

MINERAÇÃO DE DADOS NA ANÁLISE DA CRIMINALIDADE: UM ESTUDO DE CASO EM CARAPICUÍBA (2014-2016)

DATA MINING IN CRIME ANALYSIS: A CASE STUDY IN CARAPICUÍBA (2014-2016)

MINERÍA DE DATOS EN EL ANÁLISIS DEL DELITO: UN ESTUDIO DE CASO EN CARAPICUÍBA (2014-2016)

Nicole Carvalho Souza¹
Alexandre Bittencourt Faria²
Arthur Mendes Martins³
Dayane Sivil Moreira⁴

Artigo recebido em junho de 2025
Artigo aceito em dezembro de 2025

DOI: 10.26853/Refas_ISSN-2359-182X_v12n03_01

RESUMO

Este artigo explora a aplicação da mineração de dados no estudo da criminalidade e violência em Carapicuíba durante o período de 2014 a 2016, com foco na importância da compreensão desses eventos para estratégias de segurança pública. Utilizando o processo KDD, algoritmos de análise preditiva e o software minerador SIPINA Research, o presente artigo visa analisar as informações relacionadas à criminalidade e ao perfil demográfico da região. O uso do processo de Descoberta de Conhecimento em Bancos de Dados (KDD) permitiu a extração de insights valiosos a partir de suas etapas de seleção, pré-processamento e interpretação, onde foi possível transformar informações brutas em conhecimento útil e acionável, enquanto a ferramenta open source SIPINA, ainda que limitada em capacidade de processamento, facilitou apurações exploratórias, modelagem e visualização dos resultados. Ao combinar o KDD com a modelagem preditiva do algoritmo C4.5, o estudo demonstra como a examinação de dados pode ser utilizada para obter percepções e suportar decisões informadas em proteção civil. Os resultados indicaram os tipos de infrações mais frequentes, destacando padrões significativos como a relação entre idades e sexo e revelando tendências como o grau de instrução dos autores dos crimes.

Palavras-chave: KDD; Algoritmo C4.5; SIPINA.

¹ Graduanda em Análise e Desenvolvimento de Sistemas na Fatec de Carapicuíba. Fatec Carapicuíba. E-mail: nicolecarvalho6@gmail.com. OrcId: <https://orcid.org/0009-0006-5483-2592>.

² Mestrado em Engenharia Eletrônica e Computação pelo ITA . Professor de Banco de Dados, Engenharia de Software, Tecnologia da Informação e Tecnologias Web (Senai, UniBan, Faspe, UNG, FATEC e UNIP). E-mail: alexandre.faria01@fatec.sp.gov.br. Lattes: <https://lattes.cnpq.br/6565540672197089>. OrcId: <https://orcid.org/0009-0000-5013-9682>.

³ Graduando em Análise e Desenvolvimento de Sistemas na Fatec de Carapicuíba. E-mail: arthurmendes.fatec@gmail.com. arthurmendes.fatec@gmail.com.

⁴ Graduanda no curso Análise e Desenvolvimento de Sistemas na Fatec de Carapicuíba. E-mail: daysmoreira06@gmail.com. OrcId: <https://orcid.org/0009-0006-2348-3676>.

ABSTRACT

This article explores the application of data mining in the study of crime and violence in Carapicuíba during the period from 2014 to 2016, focusing on the importance of understanding these events for public security strategies. Using the KDD process, predictive analysis algorithms and the SIPINA Research mining software, this article aims to analyze information related to crime and the demographic profile of the region. The use of the Knowledge Discovery in Databases (KDD) process allowed the extraction of valuable insights from its selection, pre-processing, and interpretation stages, where it was possible to transform raw information into useful and actionable knowledge, while the tool open source SIPINA, although limited in processing capacity, facilitated exploratory investigations, modeling and visualization of results. By combining KDD with the predictive modeling of the C4.5 algorithm, the study demonstrates how data examination can be used to gain insights and support informed decisions in civil protection. The results indicated the most frequent types of infractions, highlighting significant patterns such as the relationship between ages and sex and revealing trends such as the level of education of the perpetrators.

Keywords: KDD; Algorithm C4.5; SIPINA.

RESUMEN

Este artículo explora la aplicación de la minería de datos en el estudio del crimen y la violencia en Carapicuíba durante el período de 2014 a 2016, centrándose en la importancia de comprender estos eventos para las estrategias de seguridad pública. Utilizando el proceso KDD, algoritmos de análisis predictivo y el software de minería SIPINA Research, este artículo tiene como objetivo analizar información relacionada con la criminalidad y el perfil demográfico de la región. El uso del proceso Knowledge Discovery in Databases (KDD) permitió extraer insights valiosos desde sus etapas de selección, preprocesamiento e interpretación, donde fue posible transformar información cruda en conocimiento útil y procesable, mientras que la herramienta de código abierto SIPINA, aunque limitado en capacidad de procesamiento, facilitó investigaciones exploratorias, modelado y visualización de resultados. Al combinar KDD con el modelado predictivo del algoritmo C4.5, el estudio demuestra cómo se puede utilizar el examen de datos para obtener conocimientos y respaldar decisiones informadas en protección civil. Los resultados indicaron los tipos de infracciones más frecuentes, destacando patrones significativos como la relación entre edades y sexo y revelando tendencias como el nivel de educación de los perpetradores.

Palabras clave: KDD; Algoritmo C4.5; SIPINA.

1 INTRODUÇÃO

A análise da criminalidade é uma questão crucial para a segurança pública em qualquer sociedade. No contexto específico de Carapicuíba, município do estado de São Paulo, essa observação torna-se ainda mais relevante, considerando a necessidade de compreender e abordar os padrões de delitos e violência que impactam a comunidade local.

