

# Tendências e Soluções para a Segurança Viária no Brasil: Uma Análise Abrangente de Acidentes de Trânsito entre 2013 e 2024

Arthur César Ferreira Costa, Ricardo Gomes de Farias Junior

<sup>1</sup>Faculdade Nova Roma  
Av. Adjar da Silva Casé, 800 - Indianópolis, Caruaru - PE, 55024-740

<sup>2</sup>Análise e Desenvolvimento de Sistemas

ricardogomesfarias@hotmail.com, arthurcesar252@gmail.com

**Abstract.** *The article examines the serious issue of traffic accidents in Brazil, highlighting their impact on public health and the economy. The main causes of accidents are related to the non-compliance with traffic laws, such as driving under the influence of alcohol, speeding, and lack of attention. Deficiencies in infrastructure, such as poor roads, also play a significant role. The economic impact of these high mortality rates exceeds 50 billion reais annually. It proposes to analyze traffic accidents between 2013 and 2024, identifying contributing factors and presenting solutions to reduce the number of deaths and injuries. It uses techniques such as time series analysis, predictive models (ARMA, SARIMA, ARIMA), and evaluation metrics (MSE, MAE). Grid Search is employed for hyperparameter optimization, and the data is obtained from the Open Data of the Federal Highway Police. The results of the study include tables and graphs describing the ARMA, SARIMA, and ARIMA models, as well as the values of mean absolute error and mean squared error. The work offers an in-depth understanding of traffic accident patterns and proposes targeted preventive interventions to reduce their incidence.*

**Resumo.** *O artigo examina a grave problemática dos acidentes de trânsito no Brasil, destacando seu impacto na saúde pública e na economia. As principais causas dos acidentes estão relacionadas ao descumprimento das leis de trânsito, como dirigir sob efeito de álcool, excesso de velocidade e falta de atenção. Deficiências na infraestrutura, como estradas precárias, também desempenham um papel determinante. O impacto econômico dessas altas taxas de mortalidade ultrapassa 50 bilhões de reais anualmente. Propõe analisar os acidentes de trânsito entre 2013 e 2024, identificando fatores contribuintes e apresentando soluções para reduzir o número de mortes e feridos. Utiliza técnicas como análise de séries temporais, modelos preditivos (ARMA, SARIMA, ARIMA) e métricas de avaliação (MSE, MAE). Grid Search é empregado para otimização de hiperparâmetros, e os dados são obtidos dos Dados Abertos da Polícia Rodoviária Federal. Os resultados do estudo incluem tabelas e gráficos que descrevem os modelos ARMA, SARIMA e ARIMA, bem como os valores de erro médio absoluto e erro médio quadrático. O trabalho oferece uma compreensão aprofundada dos padrões de acidentes de trânsito e propõe intervenções preventivas direcionadas para reduzir sua incidência.*

# 1. Introdução

A segurança no trânsito é um tema de extrema importância para o bem-estar social e econômico, especialmente no contexto brasileiro, onde os acidentes viários representam um desafio significativo.

As estatísticas alarmantes revelam que o Brasil ocupa o terceiro lugar no ranking mundial de mortes no trânsito, com cerca de 392 mil óbitos registrados entre 2010 e 2019. Regiões como o Nordeste e o Norte experimentaram um aumento significativo nas taxas de mortalidade, enquanto outras, como o Sul e o Sudeste, apresentaram quedas [PACHECO 2024].

As principais causas dos acidentes estão relacionadas ao desrespeito às leis de trânsito, tais como o consumo de álcool antes de dirigir, o excesso de velocidade e a falta de atenção ao volante. A eficácia de leis como a Lei Seca, destinada a reduzir a mortalidade no trânsito, depende tanto da fiscalização rigorosa quanto da conscientização da população [Carvalho et al. 2023].

Além disso, as deficiências na infraestrutura viária desempenham um papel crucial na ocorrência de acidentes, com problemas como falhas no projeto das vias, sinalização inadequada e más condições do pavimento contribuindo para a insegurança nas estradas [CNT 2024].

O impacto econômico dos acidentes de trânsito é significativo, ultrapassando 50 bilhões de reais anualmente, incluindo despesas previdenciárias, perda de renda familiar, custos hospitalares e prejuízos patrimoniais [Carvalho et al. 2023].

Em um estudo recente conduzido pela Comissão Nacional de Desenvolvimento (CND), foi revelado um cenário alarmante em relação à qualidade das estradas no Brasil. Este estudo destacou que, sete das dez piores estradas do país estão localizadas na região Nordeste. Essa constatação levanta sérias preocupações sobre a infraestrutura viária na região e levanta questões urgentes sobre a segurança e eficiência das viagens rodoviárias na área. [CNT 2024]. Verificar Tabela1.

