

Análise de dados de violência contra a mulher no estado Rio Grande do Sul com o uso de Aprendizado de Máquina

Renata Portella Rittmann¹,

Orientador: Duncan Dubugras Alcoba Ruiz¹

¹Escola Politécnica - Pontifícia Universidade Católica do Rio Grande Do Sul (PUCRS)

renata.rittmann@edu.pucrs.br, duncan.ruiz@pucrs.br

***Resumo.** Este artigo descreve os modelos desenvolvidos para a análise de dados de violência contra a mulher no estado do Rio Grande do Sul. Os dados analisados foram coletados a partir de informações disponibilizadas em portais do governos estadual e municipais, para auxiliar na predição de casos como lesão corporal, ameaças, estupros e feminicídios. O desenvolvimento deste artigo tem como objetivo identificar possíveis indicadores desses casos para possibilitar a predição de novos casos e uma redução do número de subnotificações.*

1. Introdução

A análise de dados com o uso do aprendizado de máquina tem sido cada vez mais utilizada para aprimorar e facilitar as decisões em questões de saúde pública e de segurança [6]. Um exemplo deste, é a sua utilização para prevenção de casos de violência no estado do Rio Grande do Sul, auxiliando na alocação de recursos de forma otimizada pelos órgãos responsáveis. Porém, até o dia de hoje, não são encontrados estudos na área da computação que busquem auxiliar especificamente a questão de proteção e prevenção de casos de violência contra a mulher.

A violência contra a mulher é um problema de saúde pública, um fenômeno social que atinge indivíduos e populações [4]. De acordo com a Convenção Interamericana para Prevenir, Punir e Erradicar a Violência contra a Mulher, conhecida como a Convenção de Belém do Pará [8], a violência contra a mulher é definida como qualquer ação ou conduta, baseada no gênero, que cause morte, dano ou sofrimento físico, sexual ou psicológico à mulher, tanto no âmbito público quanto no privado. Entre os desafios enfrentados na análise de dados de violência contra a mulher estão os casos de subnotificação, onde mulheres vítimas de agressão não realizam a denúncia contra seus agressores.

O objetivo deste trabalho é auxiliar na compreensão sobre a violência contra a mulher no estado do Rio Grande do Sul, a partir da análise de dados disponibilizados pelo estado e pelos municípios. Além disso, o propósito deste trabalho, é também, facilitar as pesquisas e alocação de recursos para tentar reduzir o número de casos e também de subnotificações. Para que isso seja possível, um modelo é treinado utilizando o Aprendizado de Máquina (AM) para que este seja capaz de realizar uma análise preditiva do número de casos de violência contra a mulher.

Nas próximas seções serão apresentados detalhes da implementação realizada, desenvolvimento do modelo e das decisões tomadas para cada etapa deste trabalho. Inici-

almente, é feita a apresentação de trabalhos relacionados que ajudaram com algumas decisões, seguindo para o processo de desenvolvimento do modelo e detalhando cada passo deste processo. Por fim, concluindo com a apresentação das características que determinaram a escolha da implementação, discutindo o que foi feito, como foi implementado e apresentando os resultados obtidos e a interpretação deles.

2. Trabalhos relacionados

A análise de dados com Aprendizado de Máquina tem sido muito explorada por diversas áreas, inclusive para os problemas de segurança e saúde como violência doméstica, criminalidade, entre outros. Um exemplo foi proposto por Alfeu Buriti Pereira Júnior e Francisco Icaro do Nascimento Ribeiro [13] onde realizaram um mapeamento sistemático para facilitar a análise de dados de violência doméstica com Aprendizado de Máquina. O artigo descreve também como foram feitas as buscas por dados, artigos e a seleção de dados. O objetivo deste foi destacar os dados mais relevantes para realizar a análise preditiva de violência doméstica, facilitando assim a identificação e realizando avaliações de risco, podendo ser usados para supervisão de agressores. Este artigo auxiliou na etapa inicial de mapeamento dos dados que seriam importantes para o desenvolvimento desta proposta. Contribuiu também na melhoria dos dados utilizados para a versão final, utilizando etapas do processo descrito no artigo e inclusive nos resultados obtidos pelos pesquisadores.

