

Ferramenta para análises descritivas e preditivas em dados criminais: Um estudo de caso em Minas Gerais

Yan Andrade
yandrade123@aluno.ufsj.edu.br
UFSJ
Minas Gerais, Brasil

Wagner Meira Jr.
meira@dcc.edu
UFMG
Minas Gerais, Brasil

Gabriel Amarante
gabriel.amarante@dcc.ufmg.br
UFMG
Minas Gerais, Brasil

George Teodoro
george@dcc.edu
UFMG
Minas Gerais, Brasil

Matheus Pimenta
matheuspimenta@dcc.ufmg.br
UFMG
Minas Gerais, Brasil

Leonardo Rocha
lcrocha@ufsj.edu.br
UFSJ
Minas Gerais, Brasil

Renato Ferreira
renato@dcc.edu
UFMG
Minas Gerais, Brasil

ABSTRACT

This article examines the underutilization of detailed criminal data, collaborating with the Military Police of Minas Gerais, Brazil. We propose a new methodology, materialized in a tool, that is able to transform raw data into strategic information for public security decision-making. The tool evaluation unfolds in three phases: characterizing the data, a descriptive analysis of a real case study, and a predictive analysis. This work highlights the untapped potential in detailed criminal data, emphasizing the pivotal role of precise analysis in deciphering complex dynamics. Collaborating with law enforcement aims to bridge the gap between data abundance and actionable insights for effective public security strategies.

KEYWORDS

Modelos Descritivos e Preditivos de Crimes

1 INTRODUÇÃO

A capacidade das organizações de segurança pública para coletar e armazenar dados detalhados de eventos criminais tem experimentado um notável crescimento, resultando na geração diária de grandes volumes. Esses dados, enriquecidos com detalhes espaciais e temporais, oferecem um panorama abrangente da dinâmica criminal. No entanto, dentro desse cenário com excesso de informações, emerge uma problemática, a subutilização desses dados [6].

Neste trabalho apresentamos uma metodologia de caracterização de dados criminais que permite a sua transformação em informações que auxiliem na elaboração de táticas assertivas de combate e prevenção de crimes. Nossa metodologia visa compreender a dinâmica de ocorrência de crimes sob duas perspectivas diferentes, temporal e espacial. A partir dessa compreensão, avaliamos também nesse trabalho a aplicação de modelos de regressão que sejam capazes de correlacionar informações das duas perspectivas consideradas,

auxiliando na predição de onde e quando os crimes ocorrerão no futuro. Toda a metodologia é materializada em uma ferramenta focada em uma interface simples e intuitiva de ser utilizada pelos agentes públicos responsáveis pela segurança.

Instanciamos nossa metodologia a partir dos dados relacionados ao Brasil, utilizando os registros criminais armazenados pela Polícia Militar do Estado de Minas Gerais (PMMG). Dentre os resultados obtidos a partir da instanciação de nossa ferramenta nesta base de dados, destacamos alguns que demonstram o potencial de nosso instrumento. Em uma análise mais específica, comparamos o mapa de crimes antes, durante e após o Carnaval, o mais famoso festival popular brasileiro. Concentramos nossa análise em Belo Horizonte, a capital do Estado de Minas Gerais. Observamos uma mudança significativa no comportamento criminoso durante o Carnaval, com um aumento considerável nos crimes, concentrados em locais onde ocorrem eventos específicos (blocos de carnaval), com uma grande concentração de turistas. Identificar e caracterizar esse comportamento atípico é essencial para desenvolver estratégias específicas de combate e prevenção, auxiliando na alocação mais eficiente de agentes de segurança pública.

Além desta análise descritiva, detalhamos uma análise de dados preditiva em nossos resultados. Também, considerando a cidade de Belo Horizonte durante janeiro, treinamos e avaliamos dois modelos de previsão de crime, um temporal e um espaço-temporal e ambos demonstraram desempenho adequado, apresentando valores previstos próximos aos dados reais. Esses modelos de previsão permitem que a segurança pública desenvolva ações preventivas com mais assertividade. Integrar todos os módulos da ferramenta proposta oferece uma abordagem com análises descritivas e preditivas completas, contribuindo de forma mais eficaz para a melhoria da segurança pública. Esse artigo representa uma extensão de um trabalho previamente publicado [1], no qual acrescentamos uma estratégia preditiva espaço-temporal de crimes.

