

ARTIGO DE PESQUISA/RESEARCH PAPER

Um Framework para Avaliação de Modelos de Previsão Espaço-Temporal de Crimes

Evaluating Crime Prediction in Space and Time: A Practical Framework

Yan Andrade  [Universidade Federal de São João del Rei | yandraade123@aluno.ufsj.edu.br]

George Teodoro  [Universidade Federal de Minas Gerais | ggeorge@dcc.ufmg.br]

Renato Ferreira  [Universidade Federal de Minas Gerais | renato@dcc.ufmg.br]

Leonardo Rocha  [Universidade Federal de São João del Rei | lcrocha@ufsj.edu.br]

 Departamento de Computação, Universidade Federal de São João del Rei, Rodovia 494, s/n, Bairro Colônia do Bengo, São João del-Rei, MG, 36301-360, Brasil.

Resumo. Modelos preditivos espaço-temporais são fundamentais para prever locais e momentos de crimes, auxiliando a segurança pública. No entanto, a ausência de critérios padronizados de avaliação dificulta a comparação entre diferentes abordagens. Para superar essa limitação, apresentamos o STEval, um framework flexível de avaliação. A robustez do STEval foi validada por meio de experimentos que variaram a granularidade temporal e espacial. Os resultados indicaram que nenhum dos modelos supera os demais em todos os cenários. O framework auxilia na seleção e otimização de modelos por meio de análises detalhadas. Além disso, o STEval expõe as forças e fraquezas de cada modelo, incentivando avanços na pesquisa sobre previsão de crimes.

Abstract. Spatio-temporal predictive models are essential for forecasting crime locations and times, supporting public security in resource management. However, the absence of standardized evaluation criteria limits the comparison of different approaches. To overcome this, we present STEval, a flexible evaluation framework. STEval's robustness was validated through experiments varying time and space granularity. Results showed that no model outperforms others in all scenarios, emphasizing context-specific suitability. The framework aids in selecting and optimizing models through detailed analyses. Additionally, STEval exposes each model's strengths and weaknesses, fostering improvements in crime prediction research.

Palavras-chave: Previsão de Crime, Modelos Espaço-Temporais, Segurança Pública, Avaliação de Modelos Preditivos

Keywords: Crime Forecasting, Spatio-Temporal Models, Public Safety, Predictive Model Evaluation

Recebido/Received: 16 May 2025 • **Aceito/Accepted:** 30 June 2025 • **Publicado/Published:** 11 July 2025

1 Introdução

A segurança impacta diretamente o bem-estar pessoal e a estabilidade econômica, sendo a insegurança um fator que reduz a qualidade de vida, limita a mobilidade e afeta a confiança dos investidores, desestabilizando os mercados. A prevenção do crime continua sendo um desafio para os governos, exigindo estratégias baseadas em fatores sociais, demográficos e educacionais Mandalapu *et al.* [2023]. Paralelamente, as agências de segurança pública coletam grandes volumes de dados criminais, enriquecidos com informações espaciais e temporais, viabilizando o desenvolvimento de modelos avançados para previsão e prevenção de crimes otimizando o planejamento de recursos. As abordagens mais utilizadas incluem: (1) modelos de regressão da Silva *et al.* [2020]; (2) agrupamento/séries temporais Butt *et al.* [2021]; (3) métodos baseados em kernel Villegas *et al.* [2022]; e (4) arquiteturas neurais Li *et al.* [2022]. No entanto, a ausência de critérios padronizados dificulta comparações e avanços na área Du and Ding [2023]. Para suprir essa lacuna, propomos o STEval, uma estrutura unificada para avaliar esses modelos em condições experimentais consistentes.

O STEval é dividido em quatro módulos principais. O primeiro realiza a preparação dos dados, incluindo a limpeza, organização e pré-processamento dos mesmos. O segundo define a estrutura espacial, dividindo a área de interesse em

regiões de tamanho igual (malhas). O terceiro é responsável pelo treinamento dos modelos preditivos. Implementamos quatro algoritmos preditivos relacionados a cada abordagem (i.e., Regressão Extra Tree da Silva *et al.* [2020], Estimação de Densidade Espaço-temporal por Kernel - STKDE Zhou and Matteson [2015], Aprendizado Auto-supervisionado em Hipergrafos Espaço-temporais - STHSL Li *et al.* [2022], e modelos auto-regressivos integrados de médias móveis sazonais - SARIMA Catlett *et al.* [2019]) e oferecemos uma estrutura modular e extensível, permitindo a implementação de novas soluções. O quarto módulo é responsável pela avaliação dos modelos. Ele permite configurar diferentes cenários, variando a granularidade temporal e espacial, além de considerar diferentes métricas (ou seja, MSE, RMSE, MAE, Hit-Rate, PAI, PEI e ALS).

