



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO ACADÊMICO DO AGRESTE
NÚCLEO DE TECNOLOGIA
CURSO DE ENGENHARIA DE PRODUÇÃO

GABRIEL HENRIQUE DE MELO FONSECA

**APLICAÇÃO DO MODELO SARIMA E HOLT-WINTERS PARA PREVISÃO DE
CRIMES NO ESTADO DE PERNAMBUCO**

Caruaru
2023

GABRIEL HENRIQUE DE MELO FONSECA

**APLICAÇÃO DO MODELO SARIMA E HOLT-WINTERS PARA PREVISÃO DE
CRIMES NO ESTADO DE PERNAMBUCO**

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Produção da Universidade Federal de Pernambuco, na modalidade monografia, como requisito parcial para a obtenção do título de Bacharel em Engenharia de Produção.

Área de concentração: Pesquisa Operacional

Orientador: Prof. Dr. Caio Bezerra Souto Maior

Caruaru

2023

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Fonseca, Gabriel Henrique de Melo.

APLICAÇÃO DO MODELO SARIMA E HOLT-WINTERS PARA
PREVISÃO DE CRIMES NO ESTADO DE PERNAMBUCO / Gabriel
Henrique de Melo Fonseca. - Caruaru, 2023.

33

Orientador(a): Caio Bezerra Souto Maior

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de
Pernambuco, Centro Acadêmico do Agreste, Engenharia de Produção, 2023.
8,5.

1. Holt-Winters . 2. SARIMA. 3. CVLIs. 4. Apreensão de Armas. 5.
Previsões. I. Souto Maior, Caio Bezerra. (Orientação). II. Título.

620 CDD (22.ed.)

GABRIEL HENRIQUE DE MELO FONSECA

**APLICAÇÃO DO MODELO SARIMA E HOLT-WINTERS PARA PREVISÃO DE
CRIMES NO ESTADO DE PERNAMBUCO**

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Produção da Universidade Federal de Pernambuco, na modalidade monografia, como requisito parcial para a obtenção do título de Bacharel em Engenharia de Produção.

Aprovada em: 20/09/2023

BANCA EXAMINADORA

Prof. Dr. Caio Bezerra Souto Maior (Orientador)
Universidade Federal de Pernambuco

Prof. Dr. Walton Pereira Coutinho (Examinador interno)
Universidade Federal de Pernambuco

Profa. Dra. July Bias Macedo (Examinadora externa)
Universidade Federal de Pernambuco

AGRADECIMENTOS

Quero agradecer primeiramente a Deus por todo discernimento e sabedoria na decisão e construção desse trabalho e finalização do curso. Agradecer também a Nossa Senhora por ter me ungido com seu manto sagrado durante essa jornada.

A minha família, por todo o apoio, motivação e paciência. Especialmente aos meus pais, Sandro e Andréa, por terem me mostrado como é ser um homem de Deus e lutar por tudo o que sonho independente do desafio. Pai, minha inspiração e referência em tudo na minha vida, por ter me mostrado que com a disciplina e fé em Deus tudo é possível e alcançável. Minha mãe, que com seu jeito singular e feliz me mostrou que em certos momentos deve-se parar pedir orientação a Deus e seguir em frente. Meu irmão Gui, que me ajudou fornecendo equipamento para a formação desse trabalho.

A minha futura esposa Maria Tereza, que soube dividir comigo essa construção e me ajudou em tudo para que isso fosse possível.

Aos meus amigos que fiz durante o curso, que irei levar para vida e que me ajudaram a suportar as adversidades e irmos até o fim desse processo.

Aos professores do curso que tive a honra de presenciar suas aulas e aprender mais sobre essa incrível área. Em especial a Prof. Caio meu orientador, que se disponibilizou a me orientar e a me fornecer ideias primordiais, mesmo em um intervalo curto e com alguns desafios, mas que sem ele não seria possível. Ao Prof. Thyago que através das aulas e o grupo de pesquisa GPSID me mostrou a ideia que se construiu todo esse trabalho.

A SEPLAG pelo fornecimento dos dados, os quais foram utilizados, e as reuniões que permitiu entender mais sobre a problemática e os comportamentos dos dados, meus sinceros agradecimentos ao apoio.

Por último, quero agradecer a UFPE-CAA por me acolher e me formar o profissional que estou sendo hoje, sempre buscando inovações, persistência e sempre pensando à frente.

RESUMO

A criminalidade é uma questão pertinente no cotidiano dos governos e municípios do Brasil. Considerando os dados disponibilizados pelas instituições públicas, é possível interpretá-los e aplicar técnicas de estimação de ocorrências futuras que possam ajudar nas tomadas de decisões e mitigar a criminalidade. Dessa forma, este estudo tem como objetivo a previsão de crimes em grandes cidades do estado de Pernambuco, utilizando modelos clássicos de previsão. Assim, realizou-se uma análise e tratamento dos dados, para então previsão dos crimes e verificação de relações entre eles. Depois utilizando o SARIMA e Holt-Winters foram identificados a frequência para cada série do conjunto de a partir da Função de Autocorrelação (ACF) e Função de Autocorrelação Parcial (PACF). Após a aplicação dos modelos de previsão, foram observados as métricas dos erros e verificado os melhores modelos. Com isso, foi identificado que o modelo SARIMA foi o mais adequado para a previsão apreensão de armas e o modelo HW foi o mais adequado para a ocorrência de Crimes Violentos Letais e Intencionais (CVLI), como também observou uma relação de antagonismo entre as previsões das duas séries.

Palavras-chaves: Holt-Winters; SARIMA; CVLIs; Apreensão de Armas; Previsões.

ABSTRACT

Crime is a pertinent issue in the daily lives of governments and municipalities in Brazil. Considering the data made available by public institutions, it is possible to interpret them and apply techniques to estimate future occurrences that can help in decision-making and mitigate crime. Thus, this study aims to predict crimes in large cities in the state of Pernambuco, using classical prediction models. Thus, an analysis and treatment of the data was carried out, in order to predict the crimes and verify the relationships between them. Then, using SARIMA and Holt-Winters, the frequency for each series of the set of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) were identified. After applying the prediction models, the error metrics were observed and the best models were verified. Thus, it was identified that the SARIMA model was the most appropriate for the prediction of weapons seizure and the HW model was the most appropriate for the occurrence of Lethal and Intentional Violent Crimes (CVLI), as well as observing a relationship of antagonism between the predictions of the two series.

Keywords: Holt-Winters; SARIMA; CVLIs; Seizure of Weapons; Forecasts.