Carapicuíba enfrenta desafios significativos relacionados à criminalidade e à violência, refletindo em dinâmicas socioeconômicas complexas, densidade populacional e outros fatores ambientais. Segundo um levantamento realizado por Moreno et al. (2023) em 2023, com base nos registros de roubos e furtos da Secretaria da Segurança Pública (SSP) do mesmo ano, 138 ruas e avenidas foram identificadas como locais com maior incidência desses crimes na região da Grande São Paulo, sendo a cidade encontrada em quarto lugar nesse ranking. Essa apuração retrata a complexidade dos padrões de criminalidade que afetam a população, exigindo um exame detalhado e atualizado para desenvolver estratégias eficazes de segurança pública.

Diante da quantidade de dados disponíveis, a mineração surge como uma ferramenta poderosa para extrair descobertas e padrões ocultos. O processo de Descoberta de Conhecimento em Bancos de Dados, conhecido como KDD (*Knowledge Discovery in Databases*), oferece uma metodologia sistemática para transformar o conjunto em conhecimento útil e acionável.

É crucial ressaltar que, conforme o Manual de Interpretação dos Dados Criminais (Coordenadoria de Análise e Planejamento, 2005, p. 4), é fundamental compreender as limitações dos dados oficiais de criminalidade, pois estão sujeitos a falhas e distorções em todas as etapas, desde a detecção até o registro dos crimes, entre outros cuidados na apreciação das informações, que serão detalhados de forma abrangente neste documento.

O problema central deste estudo reside na carência de investigações detalhadas sobre a criminalidade em Carapicuíba. Essa falta de foco dificulta a implementação de políticas eficazes de segurança pública. Assim, este estudo visa demonstrar como realizar uma análise abrangente dos eventos de criminalidade e violência em Carapicuíba, utilizando como exemplo os boletins de ocorrência registrados entre o ano de 2014 e 2016.

Para isso, será utilizada a técnica de *data mining*, com ênfase no processo KDD e no algoritmo C4.5, para analisar e interpretar os elementos coletados sobre a criminalidade, buscando obter insights relevantes e significativos.

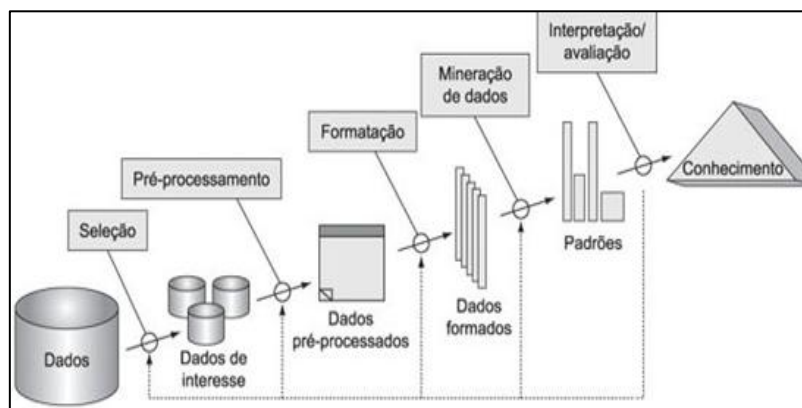
2 REFERENCIAL TEÓRICO

O referencial teórico deste estudo baseia-se no KDD, uma abordagem sistemática para retirar conhecimento útil e acionável a partir de grandes conjuntos de dados.

2.1 Processo de KDD

O processo de KDD, formalizado em 1989, refere-se à busca de conhecimento a partir de uma base de informações. Segundo TEÓFILO (2015), o KDD é um processo iterativo e iterativo, composto por várias etapas não triviais, que possui o intuito de identificar padrões compreensíveis, válidos, novos e potencialmente úteis, composto por várias etapas interdependentes que são essenciais para o sucesso da curadoria de dados (Figura 1).

Figura 1 – Visualização das etapas do Processo de KDD



Fonte: Teófilo, Daniel (2015)

Conceito e importância, permite que organizações e pesquisadores identifiquem conhecimentos valiosos que podem ser utilizados para tomar decisões informadas e estratégicas. A sua importância reside na capacidade de lidar com a complexidade dos dados, permitindo a extração de informações significativas.

Etapas do processo:

- Seleção:** Os elementos relevantes para a análise são selecionados a partir de diversas fontes de dados disponíveis. A seleção das informações é crucial para garantir a qualidade e a relevância do estudo;
- Pré-processamento:** Os dados coletados podem conter inconsistências, ruídos e valores ausentes que precisam ser tratados antes da análise. Esta etapa envolve a limpeza, integração e transformação para prepará-los para a mineração;
- Transformação:** Os dados são transformados em um formato adequado para a pesquisa. Isso pode incluir normalização, redução de dimensionalidade e a criação de novas variáveis com base nas características existentes;
- Mineração de dados:** Nesta fase, os algoritmos de mineração são aplicados aos dados para identificar padrões, tendências e relações. Existem diferentes técnicas de mineração, como árvores de decisão, redes neurais e *clustering*, que podem ser utilizadas dependendo dos objetivos da análise;
- Interpretação/Avaliação:** Nesta etapa, os resultados são interpretados e avaliados em relação aos objetivos da análise. Aqui são extraídos insights expressivos dos padrões identificados e valida-se a qualidade dos modelos desenvolvidos.

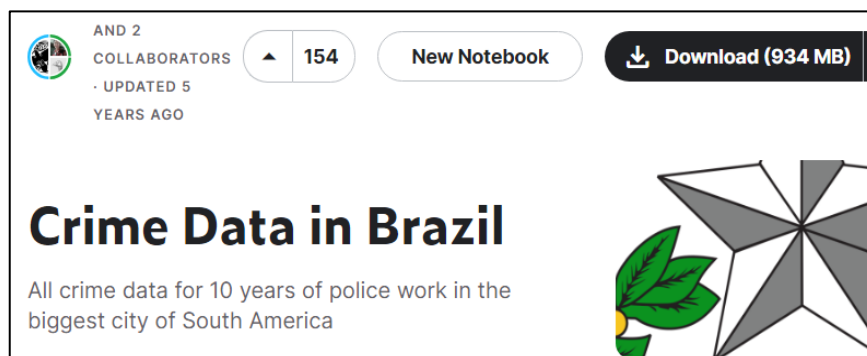
3 MÉTODO

O estudo foi conduzido em diversas etapas, detalhadas em seguida.

