57º SBPO
SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL
Gramado • RS
DE 5 A 9 DE OUTUBRO DE 2025

APLICAÇÃO DE TÉCNICAS DE MACHINE LEARNING PARA PREVISÃO DE HOTSPOTS DE MORTES VIOLENTAS INTENCIONAIS NO MUNICÍPIO DE CARUARU, INTERIOR DE PERNAMBUCO.

Luann Bruno Vidal de Andrade
Programa de Pós-Graduação em Engenharia de Produção – PPGEP/CAA, Universidade Federal de Pernambuco (UFPE), Centro Acadêmico do Agreste
Av. Marielle Franco, s/n, Bairro Nova Caruaru, Caruaru – PE, 55014-900, Brasil
luann.andrade@ufpe.br

Lúcio Camara e Silva
Programa de Pós-Graduação em Engenharia de Produção – PPGEP/CAA, Universidade Federal de Pernambuco (UFPE), Centro Acadêmico do Agreste
Av. Marielle Franco, s/n, Bairro Nova Caruaru, Caruaru – PE, 55014-900, Brasil
lucio.silva@ufpe.br

RESUMO

Este artigo apresenta um modelo preditivo para identificação de hotspots de Mortes Violentas Intencionais (MVI) no município de Caruaru, interior de Pernambuco, por meio da aplicação de algoritmos de aprendizado de máquina a dados temporais e geoespaciais. Os dados utilizados são referentes aos casos reais de MVI no município, e abrange o período de 2017 a 2024. A modelagem foi conduzida com diferentes algoritmos de machine learning, usando de técnicas como otimização de hiperparâmetros e balanceamento das classes. Os resultados demonstraram desempenho satisfatório, com destaque para o modelo XGBoost, com um F1-score médio de 75% e recall de 80% na previsão mensal dos hotspots em 2024. Aproximadamente 26% dos homicídios registrados no ano ocorreram em áreas previamente identificadas como hotspots, evidenciando o potencial da abordagem como ferramenta preventiva.

PALAVRAS CHAVE. Segurança pública, Aprendizado de máquina, Previsão de hotspots.

(1) AS&DS – Análise e Ciência de Dados; (2) D&SP – PO em Defesa e Segurança Pública; (3) OA – Outras Aplicações em PO.

ABSTRACT

This article presents a predictive model for identifying hotspots of Intentional Violent Deaths (MVI) in the municipality of Caruaru, in the interior of Pernambuco, through the application of machine learning algorithms to temporal and geospatial data. The data used refer to real cases of MVI in the municipality, and cover the period from 2017 to 2024. The modeling was conducted with different machine learning algorithms, using techniques such as hyperparameter optimization and class balancing. The results demonstrated satisfactory performance, with emphasis on the XGBoost model, with an average F1-score of 75% and recall of 80% in the monthly prediction of hotspots in 2024. Approximately 26% of the homicides registered in the year occurred in areas previously identified as hotspots, highlighting the potential of the approach as a preventive tool.

KEYWORDS. Public Secutiry. Machine learning. Hotspot prediction.

(1) AS&DS- Data Science and Analytics; (2) D&SP – OR in Defense and Public Security; (3) OA – Other Applications in OR

DOI: 10.59254/sbpo-2025-212301
https://proceedings.science/p/212301?lang=pt-br

1. Introdução

A violência no Brasil alcança níveis alarmantes, chegando a um total de 46.328 casos de Mortes Violentas Intencionais (MVI) registradas no ano de 2023 [Fórum Brasileiro de Segurança Pública, 2024]. Os altos índices de homicídios no Brasil podem ser atribuídos a diversos fatores, como o período de ditadura militar, crescente desigualdade, tráfico de drogas, disponibilidade de armas, entre outros [Imbusch et al., 2011]. Além desses desafios, os órgãos de segurança pública sofrem com a falta de investimentos em infraestrutura e modernização, resultando na falta de profissionais de segurança pública, além da obsolescência nos meios de atuação. Esse contexto evidencia a necessidade do aprimoramento de estratégias e ferramentas de gestão da segurança pública.

Em resposta, políticas públicas foram criadas com o objetivo de promover maior integração e efetividade na gestão da segurança pública. Em 2018 o governo federal instituiu o Sistema Único de Segurança Pública (SUSP), pela Lei nº 13.675/2018, que tem como objetivo integrar os diversos órgãos de segurança pública em todo o território nacional, por meio de compartilhamento de informações e a atuação coordenada entre federação, estados e municípios. Em nível estadual, o Pacto pela Vida, implementado em Pernambuco no ano de 2007, tornou-se referência por se tratar do primeiro plano de segurança pública implementado em um estado do Nordeste com foco na redução de homicídios por meio de metas claras, monitoramento sistemático e uso de dados para orientar ações policiais e sociais [Cerqueira et al., 2017; Silveira Neto et al., 2022]. Recentemente, o governo estadual lançou o programa Juntos pela Segurança, que expande essa lógica integrativa e baseada em evidências, com o objetivo de aprimorar a atuação estratégica das forças de segurança em conjunto com políticas sociais e urbanas.