LISTA DE FIGURAS

Figura 1 -	Taxa de Crimes Violentos Letais e Intencionais 2016-2020	11
Figura 2 -	Série histórica de 2018 a 2021, distribuídos em semanas	22
Figura 3 -	ACF e PACF da série homicídios	22
Figura 4 -	ACF e PACF da série apreensão de armas	23
Figura 5 -	Boxplot dos dados	24
Figura 6 -	Previsão da série homicídios em relação a série real	26
Figura 7 -	Previsão da série apreensão de armas em relação a série real	27
Figura 8 -	Previsões HW e SARIMA e respectivamente nas janelas de teste, no ano de 2021	

LISTA DE TABELAS

Tabela 1 -	Resumo dos dados	23
Tabela 2 -	Relação dos dados de covariância e correlação	25
Tabela 3 -	Métricas de erro das séries apreensão de armas e homicídios	28

SUMÁRIO

1	INTRODUÇÃO	10
1.1	OBJETIVOS	13
1.1.1	Geral.....	13
1.1.2	Objetivos específicos	13
2	REVISÃO DE LITERATURA	14
3	REFERENCIAL TEÓRICO.....	16
3.1	SÉRIES TEMPORAIS.....	16
3.1.1	Estacionaridade.....	16
3.1.2	Sazonalidade.....	16
3.1.3	Tendência.....	16
3.2	MODELOS CLÁSSICOS DE PREVISÃO.....	17
3.2.1	Modelo de Suavização Exponencial Simples (SES)	17
3.2.2	Modelo de Holt-Winters (HW)	17
3.2.3	ARIMA.....	18
3.3	COVARIÂNCIA E CORRELAÇÃO	19
3.3.1	Covariância.....	20
3.3.2	Correlação	20
4	ESTUDO DE CASO	21
4.1	DESCRIÇÃO DOS RESULTADOS	21
4.2	ESTATÍSTICA DESCRITIVA	23
5	DISCUSSÃO DOS RESULTADOS	26
5.1	PREVISÕES	26
5.2	MÉTRICAS DE ERRO	27
6	CONCLUSÃO	29
	REFERÊNCIAS.....	31

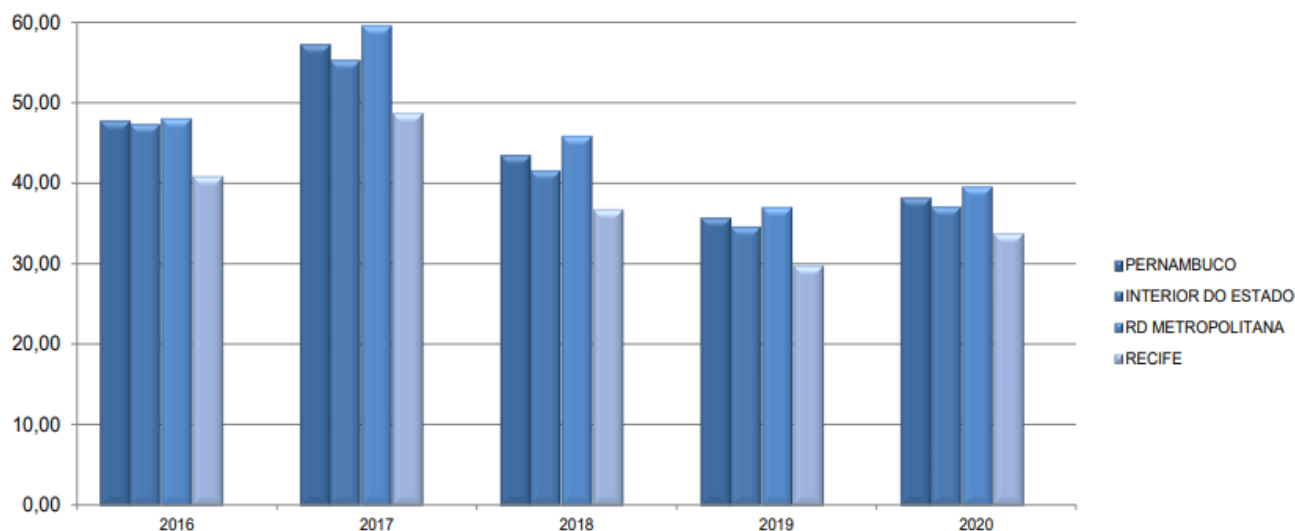
1 INTRODUÇÃO

A criminalidade é uma pauta prioritária dos governos, municipais estaduais, principalmente antes, durante e no final da pandemia, em que muitos padrões e movimentações da população foram alterados. Porém ainda é muito cedo para identificarmos ações de causa e efeito nos índices de criminalidade causados pela pandemia. No entanto, o combate à criminalidade nunca foi fácil de administrar, principalmente em regiões estratégicas economicamente e socialmente, como a região metropolitana, Recife e Olinda.

Segundo o Anuário Brasileiro de Segurança Pública (2020), o Brasil registrou algumas modificações nos padrões de crimes, por exemplo, em relação aos crimes de homicídios, que tem crescido desde o último trimestre de 2019. Os roubos registrados no primeiro semestre de 2020, apresentou redução de 24,2%, o que pode ser explicado pelo fato da população ter se isolado durante a pandemia. Os registros de apreensão de armas parecem não ter sido impactado significativamente pela pandemia.

Em relação à criminalidade no estado de Pernambuco, e em específico a Região Metropolitana de Recife (RMR) é uma importante região do Brasil. Segundo dados do IBGE (2020), a RMR possui uma população de 3,69 milhões de pessoas em quatorze municípios, sendo a quinta maior região metropolitana do Brasil de um total de 27. As Estatísticas da Criminalidade Violenta em Pernambuco de 2016 mostra que no ano de 2015, dentre todas as regiões de Pernambuco, os crimes violentos ao patrimônio cometidos na RMR correspondiam 57.974 de todos os 84.828 cometidos em Pernambuco, ou seja, 68,34%; já em 2016 esse número salta para 77.603 de um total de 115.164 (67,34%); para os crimes violentos letais intencionais, no ano de 2015, a RMR correspondia a 41,96% do total e em 2016 sobe para 42,13%, como observado na figura 1.

Figura 1: Taxa de Crimes Violentos Letais e Intencionais 2016-2020



Fonte: SDS - INFOPOL

Considerando os índices, a acurada estimação deles para evitá-los ou ao menos mitigá-los é interessante e útil para a sociedade. Dessa forma, podem ser utilizados métodos e técnicas preditivas que possibilitam a estimação de ocorrências futuras a partir de dados históricos uma vez que estão disponíveis pelas instituições públicas oficiais. Dentre as vantagens, pode-se destacar o auxílio na tomada de decisão para que sejam feitas medidas que minimizem os crimes contra o patrimônio (ex. o reforço de viaturas em um determinado local para a diminuição dos assaltos). Ainda, também seria útil favorecer a melhoria de práticas operacionais internas e externas garantindo maior efetividade nos serviços relacionados a esta área.

Porém estes dados, não levam em consideração estimações futuras e a adição de fatores como sazonalidade e tendência relacionados.

A previsão de demanda consegue fornecer informações sobre ocorrências futuras, possibilitando que tomadas de decisões sejam mais assertivas, tendo como base informações históricas e que minimizem os erros (Krajewski et al., 2009; Tubino, 2009).

