Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Douglas Branco Pessanha Lopes

Interpretação de Modelo de Previsão do Número de Homicídios no Brasil

Curitiba
2024

Douglas Branco Pessanha Lopes

Interpretação de Modelo de Previsão do Número de Homicídios no Brasil

Monografia apresentada ao Programa de Especialização em *Data Science* e *Big Data* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Luis Eduardo Soares de Oliveira

Curitiba
2024

Interpretação de Modelo de Previsão do Número de Homicídios no Brasil

Prediction Model Interpretation for the Number of Homicides in Brazil

Douglas Branco Pessanha Lopes[1], Luis Eduardo Soares de Oliveira[2]

[1]Aluno do programa de Especialização em Data Science & Big Data, douglaspessanha@protonmail.com
[2]Professor do Departamento de Informática - DINF/UFPR, lesoliveira@inf.ufpr

Este estudo tem como objetivo analisar e prever o número de homicídios no Brasil utilizando modelos de aprendizado de máquina, além de interpretar os resultados através do framework SHAP(SHapley Additive exPlanations). Utilizamos dados sobre homicídios e indicadores socioeconômicos extraídos do Atlas da Violência, IPEA, IBGE e Banco Mundial, abrangendo o período de 1992 a 2021.

Inicialmente, realizamos uma análise descritiva dos dados, identificando padrões e tendências ao longo dos anos. Observamos um aumento significativo nos homicídios, com variações nas taxas de pobreza, desemprego, IDH médio, entre outros indicadores. Em seguida, aplicamos modelos de aprendizado de máquina, incluindo Random Forest, Support Vector Regression (SVR) e Gradient Boosting, para prever o número de homicídios. O Gradient Boosting apresentou o melhor desempenho, com menores valores de RMSE e MAE.

Para interpretar os modelos, utilizamos o framework SHAP, que identificou a taxa de pobreza, taxa de suicídios, IDH médio e taxa de óbitos no transporte como os atributos mais importantes. Surpreendentemente, a análise revelou uma correlação negativa entre a taxa de pobreza e o número de homicídios, sugerindo a necessidade de uma análise mais aprofundada e multidisciplinar.

Palavras-chave: modelos preditivos, violência, SHAP, interpretação

This study aims to analyze and predict the number of homicides in Brazil using machine learning models, and to interpret the results through the SHAP (SHapley Additive exPlanations)framework. We utilized data on homicides and socioeconomic indicators extracted from the Atlas of Violence, IPEA, IBGE, and the World Bank, covering the period from 1992 to 2021.

Initially, we performed a descriptive analysis of the data, identifying patterns and trends over the years. We observed a significant increase in homicides, with variations in poverty rates, unemployment, average HDI, among other indicators. Subsequently, we applied machine learning models, including Random Forest, Support Vector Regression (SVR), and Gradient Boosting, to predict the number of homicides. Gradient Boosting showed the best performance, with lower RMSE and MAE values.

To interpret the models, we used the SHAP framework, which identified poverty rate, suicide rate, average HDI, and transportation death rate as the most important attributes. Surprisingly, the analysis revealed a negative correlation between the poverty rate and the number of homicides, suggesting the need for a deeper and multidisciplinary analysis.

Keywords: Predictive models, violence, SHAP, interpretation

1. Introdução

O Brasil é um dos países com o maior número de homicídios em termos absolutos, além de apresentar uma das maiores taxas por 100 mil habitantes. O país ocupa a primeira posição no ranking absoluto e a décima quarta na taxa de homicídios por cem mil habitantes [20], com dados de 2021. Esses dados destacam a importância da análise e estudo do tema, pois trata-se de um problema social gravíssimo que afeta toda a coletividade. No caso brasileiro, os números são significativamente maiores do que em outros países ou do que poderia ser considerado normal, se é que se pode falar em normalidade nesse contexto.

