

DOI: <https://doi.org/10.23925/ddem.v.3.n.15.72333>

Licença Creative Commons Atribuição 4.0 Internacional

PREVISÃO DE ROUBOS DE VEÍCULOS NO RIO DE JANEIRO UTILIZANDO ALGORITMOS DE MACHINE LEARNING

VEHICLE THEFT PREDICTION IN RIO DE JANEIRO USING MACHINE LEARNING ALGORITHMS

Gustavo de Souza Coelho¹Lisleandra Machado²Domingos Sávio da Cunha Garcia³Leonardo Amorim de Araújo⁴

RESUMO

As previsões de variáveis criminais são úteis para o desdobramento tático de recursos, alocação de pessoal e planejamento estratégico, aumentando a consciência tática e estratégica das forças policiais. Nos Estados Unidos, métodos preditivos são amplamente utilizados para antecipar tendências e identificar locais e indivíduos propensos à futura vitimização. No entanto, a literatura nacional ainda carece de modelos capazes de antecipar tendências criminais, com algumas exceções, como a ferramenta CrimeVis e outros estudos com técnicas de previsão mais modestas. A evolução e popularização da inteligência artificial (IA), especialmente do machine learning, permitem previsões mais rápidas, baratas e precisas, auxiliando na resolução de problemas anteriormente. A aplicação de IA pode processar e analisar grandes quantidades de informações rapidamente, melhorando a tomada de decisão. O objetivo principal deste trabalho é avaliar qual abordagem de modelagem – árvores de decisão ou redes neurais – oferece maior precisão na previsão de roubos de veículos no município do Rio de Janeiro, contribuindo para o desenvolvimento de ferramentas preditivas que possam auxiliar autoridades e órgãos de segurança pública na alocação eficiente de recursos e no planejamento estratégico de políticas

¹ Bacharel em Ciência da Computação. gscoelho0407@gmail.com. <https://orcid.org/0009-0006-3113-3609>.

² Professora e Pesquisadora pelo CNPQ, FAPEMIG, FUNDEP e CAPES. Possui graduação em Direito, Administração de Empresas, Engenharia de Produção, Pedagogia Doutora em Engenharia de Produção pela UNIMEP e mestra em Engenharia de Produção pela UFSC - Universidade Federal de Santa Catarina. Atualmente é coordenadora do Curso de graduação em Engenharia Ferroviária e Metroviária. Tem profundos conhecimentos em Data Science y Analytics, Digital Business (Business Intelligence). Professora no Instituto Federal de Educação Ciência e Tecnologia do Sudeste de Minas Gerais - Juiz de Fora, MG. Desde 2002, é avaliadora ad hoc de cursos de graduação (INEP/MEC). lisleandra.machado@ifsudestemg.edu.br. <https://orcid.org/0000-0002-7761-8023>.

³ É doutor em Economia Aplicada, com área de concentração em História Econômica, pela Universidade Estadual de Campinas (2005) e professor adjunto C10 da Universidade do Estado de Mato Grosso - UNEMAT, lotado no Departamento de História do campus de Cáceres desde 1995. Tem experiência na área de História, com ênfase em História Econômica e História Política do Brasil no século XIX, atuando principalmente nos seguintes temas: história política da fronteira oeste do Brasil no longo século XIX; geopolítica e relações internacionais: o Brasil e o Prata no século XIX; economia e sociedade na fr. domingos.garcia@unemat.br. <https://orcid.org/0000-0002-8754-6780>.

⁴ Possui graduação em Engenharia Civil pela Universidade Federal de Juiz de Fora (1978), mestrado em engenharia de transportes pela Washington University in Saint Louis (1986), doutorado em Engenharia de Transportes pela Universidade Federal do Rio de Janeiro (2003), ensino-fundamental-primeiro-graupelo Ginásio Pio X (1968) e ensino-medio-segundo-graupelo Instituto Metodista Granbery (1971). Atualmente é professor efetivo do Instituto Federal do Sudeste de Minas Gerais. leonardo.araujo@ifsudestemg.edu.br. <https://orcid.org/0000-0003-2722-7531>.

de prevenção. A precisão das modelagens será comparada utilizando o erro médio quadrático (Mean Squared Error - MSE).

Palavras-chave: redes neurais; árvore de decisão; modelos preditivos; Segurança Pública; Roubo de Veículos.

ABSTRACT

Predictions of crime variables are useful for tactical resource deployment, personnel allocation, and strategic planning, increasing the tactical and strategic awareness of police forces. In the United States, predictive methods are widely used to anticipate trends and identify locations and individuals prone to future victimization. However, national literature still lacks models capable of anticipating crime trends, with some exceptions, such as the CrimeVis tool and other studies with more modest prediction techniques. The evolution and popularization of artificial intelligence (AI), especially machine learning, allows for faster, cheaper, and more accurate predictions, helping to solve problems earlier. The application of AI can process and analyze large amounts of information quickly, improving decision-making. The main objective of this work is to evaluate which modeling approach – decision trees or neural networks – offers greater accuracy in predicting vehicle thefts in the city of Rio de Janeiro, contributing to the development of predictive tools that can assist authorities and public security agencies in the efficient allocation of resources and in the strategic planning of prevention policies. The accuracy of the models will be compared using the Mean Squared Error (MSE).

