

## CADERNOS DO IME – Série Estatística

Universidade do Estado do Rio de Janeiro - UERJ  
ISSN on-line 2317-4536 / ISSN impresso 1413-9022 - v.57, p.52-71, 2024  
DOI: 10.12957/cadest.2024.91708

# APLICAÇÃO DE MODELOS PARA A PREVISÃO DE OCORRÊNCIAS RODOVIÁRIAS: UM ENFOQUE NO CONTEXTO URBANO DO RIO DE JANEIRO

Emerson Gonçalves Araújo  
Universidade do Estado do Rio de Janeiro - UERJ  
[emerson.g.1322@gmail.com](mailto:emerson.g.1322@gmail.com)

José Vitor Damasceno Leite Guimarães  
Universidade do Estado do Rio de Janeiro - UERJ  
[josevdlguimaraes@gmail.com](mailto:josevdlguimaraes@gmail.com)

Marcello Montillo Provenza  
Universidade do Estado do Rio de Janeiro - UERJ  
[mprovenza@gmail.com](mailto:mprovenza@gmail.com)

Jorge Luiz de Jesus Goulart  
Universidade do Estado do Rio de Janeiro - UERJ  
[jorge.goulart@uerj.br](mailto:jorge.goulart@uerj.br)

### Resumo

*Este estudo tem como objetivo analisar a evolução temporal dos acidentes rodoviários no estado do Rio de Janeiro. O período analisado abrangeu desde janeiro de 2017 até dezembro de 2022, com ênfase no ano de 2022 como conjunto de teste, para realizar previsões com diferentes modelos preditivos. A base de dados foi obtida junto à Polícia Rodoviária Federal e à Associação Brasileira de Concessionárias de Rodovias. Cinco modelos foram aplicados e, com base nas métricas de previsão, constatou-se que o modelo Prophet, ao incorporar o índice ABCR como um regressor, demonstrou o melhor ajuste aos dados.*

**Palavras-chave:** Séries temporais, Modelos de previsão, Acidentes rodoviários, Testes estatísticos.

## 1. Introdução

Os acidentes de trânsito representam um desafio significativo em todo o Brasil, afetando não apenas a mobilidade, mas também a saúde pública. De acordo com a Associação Brasileira de Medicina de Tráfego, os acidentes rodoviários são o segundo maior problema de saúde pública no país. Entre 1995 e 2003 ocorrem cerca de um milhão de acidentes, resultando em mais de 30 mil mortes e 400 mil feridos (Bayerl, Santos, Braga, 2006). Esses números refletem a urgência de abordar essa questão de forma contínua e eficiente.

Neste cenário, o presente estudo se propõe a realizar uma análise abrangente da série temporal relacionada ao número de acidentes ocorridos no estado do Rio de Janeiro no intervalo temporal compreendido entre os anos de 2017 e 2022. A motivação central da pesquisa reside não apenas em compreender o comportamento intrínseco desses incidentes ao longo do período em questão, mas também em explorar e testar diferentes modelos preditivos. Essa abordagem visa identificar qual modelo demonstra um ajuste mais eficaz à série temporal, fornecendo, assim, insights valiosos para a prevenção e possível redução dos acidentes rodoviários no referido estado.

A atenção dedicada aos acidentes de trânsito no Rio de Janeiro remonta a períodos não recentes, como é documentado no artigo intitulado "Acidentes de Tráfego no Estado do Rio de Janeiro: Causas e Prevenção", publicado na Revista do Serviço Público em novembro de 1954 pelo autor Hilton Jesus Gradet. Nesse trabalho seminal, Gradet destacou de forma contundente a preocupante proeminência da cidade do Rio de Janeiro nos registros estatísticos de acidentes de trânsito daquela época. A análise detalhada dessas ocorrências já era uma prática estabelecida há décadas, evidenciando a persistência do problema e a necessidade contínua de abordagens preventivas e corretivas (Gradet, 1954).

Essa reflexão histórica lembra que as questões relacionadas à segurança viária são de longa data e requerem uma abordagem contínua para sua mitigação eficaz. O reconhecimento desses padrões históricos também destaca a importância de se construir sobre o trabalho pioneiro de pesquisadores como Gradet e de continuar a desenvolver estratégias inovadoras e baseadas em evidências para enfrentar os desafios atuais de segurança no trânsito.

A Polícia Rodoviária Federal (PRF) assume uma posição de destaque na promoção da segurança viária e na eficaz prevenção de acidentes. Desde sua criação em 1928, seguida pela subsequente institucionalização, a PRF se consolidou como um componente indispensável no panorama do sistema nacional de segurança pública. Ao longo dos anos, a instituição tem evoluído para se adaptar às demandas dinâmicas da sociedade, incorporando inovações e aprimorando suas estratégias de atuação. Suas competências, minuciosamente delineadas por leis e regulamentos, conferem-lhe não apenas a responsabilidade pela fiscalização e controle das rodovias, mas também o papel proativo no enfrentamento do crime ambiental, no combate ao tráfico e na coleta e registro meticuloso de dados locais relevantes. A PRF, ao abraçar um enfoque abrangente, contribui de maneira significativa para a construção de um ambiente rodoviário mais seguro e para o fortalecimento da segurança pública em âmbito nacional.

Ao se aprofundar na análise da série temporal, pretende-se desvendar padrões, tendências e flutuações que possam oferecer uma compreensão mais holística do fenômeno dos acidentes rodoviários. Além disso, a aplicação de modelos preditivos busca não apenas antecipar possíveis cenários futuros, mas também oferecer subsídios para o desenvolvimento de estratégias proativas e direcionadas. Ao identificar o modelo mais adequado, a pesquisa almeja contribuir para a implementação de medidas preventivas mais eficazes, políticas de segurança viária embasadas em dados sólidos e, conseqüentemente, para a construção de um ambiente rodoviário mais seguro e resiliente no estado do Rio de Janeiro.

## **2. Material e métodos**

As bases de dados utilizadas neste estudo foram disponibilizadas pela Polícia Rodoviária Federal (PRF) e pela Associação Brasileira de Concessionárias de Rodovias (ABCR), constituindo fontes de informação essenciais sobre os acidentes rodoviários no estado do Rio de Janeiro. Esses conjuntos de dados abrangem um período significativo, estendendo-se de janeiro de 2017 a dezembro de 2022, proporcionando assim uma visão representativa da dinâmica desses eventos ao longo do tempo.