Com a pandemia de COVID-19, o comportamento da população foi alterado, como também as ocorrências criminais dos estados do Brasil. Dessa forma, entender o comportamento, interpretá-lo e projetá-lo no futuro se faz necessário para uma gestão e decisões eficazes.

Por tanto, este trabalho visa o apoio às autoridades no combate à criminalidade, ao buscar prever as ocorrências de Crimes Violentos Letais e Intencionais (CVLI) no estado de

Pernambuco e Apreensão de Armas. O tratamento dos dados, análises, interpretações gráficas e as previsões foram feitos utilizando o software R. Especificamente, o modelo Autorregressivo Integrado de Médias Móveis com Sazonalidade (SARIMA) e Holt-Winters, foi aplicado para identificar, o melhor modelo de previsão, o qual se adequavam a cada série de dados e que apresentavam menores erros. Foi identificado que o modelo SARIMA se adequou melhor a séries Apreensão de Armas e o modelo HW aos homicídios. Após a escolha do melhor modelo para ambas as séries, foi observado brevemente, que ao passo que o a previsão de apreensão de armas aumentava apreciação de homicídios diminuía e vice-versa.

1.1 OBJETIVOS

1.1.1 Geral

Aplicar modelos de previsão para crimes violentos letais e intencionais (CVLI), identificar relações entre tipos de crimes e fornecer modelos preditivos robustos ao contexto da criminalidade no estado de Pernambuco.

1.1.2 Objetivos específicos

- Revisão de literatura;
- Coleta e organização dos dados homicídios e apreensão de armas;
- Decomposição de eventuais tendências e sazonalidade na série histórica de CVLI;
- Aplicação dos modelos de previsão SARIMA e algoritmos ad-hoc de suavização exponencial Holt-Winters;
- Realizar a identificação, estimação e Diagnóstico dos modelos de previsão SARIMA;
- Comparação dos modelos e análise de relações entre os dois dados;

2 REVISÃO DE LITERATURA

Alguns estudos já abordaram a problemática do comportamento dos crimes durante a pandemia do COVID-19. Inclusive foram encontrados trabalhos que fizeram a associação de natureza de crimes diferentes e suas ralações de correlação/covariância.

Um deles foi um trabalho realizado no estado do Ceará, antes e durante o isolamento social da pandemia, durante o ano de 2017 a 2020, com o objetivo de analisar a tendência temporal mensal da mortalidade por homicídios. Foram encontrados um aumento significativo nas taxas de homicídios, onde foi comparado com três anos anteriores, porém não foram levados em conta fatores que pudessem impactar o resultado para além do isolamento social. (VIRGÍLIO et al., 2022)

Outro trabalho avaliou a influência das medidas de distanciamento social no padrão de crimes em Curitiba. Os dados utilizados foram os de crimes de furto, roubo, estelionato, violência doméstica e drogas, dentro de um período de janeiro de 2016 e dezembro de 2020. Os valores foram decompostos e calcularam o valor esperado para cada intervalo, onde foram encontrados valores de déficit de 6,78% e estelionato um superávit de 10,72%, por exemplo, após o decreto de isolamento social. Porém houve algumas discordâncias em relação as análises das séries temporais e as de regressão. Sendo possível perceber que alguns crimes não estão relacionados diretamente às medidas de isolamento, o que faltou incluir esses fatores impactantes nas ocorrências criminais.(LEAL; DE; 2022, [s.d.])

Foram encontrados estudos de séries temporais de crimes no estado de Pernambuco, utilizando modelos clássicos de previsão. Um em específico utilizou o *Autoregressive Moving Average* (ARMA), identificando características como tendência e sazonalidade para os dados de Crimes Violentos Contra o Patrimônio. O estudo foi feito no município do Recife, utilizando séries temporais, onde foi comparado em um intervalo de 10 dias ao longo dos anos de 2019 e 2020. O presente estudo também utilizou o modelo Hol-Winters, o qual conseguiu representar o melhor modelo. Os resultados obtidos mostraram que houve sim uma diminuição nos CVP dentro do período de isolamento social em Recife. (COSTA, 2022).

O estudo encontrado Tourinho Peres et al. (2012) verificou a associação entre duas variáveis de natureza distintas (homicídios e indicadores específicos de segurança) fornecidos pelo estado de São Paulo, correspondente o ano de 1996 e 2008. Utilizaram o software Stata e modelos de regressão binomial. O acesso a armas de fogo, não apresentou indícios de estra diretamente relacionados a mortes por homicídios. Dentre outros resultados, foi encontrado uma forte relação entre as varáveis correspondentes as mortes (homicídios) e as ocorrências de

aprisionamento-encarceramento somado a atividade policial.

O similar trabalho, Lins & Rocha (2012), estudou os CVLI's registrado em Pernambuco, entre janeiro de 2007 a março de 2012, fornecidos pela INFOPOL (sistema de informação da Secretaria de Defesa Social SDS/PE). Utilizando o modelo SARIMA, percebeu-se um ciclo de doze meses, ou seja, apresentando características peculiares em cada mês. O que foi percebido nos resultados, em que se identificou o aumento de ocorrências de crimes nos finais de semana grande número de uso de armas de fogo no estado de Pernambuco, entre outros resultados.

Um outro trabalho, teve como objetivo analisar e relacionar o nível econômico de uma determinada região de São Paulo com a taxa de homicídios, entre os anos de 2002 a 2016. Os dados fornecidos foram da Segurança Pública do Estado de São Paulo, DIEESE e a Fundação Sistema Estadual de Análise Dados (SEADE). Utilizou-se modelos de vetores autorregressivos (VAR) e vetores de correção de erros (VEC). A princípio, resultados obtidos indicaram uma relação entre duas variáveis, sendo elas prisões e homicídios. Ao decorrer do intervalo foi observado outro comportamento de relação a longo prazo entre outras variáveis, tendo agora como relacionadas a taxa de desemprego, prisões e armas de fogo apreendidas. (DUTRA et al. 2017).

Foi possível observar na revisão de literatura as análises das séries temporais e suas respectivas relações e como é necessário a elaboração de trabalhos com essa temática, para o entendimento e fornecimento de informações da criminalidade nas regiões do Brasil.

3 REFERENCIAL TEÓRICO

3.1 SÉRIES TEMPORAIS

Uma série temporal é um conjunto de observações distribuídas ao longo do tempo, que seguem intervalos periódicos, de acordo Brockewll & Davis (2002). Podem ser classificadas como:

- Discretas: conjunto de tempos discretos, intervalo fixo (ex. quantidade vendas ao longo do mês);
- Contínua: são obtidas de forma contínua de acordo com as observações em intervalos de tempo quaisquer (ex. $T = [0,1]$).

Através da análise da série temporal pode ser feito previsões, descrever comportamentos e outras informações atribuídas aos dados, que possibilita definir objetivos e tomar decisões baseado em informações obtidas anteriormente, em um determinado intervalo de tempo, segundo Morettin & Toloi (2006) e Ehlers (2007).

