

Análise e predição da "mancha criminal" no Estado do Rio de Janeiro, utilizando técnicas de aprendizagem de máquina: um estudo de caso

Analysis and prediction of the "criminal stain" in the State of Rio de Janeiro, using machine learning techniques: a case study

Marcelo Fabiano Nascimento de Oliveira¹

Resumo

A gestão da política de Segurança Pública consiste de decisões que visam estabelecer a Paz Social por meio da definição de objetivos e estratégias de atuação e de alocação dos recursos públicos. Dessa forma, qualquer iniciativa que possibilite melhoria dos objetivos e das estratégias, já estará contribuindo bastante para a melhor utilização de recursos públicos, atendendo ao bem comum, como também tornando a política mais eficiente e eficaz. Assim, neste trabalho, o objetivo foi construir um modelo de previsão que pudesse agrupar as diversas regiões do Estado do Rio de Janeiro, relacionando e/ou identificando os tipos de crimes mais incidentes possibilitando uma tomada de decisão mais inteligente e que visem, de fato, o combate à criminalidade.

Palavras-chave: Aprendizado de Máquina, Método Não-Supervisionado, K-Means, Mancha Criminal, Instituto de Segurança Pública do Estado do Rio de Janeiro(ISP-RJ).

Abstract

The management of Public Security Policy consists of decisions that aim to establish Social Peace through the definition of objectives and strategies for action and allocation of public resources. Thus, any initiative that enables the improvement of objectives and strategies will already be contributing significantly to the better use of public resources, serving the common good, as well as making the Policy more efficient and effective. Thus, in this work, the objective was to build a forecasting model that could group the various regions of the State of Rio de Janeiro, relating and/or identifying the most frequent types of crimes, enabling more intelligent decision-making that aims, in fact, at combating crime.

Keywords: Machine Learning, Unsupervised Method, K-Means, Criminal Spot, Public Security Institute of the State of Rio de Janeiro (ISP-RJ).

¹ Doutorando da área de Cidades Inteligentes e Aprendizado de Máquina. Universidade Federal Fluminense (UFF) - Brasil - marcelofabiano@gmail.com

1. INTRODUÇÃO

Atualmente no Brasil, vivemos uma escalada de criminalidade sem precedentes, os quais têm raízes em diversos fatores importantes, dentre eles, a desigualdade social, a falta de oportunidades e a falta de recursos que levam muitos indivíduos a realizarem atividades criminosas como forma de sustento (JUSBRASIL,2024). Segundo matéria publicada em CNN, o número de registros de roubos real é 3 vezes maior do que o registrado em delegacias, mostrando que há um longo caminho ainda a ser percorrido. Hoje, a segurança é considerada uma das principais preocupações e a questão continua a crescer em intensidade e complexidade. A segurança é um aspecto que recebe prioridade máxima por políticos e governos em todo o mundo e têm como objetivo reduzir incidência de crimes (David, 2006).

O Estado do Rio de Janeiro, desde a década de 90, registra dados sobre diversos tipos de crimes nas suas diversas modalidades. Para isso, criou o *Instituto de Segurança Pública (ISP)* (ISP,2024), que coleta, organiza, tabula e disponibiliza todos os dados no formato aberto, através de portal acessado pela internet. Todas as bases de dados utilizados neste trabalho são oriundos dessa fonte, consistindo-se em um arquivo CSV contendo 54 atributos contínuos, dentre eles, registros de homicídio doloso, de lesão corporal e de latrocínio, divididos pelas principais regiões do estado, como Baixada, Capital e Niterói. Ao todo, o arquivo possui 34301 registros.

O processo de análise de crimes possui vantagens significativas como, por exemplo, o emprego de massa de dados crescente sobre crimes, com o objetivo extrair informações que ajudem nas políticas de enfrentamento da criminalidade; a redução de tempo no processo de análise, encurtando a identificação de responsáveis e a aplicação da lei, na identificação antecipada das novas tendências de crimes, na previsão do crime com base na distribuição espacial e na detecção de crime. O conhecimento que se obtém com abordagens de mineração de dados são uma ferramenta muito útil que pode ajudar e apoiar as forças policiais (Keyvanpour et al.,2011).