A violência pode ser estudada sob diversas perspectivas, como a criminologia, geralmente associada ao direito, a sociologia, a psicologia, bem como a estatística, entre outras. Neste trabalho, utilizam-se modelos

de aprendizado de máquina para lançar mais luz sobre o assunto. Dessa forma, foram criados modelos para predizer o número de homicídios no Brasil com base em dados sobre violência e indicadores socioeconômicos.

Foram utilizados dados de 1992 até 2021, extraídos do Atlas da Violência do IPEA [8]. A variável alvo é o número de homicídios em todo o território nacional, e as variáveis preditoras incluem indicadores socioeconômicos.

Outro problema abordado em nosso trabalho é a interpretação dos algoritmos de aprendizado de máquina. Algumas ferramentas preditivas da estatística clássica, como a regressão linear, permitem fácil interpretação dos resultados pelo coeficiente dos parâmetros. Já no aprendizado de máquina, nem sempre isso é possível, pois diversos modelos não possuem parâmetros em que se permite evidenciar como cada variável contribuiu para o resultado do modelo.

Alguns modelos são acusados de reproduzir vieses de forma oculta, embora apresentem-se como neutros ou objetivos. A literatura relata casos de preconceitos raciais e de gênero nos resultados de alguns modelos. A não interpretabilidade dificulta verificar como o algoritmo chegou nos resultados enviesados. Portanto, o uso de frameworks como o SHAP (Shapley Additive Explanations) é fundamental para tornar esses modelos mais compreensíveis e identificar possíveis vieses.

Este estudo tem como objetivos prever o número de homicídios e verificar os resultados do SHAP na interpretação dos modelos, comparar os modelos de aprendizado de máquina e analisar a relação entre o desempenho do modelo e sua interpretabilidade.

A estrutura deste artigo é a seguinte: a Seção 2 apresenta a revisão da literatura, a Seção 3 detalha a metodologia utilizada, a Seção 4 apresenta os resultados obtidos, a Seção 5 discute os achados e suas implicações e sugestões para trabalhos futuros.

2. Revisão de Literatura

A violência e a criminalidade têm sido temas centrais nas Ciências Sociais, especialmente no contexto latino-americano, onde desigualdades sociais e jurídicas são significativas. Nóbrega Júnior [14] revisa a literatura sobre criminalidade violenta, destacando teorias contemporâneas e testes empíricos que identificam variáveis explicativas para a violência e delinquência. As principais teorias incluem a Teoria da Desorganização Social, a Teoria do Aprendizado Social, a Teoria do Controle Social, a Teoria do Autocontrole, a Teoria da Anomia, a Teoria Econômica da Escolha Racional e a Teoria das Janelas Quebradas. Essas teorias oferecem uma base importante para entender os determinantes sociais da criminalidade.

O artigo "Homicídios no Brasil na última década: uma revisão integrativa"[15] enfatiza a importância de compreender a violência homicida sob a perspectiva da saúde pública. São analisados estudos que abordam os determinantes sociais da violência, as implicações para a saúde pública e as estratégias de prevenção. A revisão destaca a necessidade de políticas integradas e a importância de dados robustos para a formulação de intervenções eficazes. Homicídios são apresentados como um problema de saúde pública global, com impactos significativos na mortalidade e morbidade. A pobreza, desigualdade e falta de oportunidades são identificados como fatores críticos na violência homicida.

2.1. Vieses em Modelos de Machine Learning

Os vieses em algoritmos de aprendizado de máquina são uma preocupação significativa, especialmente em áreas sensíveis como justiça criminal e decisões financeiras. Verma e Rubin [21] discutem métodos para detectar e mitigar esses vieses, destacando a necessidade de tornar os modelos mais justos e transparentes. Eles sugerem que muitos modelos, embora aparentemente neutros, podem reproduzir preconceitos ocultos presentes nos dados de treinamento. Zemel et al. [22] propõem um algoritmo de aprendizado para classificações justas que visa alcançar tanto a justiça de grupo quanto a justiça individual, ao mesmo tempo em que ofusca informações sobre a pertença a grupos protegidos. Esse método busca garantir que indivíduos semelhantes sejam tratados de forma semelhante, evitando discriminação oculta.