3.1.1 Estacionaridade

Uma série temporal pode ser identificada como estacionária ou não-estacionária. Uma série estacionária é aquela que ao longo do tempo ela é distribuída aleatoriamente em torno de uma média constante, de acordo com Morettin & Toloi, sendo na maioria das vezes não-estacionárias.

3.1.2 Sazonalidade

De acordo com Ehlers (2007), a sazonalidade são comportamentos de pico ou vale periódicos, ao longo de determinado intervalo de tempo. Provocados por uma componente sazonal, podendo ser aditiva (flutuações constantes e independentes) ou multiplicativa (flutuações dependentes).

3.1.3 Tendência

Tendência é um comportamento geral dos dados, em que pode ser verificado visualmente no seu gráfico, um padrão o qual a série pode seguir futuramente, apresentando uma modificação a longo prazo no nível da série (Pal & Prakash, 2017).

3.2 MODELOS CLÁSSICOS DE PREVISÃO

A partir desta seção, serão apresentados os modelos clássicos de previsão, a caracterização do método, situações que devem ser usados e cada um de seus parâmetros.

3.2.1 Modelo de Suavização Exponencial Simples (SES)

O SES é utilizado em séries que não apresentam tendência e sazonalidade, logo a série é do tipo mais simples, onde os valores da série flutuam aleatoriamente em torno de um determinado valor. Segundo Brown (1959, 1963), tendo uma série em que seus valores mantêm constante sobre um nível médio, uma suavização exponencial simples pode ser aplicada para a previsão de ocorrências futuras. Sua representação é dada por:

$$a_t = \alpha x_t + (1 - \alpha)a_{t-1}, \text{ onde } 0 < \alpha < 1 \quad (1)$$

Onde a_t é o nível da série, que é a estimativa da média, o qual é média ponderada do valor atual de x_t e do valor anterior do nível a_{t-1} .

3.2.2 Modelo de Holt-Winters (HW)

O método HW é uma expansão do método SES, onde é considerado o nível da série, a tendência e sazonalidade. Segundo Morettin & Toloi (2006), todas as características da série devem ser levadas em consideração, já que podemos ter uma sazonalidade com componentes multiplicativo e outra com componente aditivo.

- Multiplicativa:

$$a_t = \alpha \left(\frac{x_t}{s_{t-p}} \right) + (1 - \alpha)(a_{t-1} + b_{t-1}) \quad (2)$$

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1} \quad (3)$$

$$s_t = \gamma \left(\frac{x_t}{a_t} \right) + (1 - \gamma)s_{t-p} \quad (4)$$

Onde a_t , b_t e s_t representam respectivamente o nível, inclinação e sazonalidade. Os coeficientes α , β e γ estão relacionados com os coeficientes de suavização; p é o passo sazonal anterior e x_t o valor em questão.

Para a previsão considerando o efeito multiplicativo:

$$X_{t+k} = (a_t + kb_t)s_{t+k-p} \quad (5)$$

- Onde X_{t+k} , é identificada como previsão e k é o intervalo de tempo. Aditiva:

$$a_t = \alpha(x_t - s_{t-p}) + (1 - \alpha)(a_{t-1} + b_{t-1}) \quad (6)$$

$$s_t = \gamma(x_t - a_t) + (1 - \gamma)s_{t-p} \quad (7)$$

Para a previsão considerando o efeito aditivo:

$$X_{t+k} = a_t + kb_t + s_{t+k-p} \quad (8)$$

De acordo com Brockwell & Davis (2002) o modelo escolhido deve ser o que apresentar menor soma dos erros quadrados da previsão após aplicação do modelo.

3.2.3 ARIMA

O ARIMA é um modelo de série temporal amplamente utilizado, o qual foi comparado no passado com outros modelos de séries temporais, para a previsão de consumo de eletricidade da Grécia, e o modelo ARIMA foi o mais eficiente durante esta estimação, segundo Pappas et al. (2008). Este modelo utiliza como base a metodologia Box e Jenkins (1976), que tem como expressão do modelo ARIMA:

$$\varphi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t \quad (9)$$

Onde X_t precisar ser estacionário, caso não seja é necessário diferenciá-lo (EHLERS, 2007).

Esta sigla “ARIMA” representa os componentes do modelo, que são elas:

- AR: “ p ” o número de termos autorregressivos;
- IA: “ d ” a média integrada;
- MA: “ q ” a média móvel da série temporal;

Para a aplicação do modelo ARIMA, antes deve ser analisado e interpretado as funções de autocorrelação (ACF) e autocorrelação parcial (PACF), que serão utilizados para a identificação dos parâmetros “ p ” e “ q ”.

Para a estimação dos parâmetros temos:

- Ajuste por mínimos quadrados (tomando se os resíduos);
- Ajuste por suavização exponencial (tomando os resíduos);
- Diferenciação dos dados nos *lags* adequados;

Vale salientar que a remoção de eventuais tendências e sazonalidade devem ser feitas antes do processo de estimação dos parâmetros. Durante a identificação e aplicação do modelo ARIMA é fundamental que a série seja estacionária, caso contrário a diferenciação é realizada para a remoção da tendência e estabilização da variância (LI; HU, 2012).

Para a estimativa dos parâmetros, utiliza-se o Critério de Informação de Akaike (AIC) (LEMAIC, 2008), o qual mensura a qualidade de um modelo estatístico visando também a sua simplicidade onde é escolhido o que apresenta menor AIC e o Critério de Informação Bayesiano de Schwarz (BIC) (AZADEH; SABERI; SERAJ, 2010), que está diretamente relacionado com o AIC.

Na etapa de diagnóstico, são analisados os erros, os resíduos, precisão do modelo, como também sua aderência à normalidade. Em relação aos erros, serão verificados: o erro quadrático médio (RMSE), erro absoluto médio (MAE) e também será analisado o AIC. Caso o AIC seja igual deve-se identificar o modelo que apresente o menor erro quadrático médio.

3.2.4 SARIMA

O SARIMA foi desenvolvido através da generalização do modelo ARIMA, apresentando uma autocorrelação sazonal. E assim como o ARIMA é utilizado para modelos não estacionários, segundo Box & Jenkins (1976). O qual adiciona no modelo ARIMA um parâmetro sazonal multiplicativo, sendo descrito como SARIMA $(p, d, q)(P, D, Q)$. Representado pela seguinte equação (2.1).

$$\varphi(B)\varphi(B^S)\Delta^d\Delta_S^D X_t = \theta(B)\vartheta(B^S)\varepsilon_t \quad (10)$$

Onde $\Delta_S^D X_t = (1 - B^S)^D X_t$ é a ordem sazonal, e ainda $\varphi(B)$ é o coeficiente sazonal autorregressivo. O $\vartheta(B^S)$ é também o coeficiente sazonal porém das médias móveis (BOX; JENKINS, 1976; MARTINEZ; SILVA, 2011).

