

Técnicas para predição de crimes utilizando dados oficiais considerando tempo e espaço

Heytor Norberth Leite da Silva¹, Saul Sousa da Rocha², Glauber Dias Gonçalves²

¹Universidade Federal do Piauí (UFPI) - CSHNB

²Universidade Federal do Piauí (UFPI)
Programa de Pós-Graduação em Ciência da Computação

heytonorberth2@gmail.com¹, {saul.rocha2001, ggoncalves}@ufpi.edu.br²

Abstract. *This paper proposes the use of official data to predict crime rates in metropolitan areas, with the aim of providing support for more effective public policies. To this end, a dataset was compiled that covers specific characteristics of urban regions, correlating them with official crime rates over time. The methodology adopted included the application of regression models and time series, which were evaluated for their ability to identify seasonal patterns and trends in crime occurrence. The results obtained demonstrate the effectiveness of these models in predicting crime rates, highlighting the importance of analytical tools in proactive decision-making by authorities and communities, aiming at preventing and combating crime.*

Resumo. *Este trabalho propõe a utilização de dados oficiais para prever as taxas de criminalidade em áreas metropolitanas, com o objetivo de fornecer subsídios para políticas públicas mais eficazes. Para isso, foi compilado um conjunto de dados que abrange características específicas de regiões urbanas, correlacionando-as com as taxas oficiais de criminalidade ao longo do tempo. A metodologia adotada incluiu a aplicação de modelos de regressão e séries temporais, que foram avaliados quanto à sua capacidade de identificar padrões sazonais e tendências na ocorrência de crimes. Os resultados obtidos evidenciam a eficácia desses modelos na previsão das taxas de criminalidade, ressaltando a importância de ferramentas analíticas na tomada de decisões proativas por parte de autoridades e comunidades, visando a prevenção de crimes.*

1. Introdução

Redução e controle de taxas de criminalidade são desafios enfrentados nos grandes centros urbanos do mundo. Esta é uma questão ainda mais grave nos países em desenvolvimento. No Brasil, por exemplo, foram registradas 40.464 mortes por crimes violentos em 2023, contra 42.190 em 2022, 1.726 mortes a menos. A taxa de mortes decorrentes de crimes violentos por 100 mil habitantes varia de 1,41 a 1,69 no cenário nacional, enquanto nos estados do Nordeste essa taxa chega a mais de 3 mortes.

Fontes alternativas de dados que retratam características do espaço urbano e de seus habitantes podem contribuir para a análise e compreensão de problemas das cidades, como taxas de criminalidade. Dados coletados de fontes confiáveis, i.e., a Secretarias de Segurança Pública ou Instituto Brasileiro de Geografia e Estatística (IBGE), podem

oferecer visões detalhadas dos crimes ocorridos ao longo do tempo e em diferentes regiões da cidade. Esses dados permitem a aplicação de técnicas avançadas de aprendizado de máquina para modelar, prever e entender padrões de criminalidade, fornecendo subsídios para a formulação de políticas públicas mais eficazes.

Nesta pesquisa, utilizamos dados temporais e espaciais de crimes registrados em São Paulo para desenvolver modelos preditivos que identificam tendências e sazonalidades. A análise de dados geoespaciais combinada com algoritmos de aprendizado de máquina pode revelar insights sobre a distribuição espacial dos crimes, além de identificar áreas de maior risco e momentos de maior incidência criminosa. Trabalhos anteriores já demonstraram a eficácia de tais abordagens em diferentes contextos urbanos [Chainey and Ratcliffe 2008], [Wang and Xu 2013]. Adicionalmente, a integração de análises temporais permite entender melhor como fatores sazonais e eventos específicos influenciam as taxas de criminalidade. Modelos de séries temporais, como ARIMA e SARIMA, são ferramentas essenciais para capturar essas dinâmicas e prever flutuações futuras [Gorr et al. 1994, Kamarianakis and Prastacos 2005].

