

**MODELAGEM PREDITIVA DE ROUBOS EM CAMPO
GRANDE (MS) COM ALGORITMOS
SUPERVISIONADOS DE REGRESSÃO**

***PREDICTIVE MODELING OF ROBBERIES IN CAMPO
GRANDE (MS) USING SUPERVISED REGRESSION
ALGORITHMS***



**MODELAGEM PREDITIVA DE ROUBOS EM CAMPO GRANDE (MS)
COM ALGORITMOS SUPERVISIONADOS DE REGRESSÃO**
***PREDICTIVE MODELING OF ROBBERIES IN CAMPO GRANDE (MS)
USING SUPERVISED REGRESSION ALGORITHMS***

Wesley Fabricio Souza Silva¹
wesley.fabricio@ufms.br

Jamil Roberto Dagher Neto²
jamildagher@gmail.com

RESUMO

Este estudo investiga a aplicação de algoritmos de regressão supervisionada na previsão mensal de roubos na cidade de Campo Grande (MS), com base em dados operacionais da Polícia Militar de Mato Grosso do Sul (PMMS) no período de 2018 a 2023. A metodologia adotada consistiu na comparação entre diferentes técnicas de aprendizado de máquina, incluindo Regressão Linear, Ridge, Lasso, Árvore de Decisão, Random Forest e Rede Neural Multicamadas (MLP), e métodos estatísticos tradicionais baseados em médias e medianas. O objetivo foi verificar se os modelos supervisionados apresentam desempenho superior em relação aos *baselines* simplificados. A base de dados, composta por registros mensais agregados para todo o município, refletiu com fidelidade o comportamento das ações preventivas e sua relação com os registros de roubos. A avaliação dos modelos utilizou as métricas MAE (Erro Absoluto Médio), MSE (Erro Quadrático Médio) e R² (Coeficiente de Determinação). A validação empírica, realizada com dados reais de 2023, evidenciou a superioridade da abordagem baseada em aprendizado supervisionado, com destaque para o modelo Random Forest, que apresentou o menor MAE (10,43), e para a Rede Neural Multicamadas (MLP), que obteve o menor MSE (174,04) e o maior R² (0,44). Os resultados reforçam o potencial dos algoritmos supervisionados como ferramenta de apoio à tomada de decisão na segurança pública, promovendo uma gestão orientada por dados e mais capaz de antecipar cenários críticos.

Palavras-chave: Previsão de Crimes; Regressão Supervisionada; Segurança Pública; Aprendizado de Máquina; Policiamento Preditivo.

ABSTRACT

This study investigates the application of supervised regression algorithms for the monthly prediction of robberies in the city of Campo Grande (MS), based on operational records from the Military Police of Mato Grosso do Sul (PMMS) covering the period from 2018 to 2023. The adopted methodology consisted of comparing various machine learning techniques, including Linear Regression, Ridge, Lasso, Decision Tree, Random Forest, and Multilayer Perceptron (MLP), with traditional statistical methods relying on means and medians. The objective was to assess whether supervised models outperform simplified baseline approaches. The dataset, composed of monthly aggregated records for the entire municipality, reliably reflected the behavior of preventive policing actions and their relationship with reported robberies. Model evaluation was conducted using MAE (Mean Absolute Error), MSE (Mean Squared Error), and R² (Coefficient of Determination). Empirical validation with real data from 2023 demonstrated the superiority of the supervised learning approach, with Random Forest achieving the lowest MAE (10.43) and the MLP model achieving the lowest MSE (174.04) and highest R² (0.44). The findings highlight the potential of supervised algorithms as decision-support tools

¹Mestre em Computação Aplicada (UFMS), Especialista em Engenharia de Software (UCDB), Especialista em Engenharia de Machine Learning (UNOPAR), Graduado em Análise e Desenvolvimento de Sistemas (UNIGRAN). Policial Militar do Estado de Mato Grosso do Sul. Lattes: <http://lattes.cnpq.br/7695444372834695>. ORCID: 0009-0008-1988-9707.

²Especialista em Gestão de Riscos e Cibersegurança (FOCUS), Especialista em Segurança da Informação (FOCUS), Algoritmos e Estruturas de Dados (FOCUS), Graduação em Sistemas Para Internet (IFMS). Policial Militar do Estado de Mato Grosso do Sul. Lattes: <http://lattes.cnpq.br/9015917187708078>. ORCID: 0009-0007-4959-3029.

in public security, promoting a data-driven approach to management and enhancing the ability to anticipate critical scenarios.

Keywords: Crime Prediction; Supervised Regression; Public Security; Machine Learning; Predictive Policing.

1. INTRODUÇÃO

A segurança pública configura-se como um dos pilares fundamentais para a preservação da ordem social e da qualidade de vida, especialmente nos centros urbanos brasileiros, marcados por desigualdades, dinâmicas territoriais complexas e variabilidade nos padrões de criminalidade. Entre os crimes patrimoniais, o roubo se destaca tanto pela sua frequência quanto pela sua imprevisibilidade, gerando elevados níveis de medo e sensação de insegurança na população.

A cidade de Campo Grande (MS), composta por sete regiões urbanas com características demográficas e socioeconômicas heterogêneas, ilustra bem esse cenário. Nesse contexto, a Polícia Militar do Estado de Mato Grosso do Sul (PMMS) realiza rotineiramente ações de policiamento preventivo, tais como abordagens, rondas, mandados cumpridos e prisões em flagrante. No entanto, os desafios contemporâneos da segurança pública exigem mais do que presença ostensiva: demandam inteligência preditiva e capacidade de antecipação estratégica.