Table 1. As 10 piores estradas do Brasil					
Posição	Rodovia	Estado	Região	Gestão	Classificação
510º	AM-010	AM	Norte	Pública	Péssimo
509º	BA-142	BA	Nordeste	Pública	Péssimo
508º	PE-096	PE	Nordeste	Pública	Péssimo
507º	PE-545	PE	Nordeste	Pública	Péssimo
506º	MA-034	MA	Nordeste	Pública	Péssimo
505º	PI-141	PI	Nordeste	Pública	Péssimo
504º	AC-010	AC	Norte	Pública	Péssimo
503º	GO-142	GO	Centro-Oeste	Pública	Péssimo
502º	PB-400	PB	Nordeste	Pública	Péssimo
501º	PB-386	PB	Nordeste	Pública	Péssimo

Fonte: CNT

Diante dessa realidade preocupante, este projeto se propõe a analisar os acidentes de trânsito no Brasil entre 2013 e 2024, buscando identificar os principais fatores que

contribuem para esse cenário e apresentar soluções para reduzir o número de mortes e feridos. Para isso, serão identificados os meses com maior incidência de acidentes por meio de análise de série temporal, aplicando modelos de regressão. O objetivo é promover ações preventivas direcionadas aos meses corretos, de forma que as intervenções aplicadas resultem na redução do número de acidentes.

## 2. Referencial Teórico

### 2.1. Séries Temporais

O principal objetivo da análise de séries temporais consiste em elaborar modelos matemáticos que ofereçam descrições plausíveis para conjuntos de dados amostrais.

Com o intuito de proporcionar um contexto estatístico para descrever a natureza de dados que aparentemente variam de forma aleatória ao longo do tempo, partimos do pressuposto de que uma série temporal pode ser definida como uma coleção de variáveis aleatórias organizadas de acordo com a sequência temporal em que são observadas. [Pérez 2024]

Algumas das questões e desafios relacionados ao uso de técnicas que empregam séries temporais podem ser mais claramente elucidados ao considerar dados experimentais reais provenientes de diversas áreas temáticas. [Pérez 2024]

A utilização mais comum das séries temporais é a previsão de valores futuros. Nesses cenários, buscamos identificar padrões na série que nos capacitem a antecipar a direção na qual os dados estão se encaminhando.

### 2.2. Modelos Preditivos

Os modelos de Autorregressão e Médias Móveis, também conhecidos como modelos ARMA (ou modelos Box-Jenkins), são amplamente reconhecidos como abordagens clássicas para análise de séries temporais. Uma série temporal é caracterizada como um modelo ARMA ( $p, q$ ) caso atenda às seguintes condições:

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \sum_{j=1}^q \theta_j a_{t-j} + a_t \quad (1)$$

Onde  $\phi_1, \dots, \phi_p$  são os parâmetros da parte AR,  $\theta_1, \dots, \theta_q$  são os parâmetros da parte MA e  $\varepsilon$  é um ruído branco. O modelo ARMA é uma ferramenta utilizada para prever valores futuros da série, composto por duas partes: uma parte autoregressiva (AR) e uma parte de média móvel (MA) [de Miranda Esquivel 2012].

Uma série temporal  $Y_t$  segue um modelo de autoregressiva de médias móveis integrada, se a  $d$ -ésima diferença  $W_t = \Delta^d Y_t$  é um processo ARMA estacionário. Se  $W_t$  segue um modelo ARMA ( $p, q$ ), dizemos que  $Y_t$  é um processo ARIMA( $p, d, q$ ). Para fins práticos, geralmente podemos tomar  $d = 1$  ou no máximo 2. Suponha o seguinte modelo ARMA( $p, q$ ):

$$\Theta_p(L)Y_t = \Theta_q(L)a_t \quad (2)$$

Onde o polinômio AR pode ser fatorado em termos de suas  $p$  raízes  $L_1, \dots, L_2$

$$\Theta_p(L) = (1 - L_1^{-1}L)(1 - L_2^{-1}L) \cdots (1 - L_p^{-1}L) \quad (3)$$

O modelo SARIMA (p, d, q)  $\times$  (P,D, Q) é definido como;

$$\Phi_p(B_s)\phi_p(B)\nabla_s^D\nabla_dX_t = \Theta_Q(B_s)\Theta_q(B)\epsilon_t \quad (4)$$

Onde, generaliza todos os modelos da família ARIMA e permite modelar séries estacionárias, bem como séries sazonais e não estacionárias, bem com séries sazonais e não sazonais.