Os resultados alcançados mostram que o modelo Arima com os ajustes e eficiência aumentada por meio do uso do método de box and jenkins, com melhor desempenho em comparação aos demais modelos para previsão de crimes na cidade de São Paulo considerando dados oficiais de 2022 e 2023, levando em conta a granularidade de semanas.

Em suma, esse artigo traz as seguintes contribuições:

- Aprimoramento de estratégias de segurança pública explorando modelos de aprendizagem de máquina e regressão para prever taxas de criminalidade.
- Avaliação de desempenho dos modelos considerando atributos de tempo e espaço extraídos de fontes de dados oficiais.

As próximas seções desse artigo estão com a seguinte organização. Na Seção 2, discutimos trabalhos relacionados. A descrição detalhada sobre a coleta e o processamento de dados para esse trabalho estão na Seção 3. Nossas análises e resultados são discutidos na Seção 4 ao passo que nossas considerações finais são apresentadas na Seção 5.

2. Trabalhos Relacionados

Nesta seção discutimos os trabalhos relacionados organizados em duas partes. Primeiramente, discutimos os trabalhos que utilizaram características de espaço como objetivo principal para predição de crime. A seguir, são discutidos os trabalhos que utilizam características de tempo para atingir o objetivo em questão.

2.1. Trabalhos que utilizam características de espaço

No trabalho apresentado por [Sathyadevan and Devan 2014] é descrito um sistema capaz de prever regiões que possuem maior probabilidade de ocorrer um crime. Os autores utilizaram algoritmo de Naive Bayes para realizar o treinamento com dados relacionados a vandalismo, assassinato, roubo, abuso sexual etc. Os testes realizados mostraram que a utilização de Naive Bayes apresentou 90% de acurácia. Para obter melhores resultados na previsão, idealizam encontrar mais atributos criminais dos locais, em vez de fixar certos atributos. Dado que o tempo é um fator importante no crime, eles têm a ciência de que têm que prever não só as regiões propensas ao crime, mas também o momento adequado.

A pesquisa realizada em [SANTOS 2020] utiliza mapas de criminalidade criados a partir de um Sistema de Informação Geográfica (SIG), surgem como uma ferramenta muito importante no auxílio do combate à violência, servindo de base para visualização de áreas problemáticas, análise espacial e de rede, controle e prevenção do aumento da criminalidade urbana. O objetivo do trabalho é que através desse estudo possa se mostrar a importância da aplicação do (SIG) aos gestores, fazendo com que os mesmos enxerguem o benefício do trabalho executado pela estatística e planejamento para realizar a redução dos números de crimes.

O trabalho descrito em [Ingilevich and Ivanov 2018] apresenta características do ambiente que podem contribuir para o aumento de crimes. Os autores discutem que tamanho da população, estações de polícia e até escolas podem estar diretamente relacionadas com o aumento ou diminuição do número de crimes em uma determinada área, bairro etc. Os autores obtiveram uma melhor acurácia no trabalho utilizando modelos de boosting, tais resultados podem ajudar a polícia a encontrar os melhores lugares para colocar uma estação policial e assim reduzir a taxa de criminalidade.

2.2. Trabalhos que utilizam características de tempo

Em anos recentes, no trabalho [Nery et al. 2019] foi analisado a trajetória de crimes na cidade de São Paulo ao longo dos anos e a sua distribuição geográfica, contrastando teorias que estipulam de forma estática bairros violentos e não-violentos, regiões centrais e periferia. Em [Weisburd et al. 2012] realizou-se um trabalho baseado no historial de 16 anos de crimes nas cidades de Seattle e Washington nos EUA, se tornando uma referência importante sobre afunco de pesquisa em criminologia com análises de crimes por regiões. A partir daí a área vem progredindo e conteúdos gerados por usuários (UGC) por meio de serviços de localização com base em redes sociais é um tipo de dado que favoreceu para estabilizar análises de crimes em pequena granularidade por regiões da cidade [Silva et al. 2019]. No artigo [Araújo et al.] é introduzido um framework que ajuda a melhorar o planejamento de patrulha já que o mesmo pode fornecer precisamente lugares e horários que são mais propícios a serem perigosos. O *framework* oferece a possibilidade de gerar novas predições a cada novo intervalo de tempo, dessa forma novos lugares são preditos.