Keywords: neural networks; decision trees; predictive models; Public Security; Vehicle Theft.

INTRODUÇÃO

A violência urbana é um fenômeno complexo que afeta diversas cidades ao redor do mundo, sendo o Rio de Janeiro uma das metrópoles brasileiras mais impactadas por esse problema. O crescimento populacional desordenado, as desigualdades socioeconômicas e a presença de grupos criminosos organizados contribuem para a alta taxa de criminalidade. Entre os tipos de crimes mais comuns, o roubo de veículos se destaca como uma das ocorrências mais frequentes e preocupantes. De acordo com o Instituto de Segurança Pública do Rio de Janeiro (ISP), em 2023 foram registrados mais de 20.000 roubos de veículos na capital fluminense, número que reflete a persistente crise de segurança pública no estado (Instituto de Segurança Pública, 2023).

O fenômeno da violência urbana no Rio de Janeiro também está intrinsecamente ligado à fragmentação espacial e social da cidade, que se caracteriza pela existência de áreas controladas por facções criminosas. Essas organizações exercem influência em comunidades periféricas e muitas vezes impõem um "poder paralelo", gerando um cenário de conflito constante com forças de segurança pública. Essa conjuntura torna a previsão de crimes um

desafio estratégico, já que a dinâmica de atuação das quadrilhas é muitas vezes imprevisível e heterogênea em termos espaciais e temporais.

O impacto da violência urbana, especialmente o roubo de veículos, não se limita aos danos materiais. Esse tipo de crime gera medo, afeta a mobilidade urbana e provoca a retração de atividades econômicas, além de sobrecarregar o sistema de segurança pública. Nesse contexto, soluções tecnológicas que visam à prevenção e previsão de crimes, como o uso de algoritmos de machine learning, despontam como alternativas promissoras para a gestão eficiente de recursos e a criação de estratégias de segurança mais eficazes. A aplicação de técnicas de análise preditiva pode contribuir para a antecipação de ocorrências, otimizando a alocação de policiamento e reforçando a prevenção.

As melhores práticas em segurança pública demandam um conhecimento prévio da dinâmica criminal, o que auxilia na formulação de políticas de segurança e na organização das forças policiais, permitindo que estas superem um papel reativo ao crime e se tornem mais proativas (Roeder et al., 2015). A realização de previsões de variáveis criminais serve a diversos propósitos, incluindo o desdobramento tático de recursos policiais, alocação de pessoal e planejamento estratégico (Meijer e Wessels, 2019). O objetivo dessas previsões é fornecer uma ferramenta que aumente a consciência tática e estratégica das forças policiais, permitindo um planejamento de policiamento mais proativo e eficaz. Diante dessas possibilidades teóricas, a aplicação de métodos preditivos para antecipar tendências e identificar locais e indivíduos propensos à futura vitimização é amplamente adotada por departamentos de polícia nos Estados Unidos (Perry et al., 2013).

Apesar do sucesso internacional dessa prática, a literatura nacional carece de trabalhos propondo modelos capazes de antecipar tendências criminais. Silva et al. (2017) propõem o CrimeVis, uma ferramenta de visualização de dados capaz de relacionar informações socioeconômicas e criminais, enquanto outros estudos empregam técnicas de previsão e métodos de avaliação mais modestos em comparação com o que é praticado na literatura internacional (Provenza, Costa; Silva, 2015; Tashman, 2000).

Com a evolução da inteligência artificial (IA) em todo o mundo e a democratização da sua utilização garante a capacidade de atuação em diversos setores. Um dos campos de aplicação da IA é o aprendizado de máquina, utilizado para extrair resultados e identificar padrões, conhecido como machine learning. A popularização dessa técnica está gerando uma redução de custo da previsão, assim podemos prever mais rapidamente, de forma mais barata e

com maior acuracidade, permitindo usar previsões para resolver problemas que antes não eram possíveis. (Agrawal, 2018)

Utilizar o machine learning para auxiliar na previsão pode ser uma alternativa excelente para aprimorar a precisão dos resultados. A aplicação de inteligência artificial consegue processar e analisar rapidamente uma grande quantidade de informações, auxiliando na tomada de decisão (Alves; Prado, 2022).

O objetivo principal deste trabalho é avaliar qual abordagem de modelagem – árvores de decisão ou redes neurais – oferece maior precisão na previsão de roubos de veículos no município do Rio de Janeiro, contribuindo para o desenvolvimento de ferramentas preditivas que possam auxiliar autoridades e órgãos de segurança pública na alocação eficiente de recursos e no planejamento estratégico de políticas de prevenção. A precisão das modelagens será comparada utilizando o erro médio quadrático (Mean Squared Error - MSE).