A aplicação de algoritmos de aprendizado de máquina supervisionado, especialmente técnicas de regressão, tem se mostrado promissora em diversos domínios para a previsão de valores numéricos contínuos com base em dados históricos. Essas técnicas são amplamente empregadas em diferentes áreas, como na previsão de preços de imóveis, aplicando modelos como regressão linear, regressão de árvores e redes neurais para estimar valores de mercado com base em características estruturais e localização dos imóveis (Antipov; Pokryshevskaya, 2012; Bin *et al.*, 2021). Também são utilizadas na estimativa de demanda energética, onde modelos supervisionados auxiliam na previsão do consumo futuro, possibilitando otimização na geração e distribuição de energia elétrica (Deb *et al.*, 2017; Khuntia *et al.*, 2021). Na antecipação de indicadores econômicos, técnicas de regressão têm sido aplicadas na previsão de inflação, PIB e índices de mercado financeiro, contribuindo para análises de risco e planejamento econômico (Medeiros *et al.*, 2021; Kou *et al.*, 2021). No campo da previsão de volumes criminais, estudos vêm demonstrando que modelos de aprendizado de máquina são capazes de antecipar a incidência de determinados tipos de crime com base em padrões históricos e variáveis contextuais, permitindo o aprimoramento das estratégias de segurança pública (Li *et al.*, 2020; Jindal *et al.*, 2021; Aziz *et al.*, 2022), como se propõe no presente estudo.

No campo da segurança pública, embora essa abordagem ainda seja incipiente no Brasil, particularmente no que se refere à previsão quantitativa de crimes como o roubo, ela apresenta um

potencial significativo para subsidiar o planejamento estratégico e a alocação de recursos operacionais. A partir da análise de séries temporais de ações preventivas sistematicamente registradas pela PMMS, torna-se viável investigar a possibilidade de estimar a quantidade futura de roubos com base na frequência e no tipo de atuação policial. Essa perspectiva está alinhada com a literatura sobre policiamento orientado por dados (Batista, 2021), que enfatiza o valor das evidências empíricas na formulação de decisões mais eficazes e na promoção de práticas preventivas orientadas por resultados.

Ainda que a relação entre ações preventivas e a variação nos índices criminais envolve múltiplos fatores e mediações contextuais, como desorganização social, eficácia coletiva e estrutura urbana, os modelos supervisionados podem ser capazes de capturar parte relevante dessa dinâmica. Ao identificar padrões históricos de correlação, tais modelos não têm como objetivo explicar integralmente as causas da criminalidade, mas sim prever cenários com maior risco relativo, subsidiando uma atuação policial mais eficiente.

A crescente digitalização dos registros operacionais viabiliza a coleta sistemática de dados sobre as ações de policiamento, abrindo caminho para o uso de técnicas analíticas avançadas. Nesse sentido, a construção de modelos matemáticos para previsão de roubos representa não apenas um avanço técnico, mas também uma evolução institucional rumo a uma segurança pública mais proativa, racional e baseada em evidências.

Este estudo tem como objetivo aplicar e comparar diferentes algoritmos de regressão supervisionada, incluindo Random Forest, Regressão Linear, Lasso, Ridge e Redes Neurais, no contexto da segurança pública de Campo Grande (MS). O foco central consiste em verificar se esses modelos supervisionados apresentam desempenho superior em relação a métodos estatísticos simplificados, como médias e medianas históricas, utilizados como linhas de base (*baselines*). A pesquisa também busca validar empiricamente os modelos com dados reais de 2023, utilizando métricas de erro preditivo e capacidade explicativa, contribuindo para a consolidação de uma cultura institucional orientada por dados no âmbito da Polícia Militar.

2. FUNDAMENTAÇÃO TEÓRICA – REVISÃO BIBLIOGRÁFICA

O uso de técnicas de aprendizado de máquina (*machine learning*) na área de segurança pública tem se consolidado como uma abordagem promissora para análise e previsão de padrões criminais. Essa evolução reflete a transformação digital da gestão pública, impulsionada pela crescente disponibilidade de dados e pela necessidade de respostas mais eficientes frente à complexidade dos fenômenos criminais. Segundo Mitchell (1997), o aprendizado de máquina refere-se a sistemas

computacionais capazes de aprimorar seu desempenho em tarefas específicas a partir de dados, sem necessidade de programação explícita.

No campo do aprendizado supervisionado, a regressão destaca-se pela capacidade de prever valores numéricos contínuos, como o volume de crimes em determinada localidade. Modelos como Regressão Linear, Ridge, Lasso, Árvores de Decisão, Random Forest e Redes Neurais Multicamadas (MLP) são frequentemente empregados, sendo avaliados por métricas como o Erro Absoluto Médio (MAE), Erro Quadrático Médio (MSE) e Coeficiente de Determinação (R^2), que mensuram respectivamente a média dos erros, a penalização para grandes desvios e a proporção da variabilidade explicada (Bishop, 2006; Hastie, Tibshirani e Friedman, 2009).

Estudos nacionais e internacionais ilustram a aplicabilidade desses métodos no contexto criminal. Shingleton (2012) utilizou regressão linear múltipla para estimar taxas criminais em Salinas, Califórnia, alcançando um R^2 de 0,75, o que demonstra boa capacidade explicativa, embora limitada por fatores externos, como variações econômicas e demográficas. De forma semelhante, Awal et al. (2016) aplicaram regressão linear simples aos dados criminais de Bangladesh, identificando correlação positiva entre crescimento populacional e aumento da criminalidade, embora sem apresentar métricas formais que permitam comparação mais robusta.