Não é à toa que a pesquisadora Nath (2007) menciona que solucionar crimes é uma tarefa complexa que exige inteligência e experiência humanas e a mineração de dados é uma técnica que pode ajudá-los com problemas de detecção de crimes. A ideia aqui é tentar capturar anos de experiência humana em modelos de computador por meio de mineração de dados. No cenário atual, os criminosos estão se tornando tecnologicamente sofisticados para cometer crimes (Amarnatham, 2003). Portanto, a polícia precisa de ferramentas de análise para capturar criminosos e permanecer à frente na corrida eterna entre os criminosos e as autoridades. A polícia deve usar as tecnologias atuais (Corcoran et al.,2003) para obterem a tão necessária vantagem. Assim, qualquer pesquisa que possa ajudar na resolução mais rápida de crimes se pagará (Kiani et al.,2015).

Para cumprir esse objetivo, este artigo está dividido em 7 seções, dentre elas, Referencial Teórico, Trabalhos Relacionados, Metodologia, Análise Exploratória, Limitações e Treabalhos Futuros, Dilema Ético, Resultados e Discussões e Referências Bibliográficas.

2. REFERENCIAL TEÓRICO

Esta seção tem o objetivo de apresentar os principais conceitos utilizados neste artigo, sobretudo, o Aprendizado de Máquina Não-Supervisionado, Clustering(Agrupamento) e um dos seus algoritmos mais representativos, o K-Means.

2.1. APRENDIZADO DE MÁQUINA NÃO-SUPERVISIONADO, CLUSTERING (AGRUPAMENTO) e K-MEANS

Segundo Monard & Baranauskas (2003), o Aprendizado de Máquina é uma área da Inteligência Artificial (IA) cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. O Aprendizado Não-Supervisionado, também conhecido como aprendizado por observação e descoberta ou por análise exploratória de dados, explora um conjunto de dados de entrada que é composto por exemplos não rotulados, ou seja, dados em que não há a informação sobre a classe associada a cada exemplo (Metz, 2006). Agrupamento (*Clustering*) é uma forma de modelagem de dados que se baseia na construção de clusters, que são conjuntos de dados que exibem uma propriedade importante, de que os elementos pertencentes a um mesmo conjunto apresentam maior semelhança entre si, que os elementos pertencentes a qualquer outro conjunto, levando em consideração a aplicação de critérios de similaridade (Lachi & Rocha, 2005). O K-Means é um dos principais recursos do Agrupamento, além de um dos mais utilizados devido a facilidade de uso. Uma medida muito importante para avaliar a obtenção dos grupos é o coeficiente Silhouette Score, que mede o quão bem cada instância de dados se encaixa em seu próprio cluster em comparação com os clusters vizinhos mais próximos, combinando a ideia de separação entre clusters com a coesão interna de cada cluster, além de fornecer uma visão abrangente da qualidade da clusterização.

Ainda segundo os pesquisadores Monard & Baranauskas (2003), o Aprendizado de Máquina é uma área da Inteligência Artificial (IA) cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. Em Ludermir (2021) citando o lendário pesquisador Mitchel (1997), menciona que o objetivo principal e primordial do Aprendizado de Máquina é a construção de programas que melhorem seu desempenho por meio de exemplos, sendo necessário para isso, a utilização de uma grande quantidade de exemplos para gerar o conhecimento do computador.

O Aprendizado Não-Supervisionado explora um conjunto de dados de entrada que é composto por exemplos não rotulados, ou seja, dados em que não há a informação sobre a classe associada a cada exemplo (Metz, 2006). Já Monard & Baranauskas (2003) apresenta o Aprendizado Não-Supervisionado como o indutor que analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira, formando os chamados agrupamentos ou clusters (Cheeseman et al., 1996).

A partir das construções mencionadas, podemos observar a complexidade e relevância do tema. A próxima seção menciona uma das principais técnicas utilizadas no Agrupamento: K-Means.

2.2. AGRUPAMENTO (*CLUSTERING*)

Agrupamento, termo traduzido do inglês *Clustering*, é uma forma de modelagem de dados que se baseia na construção de clusters, que são conjuntos de dados que exibem uma propriedade importante, de que os elementos pertencentes a um mesmo conjunto apresentam maior semelhança entre si, que os elementos pertencentes a qualquer outro conjunto, levando em consideração a aplicação de critérios de similaridade (Lachi and Rocha, 2005).

O objetivo do Agrupamento é dividir os dados em várias classes ou clusters e agrupar os dados de acordo com a distância euclidiana e com a proximidade entre os dados (Chong et al., 2021) conforme a Figura 1.