Kamishima et al. [9] abordam a questão da equidade em classificadores de aprendizado de máquina, propondo o "Prejudice Remover Regularizer". Este método aplica regularização a modelos de previsão, como a regressão logística, para reduzir o preconceito indireto sem comprometer significativamente a precisão do modelo. Os autores demonstram que a simples remoção de variáveis sensíveis dos modelos não é suficiente para evitar discriminação, pois informações sensíveis podem influenciar indiretamente as previsões. Experimentos mostraram que o método deles melhora a equidade das previsões, mantendo um bom

equilíbrio entre precisão e justiça no modelo de aprendizado de máquina.

Gagliardi Cozman e Kaufman [4] discutem a diversidade de origens de vieses nos sistemas de inteligência artificial e os caminhos para sua mitigação. Eles enfatizam que os vieses podem surgir em várias etapas do processo de desenvolvimento do modelo, desde a coleta de dados até a definição de variáveis e hiperparâmetros. A falta de representatividade demográfica nas bases de dados de treinamento é uma causa comum de vieses, que pode levar a discriminação de grupos minoritários. O artigo sugere diversas abordagens para mitigar esses vieses, incluindo a criação de bases de dados mais diversificadas e o uso de técnicas de interpretabilidade como o SHAP para compreender melhor o impacto das variáveis nos modelos.

2.2. Frameworks de Interpretação de Modelos

Lundberg e Lee [12] apresentam o framework SHAP (SHapley Additive exPlanations), que unifica diversas metodologias de atribuição de importância de variáveis e oferece uma solução única baseada na teoria dos jogos para a interpretabilidade de modelos preditivos. O SHAP atribui um valor de importância a cada *feature* de uma predição específica, permitindo uma melhor compreensão do comportamento do modelo. Doshi-Velez e Kim [6] exploram diferentes métodos para explicar e interpretar modelos de aprendizado de máquina. Eles destacam a importância de modelos interpretáveis, especialmente em contextos de alto risco como saúde e justiça criminal. O artigo enfatiza a necessidade de um equilíbrio entre a complexidade do modelo e sua interpretabilidade, argumentando que a transparência é crucial para a confiança do usuário e para a aceitação dos modelos de IA na prática.

2.3. Predição de Homicídios

O artigo "Crime Analysis and Prediction Using Data Mining"[18] explora o uso de técnicas de mineração de dados para a análise e predição de crimes, com foco em homicídios. A pesquisa utiliza algoritmos de aprendizado de máquina para identificar padrões e prever ocorrências de homicídios, destacando a importância da análise de dados para a segurança pública. São discutidos diferentes modelos preditivos e sua eficácia na previsão de crimes. Técnicas de mineração de dados podem identificar padrões relevantes na ocorrência de homicídios, e modelos de aprendizado de máquina, como redes neurais e árvores de decisão, são eficazes na predição de crimes. A análise preditiva pode ajudar na formulação de políticas de segurança pública mais eficazes e direcionadas.

2.4. Principais Contribuições da Literatura

1. Framework SHAP: Lundberg e Lee [12] oferecem um método robusto e teoricamente fundamentado para a interpretação de modelos complexos, utilizando valores de Shapley para atribuir importância às variáveis.

2. Mitigação de Vieses: Verma e Rubin [21], Zemel et al. [22], Kamishima et al. [9] e Gagliardi Cozman e Kaufman [4] discutem técnicas para garantir que os modelos de aprendizado de máquina não perpetuem ou amplifiquem vieses existentes nos dados de treinamento.

3. Predição de Crimes: Estudos aplicam modelos de aprendizado de máquina na predição de homicídios, demonstrando a utilidade prática dessas técnicas para a segurança pública.