MATERIAL E MÉTODOS

Coleta e preparação dos dados

Os dados foram obtidos através de data mining do site do Instituto de Segurança do Estado do Rio de Janeiro. A série histórica de crimes foi extraída no período entre janeiro de 2003 e fevereiro de 2024, abrangendo um total de 254 observações. Conforme a Tabela 1, entre as variáveis incluídas na base de dados estão:

Outros tipos de crimes: Roubos de transeunte, roubos de celular, roubos em coletivo, roubos de carga, roubos em residência, entre outros.

Variáveis temporais: Mês e ano de cada observação.

Variáveis geoespaciais: Localização dos crimes por área da cidade (exemplo: zona norte, zona sul, etc.).

Fatores sócio demográficos: Alguns fatores socioeconômicos das regiões incluídas, como renda média e densidade populacional, foram incorporados ao conjunto de dados para análise.

Tabela 1. Primeiras observações da base de dados

	roubo_transeunte	roubo_celular	roubo_em_coletivo	roubo_rua	roubo_carga	roubo_comercio	roubo_residencia	roubo_veiculo
1	1427	966	658	3051	357	589	172	3236
2	1532	963	691	3186	443	720	170	3169
3	1537	982	683	3202	310	640	178	2923
4	1529	1007	624	3160	303	702	169	2966
5	1558	1072	663	3293	345	712	188	3167
6	1426	963	540	2929	312	643	140	2686
7	1395	1018	426	2839	324	763	174	2541
8	1491	1040	349	2880	306	582	164	2520
9	1504	1202	377	3083	271	571	188	2524
10	1457	1277	345	3079	363	563	143	2701

Showing 1 to 10 of 254 entries, 8 total columns

Fonte: ISP-RJ – Instituto de Segurança do Rio de Janeiro (Elaborado pelo autor junho/2024)

Pré-processamento dos Dados

Antes da modelagem, o conjunto de dados foi submetido a um processo de pré-processamento, composto pelas seguintes etapas:

Tratamento de Dados Faltantes: as observações que apresentavam valores ausentes foram inicialmente investigadas. Caso o dado ausente fosse pontual e de uma variável crítica, a imputação foi feita utilizando técnicas como interpolação linear para variáveis temporais e médias móveis para variáveis de crimes. Se a ausência fosse extensa em algumas variáveis menos críticas, as observações foram removidas do conjunto de dados.

Codificação de Variáveis Categóricas: as variáveis categóricas, tipos de crimes, foram transformadas em variáveis *dummy*, adequando-as ao formato exigido pelos algoritmos de machine learning;

Normalização: Para o treinamento das redes neurais, os dados numéricos foram normalizados utilizando a técnica de escalonamento min-max. Essa técnica reescala os valores das variáveis para um intervalo de 0 a 1, conforme a seguinte fórmula:

$$X_{normalizado} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Onde X representa o valor original da variável, X_{min} e X_{max} são o valor mínimo e máximo da variável, respectivamente.

Análise Descritiva dos Dados

A Tabela 2 apresenta as estatísticas descritivas das variáveis utilizadas no estudo. A análise dessas estatísticas permite compreender a distribuição dos dados e possíveis padrões que possam impactar a previsão de roubos de veículos.

Tabela 2: Estatísticas descritivas das variáveis do conjunto de dados

Variável	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
Roubo Transeunte	1.384	3.344	4.751	4.744	6.136	9.669
Roubo Celular	301	589,2	900	1.058,2	1.366	2.626
Roubo Coletivo	303	547	704,5	765,5	892	1.628
Roubo Rua	2.839	4.840	6.064	6.568	7.768	13.833
Roubo Carga	131	287,2	362	428	503,5	1.329
Roubo Comércio	122	362,2	423	435,1	535	774
Roubo Residência	38	91	111	113,5	136	188
Roubo Veículo	1.165	1.981	2.466	2.594	2.976	5.358

(Fonte: Elaborado pelo autor com base nos dados do ISP-RJ)

A variável roubo_veiculo, que é o alvo do estudo, apresenta uma média de 2.594 registros por período e um máximo de 5.358 ocorrências. As variáveis preditoras apresentam distribuições distintas, sendo roubo_rua a variável com maior incidência média (6.568 registros) e roubo_residencia a de menor incidência (113,5 registros).

A análise dos quartis revela que os dados possuem distribuição assimétrica para algumas variáveis, como roubo_celular e roubo_carga, que apresentam valores máximos muito acima da média, indicando a presença de picos sazonais.

Correlação entre as Variáveis

Para compreender as relações entre as variáveis, foi gerada uma matriz de correlação entre roubo veículo e os demais crimes registrados. A Tabela 3 apresenta os coeficientes de correlação de Pearson.