3.1 Coleta

Inicialmente, os dados de criminalidade foram obtidos, através do site Kaggle, com a localidade abrangendo todos os municípios do estado de São Paulo, no período de 2007 a 2016.

Figura 2 - Dataset em Kaggle



Fonte: Kaggle (2019)

3.2 Pré-processamento

Os dados coletados foram submetidos a um processo de pré-processamento para garantir qualidade e prepará-los para análise. Esse preparo consistiu na conversão dos arquivos .csv para .xlsx e posterior filtragem, deixando somente boletins de ocorrência de Carapicuíba. O processo incluiu a limpeza dos dados para remoção de informações irrelevantes ou inconsistentes, tratamento de valores ausentes e, a fim de uma predição mais clara e objetiva, foi realizada a simplificação das rubricas, agrupando-as em categorias amplas. O Quadro 1 demonstra a nova categorização dentro da base de dados.

Quadro 1 – Simplificação das rubricas

Categoria	Descrição
Drogas	Drogas sem autorização ou em desacordo (Art.33, caput)
Drogas	Oferecer droga a pessoa de seu relacionamento (Art.33,§3º)
Drogas	Induzir, instigar ou auxiliar alguém ao uso indevido de droga (Art.33,§2º)
Drogas	Tráfico de entorpecente (Art. 12)
Estupro	Estupro de vulnerável (art.217-A)
Estupro	Estupro (art.213)
Furto	Furto (art. 155)
Furto	Furto qualificado (art. 155, §4o.)
Furto	Furto de coisa comum (art. 156)
Homicídio	Homicídio qualificado (art. 121, §2o.)
Homicídio	Homicídio simples (art. 121)
Homicídio	Homicídio culposo (art. 121, §3o.)
Homicídio	Homicídio culposo na direção de veículo automotor (Art. 302)
Lesão Corporal	Lesão corporal (art 129 § 9º)
Lesão Corporal	Lesão corporal (art. 129)
Lesão Corporal	Lesão corporal culposa (art. 129. §6o.)
Lesão Corporal	Lesão corporal culposa na direção de veículo automotor (Art. 303) (“Art. 303 - CTB Digital”)
Lesão Corporal	Lesão corporal de natureza 'GRAVÍSSIMA' (art. 129, §2o.)
Lesão Corporal	Lesão corporal de natureza GRAVE (art. 129, §1o.)
Roubo	Roubo (art. 157)

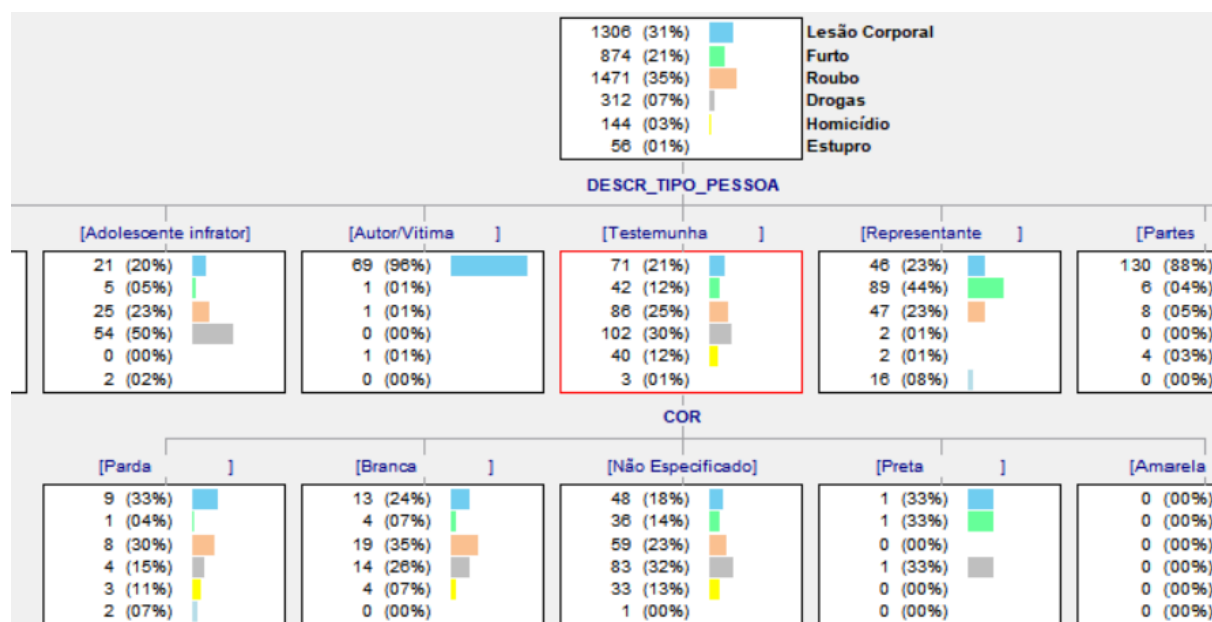
Fonte: Elaboração própria

3.3 SIPINA

O SIPINA é umas das ferramentas iniciais de previsão estatística desenvolvida para auxiliar empresas e instituições a tomarem decisões estratégicas embasadas em dados. Com uma interface simples e uma variedade de algoritmos de aprendizado de máquina, o software permite usuários explorarem um conjunto limitado de dados, identificar padrões e tendências,

e fazer previsões precisas sobre o comportamento futuro. Além do mais, o SIPINA oferece recursos avançados de visualização de dados, facilitando a interpretação e comunicação dos aprendizados gerados, tornando-o indispensável para a busca inicial por aprimorar sua capacidade de previsão e antecipação de cenários (Figura 3).