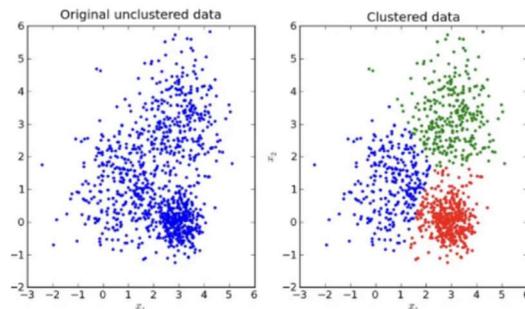


Figura 1 – Exemplo de Clustering
Fonte: encord.com (2023)

2.3. K-Means

O K-means é um dos principais recursos do Agrupamento, além de um dos mais utilizados devido à facilidade de uso. Para Chong et al. (2021), embora o K-means seja amplamente utilizado, ele ainda apresenta alguns defeitos como o fato de que quando a seleção do valor K é inadequada, é fácil fazer o algoritmo cair no mínimo local. Além do fato de ser suscetível a valores discrepantes e se os outliers forem usados como centróides iniciais, a eficiência do algoritmo será significativamente reduzida. A Figura 2 apresenta o cálculo das distâncias quadradas dos centros dos agrupamentos. Desta forma, x_i é o ponto que pertence ao cluster C_k e μ_k representa a média do valor atribuído ao cluster C_k . Cada observação x_i é designada a um cluster de forma que a soma dos quadrados da distância da observação em relação ao seu cluster central (μ_k) é mínima (Smolski, 2024).

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2$$

Figura 2 – Cálculo da distância quadrada dos centros.
Fonte: Chong et al., 2021

Uma medida muito importante para avaliar os grupos obtidos é o coeficiente da *Silhouette Score*, que mede o quão bem cada instância de dados se encaixa em seu próprio cluster em comparação com os clusters vizinhos mais próximos, combinando a ideia de separação entre clusters com a coesão interna de cada unidade, além de fornecer uma visão mais abrangente da qualidade da clusterização. A forma de cálculo do Silhouette é apresentada na Figura 3, onde b_i é a distância *interclusters* definida como a média da distância para o cluster mais próximo do ponto i e a_i é a distância *intracluster* definida como a média da distância para todos os outros pontos dentro do mesmo cluster. O valor resultante do cálculo está entre -1 e 1, sendo que, quanto mais próximo de 1 significa que os clusters são densos e bem separados, quando está próximo de 0 significa que os clusters estão sobrepostos e finalmente valores negativos significam que a clusterização está incorreta.

$$S_i = \frac{b_i - a_i}{\text{Max}(a_i, b_i)}$$

Figura 3 – Cálculo do Silhouetter Score.
Fonte: Autor, 2024

Na próxima seção, apresentamos alguns trabalhos relacionados que fizeram uso de bases de dados de crimes a aplicaram diversas técnicas para processá-los, dentre eles, o Agrupamento.

3. TRABALHOS RELACIONADOS

No trabalho de Agarwal (2013), o interesse se concentrou na análise de crimes de homicídio registrado pela polícia na Inglaterra e no País de Gales, utilizando o K-means. A pesquisa conclui que o número de crimes reduziu, dentro do período de observação da massa de dados que foi de 1990 a 2011. Essa abordagem vai ao encontro da análise realizada com os dados do ISP e nos serviu de referência para a pesquisa.

Em De Bruin et al. (2006) são identificados quatro fatores vitais para a ocorrência de crimes: a natureza do crime, a frequência, a duração e a gravidade. A pesquisa extraiu e criou perfis digitais para todos os infratores, identificando classes de criminosos, utilizando uma abordagem conhecida como Progressive Multi Dimensional Scaling (PDS) proposta por (Williams & Munzner, 2004).

Já Baboo et al. (2011) mencionam a preocupação com a segurança nacional, sobretudo após os eventos de 26 de novembro de 2008, em que foram realizados diversos atentados terroristas em Mumbai. A pesquisa concentrou-se na utilização das técnicas MV (Missing Value) e Apriori, com algumas melhorias para auxiliar no processo de preenchimento do valor faltante e identificação de padrões de crime, permitindo obter conhecimento a partir dos registros criminais, aumentando a precisão preditiva.