Podemos ver da mesma forma, exemplos de pesquisas da mesma área de inferência preditiva utilizando uma abordagem de aprendizado de máquina de outro tema, neste caso proposto por Henrique Joner [12] para identificação geoespacial da criminalidade em Porto Alegre. O artigo descreve quatro tipos de modelos que foram testados para a predição de eventos criminosos e do nível de criminalidade em cada localidade. Os modelos testados foram de regressão, classificação, redes neurais profundas e *long-short-term memory*. Todos obtiveram resultados bons para prever o número de crimes nos dias seguintes, e resultados excelentes para previsão do risco de criminalidade. O objetivo deste artigo foi construir modelos que tenham capacidade de prever a criminalidade de forma eficaz, o qual foi atingido. Esta pesquisa foi utilizada também como base para o desenvolvimento do projeto com o objetivo de auxiliar na tomada de decisões na implementação do modelo, para que fosse possível prever o número de casos de violência contra a mulher nos próximos anos.

Uma pesquisa desenvolvida recentemente pelos estudantes e pesquisadores Thiago Cunha Montenegro e Fábio Jorge Almeida Moraes da Universidade Federal de Campina Grande, estuda os índices de violência doméstica durante a pandemia do coronavírus [15]. A pesquisa demonstrou através de coletas de dados públicos e aplicação de técnicas de análise e visualização de dados, os impactos causados pelo isolamento social no número de casos de violência contra a mulher no Brasil. Desta forma, auxiliou a identificação de novos parâmetros que devem ser considerados a partir do início do isolamento social causado pelas medidas de prevenção do coronavírus para dados a partir do ano de 2020.

Existem outras pesquisas relacionadas com o tema de violência contra a mulher que realizam a análise de dados, como por exemplo, no estudo realizado por Alana Marques de Moraes, Aline Marques de Moraes e Dynnah Hanna Max P. G. da Silva [2], onde foram discutidos os resultados do planejamento e desenvolvimento de uma plataforma on-

line com o objetivo de apresentar de maneira mais intuitiva dados sobre a violência contra a mulher no Brasil. Este projeto destacou a importância do tema ainda pouco debatido e com baixo número de pesquisas, no qual procuravam auxiliar na busca de soluções efetivas que impactem na diminuição de casos, através da identificação de fatores que possam auxiliar na prevenção de casos de violência.

3. Modelo

Este trabalho foi desenvolvido adotando uma série de processos para que fosse possível atingir o resultado esperado do modelo de prever o número de casos de violência contra a mulher no Rio Grande do Sul, para cada um dos tipos de violência avaliados no contexto deste projeto. Para o desenvolvimento deste modelo em etapas, mostrou-se necessária a realização de um planejamento prévio de cada uma destas etapas definindo seus respectivos escopos e objetivos, que serão detalhados nesta seção.

3.1. Etapas para o desenvolvimento

O processo de desenvolvimento do modelo pode ser visto na Figura 1. Este processo é iniciado a partir do momento em que os dados são adquiridos em diversas fontes para em seguida unir em conjuntos de dados, preparando e separando os dados que podem ser relevantes para serem analisados. Assim será possível por fim atingir o objetivo desta primeira etapa, de coletar diversas informações que podem corroborar para a criação dos *datasets* que tenham um bom desempenho para o treinamento do modelo.

A partir deste ponto a segunda etapa é iniciada, realizando a análise exploratória dos *datasets* gerados. É iniciado um tratamento no conjunto de dados que faz parte do *script* final do modelo, analisando as informações e fazendo uma seleção entre os diferentes tipos de dados. Assim o objetivo desta etapa será atingido com a limpeza e seleção das linhas e colunas que serão utilizadas de fato nos *datasets* para a próxima etapa, no desenvolvimento do modelo.