3.3 COVARIÂNCIA E CORRELAÇÃO

Para analisarmos a associação entre duas variáveis, segundo Silva (2009), deve-se identificarmos a covariância e correlação entre elas. Ambas as estatísticas estão relacionadas e

forneirão o sinal, o qual pode ser negativo ou positivo; e a intensidade da associação entre as variáveis.

3.3.1 Covariância

Em relação a covariância, está relacionada especificamente com a unidade de medida levando em consideração também o sinal, se identificado que o momento conjunto de primeira ordem da variável X e Y for positivo, elas variam no mesmo sentido, porém, sendo negativo as variáveis possuem uma relação de interdependência inversa, variando em sentidos opostos.

3.3.2 Correlação

A correlação é uma medida adimensional, que varia de +1 (perfeita positiva) e -1 (perfeita negativa), medindo o grau de associação entre X e Y. Caso o coeficiente seja igual a 0 significa que as variáveis não dependem uma da outra. Sendo importante destacar que não podemos concluir que X causa Y ou vice-versa, nem que possuem uma relação de efeito entre elas.

4 ESTUDO DE CASO

Visando o combate à criminalidade e auxílio as políticas públicas contra os crimes, foi elaborado esse estudo de caso.

Este trabalho considerou dados reais fornecidos pela Secretaria de Planejamento e Gestão de Pernambuco (SEPLAG) durante os anos de 2018 a 2021. Toda sua coleta, organização e análise é apresentada nesta seção. Os dados analisados serão as ocorrências de CVLI, especificamente os homicídios, e apreensão de armas.

A partir de análises, foi encontrada a frequência semanal de cada série, sendo elas 17 semanas para a série de homicídios e 19 para a série de apreensão de armas. Para a comparação entre esses dois tipos de dados em um mesmo intervalo de tempo foi atribuído a mesma frequência de 19.

Por fim, aplicou-se as previsões e foi selecionado os melhores modelos para que pudessem ser comparados entre si ao longo de 5 meses, de agosto até dezembro de 2021.

4.1 DESCRIÇÃO DOS DADOS

Os dados CVLI fornecidos estavam separados por dia, porém após a identificação de que em alguns dos dias não apresentavam ocorrências ou ocorrências baixas, foram agrupados em semanas, para que todo o intervalo de tempo permanecesse preenchido e com uma quantidade considerável de ocorrências.

O conjunto de dados CVLI apresenta a ocorrência de crimes registrados no estado de Pernambuco de 2018 até 2021, totalizando 209 semanas. Ao todo são 26 Áreas Integradas de Segurança (AIS) em Pernambuco, porém a região metropolitana corresponde de 1 a 5 sendo elas: Santo Amaro, Espinheiro, Boa viagem, Várzea e Apipucos. Essas regiões foram agrupadas em um único conjunto de dados, visto que são as principais regiões de ocorrências desses crimes.

Especificamente os dados CVLI são divididos nos seguintes tipos de crimes:

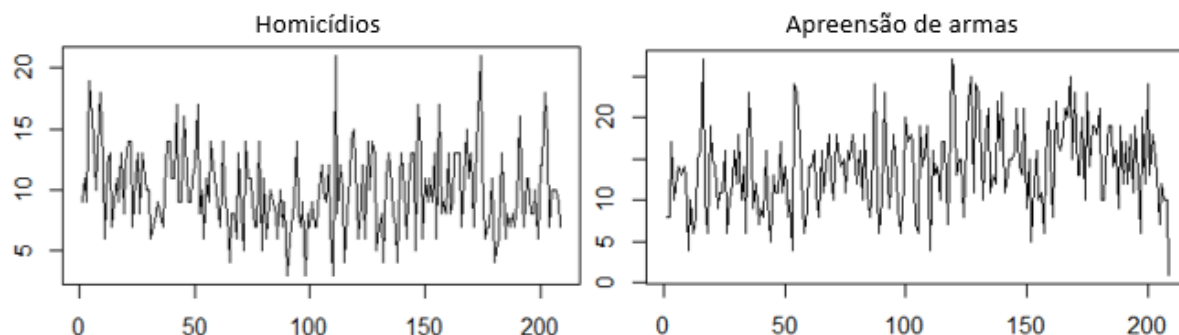
- Homicídios
- Latrocínios
- Lesões
- Feminicídios

O conjunto de dados homicídios e apreensão de armas foram escolhidos para análise pelo fato de possuírem um maior número de ocorrências e possuírem em todas as semanas dados. Vale ressaltar que, a seleção desses dois dados gerou uma análise de comparação em

que os dois podem ou não estar relacionados diretamente, o que vai ser discutido mais adiante.

A princípio foi visualizado as duas séries temporais e observado se possuem alguma tendência ou sazonalidade apenas na análise do gráfico.

Figura 2: Série histórica de 2018 à 2021, distribuídos em semanas

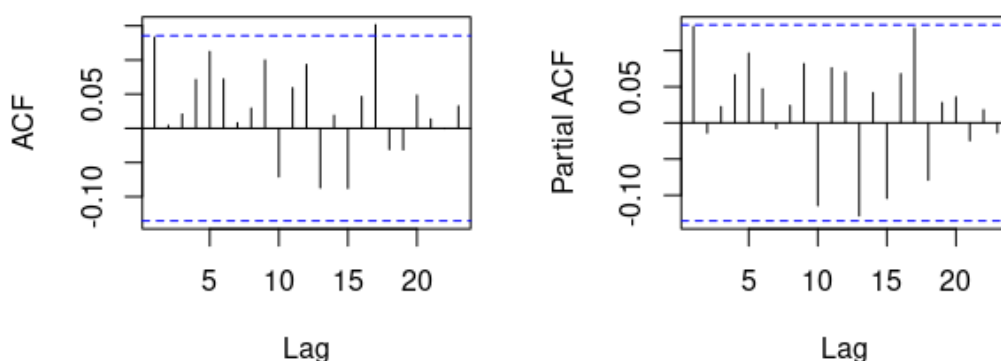


Fonte: autor (2023)

A princípio, analisando a figura 2, os dados não apresentaram sinais claros de comportamentos de tendência nem de sazonalidade. Porém é possível perceber que a série de homicídios possui um pico que pode representar um outlier e que pode considerar alguma informação relevante. Já a série de apreensão de armas possui um pouco mais de centralização em torno da média.

Desse modo, utilizou-se o ACF e o PACF para identificação dos parâmetros de tendência e sazonalidade.

