

Planejamento Estratégico do Posicionamento Policial em Estádios de Futebol com Análise Multicritério e Clustering Não Supervisionado

**Antonio Marcos de Lima¹, Thyago Celso Cavalcante Nepomuceno¹ e Jean
Turet¹**

Universidade Federal de Pernambuco. Programa de Pós Graduação em
Engenharia de Produção - CAA. Av. Marielle Franco, s/n - Km 59 – Caruaru -
PE

Resumo. *Este estudo propõe uma metodologia que combina auxílio à decisão multicritério (MCDA) e análise de cluster para otimizar a localização de unidades policiais e estacionamentos em Recife, Brasil, durante eventos esportivos de risco. O algoritmo K-means agrupa ocorrências criminais em clusters homogêneos, e o método PROMETHEE hierarquiza as melhores localizações com base em três critérios: número de instalações (N), distância média das ocorrências (D) e nível de cobertura (C). A análise de dados criminais identifica áreas de alto risco, auxiliando na alocação estratégica de recursos e no aprimoramento da segurança pública que impacta diretamente a sociedade.*

Abstract. *This study proposes a methodology combining multicriteria decision aid (MCDA) and cluster analysis to optimize the placement of police units and parking facilities in Recife, Brazil, during high-risk sports events. The K-means algorithm groups crime occurrences into homogeneous clusters, while the PROMETHEE method ranks the best locations based on three criteria: number of facilities (N), average distance of occurrences (D), and coverage level (C). Crime data analysis identifies high-risk areas, supporting strategic resource allocation and enhancing public security.*

1. Introdução

O futebol é um evento social que congrega indivíduos de diversas origens culturais, sociais e econômicas. Este fenômeno é impulsionado por uma gama variada de motivações e paixões, tanto subjetivas quanto objetivas [Murad, 2017]. No entanto, a prática do futebol frequentemente resulta em incidentes de violência e delitos, tanto nos estádios quanto nas áreas circundantes, particularmente em países onde o futebol possui uma relevância cultural significativa [Palhares, 2012]. Os locais esportivos são comumente palco de confrontos entre torcidas rivais, bem como de vandalismo, furtos e outras manifestações de violência. Pesquisas pertinentes têm analisado o comportamento violento dos torcedores no Brasil [NEPOMUCENO et al., 2017; NEPOMUCENO et al., 2022].

Para mitigar e prevenir a violência nas proximidades dos estádios, é fundamental a adequada localização das unidades policiais. A proximidade dessas unidades assegura a

proteção tanto dos torcedores quanto dos residentes da área, permitindo uma resposta rápida a conflitos ou atos violentos. Além disso, a presença policial tem o potencial de desencorajar a ocorrência de incidentes violentos. Este estudo propõe um método híbrido de análise espacial e multicritério para identificar locais estratégicos para a instalação de unidades policiais. Considera-se uma unidade policial como um veículo com até quatro policiais, com o objetivo de responder de maneira eficiente e rápida às demandas.

O método híbrido sugerido combina as ferramentas PROMETHEE e K-means para identificar locais estratégicos para a instalação de unidades policiais, com o intuito de aprimorar a segurança pública e reduzir a violência nas áreas ao redor dos estádios em Recife, Brasil. Este método é eficaz com diversos idiomas e softwares e possui uma base estatística robusta. A necessidade de uma análise exploratória, sem categorias ou rótulos pré-definidos, justifica a utilização do clustering não supervisionado. A complexidade e dinamicidade das atividades criminosas em Recife tornam os dados de criminalidade difíceis de definir. Os métodos de clustering não supervisionado são adaptáveis a diferentes tipos de dados, permitindo a identificação de similaridades e diferenças entre as categorias de crimes e suas localizações.

2. Problema de localização de instalações (*Facility Location Problem*)

Em áreas como logística, transporte e gestão de serviços públicos, a localização de novas unidades de serviço ou equipamentos é um problema de grande relevância, conhecido como o problema de localização de instalações (FLP, do inglês Facility Location Problem). O objetivo central desse problema é determinar a melhor estratégia para posicionar um número específico de serviços, com o intuito de minimizar custos e maximizar a eficiência do serviço prestado [FERNANDES et al., 2005].