Com relação à etapa de desenvolvimento do modelo, uma nova abordagem foi desenvolvida a partir dos resultados obtidos na proposta deste projeto. Para que isso fosse possível foram utilizadas diferentes bibliotecas na implementação visando melhores resultados para os conjuntos de dados atualizados. Por meio de testes realizados em diferentes períodos, foram analisados os resultados médios de validações cruzadas para garantir a qualidade do modelo.

Na última etapa é realizada a avaliação do modelo, seguida da análise e interpretação de todos os resultados obtidos a partir do modelo e de pesquisas. Esta etapa tem como objetivo avaliar se o modelo desenvolvido, em conjunto com os *datasets* utilizados atenderam as necessidades e expectativas para auxiliar com problema de proteção e prevenção de casos de violência contra a mulher.

3.2. Aquisição e tratamento inicial dos dados

Para a aquisição dos dados brutos foi realizada uma busca em portais oficiais do governo do país [16], do estado do Rio Grande do Sul [17] e também de algumas cidades para buscar mais detalhes de algumas informações disponibilizadas nos *datasets*, como por exemplo ano de implantação de um serviço na cidade. Nas primeiras buscas foram encontrados diversos dados sobre a violência contra a mulher no portal da Secretaria de



Figura 1. Diagrama do processo para desenvolvimento do modelo

Fonte: a autora

Segurança Pública do RS [18], com *datasets* anuais separados por tipo de violência para cada cidade do RS. Por conta disto tornou-se necessário escolher pontos principais a serem trabalhados e avaliados como áreas e períodos para este projeto.

A partir destas pesquisas, em um primeiro momento foi gerado um *dataset* que contém as informações gerais de cada cidade do Rio Grande do Sul, tais como: nome da cidade, sua respectiva mesorregião e microrregião. Em conjunto com essas informações foram coletados na plataforma do IBGE [11] os dados sobre a população no último censo, PIB per capita e índice de desenvolvimento humano municipal (IDHM). Foram utilizados também o ano de implantação da patrulha Maria da Penha (PMP) [1] e os últimos dados do ano de inauguração e quantidade de locais de atendimento à mulher¹ por cidades [19]. O ano de implantação da PMP por cidade é disponibilizado no *dataset* do observatório da mulher disponibilizado pelo estado [1], que também disponibiliza as informações sobre os locais de atendimento, como quais cidades apresentam esse serviço e quais, porém não informam o ano de instauração por cidade. A informação do ano de implantação de cada local de atendimento, de cada uma das cidades que possuem, foi adquirida através de pesquisas nos portais da cidade e também de jornais locais que noticiaram a inauguração do serviço.

Para completar as pesquisas foram coletados os *datasets* disponibilizados pelo governo do estado [18] que apresentam os dados sobre violência contra a mulher distribuídos por cidades para cada ano. Os principais dados arrecadados destes casos são alguns diferentes tipos de violência como o número de ameaças, lesão corporal, estupro, feminicídio tentado e feminicídio consumado.

Por fim, os dados de violência foram agrupados com os dados de cada cidade, gerando um novo *dataset* completo com todas as informações disponíveis das cidades por ano. Essas informações foram agrupadas em outros conjuntos de dados complementares com a divisão por microrregiões. Portanto, para cada ano existem dois *datasets* com os números de casos de violência contra a mulher, um dividido por cidades (total de 497 cidades) e outro por microrregiões, que são 35 no total como é possível ver na Figura 2.

3.3. Análise exploratória

A partir dos *datasets* gerados na etapa anterior é iniciada a análise exploratória dos dados obtidos. O tratamento inicial é realizado, validando o formato dos dados de cada coluna, padronizando os valores e excluindo valores nulos.

¹Locais de atendimento considerados no estudo: Delegacia Especializada no Atendimento à Mulher (DEAM), Sala das Margaridas, Posto Policial de Proteção à Mulher (PPPM) e Cartório Especializado no Atendimento à Mulher.