2 TRABALHOS RELACIONADOS

A análise de dados criminais é objeto de investigação de diversos trabalhos envolvendo a exploração de diferentes técnicas de mineração de dados para identificar padrões e relações entre eventos criminais. Marzan et al. [5] propôs o uso de algoritmos de regras de associação para encontrar correlações entre atributos descritivos e tipos de crime. Os autores identificaram regras que correlacionaram atributos como dia da semana, condições climáticas, feriado, horário do dia, tipo de crime com a localidade da ocorrência. Notavelmente, a maior parte dos crimes tendia a ocorrer em dias sem chuva e em áreas residenciais. As regras encontradas foram contrastadas com um heatmap de ocorrências apresentando uma grande correspondência. Baseado nessa última observação, optamos em nossa ferramenta pelo uso de *heatmaps*.

Outro conjunto de técnicas muito utilizado em caracterizações criminais são as de clusterização e detecção de *outliers*. Walter et al. [8] analisou padrões espaço-temporais de crimes em micro locais (ruas) em seis grandes cidades dos EUA utilizando algoritmos de clusterização. Os autores identificam os micro locais de maior incidência criminal em diferentes momentos no tempo e realizam uma avaliação temporal comparativa entre eles. Em nossa ferramenta, realizamos um agrupamento georreferenciado de ocorrências criminais, identificando e destacando os pontos de maior incidência, o qual chamamos de pontos críticos, no mapa.

Outra linha crescente de trabalhos que visam construir modelos preditivos espaço-temporais [2], com destaque para as séries temporais. Estas são notáveis pela indexação cronológica de suas observações e, conseqüentemente, cruciais na análise de tendências e padrões ao longo do tempo. Esta abordagem divide-se em três componentes principais: a Tendência, que reflete a direção estrutural de longo prazo dos dados; a Sazonalidade, relacionada a padrões repetitivos em intervalos fixos; e o Ruído, representando resíduos após a remoção de sazonalidade e tendência, expressando variação não sistemática nos dados. Catlett et al. [3] empregou o modelo ARIMA [9] para realizar regressão temporal, explorando diferentes parâmetros e seus impactos. Focando na região de Chicago, o objetivo era identificar áreas críticas do estado. Prieto [7] apresentou uma técnica de análise temporal para dados criminais denominada "Heartbeat", que sobrepõem as informações espaciais, temporais e de intensidade, separadas por tipo de crime, em modelo preditivo. Os autores observaram que o "Heartbeat" foi capaz de indicar momentos específicos de maior ocorrência para diferentes tipos de crime.

3 FERRAMENTA DE ANÁLISE CRIMINAL

O objetivo da ferramenta é transformar registros de ocorrências criminais em informações capazes de auxiliar na elaboração de: (1) estratégias assertivas de combate ao crime; e na (2) construção de políticas públicas de prevenção ao crime. Uma visão geral da mesma é apresentada na Figura 1, com detalhamento apresentado a seguir.

3.1 Filtragem de Cenário

O primeiro componente da ferramenta são filtros (item 1 na Figura 1) que permitem analisar diferentes cenários segmentando a representação dos dados de acordo com critérios específicos: **Intervalo de tempo**: intervalo de datas específico; **Granularidade**: diário, mensal ou anual; **Dia da semana**: finais de semana, apenas

dias de semana ou apenas algum dia específico da semana; **Período do dia**: madrugada, manhã, tarde e noite; **Bairros**: bairros específicos dentro de cada município; **Endereço**: local mais preciso; **Tipos de Crime**: roubo, furto, agressão, etc. **Departamento**: delegacias ou postos policiais; e **Unidade**: batalhão, companhia. A partir desses filtros, diversos cenários podem ser construídos. Por exemplo, investigar como é a distribuição de crimes do tipo "roubo", no bairro "X" nas "madrugadas" durante os finais de semana dos meses de "Janeiro" à "Abril", contrapondo com o cenário parecido, porém avaliando os "dias de semana".

