

**MINERAÇÃO DE DADOS NO CONTEXTO DA SEGURANÇA  
PÚBLICA: ESTUDO SOBRE VIOLÊNCIA DOMÉSTICA EM  
CAMPO GRANDE, MS**

***DATA MINING IN THE CONTEXT OF PUBLIC SECURITY: A  
STUDY ON DOMESTIC VIOLENCE IN CAMPO GRANDE, MS***



## MINERAÇÃO DE DADOS NO CONTEXTO DA SEGURANÇA PÚBLICA: ESTUDO SOBRE VIOLÊNCIA DOMÉSTICA EM CAMPO GRANDE, MS

### DATA MINING IN THE CONTEXT OF PUBLIC SECURITY: A STUDY ON DOMESTIC VIOLENCE IN CAMPO GRANDE, MS

Wesley Fabricio Souza Silva<sup>1</sup>  
wesley.fabricio@ufms.br

#### RESUMO

Com o crescimento exponencial de dados armazenados, especialmente na segurança pública, torna-se crucial aplicar técnicas avançadas de mineração de dados para apoiar a tomada de decisões. No entanto, embora o uso de mineração de dados em segurança pública esteja em expansão, a aplicação dessas técnicas no Sistema Integrado de Gestão Operacional (SIGO) de Mato Grosso do Sul para prever padrões de recorrência em violência doméstica ainda é uma área inexplorada. Este estudo preenche essa lacuna, aplicando algoritmos simbólicos, como regras de associação e classificação, para extrair padrões significativos nos registros de crimes de violência doméstica em Campo Grande (MS). Os resultados revelam maior incidência de injúria entre vítimas com alta escolaridade e uma tendência de vítimas mais jovens a romper o ciclo de violência doméstica. As regras de classificação atingiram uma acurácia de 84%, identificando que vítimas com maior escolaridade e independência financeira deixam mais facilmente relacionamentos abusivos. Um achado relevante é que vítimas com filhos do agressor registram mais de três ocorrências, sugerindo que permanecem mais tempo no ciclo de violência. Esses achados apoiam intervenções direcionadas, especialmente no Programa Mulher Segura (PROMUSE), e incentivam o uso mais amplo da mineração de dados no SIGO para melhorar a segurança pública.

**Palavras-chave:** Mineração de Dados; Extração de Conhecimento; Algoritmos Simbólicos; Segurança Pública; Violência Doméstica.

#### ABSTRACT

*With the exponential growth of stored data, especially in public security, the application of advanced data mining techniques to support decision-making has become crucial. However, while the use of data mining in public security is expanding, the application of these techniques to the Integrated Operational Management System (SIGO) of Mato Grosso do Sul to predict patterns of recurrence in domestic violence remains an unexplored area. This study addresses this gap by applying symbolic algorithms, such as association and classification rules, to extract significant patterns from records of domestic violence crimes in Campo Grande (MS). The results reveal a higher incidence of verbal abuse (injury) among victims with higher education levels and a tendency for younger victims to break the cycle of domestic violence. The classification rules achieved 84% accuracy, identifying that victims with higher education and financial independence are more likely to leave abusive relationships. A relevant finding is that victims with children from the aggressor register more than three occurrences, suggesting that they remain longer in the cycle of violence. These findings support targeted interventions, especially in the context of the Safe Woman Program (PROMUSE) and encourage broader use of data mining within SIGO to improve public security.*

**Keywords:** Data Mining; Knowledge Extraction; Symbolic Algorithms; Public Security; Domestic Violence.

<sup>1</sup> Mestre em Computação Aplicada (UFMS), Especialista em Engenharia de Software (UCDB), Graduado em Análise e Desenvolvimento de Sistemas (UNIGRAN). Policial Militar do Estado de Mato Grosso do Sul.

## INTRODUÇÃO

O volume de dados gerados e armazenados tem crescido exponencialmente, impulsionado pela evolução do poder computacional e pela digitalização de diversos processos (Katal *et al.*, 2022). Esse crescimento traz consigo tanto desafios quanto oportunidades, principalmente no que diz respeito à transformação desses dados em informações úteis para a tomada de decisões estratégicas.

