2023.8.06.0000) para uso exclusivamente científico. Identificadores foram suprimidos ou codificados; a base analisada não permite identificação das pessoas. Dados anonimizados disponíveis em: <https://tinyurl.com/35sn6wfr>.

3.2 Estruturação do conjunto de dados

A identificação de casos de revitimização seguiu o critério validado por magistrados do Tribunal de Justiça do Estado do Ceará: a primeira medida protetiva está, em regra, associada ao evento inicial de agressão; medidas posteriores - no mesmo ou em diferentes processos - indicam novos episódios ou agravamento. Caracterizam revitimização:

—registro de novas ocorrências em datas distintas após o primeiro atendimento;

- medida protetiva concedida até seis meses antes do registro da agressão, indicando histórico recente de violência;
- peticionamento de múltiplos processos ou procedimentos em até seis meses após a ocorrência;
- concessão de medidas protetivas adicionais após a primeira.

Foram consideradas 10.069 ocorrências do FONAR, registradas entre março de 2022 e setembro de 2024 em Fortaleza e Maracanaú. Após a integração, obteve-se uma base enriquecida com 8.694 registros, dos quais 1.429 correspondem a casos de revitimização. Essa base articula o contexto da agressão — capturado pelo formulário — ao histórico judicial das partes, ampliando a capacidade analítica e permitindo identificar, com maior precisão, os fatores associados à concessão de medidas protetivas e à recorrência da violência.

Do total integrado, 6.379 casos tiveram medida protetiva concedida (1.244 com revitimização; 5.135 sem). Nos 2.315 casos restantes — sem concessão, sem solicitação ou não identificados na integração —, 185 apresentaram revitimização. Identificaram-se ainda 165 casos com duas ou mais medidas no período e 64 com nova violação em até seis meses, caracterizando vitimização terciária.

O item 7b do FONAR (descumprimento prévio de medida) recebeu 922 respostas positivas, das quais 181 foram classificadas como revitimização. A integração identificou 185 concessões anteriores ao atendimento, das quais 159 sem indicação positiva no 7b. O descompasso evidencia erro de medida em ambas as fontes - autorrelato sujeito a imprecisões e registros administrativos com cobertura parcial - sugerindo que parte das respostas positivas refere-se a medidas concedidas há mais de seis meses ou a processos não recuperados na vinculação. As fontes são, portanto, complementares e imperfeitas.

3.3 Análise estatística e modelo preditivo

A análise compreendeu duas etapas: a estimação do efeito causal da medida protetiva sobre a revitimização, por pareamento via escore de propensão, e a construção de um modelo preditivo para apoio à decisão judicial.

O escore de propensão [Rosenbaum and Rubin 1983] foi estimado por regressão logística sobre as 41 covariáveis da Tabela I, após padronização. O pareamento adotou o algoritmo guloso 1:1 por vizinho mais próximo, sem reposição, no espaço do logit do escore, com *caliper* de 0,2 desvio-padrão do logit [Austin 2011]. Dois desenhos foram implementados sobre a população completa de tratados e controles: para estimar o efeito médio sobre os controles (ATC) — a revitimização esperada para os casos sem medida, caso a tivessem recebido —, e para estimar o efeito médio sobre os tratados (ATT) — a redução de revitimização atribuível à medida, entre os casos que de fato a receberam. O balanceamento das variáveis foi avaliado pela diferença média padronizada (SMD), com $|SMD| < 0,1$ como critério de equilíbrio [Stuart 2010]. Os efeitos pontuais foram calculados pelo *t*-teste emparelhado sobre os pares formados; os intervalos de confiança a 95%, por *subsampling* sem reposição com $m \approx n^{0,9}$ e 1.000 réplicas e fator de escala $\sqrt{m/n}$ procedimento consistente para estimadores de pareamento, em que o *bootstrap* padrão é inválido [Abadie and Imbens 2008].