3. Metodologia

Os dados sobre violência utilizados neste estudo foram extraídos do Atlas da Violência, publicado pelo Instituto de Pesquisa Econômica Aplicada (IPEA) [8]. Os indicadores socioeconômicos foram extraídos do IBGE [7] e do Banco Mundial [5]. O conjunto de dados abrange informações detalhadas sobre violência no Brasil no período entre 1992 e 2021, além de trazer importantes indicadores socioeconômicos. A tabela abaixo 1 contém o nome dos atributos utilizados e os tipos de dados.

Nome do Atributo	Tipo de Dados
homicídios	Numérico (int)
taxa de suicídios anos	Numérico (float)
taxa óbitos transporte ano	Numérico (float)
taxa mortes violentas ano	Numérico (float)
índice Theil	Numérico (float)
índice de Gini	Numérico (float)
taxa de pobreza	Numérico (float)
IDH médio	Numérico (float)
taxa de desemprego	Numérico (float)
crescimento PIB	Numérico (float)

Tabela 1: Nomes dos atributos utilizados e seus tipos de dados

O processamento e limpeza dos dados foram feitos com a biblioteca Pandas usando a linguagem Python.

Foi necessário transformar alguns dados para *float*, porque alguns atributos foram carregados como *string*, juntamente com a substituição da vírgula por ponto devido ao padrão de *float* usado pelas bibliotecas de aprendizado de máquina. Não haviam dados ausentes no período analisado.

Após a limpeza, foi realizada uma análise exploratória dos dados para entender a distribuição e identificar padrões relevantes. Gráficos e estatísticas descritivas foram utilizados para visualizar e resumir os dados.

Antes da modelagem, foi realizada a normalização dos dados com a classe *MinMaxScaler* do *Scikit-learn*. Essa classe transforma os dados para um intervalo entre 0 e 1. Somente a variável alvo não foi transformada.

Na fase de modelagem, três modelos de aprendizado de máquina foram utilizados: *Random Forest*, *Support Vector Regression* (SVR) e *Gradient Boosting*. Esses modelos foram escolhidos por sua capacidade de lidar com dados complexos e detectar padrões não lineares. De forma resumida, o SVR é uma extensão do algoritmo *Support Vector Machine* (SVM) para problemas de regressão, tentando encontrar uma função que desvie o mínimo possível de um conjunto de pontos de treinamento, dentro de um limite de margem especificado. O *Gradient Boosting* é um algoritmo de *ensemble* que cria uma série de modelos preditivos simples (geralmente árvores de decisão) de forma sequencial, cada modelo corrigindo os erros do modelo anterior, melhorando a precisão progressivamente. Já o *Random Forest* é um algoritmo de ensemble que utiliza múltiplas árvores de decisão para realizar previsões, cada árvore sendo treinada em um subconjunto diferente dos dados, e suas previsões são combinadas para obter o resultado final.

A validação cruzada foi realizada utilizando *K-fold cross-validation* para garantir a robustez dos modelos. A otimização de hiperparâmetros foi realizada utilizando *Grid Search*, um método sistemático para encontrar a melhor combinação de hiperparâmetros que melhora o desempenho dos modelos. As métricas utilizadas para avaliar os modelos foram RMSE (Raiz do Erro Quadrático Médio) e MAE (Erro Médio Absoluto).

O framework SHAP (SHapley Additive exPlanations) foi utilizado para interpretar os modelos de aprendizado de máquina. O SHAP atribui um valor de importância a cada atributo de uma predição específica, permitindo entender como cada variável contribui para as previsões do modelo. Após treinar os modelos, os valores SHAP foram calculados para cada predição, ajudando a identificar quais atributos tinham maior impacto nas previsões. Foram geradas diversas visualizações, como gráficos de dependência e gráficos de barras de importância de atributos, para facilitar a interpretação dos resultados.

Por fim, foi feita uma seleção de atributos com base nos mais importantes apontados pelo SHAP. O modelo com os melhores resultados foi treinado novamente com esses atributos para verificar se havia melhoras no desempenho do modelo.