Nesse cenário, a aplicação de técnicas de mineração de dados surge como uma solução essencial para lidar com a complexidade e o volume massivo de informações, possibilitando a extração automática de padrões e *insights* valiosos (Tan *et al.*, 2016). Esses padrões podem não apenas automatizar tarefas e otimizar processos, mas também servir de base para decisões embasadas em evidências, contribuindo para melhorias em diversas áreas, como a segurança pública (Vercellis, 2009; Garcia *et al.*, 2019).

No campo da segurança pública, particularmente em relação à violência doméstica, a mineração de dados tem se mostrado uma ferramenta poderosa para identificar padrões ocultos e fornecer *insights* que seriam difíceis de detectar por meio de métodos tradicionais (Karakurt *et al.*, 2017; Fórum Brasileiro de Segurança Pública, 2022). Enquanto a mineração de dados refere-se ao processo de explorar grandes volumes de dados para identificar padrões e informações úteis, o aprendizado de máquina, uma subárea da inteligência artificial, utiliza algoritmos que aprendem com os dados para realizar previsões ou classificações.

Ambos são complementares e frequentemente usados juntos para análises descritivas e preditivas. Por exemplo, Neubauer *et al.* (2023) realizaram uma revisão sistemática que demonstrou como métodos computacionais, como redes neurais e algoritmos baseados em regras, são utilizados para identificar vítimas de risco e prevenir recorrência de violência.

Da mesma forma, trabalhos como os de Júnior e Ribeiro (2019) adquiriram modelos baseados em árvores de decisão para analisar fatores associados à violência, destacando sua aplicabilidade para orientar políticas públicas. Além disso, Costa (2023) aplicou técnicas de aprendizado supervisionado, como regressão logística e árvores de decisão, para prever a chance de uma mulher ser vítima de violência em um contexto brasileiro.

Esse estudo ressalta a relevância de fatores socioeconômicos, como escolaridade e emprego, na probabilidade de vítimas registrarem novas ocorrências. Esses resultados reforçam o potencial da mineração de dados em métodos tradicionais de análise complementares, fornecendo *insights* mais precisos e orientados por dados.

Apesar do grande volume de dados armazenados no Sistema Integrado de Gestão Operacional (SIGO), a ausência de técnicas avançadas de mineração de dados e de um departamento especializado em análise de dados restringe a capacidade de transformar essas informações em *insights* estratégicos.

Essa limitação impede que padrões críticos sejam identificados, como os fatores associados à recorrência de violência doméstica, dificultando a alocação eficiente de recursos e o desenvolvimento de políticas públicas embasadas em evidências. Por exemplo, sem ferramentas modernas para prever o comportamento de agressores reincidentes ou identificar as vítimas mais vulneráveis, programas como o PROMUSE enfrentam dificuldades em priorizar casos e promover intervenções direcionadas.

O Programa Mulher Segura (PROMUSE) desempenha um papel fundamental no enfrentamento à violência doméstica em Mato Grosso do Sul. Contudo, sua eficácia depende diretamente da capacidade de identificar rapidamente os casos mais graves e direcionar os recursos necessários, o que poderia ser significativamente aprimorado com o uso de ferramentas analíticas avançadas. Políticas públicas reativas, além de consumir mais recursos, tendem a ser menos eficazes na redução do ciclo de violência. Em contraste, ações baseadas em dados analíticos permitem intervenções antecipadas e mais direcionadas, aumentando a eficiência e a eficácia das estratégias de prevenção.

Diante desses desafios, este estudo propõe aplicar técnicas de mineração de dados ao SIGO para explorar padrões em crimes de violência doméstica registrados em Campo Grande (MS). A proposta visa não apenas preencher a lacuna prática identificada, mas também oferecer um modelo replicável para o aprimoramento de políticas públicas e programas como o PROMUSE. Ao explorar o potencial dos dados armazenados no SIGO com técnicas avançadas de análise, este trabalho busca contribuir para a construção de políticas públicas mais eficazes, promovendo um modelo inovador para o enfrentamento da violência doméstica no Brasil.

## **METODOLOGIA**

Nesta seção, descrevem-se as etapas essenciais seguidas no desenvolvimento deste estudo, com o objetivo de atingir os resultados propostos. A metodologia CRISP-DM (*Cross-Industry Standard Process for Data Mining*) serviu como a estrutura fundamental que guiou todo o processo de extração de conhecimento a partir da base de dados, assegurando uma abordagem padronizada e eficiente na análise dos dados (Shearer, 2000). As subseções seguintes apresentam o entendimento do problema de negócio, a compreensão e preparação dos dados, a modelagem aplicada com os algoritmos de classificação e associação, além da avaliação e a validação dos modelos gerados.