Avaliamos o STEval utilizando dados de criminalidade do Brasil, um dos 20 países com as maiores taxas de criminalidade no mundo. Especificamente, analisamos 40 milhões de registros criminais da Polícia Militar do Estado de Minas Gerais (PMMG), abrangendo 853 municípios e incluindo crimes que vão desde crimes violentos (furto, roubo, homicídio) até incidentes menores de segurança pública (discussões, agressões verbais). Para avaliar a flexibilidade e robustez do STEval, realizamos dois cenários principais: (1) variando a granularidade temporal para testar a precisão preditiva em diferentes intervalos de tempo; (2) ajustando a resolução es-

pacial para avaliar a sensibilidade a mudanças no tamanho da malha. Os resultados da avaliação confirmam a capacidade do STEval de lidar com esses cenários variados de forma eficaz. Nenhum dos modelos implementados se destacou em todos os cenários. Os testes de granularidade temporal demonstraram os pontos fortes do STHSL nas previsões de longo prazo, com precisão excepcional mesmo para intervalos temporais distantes do período dos dados de treinamento, e a adequação do SARIMA para previsões de curto prazo. Os testes de granularidade espacial confirmaram a eficácia do STKDE em resoluções mais finas. Os cenários de resiliência temporal destacaram a capacidade do STHSL de fornecer previsões altamente precisas para intervalos de médio e longo prazo, apoiado por suas estimativas probabilísticas confiáveis.

Este estudo revisa os avanços nos modelos espaço-temporais para previsão de crimes, abordando uma lacuna na literatura quanto à sua avaliação. Nossas **principais contribuições incluem:** (1) uma revisão da literatura sobre modelos de previsão de eventos urbanos e suas adaptações para a prevenção de crimes, (2) uma estrutura de avaliação aberta e extensível com implementações dos diversos modelos, e (3) uma comparação experimental abrangente dos modelos mais representativos.

2 Trabalhos Relacionados

A previsão espaço-temporal de crimes auxilia na identificação de áreas e horários de maior risco para apoiar no planejamento das forças de segurança Hu *et al.* [2018]; Zhou and Matteson [2015]. A revisão da literatura sobre modelos de previsão de crimes considera três critérios: adequação dos dados às dimensões espaciais e temporais, custo computacional e efetividade. Os principais modelos identificados incluem regressão, por sua simplicidade e eficiência; agrupamento de séries temporais, para detectar padrões regionais; modelos baseados em kernel, na estimativa de áreas de alto risco; e redes neurais, que oferecem alta precisão, mas com maior custo computacional.

Os modelos de regressão visam correlacionar variáveis preditoras com a variável alvo Alves *et al.* [2018] considerando fatores sociais, econômicos e espaciais, fornecendo suporte interpretável. Em da Silva *et al.* [2020] utilizou-se os modelos Extra Trees e Bagging Regressor para prever locais de roubos em Fortaleza, destacando a conversão eficaz dos dados em coordenadas geográficas precisas. Dentre as abordagens testadas, Random Forest mostrou eficiência na captura de interações complexas. Em nosso framework, implementamos o modelo Extra Trees, que demonstrou desempenho robusto da Silva *et al.* [2020], baixo custo computacional e alta aderência aos dados reais.

A integração de técnicas de agrupamento com a análise de séries temporais tem se mostrado eficaz. Catlett *et al.* [2019] utilizou o DBSCAN em dados de crimes de Chicago e Nova York, revelando que 40% dos crimes em Chicago ocorrem em apenas 20% da cidade. Em seguida, o modelo SARIMA foi aplicado para prever tendências sazonais, com baixos erros absolutos médios. De forma semelhante, Butt *et al.* [2021] empregou o HDBSCAN para analisar padrões criminais em Nova York com excelentes resultados. Em nosso framework, adaptamos o agrupamento para uma estrutura

em grade para maior uniformidade espacial, mantendo o SARIMA para previsão devido à sua capacidade de capturar padrões sazonais e cíclicos.

Modelos baseados em kernel são amplamente utilizados na estimativa de densidades e padrões espaço-temporais devido à sua flexibilidade na modelagem de distribuições complexas. A Estimação de Densidade por Kernel (KDE) e sua variação espaço-temporal (STKDE) são técnicas fundamentais, permitindo a atualização contínua das estimativas com base em observações passadas ponderadas Zhou and Matteson [2015]. O STKDE, foi aplicado na identificação de pontos críticos de criminalidade em Baton Rouge, demonstrando sua versatilidade Hu *et al.* [2018]. Já o Kernel Warping aprimora a estimativa ao integrar dados externos, possibilitando um mapeamento mais preciso de crimes em áreas com poucos registros Villegas *et al.* [2022]. A comparação entre as técnicas reforçou a robustez do STKDE e, por isso, também é adotado neste estudo.