3. Bases de Dados e Metodologia

Nesta seção, descrevemos as bases de dados utilizadas e detalhamos a metodologia empregada para o processamento desses dados e sua aplicação em modelos de predição. A estrutura foi organizada em três tópicos principais: Base de Dados, Métodos e Técnicas, e Metodologia.

3.1. Base de Dados

Os dados de índices criminais foram extraídos da Secretaria de Segurança Pública do Estado de São Paulo [SSP-SP 2022]. Esses dados são divulgados mensalmente e organizados em vinte e três categorias de crimes, contendo a contagem de ocorrências registradas por regiões do estado desde 2001. A área geográfica de cada região é delimitada pelos setores censitários definidos pelo [IBGE]. Um setor censitário é a unidade territorial estabelecida para fins de controle cadastral, formado por uma área contínua, situada em um único quadro urbano ou rural, com dimensão e número de domicílios que permitam o

levantamento. A Figura 1 ilustra um setor censitário na cidade de São Paulo conforme definido pelo IBGE.

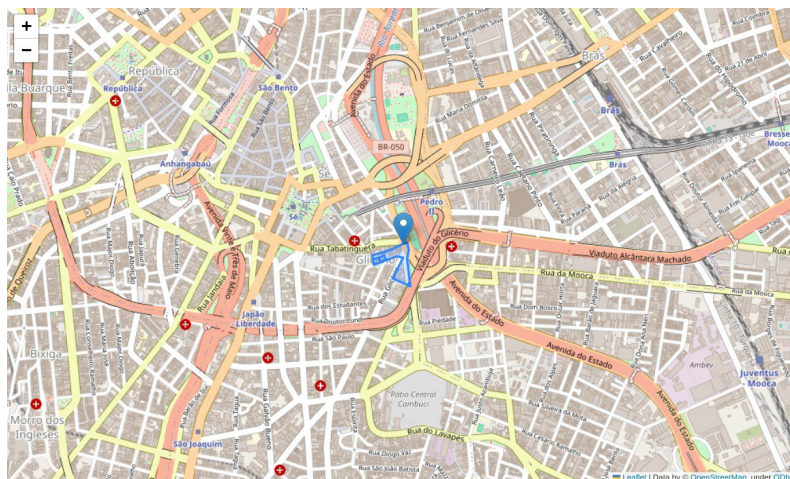


Figura 1. Exemplo de setor censitário em São Paulo [IBGE].

O foco de nosso estudo foi nas top-100 e top-1000 regiões mais violentas da cidade de São Paulo, a capital do estado, por se tratar de regiões com os maiores índices criminais. Conforme a [SSP-SP 2023], os furtos em geral na cidade passaram de 235 mil casos para 250 mil, uma alta de 6,53% entre 2022 e 2023. Já os roubos seguiram na direção contrária, como aconteceu no estado, passando de 139 mil registros em 2022 para 130 mil em 2023. A Figura 2 mostra um mapa de calor com os setores censitários com a quantidade de crimes contra patrimônio.

3.2. Métodos e Técnicas

Os métodos e técnicas utilizados para análise e modelagem dos dados criminais incluíram a aplicação de modelos de regressão e modelos de séries temporais. A seguir, detalhamos os métodos de regressão utilizados, bem como a metodologia para o desenvolvimento dos modelos preditivos.

Os modelos de regressão usados foram o RF (Random Forest) e o SVR (Support Vector Regression). O RF é um modelo de aprendizado de máquina que combina múltiplas árvores de decisão para criar um modelo robusto e geralmente melhorar a precisão preditiva. Cada árvore na floresta é treinada de forma independente em uma amostra aleatória dos dados, e suas previsões são combinadas por meio de votação ou média. O modelo é conhecido por sua capacidade de lidar com dados complexos e grandes conjuntos de dados, além de reduzir o overfitting [Breiman 2001].