Por fim, os pesquisadores Kiani et al. (2015) apresentam a classificação dos crimes agrupados com base na frequência em diferentes anos, em que foi aplicado um modelo teórico com base em técnicas de mineração de dados, como clustering, para um conjunto de dados de crimes reais registrados pela polícia na Inglaterra e pelo País de Gales, entre 1990 e 2011. Além disso, foi utilizada a técnica de atribuição de pesos aos recursos para melhorar a qualidade do modelo e remover baixos valores. O Algoritmo Genético (AG) foi usado para otimização dos parâmetros do operador *Outlier Detection* usando Ferramenta *RapidMiner*.

Todos os trabalhos apresentados fazem uso de bases de dados sobre eventos criminais, aplicando em diferentes cenários técnicas de Agrupamento (Clustering) ou uma combinação de técnicas para aumentar a predição. Entretanto, em razão dos objetivos e das características desta presente pesquisa, seguiremos a linha do trabalho de (Agarwal,2013) e (Baboo et al.,2011), trabalhando com Agrupamentos utilizando o algoritmo K-Means.

4. METODOLOGIA

A figura 4 apresenta a metodologia aplicada neste trabalho, que conta com quatro importantes etapas principais, dentre elas:



Figura 4 – Metodologia de Pesquisa.
Fonte: Autor (2024)

A. Busca dos dados históricos: Os dados foram obtidos no formato CSV, através do portal ISP-RJ e, dessa forma, puderam ser carregados para execução das pesquisas. Nesta etapa, foram utilizados o VSCODE(Versão 1.90.1), PYTHON(Versão 3.12.3) e o pacote de visualização de dados PANDAS(Versão 2.0). Isso permitiu uma avaliação geral dos dados, a verificação de dados faltantes ou irrelevantes(não presentes), além de informações básicas como número de colunas, linhas, quantidade de colunas e etc, cujo resultados da carga é apresentado na Figura 6.

B. Tratamento dos Dados: Com 55 colunas(atributos) foi preciso verificar a necessidade de alocação de todos os dados na pesquisa. Primeiramente, foi aplicada a técnica de correlação entre os atributos, eliminando aqueles em que o fator de alinhamento estivesse abaixo de 0.5 (50%), cujo resultado pode ser vista na Figura 5. Esta etapa foi necessária para otimização do tempo de processamento, já que a pesquisa foi realizada tendo à mão um Notebook Core I5 INTEL geração 10, com 256 SSD e 8GB de RAM. Por esse motivo, algumas etapas da pesquisa não conseguiram ser executadas, sendo necessário migrar para outro ambiente computacional.

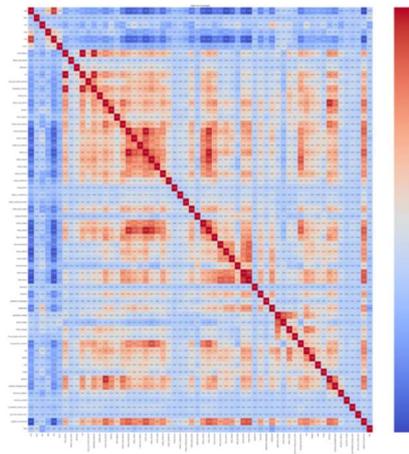


Figura 5 – Matrix de Correlação.

Fonte: Autor

C. Clusterização(k-Means): Diante dos problemas de processamento com o Core I5 INTEL, houve a necessidade de mudar o ambiente computacional para continuar a pesquisa. Dessa forma, foi escolhido um ambiente em nuvem, o Kaggle (Kaggle,2024), juntamente com os pacotes SCIKIT LEARN(Versão 1.5.0), PANDAS(Versão 2.0) e o MATPLOTLIB(Versão 3.9.0). O algoritmo K-Means foi aplicado com os parâmetros de algoritmo 'lloyd', número de clusters 6, máximo de iterações igual a 100 e inicialização ('auto') automática.

D. Resultados: Uma vez realizado o processo de geração dos grupos(Clusters), foi necessária uma visualização dos dados gerados a fim de validar a solução. Dessa forma, utilizamos T-distributed Stochastic Neighbor Embedding (TSNE), uma função do SCIKIT LEARN(Versão 1.5.0) para visualizar dados de alta dimensão, convertendo-os num espaço de duas dimensões, como pode ser visto na Figura 8.