## **COMPREENSÃO DO NEGÓCIO**

O estudo foca nos boletins de ocorrência registrados no SIGO, que armazenam informações detalhadas sobre crimes, criminosos e vítimas, incluindo dados como tipo de crime, localidade, data e hora da ocorrência, características das vítimas e situação socioeconômica. Em colaboração com a Secretaria de Justiça e Segurança Pública de Mato Grosso do Sul (SEJUSP), foi obtida uma base de dados referente a crimes de violência doméstica registrados em Campo Grande (MS) nos meses de março, abril e maio de 2021.

O objetivo é identificar padrões de comportamento entre vítimas que registram mais ou menos boletins de ocorrência contra o mesmo agressor. Compreender os fatores que mantêm essas vítimas no ciclo de violência pode fornecer *insights* valiosos para o PROMUSE, facilitando a priorização do atendimento a essas vítimas. Para alcançar esse objetivo, foram aplicados algoritmos simbólicos de mineração de dados, como regras de associação e regras de classificação, visando à extração de regras interpretáveis sobre os padrões de comportamento das vítimas.

## **COMPREENSÃO DOS DADOS**

Os dados foram disponibilizados no formato *Comma Separated Values* (CSV), contendo inicialmente 1.361 registros. Após a filtragem, focando nos casos em que o autor do crime era o cônjuge da vítima, a amostra foi reduzida para 555 registros. O Quadro 1 apresenta uma visão geral dos atributos contidos na base de dados, como o identificador do boletim de ocorrência, tipificação penal, presença de filhos em comum,

duração do relacionamento, idade da vítima, *status* de emprego, e nível de escolaridade, entre outros.

#### Quadro 1 – Base de dados.

| ATRIBUTO                     | DESCRIÇÃO  |
|------------------------------|--|
| <b>NR BO</b>                 | Identificador do boletim de ocorrência                                   |
| <b>FATO</b>                  | Tipificação penal do crime   |
| <b>FILHOS JUNTOS</b>         | Indicação de filhos em comum entre vítima e agressor                     |
| <b>TEMPO JUNTOS</b>          | Duração do relacionamento entre vítima e agressor em meses               |
| <b>DATA DO FATO</b>          | Dia da semana do crime   |
| <b>HORA DO FATO</b>          | Período do dia do crime  |
| <b>IDADE NA DATA DO FATO</b> | Idade da vítima no momento do crime                                      |
| <b>TRABALHA</b>              | <i>Status</i> de emprego da vítima                                       |
| <b>ESCOLARIDADE</b>          | Nível de escolaridade da vítima  |
| <b>TIPO DE LOCAL DO FATO</b> | Local onde ocorreu a violência   |
| <b>QTD REG ANT</b>           | Quantidade de boletins registrados anteriormente contra o mesmo agressor |

Fonte: Autoria própria.

A coleta manual de alguns desses atributos foi necessária, sendo realizada por especialistas que buscaram informações adicionais no histórico das ocorrências para preencher campos como **FILHOS JUNTOS** e **TEMPO JUNTOS**.

A análise exploratória dos dados foi conduzida para avaliar a relevância dos atributos, detectar valores ausentes e identificar exemplos redundantes. A análise revelou que alguns atributos, como **TEMPO JUNTOS**, **ESCOLARIDADE** e **TRABALHA**, apresentaram valores faltantes. Tais descobertas orientaram as escolhas metodológicas para as etapas subsequentes.

#### PREPARAÇÃO DOS DADOS

Na etapa de preparação dos dados, os valores ausentes dos atributos categóricos **ESCOLARIDADE** e **TRABALHA** foram preenchidos utilizando a moda (valor mais frequente), enquanto os valores ausentes no atributo numérico **TEMPO JUNTOS** foram substituídos pela mediana. O atributo **NR BO** foi removido por ter sido considerado irrelevante para o processo de extração de conhecimento.