Li et al. (2020) avançaram na aplicação de modelos mais sofisticados, comparando algoritmos de séries temporais e de aprendizado de máquina na previsão de crimes em Chicago. O modelo XGBoost destacou-se com R^2 de 0,96 e desempenho superior ao ARIMA, evidenciando a eficácia dos métodos baseados em árvores em cenários urbanos complexos. De modo similar, Jindal et al. (2021) demonstraram que o modelo LightGBM obteve R^2 de 0,89 na previsão criminal em Maharashtra, Índia, com desempenho superior a outros algoritmos testados.

Aziz et al. (2022) também corroboraram a eficiência dos modelos baseados em árvores. Aplicando Random Forest Regression para previsão de crimes na Índia, alcançaram R^2 ajustado de 0,96 e MAPE de 20,27%. Para a previsão específica de furtos, o modelo obteve R^2 de 0,97 e MAPE de 16,57%, demonstrando elevada acurácia.

Apesar dos avanços metodológicos, observa-se uma lacuna recorrente na literatura: a escassez de estudos que comparem diretamente modelos supervisionados com abordagens estatísticas simplificadas, como médias ou medianas históricas, práticas ainda comuns no ambiente institucional. Essa ausência compromete a avaliação do custo-benefício na adoção de modelos mais sofisticados, que, embora mais precisos, demandam maior capacidade técnica para sua compreensão e operacionalização.

Além disso, poucos trabalhos incorporam variáveis diretamente relacionadas à atuação operacional das forças de segurança, como o número de abordagens, prisões ou mandados cumpridos.

Tal omissão limita a compreensão sobre como as ações preventivas impactam a dinâmica criminal. Como alertam Lopes e Russo (2022), a adoção irrestrita de soluções algorítmicas, sem considerar os aspectos sociais, urbanos e comunitários, pode conduzir a interpretações reducionistas do fenômeno criminal.

Por fim, é necessário reconhecer que modelos preditivos não estabelecem relações de causalidade, mas sim padrões baseados em dados históricos (Breiman, 2001). Assim, este estudo posiciona-se no campo da predição aplicada, utilizando dados operacionais da PMMS para estimar a tendência de roubos, sem pretender esgotar a complexidade social do crime. A proposta, portanto, é avaliar se os ganhos estatísticos proporcionados pelos modelos supervisionados justificam sua adoção em detrimento de métodos tradicionais, contribuindo para uma gestão da segurança pública mais técnica, eficiente e orientada por dados.

3. METODOLOGIA

Este estudo adota uma abordagem quantitativa, aplicada e comparativa, estruturada para avaliar a eficácia de algoritmos de regressão supervisionada na previsão da quantidade mensal de roubos na cidade de Campo Grande (MS). O foco central consiste em verificar se esses modelos preditivos apresentam desempenho superior em relação a métodos estatísticos simplificados, como médias e medianas históricas, comumente utilizados como referência na prática institucional.

Além disso, como parte do delineamento experimental, foi conduzida uma etapa comparativa entre duas estratégias de modelagem: (i) segmentação por regiões urbanas específicas da cidade e (ii) modelagem consolidada em escala municipal. A análise indicou que a abordagem consolidada apresentou desempenho preditivo superior e maior estabilidade estatística, razão pela qual foi adotada para apresentação dos resultados finais.

O processo metodológico contempla: (i) a coleta e organização da base de dados; (ii) o pré-processamento das variáveis; (iii) a aplicação dos modelos preditivos; e (iv) a avaliação do desempenho comparativo.

3.1 Coleta e Organização dos Dados

Os dados utilizados neste estudo foram fornecidos pela Polícia Militar do Estado de Mato Grosso do Sul (PMMS), abrangendo o período de janeiro de 2018 a dezembro de 2023. A base contém registros mensais de ações preventivas realizadas em todo o município de Campo Grande. Entre as variáveis estão: número de abordagens a pessoas e veículos, prisões em flagrante, cumprimento de mandados de prisão, rondas ostensivas e recuperação de veículos furtados ou roubados. A variável de interesse (*target*) é a quantidade mensal de roubos registrados.

Para garantir consistência temporal, cada linha da base representa um mês completo de ações. As variáveis foram tratadas para garantir homogeneidade numérica, e os dados foram organizados em dois blocos: treinamento e validação cruzada (2018 a 2022) e validação externa com dados de 2023.

A Tabela 1 a seguir apresenta um recorte da base de dados utilizada, contendo os principais atributos relacionados às ações preventivas empregados na modelagem preditiva.

Tabela 1 – Amostra dos dados de ações preventivas utilizados no estudo.

Mês/Ano	Pessoas Abordadas	Veículos Abordados	Mandados Cumpridos	Prisões em Flagrante	Roubos
Jan/2018	1245	532	11	9	103
Fev/2018	1189	510	10	7	97
Mar/2018	1343	578	12	11	112
Abr/2018	1292	540	13	10	101
Mai/2018	1201	495	9	8	95

Fonte: PMMS (dados preventivos processados pelo autor).