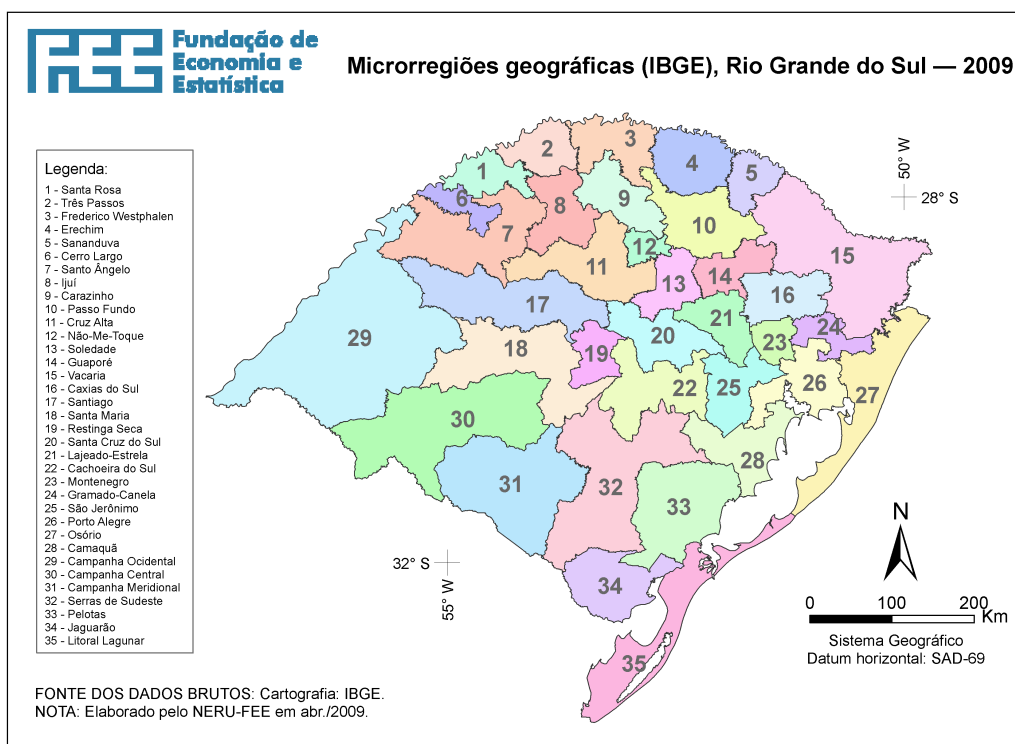


Figura 2. Mapa de Microrregiões do Rio Grande do Sul

Fonte: IBGE - Retirado de: Info Escola - Mapas [7]

Segue-se então para a análise da Correlação de *Pearson* entre as colunas, sendo gerado um conjunto apenas com atributos numéricos para a correlação. A correlação é fundamental para auxiliar na identificação de dados que impactam diretamente um no outro e que devem ser utilizados, auxiliando também na redução da dimensionalidade. A análise correlacional indica a relação entre duas variáveis lineares e os valores sempre serão entre “+1” e “-1”. O sinal indica a direção, se a correlação é positiva ou negativa, e o tamanho da variável indica a força da correlação.

As análises foram realizadas individualmente para cada um dos dois tipos de *datasets*, um que possui informações por cidades e o outro por microrregiões, que serão detalhadas em suas respectivas subseções a seguir. Em ambos, as colunas que serão avaliadas para realizar a predição no modelo são as que apresentam os números de casos de violência contra a mulher. São elas: ameaça, lesão corporal, estupro, feminicídio tentado e feminicídio consumado.