O FLP é uma questão complexa que deve considerar diversos fatores, incluindo o perfil socioeconômico da população atendida, as características das instalações, as capacidades disponíveis e a demanda por serviços. Além dos desafios técnicos e econômicos, a localização de instalações impacta diretamente a inclusão social, a equidade no acesso aos serviços públicos e a qualidade de vida da população. Em comunidades vulneráveis, a distância ou a falta de acesso adequado a serviços essenciais pode acentuar desigualdades e comprometer o desenvolvimento social.

A literatura especializada apresenta vários métodos para resolver o FLP, abrangendo desde modelos matemáticos que buscam soluções exatas até algoritmos heurísticos que oferecem soluções aproximadas em um tempo razoável. A seleção de locais para instalações tem sido abordada por muitos estudos que utilizam simulações computacionais e análises matemáticas [ADELEKE et al., 2020]. No entanto, é essencial que essas abordagens considerem fatores humanos e sociais, como a acessibilidade para grupos vulneráveis, a segurança da população e o impacto na mobilidade urbana.

A solução do FLP pode trazer diversos benefícios à sociedade, como a melhoria da qualidade dos serviços, a redução de custos, a otimização da distribuição de recursos e a elevação do bem-estar dos usuários. Um planejamento eficiente pode garantir que serviços de saúde, segurança e transporte estejam distribuídos de forma equitativa, reduzindo desigualdades regionais e promovendo maior justiça social. Embora não seja possível controlar eventos indesejáveis como roubos, assaltos e vandalismo, é desejável

que esses eventos ocorram o mais distante possível da população e que os serviços essenciais estejam mais próximos das áreas que mais necessitam deles.

Para abordar problemas relacionados à segurança pública e questões socioeconômicas, diversas ferramentas quantitativas e tecnologias da informação podem ser empregadas. Neste estudo, o algoritmo de clustering K-means é combinado com o modelo PROMETHEE para resolver o problema de localização das instalações, garantindo não apenas eficiência operacional, mas também um impacto positivo no desenvolvimento social e na segurança das comunidades atendidas [NEPOMUCENO et al., 2023; DE CARVALHO et al., 2023; BOEZI et al., 2023].

3. Materiais e Métodos

3.1 K-Means

O K-means é um algoritmo de clustering não supervisionado amplamente utilizado para agrupar dados em clusters homogêneos com base em suas características. Devido à sua simplicidade e eficiência, é um dos algoritmos mais comuns para a classificação de clusters (CABRERA-BARONA, 2020).

O algoritmo K-means opera de forma iterativa, começando com a seleção aleatória de k centróides que servem como núcleos dos clusters. Cada ponto de dado é atribuído ao cluster cujo centróide está mais próximo. Após a atribuição de todos os pontos, os centróides são recalculados com base nos dados do respectivo cluster. Esse processo é repetido até que não ocorram mais alterações nas atribuições dos pontos ou até que o número máximo de iterações seja atingido.

Embora o K-means seja um método de agrupamento eficaz e relativamente simples, ele apresenta algumas limitações, como a dependência da inicialização aleatória dos centróides e a sensibilidade na escolha do número de clusters k. Esses problemas podem ser mitigados por meio de técnicas como a inicialização através de amostragem hierárquica e a realização de múltiplas inicializações aleatórias. Para aprimorar a classificação de áreas urbanas vulneráveis, o K-means pode ser integrado ao método PROMETHEE.

3.2 PROMETHEE

O método PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluation) é uma abordagem que orienta o processo de tomada de decisão, permitindo que as opções sejam organizadas em uma hierarquia com base nas preferências do tomador de decisão (MARESCHAL et al., 2009). Desenvolvido por Brans et al. (1986), essa técnica ajuda a classificar ações de acordo com critérios estabelecidos, permitindo uma análise mais precisa e estruturada das possibilidades disponíveis. Modelos de tomada de decisão não compensatórios, como o PROMETHEE, têm sido bem utilizados na literatura para avaliar a satisfação no trabalho e a alocação de recursos (NEPOMUCENO et al., 2020; NEPOMUCENO et al., 2021).