Neste contexto, o aprimoramento das estratégias de policiamento e políticas de prevenção à violência torna-se um fator crucial para garantir o bem-estar da sociedade. Este cenário fortalece o uso de sistemas de informação especializados e abordagens baseadas na proatividade policial, que apoiam-se na análise de dados criminais, na tomada de decisões baseadas em evidências e na atuação estratégica das forças de segurança, podendo-se destacar conceitos como o ”policiamento inteligente”e o Intelligence-Led Policing (ILP). O ”policiamento inteligente, efetivo e proativo é claramente preferível a simplesmente reagir a atos criminosos”[Perry et al., 2013], com isso, o policiamento preditivo surge como uma abordagem promissora para o combate à criminalidade, baseia-se na aplicação de técnicas analíticas, especialmente técnicas quantitativas, para identificar prováveis alvos para intervenção policial e prevenção de crimes ou solucionar crimes passados, através de predições estatísticas [Perry et al., 2013].

Entre as diversas abordagens encontradas na literatura voltadas à segurança pública e à previsão de crimes, as técnicas baseadas em machine learning vêm ganhando destaque nos últimos anos. Isso pode ser associado a capacidade destes algoritmos em identificar padrões complexos na ocorrência de crimes e gerar insights valiosos acerca de variáveis relacionadas [Mandalapu et al., 2023]. Além disso, os diversos algoritmos de machine learning oferecem uma grande flexibilidade para a aplicação em diferentes contextos, com variados tipos de dados, objetivos analíticos e restrições operacionais. Estudos, como em Kounadi et al. [2020]; Dakalbab et al. [2022], mostram o machine learning como uma ferramenta promissora para a previsão de crimes, capaz de antecipar a ocorrência de crimes, possibilitando a atuação das forças de segurança de forma mais eficaz e direcionada.

Com isso, surge a pergunta norteadora deste trabalho: é possível, com base em fatores já conhecidos, identificar os locais com maior probabilidade de mortes violentas intencionais no futuro, possibilitando a atuação de forma preventiva das forças de segurança?. Neste contexto, o presente trabalho se propõe a, a partir da análise de dados históricos de MVI no contexto do município de Caruaru, Pernambuco, aplicar algoritmos de machine learning para identificar prováveis

DOI: 10.59254/sbpo-2025-212301
https://proceedings.science/p/212301?lang=pt-br

focos futuros, possibilitando a intervenção e contribuindo para a redução dos índices de violência e uso eficiente dos recursos das forças de segurança.

2. Segurança pública

A segurança pública é um dos pilares fundamentais para a promoção do bem-estar e da qualidade de vida de uma sociedade [Lima, 2024]. Segundo Murray et al. [2013], a violência se tornou um dos principais problemas de saúde pública no Brasil, com um crescimento alarmante nos números de homicídios no país entre 1980 e 2010. Em 2018, por meio da lei Nº 13.675, foi criado a Política Nacional de Segurança Pública e Defesa Social (PNSPDS) e institui-se o Sistema Único de Segurança Pública (Susp), marcando mais uma tentativa importante para a reversão do cenário da violência no país. Entre os objetivos do PNSPDS estão a priorização de políticas de redução da letalidade violenta, além do incentivo a medidas para a modernização de equipamentos, da investigação e da perícia e para a padronização de tecnologia dos órgãos e das instituições de segurança pública [Brasil, 2018].

Conforme exposto por Azevedo et al. [2011], a utilização de sistemas de informações é um fator crucial para a atuação das forças de segurança pública. Neste contexto, abordagens como a Intelligence-Led Policing (ILP) ganham força como alternativa para os métodos tradicionais de policiamento, trazendo um modelo de gestão centrado na tomada de decisão baseada em análise de dados e ”inteligência policial”, priorizando pontos críticos de criminalidade, vítimas reincidentes, infratores prolíficos e grupos criminosos [Ratcliffe, 2016]. Ao possibilitar o direcionamento das operações policiais com base em padrões criminais e tendências emergentes, os sistemas de informação se tornam peças-chave para a construção de estratégias de policiamento e de uma segurança pública mais eficaz.

O policiamento é o mecanismo colocado em ação para regular a ordem social, meio pelo qual o Estado assegura o desenrolar da vida social e o cumprimento dos postulados definidos politicamente [Fernandes, 2021]. Portanto, trata-se de uma função essencial para a manutenção da sociedade e da proteção dos direitos fundamentais, e, consequentemente, deve ser planejada e executada de forma estratégica, com base em evidências e diagnósticos precisos, de modo a garantir maior efetividade. Na literatura, é possível encontrar estudos acerca de diferentes estratégias de policiamento adotadas pelos órgãos de segurança pública, dentre elas se destaca o policiamento de hotspots, conforme apresentado por Araujo [2019]. Braga e Weisburd [2022] apontam que programas de policiamento de hotspots geram uma redução estatisticamente significativa na criminalidade. Borges et al. [2017] definem um hotspot como ”uma área ou vizinhança com um alto número de registros de crimes em um certo período de tempo comparado a outras vizinhanças da cidade”. Portanto, pode-se definir o policiamento de hotspots como a aplicação das forças de segurança de forma focalizada em áreas de alto risco. Nesse mesmo escopo de atuação baseada em evidências, emerge o conceito de policiamento preditivo, definido por Perry et al. [2013] como a ”aplicação de técnicas analíticas, especialmente técnicas quantitativas, para identificar prováveis alvos para intervenção policial e prevenção de crimes ou solucionar crimes passados através de predições estatísticas. Perry et al. [2013], ainda, classificam os métodos de predição de crimes em quatro grupos, sendo um deles voltado especificamente para a predição de lugar e período com maiores riscos de crimes. Essa perspectiva amplia o alcance do policiamento de hotspots, ao permitir não apenas a identificação retrospectiva de áreas críticas, mas também a antecipação de eventos criminosos com base em padrões históricos e dinâmicas territoriais.