3.2 Pré-processamento e Normalização

O conjunto de dados passou por uma etapa de limpeza e transformação. Variáveis categóricas foram descartadas, e colunas com baixa variabilidade ou alta colinearidade foram removidas para evitar redundância e sobreajuste. As variáveis numéricas foram padronizadas por meio do método StandardScaler, que ajusta os valores para média zero e desvio padrão um, evitando viés em algoritmos sensíveis à magnitude dos dados.

3.3 Modelagem e Abordagens Comparativas

Foram aplicados seis algoritmos de regressão supervisionada, selecionados por sua diversidade metodológica e ampla aceitação na literatura:

- **Regressão Linear:** modelo base que assume relação linear entre as variáveis;
- **Ridge Regression:** regressão com penalização L2 para evitar sobreajuste;
- **Lasso Regression:** similar ao Ridge, com penalização L1, que favorece a seleção de variáveis;
- **Árvore de Decisão:** modelo hierárquico baseado em regras sucessivas;
- **Random Forest:** técnica de ensemble composta por múltiplas árvores com alta robustez;

- **Rede Neural Multicamadas (MLP):** modelo inspirado em redes neurais biológicas, adequado para capturar padrões não lineares.

Adicionalmente, foram testadas quatro abordagens estatísticas tradicionais, utilizadas como linhas de base (*baselines*) para fins comparativos:

- **Média Global:** valor médio de roubos ao longo de todo o período de treinamento;
- **Mediana Global:** valor central da série histórica agregada;
- **Média Mensal:** média específica para cada mês do ano (ex.: todos os janeiros);
- **Mediana Mensal:** mediana por mês, representando o comportamento típico mensal.

3.4 Avaliação de Desempenho

Os modelos foram avaliados com base em três métricas amplamente consolidadas na literatura para tarefas de regressão:

- **MAE (*Mean Absolute Error*):** erro médio absoluto entre os valores reais e previstos, mensurando a média dos desvios absolutos;
- **MSE (*Mean Squared Error*):** erro quadrático médio, que penaliza de forma mais intensa os erros maiores, sendo sensível a *outliers*;
- **R² (Coeficiente de Determinação):** proporção da variância dos dados que é explicada pelo modelo, funcionando como indicador de ajuste.

As métricas foram calculadas tanto para o período de validação cruzada (2018 a 2022) quanto para uma validação externa com dados reais de 2023, permitindo uma análise comparativa robusta entre os diferentes métodos testados, tanto sob a ótica da generalização quanto da aplicabilidade prática.

Para garantir robustez estatística na avaliação dos modelos, foi empregada a técnica de validação cruzada do tipo *k-fold*, com *k* igual a 5, aplicada especificamente à tarefa de regressão. Esse método é amplamente recomendado tanto em problemas de classificação quanto de regressão, sendo particularmente eficaz na estimativa da capacidade de generalização dos modelos preditivos.

No contexto da regressão, o *k-fold* opera dividindo aleatoriamente o conjunto de dados do período de 2018 a 2022 em cinco subconjuntos aproximadamente iguais. Em cada iteração, quatro subconjuntos são utilizados para o treinamento do modelo e o subconjunto restante para teste. Esse processo se repete até que cada subconjunto tenha sido utilizado uma única vez como dado de validação. O desempenho final do modelo é então calculado pela média dos resultados obtidos nas

cinco iterações, o que confere uma avaliação mais estável e menos suscetível a flutuações decorrentes de partições específicas dos dados.

A utilização do *k-fold* na regressão é plenamente justificada, uma vez que essa técnica não se restringe a problemas de classificação. Seu objetivo principal é estimar a capacidade do modelo de generalizar para novos dados, independentemente de se tratar de previsão de valores contínuos ou categóricos. No presente estudo, a escolha do valor de $k=5$ reflete um equilíbrio entre robustez estatística e viabilidade computacional, sendo adequado para bases de dados de porte moderado, como a utilizada. Valores de k muito baixos tenderiam a gerar alta variabilidade nas estimativas, enquanto valores excessivamente altos, como no leave-one-out, implicariam custos computacionais elevados sem ganhos proporcionais na acurácia estimada.

Adicionalmente, foi realizada uma validação externa, utilizando dados completamente novos, correspondentes ao ano de 2023, que não participaram de nenhuma etapa de treinamento ou validação cruzada. Essa estratégia permite avaliar o desempenho dos modelos em um cenário prospectivo, simulando as condições reais de aplicação no contexto operacional da segurança pública, conferindo maior credibilidade e aplicabilidade prática aos resultados obtidos.

4. RESULTADOS

4.1 Avaliação por Validação Cruzada (2018–2022)

Inicialmente, os modelos supervisionados foram avaliados com base nos dados do período de 2018 a 2022, utilizando validação cruzada com o objetivo de simular o desempenho preditivo interno dos algoritmos. Essa metodologia permite uma avaliação mais robusta, ao reduzir o viés decorrente da divisão específica entre treino e teste. A intenção principal desta etapa foi verificar quais algoritmos apresentavam maior capacidade de generalização e ajuste aos padrões históricos observados nas ações preventivas e suas relações com os registros mensais de roubos. A Tabela 2 apresenta os resultados obtidos para três métricas clássicas de regressão: o erro absoluto médio (MAE), o erro quadrático médio (MSE) e o coeficiente de determinação (R^2).

Tabela 2 – Desempenho dos modelos supervisionados (validação cruzada).