O método PROMETHEE pode ser definido em dois passos. No primeiro passo, é feita uma comparação usando funções de preferência de diferentes tipos, como funções de diferença e funções de similaridade. Essas comparações geram uma matriz de pré-ordem

que indica a relação de preferência entre as alternativas em relação a cada critério individualmente. No segundo passo, a pessoa responsável pela decisão atribui pesos aos critérios de acordo com as prioridades, de modo que os critérios mais relevantes recebam pesos mais significativos. É necessário atribuir pesos a cada critério onde cada critério i deverá ter um peso p_i . Assim, o grau de sobreclassificação dado pela equação pode ser calculado da seguinte forma:

$$\pi(a, b) = \sum_{i=1}^n p_i F_i(a, b)$$

Onde:

n = número de critérios avaliados.

$\pi(a,b)$ = grau de sobreclassificação da alternativa a em relação a alternativa b.

F_i = função de preferência do critério i

p_i =peso do critério i

Com base no grau de superação de todas as alternativas, os fluxos das alternativas podem ser calculados. Existem três tipos de fluxo: fluxo positivo, fluxo negativo e fluxo líquido. O fluxo positivo $\emptyset^+(a)$ mede o quanto uma ação é preferida em relação às outras $n - 1$ alternativas. De forma análoga, o fluxo negativo $\emptyset^-(a)$ indica o quanto as $n - 1$ alternativas são preferíveis. O fluxo líquido $\emptyset(a)$ é obtido subtraindo o fluxo negativo do fluxo positivo. Isso basicamente mostra como a alternativa se comporta em comparação com as outras. Quanto maior o fluxo líquido, mais preferível é a alternativa. Os fluxos são calculados pelas equações:

$$\emptyset^+(a) = \frac{1}{n-1} \sum_{b \neq a} \pi(a, b)$$

$$\emptyset^-(a) = \frac{1}{n-1} \sum_{b \neq a} \pi(b, a)$$

$$\emptyset(a) = \emptyset^+(a) - \emptyset^-(a)$$

Onde:

$\emptyset^+(a)$ = fluxo positivo da alternativa a

$\emptyset^-(a)$ = fluxo negativo da alternativa a

$\emptyset(a)$ = fluxo líquido da alternativa a

b=são as n-1 alternativas

4. Resultados e discussões

O aprendizado virtual desempenha um papel crucial na otimização da provisão e alocação de recursos, oferecendo insights valiosos sobre o cenário socioeconômico [NEPOMUCENO, 2023; POLETO et al., 2023; JI et al., 2022]. Ao analisar diversos conjuntos de dados, como padrões de crimes, informações demográficas e indicadores econômicos, os formuladores de políticas podem tomar decisões mais informadas para enfrentar desafios específicos e alocar recursos de forma eficaz.

A SEPLAG (Secretaria de Planejamento e Gestão do Estado) disponibilizou um banco de dados contendo informações sobre casos de CVP (crimes contra o patrimônio) e suas respectivas geolocalizações em todo o estado de Pernambuco, cobrindo o período de 2018 a 2021. Este banco de dados inclui casos como furtos em instituições bancárias, furtos em transporte público, assaltos, roubos de carga, invasões domiciliares e furtos comerciais. Para uma descrição mais detalhada desses crimes georreferenciados e do contexto em que ocorrem, consulte [NEPOMUCENO e COSTA, 2019, DE CARVALHO E COSTA, 2023, BORBA et al. 2022].

Mesmo na ausência de dados sobre outros tipos de crimes, como crimes físicos, sexuais, psicológicos ou morais, a presença de dados sobre roubos pode servir como um indicador útil de violência em áreas urbanas. Isso se deve ao fato de que roubos frequentemente estão associados a atos de violência, como agressões físicas, ameaças e intimidações. A Figura 1 apresenta um gráfico de crimes no Recife, criado com o software ArcGIS versão 10.8, utilizando o OpenStreetMap como mapa base.