Tabela 3: Correlação entre roubo veículo e outras variáveis

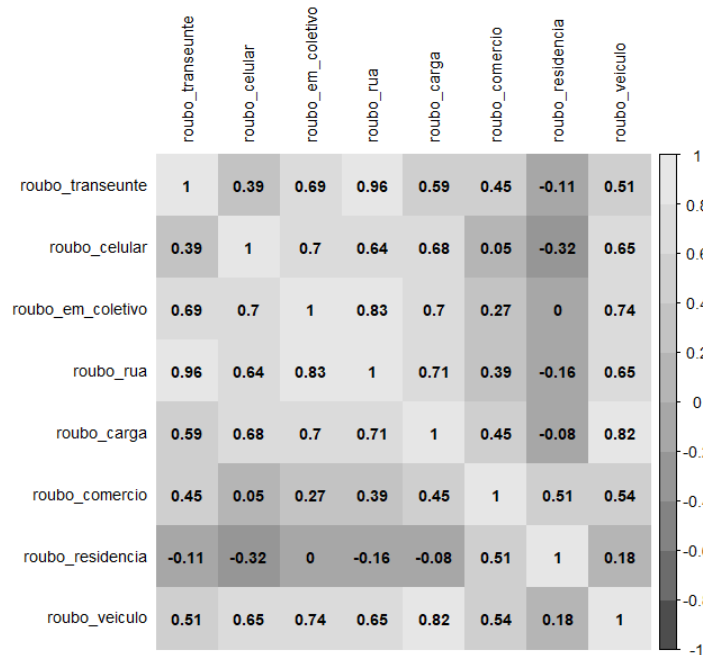
Variável	Correlação com roubo_veículo
Roubo Transeunte	0,51
Roubo Celular	0,65
Roubo Coletivo	0,74
Roubo Rua	0,65
Roubo Carga	0,82
Roubo Comércio	0,18
Roubo Residência	0,22

(Fonte: Elaborado pelo autor com base nos dados do ISP-RJ)

Os resultados indicam que roubo carga ($r = 0,82$) e roubo coletivo ($r = 0,74$) apresentam a maior correlação com roubo veículo, sugerindo que regiões com maior incidência desses crimes também registram mais roubos de veículos. Por outro lado, crimes como roubo a comércio e roubo a residência possuem correlações mais fracas (0,18 e 0,22, respectivamente), indicando menor impacto na previsão.

Para facilitar a visualização das relações entre as variáveis, a Figura 1 apresenta um *heatmap* com a matriz de correlação.

Figura 1: Matriz de Correlação



(Fonte: Elaborado pelo autor com base nos dados do ISP-RJ)

As variáveis com maior correlação foram utilizadas como principais preditoras no desenvolvimento dos modelos de **árvore de decisão** e **rede neural**, permitindo uma melhor adaptação dos algoritmos à dinâmica dos dados.

ÁRVORE DE DECISÃO

A árvore de decisão é um método baseado em uma estrutura hierárquica, onde cada nó interno representa uma variável de entrada (feature), cada ramo representa um resultado possível de uma decisão e cada nó folha representa uma previsão ou valor final (Breiman et al., 1984). O processo de divisão segue um critério de minimização de erro, de acordo com a métrica escolhida. Neste trabalho, utilizou-se o critério ANOVA para regressão, onde a divisão ocorre para minimizar a soma dos quadrados das diferenças entre os valores previstos e reais.

O procedimento seguido foi o seguinte:

1. Divisão dos Dados: O conjunto de dados foi dividido aleatoriamente em dois subconjuntos, um para o treinamento (80%) e outro para o teste (20%) para treinar o modelo e

avaliar sua performance. Para essa divisão, foi utilizada a função *SAMPLE* para selecionar aleatoriamente as amostras de treinamento e teste;

2. Treinamento: Para ajustar o modelo de árvore de decisão, foi utilizada a função *rpart* da biblioteca *rpart*. A variável "roubo_veiculo" foi definida como a variável dependente (target) e as demais variáveis preditoras foram incluídas no modelo. A técnica de particionamento utilizada foi o método **ANOVA** (*Analysis of Variance*), adequado para regressão. O hiperparâmetro **cp** (complexity parameter) foi ajustado para evitar o overfitting;

3. Predições e Avaliação: Após o treinamento, o modelo foi aplicado ao conjunto de teste e o erro médio quadrático (MSE) foi calculado para comparar a performance do modelo com as redes neurais.

Erro Médio Quadrático (MSE)

O erro médio quadrático (Mean Squared Error - MSE) é uma métrica utilizada para medir a precisão de um modelo preditivo. Ele calcula a média dos quadrados das diferenças entre os valores previstos e os valores reais:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Onde y_i são os valores reais e \hat{y}_i são os valores previstos pelo modelo. O MSE foi calculado tanto para o modelo de árvore de decisão quanto para a rede neural. Um MSE menor indica melhor precisão do modelo (Hyndman; Athanasopoulos, 2018). O MSE foi calculado para avaliar a precisão do modelo de árvore de decisão, permitindo sua comparação com a rede neural.

REDE NEURAL

Redes neurais são modelos computacionais inspirados no funcionamento do cérebro humano, compostos por camadas de neurônios artificiais. Elas são especialmente úteis para capturar padrões complexos e não lineares nos dados, sendo amplamente aplicadas em tarefas de regressão e classificação (Goodfellow; Bengio; Courville, 2016). No contexto da previsão de roubos de veículos no Rio de Janeiro, as RNAs são particularmente úteis por sua capacidade

de modelar padrões complexos e não lineares nos dados, características frequentemente encontradas em problemas de previsão criminal.