Modelo	MAE	MSE	R^2
Random Forest	9,93	173,91	0,499
Rede Neural (MLP)	10,71	174,04	0,445
Regressão Linear	10,70	195,25	0,4378
Ridge	10,70	195,25	0,4378
Lasso	10,77	197,26	0,4320
Árvore de Decisão	15,53	416,68	-0,199
Média Global	15,26	362,06	-0,0426
Mediana Global	15,76	403,51	-0,1620

Fonte: Autoria própria.

A partir dos resultados apresentados na Tabela 2, observa-se que os modelos supervisionados, especialmente o Random Forest e a Rede Neural Multicamadas (MLP), obtiveram os melhores desempenhos na etapa de treinamento e teste. O modelo Random Forest destacou-se por apresentar o menor erro absoluto médio (MAE = 9,93) e o maior coeficiente de determinação ($R^2 = 0,499$), o que indica maior precisão nas previsões e maior capacidade explicativa em relação à variabilidade dos dados. A MLP também apresentou desempenho expressivo, com MAE = 10,71 e $R^2 = 0,445$, demonstrando boa adaptação aos padrões históricos observados.

Em contrapartida, os métodos estatísticos tradicionais, como a média global e a mediana global, apresentaram desempenho significativamente inferior. Os valores negativos de R^2 registrados para essas abordagens revelam sua baixa capacidade de explicar a variabilidade nos dados de roubos a partir de tendências centrais, o que evidencia as limitações de modelos simplificados em cenários com alta variabilidade. Da mesma forma, o modelo de Árvore de Decisão isolada obteve $R^2 = -0,199$, resultado também insatisfatório, possivelmente devido à sua maior sensibilidade a flutuações sazonais e à menor robustez frente a ruídos em séries temporais.

Esses achados reforçam o potencial dos algoritmos supervisionados em capturar padrões não lineares e multivariados presentes nas ações preventivas desenvolvidas pela Polícia Militar. Ainda que os resultados analisados referem-se à fase de treinamento e validação cruzada, os modelos já demonstraram desempenho superior aos métodos estatísticos simplificados, evidenciando sua capacidade de adaptação a cenários complexos mesmo em condições controladas de teste. Isso sugere que, mesmo antes da validação externa com dados futuros, os algoritmos supervisionados oferecem um suporte mais robusto e promissor à previsão estratégica em segurança pública, contribuindo para uma atuação mais proativa e orientada por evidências.

4.2 Validação com Dados Reais (2023)

Após a etapa de validação cruzada, os modelos foram submetidos a uma validação empírica utilizando os dados reais do ano de 2023, que **não foram incluídos no processo de treinamento**. Esta etapa teve como objetivo avaliar a capacidade de generalização dos algoritmos em um cenário prático e não observado anteriormente, simulando sua aplicação real no contexto da segurança pública.

A tabela 3 apresenta os resultados obtidos na validação externa, considerando as mesmas métricas adotadas na fase anterior: erro absoluto médio (MAE), erro quadrático médio (MSE) e coeficiente de determinação (R^2).

Tabela 3 – Desempenho dos modelos com os dados reais de 2023.

Modelo	MAE	MSE	R^2
Random Forest	10,43	210,67	0,359

Modelo	MAE	MSE	R ²
Rede Neural (MLP)	10,71	174,04	0,445
Lasso	12,14	233,95	0,293
Regressão Linear	12,29	240,02	0,276
Ridge	12,29	240,02	0,276
Média Mensal	15,42	320,69	0,023
Mediana Mensal	15,11	326,67	0,042
Média Global	14,77	350,41	0,007
Mediana Global	15,13	378,49	0,092

Fonte: Autoria própria.

Conforme apresentado na tabela 3, os resultados confirmam a superioridade dos modelos supervisionados em relação aos métodos estatísticos tradicionais. A rede neural MLP obteve o maior coeficiente de determinação ($R^2 = 0,445$), demonstrando excelente capacidade de adaptação aos padrões temporais dos dados históricos. Já o modelo Random Forest registrou o menor erro absoluto médio (MAE = 10,43), destacando-se como o mais preciso em termos práticos, com menor discrepância entre os valores reais e os previstos.

Por outro lado, ainda de acordo com a Tabela 3, os modelos baseados em médias e medianas, tanto globais quanto mensais, apresentaram coeficientes de determinação próximos de zero. Isso indica que suas previsões foram, na prática, menos eficazes do que simplesmente utilizar a média do período de treinamento como uma estimativa constante. Esses resultados evidenciam as limitações dos métodos estatísticos simplificados em capturar a complexidade dos dados criminais, especialmente diante da influência de múltiplos fatores operacionais e das variações sazonais.

Diante dos resultados obtidos, destaca-se que os modelos supervisionados não apenas apresentaram métricas estatísticas superiores, mas também produziram previsões com níveis de erro absolutamente aceitáveis do ponto de vista prático. Por exemplo, o modelo Random Forest, que obteve o menor erro absoluto médio (MAE = 10,43), indica que, em média, o desvio entre o valor previsto e o número real de roubos por mês foi de aproximadamente 10 ocorrências. Considerando que os registros mensais de roubos em Campo Grande variam entre 30 e 70 casos, esse nível de erro representa uma margem de variação relativa inferior a 20%, o que é considerado altamente satisfatório em contextos operacionais marcados por alta variabilidade.

Esses resultados reforçam que, além de apresentarem desempenho estatístico superior, os algoritmos supervisionados alcançaram níveis de acurácia compatíveis com o uso prático. Isso os torna aptos a serem integrados como ferramentas de apoio à decisão tática e estratégica na gestão da segurança pública. Ao substituir métodos simplificados por modelos preditivos mais robustos, abre-

instrumentos de apoio à decisão, cuja natureza é estatística e associativa, e não determinista, especialmente em contextos multifatoriais, como o da segurança pública.