Na fase de análise e priorização territorial, foi necessário atribuir pesos distintos aos critérios considerados, como densidade populacional, incidência de CVPs, vulnerabilidade social e acessibilidade a serviços públicos. A definição desses pesos foi realizada por meio de um processo combinado de consulta técnica e revisão bibliográfica, com apoio de especialistas em segurança pública e planejamento urbano.

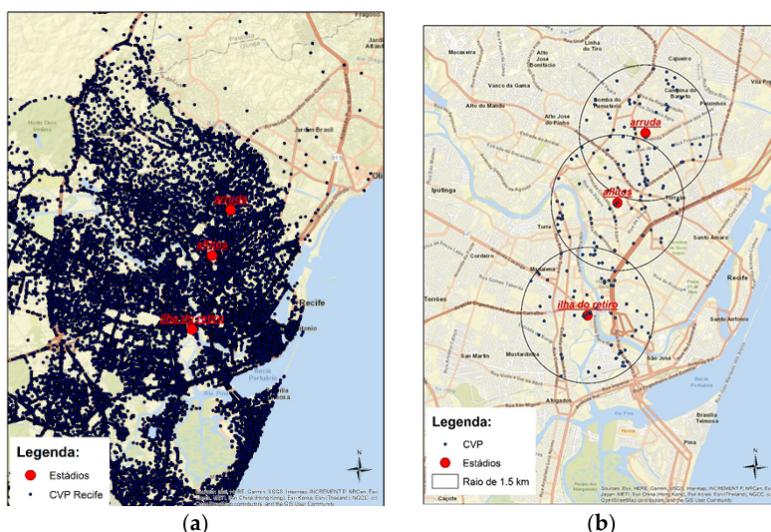


Figura 1. Ocorrências de CVP (crimes contra o patrimônio): (a) ocorrências totais no Recife; (b) ocorrências ao redor dos estádios (até 1,5 km)

A Tabela 1, apresenta a proporção de ocorrências para cada estádio com base nos tipos de crimes.

Tabela 1. Proporção de CVPs (crimes contra o patrimônio)

	Arruda	Ilha Do Retiro	Aflitos
Outros Roubos	13,5%	2,9%	6,7%
Roubo A Transeunte	76,9%	88,4%	81,3%
Roubo Em Residência	3,8%	1,4%	0,0%
Roubo A Ônibus	3,8%	0,0%	1,3%
Roubo Em Estabelecimento Comercial Ou De Serviços	1,9%	5,9%	10,7%
Roubo A Outros Transportes Coletivos	0%	1,4%	0%

Fonte: Os Autores, 2024

Para minimizar o problema da escolha inicial dos centróides, foi adotada a técnica de inicialização aleatória múltipla, que é padrão na biblioteca Scikit-learn. O algoritmo é executado com diferentes sementes para os centróides, conforme o número de vezes especificado. A semente que resulta na menor soma das distâncias quadradas dos pontos aos centróides dos clusters é escolhida. Neste estudo, cada execução do algoritmo foi realizada com 5000 sementes distintas. O processo foi conduzido em um notebook com um processador Intel(R) Core(TM) i3-3217U, com frequência de 1,80 GHz e 10 GB de memória RAM, levando 1 hora e 43 minutos para ser concluído.

Foram criados cenários para identificar a melhor alocação de instalações com base no número de instalações (N), distância média (D) e nível de cobertura (C). Os pesos para cada critério foram determinados pela técnica ROC, conforme apresentada por Barron (1992), levando em conta a ordem de preferência dos critérios. A Tabela 2 a seguir apresenta o número de instalações (localizações) para as cinco melhores alternativas classificadas para cada estádio, considerando todos os seis cenários: o primeiro cenário (C, D, N), o segundo cenário (C, N, D), o terceiro cenário (D, C, N), o quarto cenário (N, C, D), o quinto cenário (D, N, C) e o sexto cenário (N, D, C).