As arquiteturas neurais demonstram serem capazes de capturar padrões complexos em dados não lineares e de alta dimensão Han *et al.* [2020]. Modelos como redes convolucionais, recorrentes, de grafos e hipergrafos aprimoram a modelagem de conectividade e dependências. Duas abordagens se destacam: a Rede de Hipergrafo Sequencial Espaço-Temporal (STSHN) Xia *et al.* [2021], que modela dependências espaciais e temporais em dados criminais por meio de hipergrafos, e o Aprendizado Auto-Supervisionado em Hipergrafo Espaço-Temporal (STHSL) Li *et al.* [2022], que supera 15 métodos concorrentes ao integrar dependências espaciais e temporais em cenários de escassez de dados rotulados, sendo esse o escolhido para compor nosso framework.

3 Framework STEval

O Framework STEval representa a segunda principal contribuição deste trabalho, oferecendo uma abordagem abrangente e padronizada para a avaliação multiperspectiva de modelos de previsão espaço-temporal, com o objetivo de garantir consistência, rigor e ampla aplicabilidade em diferentes contextos e tipos de dados. Organizado em quatro módulos principais (**Figura 1**), o framework inclui: (1) preparação dos dados, onde os dados brutos são carregados, limpos e estruturados para modelagem; (2) geração de malha, que define a estrutura espacial e temporal do problema por meio da criação de malhas regulares para avaliação; (3) treinamento do modelo, envolvendo a aplicação e ajuste de diversos algoritmos para aprendizado a partir dos dados; e (4) avaliação do modelo, onde métricas e critérios específicos são aplicados para medir o desempenho preditivo e a robustez, tudo gerenciado por um arquivo de configuração que define o ambiente experimental.

3.1 Preparação dos Dados

O módulo de preparação de dados é estruturado em três etapas principais: carregamento, pré-processamento e divisão do conjunto de dados em conjuntos de treinamento e teste. Seu principal objetivo é organizar e limpar os dados, garantindo que estejam prontos para o processamento subsequente. Os dados podem ser importados de diretórios locais ou bancos de dados, com suporte para consultas SQL personalizadas a fim de filtrar registros com base em atributos definidos no arquivo

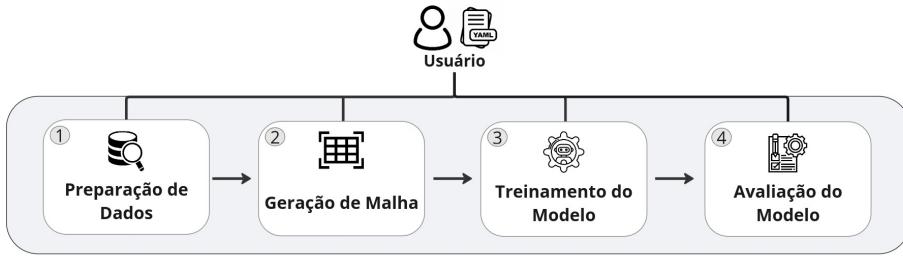


Figura 1. STEval framework

de configuração. Para uma avaliação adequada, os conjuntos de dados devem incluir atributos de latitude, longitude e timestamp. Além disso, a região de estudo deve ser fornecida por meio de um *shapefile* ou pela especificação direta dos limites geográficos da área, utilizando os valores mínimo e máximo de latitude e longitude. Após o carregamento, os dados passam por um processo de pré-processamento para garantir a consistência. Isso inclui a remoção de valores nulos e a padronização das coordenadas geográficas para o sistema de referência espacial EPSG:3857, uma projeção cartográfica métrica amplamente utilizada para análise espacial Battersby *et al.* [2014]. Por fim, os dados são divididos em conjuntos de treinamento e teste conforme os intervalos de datas especificados no arquivo de configuração, que define as datas de início e término de cada período.

3.2 Geração de Malha

Esse módulo divide a região de interesse em malhas de tamanho igual, padronizando o treinamento e a avaliação dos modelos por meio da geração de uma malha espacial que garante consistência na estrutura espacial utilizada durante a modelagem. A malha é criada a partir de um *shapefile* que define a região-alvo ou de limites de latitude e longitude especificados no arquivo de configuração. Por padrão, as células são quadradas e cada célula representa uma unidade espacial uniforme para segmentação de dados e análise em múltiplas escalas. Embora as malhas quadriculadas sejam o padrão, o framework também oferece suporte a formatos alternativos, como malhas hexagonais ou personalizadas, para atender a requisitos específicos de estudo.