O SVR é uma técnica de regressão baseada em máquinas de vetores de suporte, adaptada da SVM (Support Vector Machine) para problemas de regressão. O objetivo do SVR é encontrar uma função que se ajuste aos dados, mantendo as violações de margem (erro) dentro de uma tolerância especificada. Ele é eficaz em espaços de alta dimensionalidade e pode lidar com dados não-lineares usando kernels [Drucker et al. 1997].

Os modelos ARIMA (AutoRegressive Integrated Moving Average) são amplamente utilizados na análise de séries temporais para entender e prever pontos futuros de

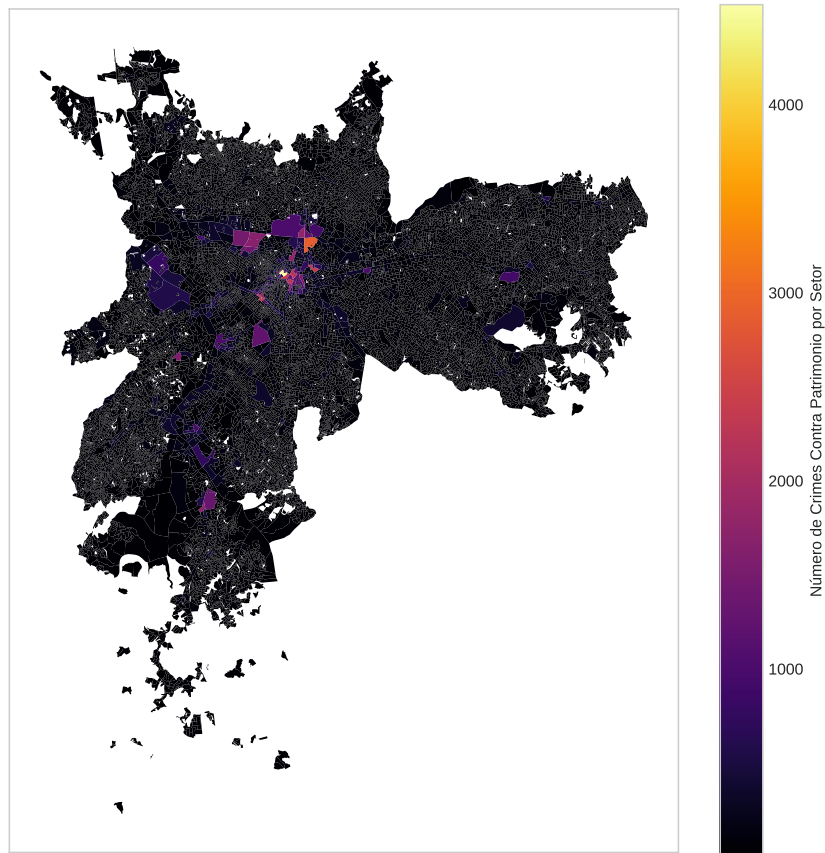


Figura 2. Mapa de calor dos crimes contra o patrimônio por setor.

uma série de dados. Esses modelos são uma generalização dos modelos ARMA (Auto-Regressive Moving Average), que combinam duas abordagens principais: auto-regressão (AR) e médias móveis (MA) [Box and Jenkins 1970]. A parte auto-regressiva (AR) de um modelo ARIMA indica que a variável de interesse é regredida em seus próprios valores defasados. Isso significa que os valores anteriores da série são usados para prever os valores futuros. A parte de média móvel (MA) sugere que o erro de previsão é uma combinação linear de erros passados. A parte integrada (I) do modelo ARIMA refere-se ao número de vezes que os dados são diferenciados para alcançar a estacionariedade [Chatfield 2003].

3.3. Metodologia

Para o treinamento dos modelos RF e SVR, uma série temporal foi gerada considerando um período de observação de 4 semanas. Para prever se um crime ocorrerá em uma data específica, utilizamos as 4 semanas anteriores, complementadas pelas semanas correspondentes do ano anterior. Esta abordagem permite que o modelo aprenda não apenas com as tendências recentes, mas também considere padrões sazonais.