3.3.1. Análise por cidades

Os *datasets* que foram gerados por cidades utilizaram a planilha inicialmente desenvolvida, citada anteriormente, com diversas informações de cada uma das 497 cidades do Rio Grande do Sul. Nestes conjuntos de dados foram utilizadas as seguintes colunas, para a realização da análise, com seus respectivos tipos de informações:

- Cidade: nome, informação removida para treino do modelo;
- População: número total de habitantes da cidade. Tipo final: valor inteiro;

- PIB per capita²: valor em reais. Tipo final: arredondado para valor inteiro;
- IDHM (Índice de Desenvolvimento Humano Municipal): valor original entre 0 e 1 com três casas decimais. Tipo final: convertido para inteiro entre 0 e 100;
- PMP (Patrulha Maria da Penha): booleano indicando se possui (1) ou não (0);
- Locais de atendimento: booleano indicando se possui (1) ou não (0).

Durante a análise exploratória manual dos dados já foi possível identificar cidades onde o número de casos de violência registrados eram baixos, quando comparadas com cidades vizinhas com condições socioeconômicas similares. Desta forma, estas cidades podem representar um alto número de subnotificações, que possivelmente estão relacionados com a ausência das opções de apoio que são a patrulha Maria da Penha e os locais de atendimento à mulher. Este entendimento foi resultado de comparações de cidades que não apresentavam nenhuma das opções de apoio a mulher no ano de 2019, por exemplo, e que no ano de 2020 passaram a contar com pelo menos uma das opções de apoio. Nestas cidades foi possível observar um aumento no número de casos relatados.

Foi possível identificar correlações nos níveis moderado, forte e muito forte com a análise da correlação de Pearson dos dados de um dos *datasets* gerados, apresentado na Figura 3. Este diagrama auxilia a ilustrar o exemplo citado anteriormente do impacto da ausência de sistemas de apoio, onde há uma correlação entre moderada e forte com valores de “0.6” a “0.75” para o período avaliado. Esta correlação foi verificada entre o sistema de apoio e o número de casos registrados de lesão corporal e ameaças.

Além disso foram identificadas correlações muito fortes entre a população da cidade e os casos de estupro, feminicídio tentado e feminicídio consumado, e também entre os casos de estupro e feminicídio tentado. A partir das informações coletadas no diagrama da Figura 3 notou-se também a correlação fraca dos dados como PIB per capita e o IDHM, que foram removidos para a realização do treino do modelo.

3.3.2. Análise por microrregiões

Os conjuntos de dados gerados por microrregiões também utilizaram a planilha com detalhes de todas as cidades do Rio Grande do Sul. Para que isso fosse possível, as informações de cada cidade utilizadas foram agrupadas em um único *dataset* para analisar cada uma das 35 microrregiões que podem ser vistas na Figura 2. Neste *dataset* foram utilizadas as seguintes colunas para a realização da análise, com seus respectivos tipos de informações:

- Microrregião: nome, informação removida para treino do modelo;
- Número de cidades: total de cidade agrupadas em cada microrregião. Tipo final: valor inteiro;
- População: número total de habitantes de todas as cidades da microrregião somadas. Tipo final: valor inteiro;
- PMP (Patrulha Maria da Penha): informação somada de quantas cidades possuem a patrulha. Tipo final: valor inteiro;

²O PIB é a soma de todos os bens e serviços finais produzidos por um país, estado ou cidade, geralmente em um ano. Para o cálculo do PIB, são utilizados diversos dados; alguns produzidos pelo IBGE, outros provenientes de fontes externas [10].