Em um modelo ARIMA, os termos estão incluídos autoregressivos (p), diferenciação da variável (d) e termos de média móvel (d). No entanto, o modelo SARIMA inclui termos sazonais autorregressivos (P), diferenciação sazonal (D) e média móvel sazonal (Q), ou seja, o SARIMA contém fatores sazonais e não sazonais em um modelo multiplicativo. [Ulloa 2022]

### 3. Métricas e Técnicas

#### 3.1. Grid Search

Grid Search é uma técnica de otimização de hiperparâmetros amplamente utilizada em aprendizado de máquina. Consiste em uma busca exaustiva sobre um conjunto especificado de valores de hiperparâmetros, com o objetivo de encontrar a combinação que maximiza o desempenho do modelo segundo uma métrica de avaliação pré-definida. A técnica é frequentemente implementada com validação cruzada, garantindo que a avaliação da performance do modelo seja robusta e menos suscetível a variações nos dados de treino.

#### 3.2. Erro Quadrado Médio (MSE)

O Erro Quadrado Médio (MSE) é uma métrica que calcula a média dos quadrados das diferenças entre os valores reais ( $y_i$ ) e os valores previstos ( $\hat{y}_i$ ). É utilizado para avaliar a precisão de modelos de regressão.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

#### 3.3. Erro Absoluto Médio (MAE)

O Erro Absoluto Médio (MAE) calcula a média das diferenças absolutas entre os valores reais ( $y_i$ ) e os valores previstos ( $\hat{y}_i$ ). É preferido em alguns casos por ser menos sensível a valores extremos em comparação ao MSE.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

## **4. Trabalho Relacionados**

### **4.1. Aprendizado de Máquina Utilizando Agrupamento e Regressão na Previsão de Locais de Acidentes de Trânsito em Zonas Urbanas**

O artigo aborda a crescente preocupação com a segurança viária em meio à rápida urbanização das cidades brasileiras. Com o aumento da mobilidade urbana, os acidentes de trânsito se tornam uma questão crucial, resultando de uma variedade de fatores, incluindo congestionamentos, erros humanos e deficiências na infraestrutura. Este trabalho propõe uma solução para prever locais de acidentes em áreas urbanas, utilizando técnicas de agrupamento (K-Means) e regressão (KNN regressor). A ideia é que ao agrupar os dados históricos de acidentes, seja possível realizar regressões em grupos específicos, reduzindo assim o erro e aumentando a precisão na identificação de locais propensos a acidentes.

### **4.2. Acidentes de trânsito no Brasil: dados e tendências**

O artigo "Acidentes de trânsito no Brasil: dados e tendências" publicado no SciELO analisa a mortalidade por acidentes de trânsito no Brasil entre 1978 e 1989, destacando um aumento significativo das taxas de mortalidade, frequentemente superiores a 20 por 100.000 habitantes. O estudo aponta que os acidentes de trânsito representam cerca de 30% das mortes por causas externas, sublinhando a gravidade desse problema de saúde pública. Além disso, observa-se uma variação nas tendências entre as capitais dos estados, com algumas apresentando estabilidade ou até mesmo declínio nas taxas de mortalidade.

O estudo fornece uma base histórica importante para entender as mudanças e persistências nos padrões de acidentes de trânsito ao longo do tempo.

### **4.3. Acidentes de trânsito no Brasil de 1998 a 2010: muitas mudanças e poucos resultados**

O artigo "Acidentes de trânsito no Brasil de 1998 a 2010: muitas mudanças e poucos resultados" examina a evolução dos acidentes de trânsito nesse período, destacando o aumento no número de mortes e na taxa de mortalidade. Apesar das novas leis de trânsito e melhorias na segurança dos veículos, os resultados foram limitados, principalmente devido ao crescimento no uso de motocicletas e ao consumo de álcool. O estudo critica a eficácia das medidas públicas para controlar e reduzir os acidentes.