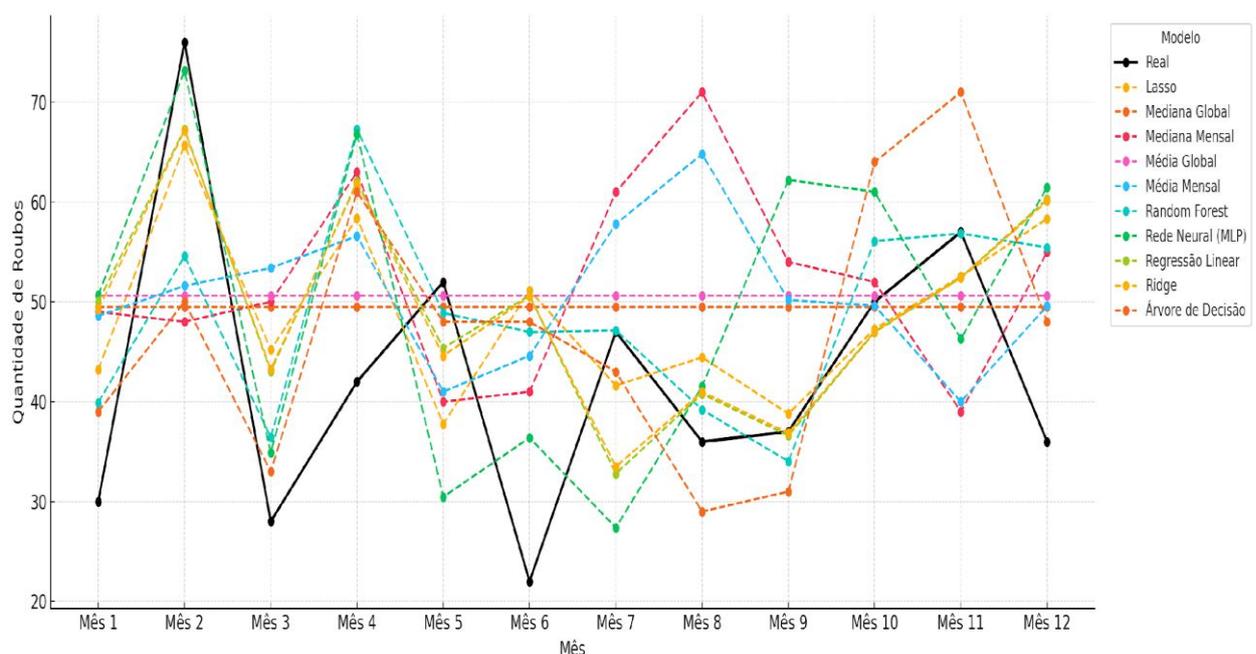
A realização da validação externa é particularmente relevante, pois demonstra a robustez e a capacidade de generalização dos modelos supervisionados quando aplicados a dados reais e inéditos. O bom desempenho alcançado, especialmente pelos modelos Random Forest e MLP, indica que essas técnicas possuem potencial concreto de aplicação no planejamento tático das forças de segurança, possibilitando a antecipação de cenários de maior risco e a orientação mais estratégica das ações policiais. Trata-se de um avanço importante rumo a uma gestão mais inteligente, orientada por dados e voltada à redução de danos e alocação mais eficiente dos recursos operacionais.

4.3 Comparativo Mensal de Previsões (2023)

Para complementar a análise empírica apresentada na subseção anterior, foi realizada uma comparação direta entre os valores reais de roubos registrados em cada mês de 2023 e as respectivas previsões geradas pelos modelos supervisionados, bem como pelas abordagens estatísticas tradicionais. Essa etapa visa avaliar, de forma prática e sequencial, a proximidade entre previsões e observações reais, permitindo identificar o modelo com melhor desempenho mês a mês.

A figura 2 apresenta a evolução dos roubos ao longo dos 12 meses de 2023, comparando a linha de valores reais (preta e contínua) com as curvas preditivas produzidas por cada modelo testado (linhas tracejadas coloridas).

Figura 2 – Evolução Mensal: Roubos Reais vs Previsões (2023).



Fonte: Autoria própria.

A análise visual da figura 2 evidencia que o modelo Random Forest, representado pela linha azul-claro, manteve trajetória consistentemente próxima dos dados reais, com destaque para os meses de janeiro, abril, agosto, setembro e novembro. Sua capacidade de acompanhar oscilações abruptas, como a queda entre fevereiro e março e o aumento observado no final do ano, demonstra boa generalização em contextos instáveis e não lineares.

A Rede Neural Multicamadas (MLP), embora tenha apresentado desempenho satisfatório em meses como fevereiro e dezembro, revelou maior variabilidade nas demais previsões, indicando uma sensibilidade acentuada às flutuações de padrão. Sua performance sugere que, embora tenha potencial para captar relações complexas, ainda carece de maior regularidade em séries temporais mais voláteis.

As abordagens estatísticas tradicionais, como média global, mediana global, média mensal e mediana mensal, exibem comportamentos quase estacionários, representados por linhas horizontais com pouca ou nenhuma variação ao longo do tempo. Isso reflete sua incapacidade de reagir às oscilações reais do fenômeno criminal, resultando em previsões sistematicamente distantes dos valores observados, principalmente nos meses com variações mais acentuadas.