5. ANÁLISE EXPLORATÓRIA

A Figura 6 apresenta os dados em sua versão original, em que é possível observar o cruzamento dos dados de "*Roubo de Celular*" e "*Roubo a Transeunte*", subdivididos nas diversas regiões da pesquisa: Interior, Baixada Fluminense, Grande Niterói e Capital. É

possível observar, também, que estamos diante de um problema, em princípio, não linearmente separável. Como próximo passo importante do Agrupamento, é preciso determinar a quantidade de grupos a serem utilizados na pesquisa, o que é conseguido com o conhecido método do cotovelo, visto na próxima seção.

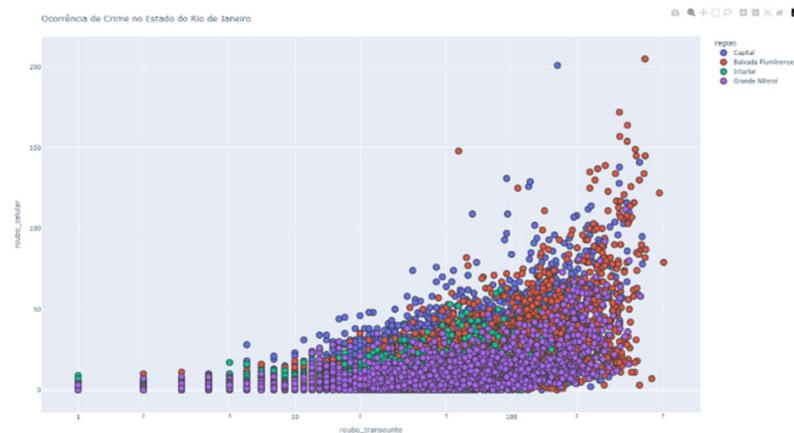


Figura 6 – Dados Originais do ISP sobre Crimes.
Fonte: Autor

5.1. MÉTODO DO COTOVELO

O método “elbow” é conhecido como método do cotovelo e, basicamente, o que ele faz é testar a variância dos dados em relação ao número de clusters. O valor indicado pelo “cotovelo”, no gráfico, significa que não existe ganho marginal em relação ao aumento de clusters. Este método é bastante eficaz para a determinação dos clusters utilizados que, em última aceção, determina também o número de grupos construídos pelo algoritmo K-Means.

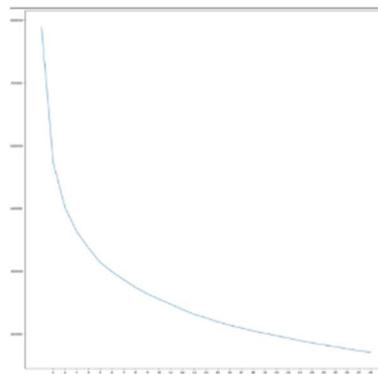


Figura 7 – Dados Originais do ISP sobre Crimes.
Fonte: Autor (2024)

Assim, a Figura 7 apresenta o resultado da curva depois da construção pelo método do cotovelo. Segundo a análise, vamos assumir o valor de $K = 6$ para este trabalho, com Silhouette Score de 0.2836536879977672 ou $28,4\%$ e Coesão Final de 464225.1741986243 , sendo que a inicial foi de 651983.5408347144 .

5.2. K-MEANS E ANÁLISE DE GRUPOS

Através do método do “cotovelo”, chegamos a definição de 6 grupos importantes para aplicação nesta pesquisa, apresentados na figura 8 abaixo, a qual usa um padrão de cores para diferenciar todos os grupos criados.