## **5. Metodologia**

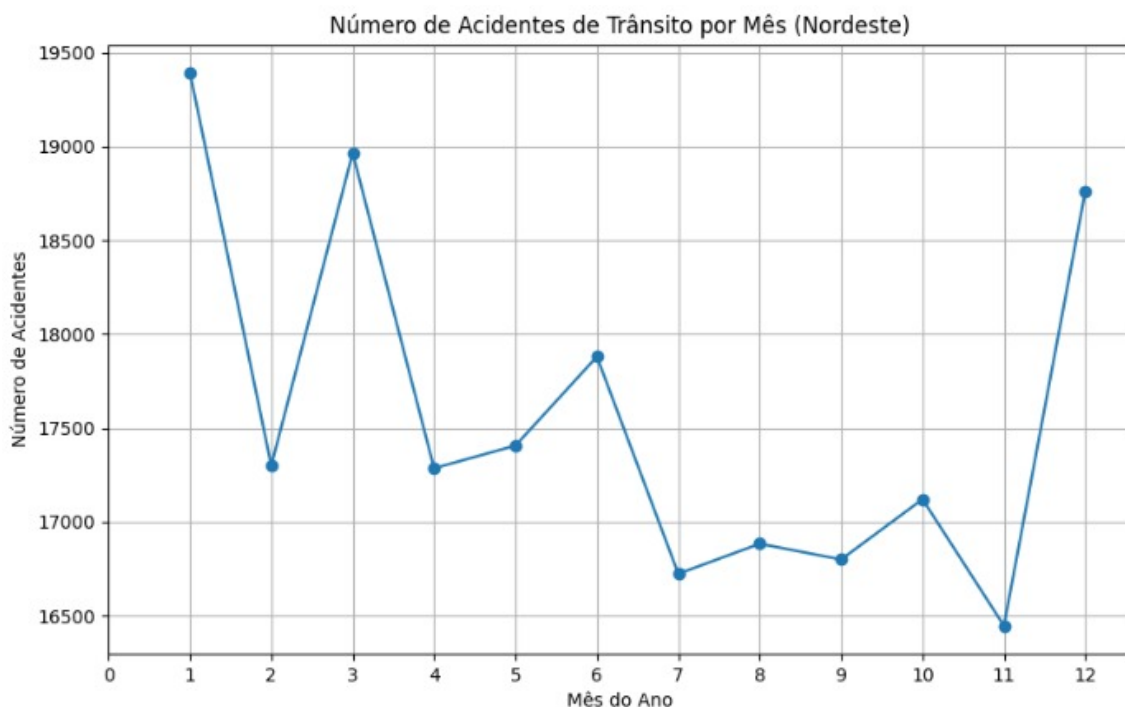
### **5.1. Entendimento dos dados**

Os dados utilizados neste estudo foram obtidos dos **Dados Abertos da Polícia Rodoviária Federal (PRF)**, disponíveis em <https://www.gov.br/prf/pt-br/acesso-a-informacao/dados-abertos/dados-abertos-da-prf>.

### **5.2. Pré-processamento**

Esses dados passaram por um processo de pré processamento a fim de analisar aquilo que seria interessante para o projeto.

Em nosso caso, a base de dados utilizada é possível devido à Lei de Acesso à Informação (LAI), na Instrução Normativa SLTI nº 4, de 13 de abril de 2012 (que institui



**Figure 1. Número de acidentes por mês**

Fonte: De autoria própria.

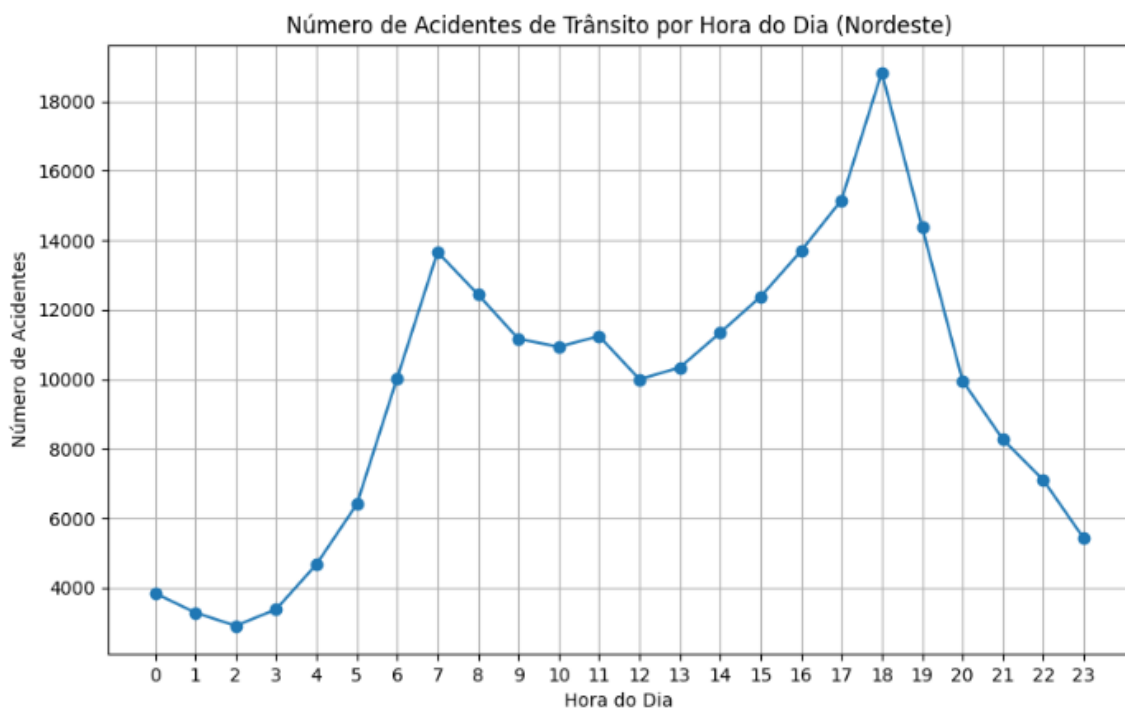
a Infraestrutura Nacional de Dados Abertos), no Decreto nº 8.777, de 11 de maio de 2016 (que institui a Política de Dados Abertos no Executivo Federal).