Caruaru é um dos principais e mais populosos municípios de Pernambuco. Semelhante ao cenário nacional, Caruaru enfrenta desafios em relação à violência e segurança pública. De acordo com dados da Secretaria de Defesa Social de Pernambuco, o município registrou 127 MVIs no ano ano de 2023, resultando em uma taxa de 33,16 MVIs a cada 100 mil habitantes, 45% superior à

DOI: 10.59254/sbpo-2025-212301
https://proceedings.science/p/212301?lang=pt-br

taxa nacional [Secretaria de Defesa Social de Pernambuco, 2025], [Fórum Brasileiro de Segurança Pública, 2024]. Esses dados indicam uma grave deficiência no aspecto da segurança no município, especialmente no que diz respeito à proteção à vida.

2.1. Machine learning para previsão de crimes

Na literatura, é possível encontrar uma variedade de estudos no campo da segurança pública e da previsão de crimes, baseados em abordagens metodológicas distintas. Embora cada uma dessas abordagens contribua com perspectivas válidas e complementares ao tema, essa diversidade evidencia a inexistência de um consenso quanto à melhor estratégia a ser adotada para lidar com a complexidade dos fenômenos criminais. Nesse contexto, o uso de técnicas de machine learning tem ganhado destaque como uma abordagem promissora. De acordo com Zhang et al. [2022], ferramentas de machine learning podem ser aplicadas à previsão de crimes de forma que proporcione insights significativos para o entendimento dos fatores influenciadores dos crimes. No contexto da segurança pública, o machine learning pode ser aplicado de diferentes maneiras, desde a previsão de áreas de risco até a identificação de padrões comportamentais, com o uso de distintos tipos de dados, algoritmos, técnicas e objetivos, o que reforça sua flexibilidade e adaptabilidade a diferentes contextos.

Modelos baseados em árvores de decisão, como Random Forest e XGBoost, têm sido amplamente utilizados neste contexto por sua robustez, como evidenciado nos trabalhos de Zhang et al. [2022], Campedelli [2022], Araujo [2019] e Alsubayhin et al. [2024]. Além disso, algoritmos como o Suport Vector Machine (SVM), utilizado por Kianmehr e Alhajj [2006], e abordagens comparativas como a de Wheeler e Steenbeek [2021], que analisam o desempenho do Random Forest frente a métodos tradicionais como estimativa de densidade de kernel (KDE) e modelagem de risco de terreno (RTM), reforçam o potencial das técnicas de aprendizado supervisionado em prever áreas de risco com base em dados históricos.

3. Metodologia

O presente trabalho foi desenvolvido a partir de uma base de dados com informações sobre os casos de MVI no município de Caruaru fornecida pela Secretaria de Segurança Municipal de Caruaru (SSM). Os dados abrangem o período de janeiro de 2017 a dezembro de 2024, e reúne informações detalhadas sobre os crimes, como georreferenciamento, data, além de variáveis do perfil do crime, como a motivação e objeto utilizado, e do perfil da vítima, como antecedentes criminais, sexo e uso de narcóticos na hora da morte. Para a análise espacial dos dados, foi utilizado um arquivo GeoJSON com os limites dos bairros e da área territorial do município. Para fins de simplificação do modelo, a análise foi delimitada à zona urbana do município, definida pela extensão dos bairros oficialmente reconhecidos.

3.1. Preparação dos dados

Ao total a base de dados possui 1167 casos de MVI registrados ao longo do período mencionado, no entanto, foram descartados os casos onde não foi informado o georreferenciamento, e os casos na zona rural do município. Após a filtragem, restaram 945 casos válidos para análise.

Utilizando a linguagem de programação Python, com a biblioteca de geoprocessamento GeoPandas, foi criado uma malha espacial composta por quadrantes regulares de 800 metros de lado, delimitado à zona urbana do município. Conforme sugerido por Limeira et al. [2022], a escolha do tamanho dos quadrantes deve se adequar a realidade do estudo. Neste trabalho, a extensão de 800 metros foi escolhida visando equilibrar o nível de detalhamento da área estudada com a viabilidade computacional do modelo, além do bom enquadramento de áreas chaves do centro do município, como os bairros Nossa Senhora das Dores e Centenário, e Parque 18 de Maio. Ao restringir a análise à zona urbana, foi possível reduzir a quantidade de quadrantes de 1584 para

DOI: 10.59254/sbpo-2025-212301
https://proceedings.science/p/212301?lang=pt-br

213, permitindo uma redução considerável no tempo de processamento e eliminando diversos quadrantes sem registros de MVI no período analisado. Em seguida, os casos MVIs foram agregados espacialmente, permitindo a identificação do quadrante para os dados de cada caso de MVI, individualmente.

Conforme descrito no trabalho de Imbusch et al. [2011], é possível identificar diversos fatores que contribuem para os índices de violência nos países da América Latina como, períodos de ditadura militar, desigualdade, tráfico de drogas, disponibilidade de armas de fogo, crime organizado, baixo crescimento econômico, entre outros. Limeira et al. [2022] utiliza os seguintes fatores de risco para a determinação de áreas de risco de homicídios: lugares de ocorrências de arma de fogo; zonas de atuação de gangues e tráfico de drogas; pontos de incidência de homicídios; locais de violência interpessoal e de conflitos violentos cotidianos; áreas de vulnerabilidade: aglomerados subnormais, vilas e áreas de risco. Considerando esse contexto e com base nas dados disponíveis em cada registro de MVI, foram extraídas quatro variáveis explanatórias com o intuito de captar essas dinâmicas: MVIs utilizando arma de fogo, MVIs relacionados à narcóticos, MVIs motivados por atividades criminais (de acordo com a portaria nº 6195, de 30/12/2019 da SDS), número de vítimas que possuíam antecedentes criminais.