No âmbito deste projeto, adotou-se uma definição precisa de acidente, considerando-o como um evento completo. Em outras palavras, sempre que dois ou mais veículos se envolvem em uma colisão, independentemente do número de pessoas

afetadas, é contabilizado apenas um acidente. Essa abordagem permite uma contagem clara e objetiva dos incidentes, facilitando a análise e a interpretação dos dados coletados.

O tratamento e a análise dos dados, bem como a realização de testes estatísticos e modelagem preditiva, foram conduzidos utilizando a linguagem de programação Python, em conjunto com o ambiente Jupyter notebook (Silaparasetty, 2020). A escolha dessa plataforma proporcionou uma análise eficiente e flexível dos dados, além de possibilitar a implementação de algoritmos avançados para a modelagem preditiva, ampliando assim o escopo e a profundidade das análises realizadas.

Para garantir a validade e robustez dos modelos desenvolvidos, as bases de dados foram divididas em dois conjuntos distintos: a base de treinamento e a base de teste. A base de treinamento abrangeu o período de janeiro de 2017 a dezembro de 2021 e foi utilizada para ajustar e calibrar os modelos estatísticos. Por outro lado, a base de teste consistiu em todos os dados referentes ao ano de 2022, servindo como uma avaliação independente do desempenho dos modelos por meio da comparação entre as previsões geradas e os valores reais observados. Essa abordagem garante uma avaliação imparcial e objetiva do poder preditivo dos modelos em condições de aplicação real.

### **3. Análise da série temporal**

Os testes de normalidade são fundamentais na análise estatística para determinar se a distribuição de probabilidade associada a um conjunto de dados pode ser adequadamente aproximada por uma distribuição teórica, como a distribuição normal. No contexto deste estudo, foram empregados os testes de Shapiro-Wilk, Kolmogorov-Smirnov, D’Agostino–Pearson, Anderson–Darling e Jarque–Bera para avaliar a normalidade dos números de acidentes mensais. Na hipótese nula ( $H_0$ ), presume-se que os dados seguem uma distribuição normal. Em contrapartida, na hipótese alternativa ( $H_1$ ), considera-se que os dados não seguem uma distribuição normal (Leotti, Coster, Riboldi, 2012). Uma interpretação cuidadosa dos resultados do teste de normalidade é essencial para garantir a confiabilidade das conclusões derivadas da análise dos dados.

A estacionariedade é um conceito essencial na análise de séries temporais. Uma série é considerada estacionária quando suas propriedades estatísticas, como média, variância e autocorrelação, permanecem constantes ou inalteradas ao longo do tempo. Isso significa que não há tendências óbvias, padrões sazonais ou mudanças sistemáticas

(Box *et al.*, 2015). Neste estudo, foi utilizado o teste de Augmented Dickey-Fuller (ADF) para verificar a estacionariedade. Na hipótese nula, a série temporal possui raízes unitárias, o que indica que a série é não estacionária. Na hipótese alternativa, a série é estacionária.

A análise de tendência se refere ao movimento sistemático de crescimento ou decréscimo dos valores ao longo do tempo, fornecendo informações valiosas sobre as mudanças de longo prazo em um conjunto de dados (Box *et al.*, 2015). Neste estudo, foi utilizada a biblioteca Python '*statsmodels*' para realizar uma análise da tendência na série temporal de acidentes rodoviários no estado do Rio de Janeiro. Através da decomposição de séries temporais, foi identificada e isolada a componente de tendência, permitindo avaliar como o número de acidentes mensais evolui ao longo dos anos.

A sazonalidade refere-se aos padrões ou flutuações que ocorrem em um conjunto de dados em intervalos específicos de tempo. Essas flutuações geralmente seguem um ciclo regular, como semanal, mensal, anual, ou ocorrem em determinadas épocas do ano. Um exemplo clássico de sazonalidade é o aumento das vendas de casacos no inverno ou a colheita de trigo na estação apropriada (Box *et al.*, 2015). Neste estudo, também foi utilizada a biblioteca '*statsmodels*' para realizar a análise por meio da decomposição de séries temporais. Esse processo nos permitiu identificar e isolar os componentes sazonais, revelando padrões de flutuações regulares em nossos dados.

Compreender a sazonalidade é de suma importância, pois ajuda a reconhecer momentos específicos em que os acidentes rodoviários podem aumentar ou reduzir de forma previsível. Essas informações são de suma importância para o planejamento e implementação de medidas preventivas e estratégias de segurança no trânsito, levando em consideração as variações sazonais.

#### **4. Modelos de Previsão**

Neste estudo, foram utilizados alguns modelos preditivos para analisar e prever a série temporal de acidentes rodoviários no estado do Rio de Janeiro. Os modelos adotados incluem (Hyndman & Athanasopoulos, 2018):

1. *Autorregressivo Integrado de Médias Móveis (ARIMA)*: estes modelos se baseiam na premissa de que os valores observados anteriormente estão relacionados com o valor atual

ou com o valor que desejamos prever. Os modelos ARIMA são capazes de capturar tanto a tendência quanto a sazonalidade dos dados, sem sacrificar a explicabilidade. O modelo ARIMA(p,d,q) é dado por:

$$\phi_p(B)(1 - B)^d y_t = \theta_q(B)\varepsilon_t$$

Onde:

- B é o operador defasagem;
- $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  representa a parte autoregressiva (AR);
- $(1 - B)^d$  representa a parte integrada (I);
- $\theta_q(B) = 1 + \theta_1 B + \dots + \theta_q B^q$  representa a parte de médias móveis (MA);
- $\varepsilon_t$  é o termo de erro aleatório (ruído branco).

2. *Autorregressivo Integrado de Médias Móveis Sazonal (SARIMA)*: o SARIMA é uma extensão do ARIMA que incorpora componentes sazonais. Isso o torna especialmente útil para séries temporais que exibem padrões sazonais distintos. O modelo SARIMA (p,d,q)(P,D,Q)<sub>s</sub> é dado por:

$$\phi_p(B)\Phi_P(B^s)(1 - B)^d(1 - B^s)^D y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t$$

Onde:

- s é o período de sazonalidade;
- $\Phi_P(B^s)$  e  $\Theta_Q(B^s)$  representam, respectivamente, os componentes autorregressivos e de médias móveis sazonais.