Os dados classificados como abertos podem ser utilizados de várias formas, seja pelo próprio governo ou pela sociedade, como, por exemplo, no desenvolvimento de aplicativos, que exibem informações de forma gráfica e interativa. [Federal 2024]

### 5.3. Passos de Pré-processamento

1. **Leitura dos Dados:** Os conjuntos de dados foram importados e lidos a partir de arquivos CSV de 2013 a 2024.
2. **Concatenação dos Dados:** Os conjuntos de dados foram concatenados em um único DataFrame.
3. **Análise de Dados Faltantes:** Foi realizada uma análise para identificar e quantificar os valores nulos em cada coluna.
4. **Remoção de Colunas Desnecessárias:** Colunas irrelevantes para a análise foram removidas do DataFrame, como 'delegacia', 'regional', entre outras.
5. **Filtragem por Região:** Os dados foram filtrados para incluir apenas informações relacionadas aos estados do Nordeste do Brasil.
6. **Divisão da Base de Dados:** Os dados foram divididos em conjuntos de treino e teste para permitir a avaliação do desempenho do modelo. Utilizou-se uma proporção de 70% para treino e 30% para teste, garantindo que o modelo fosse treinado com uma quantidade significativa de dados e testado de forma independente para avaliar sua generalização.

Afim de analisar padrões temporais foi feito uma contagem de acidentes por hora do dia, assim como mostra a Figura 2.



**Figure 2. Números de acidentes por hora do dia**

Fonte: De autoria própria.

Depois partimos para uma filtragem mais abrangente, fazendo outra contagem por mês, exibindo assim o número de acidentes por mês, com indica na Figura 1.

#### 5.4. Modelagem

A fase de modelagem no processo CRISP-DM (Cross-Industry Standard Process for Data Mining) é crucial para desenvolver e construir os modelos preditivos. Refinamos os modelos por meio de treino e teste, gerando os gráficos essenciais para a visualização do desempenho. Em seguida, foi empregado o Grid Search para identificar os hiper-parâmetros.

Melhor ordem ARMA encontrada: (3, 0, 3)

Melhor ordem ARIMA encontrada: (2, 1, 3)

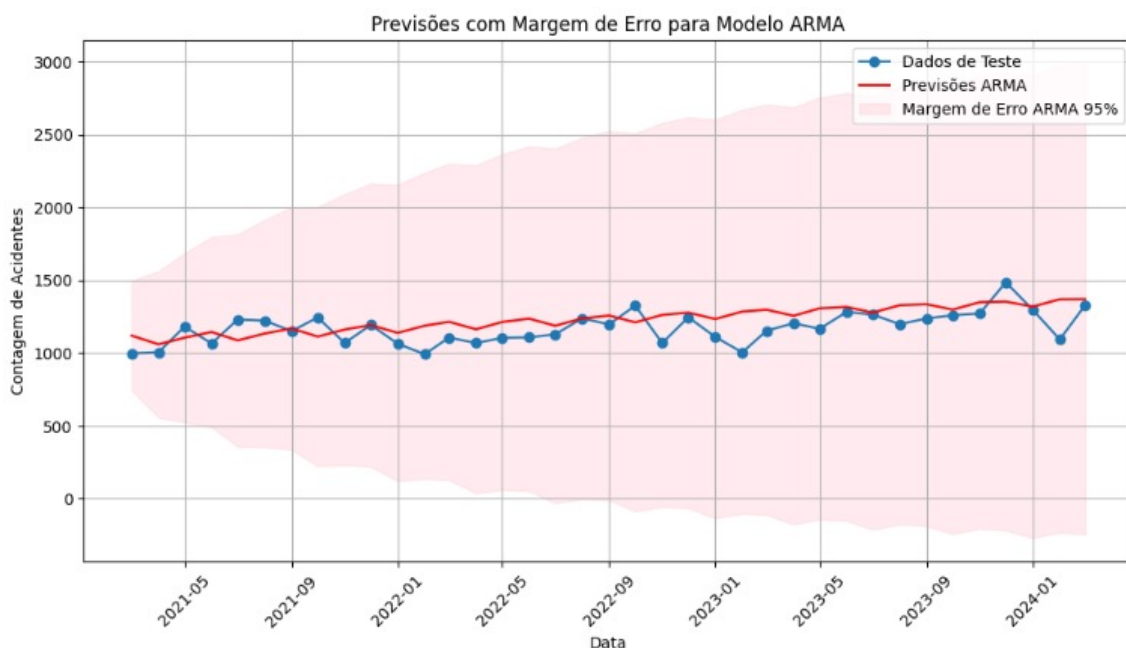
Melhor ordem SARIMA encontrada: (0, 1, 0) com ordem sazonal (0, 1, 1, 12)