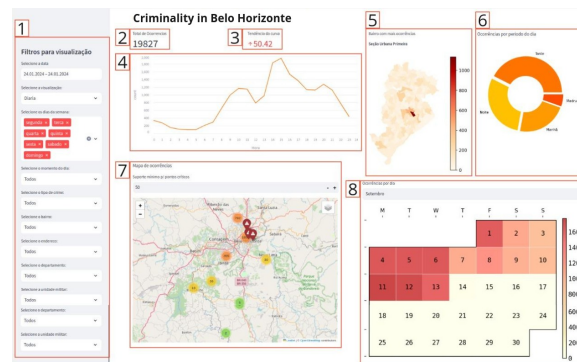


Figura 1: Tela inicial da ferramenta

3.2 Análises de Tendências

Esse módulo centra-se em fornecer uma compreensão abrangente das tendências temporais dos eventos. O primeiro campo (item 2 na Figura 1) exibe o número total de incidentes com base nos filtros utilizados, permitindo ver imediatamente o quantitativo de incidentes. O segundo componente (item 4 na Figura 1) destaca visualmente o comportamento do evento, permitindo análises cumulativas ao longo do tempo. Para essa curva, a ferramenta utiliza um algoritmo de regressão linear de pontos, utilizando a derivada da função resultante para determinar se a tendência está aumentando ou diminuindo (item 3 na Figura 1).

3.3 Análise de Distribuição

Esse módulo foi projetado para fornecer uma visão detalhada dos aspectos espaciais e temporais dos eventos criminais, incluindo distribuição atualizada por bairro (item 5 na Figura 1) e horário do dia (item 6 na Figura 1). Cada uma dessas distribuições representa aspectos criticamente importantes dos dados, permitindo uma análise mais profunda das nuances espaciais e temporais dos eventos.

3.4 Mapa de Calor Temporal

Esse módulo combina dois conceitos, um mapa de calor (item 7 na Figura 1) que destaca as áreas de maior concentração e um calendário (item 8 na Figura 1) que exibe os dias do mês, proporcionando uma representação visual dinâmica da concentração de incidentes em cada dia do mês, permitindo uma análise comparativa de padrões temporalmente recorrentes.

3.5 Mapa de Eventos em Camadas

Um dos pilares fundamentais da nossa ferramenta é a capacidade de analisar dados com base nas suas propriedades georreferenciadas.

Para isso, adotamos uma apresentação cartográfica por meio de mapas (item 7 na Figura 1). Implementamos quatro camadas distintas para fornecer informações sobre a geografia do crime visando proporcionar uma experiência rica e completa em análise espacial de eventos. A primeira camada demarca os bairros da cidade (Figura 2 a). A segunda é um mapa de calor (Figura 2 b) destacando densidade de incidência criminal. A terceira camada marca os pontos que aparecem no mapa (Figura 2 c agregando data, hora e tipo de evento). A quarta camada destaca pontos críticos (Figura 2 d), identificados por meio de um processo de agrupamento baseado em um limite de distância máxima entre pontos. É importante ressaltar que todas essas camadas estão integradas em um único mapa, permitindo que as camadas sejam modificadas e combinadas conforme necessário.

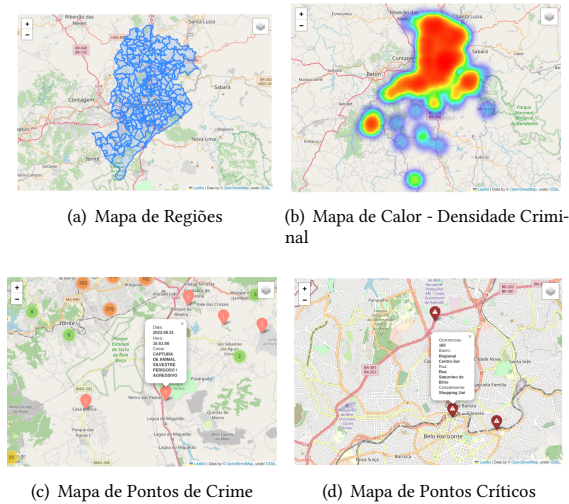


Figura 2: Detalhamento das Camadas do Mapa de Eventos

3.6 Modelos de Regressão

A aplicação de modelos de regressão para *forecasting* tem se mostrado uma ferramenta promissora. Do ponto de vista da análise temporal, destacamos o modelo ARIMA[4]. Sua versatilidade reside na habilidade de capturar tendências, sazonalidades e padrões temporais em uma variedade de domínios, contribuindo significativamente para análises preditivas temporais. Do ponto de vista da predição espaço-temporal, destacamos o st-KDE (spacial-temporal KDE) [10] que calcula a densidade espacial contínua de um evento (i.e. crimes), em cada instante de tempo por meio de um somatório das observações passadas de forma ponderada (i.e. *time-series*). Ambos os modelos estão implementados em nossa ferramenta e têm se mostrado, conforme veremos na seção seguinte, muito promissores no auxílio às forças policiais no planejamento estratégico e alocação eficiente de recursos.