Tabela 2. Melhores alternativas em cada cenário

Estádio	Cenários						Ranking
	1	2	3	4	5	6	
Aflitos	30 Locais	11 Locais	30 Locais	7 Locais	26 Locais	1 Local	1º
	26 Locais	10 Locais	29 Locais	8 Locais	28 Locais	2 Locais	2º
	29 Locais	12 Locais	28 Locais	9 Locais	27 Locais	7 Locais	3º
	28 Locais	13 Locais	27 Locais	10 Locais	29 Locais	3 Locais	4º
	27 Locais	14 Locais	26 Locais	1 Locais	30 Locais	6 Locais	5º
Ilha do Retiro	30 Locais	10 Locais	30 Locais	7 Locais	30 Locais	1 Local	1º
	29 Locais	11 Locais	29 Locais	6 Locais	29 Locais	2 Locais	2º
	28 Locais	12 Locais	28 Locais	8 Locais	26 Locais	6 Locais	3º
	27 Locais	9 Locais	27 Locais	9 Locais	28 Locais	7 Locais	4º
	26 Locais	14 Locais	26 Locais	5 Locais	27 Locais	3 Locais	5º
Arruda	30 Locais	8 Locais	30 Locais	6 Locais	30 Locais	4 Locais	1º
	28 Locais	10 Locais	29 Locais	7 Locais	28 Locais	1 Local	2º
	29 Locais	11 Locais	28 Locais	5 Locais	29 Locais	5 Locais	3º
	27 Locais	7 Locais	27 Locais	4 Locais	27 Locais	6 Locais	4º
	26 Locais	9 Locais	26 Locais	8 Locais	26 Locais	7 Locais	5º

Fonte: Os Autores, 2024

Importante frisar que a utilização de dados criminais georreferenciados requer uma reflexão ética cuidadosa. É fundamental considerar o risco de viés nos dados, uma vez que áreas com maior policiamento podem apresentar registros inflacionados de crimes (overpolicing), enquanto regiões com menor presença policial ou desconfiança institucional podem sofrer com subnotificação. Além disso, há preocupações relevantes quanto à privacidade, especialmente em contextos de exposição indevida de comunidades específicas. Para mitigar esses riscos, foram adotadas medidas como a anonimização dos dados sensíveis, a agregação por áreas geográficas maiores (em vez de pontos individuais) e a validação cruzada com fontes independentes para reduzir os efeitos da subnotificação. Também foram considerados indicadores socioeconômicos no modelo analítico, de modo a evitar análises unidimensionais e promover uma abordagem mais justa e contextualizada da criminalidade urbana. Essas precauções visam garantir que as análises resultantes respeitem os princípios de justiça social, transparência e responsabilidade no uso de dados.

5. Conclusão

Os resultados apresentam perspectivas interessantes considerando os três critérios e seis opções em ordem de preferência (cenários). Entre os cenários, apenas dois obtiveram um número equilibrado de instalações. Os demais cenários não são recomendados para um tomador de decisão em nenhuma das etapas, pois esses cenários implicam um número alto ou insuficiente de instalações. Entre os cenários recomendados no estádio Ilha do Retiro, a segunda melhor alternativa resolveu o problema das interseções com rios.

O método proposto tem a vantagem de considerar simultaneamente múltiplos critérios, como fatores socioeconômicos, dados demográficos e taxas de criminalidade. Essa integração melhora a identificação de padrões e relações dentro dos dados de criminalidade, auxiliando na tomada de decisões abrangentes para a aplicação da lei e formuladores de políticas. A combinação de clustering não supervisionado com auxílio à decisão multicritério aborda a subjetividade, equilibrando critérios subjetivos com insights baseados em dados. Além disso, ajuda a identificar trade-offs em estratégias de prevenção ao crime e revela interações complexas dentro dos dados, levando a uma análise de crime mais eficaz. Essa integração garante um processo de tomada de decisão consistente e informado para enfrentar os desafios da criminalidade.