#### 5.5. Validação

Após a obtenção dos dados gerados, é crucial avaliá-los para determinar sua eficácia na resolução do problema de negócio proposto. Essa avaliação é realizada para conduzir as previsões e calcular as métricas de erro necessárias para validar a precisão e a robustez do modelo.

### 6. Resultados

Com base nos gráficos gerados e nos valores calculados para o erro médio quadrático (MSE) e o erro médio absoluto (MAE) (Tabela 2), podemos observar que o modelo ARMA se destacou como o mais eficaz em nossa análise. Este modelo apresentou um



**Figure 3. Previsões do Modelo ARMA**

Fonte: De autoria própria.

MSE de 13611.7198 e um MAE de 96.7759. Vale ressaltar que o MSE visa quantificar a diferença média entre os valores previstos e os valores reais, enquanto o MAE fornece uma medida direta da discrepância média entre esses dois conjuntos de dados. Em outras palavras, quanto menor os valores obtidos para o MSE e o MAE, mais precisa é a previsão dos dados realizada pelo modelo.

Para este trabalho, usamos o ambiente de produção Google Colab, utilizando a linguagem Python. Os códigos utilizados estão disponíveis para acesso em: <https://github.com/Ricardo-Gomes-Junior/-DS---AVALIACAO-N2>

Modelo	Mean Squared Error (MSE)	Mean Absolute Error (MAE)
ARMA	13611.7198	96.7759
ARIMA	23480.0860	122.0141
SARIMA	71989.5774	224.0296

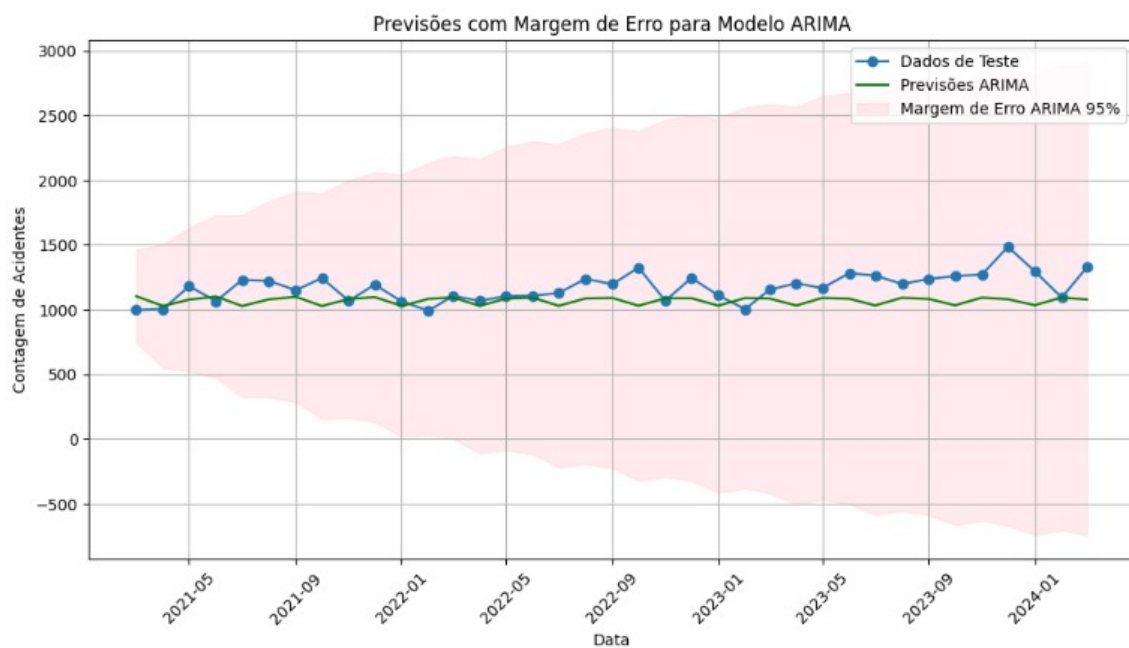
**Table 2. Média dos erros para os modelos ARMA, ARIMA e SARIMA**

Fonte: De autoria própria.