O atributo **ESCOLARIDADE** foi padronizado em três níveis: **Fundamental**, **Médio** e **Superior**. O atributo **QTD REG ANT** foi discretizado em duas categorias: **Entre\_1\_e\_3\_registros** (para vítimas com até 3 registros) e **Mais\_que\_3\_registros** (para vítimas com mais de 3 registros).

## MODELAGEM

A modelagem dos dados foi realizada com o uso de algoritmos simbólicos que priorizam a clareza dos resultados e a interpretabilidade dos modelos, essencial para a aplicabilidade prática dos achados. Para as regras de associação, foi utilizada a biblioteca **mlxtend** com o algoritmo **Apriori**, amplamente reconhecido por sua eficiência em identificar padrões frequentes em grandes bases de dados (Agrawal *et al.*, 1994). O suporte mínimo foi definido em 0,10 e a confiança mínima em 0,20, após experimentações para garantir a relevância das regras extraídas.

Na extração das regras de classificação, utilizou-se o *software* **Weka**, uma ferramenta de mineração de dados de código aberto. O algoritmo **J48.PART** foi escolhido por sua capacidade de gerar árvores de decisão claras e interpretáveis (Quinlan, 1993), com uma divisão de 70% dos dados para treino e 30% para teste.

Os **atributos descritivos** utilizados na classificação foram **FILHOS JUNTOS**, **TEMPO JUNTOS**, **IDADE**, **TRABALHA** e **ESCOLARIDADE**, que fornecem informações contextuais sobre as vítimas e suas relações. O atributo **preditivo (ou alvo)** foi **QTD REG ANT**, representando a quantidade de boletins de ocorrência registrados pela vítima contra o mesmo agressor. O objetivo da classificação é prever o número de registros (entre 1 e 3, ou mais de 3), com base nos atributos descritivos fornecidos.

A distribuição das classes mostrou que 75,5% das vítimas registraram entre 1 e 3 boletins de ocorrência contra o agressor (**Entre\_1\_e\_3\_registros**), enquanto 24,5% das vítimas registraram mais de 3 boletins (**Mais\_que\_3\_registros**). Essa distribuição demonstra um certo desbalanceamento de classes, com a maioria das vítimas pertencendo à classe **Entre\_1\_e\_3\_registros**, o que foi levado em consideração na avaliação dos modelos.

## AVALIAÇÃO

Os modelos de classificação foram avaliados utilizando o método de validação cruzada **K-fold cross-validation**, com **10 folds**, a fim de assegurar a generalização dos resultados e minimizar o risco de sobreajuste (*overfitting*) aos dados de treino. Esse método distribui os dados em subconjuntos (*folds*) e, em cada iteração, utiliza um *fold* para teste e os outros para treino, garantindo que o modelo seja testado de maneira robusta e em diferentes combinações de dados.



As métricas tradicionais de desempenho utilizadas para a avaliação dos modelos de classificação incluem **acurácia**, **precisão**, **revocação**, **F1** e **AUC-ROC**. Estas métricas são amplamente empregadas para medir a eficácia de modelos preditivos, cada uma fornecendo uma perspectiva única sobre a performance do modelo em termos de previsibilidade e capacidade de evitar erros.

Para as regras de associação, foi utilizada a métrica de **elevação (lift)** para ranquear as regras extraídas. A elevação é uma medida que quantifica a força da correlação entre o antecedente e o consequente de uma regra, comparando a ocorrência conjunta dos itens com sua ocorrência independente. Em outras palavras, a elevação indica o quanto a presença de um item (ou conjunto de itens) aumenta a probabilidade de ocorrência de outro item em relação ao que seria esperado por acaso. Um valor de elevação maior que 1 indica uma correlação positiva significativa, sugerindo que os itens têm maior chance de ocorrerem juntos. Já valores próximos de 1 indicam que os itens são independentes, ou seja, a regra pode não ter relevância prática. Essa análise é essencial para assegurar que os padrões extraídos sejam estatisticamente significativos e não meras coincidências.

A utilização dessas métricas fornece uma avaliação abrangente da qualidade dos modelos de classificação e das regras de associação, garantindo que os resultados gerados sejam não apenas precisos, mas também úteis e aplicáveis ao contexto de análise dos crimes de violência doméstica.

## IMPLANTAÇÃO

Os padrões e o conhecimento extraído serão disponibilizados para as autoridades competentes e para a sociedade, com o intuito de auxiliar na compreensão dos padrões de violência doméstica. Espera-se que esse conhecimento apoie a tomada de decisões no combate a esse tipo de crime, particularmente no contexto do PROMUSE, permitindo intervenções mais eficazes e preventivas.