Seguindo, para cada quadrante, os dados foram agregados temporalmente, de acordo com o mês e ano, formando uma série temporal mensal para cada quadrante, uma granularidade comumente utilizada pelos gestores e autoridades do município, sendo esta, também, a frequência com que os dados dos indicadores criminais são divulgados pela SDS-PE. Conforme a metodologia utilizada por Araujo [2019], foram extraídas os componentes de tendência e a sazonalidade para estas séries temporais, criando novas variáveis explanatórias. Além disso, para cada variável explanatória, foram incorporadas os valores defasados correspondentes aos últimos cinco meses, de modo que, para cada observação, o modelo tenha acesso aos valores das variáveis explicativas referentes ao mês atual e aos cinco meses anteriores.

Neste trabalho, um quadrante foi classificado como hotspot caso atendesse a pelo menos um dos seguintes critérios: (i) ter registrado quatro ou mais casos de MVI nos últimos seis meses; (ii) apresentar uma taxa de crescimento de, no mínimo, 100% no número de MVIs nos últimos seis meses em relação aos seis meses anteriores a estes; (iii) ter registrado três ou mais casos de MVI relacionados a atividades criminais nos últimos seis meses; conforme representado na Equação 1. Os parâmetros utilizados para a identificação dos hotspots foram definidos pelos autores em conjunto com gestores da Secretaria de Segurança do município.

$$\text{hotspot}_t = \begin{cases} \text{Sim}, & \text{se } \sum_{i=0}^{5} \text{MVI}_{t-i} \geq 4 \\ & \quad \text{ou } \left(\frac{\sum_{i=0}^{5} \text{MVI}_{t-i} - \sum_{i=6}^{11} \text{MVI}_{t-i}}{\sum_{i=6}^{11} \text{MVI}_{t-i}} \geq 1.0 \right) \\ & \quad \text{ou } \sum_{i=0}^{5} \text{MVI_ac}_{t-i} \geq 3 \\ \text{Não}, & \text{caso contrário} \end{cases} \tag{1}$$

A partir desta definição, para cada período, cada quadrante foi rotulado como hotspot de MVI ou não, sendo essa a variável dependente do modelo. Como o objetivo deste trabalho é prever se determinado quadrante será um hotspot no próximo período, a variável dependente foi deslocada um período para trás, fazendo com que cada linha da base de dados reflita a condição futura do quadrante, permitindo ao modelo aprender padrões antecedentes ao surgimento de um hotspot.

Com isso, formou-se uma base de dados estruturada conforme ilustrado na Figura 1, onde o termo ”lag”representa o valor da variável defasado por n períodos.

DOI: 10.59254/sbpo-2025-212301
https://proceedings.science/p/212301?lang=pt-br

Figura 1: Estrutura dos dados utilizados no modelo

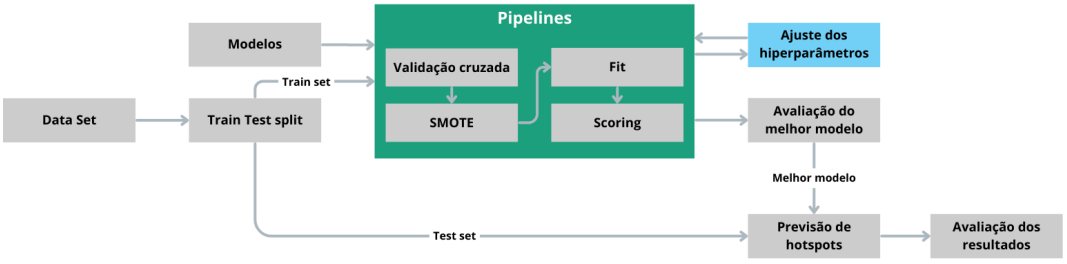
Quadrante	Mês-ano	MVI, X1, X2, ..., Xn					T	S	hotspot no próximo período?
		Mês atual	lag 1	lag2	...	lag5			
Q100	2017-01	3	1	0	...	0	1	0,15	VERDADEIRO
Q100	2017-02	4	1	0	...	0	1	0,22	FALSO
Q100	2017-03	3	0	0	...	0	1	0,14	FALSO
Q102	2017-01	0	0	0	...	0	0	0,25	FALSO
Q102	2017-02	0	0	0	...	0	1	0,30	FALSO
Q102	2017-03	1	0	1	...	1	1	0,10	FALSO

Fonte: os autores (2025).

3.2. Modelo de previsão de hotspots

Conforme o apresentado na secção 2.1, para o problema de previsão de hotspots, foram selecionados quatro modelos de machine learning, sendo eles: XGBoost (XGB), Random Forest (RF), Support Vector Classifier (SVC) e Regressão Logística (LR). A criação do modelo de previsão de hotspots seguiu o processo ilustrado no fluxograma da Figura 2.

Figura 2: Fluxograma da criação do modelo de machine learning para previsão de hotspots



Fonte: os autores (2025).