3.3 Treinamento do Modelo

O terceiro módulo se concentra no treinamento de modelos preditivos por meio de uma estrutura modular e extensível que padroniza e agiliza o desenvolvimento de novos modelos usando um modelo de classe base. Atualmente, implementa quatro modelos distintos—STKDE, STHSL, Árvore de Decisão e SARIMA—conforme descrito na Seção 2, selecionados para representar as quatro principais classes de abordagens preditivas espaço-temporais e adaptados para previsão em malhas predefinidas. Além disso, o módulo oferece flexibilidade na configuração da granularidade temporal das previsões, permitindo que os pesquisadores selezionem intervalos como horas, dias ou semanas, de acordo com as necessidades experimentais e a resolução dos dados. Isso possibilita tanto previsões de curto prazo, como padrões horários, quanto a identificação de tendências de longo prazo ao longo de semanas.

3.4 Avaliação do Modelo

O módulo de avaliação analisa os modelos preditivos com base na malha gerada, na granularidade temporal selecionada e nos dados de teste. Ele compara as contagens de crimes previstas e reais para cada célula da malha, permitindo uma análise detalhada do desempenho. Para ilustrar a flexibilidade do framework, diversos cenários podem ser configurados ajustando parâmetros como granularidade temporal (por exemplo, horária, diária, semanal), resolução espacial (por exemplo, tamanhos de malha de 100 a 1000 metros) e tamanho do conjunto de dados de treinamento. Além disso, os experimentos podem avaliar a robustez do modelo sob diferentes condições, como lacunas de previsão e variações no tamanho da janela de treinamento, proporcionando uma ampla gama de insights. O módulo aplica métricas de erro, como MSE (Erro Quadrático Médio), RMSE (Raiz do Erro Quadrático Médio) e MAE (Erro Absoluto Médio) Santos-Marquez [2022], para quantificar a precisão preditiva. O desempenho espacial é analisado por meio de métricas como Hit-Rate, PAI (Índice de Precisão Preditiva), PEI (Índice de Eficiência Preditiva) Du and Ding [2023] e ALS (Pontuação Média do Logaritmo) Zhou and Matteson [2015], que avaliam a capacidade do modelo de identificar áreas com alta incidência de crimes e a qualidade das previsões probabilísticas. Juntas, essas métricas fornecem uma avaliação abrangente da precisão estatística e do desempenho espaço-temporal. Além disso, o módulo aplica o teste de postos sinalizados de Wilcoxon Vinagre *et al.* [2021] para validar estatisticamente as diferenças entre modelos ou cenários, garantindo a robustez e confiabilidade dos resultados.

3.5 Arquivos de Configuração

As configurações do STEval são centralizadas em um único arquivo YAML, organizado em três seções principais: `database`, `evaluation` e `models`. A seção `database` define filtros aplicados aos dados, utilizando nomes de colunas como chaves e critérios de seleção como valores, permitindo restrições por atributos como estado, intervalo de datas e área espacial. Além disso, a subseção `columns` especifica os atributos a serem mantidos, exigindo pelo menos longitude, latitude e timestamp para análises espaço-temporais. A seção `evaluation` estabelece parâmetros para avaliação dos modelos, como métricas (HR, MSE, RMSE, ALS e PAI), períodos de treinamento e teste (`train_end_date` e `test_end_date`), granularidade temporal (`temporal_granularity`) e resolução espacial (`grid_size`). A seção `models` lista os modelos a serem treinados e avaliados, cada um definido por sua classe e respectivos hiperparâmetros. Exemplos incluem o `STKDEModel`, que especifica `days_past` e `grid_size_optimization`;

o STARIMA, que define configurações de ordem sazonal; e o ExtraTreesRegressor, que ajusta parâmetros como `n_estimators` e `max_depth`. Essa estrutura modular facilita a personalização dos experimentos, permitindo que pesquisadores configurem facilmente os parâmetros essenciais para análise e avaliação dos modelos no STEval.

```

database:
  filters:
    state: 'MG'
    start_date: '2024-01-01'
    end_date: '2024-08-30'
    neighborhood: ['BELO HORIZONTE']
    crimes: ['theft']
    columns: ['id', 'longitude', 'latitude',
    ← 'timestamp']

evaluation:
  metrics: ['HR', 'MSE', 'RMSE', 'ALS', 'PAI']
  train_end_date: '2024-06-30'
  test_end_date: '2024-07-31'
  temporal_granularity: '1D'
  grid_size: 1000

models:
  STKDEModel:
    days_past: 180
    grid_size_optimization: 1500
  STARIMA:
    seasonal_order: [1, 1, 1, 7]
  STHSLModel:
    batch: 1
  ExtraTreesRegressor:
    n_estimators: 2
    max_depth: 25

```

Configuração 1. Exemplo de arquivo de configuração YAML.