O modelo de Árvore de Decisão apresentou desempenho irregular: embora tenha se aproximado dos dados reais em períodos pontuais (como março e julho), registrou distorções significativas em meses como agosto e outubro. Essa instabilidade pode ser atribuída à sua alta sensibilidade a pequenas variações nos dados de entrada, o que compromete sua robustez em contextos operacionais com elevada dinamicidade.

A análise gráfica reforça o papel estratégico de modelos supervisionados mais robustos, como o Random Forest e, em menor escala, a MLP, na produção de previsões aderentes à realidade mensal, característica fundamental para o direcionamento tático e a gestão eficiente de recursos no âmbito da segurança pública.

Do ponto de vista técnico, a capacidade adaptativa dos modelos supervisionados se destaca como diferencial crítico. O Random Forest, por combinar múltiplas árvores de decisão e realizar agregação por votação, apresenta alta resiliência a ruídos e variações sazonais. A MLP, por sua vez, é capaz de capturar padrões não lineares complexos, desde que adequadamente calibrada e ajustada. Em contraste, os métodos estatísticos convencionais, por não incorporarem mecanismos de aprendizado contínuo, demonstram baixa responsividade às mudanças estruturais nos dados, o que os torna menos eficazes em ambientes operacionais dinâmicos como o da criminalidade urbana.

5. CONSIDERAÇÕES FINAIS

Os resultados deste estudo demonstram de forma consistente a superioridade dos algoritmos de regressão supervisionada na tarefa de previsão de roubos em Campo Grande, Mato Grosso do Sul, tanto em métricas agregadas quanto na acurácia das estimativas mensais. Mais do que comprovar o desempenho estatístico dos modelos, o estudo validou sua aplicabilidade prática como ferramenta de apoio ao planejamento tático e à tomada de decisão operacional no âmbito da Polícia Militar. Dentre os modelos testados, o Random Forest destacou-se por apresentar o menor erro absoluto médio, enquanto a Rede Neural Multicamadas (MLP) obteve o maior poder explicativo, medido pelo coeficiente de determinação (R^2), na validação externa com dados reais de 2023. Esses achados reforçam o potencial das técnicas de aprendizado de máquina como instrumentos de apoio à gestão da segurança pública, promovendo a modernização das estratégias de combate à criminalidade.

Por outro lado, a análise também revelou que abordagens estatísticas tradicionais, como médias e medianas globais ou mensais, foram insuficientes para capturar a variabilidade dinâmica do fenômeno criminal. Tais métodos apresentaram coeficientes de determinação próximos de zero e erros consideravelmente superiores, o que reafirma sua limitação para contextos operacionais complexos, caracterizados pela interação não linear de múltiplas variáveis, fortemente dependentes dos contextos sociais, territoriais e temporais.

O experimento comparativo entre a modelagem segmentada por áreas específicas e a abordagem consolidada da cidade como um todo indicou que, ao menos com a base de dados disponível, a modelagem em escala municipal plena proporciona maior estabilidade estatística e melhor desempenho preditivo. Esse resultado decorre diretamente do maior volume de observações e da diversidade comportamental presente no conjunto consolidado, elementos fundamentais para a construção de modelos mais robustos, estáveis e generalizáveis.

Importante destacar, entretanto, que os valores de R^2 obtidos, embora superiores aos das abordagens tradicionais, foram relativamente modestos, atingindo no máximo 0,44. Esse resultado evidencia que, embora os modelos sejam úteis para apoiar o planejamento operacional, há uma parcela significativa da variabilidade dos registros de roubos que permanece não explicada. Esse comportamento sugere que fatores não observados na base atual, como variáveis socioeconômicas, demográficas, geoespaciais, ambientais e aspectos relacionados à dinâmica urbana, exercem influência relevante sobre os padrões criminais.

Neste contexto, o presente estudo demonstra que ações preventivas, como abordagens de pessoas e veículos, rondas ostensivas e cumprimento de mandados de prisão, apresentam correlação estatisticamente significativa com a variação mensal dos registros de roubos. Essa constatação oferece subsídios concretos para o redirecionamento tático de recursos, contribuindo para uma gestão mais

orientada por evidências. Por outro lado, os resultados também reforçam que a criminalidade é um fenômeno multifacetado, influenciado por determinantes sociais, econômicos, territoriais e estruturais que extrapolam as variáveis operacionais isoladas.

Adicionalmente, o potencial desses modelos pode ser significativamente ampliado com a incorporação de variáveis contextuais, como indicadores socioeconômicos (nível de renda e taxa de desemprego), dados demográficos, crescimento populacional, Índice de Desenvolvimento Humano (IDH), estrutura urbana, indicadores de eficácia coletiva e dados ambientais, incluindo clima, temperatura e padrões sazonais. A integração dessas dimensões não apenas enriquece a base informacional dos algoritmos, como também favorece a interpretação dos resultados, aspecto essencial para sua adoção prática, ética e transparente no âmbito das políticas públicas. Dessa forma, a combinação entre robustez preditiva e contextualização analítica oferece uma base mais sólida para uma gestão da segurança pública mais inteligente, eficiente e sensível às realidades locais.