Foram comparados quatro classificadores - XGBoost, Random Forest, Linear Discriminant Analysis e Logistic Regression - em duas configurações de variáveis: (i) conjunto completo (FONAR e histórico judicial) e (ii) somente FONAR. O desempenho foi avaliado por AUC sob validação cruzada estratificada em cinco *folds*, precedida de embaralhamento (*shuffle*), reportando-se a média e o intervalo de confiança a 95%. Os hiperparâmetros do XGBoost foram otimizados com Optuna², maximizando ROC-AUC. Para o melhor modelo, reportam-se ainda *recall* da classe positiva, G-mean e F1, além da análise SHAP [Lundberg and Lee 2017] para quantificar a contribuição de cada variável.

²<https://optuna.org/>

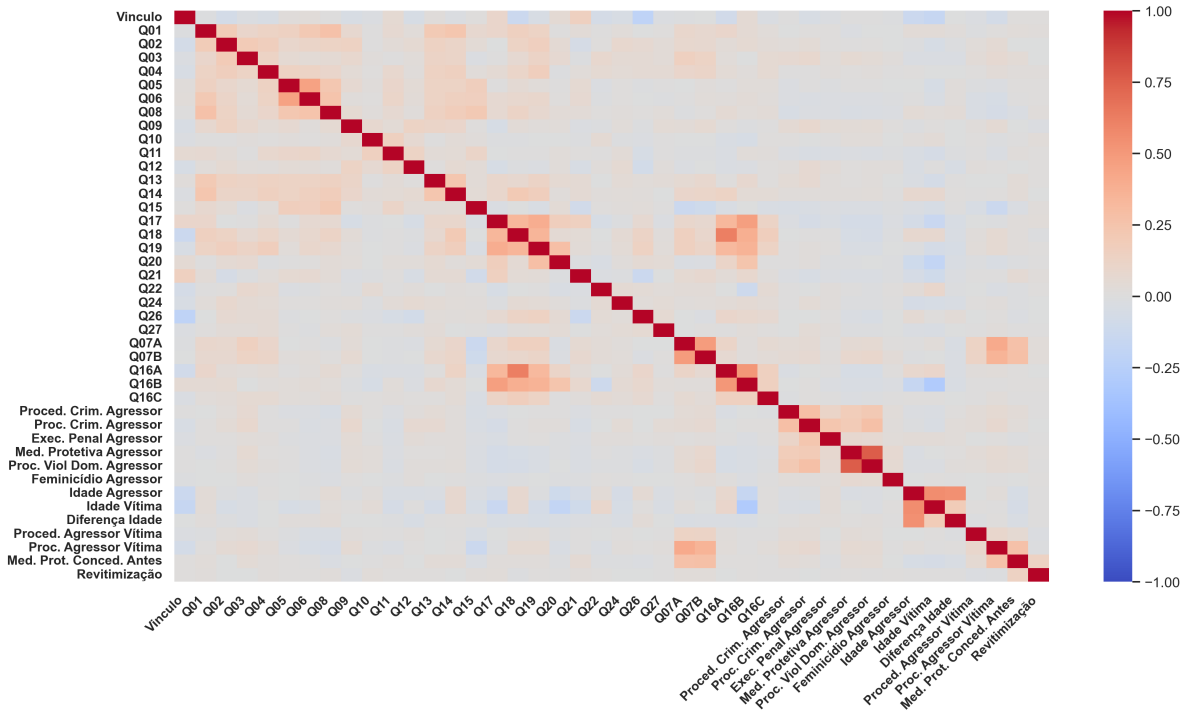


Fig. 1. Matriz de correlação

4. RESULTADOS

A análise etária mostra predominância de adultos jovens e de meia-idade: 59,17% das vítimas têm entre 25 e 50 anos e 53,87% dos agressores, entre 30 e 55. A revitimização concentra-se em relações com diferença etária entre 0 e 3 anos, reduzindo-se à medida que a diferença aumenta, o que sugere maior intensidade relacional em idades próximas.