O procedimento seguido foi o seguinte:

1. Normalização dos dados:

Para a modelagem da rede neural, o conjunto de dados foi normalizado utilizando o escalonamento min-max. Esta técnica de normalização transforma os dados para que seus valores fiquem entre um mínimo e um máximo, geralmente entre zero e um, subtraindo-se o valor mínimo de cada variável e dividindo pelo intervalo (max - min) (Han; Kamber; Pei, 2011).

2. Divisão dos dados:

Os dados foram divididos em conjunto de treinamento (80%) e teste (20%). O conjunto de treino foi utilizado para ajustar o modelo de rede neural, enquanto o conjunto de teste foi empregado para avaliar o desempenho da rede em prever novos dados. A rede neural foi treinada utilizando a biblioteca *neuralnet* do *RStudio*.

3. Definição da Arquitetura de Rede:

A arquitetura especificada da rede neural consistia em quatro camadas ocultas, contendo 5, 4, 3 e 2 neurônios, respectivamente, e a saída da rede foi configurada como linear, apropriada para problemas de regressão.

4. Função de Ativação e Saída Linear:

A função de ativação utilizada nas camadas ocultas foi a função sigmoide, padrão na biblioteca *neuralnet*. Entretanto, como o problema é de regressão, a camada de saída foi configurada para ser linear, o que permite à rede neural realizar previsões contínuas.

Para avaliar o desempenho da rede neural, utilizou-se o conjunto de teste. O processo de avaliação envolveu a aplicação do modelo treinado sobre os dados de teste e a comparação das previsões geradas com os valores reais da variável roubo veículo.

O erro quadrático médio (**MSE**) foi utilizado como métrica para avaliar a precisão do modelo, sendo definido como a média das diferenças ao quadrado entre os valores previstos pela rede neural e os valores reais.

Após o treinamento da rede neural e a obtenção das previsões, gerou-se um gráfico comparando as previsões com os valores reais da variável roubo veículo. Esse gráfico permite visualizar a proximidade entre os dados reais e os valores previstos pela rede neural.

No gráfico, a linha vermelha representa os valores reais da variável roubo veículo, enquanto a linha azul corresponde aos valores previstos pela rede neural. Quanto mais próximas estiverem as duas linhas, melhor o desempenho da rede neural.

Após a visualização da rede neural, o modelo ajustado foi utilizado para realizar as previsões no conjunto de teste. As previsões e os dados reais foram reescalados para os intervalos originais da base de dados e o MSE das previsões reescaladas foi calculado, comparando-os com os valores reais reescalados.

RESULTADOS

1. Interpretação do Modelo de Árvore de Decisão

A Figura 2 apresenta a árvore de decisão ajustada no estudo, utilizada para prever o número de roubos de veículos no Rio de Janeiro com base em variáveis relacionadas a outros tipos de crimes. A estrutura hierárquica do modelo identifica os principais fatores explicativos, indicando como diferentes tipos de roubo influenciam a variável dependente.

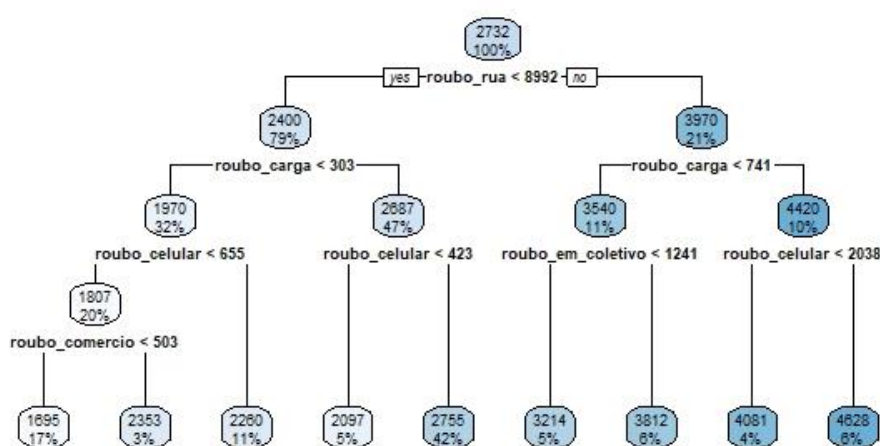


Figura 2: Árvore de decisão para previsão de roubos de veículos no Rio de Janeiro, mostrando os critérios de divisão baseados em variáveis como roubos de rua, carga, celular e comércio.

Fonte: Resultados originais da própria pesquisa (Elaborado pelo autor junho/2024)

A análise da árvore de decisão revela que a variável roubo de rua é o principal fator discriminante, representando o primeiro nível de decisão. Observa-se que:

- Quando o número de roubos de rua é inferior a 8.992, a segmentação avança para variáveis como roubo de carga e roubo de celular, identificando padrões mais específicos.
- Nos casos em que roubo de rua é elevado, o modelo considera outras variáveis, como roubo em coletivo e roubo de celular, para refinar a previsão.