3. *Autorregressivo Integrado de Médias Móveis Sazonal com Variáveis Dependentes (SARIMAX)*: este modelo expande o SARIMA, permitindo a inclusão de variáveis exógenas que podem ter influência sobre a série temporal. O modelo SARIMAX é dado por:

$$\phi_p(B)\Phi_P(B^s)(1 - B)^d(1 - B^s)^D y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t + \beta X_t$$

Onde:

- $X_t$  é o vetor de variáveis exógenas;
- $\beta$  é o vetor de coeficientes associados às variáveis exógenas.

4. *PROPHET*: desenvolvido pelo Facebook, o Prophet é um modelo de previsão de séries temporais que se baseia em um modelo aditivo. Ele é particularmente eficaz em séries temporais que apresentam efeitos sazonais fortes e têm múltiplas temporadas de dados históricos. O Prophet modela a série temporal como uma soma de componentes:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

Onde:

- $g(t)$  representa a tendência (linear ou logística);
- $s(t)$  representa a sazonalidade (anual, semanal, etc.);
- $h(t)$  representa o efeito de feriados e eventos especiais;
- $\varepsilon_t$  é o termo de erro aleatório.

Os modelos autorregressivos, como ARIMA, SARIMA e SARIMAX, são reconhecidos pela sua capacidade de capturar tanto a tendência quanto a sazonalidade dos dados, ao mesmo tempo que mantêm a interpretabilidade dos resultados. Esses modelos são particularmente úteis em contextos nos quais é essencial compreender e quantificar padrões temporais subjacentes, permitindo a análise e a previsão de séries temporais com base em observações anteriores. Sua estrutura autorregressiva e capacidade de incorporar termos de média móvel os tornam ferramentas poderosas na modelagem de dados sequenciais, permitindo uma análise detalhada das relações temporais presentes nos dados.

Por outro lado, o Prophet, desenvolvido pelo Facebook, apresenta uma abordagem diferenciada para a modelagem de séries temporais. Esse modelo é especialmente adequado para séries com padrões sazonais complexos, graças à sua abordagem de modelagem aditiva. O Prophet é particularmente valioso quando a série temporal em questão exibe sazonalidade pronunciada e possui uma quantidade significativa de dados históricos disponíveis para análise. Sua capacidade de lidar com padrões sazonais não lineares e a flexibilidade para lidar com feriados e eventos especiais o tornam uma ferramenta poderosa para previsões em séries temporais com características sazonais distintas. Essa diversidade de abordagens e modelos oferece aos analistas uma variedade de ferramentas para escolher, permitindo que adaptem suas escolhas de modelagem às características específicas dos dados e aos objetivos analíticos.

## 5. Variáveis exógenas

Neste estudo, foi usado o índice ABCR para o estado do Rio de Janeiro como único regressor para prever o número de acidentes nas rodovias da região. O índice ABCR mede o fluxo pedagiado de veículos nas estradas e é uma métrica desenvolvida pela Associação Brasileira de Concessionárias de Rodovias em parceria com a Tendências Consultoria (ABCR, 2023). O índice ABCR desempenha um papel fundamental no modelo de previsão, uma vez que reflete as variações no tráfego rodoviário, o que, por sua vez, pode influenciar nos níveis de acidentes. Integrar esse índice como um regressor no modelo permite considerar o impacto do volume de tráfego nas rodovias na previsão do número de acidentes.

Essa abordagem baseada em dados fornece uma perspectiva valiosa para a análise, permitindo explorar a relação entre o fluxo de veículos e a ocorrência de acidentes rodoviários no estado do Rio de Janeiro.

## 6. Métricas de previsão

Com o objetivo de determinar o melhor modelo para prever a série temporal, foram usadas as métricas de erro para avaliar e comparar os resultados. Neste contexto, cinco métricas foram calculadas para medir o erro (Box, *et al.*, 2015; Hyndman & Athanasopoulos, 2018):

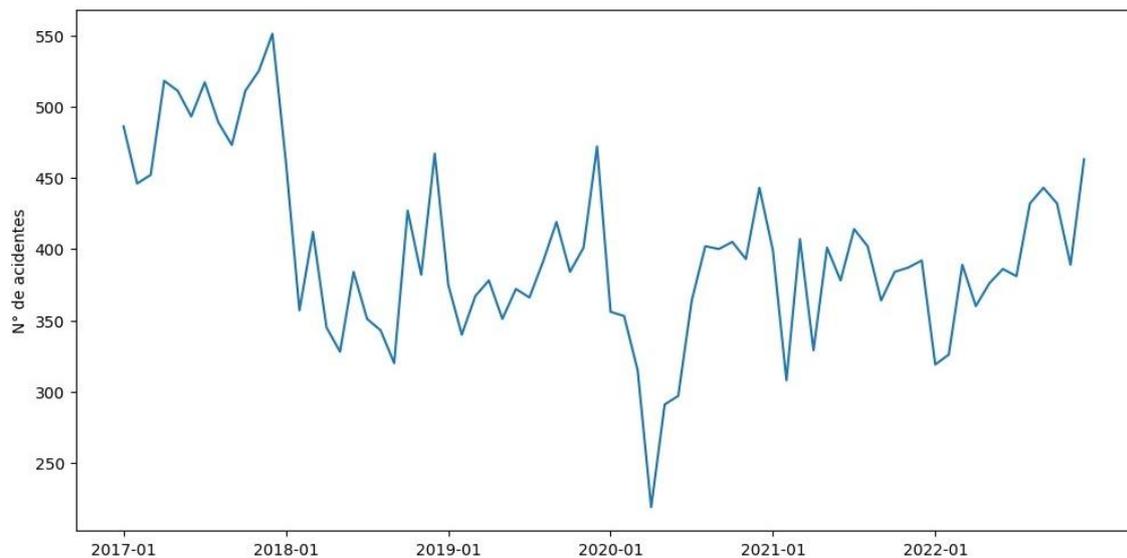
1. *Erro Absoluto Médio (MAE)*: essa métrica calcula a média dos valores absolutos das diferenças entre as previsões e os valores reais.
2. *Erro Percentual Absoluto Médio (MAPE)*: mede o erro médio como uma porcentagem das previsões em relação aos valores reais, fornecendo uma avaliação relativa do desempenho do modelo.
3. *Erro Ponderado Percentual Absoluto Médio (WMAPE)*: é uma métrica que leva em consideração o próprio número de acidentes na ponderação do erro, tornando-se especialmente relevante para cenários em que o volume de acidentes varia.