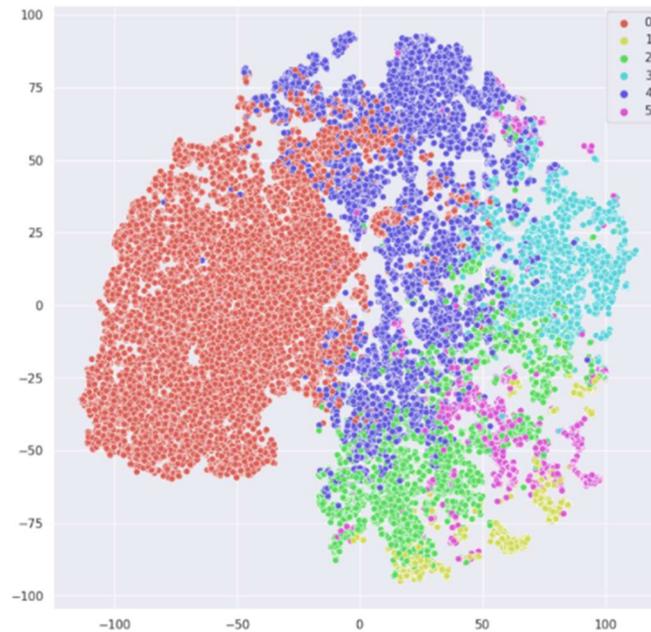


Figura 8 – Grupos Criados após o Clustering(Agrupamento).
Fonte: Autor (2024)

Dentre eles, os que se mostraram mais relevantes para as regiões foram os grupos 1 e 3, vistos abaixo (Figura 9).

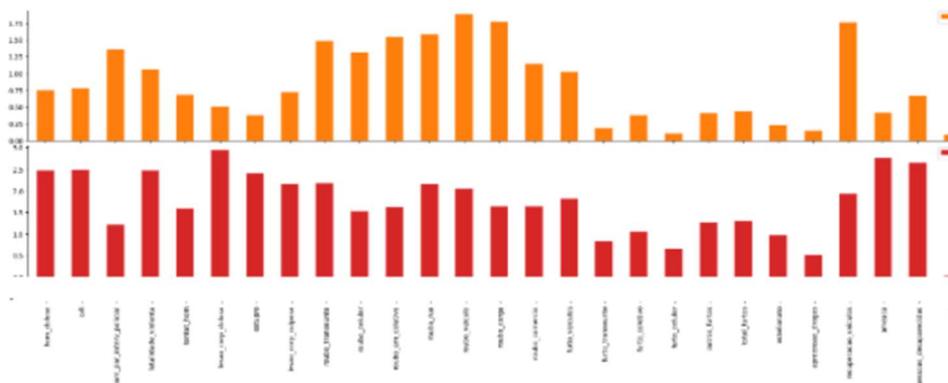


Figura 9 – Grupos Criados após o Clustering(Agrupamento).
Fonte: Autor (2024)

O Grupo 1 (Figura 9, em amarelo) retrata a situação em que os crimes mais observados ou de maior importância são o *Homicídio Doloso*, os *Crimes Violentos Letais Intencionais(CVLI)*, os de *Letalidade Violenta*, a *Lesao Corporal Dolosa*, a *Ameaça e Pessoas Desaparecidas*. O Grupo 3 (Figura 9, em vermelho) retrata a situação em que os fatores como *Lesao Corporal Dolosa*, *Roubo Traseunte*, *Roubo Celular*, *Roubo em Coletivo*, *Roubo Rua*,

Roubo em Coletivo, Roubo Carga e Recuperação de Veículos são mais relevantes. Com os grupos identificados, podemos realizar uma validação a fim de verificar quais regiões estão acometidas de tais crimes, sendo apresentados na próxima seção.

5.3. VALIDAÇÃO DE MODELO

Para validar os grupos, selecionamos uma amostra de 20 exemplos de ocorrências de crimes, juntamente, com suas cidades de origem, observados nas figuras 10 e 11. Como é possível observar, as regiões da Grande Niterói, Capital e Baixada Fluminense podem ser classificadas pelo Grupo 3, em que a incidência de roubos é maior. Já o Interior do estado é mais penalizado com episódios de *Lesões Corporais* e *Homicídios*, que já ensejam maior gravidade.

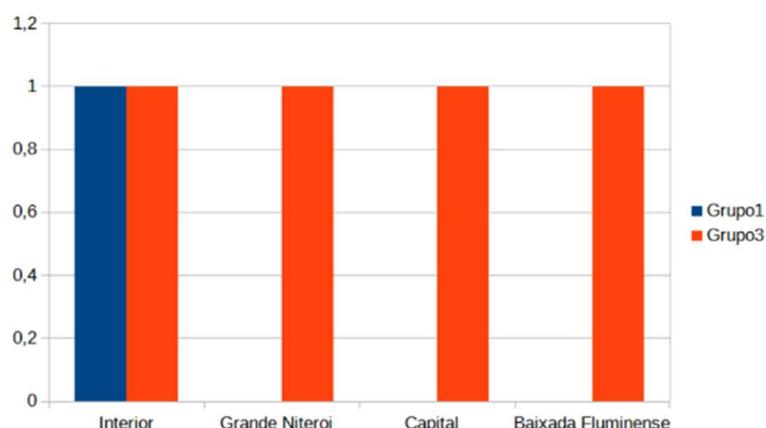


Figura 10 – Grupos 1 e 3 dos Dados Criminais.
Fonte: Autor (2024)