Inicialmente, os dados foram divididos entre os grupos de treino e teste. Os dados entre janeiro de 2017 e novembro de 2023 foram usados para o treinamento dos modelos, enquanto que os dados de dezembro de 2023 a novembro de 2024 foram utilizados para a validação da abordagem e avaliação dos resultados. Na etapa de escolha do modelo mais adequado, os modelos escolhidos são treinados utilizando os dados de treino, seguindo o *pipeline* da Figura 2. Para garantir mais robustez nos resultados e mitigar o *over* e *underfitting*, foi empregada a técnica de validação cruzada estratificada repetida (Repeated Stratified KFold) da biblioteca Scikit-learn. O método consiste em particionar os dados de forma estratificada múltiplas vezes, preservando a proporção de classes da variável dependente para cada partição. Essa abordagem é especialmente útil em casos onde há desbalanceamento entre as classes da variável dependente, como é o caso deste trabalho, e permite reduzir a variabilidade dos modelos, tendo em vista que os resultados são convertidos em uma média de várias execuções do modelo, resultando em uma avaliação mais confiável.

Pela natureza do problema, a variável dependente (hotspot ou não hotspot) apresenta um forte desbalanceamento entre as classes, a maioria dos quadrantes não são hotspots. Essa assimetria cria uma dificuldade na previsão de quadrantes que são hotspots, já que, durante o treinamento, os modelos têm poucos exemplos dessa classe minoritária, e consequentemente tendem a favorecer a classe majoritária. Para contornar esse fator, foi utilizado o método SMOTE (Synthetic Minority Over-sampling Technique, ou Técnica de sobreamostragem de minoria sintética), criando artifici-

DOI: 10.59254/sbpo-2025-212301
https://proceedings.science/p/212301?lang=pt-br

almente novos exemplos da classe minoritária. Com isso, busca-se fornecer aos modelos exemplos suficientes para aprender os padrões associados aos hotspots e, assim, melhorar sua capacidade de previsão.

Utilizando otimização bayesiana, através da biblioteca Optuna, os hiperparâmentros dos modelos propostos foram ajustados buscando otimizar o F1-score dos modelos. Após a otimização, o modelo com o melhor F1-score médio foi selecionado para as etapas posteriores. O modelo selecionado foi aplicado ao grupo de teste, correspondente aos dados do período de dezembro de 2023 a novembro de 2024. Por fim, os resultados foram avaliados.

4. Resultados e discussões

Nesta seção, são apresentados os resultados dos modelos de machine learning propostos para a previsão de hotspots de MVI, no município de Caruaru. As análises se concentram na capacidade preditiva dos modelos, no desempenho mês a mês ao longo do ano de 2024 e nos padrões de acertos e erros.

4.1. Desempenho geral dos modelos

A Tabela 1 resume o desempenho médio dos modelos Random Forest, XGBoost, SVC e Regressão Logística, avaliados com base no F1-score, assim como os respectivos parâmetros otimizados para cada modelo. Conforme os resultados obtidos na avaliação, o modelo XGBoost apresentou o melhor resultado, e com isso, foi escolhido para implementação junto aos dados de teste, referentes ao ano de 2024.

Tabela 1: Desempenho e parâmetros dos modelos

Modelo	F1-score	Parâmetros
XGB	80,17%	`n_estimators = 973, learning_rate = 0.0636, max_depth = 10, subsample = 0.6657, colsample_bytree = 0.8331, min_child_weight = 8`
RF	77,42%	`n_estimators = 201, max_depth = 19, min_samples_split = 8, min_samples_leaf = 2, max_features = 'sqrt'`
SVC	54,45%	`C = 0.9575`
LR	57,42%	`C = 0.1684`

Fonte: os autores (2025).

Os resultados obtidos pelo modelo são apresentados na Tabela 2, onde, para cada mês, é disposto o número de hotspots observados nos dados reais, o número de previsões acertadas (TP), o número de previsões incorretas (FP), e o número de hotspots reais não previstos (FN). A partir desses valores, foram calculadas as métricas de avaliação: precisão, recall e F1-score. O modelo alcançou uma precisão média de 71%, recall de aproximadamente 80% e F1-score de 75%, indicando um grande potencial preditivo para modelo desenvolvido.

4.2. Discussão dos resultados

Conforme os resultados da Tabela 2, apesar das métricas apresentarem resultados satisfatórios, nota-se uma tendência maior do modelo a cometer o erro de tipo 1 — isto é, classificar como hotspot uma área que, na realidade, não seria. Esta perspectiva é refletida na diferença entre as métricas de precisão e recall: enquanto o recall foi de aproximadamente 80%, indicando boa capacidade de identificar corretamente os hotspots reais, a precisão foi cerca de 71%, revelando que parte das áreas classificadas como hotspots não se concretizaram. Na prática, esse comportamento pode se traduzir em um uso ineficiente de recursos por parte das forças de segurança pública, ao

DOI: 10.59254/sbpo-2025-212301
https://proceedings.science/p/212301?lang=pt-br

Tabela 2: Desempenho mensal da previsão de hotspots em 2024

Ano-mês	Reais	TP	FP	FN	Precisão	Recall	F1-score
2024-jan	5	2	3	3	40,00%	40,00%	40,00%
2024-fev	6	5	1	1	83,33%	83,33%	83,33%
2024-mar	7	7	2	0	77,78%	100,00%	87,50%
2024-abr	7	7	0	0	100,00%	100,00%	100,00%
2024-mai	5	4	1	1	80,00%	80,00%	80,00%
2024-jun	5	3	1	2	75,00%	60,00%	66,67%
2024-jul	7	5	2	2	71,43%	71,43%	71,43%
2024-ago	6	5	2	1	71,43%	83,33%	76,92%
2024-set	6	5	2	1	71,43%	83,33%	76,92%
2024-out	8	7	1	1	87,50%	87,50%	87,50%
2024-nov	6	5	3	1	62,50%	83,33%	71,43%
2024-dez	6	4	6	2	40,00%	66,67%	50,00%
Total	**74**	**59**	**24**	**15**	**71,08%**	**79,73%**	**75,16%**

Fonte: os autores (2025).

deslocar atenção e efetivo para regiões que não apresentavam risco real. Portanto, os resultados do modelo devem ser interpretados com cautela, especialmente no que se refere à tomada de decisões operacionais baseadas unicamente nas previsões.