4 Avaliação do STEval: Resultados & Discussões

O STEval é avaliado aqui por meio de experimentação detalhada, comparando quatro modelos espaço-temporais com um conjunto de dados reais de crimes no Brasil, o que corresponde à nossa terceira principal contribuição. Nesta comparação, configuramos dois cenários que nos permitem explorar os modelos sob diferentes condições e restrições, proporcionando uma análise detalhada e padronizada.

4.1 Conjunto de Dados

Foram utilizados dados criminais do Brasil, país com altas taxas de criminalidade segundo o Índice Global de Crime Organizado¹. A análise focou em registros da Polícia Militar de Minas Gerais (PMMG), vinculada à Secretaria de Segurança Pública de Minas Gerais (SSPMG), no âmbito de um projeto de pesquisa em parceria com universidades. O estudo concentrou-se em furtos em Belo Horizonte, com dados do primeiro semestre de 2024 para treinamento e julho de 2024 para teste. A granularidade adotada foi diária, com células espaciais de 1.000 metros, permitindo avaliar metodologias sob diferentes condições para uma análise robusta do desempenho dos modelos. Consideramos os dados de furtos para Belo Horizonte, um total de 72.191 instâncias.

4.2 Cenários de Avaliação

Projetamos dois cenários para avaliar os modelos de previsão espaço-temporal no STEval: (1) Teste de Granularidade Temporal, que mede o desempenho em previsões de 8 horas, diárias (1 dia) e semanal (7 dias), analisando a adaptabilidade a diferentes escalas temporais; e (2) Teste de Granularidade Espacial, que avalia a sensibilidade à resolução espacial, variando o tamanho da célula de previsão (`grid_size`) de 100 a 1000 metros. As avaliações utilizam as métricas HR, ALS, MSE, RMSE e PAI, além do tempo de treinamento, com resultados calculados pela média de dez execuções e desvio padrão. A comparação estatística é feita pelo teste de Wilcoxon com postos sinalizados, destacando os melhores desempenhos em negrito. Os experimentos foram realizados em uma máquina com 62 GB de RAM e duas GPUs NVIDIA GeForce RTX 4090.

4.3 Análise dos Resultados

4.3.1 Teste de Granularidade Temporal

A **Tabela 1** apresenta os resultados para o cenário de granularidade temporal, mostrando que o modelo STHSL alcança as menores métricas de erro (MSE e RMSE) em granularidades mais finas, como 8 horas, enquanto o modelo STKDE supera os demais à medida que a granularidade aumenta para 1 dia e 7 dias, indicando sua eficácia em cenários com maior agregação temporal. No entanto, o erro do STHSL aumenta com granularidades maiores, demonstrando suas limitações em escalas temporais mais amplas. Em termos de métricas de desempenho espacial (HR, ALS e PAI), o modelo STARIMA se destaca em granularidades menores (8 horas e 1 dia) para a detecção de hotspots de curto prazo, enquanto o STHSL mantém resultados consistentes em todas as granularidades, com desempenho superior em 7 dias, evidenciando sua capacidade de modelar dependências locais e globais ao longo do tempo. Embora o modelo Extra Tree Regressor apresente tempos de execução mais rápidos em todas as granularidades, sua precisão é inferior nas granularidades mais finas, mas melhora substancialmente nas maiores, tornando-se uma opção viável para previsões de longo prazo em que a eficiência computacional é crucial.

4.3.2 Teste de Granularidade Espacial

A **Tabela 2** mostra que o modelo STKDE apresenta consistentemente os menores erros em todas as escalas espaciais analisadas (100 a 1000 metros), demonstrando sua eficácia na captura de padrões de densidade e adaptabilidade a diferentes granularidades. No entanto, à medida que a granularidade aumenta, a concentração de crimes por célula cresce, resultando em erros globais mais altos devido à agregação espacial.

O STKDE se destaca na identificação de hotspots em granularidades mais finas, enquanto o STARIMA tem melhor desempenho em escalas maiores (a partir de 300 metros), sendo mais adequado para contextos espaciais agregados. Já o STHSL, apesar de limitado a granularidades acima de 500 metros devido ao seu alto custo computacional, distribui eficientemente as probabilidades, capturando padrões espaciais gerais. Em relação ao tempo de execução, todos os modelos—exceto o Extra Tree Regressor—apresentam maior tempo de processamento em granularidades mais finas (como 200 metros), com redução significativa conforme a granularidade aumenta.