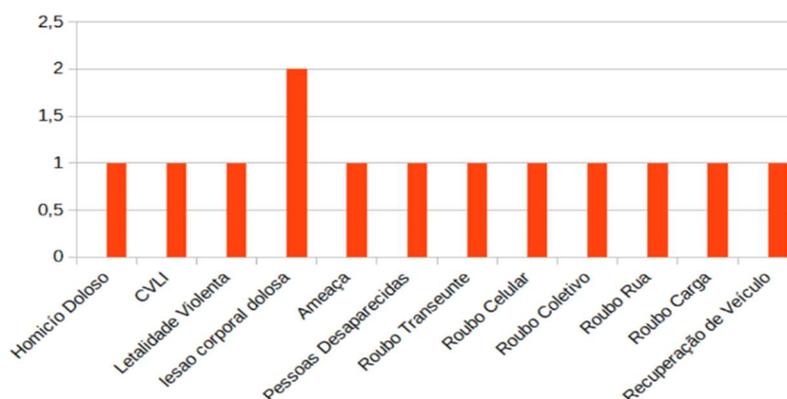


Figura 11 – Cruzamento simples dos dados dos grupos.
Fonte: Autor (2024)

Um simples levantamento dos crimes identificados nos grupos, revela que *Lesão Corporal Dolosa* é um evento importante, pois, alcança valor 2, ou seja, figura em ambos os grupos, como pode ser observado na figura 11. Dessa forma, um gestor de segurança pública, de posse dessa informação, poderia criar estratégias para combater esse crime mais prioritário.

6. RESULTADOS E DISCUSSÕES

Neste artigo, nosso objetivo foi aplicar a técnica de agrupamento K-Means em um conjunto de dados sobre ocorrências criminais, extraídos do *Instituto de Segurança Pública do Estado do Rio de Janeiro (ISP-RJ)*, na forma de um arquivo CSV. Identificamos 4 regiões de interesse importantes na forma da Baixada Fluminense, da Grande Niterói, da Capital e do Interior.

Mencionamos alguns estudos importantes na área, como o trabalho de (Agarwal 2013), no qual o interesse se concentrou na análise de crimes de homicídio registrados pela polícia da Inglaterra e do País de Gales, considerando o tipo de crime de homicídio.

A metodologia aplicada nesta pesquisa, constituiu-se de 4 etapas importantes: Busca de dados históricos, Tratamento de dados, Aplicação de técnicas de Clustering e Apresentação dos resultados. Através do método do cotovelo, o hiperparâmetro k foi definido com o valor 6, uma vez que o Silhouette Score apresentou o maior valor em 28,34% e Coesão final de 464225.1741986243, correspondendo aos 6 grupos utilizados, composto por diferentes tipos de ocorrências criminosas.

A validação foi realizada utilizando-se 20 exemplos de diversas regiões, que apresentaram incidência de diferentes tipos criminais. Munido dessas informações, um gestor de segurança pública poderia realizar a destinação de recursos (a principal proposta desse trabalho) em áreas, setores e combate às necessidades mais visíveis e importantes sobre o tema. Por último, apresentamos as limitações do trabalho e uma faceta do dilema ético advindo da abordagem proposta.

7. LIMITAÇÕES E TRABALHOS FUTUROS

1. Os dados escolhidos para categorizar a “Mancha Criminal” são oriundos do conjunto de dados do ISP-RJ, porém, em quantidade inferior ao total de atributos disponíveis. Assim, para uma análise mais completa, faz-se necessária a introdução de novos atributos sobre o tema, incluindo outras bases de dados complementares.

2. Este trabalho utilizou o K-Means, o principal recurso da técnica de Método Não-Supervisionado. Logicamente, é interessante a aplicação de outros algoritmos e/ou de outras técnicas, como a Supervisionada, com o intuito de capturar mais características importantes sobre este assunto.

8. REFERÊNCIAS BIBLIOGRÁFICAS

AGARWAL, JYOTI; NAGPAL, R. S. R. (2013). Crime analysis using k-means clustering. In Agarwal-Nagpal-Sehgal, editor, *International Journal of Computer Applications*, pages v. 83, n. 4. Academia.