Neste estudo, desenvolvemos dois modelos preditivos distintos para avaliar a ocorrência de crimes em datas específicas, conforme descrito na Equação 1.

Seja Y_t o rótulo (indicando a ocorrência de um crime em uma região) na data t .

- **Rótulo:** Y_t (ocorrência de um crime em uma região na semana t).
- **Atributos do Ano Corrente:** Atributos coletados nas 4 semanas anteriores à semana t do ano corrente, denotados por $X_{t-4:t-1}$.
- **Atributos do Ano Anterior:** Atributos coletados nas mesmas 4 semanas do ano anterior, denotados por $X_{t-56-4:t-56-1}$.

A entrada do modelo para prever o rótulo Y_t é:

$$Y_t = [X_{t-4:t-1}, X_{t-56-4:t-56-1}] \quad (1)$$

Já os modelos ARIMA são denotados como ARIMA(p,d,q), onde p é a ordem do modelo auto-regressivo, d é o grau de diferenciação, e q é a ordem do modelo de média móvel. Para séries temporais que exibem sazonalidade, os modelos são estendidos para ARIMA(p,d,q)(P,D,Q)m, onde m é o número de períodos em cada temporada e P, D, e Q são os termos sazonais de auto-regressão, diferenciação e média móvel, respectivamente [Hamilton 1994].

A estimativa e a identificação de modelos ARIMA são frequentemente realizadas usando a metodologia de Box-Jenkins, proposta por [Box and Jenkins 1970]. Segundo [Shumway and Stoffer 2017], este método envolve um processo iterativo de modelagem que inclui a identificação do modelo, a estimação dos parâmetros e a verificação do modelo. A metodologia de Box-Jenkins é conhecida por sua rigorosidade e pela capacidade de produzir modelos precisos e robustos para previsão de séries temporais.

- **Identificação:** Nesta fase, os dados são analisados para determinar se são estacionários e para identificar a estrutura apropriada do modelo (valores de p, d, q, P, D, Q e m). Ferramentas como o gráfico de autocorrelação (ACF) e o gráfico de autocorrelação parcial (PACF) são usadas para esta finalidade.
- **Estimação:** Uma vez identificado o modelo, os parâmetros são estimados usando métodos como máxima verossimilhança ou mínimos quadrados.
- **Verificação:** Após a estimação, o modelo é verificado para garantir que ele se ajusta bem aos dados. Isso envolve a análise dos resíduos (erros de previsão) para verificar se são ruído branco, ou seja, se não exibem autocorrelação significativa.

4. Resultados Experimentais

Nesta seção apresentamos nossos resultados sobre as bases de dados de crimes, características de espaço e tempo apresentadas na seção anterior. Primeiramente, descrevemos as configurações utilizadas nos experimentos, considerando métodos de regressão e de séries temporais, também as métricas de desempenho para predição. A seguir, discutimos os resultados alcançados em termos de desempenho de diferentes métodos e importância dos atributos utilizados na predição.

4.1. Métricas de avaliação

O nosso objetivo nesses experimentos é investigar o potencial de predição usando as características de tempo e espaço utilizando dados reais, onde podemos ter um entendimento melhor do que acontece em determinadas regiões da cidade, para explicar taxas de ocorrências de crimes por categoria e por região da cidade.

Todas as predições para cada categoria crime foram utilizadas para avaliar o desempenho do modelo. O desempenho dos modelos de predição foi avaliado através das métricas média do erro absoluto (MAE), média do erro relativo (MRE). Especificamente essas métricas foram calculadas da seguinte forma:

$$MAE = \frac{\sum_i |y_i - \hat{y}_i|}{n}; MSE = \frac{\sum_i (y_i - \hat{y}_i)^2}{n}, \quad (2)$$

onde y_i e \hat{y}_i representam os valores reais e preditos para a taxa de crime na i -ésima região, ao passo que \bar{y} representa a média da taxa de crime considerando todas as n regiões para a construção do modelo da cidade de São Paulo. Nossos experimentos foram realizados com as implementações de métodos de regressão da biblioteca *scikit-learn* da linguagem *python* [Pedregosa et al. 2011]. Utilizamos modelos de regressão linear e modelos de série temporais para construção do modelo Y_t mostrado na Equação 1.