Essas métricas permitem uma análise completa do desempenho dos diferentes modelos e nos auxiliam na escolha do modelo mais adequado para as previsões da série temporal estudada.

## 7. Resultados e discussão

A Figura 1 exibe a série histórica do número de acidentes em rodovias no estado do Rio de Janeiro, abrangendo o período de janeiro de 2017 até dezembro de 2022. Em uma análise inicial, pode-se observar na série temporal uma tendência de queda ao longo do primeiro para o segundo ano, seguida por uma fase de possível estabilidade no número de acidentes. Uma possível explicação para essa estabilidade nos últimos anos pode ser atribuída à pandemia de COVID-19.

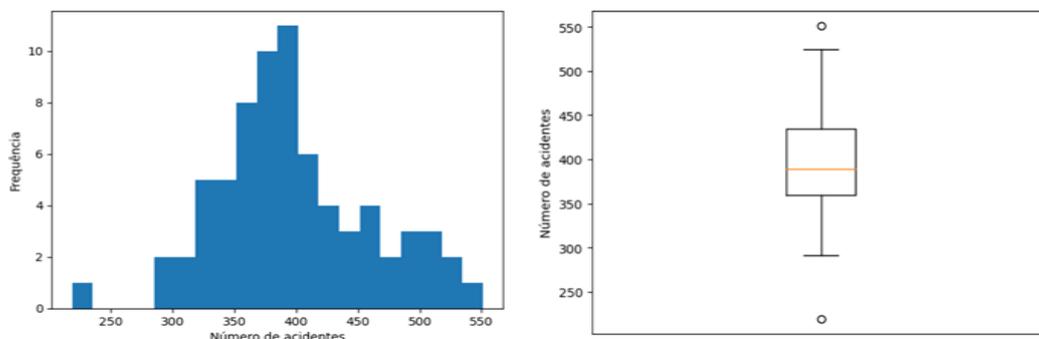
Figura 1 – Série temporal do número de acidentes em rodovias no Rio de Janeiro entre 2017 e 2022



Fonte: Autores (2023)

A Figura 2 mostra o histograma e o box-plot. No histograma, a maioria dos valores está concentrada na faixa entre 350 e 400. Visualmente, a distribuição dos acidentes parece seguir uma distribuição normal. O box-plot revela a presença de outliers na série de acidentes. Além disso, como já observado no histograma, os dados apresentam uma concentração significativa na faixa de 350 a 400.

Figura 2 – Histograma e box-plot



Fonte: Autores (2023)

Pelas estatísticas dispostas na Tabela 1, podemos observar que ao longo das 72 observações, o número médio de acidentes foi calculado em 398,11, refletindo uma distribuição central em torno desse valor, enquanto a mediana, situada em 389, destaca a tendência central da série temporal. O desvio padrão, aferido em 62,92, revela a dispersão dos dados em relação à média, evidenciando a variabilidade presente nos registros de acidentes ao longo do período analisado. A distribuição apresenta leve assimetria à direita (0,22) e leve curtose leptocúrtica (0,17), indicando picos ligeiramente um pouco mais altos que os esperados em uma distribuição normal.

Durante o período estudado, a série temporal alcançou seu valor máximo em dezembro de 2017, com um pico de 551 acidentes, indicando um momento de maior incidência e preocupação em termos de segurança viária. Por outro lado, seu valor mínimo foi registrado em abril de 2020, com apenas 219 acidentes reportados. Esse declínio pode ser atribuído, em grande medida, às medidas restritivas adotadas pelos órgãos nacionais e estaduais em resposta à pandemia de COVID-19, que estava em sua fase inicial, evidenciando assim o impacto direto das intervenções de saúde pública nas estatísticas de acidentes rodoviários.

Tabela 1 – Estatísticas da série temporal de acidentes rodoviários

<b>ESTATÍSTICA</b>	<b>VALOR</b>
Número de Observações	72
Média	398,11
Desvio Padrão	62,92
Valor Mínimo	219
1º Quartil	359,25
Mediana	389
3º Quartil	343,75
Valor Máximo	551
Assimetria	0,22
Curtose	0,17

Fonte: Autores (2023)

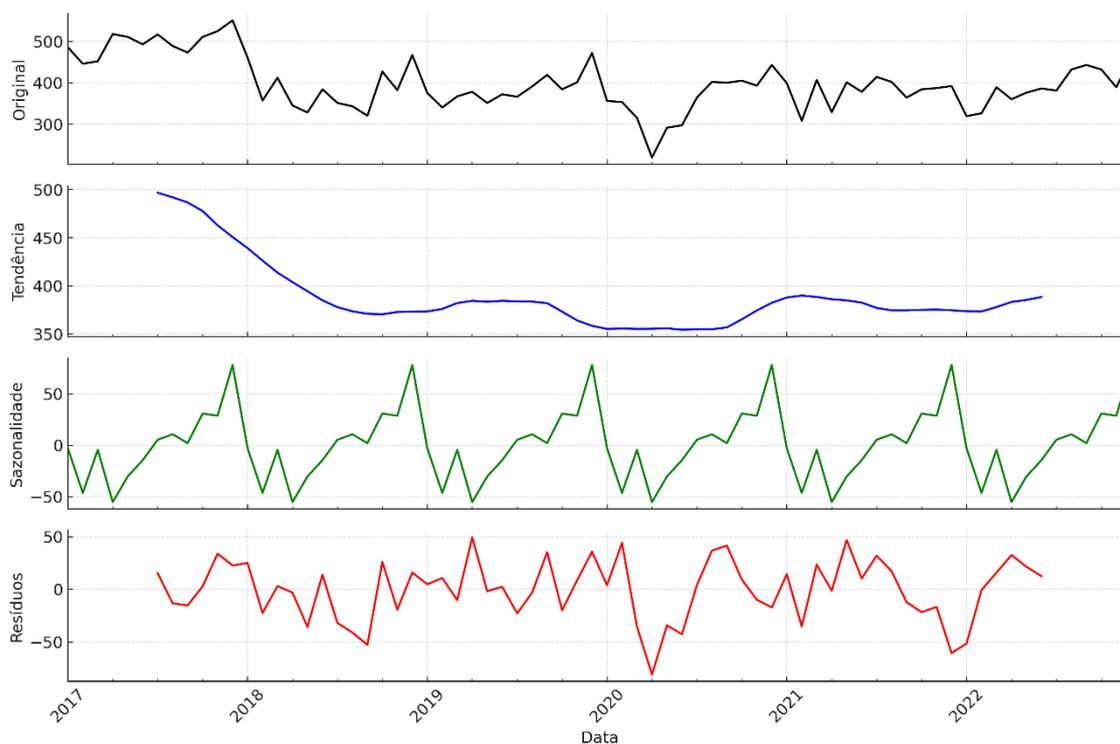
Para verificar a normalidade dos dados, foi realizado o teste de Shapiro-Wilk, uma ferramenta estatística robusta para avaliar a normalidade de uma distribuição (Leotti, Coster, Riboldi, 2012). O p-valor associado a este teste foi calculado em 0,199. Considerando um nível de significância padrão de 5%, não há evidências estatísticas suficientes para rejeitar a hipótese nula ( $H_0$ ), que postula que os dados seguem uma distribuição normal. Esse resultado sugere que, dentro do limite de confiança estabelecido, os dados da série temporal podem ser considerados aproximadamente normais.