Figura 3 – Interface SIPINA: Árvore de decisão



Fonte: Elaboração própria

Como o SIPINA é um sistema legado, que não recebe mais suporte, atualizações ou mantém suas documentações ativas, ele desempenha bem apenas com um conjunto menor de dados e, por conta disso, a avaliação que inicialmente seria feita com fundamento nos 10 anos disponíveis foi reduzida para focar nos últimos 3 anos (2014, 2015 e 2016). Ainda que, com limitações visíveis, tal ferramenta identifica padrões e tendências a partir da aplicação de seus algoritmos disponíveis e apresenta flexibilidade na personalização de acordo com as necessidades específicas do usuário.

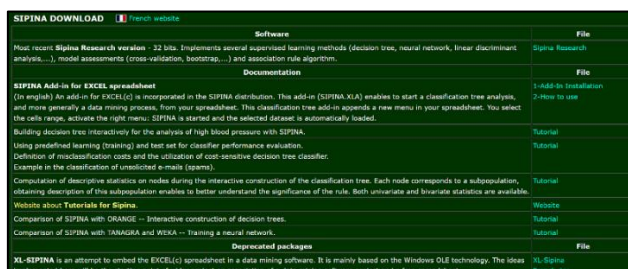
Dentre as vantagens da utilização do SIPINA, encontra-se a tomada de decisão mais informada e precisa, com o fornecimento de previsões baseadas em estatísticas sólidas e diagnósticos robustos. As previsões podem resultar em economia de tempo e recursos, além de minimizar riscos ao antecipar tendências e padrões futuros. A utilização do SIPINA na análise promove uma cultura organizacional orientada por dados, onde as decisões são fundamentadas em evidências, gerando confiança e impulsionando o crescimento sustentável.

3.3.1 Instalação e ajustes

Passos a seguir:

1. Fazer o download na plataforma SIPINA (Figura 4).

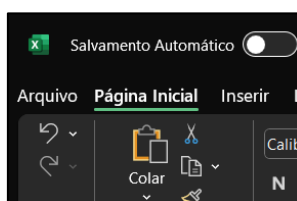
Figura 4 – Instalando a ferramenta SIPINA



Fonte: Elaboração própria

2. Adicionando o SIPINA ao Excel (Figura 5).

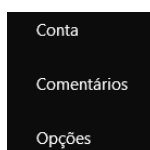
Figura 5 – No Excel, selecionando o menu "Arquivo"



Fonte: Elaboração própria

3. Selecionando "Opções" (Figura 6).

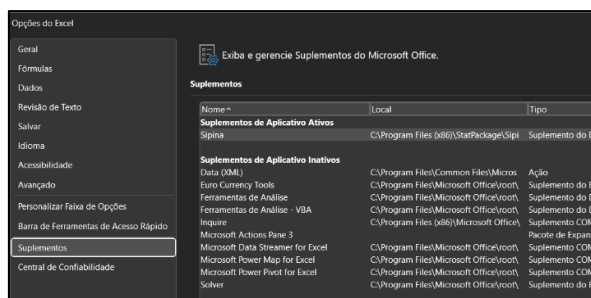
Figura 6 – No menu decorrente da opção "Arquivo", selecionando "Opções"



Fonte: Elaboração própria

4. Selecionar "suplementos" (Figura 7).

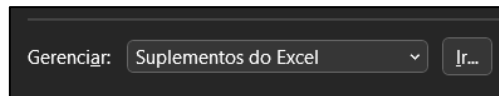
Figura 7 – Selecionando a guia "suplementos"



Fonte: Elaboração própria

5. Selecionar "Suplementos do Excel" (Figura 8).

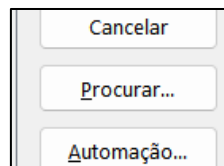
Figura 8 – Na opção 'gerenciar', selecionar 'Suplementos do Excel'



Fonte: Elaboração própria

6. Clicar em "procurar" (Figura 9).

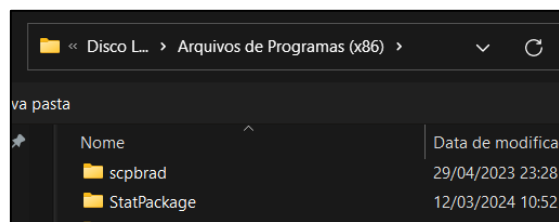
Figura 9 – "Procurar"



Fonte: Elaboração própria

7. Buscar, nos arquivos de programa (Figura 10).

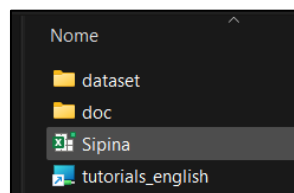
Figura 10 – Abrir a pasta StatPackage



Fonte: Elaboração própria

8. Selecionar arquivo .xla (Sipina) dentro da pasta StatPackage (Figura 11).

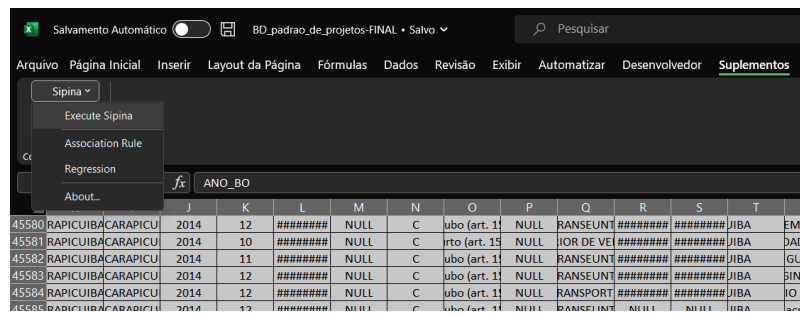
Figura 11 – Arquivo .xla



Fonte: Elaboração própria

9. Selecionar a tabela por completo, "Suplementos", "Sipina" e por fim, "Executar" (Figura 12).

Figura 12 – "Suplementos" > "Sipina"> "Executar"



Fonte: Elaboração própria

3.4 Árvore de Decisão

A análise foi realizada utilizando o algoritmo de árvore de decisão C4.5, algoritmo criado por J. Ross Quinlan no ano de 1993, disponível no software SIPINA. A árvore de decisão é uma técnica de análise preditiva usada na mineração de dados baseada na estrutura de uma árvore, realizando a classificação de acordo com regras de decisão aprendidas a partir do conjunto de treinamento. Utilizando essa abordagem foram explorados padrões temporais, geográficos e de ocorrência de crimes em Carapicuíba ao longo do período estudado. Ainda, foram desenvolvidos modelos preditivos fundados neste algoritmo para estimar a incidência de crimes em diferentes áreas e períodos, visando auxiliar na tomada de decisão das autoridades responsáveis pela segurança pública na região.