4.2. Resultados obtidos

Nesta seção, comparamos o desempenho dos diferentes métodos de regressão e séries temporais. A seguir, discutimos os resultados obtidos com a interpretação de alguns gráficos gerados para facilitar a compreensão e obter mais clareza. Para interpretação e melhor entendimento foi gerado gráficos de Erro Bar e de ECDF(Empirical Cumulative Distribution Function).

Os gráficos de erro bar são usados para representar visualmente a variabilidade dos dados em torno de uma medida central, como a média. As barras de erro estendem-se acima e abaixo da medida central, indicando a faixa de variação dos dados. O intervalo de confiança, muitas vezes representado pelas barras de erro, proporciona uma estimativa da incerteza associada à medida central, geralmente com uma confiança estatística específica, como 95%. Quanto mais largas as barras de erro, maior é a variabilidade dos dados e menor é a precisão da estimativa da medida central.

Por outro lado, os gráficos ECDF são usados para mostrar como os dados estão distribuídos ao longo de uma faixa de valores observados. No gráfico ECDF, os valores observados são ordenados em ordem crescente no eixo horizontal, enquanto o eixo vertical representa a fração acumulada de observações que são menores ou iguais a cada valor no eixo horizontal. Isso permite visualizar rapidamente a distribuição cumulativa dos dados e identificar padrões, como a concentração de dados em certos intervalos ou a presença de outliers.

Na figura 3, o modelo RF mostra inicialmente uma superestimação das previsões, pois como o cálculo usado para o ecdf foi (real-predito), ele tem boa parte da sua linha na parte negativa do eixo x, o que nos diz que ele predisse valores maiores que os reais, ao longo da sua linha ele passa para o lado positivo do eixo x, mas permanece pouco tempo próximo ao 0, que é o centro do eixo. Já o SVR superestima bem menos que o RF, só que boa parte da sua linha está do lado positivo do eixo x, o que indica subestimação das suas predições. E o modelo AR pode ser considerado o melhor, pois por mais que ele subestime algumas predições, foi o modelo em que a linha ficou constantemente mais próxima de 0 no eixo x, estudos indicam que quanto mais próxima a curva do modelo estiver da linha zero no gráfico ECDF, melhor ele está prevendo os dados. Isso significa que as previsões são muito próximas dos valores reais.