## RESULTADO E DISCUSSÃO

Nesta seção, são apresentados os resultados obtidos a partir das regras de classificação e associação, acompanhados de uma análise detalhada sobre o conhecimento extraído dos dados de violência doméstica.



## RESULTADOS DAS REGRAS DE CLASSIFICAÇÃO

Para avaliar as regras de classificação, utilizaram-se as métricas tradicionais de desempenho, como acurácia, precisão, revocação, F1 e AUC-ROC. Os resultados obtidos com o algoritmo J48.PART estão sintetizados na Tabela 1. A acurácia de 84% indica um nível satisfatório de precisão na identificação dos padrões de comportamento das vítimas de violência doméstica.

**Tabela 1 – Resultado das métricas de avaliação da classificação com o algoritmo J48.PART.**

| Métrica   | Valor da medida |
|-----------|-----------------|
| Acurácia  | 84%             |
| Precisão  | 79%             |
| Revocação | 84%             |
| F1        | 80%             |
| AUC-ROC   | 61%             |

Fonte: Autoria própria.

A acurácia de 84% demonstra que o modelo é eficiente na classificação correta das instâncias, enquanto a revocação de 84% revela que o modelo tem um bom desempenho na identificação de vítimas que registraram boletins de ocorrência. Embora a AUC-ROC de 61% indique um nível moderado de capacidade discriminativa, as outras métricas corroboram a robustez do modelo para a finalidade proposta.

O algoritmo J48.PART foi capaz de extrair regras interpretáveis sobre os perfis das vítimas. Um exemplo significativo é a regra que revela que vítimas com idade menor ou igual a 23 anos tendem a registrar entre 1 e 3 boletins de ocorrência contra o mesmo agressor. Este padrão foi encontrado em 91 casos, com uma taxa de erro de apenas 3 exemplos:

idade <= 23: Entre\_1\_e\_3\_registros (91.0/3.0)

Além disso, as vítimas com escolaridade média, filhos em comum com o agressor e uma profissão também estão entre aquelas que registram de 1 a 3 boletins de ocorrência. Este achado sugere que o contexto socioeconômico desempenha um papel fundamental na dinâmica das denúncias, influenciando a decisão de reportar ou não novos incidentes:

escolaridade = MÉDIO AND possuiProfissao = S AND

filhosJuntos = S: Entre\_1\_e\_3\_registros (26.0/3.0)

Outro padrão interessante é o das vítimas com idade inferior a 54 anos e sem profissão, que tendem a registrar mais de 3 boletins de ocorrência. Isso sugere uma maior vulnerabilidade socioeconômica dessas vítimas, possivelmente relacionada à dependência financeira ou social:

Possui Profissão = N AND idade <= 53: Mais\_que\_3\_registros (6.0)

Quando nenhuma das regras anteriores se aplica, o algoritmo categoriza as vítimas na classe de registro de entre 1 e 3 boletins de ocorrência, refletindo a predominância desse comportamento na base de dados:

Entre\_1\_e\_3\_registros (148.0/37.0)

Esses padrões reforçam a relevância de variáveis como idade, escolaridade, filhos em comum e *status* de emprego na determinação da frequência com que as vítimas registram boletins de ocorrência. Esses fatores são cruciais para a compreensão do comportamento das vítimas no ciclo da violência.

## RESULTADOS DAS REGRAS DE ASSOCIAÇÃO

Na Tabela 2, são apresentados os 10 *itemsets* mais frequentes, revelando padrões significativos nos dados. Os resultados mostram que a residência é o local mais comum para a ocorrência dos crimes, e que a maioria das vítimas registra entre 1 e 3 boletins de ocorrência. Além disso, observa-se que vítimas com nível de escolaridade fundamental e que têm filhos com o agressor formam um grupo significativo dentro da base de dados, reforçando a necessidade de intervenções mais direcionadas.