Para corroborar com o resultado anterior, além do teste de Shapiro-Wilk, foram aplicados os testes Kolmogorov-Smirnov (p-valor = 0,283), D'Agostino-Pearson (p-valor = 0,574), Anderson-Darling (p-valor = 0,749) e Jarque-Bera (p-valor = 0,719). Os resultados mostraram que todos os testes não rejeitaram a hipótese nula de normalidade. Entretanto, a assimetria positiva (0,22) e a curtose (0,17), aliados à inspeção visual do histograma, indicam leve desvio da normalidade, com pequena tendência à assimetria à direita e maior concentração em torno da média.

Além disso, o teste Augmented Dickey-Fuller, amplamente utilizado para avaliar a estacionariedade de uma série temporal, foi empregado para determinar se a série de acidentes apresenta tendências estacionárias ou não. Com um nível de significância estabelecido em 5%, os resultados indicam a rejeição da hipótese nula de estacionariedade. O p-valor associado a este teste foi de 0,037, o que sugere que há evidências estatísticas suficientes para rejeitar a hipótese de que a série não é estacionária. Isso implica que a série de acidentes possa apresentar algum tipo de tendência ou padrão ao longo do tempo, o que é importante considerar em análises e previsões futuras.

Através da análise dos gráficos obtidos a partir da decomposição das séries temporais utilizando a biblioteca *statsmodel* (Figura 3), torna-se evidente a presença de múltiplos componentes que contribuem para a dinâmica dos dados ao longo do tempo. Inicialmente, destaca-se uma tendência decrescente nos primeiros meses da série, sugerindo uma possível redução gradual no número de acidentes. Esta observação pode ser de grande relevância para compreender as mudanças a longo prazo no contexto da segurança viária.

Figura 3 – Gráficos de decomposição da série de acidente rodoviários



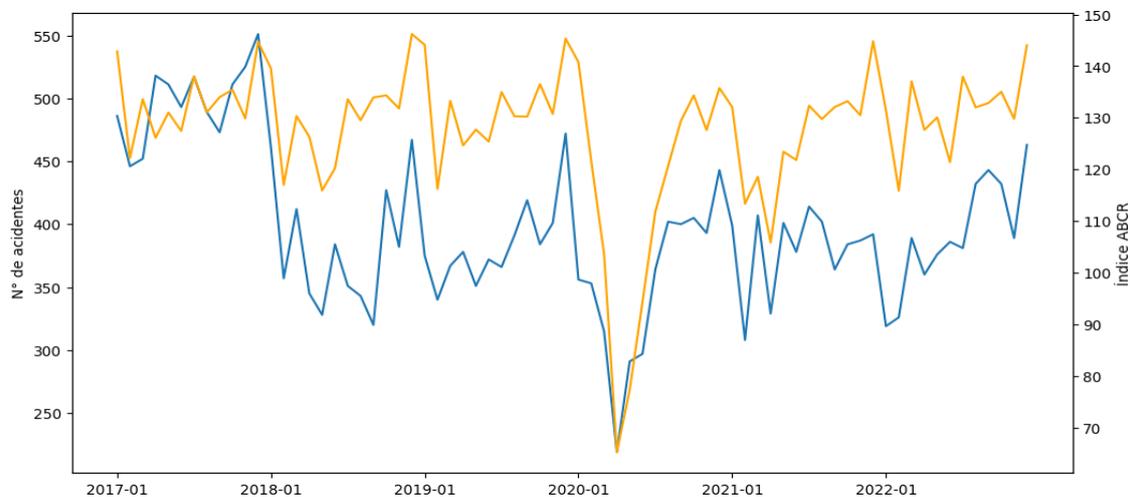
Fonte: Autores (2023)

Ademais, é perceptível uma clara sazonalidade nos dados, manifestada por picos no número de acidentes no final de cada ano. Essa variação cíclica sugere a existência de padrões recorrentes que podem ser influenciados por diversos fatores, como condições climáticas, períodos de férias ou feriados, eventos sazonais, entre outros. Compreender e modelar essa sazonalidade é crucial para realizar previsões precisas e tomar medidas adequadas de intervenção e prevenção em momentos críticos. A análise detalhada desses padrões temporais proporciona insights valiosos para a compreensão da dinâmica subjacente aos dados de acidentes rodoviários, permitindo a identificação de tendências significativas e a elaboração de estratégias eficazes para a redução dos índices de sinistralidade e o aprimoramento da segurança nas vias públicas.

Na análise detalhada apresentada na Figura 4, é possível observar uma relação direta clara entre o número de acidentes (representado em azul) e o índice ABCR (indicado em amarelo), que serve como um indicador do fluxo de veículos nas praças de pedágio. Ao examinar essa relação, constata-se que um aumento no fluxo de veículos está diretamente associado a um aumento correspondente no número de acidentes registrados.