Essas divisões indicam que crimes relacionados a locais públicos e ao transporte coletivo têm maior impacto na previsão de roubos de veículos. Além disso, a presença de variáveis como roubo de carga e roubo de comércio destaca fatores socioeconômicos que podem influenciar a dinâmica do crime.

2. COMPARAÇÃO DOS VALORES REAIS E PREDITOS PELA ÁRVORE DE DECISÃO

A Figura 3 apresenta a comparação entre os valores reais e os valores previstos pela árvore de decisão. O desempenho apresentado no gráfico destaca uma limitação comum das árvores de decisão: sua dificuldade em capturar padrões complexos ou variações pequenas nos dados. Isso pode ser útil em contextos onde decisões mais gerais são suficientes, mas prejudicial para análises detalhadas.

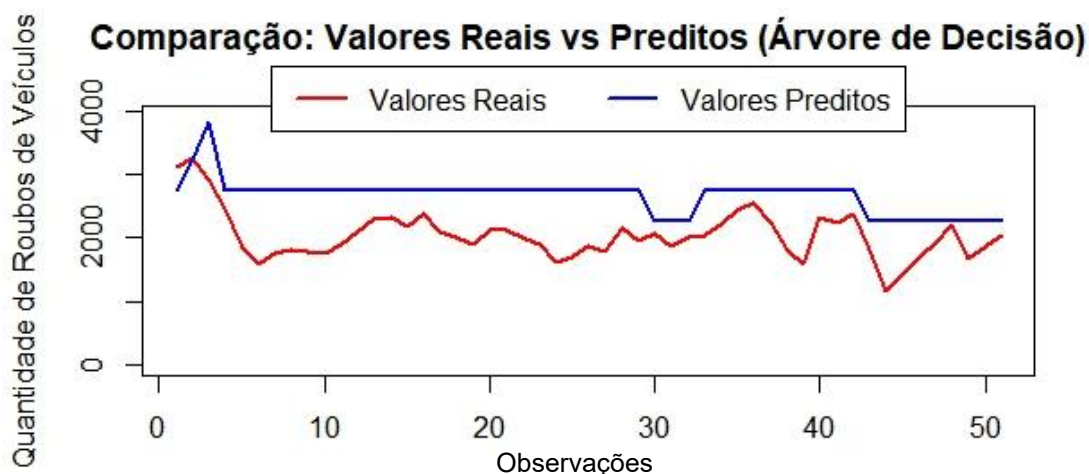


Figura 3: Comparação em valores reais e preditos pela modelo de árvore de decisão
 Fonte: Resultados originais da própria pesquisa (Elaborado pelo autor junho/2024)

3. Visualização da Rede Neural

A Figura 4 ilustra a arquitetura da rede neural utilizada no estudo, destacando as camadas e neurônios empregados no modelo.

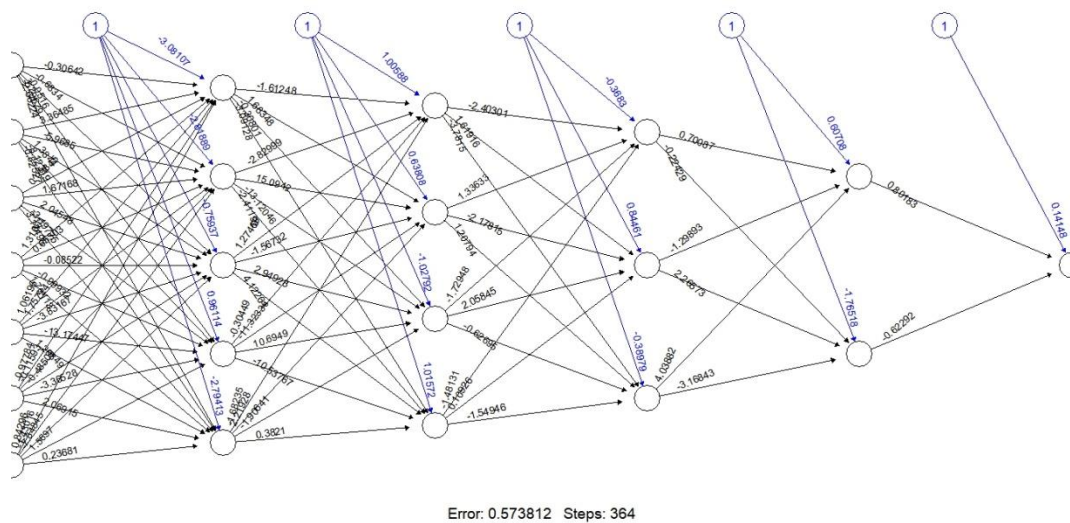


Figura 4: Visualização da rede neural modelada no Rstudio.