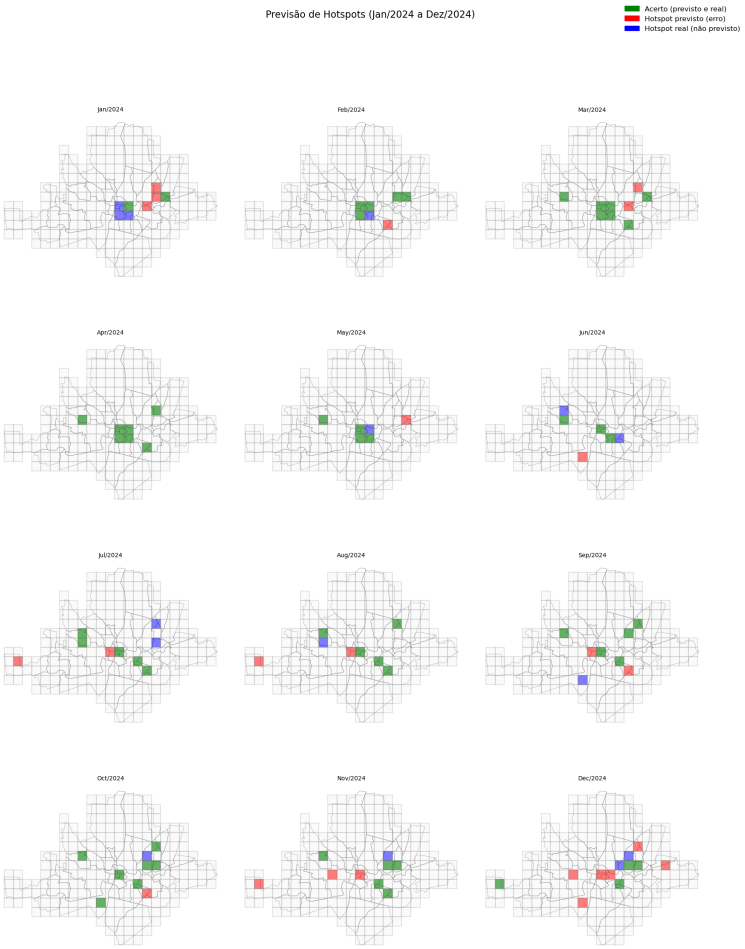
No entanto, cabe destacar que, no contexto da segurança pública e crimes contra a vida, os custos associados a erros do tipo II (não identificar um hotspot real) são, muitas vezes, mais impactantes. Essa lógica está alinhada com a filosofia de atuação de políticas públicas adotadas pelo Estado de Pernambuco, como os programas Pacto pela Vida e Juntos pela Segurança, que utilizam abordagens baseadas na análise sistemática de dados para a redução dos indicadores criminais, com foco principal na diminuição dos homicídios e preservação da vida [Silveira Neto et al., 2022]. Por essa razão, a estratégia adotada para a avaliação o modelo, priorizar a sensibilidade (recall) em relação à precisão, é coerente com a lógica de minimização de crimes contra a vida, ainda que implique em um número maior de falsos positivos. Esse equilíbrio entre sensibilidade e precisão deve ser constantemente avaliado conforme os objetivos e recursos disponíveis.

Seguindo esta perspectiva, e a partir da análise dos hotspots previstos, conforme representado na Figura 3, é possível observar a distribuição dos casos de MVI dentro dos quadrantes classificados como hotspots pelo modelo. A Figura 4 apresenta a frequência com que cada quadrante foi classificado como hotspot de MVI no período de janeiro de 2017 a dezembro de 2023. Nota-se que algumas áreas centrais do município mantêm alta recorrência de registros, o que indica a presença de pontos críticos persistentes ao longo dos anos. No entanto, a distribuição espacial não é estática: diversos quadrantes exibem menor frequência histórica, mas foram previstos como hotspots pelo modelo em 2024, e parte desses foi efetivamente confirmada nos dados reais. Essa característica sugere que o modelo não se limita a reproduzir padrões sazonais ou áreas historicamente críticas, mas também identifica potenciais emergências de novos focos de homicídios.

No total de 99 casos considerados do ano de 2024, 26 ocorreram em áreas previamente identificadas como hotspots, o que demonstra a capacidade preditiva do modelo em antecipar zonas críticas de violência. Sob a ótica do policiamento preditivo, esse resultado indica que, caso medidas efetivas de intervenção tivessem sido aplicadas com base nas previsões, até 26% dos homicídios poderiam ter sido evitados. Embora este seja um cenário idealizado, ele reforça o potencial do modelo como ferramenta de apoio à formulação de políticas públicas voltadas à redução da violência

DOI: 10.59254/sbpo-2025-212301
https://proceedings.science/p/212301?lang=pt-br

Figura 3: Mapa de hotspots reais e previstos (2024)



Fonte: os autores (2025).

letal e à otimização do emprego de recursos operacionais.