Quanto ao vínculo afetivo, a violência ocorre majoritariamente em relações conjugais estáveis — união estável (24,63%) e casamento civil (15,97%) — e em vínculos recentemente rompidos. Entre os 8.694 casos integrados, "união estável" reúne o maior volume (2.124 ocorrências) com a menor taxa de revitimização (13,98%), enquanto "separados de fato há menos de 6 meses" combina o menor volume (771 casos) com a maior taxa (20,10%). O risco, portanto, não é proporcional à frequência: depende do contexto.

A matriz de correlação da Figura 1 confirma o caráter multifatorial do fenômeno. As correlações das variáveis com a revitimização são, isoladamente, fracas (máximo de 0,13, para medida protetiva concedida anteriormente), mas há núcleos relevantes entre os próprios preditores: processos de violência doméstica e medida protetiva contra o agressor (0,76); filhos e testemunho de violência (0,62); idades da vítima e do agressor (0,56); e descumprimento de medida e ocorrência anterior (0,49). Isso indica que a reincidência decorre do acúmulo de condições de risco correlacionadas, e não de um único fator isolado. Esse padrão não implica que múltiplas medidas elevem o risco: indica que as medidas tendem a ser aplicadas em casos já mais graves — efeito de seleção.

A taxa de revitimização (TR) é a proporção de casos com novo episódio de violência sobre o total do grupo: $TR = N_r / (N_r + N_{\bar{r}})$ em que N_r e $N_{\bar{r}}$ são, respectivamente, os casos com e sem revitimização. A TR bruta foi de 16,44% (1.429/8.694); entre os casos com medida, 19,50% (1.244/6.379); entre os sem medida, 7,99% (185/2.315). Essa comparação é enviesada, pois os casos com medida concentram maior gravidade.

Aplicando o procedimento descrito na metodologia, o efeito médio sobre os controles (ATC) foi de 0,1246 (IC 95% [0,1089; 0,1412]) sobre 2.248 pares, e o efeito médio sobre os tratados (ATT) foi de 0,1226 (IC 95% [0,1064; 0,1399]) sobre 2.243 pares — ambos com equilíbrio integral das 41 covariáveis ($|SMD| < 0,1$ em 100% dos casos após o pareamento; SMD médio de 0,0146 no ATT e equivalente no ATC). Os dois efeitos são consistentes em direção e magnitude: a medida protetiva, quando concedida a casos com perfil comparável ao dos que a receberam de fato, reduziria a probabilidade de revitimização em aproximadamente 12 pontos percentuais.

A Tabela II detalha a composição dos pares. Em ambos os cenários, o grupo tratado apresenta TR substancialmente superior à do controle pareado (19,79% vs. 7,53% no ATT; 20,15% vs. 7,70% no ATC), padrão que reflete a seleção de casos de maior risco para concessão da medida.

Estimador	Grupo	Com revitimização	Sem revitimização	Com e sem revitimização	TR
ATT	Tratado	444	1.799	2.243	19,79%
ATT	Controle	169	2.074	2.243	7,53%
ATC	Tratado	453	1.795	2.248	20,15%
ATC	Controle	173	2.075	2.248	7,70%

Table II. Comparação dos resultados dos cenários após a aplicação do método PSM

Como caracterização descritiva do subgrupo — e não como estimativa de efeito —, examina-se por fim quantos dos 185 casos revitimizados sem medida tinham perfil semelhante ao dos casos protegidos. Aplicando o *caliper* aos 185 casos, 183 (98,9%) encontraram par tratado dentro do limiar (PS entre 0,25 e 0,89, mediana 0,66): 43,2% com $PS \geq 0,70$, 48,6% entre 0,40–0,70 e 8,1% abaixo de 0,40. A concentração em faixas elevadas sugere subidentificação de risco no momento da decisão, motivando o modelo preditivo descrito a seguir.