**Tabela 2 – Conjuntos de itens frequentes.**

| Suporte | Conjuntos de itens  |
|---------|---|
| 0.90    | (tipo Local = RESIDÊNCIA)                                       |
| 0.86    | (qtdRegAnt = Entre_1_e_3_registros)                             |
| 0.77    | (tipoLocal = RESIDÊNCIA, qtdRegAnt = Entre_1_e_3_registros)     |
| 0.65    | (escolaridade = FUNDAMENTAL)                                    |
| 0.58    | (escolaridade = FUNDAMENTAL, tipoLocal = RESIDÊNCIA)            |
| 0.57    | (fato = AMEAÇA)   |
| 0.56    | (filhosJuntos = S)  |
| 0.56    | (escolaridade = FUNDAMENTAL, qtdRegAnt = Entre_1_e_3_registros) |
| 0.52    | (possui profissão = S)  |
| 0.52    | (fato = AMEAÇA, tipo Local = RESIDÊNCIA)                        |

Fonte: Autoria própria.

Os resultados das regras de associação, apresentados na Tabela 3, indicam que, em situações em que são registrados mais de 3 boletins de ocorrência, as vítimas geralmente têm filhos com o agressor, e o tipo de violência mais comum é a injúria. Este padrão é observado predominantemente aos domingos, durante o período da tarde, refletindo uma possível correlação com a dinâmica familiar e a maior interação entre as vítimas e os agressores nesses momentos. Além disso, pode haver uma conexão com os horários de partidas de futebol, muito populares no Brasil, que tradicionalmente ocorrem aos domingos à tarde. Durante esses momentos, a tensão em ambientes familiares pode aumentar, especialmente em situações em que o agressor está assistindo ao jogo, o que pode gerar conflitos exacerbados por frustrações, consumo de álcool ou apostas esportivas, criando um cenário propício para o agravamento da violência doméstica.

**Tabela 3 – Regras de associação cujo êxito são vítimas que têm filhos com o agressor.**

| Antecedentes  | Consequentes        | Suporte | Confiança | Elevação |
|---|---------------------|---------|-----------|----------|
| (qtdRegAnt = Mais_que_3_registros)                        | (filhos Juntos = S) | 0.10    | 0.71      | 1.27     |
| (tipoLocal = RESIDÊNCIA, período = TARDE, fato = INJÚRIA) | (filhos Juntos = S) | 0.10    | 0.71      | 1.26     |
| (período = TARDE, fato = INJÚRIA)                         | (filhos Juntos = S) | 0.11    | 0.69      | 1.22     |
| (diaSemana = Domingo)                                     | (filhos Juntos = S) | 0.10    | 0.66      | 1.17     |

**Fonte:** Autoria própria.

Este achado pode estar correlacionado com os resultados de uma pesquisa do Fórum Brasileiro de Segurança Pública (2022), que demonstrou que os dias de jogos do Campeonato Brasileiro estão associados a um aumento de 23,7% nas ameaças contra mulheres e um aumento de 20,8% em lesões corporais dolosas no contexto de violência doméstica. A pesquisa sugere que o futebol atua como um catalisador de comportamentos agressivos, exacerbando valores patriarcais e masculinidades tóxicas, especialmente quando há frustrações relacionadas aos resultados dos times. A dinâmica familiar e o contexto cultural, em que o futebol pode intensificar tensões preexistentes, se refletem nos dados, destacando a importância de considerar esses fatores em estratégias de prevenção de violência doméstica (Fórum Brasileiro De Segurança Pública, 2022).

Outro achado significativo, mostrado na Tabela 4, é a relação entre o nível educacional superior e a ocorrência de injúria. Aproximadamente 60% dos casos de injúria ocorreram dentro do ambiente doméstico para vítimas com escolaridade superior. Esses dados indicam uma correlação importante entre o nível de escolaridade e a violência verbal, ressaltando a necessidade de intervenções mais específicas para esse grupo.

**Tabela 4 – Regras de associação cujo êxito são vítimas com nível superior.**

| Antecedentes                                     | Consequentes     | Suporte | Confiança | Elevação |
|--|------------------|---------|-----------|----------|
| (escolaridade = SUPERIOR, tipoLoca = RESIDÊNCIA) | (fato = INJÚRIA) | 0.10    | 0.60      | 1.23     |
| (escolaridade = SUPERIOR)                        | (fato = INJÚRIA) | 0.11    | 0.58      | 1.19     |

Fonte: Autoria própria.