## 7. Conclusão

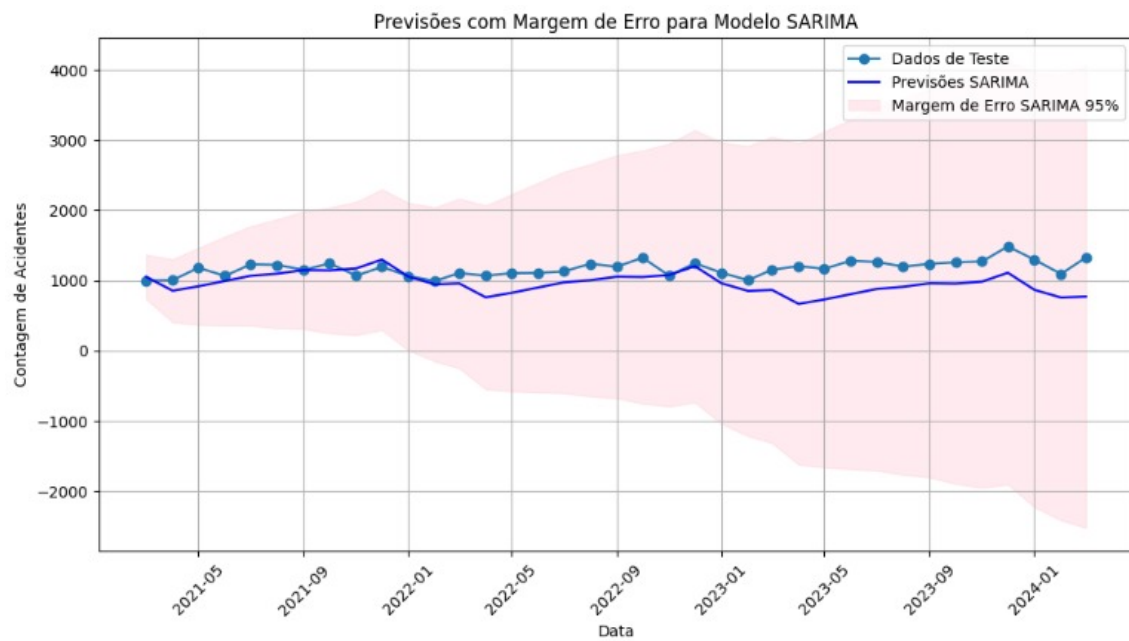
De acordo com o estudo realizado, revelou-se um panorama preocupante, especialmente nas regiões Nordeste e Norte do País, que apresentam as maiores taxas de mortalidade no trânsito. A análise dos dados evidenciou a relação entre os acidentes e diversos fatores, como o desrespeito às leis de trânsito, deficiências na infraestrutura e a influência de fatores sazonais.





**Figure 4. Previsões do Modelo ARIMA**

Fonte: De autoria própria.



**Figure 5. Previsões do Modelo SARIMA**

Fonte: De autoria própria.

Através da aplicação de técnicas de análise de séries temporais e modelos preditivos (ARMA, ARIMA e SARIMA), foi possível identificar os meses com maior incidência de acidentes, com destaque para janeiro, março e dezembro. O modelo ARMA se destacou como o mais eficaz na previsão do número de acidentes, apresentando os menores valores de erro médio quadrático (MSE) e erro médio absoluto (MAE).

Com base nos resultados obtidos, este trabalho propõe as seguintes medidas para reduzir o número de acidentes de trânsito no Brasil:

1. **Campanhas educativas:** Intensificar campanhas de conscientização sobre os riscos do desrespeito às leis de trânsito, com foco em comportamentos como dirigir sob a influência de álcool, excesso de velocidade e uso inadequado do celular ao volante.
2. **Fiscalização rigorosa:** Reforçar a fiscalização das leis de trânsito, punindo com severidade os infratores e promovendo ações educativas em conjunto com as blitzes.
3. **Melhorias na infraestrutura:** Investir na melhoria das condições das estradas, sinalização adequada e implementação de medidas de segurança, como radares e lombadas eletrônicas.
4. **Planejamento urbano:** Implementar medidas de planejamento urbano que priorizem a segurança dos pedestres e ciclistas, como a criação de ciclovias e calçadas adequadas.
5. **Investimento em tecnologia:** Investir em pesquisas e desenvolvimento de novas tecnologias para auxiliar na segurança viária, como sistemas de alerta de fadiga ao volante e veículos autônomos.