¹<https://ocindex.net/rankings?f=rankings&view=List>

Tabela 1. Resultados do teste de granularidade temporal - Eles demonstram as vantagens do STHSL em previsões de longo prazo, com precisão excepcional mesmo para intervalos fora do domínio de treinamento, e a adequação do SARIMA para previsões de curto prazo.

Granularidade	Modelo	HR	ALS	MSE	RMSE	PAI	Tempo (s)
8H	STKDE	0,6156	-5,7323	0,1026	0,3203	6,1555	130,2977
	STARIMA	0,6363	-5,7935	0,1038	0,3222	6,3633	114,5361
	Extra Tree Reg.	0,1743	-18,2408	0,4570	0,6760	1,7428	4,7941
	STHSL	0,6304	-5,5822	0,1019	0,3192	6,3035	197,7921
1D	STKDE	0,6206	-6,1963	1,8590	1,3634	6,2061	135,2242
	STARIMA	0,6498	-6,1750	2,0668	1,4376	6,4979	50,2652
	Extra Tree Reg.	0,2685	-17,5584	4,5000	2,1213	2,6853	4,7768
	STHSL	0,6308	-6,1299	2,1646	1,4713	6,3076	121,3040
7D	STKDE	0,6132	-6,9497	13,8142	3,7168	6,1319	47,2801
	STARIMA	0,6488	-7,2917	22,2755	4,7197	6,4878	23,1659
	Extra Tree Reg.	0,4243	-17,1042	89,3000	9,4499	4,2430	4,1532
	STHSL	0,6611	-6,9237	37,7471	6,1439	6,6114	11,9333

Tabela 2. Resultados do teste de granularidade espacial. O símbolo “-” em algumas células indica que o algoritmo não pôde ser executado porque o consumo de memória excedeu a memória disponível na máquina. Esses resultados confirmam a eficácia do STKDE em resoluções mais finas.

Granularidade	Modelo	HR	ALS	MSE	RMSE	PAI	Tempo (s)
100	STKDE	0,9949	-8,4494	0,0141	0,1189	9,9505	4005,3209
	STARIMA	0,5871	-13,8180	0,0211	0,1452	5,8717	1634,8434
	Extra Tree Reg.	0,0265	-20,6486	0,0452	0,2126	0,2654	3,0439
	STHSL	-	-	-	-	-	-
200	STKDE	0,6946	-14,2604	0,0516	0,2272	6,9469	730,1817
	STARIMA	0,6383	-11,6229	0,0544	0,2333	6,3843	779,1182
	Extra Tree Reg.	0,0403	-20,1655	0,0858	0,2929	0,4028	4,9851
	STHSL	-	-	-	-	-	-
300	STKDE	0,6473	-11,0579	0,1270	0,3564	6,4760	469,3738
	STARIMA	0,6556	-9,8050	0,1324	0,3639	6,5585	427,1419
	Extra Tree Reg.	0,0721	-19,6754	0,2180	0,4669	0,7210	4,8639
	STHSL	-	-	-	-	-	-
400	STKDE	0,6262	-9,4283	0,2491	0,4991	6,2684	350,0731
	STARIMA	0,6496	-8,6022	0,2551	0,5051	6,5023	259,8383
	Extra Tree Reg.	0,1096	-19,1561	0,4360	0,6603	1,0968	4,8368
	STHSL	-	-	-	-	-	-
500	STKDE	0,6220	-8,3477	0,3646	0,6038	6,2260	284,7334
	STARIMA	0,6634	-7,9618	0,3772	0,6142	6,6399	176,2598
	Extra Tree Reg.	0,1433	-18,7466	0,6980	0,8355	1,4347	4,7629
	STHSL	0,6443	-6,8461	0,3676	0,6063	6,4488	845,1507
600	STKDE	0,6206	-7,6155	0,5589	0,7476	6,2062	236,1839
	STARIMA	0,6673	-7,6037	0,5856	0,7652	6,6735	124,9251
	Extra Tree Reg.	0,1788	-18,3843	1,1000	1,0488	1,7876	4,7562
	STHSL	0,6465	-6,5185	0,5868	0,7660	6,4653	454,7371
700	STKDE	0,6228	-7,1283	0,8235	0,9075	6,2508	193,3892
	STARIMA	0,6544	-7,0947	0,8468	0,9202	6,5680	95,0544
	Extra Tree Reg.	0,2107	-18,1054	1,7200	1,3115	2,1143	4,7602
	STHSL	0,6211	-6,4222	0,8401	0,9166	6,2339	309,5906
800	STKDE	0,6348	-6,7219	1,1534	1,0740	6,3478	172,1977
	STARIMA	0,6598	-6,7145	1,2286	1,1084	6,5979	73,9782
	Extra Tree Reg.	0,2302	-17,8321	2,5200	1,5875	2,3025	4,7613
	STHSL	0,6352	-6,2193	1,2789	1,1309	6,3517	219,3690
900	STKDE	0,6214	-6,4749	1,4782	1,2158	6,2321	149,8627
	STARIMA	0,6454	-6,6403	1,5645	1,2508	6,4727	61,7595
	Extra Tree Reg.	0,2470	-17,7641	3,5282	1,8784	2,4766	4,7478
	STHSL	0,5050	-5,9291	1,9495	1,3962	5,0647	170,7681
1000	STKDE	0,6198	-6,1550	1,8582	1,3632	6,1979	136,5580
	STARIMA	0,6498	-6,1750	2,0668	1,4376	6,4979	50,4199
	Extra Tree Reg.	0,2684	-17,5387	4,5500	2,1331	2,6843	4,7149
	STHSL	0,6346	-5,9533	2,5110	1,5846	6,3461	121,3384