Fonte: Resultados originais da própria pesquisa (Elaborado pelo autor junho/2024)

4. Comparação dos Valores Reais e Preditos pela Rede Neural

A Figura 5 apresenta a comparação entre os valores reais e os valores previstos pela rede neural. Observa-se que o modelo apresenta um desempenho significativamente superior em comparação com a árvore de decisão, pois é capaz de capturar tendências gerais e variações mais detalhadas nos dados.



Figura 5: Comparação em valores reais e preditos pela Rede Neural

Fonte: Resultados originais da própria pesquisa (Elaborado pelo autor junho/2024)

5. Desempenho das Abordagens

Pela Tabela 4, foram identificados os seguintes resultados:

Tabela 4. Erro Médio Quadrático (MSE) das Modelagens

Modelagem	MSE
Árvore de Decisão	500.331
Rede Neural	125.081

O Erro Médio Quadrático (MSE) é uma medida da qualidade de um modelo de previsão, onde valores menores de MSE indicam previsões mais precisas. Com base nos valores de MSE, pode-se concluir que a Rede Neural apresenta melhor precisão preditiva em comparação com a Árvore de Decisão.

PRECISÃO DO MODELO

A Rede Neural apresentou um MSE significativamente menor (125.081) em comparação com a Árvore de Decisão (500.331). Isso indica que a Rede Neural faz previsões mais próximas dos valores reais.

Capacidade de Modelagem:

As redes neurais são conhecidas por sua capacidade de modelar relações complexas e não lineares nos dados. O melhor desempenho da Rede Neural sugere que o problema pode envolver tais relações, e a rede neural foi mais eficaz em capturá-las do que a Árvore de Decisão.

Complexidade do Modelo

A rede neural utilizada tem uma arquitetura relativamente complexa, com várias camadas ocultas (5, 4, 3, 2 neurônios). Essa complexidade adicional pode ter permitido ao modelo capturar padrões mais sutis nos dados, resultando em melhores previsões (Goodfellow, Bengio, & Courville, 2016).

CONCLUSÃO

Este trabalho apresentou uma abordagem para a previsão de roubos de veículos no município do Rio de Janeiro utilizando algoritmos de *machine learning*, com foco na comparação entre dois modelos preditivos: árvores de decisão e redes neurais. A análise dos resultados mostrou que as redes neurais apresentaram um desempenho superior em termos de erro médio quadrático (MSE), evidenciando sua maior capacidade de capturar padrões complexos e interações não lineares nos dados criminais. Em particular, a rede neural apresentou um MSE significativamente menor (125.081) em comparação com a árvore de decisão (500.331), o que indica uma precisão preditiva mais elevada.

Por outro lado, a árvore de decisão demonstrou ser uma ferramenta simples e de fácil interpretação, fornecendo uma visualização clara das regras de decisão e dos critérios utilizados no processo de previsão. Sua simplicidade, no entanto, resultou em uma menor capacidade de previsão, conforme refletido pelo maior MSE. Essa característica sugere que o uso de árvores de decisão pode ser mais adequado em contextos onde a interpretabilidade e a transparência do modelo sejam mais importantes do que a precisão preditiva, como em cenários onde as decisões precisam ser explicadas com clareza aos responsáveis pela segurança pública.

Implicações Práticas

Com base nos resultados, conclui-se que, embora as redes neurais tenham se mostrado mais eficientes na previsão de roubos de veículos, sua aplicabilidade prática em certos contextos, como no ambiente policial, pode ser limitada. Modelos preditivos baseados em redes

neurais tendem a ser considerados "caixas-pretas", dificultando a interpretação direta dos resultados por parte de gestores e operadores. Em situações onde a explicabilidade e a justificativa das previsões são cruciais para tomada de decisão, como na distribuição de recursos de segurança, a árvore de decisão pode ser mais apropriada, mesmo que com menor precisão.

Por outro lado, em situações onde a precisão das previsões é primordial - como no planejamento estratégico de longo prazo e na alocação otimizada de recursos - as redes neurais podem oferecer insights valiosos. Sua capacidade de modelar interações complexas entre variáveis socioeconômicas, geoespaciais e temporais as tornam uma ferramenta poderosa para identificar padrões criminais ocultos, o que pode melhorar a eficiência de ações preventivas.

Limitações e Desafios

No entanto, o uso de redes neurais também apresenta alguns desafios. A principal limitação observada neste trabalho foi o maior custo computacional associado ao treinamento das redes, especialmente quando a arquitetura do modelo inclui múltiplas camadas ocultas. Esse aumento de complexidade requer mais tempo de processamento e poder computacional, o que pode não ser viável em contextos operacionais com recursos limitados. Além disso, a necessidade de normalização dos dados e a seleção cuidadosa da arquitetura da rede neural são passos críticos que podem exigir conhecimentos especializados.

Além disso, há o desafio de explicar os resultados das redes neurais de maneira compreensível para tomadores de decisão. Em um contexto como o da segurança pública, onde a transparência e a justificativa das decisões são essenciais, a opacidade das redes neurais pode ser uma barreira significativa para sua implementação em larga escala. Esse fator reforça a importância de uma abordagem equilibrada, onde modelos mais simples e interpretáveis, como as árvores de decisão, possam complementar o uso de técnicas mais complexas.