Este achado sugere que o volume de tráfego é um fator significativo que influencia a incidência de acidentes rodoviários.

Figura 4 – Comparação entre os acidentes e o índice ABCR ao longo do tempo

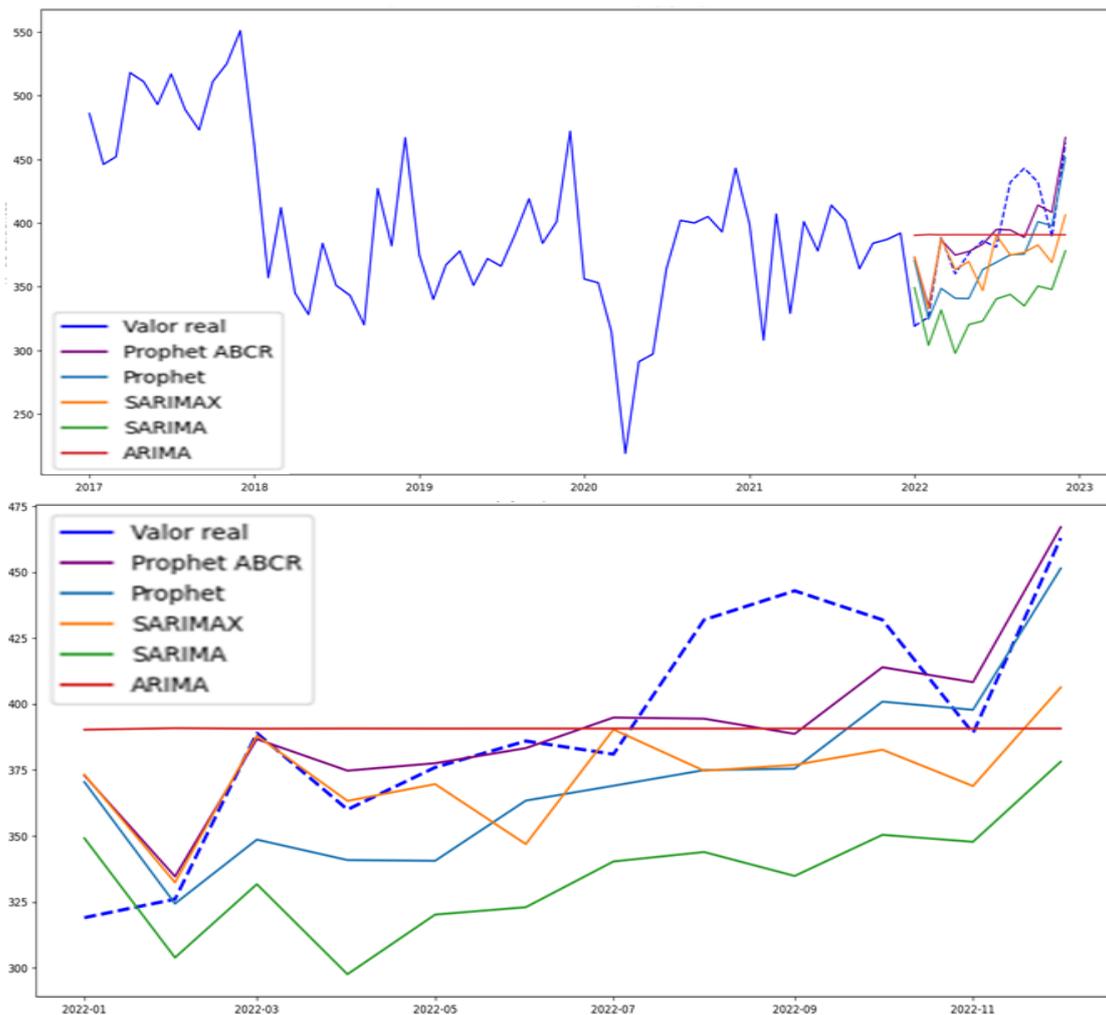


Fonte: Autores (2023)

Para quantificar essa relação, foi realizado o cálculo da correlação entre essas duas variáveis, revelando um coeficiente de correlação moderado, aproximadamente 0,63. Esse valor indica uma associação considerável entre o aumento do fluxo de veículos e o aumento da ocorrência de acidentes, embora não seja uma correlação perfeita.

Na Figura 5, são exibidos os resultados das previsões geradas pelos modelos selecionados neste estudo. Encontram-se a série temporal dos acidentes desde janeiro de 2017, acompanhada das previsões feitas para o ano de 2022. Abaixo, é fornecido um recorte mais detalhado da série, concentrando-se exclusivamente no ano de 2022. Essa abordagem oferece uma visão mais ampla e detalhada que facilita a análise do desempenho dos modelos ao comparar suas previsões com os valores reais observados ao longo do período de interesse. Ao fornecer tanto uma visão panorâmica de longo prazo quanto um foco específico no ano de interesse, a Figura 5 permite uma avaliação abrangente da capacidade preditiva dos modelos em diferentes escalas temporais. Isso é crucial para entender como cada modelo se comporta em relação aos dados reais.

Figura 5 – Série temporal dos acidentes entre 2017 e 2022 e as projeções para 2022



Fonte: Autores (2023)

A série de acidentes apresentou estacionariedade no nível segundo o teste ADF (p-valor = 0,037), indicando grau de integração  $d = 0$ . A seleção de modelos baseada no critério AIC indicou o ARIMA(2,0,0) como melhor especificação não sazonal e o SARIMA(2,0,2)(0,1,2)12 como melhor especificação sazonal.