3.5 Considerações Éticas

Todas as etapas do estudo foram realizadas seguindo os princípios éticos da pesquisa científica, garantindo o respeito aos direitos e à privacidade dos indivíduos envolvidos. As informações sensíveis relacionadas aos crimes analisados foram tratadas com confidencialidade, e os resultados foram utilizados apenas para fins acadêmicos e de pesquisa.

3.6 Limitações do Estudo

É importante destacar algumas limitações do estudo, como a disponibilidade e qualidade dos dados encontrados através do site Kaggle, bem como a complexidade e dinâmica do fenômeno da criminalidade, que podem afetar os resultados da análise. Ademais, o uso de técnicas de árvore de decisão requer cuidado na interpretação dos resultados, considerando a necessidade de ajustes e validações adicionais e deve ser feita com cuidado, considerando contextos sociais, econômicos e demográficos que podem influenciar os padrões de criminalidade.

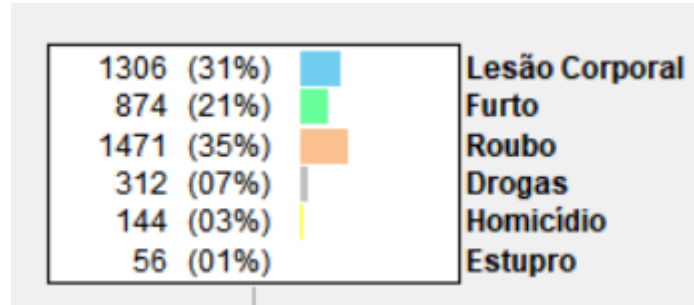
4 RESULTADOS E DISCUSSÃO

Inicia-se os resultados e discussão pelos resultados do caso estudado.

4.1 Resultados do Caso Estudado

A classificação foi gerada sob o atributo de classe rubrica. A Figura 13 mostra o destaque dos crimes por cor.

Figura 13 – Classes de crime

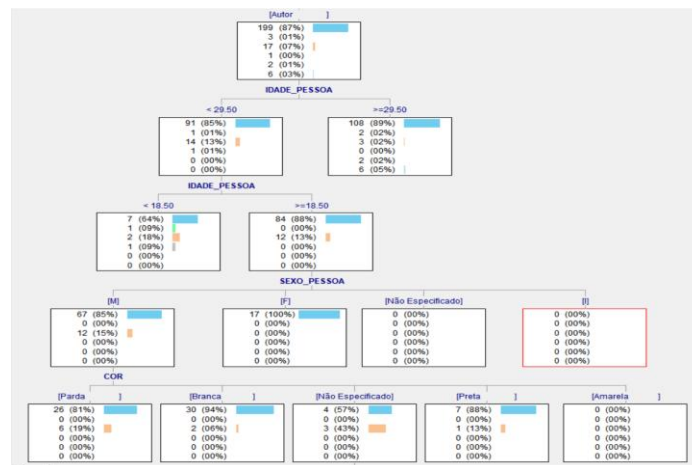


Fonte: Elaboração própria

4.1.1 Perfil demográfico dos crimes em Carapicuíba

Durante o período de 2014 a 2016, observou-se um perfil demográfico característico dos crimes em Carapicuíba. O estudo revelou que, dos boletins de ocorrência classificados pelos autores, os crimes mais recorrentes são de Lesão Corporal, onde 85% foram cometidos por jovens adultos menores de 29 anos. Além disso, constatou-se que 85% dos crimes tiveram autores do sexo masculino, em sua maioria de etnia branca (Figura 14).

Figura 14 – Árvore de característica dos autores

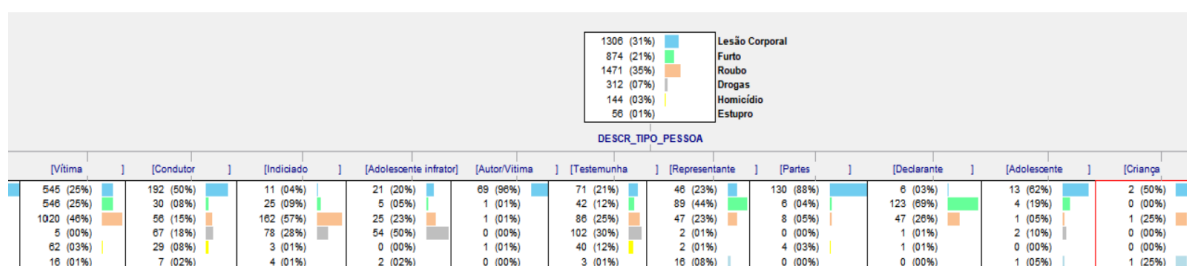


Fonte: Elaboração própria

4.1.2 Tipos de crimes mais frequentes

Os tipos de crimes mais frequentes em Carapicuíba foram analisados, destacando-se Roubo, Lesão Corporal e Furto, representando 87% do total de crimes registrados, indicando que crimes violentos são os mais comuns (Figura 15).