A implementação dessas medidas exige um esforço conjunto do governo, da sociedade civil e das empresas privadas. É fundamental que todos os setores se mobilizem para construir um trânsito mais seguro e salvar vidas.

Além das medidas acima, este estudo destaca a importância de pesquisas contínuas sobre os acidentes de trânsito, com foco em fatores específicos que contribuem para sua ocorrência. A análise aprofundada desses dados permitirá o desenvolvimento de soluções ainda mais eficazes para reduzir a mortalidade e os feridos no trânsito, construindo um futuro mais seguro para todos.

## 7.1. Trabalhos Futuros

Para futuras pesquisas sobre acidentes de trânsito no Brasil, sugere-se a realização de estudos mais aprofundados sobre a relação entre infraestrutura viária e ocorrências de acidentes, visando identificar pontos críticos que necessitam de intervenções específicas. Além disso, é importante investigar o impacto das políticas públicas de segurança viária na redução do número de acidentes e na promoção de um trânsito mais seguro.

Outra possível área de estudo seria a análise da influência de fatores como condições climáticas, horários do dia e dias da semana na ocorrência de acidentes, buscando identificar padrões comportamentais dos motoristas e desenvolver estratégias preventivas mais assertivas. Ademais, a utilização de técnicas de machine learning e inteligência artificial pode contribuir para a criação de modelos preditivos mais precisos e eficazes na prevenção de acidentes de trânsito.

Em suma, os trabalhos futuros podem se concentrar em aprofundar a compreensão dos fatores que contribuem para os acidentes de trânsito no Brasil e na proposição de medidas preventivas mais efetivas, a fim de reduzir a incidência de mortes e lesões decorrentes desses eventos.

## References

aaaa.

Carvalho, J., Homobono, K., Pacheco, H. L., Vasconcelos, C., Leal, S., Dias, N., and Silva, E. (2023). Comparação de modelos de regressão linear com aplicações em dados de mortalidade por acidentes de trânsito. *ENCICLOPEDIA BIOSFERA*, 20(46):66–77.

CNT (2024). Infraestrutura deficiente é fator preponderante na ocorrência de acidentes com vítimas. [〈https://www.cnt.org.br/agencia-cnt/infraestrutura-deficiente-causa-acidentes-com-vitimas〉](https://www.cnt.org.br/agencia-cnt/infraestrutura-deficiente-causa-acidentes-com-vitimas). Accessed: 2024-05-14.

de Miranda Esquivel, R. (2012). Análise espectral singular: Modelagem de séries temporais através de estudos comparativos usando diferentes estratégias de previsão. [〈http://repositoriosenaiba.fieb.org.br/bitstream/fieb/728/1/Disserta%C3%A7%C3%A3o%20de%20RenataEsquivelDissertacao15junh12.pdf〉](http://repositoriosenaiba.fieb.org.br/bitstream/fieb/728/1/Disserta%C3%A7%C3%A3o%20de%20RenataEsquivelDissertacao15junh12.pdf). Accessed: 2024-05-28.

Federal, P. R. (2024). Dados abertos da prf — polícia rodoviária federal. [〈https://www.gov.br/prf/pt-br/aceso-a-informacao/dados-abertos/dados-abertos-da-prf〉](https://www.gov.br/prf/pt-br/aceso-a-informacao/dados-abertos/dados-abertos-da-prf). Accessed: 2024-05-23.

PACHECO, E. L. (2024). A redução de acidentes por meio da educação para o trânsito: um enfoque na formação dos condutores. [〈https://repositorio.ifsc.edu.br/bitstream/handle/123456789/804/Emeli\\_Lemos\\_Pacheco.pdf?sequence=1&isAllowed=y〉](https://repositorio.ifsc.edu.br/bitstream/handle/123456789/804/Emeli_Lemos_Pacheco.pdf?sequence=1&isAllowed=y). Accessed: 2024-05-14.

Pérez, F. L. (2024). Características da série temporal. [〈http://leg.ufpr.br/~lucambio/STemporais/STemporaisI.html#I2〉](http://leg.ufpr.br/~lucambio/STemporais/STemporaisI.html#I2). Accessed: 2024-05-28.

Ulloa, G. C. (2022). Modelos arima, sarima e método de seleção de variáveis do tipo lasso para séries temporais. [〈https://csds2022.ufba.br/Arquivos/poster/CSDS2022\\_Poster\\_Gladys\\_Ulloa.pdf〉](https://csds2022.ufba.br/Arquivos/poster/CSDS2022_Poster_Gladys_Ulloa.pdf). Accessed: 2024-06-04.