4. Resultados

No período estudado, houve um grande aumento no número de homicídios e também na taxa de homicídios, como demonstram as figuras 1 e 2. Pode-se observar, entretanto, que entre 2016 e 2018, os números caíram.

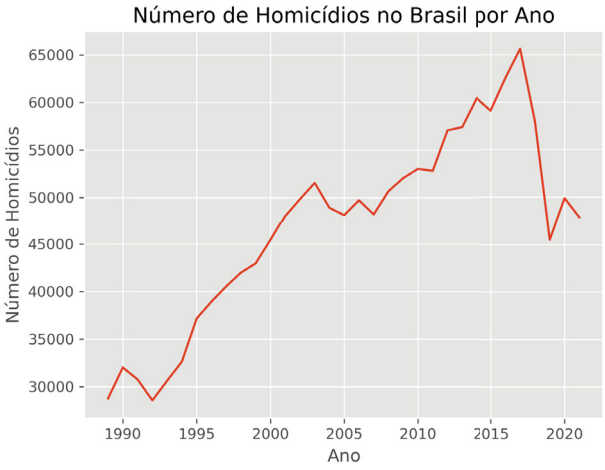

Figura 1: Homicídios por ano.

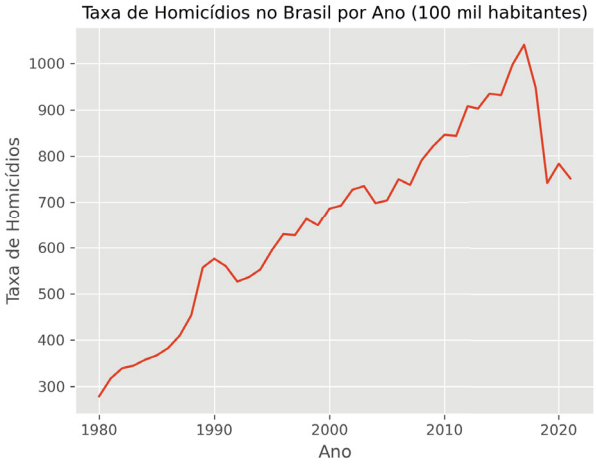

Figura 2: Taxa de Homicídios por ano.

Outras variáveis usadas no estudo apresentam um comportamento mais irregular ao longo dos anos. As figuras 3, 4, 5 e 6 mostram, respectivamente, as mortes violentas por ano, mortes de transporte por ano, taxa de crescimento do PIB e taxa de desemprego por ano.