Figura 3: ACF e PACF da série homicídios

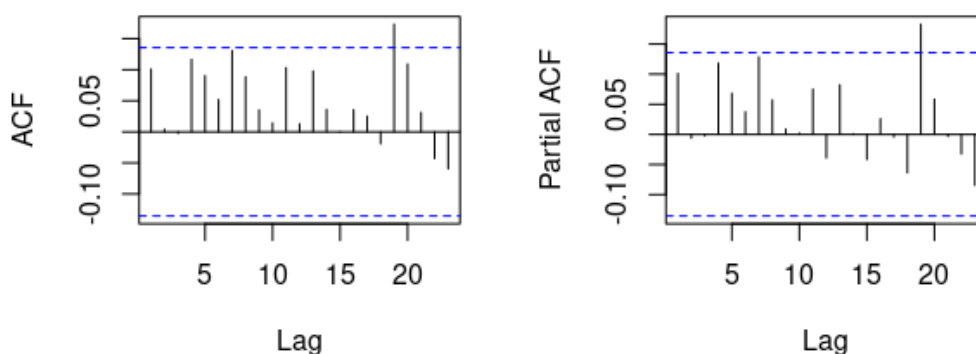


Fonte: autor (2023)

De acordo com a observação da figura 3, é possível observar em relação a série homicídios que o lag 17 ultrapassou o intervalo de confiança na função de autocorrelação

parcial chegou próximo do limite, o que é um indício de que essa série possui uma frequência igual a 17. Parâmetro que vai ser levado em consideração para a construção do modelo de previsão HW e o modelo SARIMA. Vale ressaltar que, pelo fato de estarmos trabalhando com tendência e sazonalidade foi utilizado o modelo SARIMA e não o ARIMA, pelo fato deste último não levar consideração a sazonalidade.

Figura 4: ACF e PACF da série apreensão de armas



Fonte: autor (2023)

A série apreensão de armas apresentou o lag 19 significativo, próximo do valor encontrado na série homicídios, observado na figura 4.

Desse modo, para fins de comparação, foi atribuído aos modelos das duas séries a frequência 19, para que possam ser comparados as previsões em um mesmo intervalo de tempo tendo como parâmetro a mesma frequência.

4.2 ESTATÍSTICA DESCRITIVA

Nesta seção, será abordado as principais característica estatísticas em relação as duas séries dos dados. Como também aspectos singulares que serão levadas em consideração na elaboração das previsões

Em relação as ocorrências no intervalo de 209 semanas, foi elaborado um resumo para determinar os extremos e médias.

Tabela 1: Resumo dos dados

	Min.	1° Quartil	Mediana	Média	3° Quartil	Max.
Homicídios	3	7	10	9,967	12	21
Apreensão de Armas	1	10	14	14	17	27

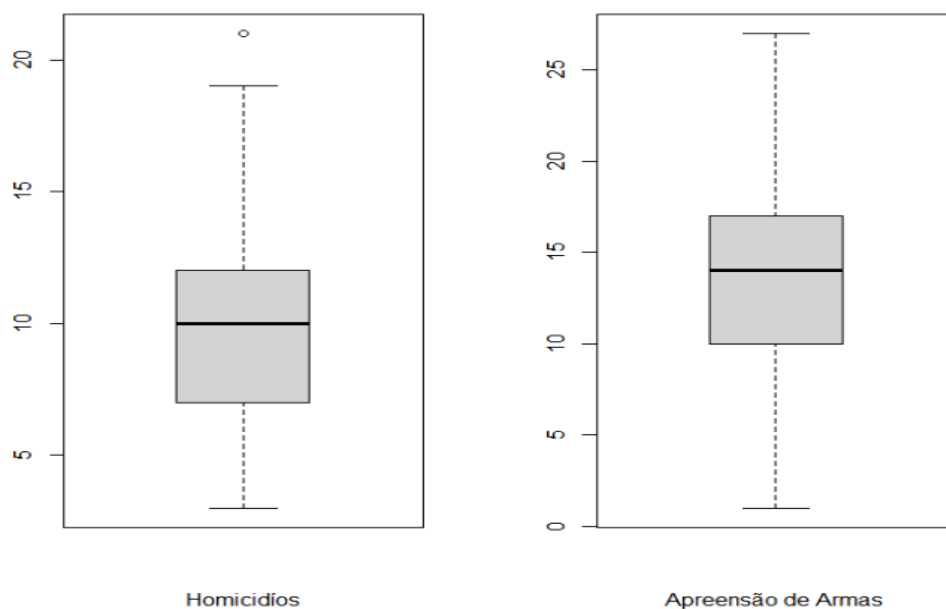
Fonte: autor (2023)

De acordo com a Tabela 1 que registrou um resumo das ocorrências semanais, é observado, por exemplo, que a série homicídios possui extremos (mínimo: 3 e máximo: 21) mais próximos da média 10, comparado com a série apreensão de armas (mínimo: 1 e máximo: 27) com média 14, logo possuindo ocorrências ou muito baixas ou muito elevadas.

Para fins de análise de identificação se a os dois dados são séries distintas, foi aplicado o teste de Kolmogorov-Smirnov. Apresentando um p-valor muito menor que 0,05 (intervalo de confiança de 5%), sendo possível afirmar que são séries diferentes, mesmo tendo valores de ocorrências parecidos e médias aparentemente próximas, com uma diferença de apenas 4 ocorrências de uma série para outra.

O boxplot relacionados aos dados homicídios foi possível encontrar um “outlier” acima do primeiro quartil.

Figura 5: Boxplot dos dados



Fonte: autor (2023)

O outliers encontrado no primeiro boxplot, é justamente o pico encontrado no gráfico da série histórica de homicídios, o qual foi considerado para análise.

Para análise a da interdependência entre as duas séries de dados foi calculado na Tabela 2 a relação conjunta e a intensidade entre ambas.

Tabela 2: Relação dos dados de covariância e correlação

	Apreensão de Armas / Homicídios
Correlação	-0,06977
Covariância	-0,26650

Fonte: autor (2023)

Em relação a intensidade da correlação identificou-se uma força fraca, aproximadamente zero, o que significa que o nível de associação entre as duas séries aparentemente não existe. Porém, a covariância registrou um valor de -0,2665, um valor pequeno, mas diferente de zero e que por apresentar um valor negativo, mostra uma pequena relação negativa de interdependência, ou seja, que os dados variam em sentidos opostos, quando uma ocorrência aumenta outra diminui e vice-versa.

5 DISCUSSÃO DOS RESULTADOS

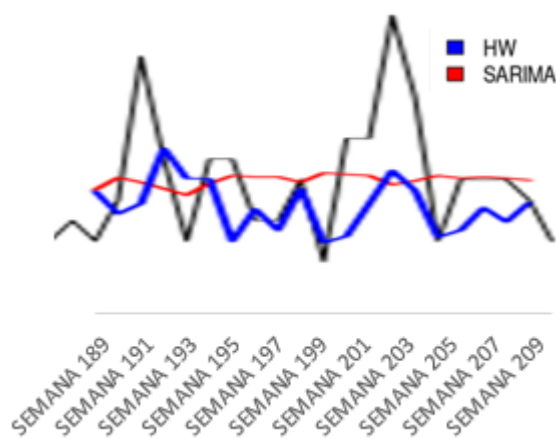
Nesta seção será feita a aplicação do método SARIMA e Holt-Winters, a comparação com os ajustes e análises de relações entre as previsões de homicídios e apreensão de armas.