Entre os algoritmos avaliados, o XGBoost apresentou o melhor desempenho médio, beneficiando-se de sua estratégia de boosting sequencial, que constrói sequencialmente árvores que corrigem os erros dos modelos anteriores por meio da otimização dos gradientes da função de perda — particularmente eficaz sob desbalanceamento de classes [Gupta 2021] [Rajendran and Sundararajan 2021].

A Tabela III apresenta os valores médios da AUC e o intervalo de confiança (95%) para cada algoritmo, considerando duas composições de variáveis: (i) o conjunto completo de variáveis; e (ii) o uso exclusivo das variáveis do FONAR. Restringir o conjunto de atributos ao FONAR reduz a AUC de todos os algoritmos para a faixa de 0,5311 a 0,5709. A integração com o histórico judicial aumenta o desempenho preditivo. No cenário com todas as variáveis, o XGBoost identifica corretamente 80,27% dos casos de reincidência (*recall* da classe positiva = 0,8027), combinando AUC = 0,6139, G-mean = 0,5330 e F1 = 0,4765. Observou-se desempenho modesto: a boa taxa de detecção da classe positiva não se traduziu em equilíbrio adequado entre sensibilidade e precisão.

Complementarmente às métricas agregadas, avaliou-se a estabilidade do modelo estratificada por tipo de vínculo entre vítima e agressor, a fim de identificar possíveis disparidades de desempenho. A AUC variou entre os subgrupos, oscilando de 0,4963 entre "ex-companheiros há menos de 6 meses" (n=956), valor próximo ao esperado pelo acaso, a 0,6567 entre "ex-namorados/noivos há mais de 6 meses" (n=226).

Classificador	Todas as variáveis	Variáveis do FONAR
XGBoost	0,6139 [0,5897, 0,6381]	0,5709 [0,5328, 0,6089]
Random Forest	0,5893 [0,5580, 0,6206]	0,5699 [0,5353, 0,6046]
Linear Discriminant Analysis	0,5704 [0,5480, 0,5929]	0,5311 [0,4998, 0,5624]
Logistic Regression	0,5740 [0,5575, 0,5905]	0,5348 [0,5037, 0,5659]

Table III. Desempenho comparativo dos algoritmos

Esse comportamento sugere que as variáveis do FONAR, quando consideradas isoladamente, não capturam integralmente os fatores de risco necessários para a discriminação adequada entre as classes. Tal limitação reforça a importância da integração com outras fontes de dados, de modo a ampliar a representação informacional e melhorar a capacidade preditiva dos modelos.

Uma análise de interpretabilidade com a biblioteca SHAP ³[Lundberg and Lee 2017] reforçou o que foi previamente descrito na matriz de correlação. "Processo Agressor Vítima" apresenta a maior magnitude de impacto, predominantemente negativo quando elevado, enquanto "Medida Protetiva Concedida Antes" exerce a influência positiva mais forte e consistente. Variáveis processuais como "Q09 – Usa Substâncias", "Processo Criminal Agressor" e "Violência Doméstica Agressor" também contribuem negativamente com magnitude moderada, reforçando a relevância do histórico processual. Já as variáveis do FONAR (Q17, Q15, Vínculo, Q01, Q18) exibem valores SHAP próximos de zero, atuando como ajustes finos ao limiar de decisão.

5. CONCLUSÃO

Este estudo apresenta evidências de que o aprendizado de máquina pode qualificar a decisão judicial sobre medidas protetivas em casos de violência contra a mulher. A análise por pareamento 1:1 indica que a medida reduziria a probabilidade de revitimização em aproximadamente 12 pontos percentuais nos casos com perfil comparável ao dos protegidos (ATC = 0,1246; ATT = 0,1226). Além disso, 43,2% dos 185 casos revitimizados sem medida apresentavam escore de propensão $\geq 0,70$, sugerindo subidentificação de risco no momento da decisão. No modelo preditivo, o XGBoost com o conjunto enriquecido atinge AUC = 0,6139 e recall de 80,27% para a classe positiva, contra AUC = 0,5709 quando restrito ao FONAR - um ganho de aproximadamente 4,3 pontos percentuais em AUC atribuível ao histórico judicial.