AMARNATHAN, L. (2003). Technological advancement: Implications for crime. *The Indian Police Journal*.

BABOO, S. S. et al. (2011). An enhanced algorithm to predict a future crime using data mining. *International Journal of Computer Applications*, 21(1):1–6.

CHEESEMAN, P. et al. (1996). Bayesian classification (autoclass): theory and results. *Advances in knowledge discovery and data mining*, p. 153–180.

CHONG, B. et al. (2021). K-means clustering algorithm: a brief review. *vol*, 4:37–40.

CNNBRASIL (2024). “Numero de roubos é 3 vezes o de casos registrados em delegacias”. Disponível em: <https://www.cnnbrasil.com.br/nacional/numero-de-roubos-e-3-vezes-o-de-casos-registrados-em-delegacias-diz-pesquisa/> Acesso em 14/06/2024.

CORCORAN, J. J., WILSON, I. D., WARE, J. A. (2003). Predicting the geo-temporal variations of crime and disorder. *International Journal of Forecasting*, 19(4):623–634.

DAVID, G. (2006). Globalization and international security: Have the rules of the game changed. In *Annual meeting of the International Studies Association, California, USA*, http://www.allacademic.com/meta/p98627_index.html.

DE BRUIN, J. S. et al. (2006). Data mining approaches to criminal career analysis. In *Sixth International Conference on Data Mining (ICDM'06)*, p. 171–177.

GOVERNO DE MINAS GERAIS (2024). “Políticas públicas ao seu alcance”. Disponível em: https://politicaspUBLICAS.almg.gov.br/temas/seguranca_publica/entenda/informacoes_gerais.html?tagNivel1=302&tagAtual=302. Acesso em 14/06/2024.

INSTITUTO DE SEGURANÇA PÚBLICA (ISP). Dados abertos. Disponível em: <https://www.ispdados.rj.gov.br>. Acesso em 14/06/2024.

JUSBRASIL (2024). “A criminalidade no brasil: Causas, consequencias e desafios”. Disponível em: <https://www.jusbrasil.com.br/artigos/a-criminalidade-no-brasil-causas-consequencias-e-desafios/1979133466/>. Acesso em: 14/06/2024.

KAGGLE (2024). Ciencia de dados. Disponível em: <https://www.kaggle.com>. Acesso em 19/06/2024.

KEYVANPOUR, M. R., JAVIDEH, M., EBRAHIMI, M. R. (2011). Detecting and investigating crime by means of data mining: a general crime matching framework. *Procedia Computer Science*, 3:872–880.

KIANI, R., MAHDAVI, S., KESHAVARZI, A. (2015). Analysis and prediction of crimes by clustering and classification. *International Journal of Advanced Research in Artificial Intelligence*, 4(8):11–17.

LACHI, R. L.; ROCHA, H. (2005). Aspectos basicos de clustering: conceitos e tecnicas. Núcleo de Informática Aplicada à Educação (Nied), UNICAMP-Instituto de Computação–Universidade Estadual de Campinas.

LUDERMIR, T. B. (2021). Inteligencia artificial e aprendizado de máquina: estado atual e tendencias. *Estudos Avançados*, 35:85–94.

METZ, J. (2006). Interpretação de clusters gerados por algoritmos de clustering hierarquico. Tese de doutorado: Universidade de São Paulo.

MITCHEL, T. (1997). Machine learning, mcgraw-hill education (ise editions).

MONARD, M. C.; BARANAUSKAS, J. A. (2003). Conceitos sobre aprendizado de maquina.

NATH, S. V. (2007). Crime data mining. In Advances and Innovations in Systems, Computing Sciences and Software Engineering, pages 405–409. Springer.

PUSCHEL, A. F. S., RODRIGUES, R. T., VALLE, V. C. L. L. (2022). O dilema etico da decisão algorítmica. A&C-Revista de Direito Administrativo & Constitucional, 22(90):207–226.

SMOLSKI (2024). Analise de clusters. Disponivel em:

<https://smolski.github.io/livroavancado/analise-de-clusters.html>. Acesso em 19/06/2024.

WILLIAMS, M.; MUNZNER, T. (2004). Steerable, progressive multidimensional scaling. In IEEE Symposium on information Visualization, pages 57–64.