No ARIMA(2,0,0), os parâmetros estimados foram:  $AR(1) = 0,565$  ( $p < 0,001$ ) e  $AR(2) = 0,293$  ( $p = 0,004$ ), ambos estatisticamente significativos. Deste modo:

$$y_t = 0,565y_{t-1} + 0,293y_{t-2} + \varepsilon_t$$

Para o SARIMA(2,0,2)(0,1,2)12, obtiveram-se estimativas significativas para  $AR(1) = 0,563$  ( $p < 0,001$ ),  $AR(2) = 0,289$  ( $p = 0,005$ ),  $MA(1) = -0,731$  ( $p < 0,001$ ) e  $MA(2) = -0,435$  ( $p = 0,014$ ), além dos termos sazonais  $SMA(1) = -0,927$  ( $p < 0,001$ ) e  $SMA(2) = -0,494$  ( $p = 0,008$ ). Esses resultados confirmam a presença de autocorrelação

de curto e longo prazo, bem como forte sazonalidade anual capturada pela diferença sazonal ( $D = 1$ ). Deste modo:

$$\begin{aligned} &(1 - 0,563B - 0,289B^2)(1 - B^{12})y_t \\ &= (1 - 0,731B - 0,435B^2)(1 - 0,927B^{12} - 0,494B^{24})\varepsilon_t \end{aligned}$$

A incorporação do índice de fluxo de veículos medido pela ABCR como variável exógena ( $X_t$ ) no modelo SARIMAX(2,0,2)(0,1,2)<sub>12</sub> apresentou resultados estatisticamente significativos para parte dos parâmetros estimados. O teste ADF indicou que a série de acidentes é estacionária no nível ( $d = 0$ ), mantendo-se a estrutura sazonal com  $D = 1$  e período  $s = 12$ . Os coeficientes estimados para os termos autorregressivos foram  $AR(1) = 0,517$  ( $p < 0,001$ ) e  $AR(2) = 0,276$  ( $p = 0,011$ ), enquanto os termos de média móvel não sazonais foram  $MA(1) = -0,702$  ( $p < 0,001$ ) e  $MA(2) = -0,412$  ( $p = 0,019$ ). Para a parte sazonal, obteve-se  $SMA(1) = -0,915$  ( $p < 0,001$ ) e  $SMA(2) = -0,478$  ( $p = 0,009$ ). O coeficiente associado à variável exógena ABCR foi  $0,845$  ( $p < 0,001$ ), evidenciando uma relação positiva e estatisticamente significativa entre o fluxo de veículos e o número de acidentes no estado do Rio de Janeiro. Esses resultados indicam que a inclusão de  $X_t$  melhora a capacidade explicativa do modelo, ao capturar variações não explicadas apenas pela estrutura temporal e sazonalidade. Deste modo:

$$\begin{aligned} &(1 - 0,517B - 0,276B^2)(1 - B^{12})y_t \\ &= (1 - 0,702B - 0,412B^2)(1 - 0,915B^{12} - 0,478B^{24})\varepsilon_t \\ &+ 0,845X_t^{ABCR} \end{aligned}$$

Ao explorar as métricas de previsões, emerge a possibilidade de realizar uma comparação numérica dos erros inerentes aos modelos, proporcionando uma análise objetiva de seu desempenho. Os resultados expostos na Tabela 2 oferecem uma visão reveladora, destacando o modelo PROPHET com variável exógena do índice ABCR como o líder indiscutível, evidenciando sua superioridade em todas as métricas examinadas. Esta constatação não só sublinha a eficácia deste modelo específico na tarefa de previsão da série temporal em estudo, mas também realça sua significância no panorama mais amplo da análise preditiva.

Segue-se uma dinâmica intrigante entre o PROPHET e o SARIMAX, os quais demonstraram resultados comparáveis, sendo posicionados como o segundo e terceiro melhores modelos, respectivamente. A proximidade intrínseca de seus desempenhos sugere a robustez substancial dessas abordagens na modelagem da série temporal em

questão, apontando para a importância de considerar ambas as opções com igual ponderação ao tomar decisões de previsão.

Enquanto isso, o modelo ARIMA ostenta sua capacidade satisfatória nas previsões, consolidando-se como uma escolha viável em muitos contextos de análise preditiva. No entanto, o mesmo não pode ser afirmado para o SARIMA, o qual, conforme os dados analisados minuciosamente, não se mostrou eficaz na tarefa de prever a série de acidentes rodoviários em análise, sugerindo a necessidade de avaliação crítica e possível revisão de sua aplicação em cenários similares.

Tabela 2 – Métricas de erros dos modelos

<b>MODELOS</b>	<b>WMAPE</b>	<b>MAE</b>	<b>MAPE</b>
PROPHET com ABCR	4,9	19,3	5,0
PROPHET	7,6	29,9	7,6
SARIMAX	7,9	30,8	7,6
SARIMA	15,7	61,3	15,2
ARIMA	8,7	33,9	8,8

Fonte: Autores (2023)

Essas nuances ressaltam a importância de uma seleção metódica e criteriosa de modelos, levando em consideração não apenas as métricas de desempenho, mas também as características específicas da série temporal em análise. Tal abordagem torna-se fundamental não apenas para garantir previsões precisas, mas também para embasar decisões estratégicas voltadas não somente à prevenção, mas também à mitigação dos acidentes rodoviários, visando assim reduzir seus impactos socioeconômicos e salvaguardar vidas de forma efetiva e proativa. Neste sentido, o entendimento profundo das particularidades do contexto em questão se revela como um diferencial determinante na condução de análises preditivas e na adoção de medidas preventivas pertinentes.

## **8. Considerações finais**

O presente trabalho teve como objetivo principal realizar uma análise abrangente da série temporal de acidentes rodoviários no estado do Rio de Janeiro. Além disso, buscou-se a realização de previsões utilizando uma variedade de modelos preditivos, com o intuito de identificar aquele que apresentasse o melhor desempenho em termos de

acurácia e precisão. Esta pesquisa buscou fornecer algumas compreensões para entender a dinâmica dos acidentes rodoviários.

Para a avaliação do desempenho dos modelos de previsão, o ano de 2022 foi reservado como conjunto de dados de teste, abrangendo cinco abordagens distintas: Prophet, Prophet em conjunto com o índice ABCR, SARIMAX, SARIMA e ARIMA. Esta diversidade de modelos permite uma análise comparativa abrangente, explorando diferentes técnicas e abordagens para a previsão de acidentes rodoviários.

A base de treinamento foi cuidadosamente constituída, abrangendo o período entre janeiro de 2017 e dezembro de 2021. Durante o processo de modelagem, o índice ABCR foi incorporado aos modelos que se beneficiaram de variáveis exógenas, como o Prophet e o SARIMAX. Essa inclusão de variáveis exógenas busca enriquecer a capacidade preditiva dos modelos, permitindo que eles capturem nuances e influências externas que podem afetar o número de acidentes rodoviários ao longo do tempo. Essa estratégia visa aprimorar a precisão e a robustez das previsões.