4 AVALIAÇÃO EXPERIMENTAL

Para avaliar a aplicabilidade da ferramenta de análise criminal proposta nesse artigo, instanciamos-a utilizando os registros criminais armazenados pela Polícia Militar do Estado de Minas Gerais (PMMG), um dos estados mais violentos do Brasil. A PMMG está subordinada à Secretaria de Segurança Pública de Minas Gerais (SSPMG), a qual mantém atualmente, um Projeto de Pesquisa, Desenvolvimento e Inovação (PD&I) com as universidades dos autores

do presente artigo. Nossa avaliação divide-se em duas partes complementares: a aplicação prática da ferramenta, permitindo uma interação dinâmica de diferentes cenários; e a aplicação e testes de modelos preditivos.

4.1 Aplicação Prática

Nessa seção visamos apresentar estudos de caso que demonstrem a aplicabilidade de ferramenta focando na compreensão do comportamento criminal durante o Carnaval de 2023 e em como a mesma pode auxiliar na tomada de decisão. A ferramenta permite identificar padrões e tendências, que podem ser usados para definir pontos de prioridades para alocação de recursos de segurança, como escala de patrulhas e definição de rotas. Analisaremos os dados do Bairro Seção Urbana Primeira, em Belo Horizonte, entre os dias 4 a 26 de fevereiro. A Figura 3 apresenta o módulo de tendências englobando o primeiro trimestre de 2023, onde fica evidente uma alta expressiva da criminalidade no mês de fevereiro. Outro ponto a destacar é a semelhança das ocorrências nos meses de janeiro e março, períodos que antecedem e sucedem o Carnaval, destacando o mês de fevereiro.



Figura 3: Módulo de tendências da ferramenta para o primeiro trimestre de 2023



Figura 4: Comparativo entre a distribuição espacial de ocorrências no bairro SUP, antes, durante e após o Carnaval 2023

A Figura 4 representa três momentos distintos: antes, durante e após o Carnaval. Nota-se um aumento significativo de crimes

na região central durante o Carnaval, Figura 4 (b), destacando a presença de pontos críticos nesse período. Esses pontos críticos apresentam alta relação com a localização dos tradicionais blocos de carnaval em BH, ressaltando, portanto, áreas de maior perigo que devem ter maior atenção. Outro ponto a destacar é que, assim como acontece em números de ocorrência, os pontos de criminalidade antes e depois do Carnaval, Figura 4(a e c), tendem a manter um padrão. A ferramenta permite distribuir recursos e escalar patrulhas de maneira mais eficiente, em lugares que demandam atenção, tudo baseado em filtros que podem definir múltiplos cenários.

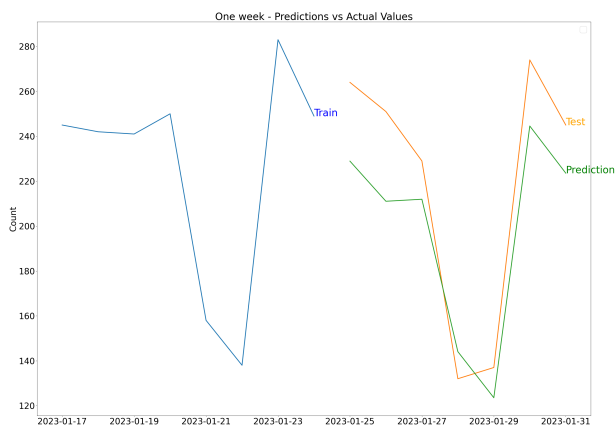


Figura 5: Modelo Arima; Em azul o último intervalo de treino, em verde a previsão (em verde-claro o intervalo de confiança) e em laranja/vermelho o valor real

Para ilustrar o funcionamento do módulo de regressão, testamos os algoritmos ARIMA e st-KDE. Utilizamos dados de furtos de 01/01/2023 a 24/01/2023 para treinar o ARIMA e dados de novembro de 2023 para o st-KDE. Previmos crimes para a semana de 25/01/2023 a 31/01/2023 com o ARIMA (Figura 5). Para o st-KDE, fizemos previsões por setores para a primeira semana de dezembro de 2023 (Figura 6 e Tabela 1¹). Ambos os modelos mostraram desempenho satisfatório, com previsões próximas aos dados reais. A integração desses modelos oferece uma abordagem completa para análises preditivas, contribuindo de forma mais eficaz para o aprimoramento da segurança pública.