Vale salientar que, através de análises dos dados foi observado que, agrupando as ocorrências em semanas e não em dias como foi fornecido, tornou-se mais efetiva as previsões, visto que foi eliminado as ocorrências iguais a zero.

5.1 PREVISÕES

A Figura 5 apresenta a previsão para o conjunto de dados “homicídios”. Vale ressaltar que a série apresentou uma frequência de 17 semanas e a série apreensão de armas uma frequência de 19 semanas. Dessa forma, foi aplicado em todas as séries uma mesma frequência, igual a 19, para que pudessem ser comparadas os seus resultados. A previsão realizada foi para o mesmo período do conjunto de teste, que corresponde de agosto de 2021 até dezembro do mesmo ano.

Figura 6: Previsão da série homicídios em relação a série real



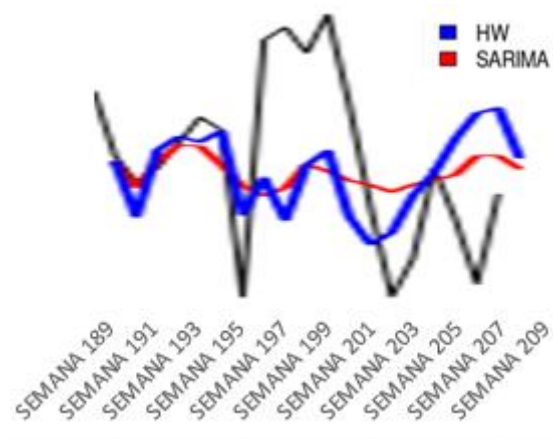
Fonte: autor (2023)

De forma visual, é possível inferir que o modelo de HW se adequa mais a série real, uma vez que, seu ajuste se aproximou consideravelmente do comportamento da série de treinamento. A previsão acompanhou o comportamento de crescimento e decrescimento, apresentando picos e vales aderentes à série real. No entanto, apresentou alguns erros, como é possível identificar no gráfico, onde a previsão apontou uma ocorrência baixa em relação a série real.

O SARIMA mostrou um comportamento constante ao longo de todo o intervalo de teste, ao passo que apresentou valores com erros associados maiores que a previsão HW,

valores que serão justificados nas métricas de erro. Em alguns momentos a previsão se mostrou antagônica e por tanto, não se adequava a série original.

Figura 7: Previsão da série apreensão de armas em relação a série real



Fonte: autor (2023)

Na Figura 6 os dois modelos nos seis primeiros pontos se adequam perfeitamente a série original. Porém ao passo que as semanas vão avançando o modelo HW oscila de forma mais extrema que o modelo SARIMA, errando ainda mais a série real, o que pode ser verificado nas métricas de erro. Ao passo que se adequava muito bem a série original em determinados pontos a diferença entre a série original era muito grande, ou seja, do mesmo modo que a previsão HW se ajustou muito bem os erros também eram elevados.

A previsão SARIMA foi mais assertiva, mesmo observando algumas diferenças visíveis na Figura 6, porém bem menores que a previsão HW. O comportamento em comparação a série original foi respeitado, principalmente no aumento ou diminuição das ocorrências.

5.2 MÉTRICAS DE ERRO

Em relação à precisão das previsões foram calculados o erro quadrático médio e erro médio absoluto separadas por modelo e métrica utilizada. Além dessa métricas foi calculado também o teste Ljung-Box, com o objetivo de analisar a hipótese nula (se os erros não são correlacionados) identificando a previsão está aderente ou não à serie real.

Sendo importante destacar que, os valores em **negrito** apresentam os melhores resultados, logo os de menor erro em relação a acurácia das previsões.

Tabela 3: Métricas de erro das séries apreensão de armas e homicídios

Métricas de erro	HW			SARIMA		
	Erro quadrático médio	Erro médio absoluto	Ljung-Box (p-valor)	Erro quadrático médio	Erro médio absoluto	Ljung-Box (p-valor)
Homicídios	3,8679	3,0122	0,9021	6,0071	4,863	0,5626
Apreensão de Armas	7,0258	5,4902	0,0006	5,9655	4,8455	0,8732

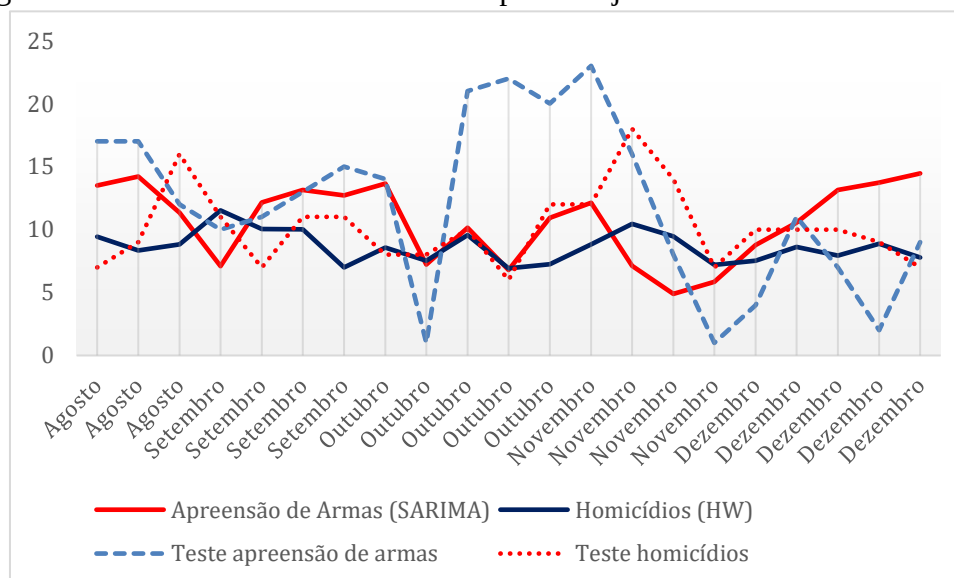
Fonte: autor (2023)

O modelo de HW foi mais preciso para os dados de homicídios. O erro quadrático médio foi aproximadamente duas vezes menor, o erro médio absoluto foi consideravelmente menor em comparação a previsão SARIMA e o p-valor superior. O que indica que o fato de o teste Ljung-Box apresentar um valor superior a 0,05 mostra que os erros não estão relacionados, deixando de rejeitar a hipótese nula, ou seja, possui melhor aderência.

Para a série apreensão de armas o melhor modelo foi o SARIMA. Em comparação com as métricas do modelo HW o erro quadrático médio e o erro médio absoluto se aproximaram, porém o SARIMA ainda foi inferior. O que pode também ser identificado que o p-valor do HW foi de 0,0006, o que significa dizer que o modelo rejeitou a hipótese nula, possuindo erros correlacionados, inviabilizando a utilização do modelo.