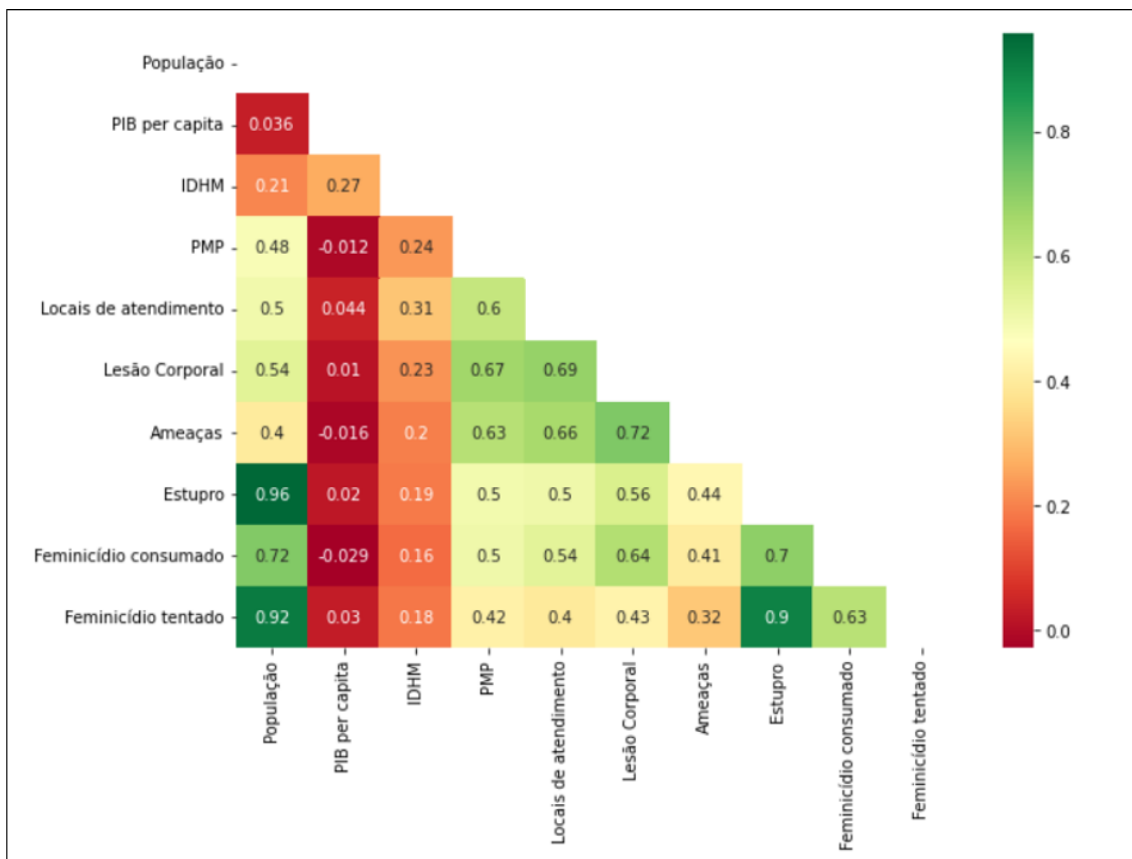


Figura 3. Correlação de Pearson dos dados do *dataset* de 2019 por cidades

Fonte: a autora

- Locais de atendimento: informação somada de quantas cidades possuem locais de atendimento. Tipo final: valor inteiro;

Assim como na análise dos *datasets* por cidades, nos *datasets* por microrregiões também foi possível identificar um baixo número de casos de violência registrados para alguns casos. Algumas dessas microrregiões não possuem nenhuma das opções de apoio em nenhuma das cidades dela, com isso são as que menos tem registros de casos de violência para todos os tipos trabalhados. Assim, indica-se a possível causa de um alto número de subnotificações.

Essas informações podem ser conferidas na Figura 4 com os dados mais recentes de 2021. Nesta são ilustradas as regiões que não possuem o serviço da patrulha Maria da Penha e nem os locais de atendimento que estão em vermelho na Figura 4, seguidas pelas microrregiões que não possuem nenhum local de atendimento em laranja e por fim as que não possuem patrulha Maria da Penha para nenhuma de suas cidades representadas em amarelo.