Esses achados dialogam diretamente com a literatura internacional, indicando que apenas um subconjunto das variáveis demonstra significância preditiva relevante [Pham et al. 2026], e a integração com dados históricos produz ganhos substanciais de desempenho [Turner et al. 2019] [Turner et al. 2021] [Grogger et al. 2021]. A análise SHAP corrobora o primeiro ponto: as variáveis de "Medida Protetiva Concedida Antes" e "Processo Agressor Vítima" concentram o sinal preditivo, enquanto a maioria dos itens do formulário atua apenas como ajuste fino; e o salto de AUC entre os cenários (i) e (ii) da Tabela III corrobora o segundo.

É importante ressaltar as limitações deste estudo. Primeiro, a qualidade do registro administrativo: a vinculação probabilística depende da consistência dos identificadores, e os códigos de movimentação processual nem sempre distinguem com precisão a concessão de medida, como ilustra o descompasso entre o item 7b do FONAR e os registros judiciais. Segundo, a causalidade exige cautela — a associação entre concessão da medida e revitimização decorre em parte da seleção de casos mais graves para tratamento; o PSM mitiga, mas não elimina esse viés. Terceiro, o desempenho preditivo, embora superior ao uso isolado do FONAR (AUC=0,6139 vs. 0,5709), permanece moderado e heterogêneo entre subgrupos de vínculo (AUC de 0,50 a 0,66), o que limita seu uso como ferramenta automatizada e reforça seu papel de apoio complementar ao julgamento técnico. Por fim, a dependência de um conjunto restrito de variáveis institucionais favorece a interpretabilidade, mas recomenda parcimônia quanto à generalização — mudanças nas práticas de registro alterariam o sinal capturado.

Em termos de política pública, os achados apontam para o monitoramento intensivo do subconjunto identificável com histórico de descumprimento, registros prévios e processos relacionados — trajetórias em que a violência tende a se concentrar. A prevenção não se esgota na medida protetiva: exposição infantil, disputas de guarda e padrões de controle reforçam a necessidade de proteção social e apoio psicossocial articulados a ela. Trabalhos futuros devem investigar calibração temporal, robustez a mudanças no protocolo, equidade do modelo e efeito de intervenções por faixa de risco.