Recomendações para Trabalhos Futuros

Para futuras pesquisas, recomenda-se explorar abordagens híbridas que combinem a interpretabilidade das árvores de decisão com a robustez das redes neurais. Por exemplo, o uso de modelos como *Random Forests* ou *Gradient Boosting* pode fornecer uma alternativa que balanceie precisão e interpretabilidade. Além disso, técnicas de *explainable AI* (XAI) poderiam ser exploradas para melhorar a explicabilidade dos modelos de redes neurais, tornando-os mais aplicáveis a cenários de segurança pública.

Outra direção promissora para trabalhos futuros é a inclusão de mais variáveis preditoras, como dados meteorológicos, eventos sociais e mudanças políticas, que podem impactar as taxas de criminalidade. A incorporação de diferentes fontes de dados poderia melhorar ainda mais a capacidade dos modelos de prever crimes e, assim, contribuir para um policiamento mais eficiente e proativo.

Em suma, este estudo demonstrou que redes neurais são ferramentas poderosas para a previsão de crimes em contextos onde a precisão preditiva é o fator mais importante. No entanto, sua complexidade e falta de interpretabilidade podem dificultar sua aplicação prática em contextos operacionais. A árvore de decisão, embora menos precise, oferece uma solução mais simples e transparente, sendo útil para casos onde a facilidade na interpretação dos resultados é essencial. A escolha do modelo ideal dependerá, portanto, do equilíbrio entre precisão, interpretabilidade e os recursos disponíveis para a implementação da solução em um contexto de segurança pública.

REFERÊNCIAS

AGRAWAL, Ajay. **The economics of artificial intelligence**. McKinsey Analytics, abril de 2018. Disponível em: <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-economics-of-artificialintelligence>. Acesso em 28/11/2025.

ALVES, P. C. P.; PRADO S. C. C. 2022. **Estudo Comparativo entre algoritmos de machine learning aplicados à previsão de series temporal do mercado financeiro**. In: IX Congresso de trabalhos de graduação faculdade de tecnologia e Mococa, 2022, São Paulo, SP, Brasil. Vol.9 N.1 A.2022.

BREIMAN, L., FRIEDMAN, J., STONE, C. J.; OLSHEN, R. A. (1984). **Classification and Regression Trees**. CRC Press.

GOODFELLO W, I., BENGIO, Y.; COURVILLE, A. (2016). **Deep Learning**. MIT Press. Disponível em: <https://www.deeplearningbook.org/>. Acesso em 28/11/2025.

HAN, J., KAMBER, M.; PEI, J. (2011). **Data Mining: Concepts and Techniques**. Elsevier.

HYNDMAN, R. J.; ATHANASOPOULOS, G. (2018). **Forecasting: principles and practice**. OTexts. Disponível em: <https://otexts.com/fpp3/>. Acesso em 28/11/2025.

Instituto de Segurança Pública. (2023). **Relatório Anual de Estatísticas de Segurança Pública no Rio de Janeiro**. Acesso em 2023, disponível em: www.isp.rj.gov.br. Acesso em 28/11/2025.

MEIJER, A.; WESSELS, M. **Predictive policing: review of benefits and drawbacks.** *International Journal of Public Administration*, v. 42, n. 12, p. 1031-1039, 2019. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/01900692.2019.1575664>. Acesso em 28/11/2025.

PERRY, W. et al. **Predictive policing: the role of crime forecasting in law enforcement operations.** [s.l.]: Rand Corporation, 2013. Disponível em: https://www.rand.org/pubs/research_reports/RR233.html. Acesso em 28/11/2025.

PROVENZA, M. M.; COSTA, J. F. da S.; SILVA, L. D. C. Análise e previsão de séries temporais do homicídio doloso no Rio de Janeiro. *Revista Cadernos de Estudos Sociais e Políticos*, v. 4, n. 7, p. 63-85, 2015. Disponível em: <https://www.e-publicacoes.uerj.br/index.php/CESP/article/view/18992/13810>. Acesso em 28/11/2025.

ROEDER, O. K. et al. **What caused the crime decline?** New York: Brennan Center for Justice, 2015. p. 142. Disponível em: https://www.brennancenter.org/sites/default/files/publications/What_Caused_The_Crime_Decline.pdf. Acesso em 28/11/2025.

SILVA, L. J. S. et al. **CrimeVis: an interactive visualization system for analyzing crime data in the State of Rio de Janeiro.** In: INTERNATIONAL CONFERENCE ON ENTERPRISE INFORMATION SYSTEMS, 19., 2017, Porto, Portugal. **Proceedings...** Porto: Iceis, 2017. p. 193-200. Disponível em: <http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0006258701930200>. Acesso em 28/11/2025.

TASHMAN, L. J. Out-of-sample tests of forecasting accuracy: analysis and review. *International Journal of Forecasting*, v. 16, n. 4, p. 437-450, 2000. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0169207000000650>. Acesso em 28/11/2025.

Recebido – 01/07/2025

Aprovado – 24/11/2025