Isso favorece aplicações que demandam respostas rápidas em escalas mais agregadas, mas limita a aplicabilidade dos modelos em grades de alta resolução, especialmente em ambientes com restrições computacionais.

4.4 Discussões e Diretrizes

De modo geral, nenhum modelo espaço-temporal superou os demais em todos os cenários, métricas ou classes. Cada um demonstrou pontos fortes em contextos específicos, sem um domínio absoluto. O STKDE demonstrou maior robustez, com menores erros em granularidades temporais médias e longas (1D e 7D), em todas as escalas espaciais e diferentes volumes de dados ao longo de janelas temporais curtas a médias (1–12 meses), além de oferecer alta precisão espacial em granularidades mais finas. O STARIMA destacou-se na identificação de áreas críticas, especialmente em granularidades espaciais intermediárias (300–1000 metros), enquanto o STHSL apresentou melhor desempenho em granularidades temporais curtas (8h), mantendo solidez em janelas mais longas (7D) e resoluções espaciais ideais (500–1000 metros). Em contrapartida, o Extra Tree Regressor teve desempenho inferior, mas com baixo custo computacional, sendo uma opção viável para cenários com restrições de recursos ou tempo. Esses resultados evidenciam a capacidade do framework em capturar variações de desempenho, reforçando seu potencial como ferramenta estratégica para seleção e otimização de modelos em aplicações reais.

5 Conclusões e Trabalhos Futuros

Este estudo apresenta o STEval, um framework padronizado de avaliação desenvolvido para suprir a ausência de critérios uniformes e replicáveis na modelagem preditiva espaço-temporal. Ao oferecer uma abordagem modular abrangendo a preparação de dados, estruturação espacial, treinamento de modelos e avaliação STEval permite análises abrangentes e comparáveis entre diferentes metodologias. Por meio de testes rigorosos com dados reais de criminalidade no Brasil, o framework demonstrou sua capacidade de capturar variações de desempenho em diversas condições espaciais e temporais. Os resultados evidenciam que nenhum modelo se destaca universalmente, ressaltando a importância da avaliação dependente do contexto. Por exemplo, o STKDE apresentou bom desempenho em cenários espaciais de alta resolução com grande volume de dados, o SARIMA adaptou-se eficazmente a distribuições de dados variáveis, o STHSL demonstrou robustez em previsões de curto prazo, e o Extra Tree Regressor ofereceu eficiência computacional, apesar de menor precisão preditiva.

O STEval representa um avanço significativo na padronização da avaliação de modelos espaço-temporais, promovendo tomadas de decisão mais confiáveis na segurança pública e em outras áreas. Como trabalhos futuros, planejamos incorporar modelos adicionais de diversas categorias metodológicas e ampliar a aplicabilidade do framework para outros fenômenos urbanos, como congestionamento de tráfego e riscos ambientais. Além disso, esforços colaborativos dentro da comunidade científica para padronizar práticas de avaliação podem fortalecer ainda mais o desenvolvimento de modelos preditivos espaço-temporais.

Declarações complementares

Financiamento

Esta pesquisa foi financiada pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) e Secretaria de Segurança Pública de Minas Gerais.