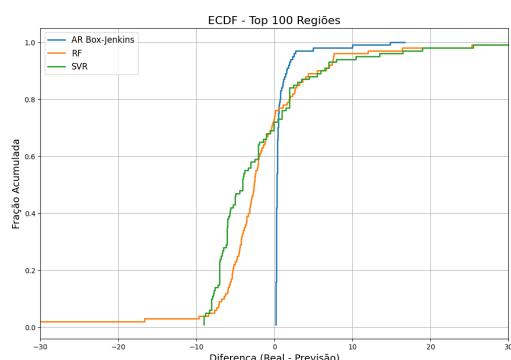


Figura 3. ECDF - Top 100

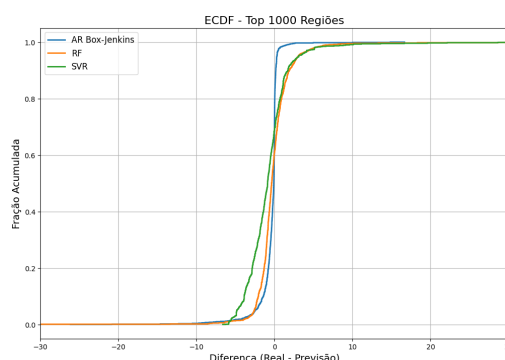


Figura 4. ECDF - Top 1000

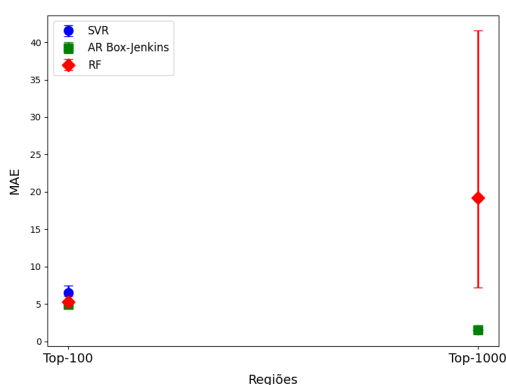


Figura 5. Error bar - MAE

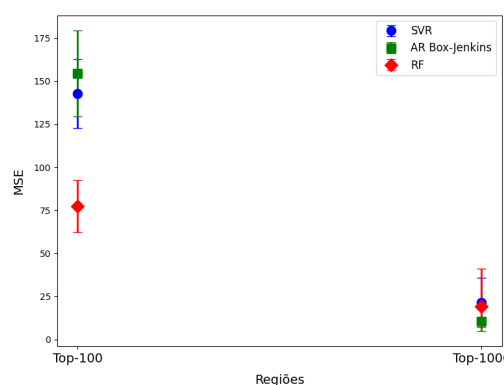


Figura 6. Error bar - MSE

Já para a figura 4, considerando as top-1000 regiões mais violentas de São Paulo a linha dos modelos em certa parte ficam até que semelhantes, mas observando de forma mais cautelosa podemos entender que o modelo RF superestima bastante novamente suas predições, já o SVR subestima mais suas predições que os demais modelos, da mesma forma do top-100 o modelo AR permanece por toda sua extensão mais próximo da linha zero no gráfico, o que pode representar uma predição mais próxima dos dados reais.

Considerando a figura 5, observamos que para as top-100 regiões mais violentas o SVR apresenta o maior valor de MAE, mas com uma diferença quase que insignificante para os demais modelos, que ficam bem ajustados no intervalo entre 4 e 8 de MAE. Já para as top-1000 regiões o RF apresenta um valor bem maior que os demais e com um intervalo de confiança consideravelmente grande para valores maiores, contudo, o AR e SVR conseguem ficar nivelados com um MAE bem baixo, inclusive menor que para as 100 regiões, levando em consideração a melhoria dos modelos com uma melhor quantidade de dados.

De acordo o observado na figura 6, top-100, o método RF apresenta o maior valor de MSE, por outro lado, o método AR Box-Jenkins e o SVR tem consideravelmente uma média de erros maiores, mas bem nivelados levando em conta até seus intervalos de confiança. Considerando os top-1000 setores, os modelos ficam bem equiparados em relação a ao erro calculado pelo MSE, que nos leva a entender que todos os métodos melhoram seu desempenho na previsão de crimes em áreas mais extensas, ou seja, quando eles possuem mais dados para serem trabalhados.

5. Conclusões e Trabalhos Futuros

Neste trabalho, investigamos o potencial de predição de modelos de regressão e séries temporais utilizando dados disponibilizados pela Secretaria de Segurança Pública de São Paulo. Nosso objetivo foi predizer taxas semanais de ocorrências de crimes nas regiões de São Paulo, focando nas top-100 e top-1000 áreas mais violentas, com dados de 2022 e 2023.

A pesquisa em computação sobre esse tema tipicamente utiliza características de tempo e espaço, além de dados demográficos e censitários, para predizer crimes. Nosso desafio foi predizer crimes unicamente baseado em dados oficiais de ocorrências nas regiões de São Paulo, com granularidade semanal. Analisamos a relação com modelos de regressão baseados em aprendizagem de máquina e séries temporais, avaliando o desempenho por meio de média de erros absolutos (MAE) e erros quadráticos (MSE).