Considerando essas duas séries e observando que pode existir ou não uma relação de correlação entre elas, foi utilizado os melhores modelos e suas respectivas janelas de teste em um mesmo gráfico para que seja analisado e verificado se possuem alguma relação ao decorrer das semanas.

Figura 8: Previsões HW e SARIMA e respectivas janelas de teste no ano de 2021



Fonte: autor (2023)

De acordo com o gráfico e a Figura 7, é possível observar o comportamento entre as janelas de teste e suas respectivas previsões. Em relação as previsões, podemos perceber que dentro do intervalo de teste, é observada oscilações inversas ao passo que a previsão de homicídios aumenta a de apreensão de armas diminui e vice-versa. O que significa que mesmo possuindo uma covariância mínima, porém diferente zero e negativa, existe uma interdependência mínima de sentidos opostos a medida que as semanas vão avançando. O que pode ser explicado pela ideia de que quanto mais são apreendidas as armas pode influenciar negativamente as ocorrências de homicídios.

6 CONCLUSÃO

O presente trabalho teve como objetivo, através dos dados históricos a aplicação do modelo SARIMA e Holt-Winters e a verificação de relações de associação entre dois indicadores de criminalidade, em uma região que possui grande importância nacional.

Em relação aos objetivos, foi aplicado os modelos de previsão nos dados CVLI e a identificação de relações entre eles, através de modelos preditivos robustos, inserindo fatores que impactaram os dados no período da pandemia. Através da revisão da literatura foram estudados trabalhos que traziam a problemática de indicadores criminais durante a pandemia nos estados do Brasil, e verificado a pequena quantidade de trabalhos ainda publicados nesse contexto.

As análises conseguidas neste trabalho, trazem uma pequena parcela do que ainda pode ser feito dentro deste contexto de pandemia, porém mostrou inicialmente um comportamento de relação entre essas variáveis de homicídios e apreensão de armas, através dos modelos gerados e interpretados.

O SARIMA foi o melhor modelo para os dados de apreensão de armas, apresentando uma precisão um pouco melhor que o modelo HW da mesma série, porém o comportamento inverso foi visto nos dados homicídios. Importante destacar, que em relação a comparação dos modelos na série apreensão de armas, o modelo apresentou um forte indício de possuir erros nos dados da previsão correlacionados, o que viabilizou o uso do modelo, sendo o SARIMA o escolhido como melhor.

Os modelos utilizados apresentaram erros, o que gerou limitações nos resultados. Outro ponto limitante foi o intervalo em que as ocorrências foram registrados, pois um intervalo maior poderia fornecer interpretações e análises mais precisas. Estudos futuros podem explorar a

comparação entre modelos clássicos de previsão, modelos de *machine learning* e outros modelos, a fim de encontrar o máximo de acurácia possível.

Ainda, relacionado aos modelos de previsão, também podem ser feitos trabalhos que estudam as relações de interdependência entre variáveis inseridas em um mesmo contexto, ou mesmo intervalo de tempo. Como também a identificação se determinados comportamentos se mantiveram após a pandemia, ou se houve alguma alteração e o porquê.

Dessa forma, podendo levar a explicações de determinados comportamentos de dados, de decrescimento ou aumento de ocorrências específicas, inversamente ou diretamente proporcional, considerando um evento de pandemia.

REFERÊNCIAS

- AZADEH, A.; SABERI, M.; SERAJ, O. An integrated fuzzy regression algorithm for energy consumption estimation with non-stationary data: a case study of Iran. **Energy**, v. 35, n. 6, p. 2351-2366, 2010.
- BOX, George E. P. e JENKINS, Gwilym. **Time Series Analysis: Forecasting and Control**. Oakland, Holden-Day, 1976.
- COSTA, R. DA. Um estudo sobre o roubo à propriedade no estado de Pernambuco no contexto da Pandemia e Quarentena de 2020. 2022.
- CONDEPE/ FIDEM. Estatísticas da Criminalidade Violenta em Pernambuco 2020. **Agência Estadual de Planejamento e Pesquisas de Pernambuco – CONDEPE/ FIDEM**, 2020. Disponível em: https://www.sds.pe.gov.br/images/media/1637086534_ESTATSTICAS%20DA%20CRIMINALIDADE%20VIOLENTA%20EM%20PERNAMBUCO%202020.pdf. Acesso em: 10, maio de 2022.
- COSTA, R. DA. Um estudo sobre o roubo à propriedade no estado de Pernambuco no contexto da Pandemia e Quarentena de 2020. 2022.
- KRAJEWSKI, L. J., RITZMAN, L. P., & MALHOTRA, M. K. **Administração de Produção e Operações** (L. B. R. Fernandes & M. S. R. de Oliveira (eds.); 8ª ed. Pearson Education do Brasil, 2009.
- LEAL, M.; DE, L. G.-J.-A. DO X. E. R. DE B. Impacto da Pandemia da COVID-19 nos Padrões de Crimes no Município de Curitiba. **sol.sbc.org.br**, [s.d.]. VIRGÍLIO, F. et al. Tendência temporal dos homicídios no Ceará, antes e durante a pandemia de Covid-19. **unichristus.homologacao.emnuvens FVA de Oliveira, ML de Oliveira Santiago, RA de Lima Nunes, MKM Beserra, FTC Sena...Journal of Health & Biological Sciences, 2022•unichristus.homologacao.emnuvens ...**, v. 10, n. 1, p. 1–6, 2022.
- LEMAIC, Mile. **Markov-Chain-Based Heuristics for the Feedback Vertex Set Problem for Digraphs**. 2008. Tese de Doutorado. Universität zu Köln.
- LI, CHUNSHIEN; HU, JHAO-WUN. A new ARIMA-based neuro-fuzzy approach and swarm intelligence for time series forecasting. **Engineering Applications of Artificial Intelligence**, v. 25, n. 2, p. 295-308, 2012.
- MARTINEZ, E. Z.; SILVA, E. A. S. DA. **Predicting the number of cases of dengue infection in Ribeirão Preto, São Paulo Sate, Brazil, using a SARIMA model**. Cadernos de Saúde Pública, v. 27, n. 9, p. 1809-1818, set. 2011.
- PAPPAS, S. Sp et al. Electricity demand loads modeling using AutoRegressive Moving Average (ARMA) models. **Energy**, v. 33, n. 9, p. 1353-1360, 2008.
- TOURINHO PERES, M. F., FELICIANO DE ALMEIDA, J., VICENTIN, D., RUOTTI, C., BATISTA NERY, M., CERDA, M., CARDIA, N. & ADORNO, S. 'Evolução dos homicídios e indicadores de segurança pública no município de São Paulo entre 1996 a 2008:

um estudo ecológico de séries temporais', *Ciência & Saúde Coletiva* n 17, v 12, 2012.

TUBINO, D. F. **Planejamento e Controle da Produção: Teoria e Prática**, 2ª ed., Atlas, 2009.