Contribuições dos autores

A concepção do estudo e a definição dos objetivos principais foram realizadas por Yan Andrade e Leonardo Rocha. A metodologia foi desenvolvida por Yan Andrade, sob a orientação de Leonardo Rocha, que também supervisionou tecnicamente o projeto. Os experimentos foram conduzidos por Yan Andrade, e as análises de dados foram realizadas conjuntamente por Yan Andrade e Leonardo Rocha. A aquisição de financiamento foi responsabilidade de Renato Ferreira e George Teodoro, que também coordenaram o projeto. Todos os autores leram e aprovaram o manuscrito final.

Conflitos de interesse

Os autores declaram que não têm nenhum conflito de interesses.

Disponibilidade de dados e materiais

Os conjuntos de software gerados durante o estudo atual estão disponíveis em: <https://github.com/YanAndrade61/STEval>
Os conjuntos de dados analisados durante o estudo atual serão disponibilizados mediante solicitação.

Referências

- Alves, L. G., Ribeiro, H. V., and Rodrigues, F. A. (2018). Crime prediction through urban metrics and statistical learning. *Physica A: Statistical Mechanics and its Applications*. DOI: 10.1016/j.physa.2018.03.084.
- Battersby, S. E., Finn, M. P., Usery, E. L., and Yamamoto, K. H. (2014). Implications of web mercator and its use in online mapping. *Cartographica*, 49(2):85–101. DOI: 10.3138/carto.49.2.2313.
- Butt, U. M., Letchmunan, S., Hassan, F. H., Ali, M., Baqir, A., Koh, T. W., and Sherazi, H. H. R. (2021). Spatio-temporal crime predictions by leveraging artificial intelligence for citizens security in smart cities. *IEEE Access*, 9:47516–47529. DOI: 10.1109/ACCESS.2021.3068306.
- Catlett, C., Cesario, E., Talia, D., and Vinci, A. (2019). Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments. *Pervasive and Mobile Computing*, 53:62–74. DOI: 10.1016/j.pmcj.2019.01.003.
- da Silva, A. R. C., de Paula Júnior, I. C., da Silva, T. L. C., de Macêdo, J. A. F., and Silva, W. C. P. (2020). Prediction of crime location in a brazilian city using regression techniques. In *IEEE 32nd International Conference on Tools with Artificial Intelligence*. DOI: 10.1109/IC-TAIS50040.2020.00059.
- Du, Y. and Ding, N. (2023). A systematic review of multi-scale spatio-temporal crime prediction methods. *ISPRS International Journal of Geo-Information*, 12(6). DOI: 10.3390/ijgi12060209.
- Han, X., Hu, X., Wu, H., Shen, B., and Wu, J. (2020). Risk prediction of theft crimes in urban communities: An integrated model of lstm and st-gcn. *IEEE Access*, 8. DOI: 10.1109/ACCESS.2020.3041924.

- Hu, Y., Wang, F., and Zhu, H. (2018). A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation. *Applied Geography*. DOI: 10.1016/j.apgeog.2018.08.001.
- Li, Z., Huang, C., Xia, L., Xu, Y., and Pei, J. (2022). Spatial-temporal hypergraph self-supervised learning for crime prediction. In 2022 IEEE 38th ICDE. DOI: 10.1109/ICDE53745.2022.00269.
- Mandalapu, V., Elluri, L., Vyas, P., and Roy, N. (2023). Crime prediction using machine learning and deep learning: A systematic review. *IEEE Access*. DOI: 10.1109/ACCESS.2023.3286344.
- Santos-Marquez, F. (2022). Spatial beta-convergence forecasting models: Evidence from municipal homicide rates in colombia. *Journal of Forecasting*, 41(2):294–302. DOI: 10.1002/for.2816.
- Villegas, J. R., Pabón, J. S. M., Dulce Rubio, M., Quintero, S., Vargas, J. G., and García, H. (2022). Spatio temporal sparsity in homicide prediction models. *IEEE Access*, 10. DOI: 10.1109/ACCESS.2022.3143858.
- Vinagre, J., Jorge, A. M., Rocha, C., and Gama, J. (2021). Statistically robust evaluation of stream-based recommender systems. *IEEE TKDE*, 33(7):2971–2982. DOI: 10.1109/TKDE.2019.2960216.
- Xia, L., Huang, C., Xu, Y., Dai, P., Bo, L., Zhang, X., and Chen, T. (2021). Spatial-temporal sequential hypergraph network for crime prediction with dynamic multiplex relation learning. In Proceedings of the Thirtieth IJCAI. DOI: 10.24963/ijcai.2021/225.
- Zhou, Z. and Matteson, D. S. (2015). Predicting ambulance demand: a spatio-temporal kernel approach. In Proceedings of the 21th ACM SIGKDD. DOI: 10.1145/2783258.2788570.