5. Considerações finais

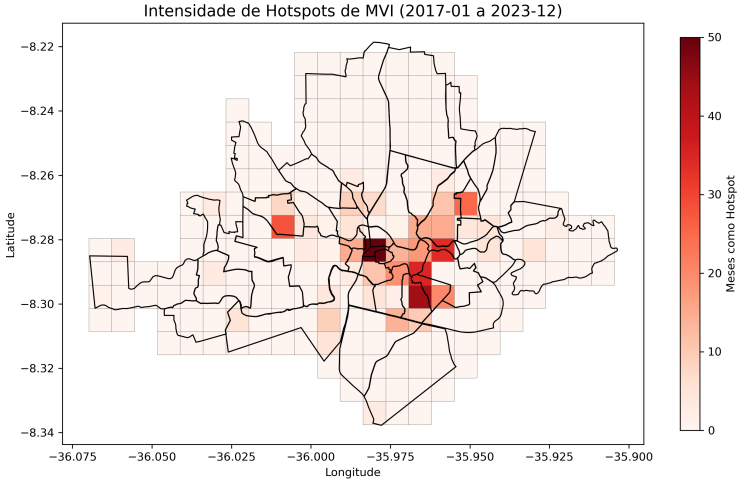
Este trabalho apresentou uma proposta de modelo preditivo para identificação de hotspots de Mortes Violentas Intencionais (MVI) no município de Caruaru, no Agreste de Pernambuco, utilizando algoritmos de aprendizado de máquina aplicados a dados geoespaciais e temporais. O modelo proposto foi desenvolvido a partir de uma base de dados reais sobre os casos de MVI registrados no município durante um período de oito anos, contendo informações detalhadas acerca de cada caso. A partir dessa estrutura, os dados foram agregados no espaço e tempo, possibilitando a aplicação de modelos de previsão direcionados ao aspecto geoespacial.

Os resultados mostraram a capacidade preditiva dos modelos e aplicabilidade da abordagem adotada. Entre os modelos testados, o modelo modelo XGBoost se mostrou o mais adequado, alcançando um F1-score de aproximadamente 75%, com recall de quase 80%. A análise mês a mês revelou um desempenho satisfatório ao longo do período de teste, ano de 2024. Além disso, cerca de 26% dos homicídios registrados no ano ocorreram em áreas antecipadas como hotspots, reforçando o potencial do modelo como ferramenta preventiva.

No entanto, observou-se uma tendência do modelo à geração de falsos positivos, o que, apesar de representar uma necessidade de cautela na utilização dos resultados, conforme argumen-

DOI: 10.59254/sbpo-2025-212301
https://proceedings.science/p/212301?lang=pt-br

Figura 4: Mapa de frequência de hotspots por quadrante



Fonte: os autores (2025).

tado, é preferível a não detectar áreas críticas, considerando o valor imensurável de uma vida, objeto de estudo deste trabalho. Nesse sentido, a avaliação do modelo prioriza a sensibilidade do modelo, alinhando-se aos princípios da preservação da vida.

Como trabalhos futuros, propõe-se: (I) O aprimoramento da definição de hotspot, incorporando abordagens baseadas em densidade espacial (KDE); (II) A integração de novas fontes de dados, como ocorrências policiais não letais, indicadores socioeconômicos e dados de infraestrutura urbana; (III) uso de delimitações espaciais alternativas ao grid regular de quadrantes, considerando fatores geográficos como, comunidades, densidade populacional, zona de comércio de drogas e presença de organizações criminosas, opiniões de autoridades e decisores, pontos de interesse, entre outros; (IV) Incorporar, no modelo, ações das forças de segurança na área de estudo durante o período considerado.

Agradecimentos

Agradecemos à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) pelo apoio financeiro.

Referências

Alsubayhin, A., Ramzan, M. S., e Alzahrani, B. (2024). Crime prediction model using three classification techniques: Random forest, logistic regression, and lightgbm. *International Journal of Advanced Computer Science and Applications*, 15(1). https://doi.org/10.14569/IJACSA.2024.0150123.

Araujo, A. (2019). *Predspot: Predicting Crime Hotspots with Machine Learning*. Dissertação (mestrado em sistemas e computação), Centro de Ciências Exatas e da Terra, Universidade Federal do Rio Grande do Norte.

Azevedo, A. L. V., Riccio, V., e Ruediger, M. A. (2011). A utilização das estatísticas criminais no planejamento da ação policial: cultura e contexto organizacional como elementos centrais à sua compreensão. *Ciência da Informação*, 40(1):9–21. ISSN 0100-1965. https://doi.org/10.1590/S0100-19652011000100001.

DOI: 10.59254/sbpo-2025-212301
https://proceedings.science/p/212301?lang=pt-br

Borges, J., Ziehr, D., Beigl, M., Cacho, N., Martins, A., Sudrich, S., Abt, S., Frey, P., Knapp, T., Etter, M., e Popp, J. (2017). Feature engineering for crime hotspot detection. In *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, p. 1–8.

Braga, A. A. e Weisburd, D. L. (2022). Does hot spots policing have meaningful impacts on crime? findings from an alternative approach to estimating effect sizes from place-based program evaluations. *Journal of Quantitative Criminology*, 38(1):1–22. https://doi.org/10.1007/s10940-020-09481-7.