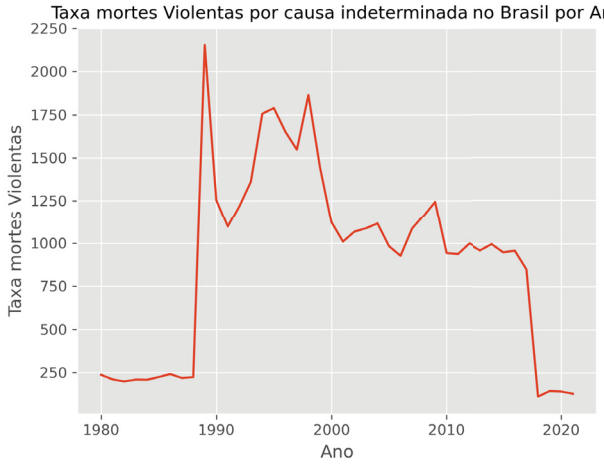

Figura 3: Taxa de Mortes Violentas por ano

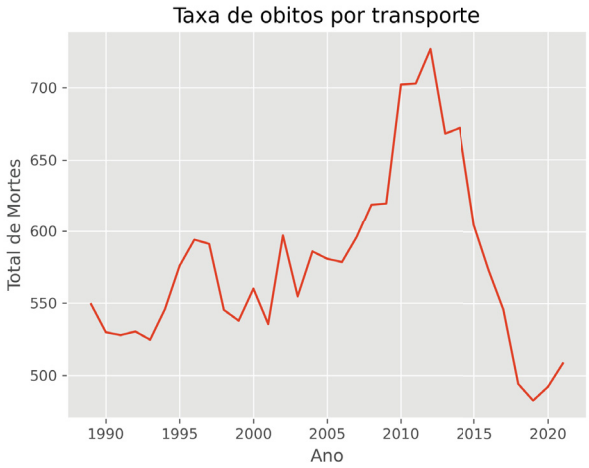

Figura 4: Taxa de Mortes de Transporte

O número de suicídios e a taxa de suicídios têm um comportamento quase linear, conforme as figuras 7 e 8. O IDH médio também apresenta um comportamento linear(figura 9), embora tenha uma leve queda após o ano de 2019. A taxa de pobreza e os indicadores de desigualdade (Gini e Theil), apesar de algumas oscilações, apresentam uma tendência de queda ao longo dos anos, como demonstra as figuras 10 e 11

Uma matriz de correlação(12) foi plotada para verificar a correlação entre as variáveis. Existe uma correlação negativa significativa entre a taxa de desemprego e o número de homicídios por ano, além de uma correlação negativa entre o índice de Gini e a taxa de suicídios. Também há uma correlação positiva entre o IDH médio e a taxa de suicídios por ano.

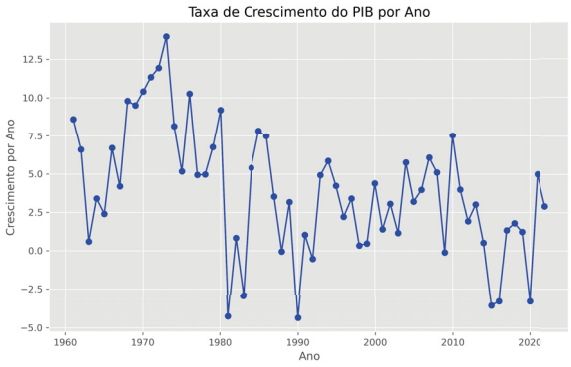

Figura 5: Taxa de Crescimento do PIB

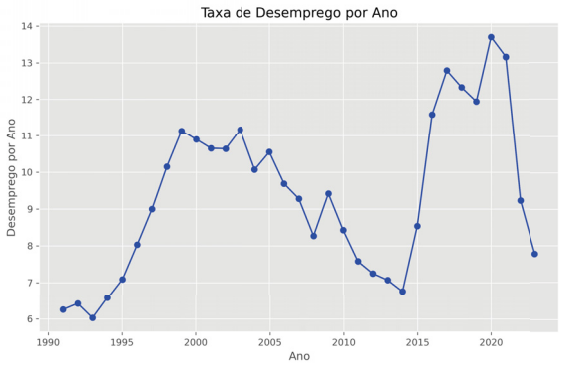

Figura 6: Taxa de Desemprego

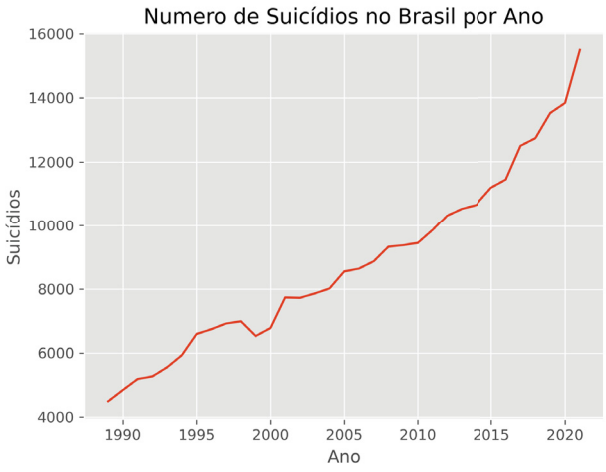

Figura 7: Número de Suicídios por ano