Figura 15 – Nós descritivos dos detentores dos registros

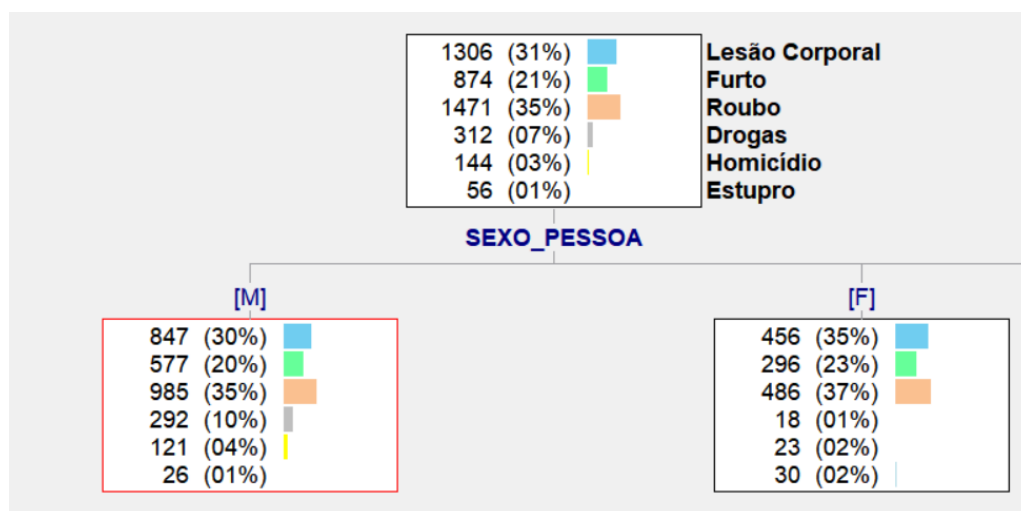


Fonte: Elaboração própria

4.1.3 Relação entre idade, sexo e tipos de crime

Revelou-se uma relação significativa entre idade, sexo e tipos de crime. Verificou-se que homens são predominantes em quase todas as categorias de crimes enquanto há menor representatividade de mulheres, especialmente em categorias de crimes violentos (Figura 16).

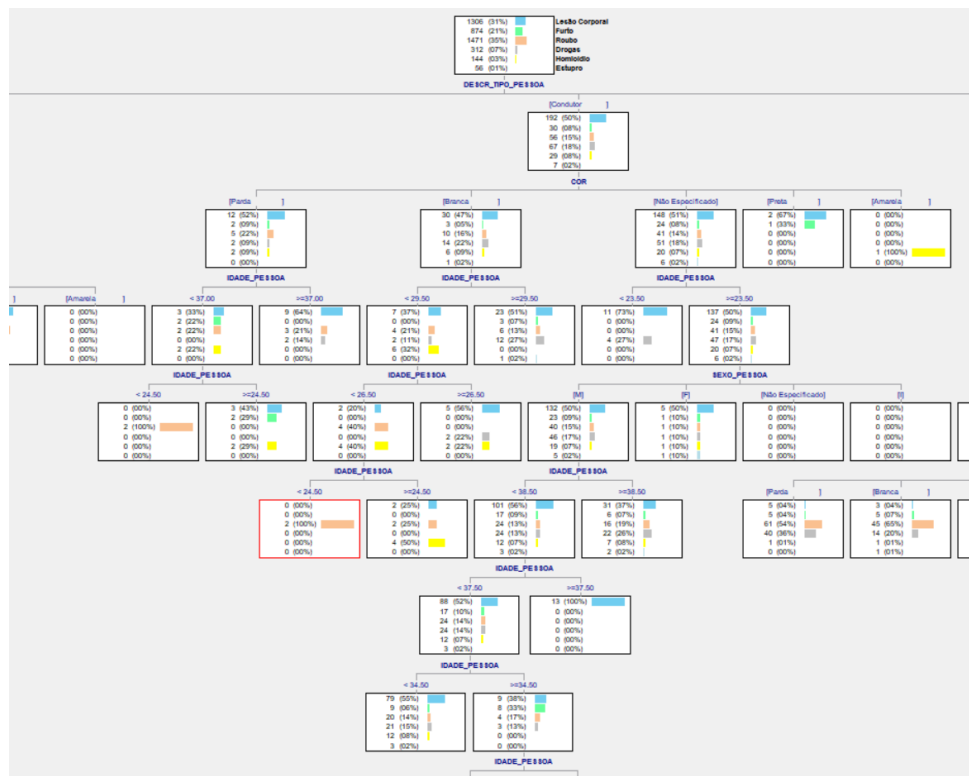
Figura 16 – Idades presentes nos crimes



Fonte: Elaboração própria

Crimes relacionados a drogas e roubo são mais comuns em indivíduos pardos abaixo de 37 anos (Figura 17).

Figura 17 – Distribuição dos nós sobre a cor da pele

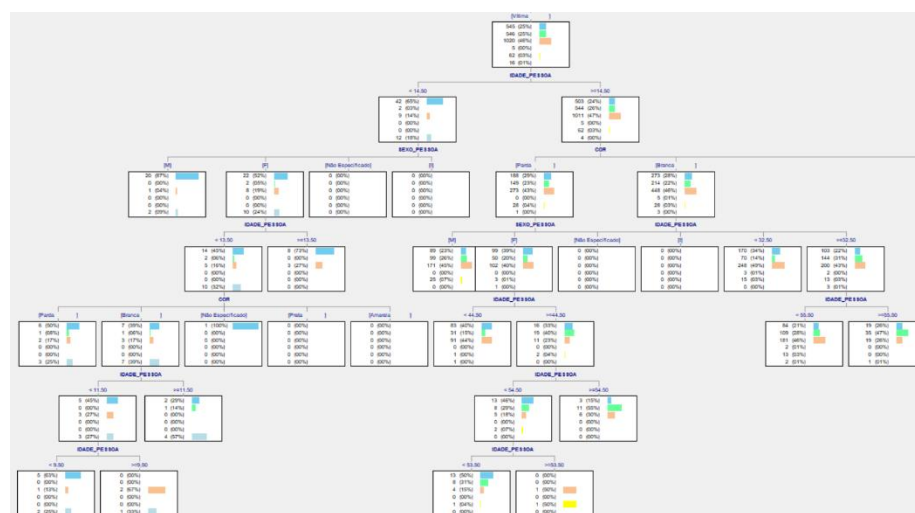


Fonte: Elaboração própria

4.1.4 Idade e vitimização

A maior parte das vítimas tem idade superior a 14.50, sendo crianças e adolescentes (< 14.50) são menos frequentemente vítimas em comparação a adultos. Considerando a cor, indivíduos brancos e pardos são os mais recorrentes e a predominância do sexo masculino se mantém nas vítimas também (Figura 18).