No plano institucional, a adoção de modelos preditivos como os aqui propostos contribui diretamente para a modernização da Polícia Militar, fortalecendo sua missão constitucional de promover a segurança preventiva. O uso de inteligência computacional na análise e antecipação de cenários críticos não apenas amplia a capacidade de resposta do Estado, como também agrega elementos de transparência, precisão e efetividade, alinhando-se às exigências de uma sociedade cada vez mais dinâmica, orientada por dados e por princípios de gestão baseada em evidências.

Como direções para pesquisas futuras, recomenda-se não apenas o enriquecimento da base de dados com variáveis contextuais e estruturais, mas também a exploração de modelos mais sofisticados, como redes neurais profundas (*deep learning*) e arquiteturas híbridas. Embora exijam maior capacidade computacional e volume de dados, esses modelos possuem elevado potencial para capturar relações não lineares mais complexas, promovendo ganhos adicionais de acurácia e aplicabilidade em ambientes urbanos multifatoriais, dinâmicos e sujeitos a constantes transformações.

REFERÊNCIAS

ANTIPOV, Evgeniy A.; POKRYSHEVSKAYA, Elena B. **Mass appraisal of residential apartments: an application of Random forest for valuation and a CART-based approach for model diagnostics.** *Expert Systems with Applications*, v. 39, n. 2, p. 1772–1778, 2012.

AWAL, Md Akhtaruzzaman; RABBI, Jannatul; HOSSAIN, Sheikh Iqbal; HASHEM, Md Mahfuzul Alam. **Using linear regression to forecast future trends in crime of Bangladesh.** In: **INTERNATIONAL CONFERENCE ON INFORMATICS, ELECTRONICS AND VISION – ICIEV, 5., 2016, Dhaka. Proceedings.** Piscataway: IEEE, 2016. p. 333–338.



AZIZ, Rabia Musheer; HUSSAIN, Aftab; SHARMA, Prajwal; KUMAR, Pavan. **Crime rate prediction using machine learning algorithms: a case study from India.** *Journal of Crime and Justice Studies*, v. 5, n. 2, p. 45–62, 2022.

AZIZ, Rabia Musheer; HUSSAIN, Aftab; SHARMA, Prajwal; KUMAR, Pavan. **Machine learning-based soft computing regression analysis approach for crime data prediction.** *Karbala International Journal of Modern Science*, v. 8, n. 1, p. 1–11, 2022.

BATISTA, Alexandre Barbosa. **Segurança pública baseada em evidências: análise, avaliação e políticas públicas.** Brasília: ENAP, 2021.

BISHOP, Christopher M. **Pattern recognition and machine learning.** New York: Springer, 2006.

BIN, Okmyung; GAO, Xiaolong; ZHAO, Zhigang. **Machine learning in housing price prediction: a review and empirical study.** *Journal of Real Estate Research*, v. 43, n. 1, p. 65–90, 2021.

BREIMAN, Leo. **Statistical modeling: the two cultures.** *Statistical Science*, v. 16, n. 3, p. 199–231, 2001.

DEB, Chandra; ZHANG, Feng; YANG, Jie; LEE, Seung E.; SHAH, Kamal W. **A review on time series forecasting techniques for building energy consumption.** *Renewable and Sustainable Energy Reviews*, v. 74, p. 902–924, 2017.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The elements of statistical learning: data mining, inference, and prediction.** 2. ed. New York: Springer, 2009.

JINDAL, Mohit; SARAN, Ritesh; YADAV, Mehak. **Machine learning-based crime prediction and analysis using ensemble methods.** *Materials Today: Proceedings*, v. 45, p. 6194–6200, 2021.

JINDAL, Rohit; SARAN, Ritesh; YADAV, Mehak. **Crime prediction using ensemble machine learning techniques.** *Journal of Big Data*, v. 8, n. 1, p. 1–25, 2021.

KHUNTIA, Swasti; DE, Soumya Kanti; SINGH, S. N.; HEMANTH, S. **Load forecasting models for power systems: an extensive review.** *Journal of Modern Power Systems and Clean Energy*, v. 9, p. 617–632, 2021.

KOU, Gang; YANG, Peiyong; CHEN, Yi. **Machine learning in financial risk management: a literature review.** *Financial Innovation*, v. 7, n. 1, p. 1–33, 2021.

LI, Xue; HUANG, Bo; EMRICH, Christopher T. **Crime forecasting using spatio-temporal pattern detection and machine learning.** *Computers, Environment and Urban Systems*, v. 81, p. 101474, 2020.

LI, Yajie; WANG, Haoxuan; HUANG, Haojun. **Crime prediction with machine learning and comparison with time series models: a case study in Chicago.** *ISPRS International Journal of Geo-Information*, v. 9, n. 10, p. 1–17, 2020.



LOPES, Leonardo de Souza; RUSSO, Eduardo Viana Vargas. **Policimento previne crime? A eficiência das atividades policiais voltadas ao controle e prevenção da criminalidade. Revista Brasileira de Segurança Pública**, v. 16, n. 1, p. 130–156, 2022.

MEDEIROS, Marcelo C.; VASCONCELOS, Gabriel F.; VEIGA, Álvaro; ZILBERMAN, Ezra. **Forecasting inflation in a data-rich environment: the benefits of machine learning methods. Journal of Business & Economic Statistics**, v. 39, n. 1, p. 98–119, 2021.

MITCHELL, Tom M. **Machine learning**. New York: McGraw-Hill, 1997.

SHINGLETON, Jonathan. **Using multiple regression analysis to predict crime rates in Salinas, California**. Bachelor's thesis (Statistics) – California State University, Monterey Bay, 2012.