A Tabela 5 destaca uma importante associação entre escolaridade e profissão. Cerca de 70% das vítimas com escolaridade superior também possuem uma profissão. Esse achado sugere que essas variáveis influenciam não apenas o padrão de denúncias, mas também a capacidade de romper o ciclo de violência doméstica.

Observou-se que vítimas com escolaridade superior e uma profissão, em 66% dos casos, registram entre 1 e 3 boletins de ocorrência, o que indica que essas vítimas tendem a romper o ciclo de violência mais rapidamente, registrando menos ocorrências em comparação a outros grupos.

**Tabela 5 – Regras de associação cujo êxito são vítimas que têm uma profissão.**

| Antecedentes   | Consequentes           | Suporte | Confiança | Elevação |
|--|------------------------|---------|-----------|----------|
| (escolaridade = SUPERIOR)                                    | (possui Profissão = S) | 0.15    | 0.70      | 1.33     |
| (escolaridade = SUPERIOR, tipoLocal = RESIDÊNCIA)            | (possui Profissão = S) | 0.13    | 0.69      | 1.31     |
| (escolaridade = SUPERIOR, qtdRegAnt = Entre_1_e_3_registros) | (possui profissão = S) | 0.11    | 0.66      | 1.26     |

Fonte: Autoria própria.

Esse padrão indica que vítimas com maior escolaridade e independência financeira, ao possuírem uma profissão, registram menos boletins de ocorrência, sugerindo que rompem o ciclo de violência mais facilmente. Esses dados são relevantes

para a formulação de estratégias mais direcionadas para esses grupos, aproveitando sua maior capacidade de sair de situações abusivas.

Por fim, a Tabela 6 destaca a recorrência de crimes de ameaça e injúria, frequentemente associados entre si. Esses dados sugerem um padrão comum de abuso verbal e ameaças, refletindo o comportamento agressivo dos agressores e o impacto contínuo na vida das vítimas.

**Tabela 6** – Regras de associação que demonstram a ameaça acompanhada de injúria.

| Antecedentes  | Consequentes  | Suporte | Confiança | Elevação |
|---|---------------|---------|-----------|----------|
| (fato = INJÚRIA, filhosJuntos = N)                                    | (fato=AMEAÇA) | 0.13    | 0.65      | 1.15     |
| (fato = INJÚRIA, filhosJuntos = N, qtdRegAnt = Entre_1_e_3_Registros) | (fato=AMEAÇA) | 0.12    | 0.65      | 1.14     |

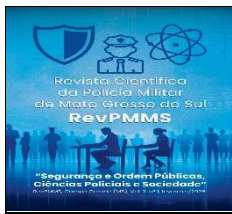
Fonte: Autoria própria.

## CONCLUSÃO

Este trabalho destaca a importância da aplicação de técnicas de mineração de dados no campo da segurança pública, especialmente na análise de crimes de violência doméstica. A pesquisa demonstrou que os dados armazenados no Sistema Integrado de Gestão Operacional (SIGO) do Mato Grosso do Sul podem ser transformados em conhecimento valioso, orientando a criação de políticas públicas e auxiliando programas como o PROMUSE na priorização do atendimento às vítimas mais vulneráveis.

Os resultados obtidos reforçam a eficácia da abordagem adotada, com uma acurácia de 84% nas regras de classificação e a identificação de padrões significativos. Um dos achados mais relevantes é a maior tendência de vítimas que têm filhos e estão em relacionamentos de longa duração a permanecerem em ciclos de violência. Isso evidencia a necessidade de estratégias específicas para essas vítimas, que, devido à sua situação familiar, enfrentam maiores dificuldades em romper o ciclo de abuso.

Outro aspecto relevante foi a constatação de que vítimas com escolaridade superior e independência financeira demonstram maior capacidade de sair de relacionamentos abusivos. Esse achado ressalta a importância de considerar fatores socioeconômicos no desenvolvimento de estratégias de intervenção, uma vez que o empoderamento financeiro e educacional pode desempenhar um papel crucial na libertação dessas vítimas do ciclo de violência.



Apesar dos resultados promissores, algumas limitações foram identificadas. A principal delas foi a necessidade de coleta manual de certos atributos devido à falta de estruturação dos dados no SIGO. Esse fator restringiu o tamanho da base de dados e aumentou o esforço necessário para análise. Para superar esses desafios, recomenda-se a adoção de técnicas mais avançadas, como a mineração de texto para extração automatizada de dados não estruturados, além da utilização de aprendizado de máquina, incluindo técnicas de aprendizado profundo, que podem ampliar a eficácia das análises e permitir previsões ainda mais precisas.