Figura 18 – Distribuição dos nós sobre idade, cor e sexo das vítimas



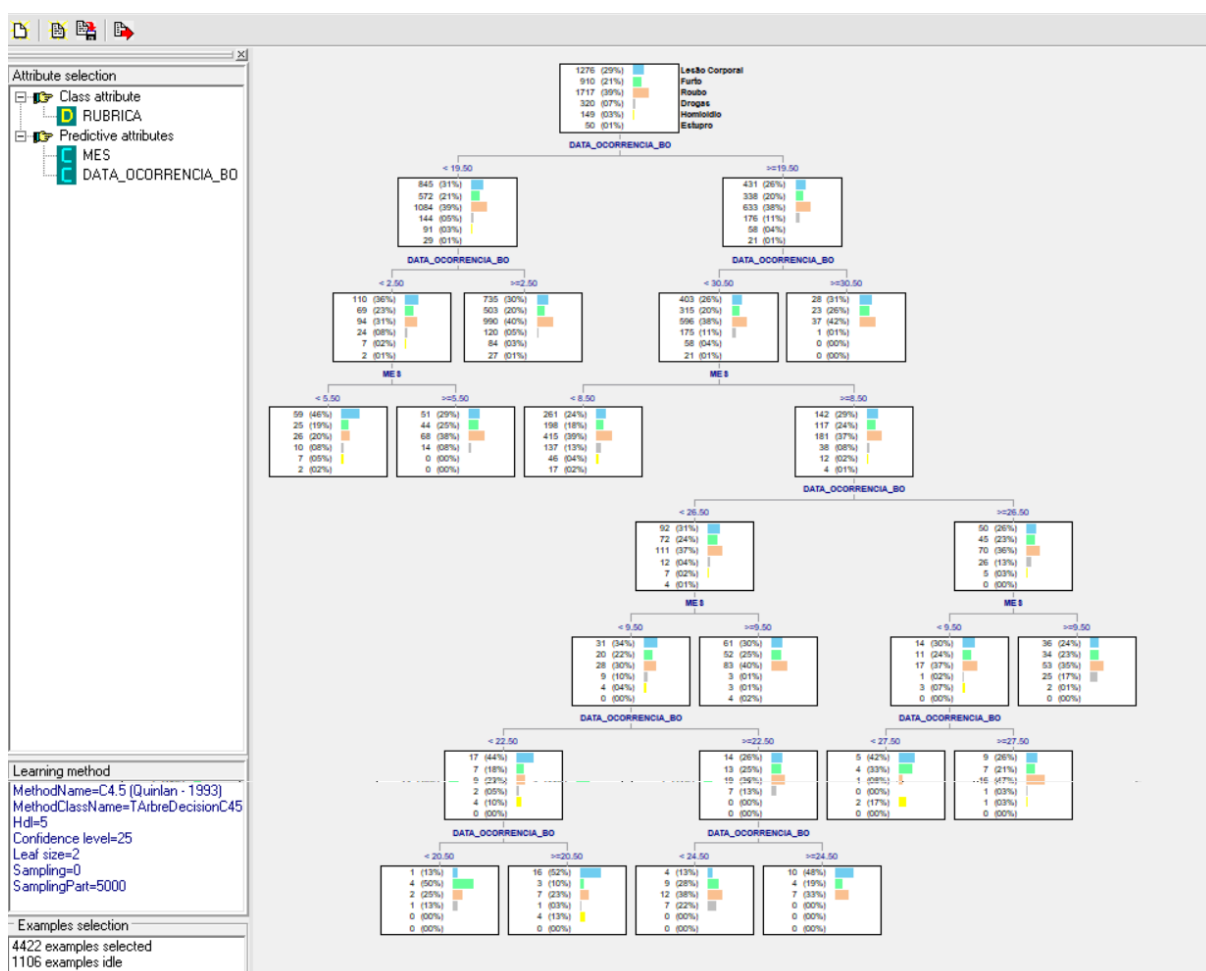
Fonte: Elaboração própria

4.1.5 Relação do período anual com mais registros

A análise também demonstrou a incidência dos crimes nos meses do ano. De acordo com o resultado (Figura 19):

- Janeiro e início de fevereiro: Existe alta incidência de lesão corporal;
- Final de fevereiro, março, abril e início de maio: Alta incidência de furto, roubo e alguns homicídios e estupro;
- Final de maio, junho, julho e início de agosto: Alta incidência de furto, roubo e crimes relacionados a drogas;
- Final de agosto, setembro, outubro e início de novembro: Alta incidência de lesão corporal e alguns homicídios e estupro;
- Final de novembro e dezembro: Alta incidência de crimes relacionados a drogas.

Figura 19 – Incidência dos crimes baseado nos meses do ano



Fonte: Elaboração própria

4.2 Discussão dos Resultados

Os resultados obtidos fornecem compreensões sobre a dinâmica da criminalidade em Carapicuíba, auxiliando na verificação dos padrões e tendências dos delitos na região. A identificação do perfil demográfico dos criminosos e dos tipos de crimes mais frequentes é fundamental para o desenvolvimento de estratégias de prevenção e combate à criminalidade.