³<https://pypi.org/project/shap/>

REFERENCES

- ABADIE, A. AND IMBENS, G. W. On the failure of the bootstrap for matching estimators. *Econometrica* 76 (6): 1537–1557, 2008.
- ALLARD, V., HIGGS, T., AND SLIGHT, M. Psychometric properties of the spousal assault risk assessment from samples of people having perpetrated intimate partner violence. *Trauma, Violence, & Abuse* 25 (5): 3443–3460, 2024.
- AUSTIN, P. C. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics* 10 (2): 150–161, 2011.
- BERK, R., HEIDARI, H., JABBARI, S., KEARNS, M., AND ROTH, A. Fairness in Criminal Justice Risk Assessments: The State of the Art. *Sociological Methods & Research* 50 (1): 3–44, February, 2021.
- CAULFIELD, K. M., GRAY, N. S., EDWARDS, A., AND SNOWDEN, R. J. The domestic abuse, stalking and harassment and honour-based violence (dash) risk assessment instrument in predicting deadly or persistent domestic abuse. *Forensic Sciences* 5 (4): 272–284, 2025.
- DA SILVA, T. D. *GESTÃO DE CONFLITOS E DE ALTERNATIVAS PENAIS NO COMBATE À VIOLÊNCIA DOMÉSTICA E FAMILIAR CONTRA A MULHER: O PROGRAMA PROTEÇÃO NA MEDIDA*. M.S. thesis, Universidade de Fortaleza (UNIFOR), Fortaleza, Ceará, Brasil, 2021.
- GRAY, A. L., MILLS, J. F., AND FORTH, A. E. Risk for violent and ipv recidivism among incarcerated men with a history of ipv perpetration: An examination of the predictive validity of the odara, dvrag, and pcl-r. *Criminal Justice and Behavior* 52 (1): 79–100, 2025.
- GROGGER, J., GUPTA, S., IVANDIC, R., AND KIRCHMAIER, T. Comparing conventional and machine-learning approaches to risk assessment in domestic abuse cases. *Journal of Empirical Legal Studies* 18 (1): 90–130, 2021.
- GUPTA, A. Xgboost versus random forest. Disponível em: <https://medium.com/geekculture/xgboost-versus-random-forest-898e42870f30>. Acesso em: 31 de dezembro de 2022, 2021.
- JÚNIOR, A. P. AND RIBEIRO, F. Análises de dados de violência doméstica com o uso de aprendizagem de máquina: um mapeamento sistemático. In *Anais da VII Escola Regional de Computação do Ceará, Maranhão e Piauí*. SBC, Porto Alegre, RS, Brasil, pp. 143–150, 2019.
- LISBOA, M., TEIXEIRA, A. L., AND PASINATO, W. *Formulário de Risk Assessment para o CNVD: um instrumento para o enfrentamento da violência doméstica contra a mulher*. <http://www.mpce.mp.br>, Brasil, 2019. info:eu-repo/grantAgreement/FCT/5876/147304/PT# UID/SOC/04647/2019.
- LUNDBERG, S. AND LEE, S.-I. A unified approach to interpreting model predictions, 2017.
- MYHILL, A., HOHL, K., AND JOHNSON, K. The ‘officer effect’ in risk assessment for domestic abuse: Findings from a mixed methods study in england and wales. *European Journal of Criminology* 20 (3): 856–877, 2023.
- PHAM, A. T., NUNES, K. L., HILTON, N. Z., ENNIS, L., AND JUNG, S. The factor structure of intimate partner violence risk. *Criminal Justice and Behavior* 53 (1): 101–118, 2026.
- RAJENDRAN, S. AND SUNDARARAJAN, P. Predicting criminal recidivism using specialized feature engineering and xgboost. *CrimRxiv* n/a (n/a): –, 2021.
- ROSENBAUM, P. R. AND RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70 (1): 41–55, 1983.
- RYLAND, H., BURGHART, M., ZHONG, S., FAZEL, S., AND YU, R. Risk assessment tools used at the policing stage for health and crime outcomes: A systematic review and meta-analysis. *Social Science & Medicine* vol. 383, pp. 118457, 2025.
- STUART, E. A. Matching methods for causal inference: A review and a guide. *Statistical Science* 25 (1): 1–21, 2010.
- TURNER, E., BROWN, G., AND MEDINA, J. Predicting domestic abuse (fairly) and police risk assessment. *CrimRxiv* vol. 31, pp. 145–157, feb 25, 2021. <https://www.crimrxiv.com/pub/nrsjkqcv>.
- TURNER, E., MEDINA, J., AND BROWN, G. Dashing hopes? the predictive accuracy of domestic abuse risk assessment by police. *The British Journal of Criminology* vol. 59, pp. 1013–1034, 08, 2019.
- VILJOEN, J. L., GOOSSENS, I., MONJAZEB, S., COCHRANE, D. M., VARGEN, L. M., JOHNSON, M. R., BLANCHARD, A. J. E., LI, S. M. Y., AND JACKSON, J. R. Are risk assessment tools more accurate than unstructured judgments in predicting violent, any, and sexual offending? A meta-analysis of direct comparison studies. *Behavioral Sciences & the Law* 43 (1): 75–113, 2025.
- WEBER, M. A., SCHNYDER, N., KIRSCHSTEIN, M. A., GRAF, M., ENDRASS, J., AND ROSSEGGER, A. The key role of base rates: systematic review and meta-analysis of the predictive value of four risk assessment instruments. *Swiss Medical Weekly* 155 (40166): n/a, 2025.