Aplicar essa metodologia em outros estádios brasileiros e considerar outros critérios não abordados neste trabalho pode ser uma extensão interessante da análise atual, além de usar a padronização de dados para verificar o impacto na geração dos clusters. Além disso, a possibilidade de aplicar um sistema multicritério que atenda a múltiplos tomadores de decisão também pode agregar valor à abordagem atual

Referências

- ADELEKE, O.J.; OLUKANNI, D.O. Facility location problems: Models, techniques, and applications in waste management. *Recycling* 2020, 5, 10.
- BOEZI, G.; COTTARELLI, A.; MANIMENDRA, I.; PAGMOZZI, D. The Efficiency of European Universities and the Role of Exogenous Factors. *Socioecon. Anal.* 2023, 1, 103–125.
- BRANS, J.P.; VINCKE, P.; MARESCHAL, B. How to select and how to rank projects: The PROMETHEE method. *Eur. J. Oper. Res.* 1986, 24, 228–238.
- CABRERA-BARONA, P.F.; BAYÓN, M.; DURÁN, G.; BONILLA, A.; MEJÍA, V. Generating and mapping Amazonian urban regions using a geospatial approach. *ISPRS Int. J. Geo-Inf.* 2020, 9, 453.
- DE CARVALHO, V.D.H.; COSTA, A.P.C.S. Exploring Text Mining and Analytics for Applications in Public Security: An in-depth dive into a Systematic Literature Review. *Socioecon. Anal.* 2023, 1, 5–55.
- FERNANDES, S.; CAPTIVO, M.E.; CLIMACO, J. Um Sistema de apoio a Decisão para Análise de Problemas de Localizaçao Bicritério; Working Paper 2/05; Centro de Investigação Operacional, Faculdade de Ciências, Universidade de Lisboa: Lisbon, Portugal, 2005. (In Portuguese)
- JI, S.; Dai, M.; Wen, H.; ZHANG, H.; ZHANG, Z.; XIA, Z.; ZHU, J. An improved virtual metrology method in chemical vapor deposition systems via multitask gaussian

processes and adaptive active learning. *Int. J. Adv. Manuf. Technol.* 2022, 122, 3149–3159.

MARESCHAL, B.; DE SMET, Y. Visual PROMETHEE: Developments of the PROMETHEE & GAIA multicriteria decision aid methods. In Proceedings of the 2009 IEEE International Conference on Industrial Engineering and Engineering Management, Hong Kong, China, 8–11 December 2009; pp. 1646–1649.

MURAD, M. A violência e o Futebol: Dos Estudos Clássicos aos Dias de Hoje; Editora FGV: Rio de Janeiro, Brazil, 2007.

NEPOMUCENO, T.C.C. Data-driven Analytics for Socioeconomic Challenges in a Contemporary World. *Socioecon. Anal.* 2023, 1, 1–4.

NEPOMUCENO, T.C.C.; DE MOURA, J.A.; E SILVA, L.C.; COSTA, A.P.C.S. Alcohol and violent behavior among football spectators: An empirical assessment of Brazilian's criminalization. *Int. J. Law Crime Justice* 2017, 51, 34–44.

NEPOMUCENO, T.C.; DARAIO, C.; COSTA, A.P.C. Multicriteria ranking for the efficient and effective assessment of police departments. *Sustainability* 2021, 13, 4251.

NEPOMUCENO, T.C.C.; DARAIO, C.; COSTA, A.P.C.S. Combining multicriteria and directional distances to decompose non-compensatory measures of sustainable banking efficiency. *Appl. Econ. Lett.* 2020, 27, 329–334.

NEPOMUCENO, T.C.C.; COSTA, A.P.C.S. Spatial visualization on patterns of disaggregate robberies. *Oper. Res.* 2019, 19, 857–886.

NEPOMUCENO, T.C.C.; DE CARVALHO, V.D.H.; SILVA, L.C.E.; DE MOURA, J.A.; COSTA, A.P.C.S. Exploring the Bedouin syndrome in the football fan culture: Addressing the hooliganism phenomena through networks of violent behavior. *Int. J. Environ. Res. Public Health* 2022, 19, 9711.

PALHARES, M.F.S.; SCHWARTZ, G.M.; TERUEL, A.P.; SANTIAGO, D.R.P.; TREVISAN, P.R.T.D.C. Lazer, agressividade e violência: Considerações sobre o comportamento das torcidas organizadas. *Mot. Rev. De Educ. Física* 2012, 18, 186–199.

POLETO, T.; DE CARVALHO, V.D.H.; DE OLIVEIRA, R.C.P. Applying Spatial Decision Support for Maternal Mortality Analysis in a Brazilian State. *Socioecon. Anal.* 2023, 1, 92–102.