As informações obtidas neste estudo têm importantes implicações para a segurança pública e o desenvolvimento de políticas de prevenção à criminalidade em Carapicuíba. Os resultados podem subsidiar a alocação de recursos e a implementação de medidas específicas para reduzir a incidência de crimes na região, promovendo assim um ambiente mais seguro e pacífico para os moradores. Sendo importante ressaltar que o presente estudo foi desenvolvido em base introdutória ao tema e para melhores análises, deve-se considerar uma base de dados ainda maior, mais recente e aplicada em conjunto com ferramentas de maior capacidade.

5 CONSIDERAÇÕES FINAIS

Na análise dos crimes de tráfico cometidos por condutores, identificamos padrões elucidativos que auxiliam o entendimento desses eventos. Os resultados revelaram que uma proporção significativa dos crimes ocorreu entre indivíduos com menos de 29 anos de idade, sendo os condutores do sexo masculino os principais infratores em todas as categorias de crimes relacionados a roubo, lesão corporal e furto. Além disso, os dados destacam a predominância de crimes relacionados a drogas e roubo com uma alta proporção de casos envolvendo infratores abaixo de 37 anos da cor parda.

Em relação aos crimes cometidos por condutores em geral, observamos uma clara disparidade entre os sexos, com uma grande maioria dos crimes sendo perpetrados por homens. Esses crimes abrangem várias categorias, incluindo tráfico de drogas, lesão corporal, homicídio e roubo. A análise dos boletins de ocorrência revelou que o roubo é o crime mais comum com 35%, seguido de lesão corporal com 31%, ambos representando a grande maioria dos registros.

O uso do processo KDD foi basal para extrair insights significativos desses vastos conjuntos de informações. Por meio das etapas do KDD, pudemos transformar dados brutos em conhecimento acionável, proporcionando uma compreensão mais profunda dos padrões de criminalidade. A ferramenta SIPINA demonstrou sua utilidade, facilitando a análise exploratória, modelagem e visualização, e permitindo a identificação de forma eficaz.

Espera-se que este estudo contribua para o cenário de abstração de conhecimento em dados, fornecendo um olhar introdutório à ciência de dados. Com uma compreensão mais clara de como informações podem ser obtidas através da mineração de dados, não apenas as autoridades, como também civis, podem desenvolver estratégias mais eficazes para prevenir e combater esses delitos.

Para pesquisas futuras, recomenda-se a expansão deste estudo para incluir interpretações mais detalhadas de outros tipos de crimes, bem como a investigação de fatores subjacentes que contribuem para o aumento ou redução da criminalidade em determinadas áreas. Além disso, o uso de técnicas mais avançadas de mineração de dados e modelagem preditiva em conjunto com uma ferramenta que ofereça maior capacidade de processamento que a utilizada neste artigo, pode oferecer pontos de vista adicionais sobre a prevenção de crimes e o desenvolvimento de políticas públicas mais eficazes na área da segurança.

6 REFERÊNCIAS

Amazon Web Services (AWS). **O que é mineração de dados?** Disponível em: <https://aws.amazon.com/pt/what-is/data-mining/>. Acesso em: 22 de março de 2024.

BUENO, Felipe. **"Violência urbana no Brasil: faltam dados para identificar com precisão os índices de homicídio."** ("Violência urbana no Brasil: faltam dados para identificar com precisão ...") Jornal da USP. Disponível em: <https://jornal.usp.br/radio-usp/violencia-urbana-no-brasil-faltam-dados-para-identificar-com-precisao-os-indices-de-homicidio/>. Acesso em: 11 de abril de 2024.

DevMedia. **Descoberta de conhecimento utilizando o processo KDD.** Disponível em: <https://www.devmedia.com.br/descoberta-de-conhecimento-utilizando-o-processo-kdd/38709>. Acesso em: 02 de abril de 2024.

Google Cloud. **O que é análise preditiva?** Disponível em: <https://cloud.google.com/learn/what-is-predictive-analytics?hl=pt-br#:~:text=A%20an%C3%A1lise%20preditiva%20%C3%A9%20o,que%20possam%20prever%20comportamentos%20futuros>. Acesso em: 20 de abril de 2024.

IBM. **Data Mining.** Disponível em: <https://www.ibm.com/br-pt/topics/data-mining>. Acesso em: 05 de abril de 2024.

KAHN, Túlio. **Manual de Interpretação de Estatística de Criminalidade.** São Paulo: Coordenadoria de Análise e Planejamento, Secretaria da Segurança Pública, 2005.

MORENO, Ana Carolina; JESUS, Cleber; JOZINO, Lucas et al. **Raio-x da criminalidade: 138 ruas estão entre as que mais tiveram registros de roubos e furtos em 2023 na Grande SP (São Paulo).** São Paulo, TV Globo, 11 mar. 2024. Disponível em: <https://g1.globo.com/sp/sao-paulo/noticia/2024/03/11/raio-x-da-criminalidade-138-ruas-estao-entre-as-que-mais-tiveram-registros-de-roubos-e-furtos-em-2023-na-grande-sp-veja-lista.ghtml>. Acesso em: 08 de abril de 2024.

REDALYC, **Análise de Algoritmos de Indução de Árvores de Decisão.** Disponível em: <https://www.redalyc.org/journal/5606/560662202029/html/>. Acesso em: 22 de março de 2024.

SAS INSTITUTE INC. **Mineração de Dados.** Disponível em: https://www.sas.com/pt_br/insights/analytics/mineracao-de-dados.html. Acesso em: 20 de abril de 2024.

SAS. **"Análises Preditivas: O que são e qual sua importância?"** Acessado em: 12 de abril de 2024. Disponível em: https://www.sas.com/pt_br/insights/analytics/analises-preditivas.html

TEÓFILO, Daniel. **KDD - Knowledge Discovery in Database.** Banco de Dados. Disponível em: <https://danielteofilo.wordpress.com/2015/02/16/kdd-knowlegde-discovery-in-database/>. Acesso em: 15 de abril de 2024.

ZANCHI, Marco et al. **Crime Data in Brazil.** Updated 5 years ago. Disponível em: <https://www.kaggle.com/datasets/inquisitivecrow/crime-data-in-brazil/code>. Acesso em: 25 de março de 2024.