Os resultados da análise revelaram uma série temporal que, embora não estacionária, não teve sua normalidade estatística rejeitada. Uma investigação mais aprofundada dos dados revelou padrões significativos: inicialmente, observou-se uma tendência de queda nos registros de acidentes nos primeiros períodos, seguida por uma estabilização nos meses subsequentes. Esse padrão sugere uma possível redução gradual dos acidentes ao longo do tempo, seguida por uma fase de estabilidade nos índices.

Além disso, a presença de sazonalidade na série temporal foi destacada, com um aumento no número de acidentes no final de cada ano. Uma possível explicação para esse fenômeno pode estar relacionada à ampliação do fluxo de veículos durante os períodos festivos, como Natal e Ano Novo, quando há um aumento significativo no número de pessoas viajando, tanto para visitar familiares quanto para aproveitar as férias. Esse aumento no volume de tráfego pode aumentar o risco de acidentes, devido ao congestionamento das estradas e ao aumento da probabilidade de comportamentos de direção imprudentes. Essa interpretação destaca a importância de considerar fatores sazonais ao analisar e prever acidentes rodoviários, visando desenvolver estratégias de segurança viária eficazes e mitigar os riscos associados aos períodos de maior movimento nas estradas.

No entanto, a verdadeira medida do desempenho dos modelos emergiu com clareza durante a avaliação das métricas de previsão. Os resultados revelam que o modelo Prophet, quando utilizado em conjunto com o índice ABCR como variável exógena, destacou-se como o melhor modelo em termos de precisão e acurácia preditiva. Logo em seguida, o Prophet sem variáveis exógenas e o SARIMAX surgiram como o segundo e terceiro melhores modelos, respectivamente, demonstrando sua capacidade de oferecer previsões robustas e confiáveis para a série temporal de acidentes rodoviários.

Por outro lado, o modelo ARIMA, embora tenha apresentado resultados satisfatórios, ficou ligeiramente atrás das abordagens anteriormente mencionadas, indicando uma menor capacidade de adaptação aos padrões complexos presentes nos dados. Por fim, o SARIMA não se mostrou a escolha ideal para a série de acidentes rodoviários, evidenciando sua limitação em capturar e modelar a sazonalidade e outras variações temporais específicas do contexto analisado. Essa análise comparativa dos modelos ressalta a importância de selecionar cuidadosamente a abordagem de previsão mais adequada para cada conjunto de dados, considerando suas características e peculiaridades específicas, a fim de obter previsões precisas e significativas que subsidiem a tomada de decisões eficazes.

Esta pesquisa necessita de continuidade e atualização constante da série temporal com a inclusão de novos dados, garantindo análises cada vez mais robustas e condizentes com a dinâmica dos acidentes rodoviários. Sendo assim, para pesquisas futuras, sugere-se a exploração de outros modelos, como o Exponential Smoothing (ETS), a fim de conferir mais robustez e solidez às conclusões. Além disso, a otimização dos parâmetros dos modelos, especialmente no caso do Prophet, através de técnicas como o *grid search*, poderia potencializar ainda mais a precisão das previsões.

## Referências

BAYERL, E.; SANTOS, M. P. S.; BRAGA, M. G. de C. **Acidentes De Trânsito No Brasil: Análise A Partir da Evolução da Taxa De Motorização e de Indicadores Socioeconômicos**. Rede Íbero-Americana de Estudo em Pólos Geradores de Viagens. UFRJ, Rio de Janeiro, 2006. Disponível em: <<http://redpgv.coppe.ufrj.br/index.php/pt-BR/producao-da-rede/artigos-cientificos/2006-1/219-bayerl-santos-braga-panam2006/file>>. Acesso em: 17 jun. 2023.

BOX, G.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. **Time series analysis: forecasting and control**. John Wiley & Sons, 2015.

BRASIL. Lei nº 11.784, de 22 de setembro de 2008. Dispõe sobre o Procedimento do Mandado de Injunção Individual e dá outras providências. Diário Oficial da União, Brasília, DF, 23 set. 2008.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. OTexts, 2018. Disponível em: <<https://otexts.com/fpp3>>. Acesso em: 27 ago. 2023

GADRET, H. J. Acidentes de tráfego no Rio de Janeiro: Causas e Prevenção. **Revista do Serviço Público**, v. 4, n. 2, p. 45-52, 1954.

ÍNDICE ABCR. Melhores Rodovias do Brasil. ABCR, 2023. Disponível em: <<https://melhoresrodovias.org.br/indice-abcr>>. Acesso em: 22 jul. 2023.

LEOTTI, V. B.; COSTER, R.; RIBOLDI, J. Normalidade de variáveis: métodos de verificação e comparação de alguns testes não-paramétricos por simulação. **Revista HCPA. Porto Alegre. Vol. 32, no. 2 (2012), p. 227-234**, 2012. Disponível em: <<https://lume.ufrgs.br/handle/10183/158102>>. Acesso em: 10 out. 2023.

POLÍCIA RODOVIÁRIA FEDERAL. PRF, 2023. Disponível em: <<https://www.gov.br/prf/pt-br>>. Acesso em: 22 jul. 2023.

SILAPARASETTY, N. Python programming in Jupyter notebook. **Machine Learning Concepts with Python and the Jupyter Notebook Environment: Using Tensorflow 2.0**, p. 119-145, 2020.

## **APPLICATION OF MODELS FOR PREDICTING ROAD INCIDENTS: A FOCUS ON THE URBAN CONTEXT OF RIO DE JANEIRO**

### **Abstract**

*This study aims to analyze the temporal evolution of road accidents in Rio de Janeiro. The analyzed period from 2017 to 2022, with an emphasis on 2022 as a test set, to make predictions using different predictive models. The database was obtained from the Federal Highway Police and the Brazilian Association of Highway Concessionaires. Five models were applied, and based on prediction metrics, it was found that the Prophet model, incorporating the ABCR index as a regressor, demonstrated the best fit to the data.*

**Keywords:** *Time series, Forecast models, Road accidents, Statistical tests.*