Brasil (2018). Disciplina a organização e o funcionamento dos órgãos responsáveis pela segurança pública; cria a política nacional de segurança pública e defesa social (pnspds) e institui o sistema Único de segurança pública (susp). *LEI Nº 13.675, DE 11 DE JUNHO DE 2018*. URL `https://www2.camara.leg.br/legin/fed/lei/2018/lei-13675-11-junho-2018-786843-publicacaooriginal-155823-pl.html`. Acesso em: 4 abr. 2025.

Campedelli, G. M. (2022). Explainable machine learning for predicting homicide clearance in the united states. *Journal of Criminal Justice*, 79:101898. ISSN 0047-2352. https://doi.org/10.1016/j.jcrimjus.2022.101898.

Cerqueira, D., Lima, R. S. d. L., Valencia, L. I., Hanashiro, O., Machado, P. H. G., e Lima, A. S. (2017). *Atlas da Violência 2017*. Instituto de Pesquisa Econômica Aplicada (IPEA), Rio de Janeiro.

Dakalbab, F., Abu Talib, M., Abu Waraga, O., Bou Nassif, A., Abbas, S., e Nasir, Q. (2022). Artificial intelligence & crime prediction: A systematic literature review. *Social Sciences & Humanities Open*, 6(1):100342. ISSN 2590-2911. https://doi.org/10.1016/j.ssaho.2022.100342.

Fernandes, A. (2021). *Ordem e segurança pública em São Paulo: análise das estratégias de policiamento da Polícia Militar Estadual*. Tese (doutorado em administração pública e governo), FUNDAÇÃO GETULIO VARGAS.

Fórum Brasileiro de Segurança Pública (2024). *ANUÁRIO BRASILEIRO DE SEGURANÇA PÚBLICA 2024*. Fórum Brasileiro de Segurança Pública, São Paulo, 18 edition. ISSN 1983-7364.

Imbusch, P., Misse, M., e Carrión, F. (2011). Violence research in latin america and the caribbean: A literature review. *International Journal of Conflict and Violence (IJCV)*, 5(1):87 – 154. https://doi.org/10.4119/ijcv-2851.

Kianmehr, K. e Alhajj, R. (2006). Crime hot-spots prediction using support vector machine. In *IEEE International Conference on Computer Systems and Applications, 2006.*, p. 952–959. https://doi.org/10.1109/AICCSA.2006.205203.

Kounadi, O., Ristea, A., Araujo, A., e Leitner, M. (2020). A systematic review on spatial crime forecasting. *Crime Science*, 9(1). ISSN 2193-7680. URL `http://dx.doi.org/10.1186/s40163-020-00116-7`. https://doi.org/10.1186/s40163-020-00116-7.

DOI: 10.59254/sbpo-2025-212301
https://proceedings.science/p/212301?lang=pt-br

Lima, J. V. d. (2024). Segurança pública eficaz, eficiente e efetiva. URL https://www.jusbrasil.com.br/artigos/seguranca-publica-eficaz-eficiente-e-efetiva/2223601024#:~:text=A%20seguran%C3%A7a%20p%C3%BAblica%20%C3%A9%20um,prote%C3%A7%C3%A3o%20dos%20direitos%20dos%20cidad%C3%A3os. Acesso em: 4 abr. 2025.

Limeira, M., Ferreira, L., Junior, M. A., Milan, M., e Donato, R. (2022). Policiamento inteligente: Proposta de modelo de localização preditiva de patrulhas policiais para a prevenção de homicídios. In *Anais do LIV Simpósio Brasileiro de Pesquisa Operacional*. https://doi.org/10.59254/sbpo-2022-157386.

Mandalapu, V., Elluri, L., Vyas, P., e Roy, N. (2023). Crime prediction using machine learning and deep learning: A systematic review and future directions. *IEEE Access*, 11:60153 – 60170. https://doi.org/10.1109/ACCESS.2023.3286344.

Murray, J., Cerqueira, D. R., e Kahn, T. (2013). Crime and violence in brazil: Systematic review of time trends, prevalence rates and risk factors. *Aggression and Violent Behavior*, 18(5):471–483. ISSN 1359-1789. https://doi.org/10.1016/j.avb.2013.07.003.

Perry, W. L., McInnis, B., Price, C. C., Smith, S., e Hollywood, J. S. (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. RAND Corporation. ISBN 9780833081483. http://www.jstor.org/stable/10.7249/j.ctt4cgdcz.

Ratcliffe, J. (2016). *Intelligence-led policing: Second edition*. Intelligence-led policing. ISBN 9781317510406. https://doi.org/10.4324/9781315717579.

Secretaria de Defesa Social de Pernambuco (2025). Portal da secretaria de defesa social de pernambuco. URL https://www.sds.pe.gov.br/. Acesso em: 4 abr. 2025.

Silveira Neto, R., Ratton, J., Menezes, T., e Monteiro, C. (2022). Políticas públicas e redução da criminalidade: uma avaliação do programa pacto pela vida do estado de pernambuco. *Economia Aplicada*, 26:175–202. https://doi.org/10.11606/1980-5330/ea136812.

Wheeler, A. P. e Steenbeek, W. (2021). Mapping the risk terrain for crime using machine learning. *Journal of Quantitative Criminology*, 37(2):445 – 480. https://doi.org/10.1007/s10940-020-09457-7.

Zhang, X., Liu, L., Lan, M., Song, G., Xiao, L., e Chen, J. (2022). Interpretable machine learning models for crime prediction. *Computers, Environment and Urban Systems*, 94:101789. ISSN 0198-9715. https://doi.org/10.1016/j.compenvurbsys.2022.101789.

DOI: 10.59254/sbpo-2025-212301
https://proceedings.science/p/212301?lang=pt-br