Em síntese, este estudo não apenas contribui para o debate sobre o uso de técnicas de descoberta de conhecimento em segurança pública, mas também oferece um exemplo concreto de como a análise de dados pode ser uma ferramenta poderosa na redução da violência doméstica. A partir dos *insights* gerados, espera-se que políticas mais eficazes possam ser implementadas, promovendo uma sociedade mais segura e justa, com decisões cada vez mais embasadas em dados.

## REFERÊNCIAS

AGRAWAL, R.; SRIKANT, R. **Fast algorithms for mining association rules in large databases**. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 20., 1994, Santiago, Chile. Proceedings [...]. San Francisco: Morgan Kaufmann, 1994. p. 487–499.

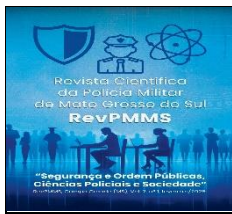
COSTA, Ana Carolina Nepomuceno. **Avaliação de predição de violência contra a mulher através de estratégias de aprendizado de máquinas**. 2023. Disponível em: [https://repositorio.ufc.br/bitstream/riufc/74007/4/2023\\_dis\\_acncosta.pdf](https://repositorio.ufc.br/bitstream/riufc/74007/4/2023_dis_acncosta.pdf). Acesso em: 27 nov. 2024.

FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. **Campeonato Brasileiro e violência doméstica**. São Paulo: Fórum Brasileiro de Segurança Pública, 2022. Disponível em: <https://forumseguranca.org.br/wp-content/uploads/2022/07/futebol-violencia-mulher.pdf>. Acesso em: 30 set. 2024.

GARCIA-VIDAL, C. et al. **Artificial intelligence to support clinical decision-making processes**. EBioMedicine, v. 46, p. 27-29, 2019. Disponível em: [https://www.thelancet.com/article/S2352-3964\(19\)30454-2/fulltext](https://www.thelancet.com/article/S2352-3964(19)30454-2/fulltext). Acesso em: 28 set. 2024.

JÚNIOR, Alfeu Buriti Pereira; DO NASCIMENTO RIBEIRO, Francisco Icaro. **Análises de dados de violência doméstica com o uso de Aprendizagem de Máquina: um Mapeamento Sistemático**. In: Anais da VII Escola Regional de Computação do Ceará, Maranhão e Piauí. SBC, 2019. p. 143-150. Disponível em: <https://sol.sbc.org.br/index.php/ercemapi/article/view/8856> Acesso em: 27 nov. 2024.





KARAKURT, G. et al. ***Mining electronic health records data: Domestic violence and adverse health effects.*** Journal of Family Violence, v. 32, p. 79-87, 2017. Disponível em: <https://link.springer.com/article/10.1007/s10896-016-9872-5>. Acesso em: 28 set. 2024.

KATAL, A.; DAHIYA, S.; CHOUDHURY, T. ***Energy efficiency in cloud computing data centers: a survey on software technologies.*** Cluster Computing, v. 26, n. 3, p. 1845-1875, 2023. Disponível em: <https://link.springer.com/article/10.1007/s10586-022-03713-0>. Acesso em: 28 set. 2024.

NEUBAUER, L. et al. ***A systematic literature review of the use of computational text analysis methods in Intimate Partner Violence research.*** Journal of Family Violence, v. 38, n. 6, p. 1205-1224, 2023. Disponível em: <https://link.springer.com/article/10.1007/s10896-023-00517-7>. Acesso em: 28 set. 2024.

QUINLAN, J. R. C4.5: ***Programs for Machine Learning.*** San Francisco: Morgan Kaufmann, 1993.

SHEARER, C. ***The CRISP-DM model: The new blueprint for data mining.*** Journal of Data Warehousing, v. 5, n. 4, p. 13-22, 2000.

TAN, P. N.; STEINBACH, M.; KUMAR, V. ***Introduction to data mining.*** New Delhi: Pearson Education India, 2018.

VERCELLIS, C. ***Business intelligence: data mining and optimization for decision making.*** Chichester: John Wiley & Sons, 2011.