Os resultados mostram que dados oficiais, considerando a divisão territorial da cidade por setores, podem predizer razoavelmente a quantidade de crimes em determinadas regiões e períodos do ano. O melhor modelo foi o ARIMA, ajustado pelo método Box-Jenkins. Os gráficos ECDF mostraram que o ARIMA teve predições mais próximas dos valores reais, tanto para as top-100 quanto para as top-1000 regiões. Nos gráficos de error bar, o ARIMA teve bom desempenho em MAE para ambos os grupos e desempenho razoável em MSE para as top-100 regiões. Portanto, o ARIMA com ajustes pelo método Box-Jenkins mostrou-se mais eficiente para a previsão de crimes em São Paulo com dados de 2022 e 2023.

Referências

- Araújo, A., Cacho, N., Bezerra, L., Vieira, C., and Borges, J. Towards a crime hotspot detection framework for patrol planning. In *2018 IEEE 20th International Conference on High Performance Computing and Communications*, pages=1256–1263, year=2018, organization=IEEE.
- Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Chainey, S. and Ratcliffe, J. (2008). *GIS and Crime Mapping*. John Wiley Sons.
- Chatfield, C. (2003). *The Analysis of Time Series: An Introduction*. Chapman & Hall/CRC, 6th edition.
- Drucker, H., Burges, C. J., Kaufman, L., Smola, A., Vapnik, V., et al. (1997). Support vector regression machines. *Advances in neural information processing systems*, 9:155–161.
- Gorr, W. L., Olligschlaeger, A. M., and Thompson, Y. (1994). Short-term forecasting of crime. *International Journal of Forecasting*, 10(3):129–142.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- IBGE. IBGE - Instituto Brasileiro de Geografia e Estatística. <https://www.ibge.gov.br/>. Acesso em: 15 jul. 2024.

- Ingilevich, V. and Ivanov, S. (2018). Crime rate prediction in the urban environment using social factors. *Procedia Computer Science*, 136:472–478.
- Kamarianakis, Y. and Prastacos, P. (2005). Space–time modeling of traffic flow. *Computers & Geosciences*, 31(2):119–133.
- Nery, M. B., Souza, A. A. L. d., and Adorno, S. (2019). Os padrões urbano-demográficos da capital paulista. *Estudos Avançados*, 33(97):5–36.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.
- SANTOS, J. M. D. R. B. D. (2020). Uso de sistema de informação geográfica para monitoramento da violência e criminalidade na Área do primeiro batalhão de polícia militar de alagoas.
- Sathyadevan, S. and Devan (2014). Crime analysis and prediction using data mining. In *2014 First international conference on networks & soft computing (ICNSC2014)*, pages 406–412. IEEE.
- Shumway, R. H. and Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples*. Springer, 4th edition.
- Silva, T. H., Viana, A. C., Benevenuto, F., Villas, L., Salles, J., Loureiro, A., and Quercia, D. (2019). Urban computing leveraging location-based social network data: a survey. *ACM Computing Surveys (CSUR)*, 52(1):1–39.
- SSP-SP (2022). Dados estatísticos do estado de são paulo. Disponível em: <http://www.ssp.sp.gov.br/estatistica/pesquisa.aspx>. Acesso em 10 de mai. 2022.
- SSP-SP (2023). Cidade de sp tem média de 21 roubos e furtos a residências por dia entre janeiro e agosto. Disponível em: <https://g1.globo.com/sp/sao-paulo/noticia/2023/09/26/cidade-de-sp-tem-media-de-21-roubos-e-furtos-a-residencias-por-dia-entre-janeiro-e-agosto.ghtml>. Acesso em: 10 de novembro 2023.
- Wang, F. and Xu, Y. (2013). Estimating o-d travel time matrix by google maps api: Implementation, advantages, and implications. *Annals of GIS*, 19(4):273–283.
- Weisburd, D., Groff, E. R., and Yang, S.-M. (2012). *The criminology of place: Street segments and our understanding of the crime problem*. Oxford University Press.