Metricas	MSE	MAE	ALS
st-KDE	0.00055	0.01079	-4.15952

Tabela 1: Resultados referente ao modelo st-KDE

5 CONCLUSÃO

Neste trabalho, apresentamos uma ferramenta para análise de dados criminais que é capaz de caracterizar os dados e transformá-los em informações que auxiliam no desenvolvimento de táticas assertivas de combate e prevenção ao crime. Nossa ferramenta compreende a dinâmica da ocorrência de crimes a partir de duas perspectivas diferentes: temporal e espacial, considerando dois tipos de análise: discretiva e preditiva, apresentando uma interface simples e intuitiva para ser utilizada por agentes públicos responsáveis pela segurança.

¹As métricas medem o erro do modelo, para o MSE e MAE quanto menor melhor, no caso do ALS quanto mais próximo de zero melhor.

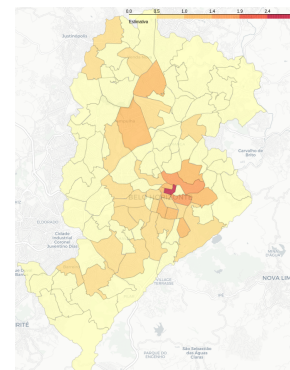


Figura 6: Modelo st-KDE; Escala da cor amarela (menor concentração) para vermelha (mais concentração).

Para avaliar a aplicabilidade da ferramenta, instanciamos-a utilizando dados criminais armazenados pela Polícia Militar do Estado de Minas Gerais (PMMG). Realizamos duas análises descritivas e preditivas, considerando dados de Belo Horizonte. Na análise descritiva, comparamos o mapa de crimes antes, durante e após o Carnaval, demonstrando uma mudança significativa no comportamento criminoso, com um aumento de crimes concentrados em pontos turísticos específicos. Na análise preditiva, treinamos e avaliamos dois modelos de previsão de crimes, um temporal e outro espaço-temporal onde ambos apresentam valores previstos próximos aos dados reais, permitindo consequentemente, que a segurança pública desenvolva ações preventivas de forma mais assertiva.

AGRADECIMENTOS

Este trabalho foi financiado por CNPq e Fapemig.

REFERÊNCIAS

- [1] Yan Andrade, Matheus Pimenta, Gabriel Amarante, Antônio Hot Faria, Marcelo Vilas-Boas, João Paulo da Silva, Felipe Rocha, Jamichel da Silva, Wagner Meira Jr., George Teodoro, Leonardo Rocha, and Renato Ferreira. 2023. A descriptive and predictive analysis tool for criminal data: A case study from Brazil. In *Computational Science and Its Applications - ICCSA 2024 (Lecture Notes in Computer Science)*. Springer.
- [2] Fateha Khanam Bappee, Amílcar Soares Júnior, and Stan Matwin. 2018. Predicting crime using spatial features. In *Canadian Conference on Artificial Intelligence*. Springer.
- [3] Charlie Catlett, Eugenio Cesario, Domenico Talia, and Andrea Vinci. 2018. A data-driven approach for spatio-temporal crime predictions in smart cities. In *2018 IEEE International Conference on Smart Computing*.
- [4] Eugenio Cesario, Charlie Catlett, and Domenico Talia. 2016. Forecasting crimes using autoregressive models. In *14th Intl Conf on Dependable, Autonomic and Secure Computing*. IEEE.
- [5] Charlie S Marzan, Maria Jeseca C Baculo, Remedios de Dios Bulos, and Conrado Ruiz Jr. 2017. Time series analysis and crime pattern forecasting of city crime data. In *Proceedings of the 1st International Conference on Algorithms, Computing and Systems*. 113–118.
- [6] Colleen McCue. 2014. *Data mining and predictive analysis: Intelligence gathering and crime analysis*. Butterworth-Heinemann.
- [7] Rafael Prieto Curiel. 2023. Weekly crime concentration. *Journal of quantitative criminology* 39, 1 (2023), 97–124.
- [8] Rebecca J Walter, Marie Skubak Tillyer, and Arthur Acolin. 2023. Spatiotemporal crime patterns across six US cities: Analyzing stability and change in clusters and outliers. *Journal of Quantitative Criminology* 39, 4 (2023), 951–974.
- [9] G Peter Zhang. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50 (2003), 159–175.
- [10] Zhengyi Zhou and David S. Matteson. 2015. Predicting Ambulance Demand: a Spatio-Temporal Kernel Approach. In *Proceedings of the 21th ACM SIGKDD*. <https://doi.org/10.1145/2783258.2788570>