A análise de correlações realizadas nos *datasets* das microrregiões do estado tiveram ótimos resultados, que podem ser vistos na Figura 5, melhores que os obtidos nas correlações entre os dados por cidades. Apenas a coluna que possui um propósito mais informativo, que é a de número de cidades agrupadas por microrregião, não apresentou correlações significativas com os demais dados. A análise resultou na exclusão da coluna

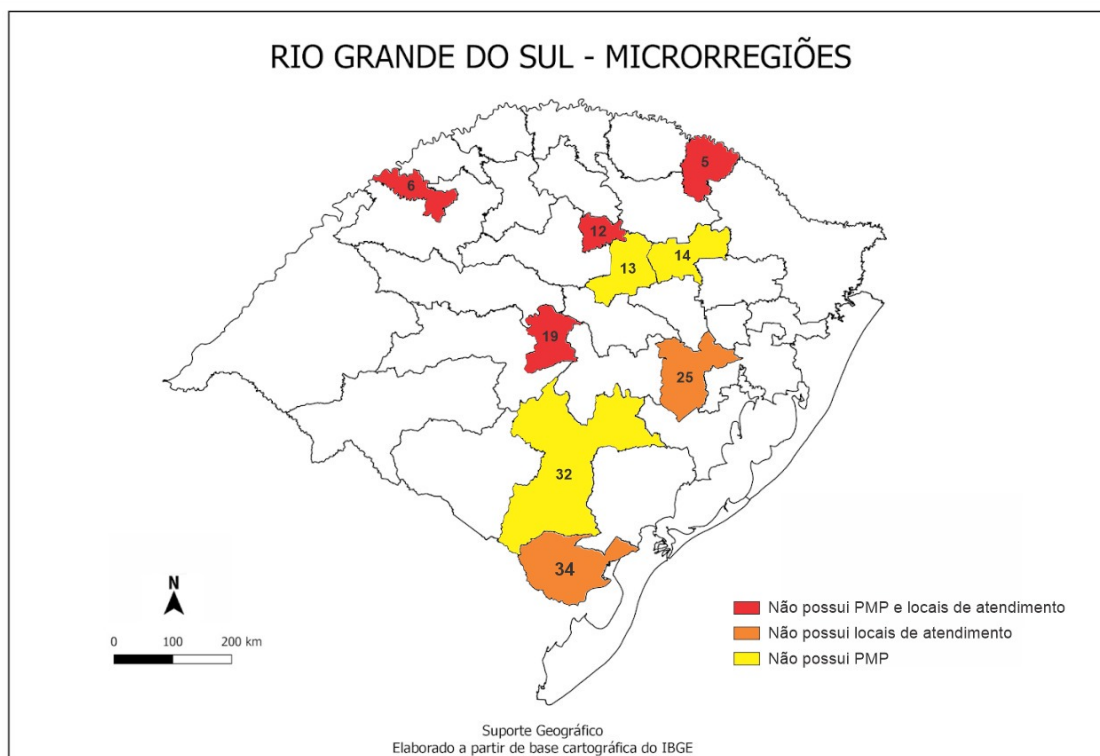


Figura 4. Microrregiões que não possuem serviços de suporte a mulher em 2021

Fonte: base cartográfica IBGE; Editado pela autora.

de número de cidades para a execução do modelo.

3.4. Desenvolvimento do modelo

A terceira etapa do processo é iniciada com a avaliação de modelos a serem desenvolvidos que possam apresentar melhores resultados para o problema em questão. O modelo escolhido de aprendizado supervisionado para o desenvolvimento foi um de regressão, tendo em vista que o objetivo é prever um valor numérico, neste caso o número de casos de violência para cada um dos tipos selecionados para este projeto, já citados anteriormente.

O algoritmo utilizado nesta implementação foi o *XGBoost* (*Extreme Gradient Boosting*), considerado atualmente um dos algoritmos mais populares na área de Aprendizado de Máquina, o qual serve para problemas de regressão e de classificação [3]. Além disso ele também é conhecido por fornecer soluções melhores que outros algoritmos para lidar com dados estruturados. Este pertence a uma família de algoritmos de aumento e usa a estrutura de aumento de gradiente (GBM) em seu núcleo. É uma biblioteca otimizada de *boosting* de gradientes distribuídos.

3.4.1. Preparo do ambiente

Para possibilitar o treinamento dos modelos escolhidos com os *datasets* gerados, foi necessário um preparo do ambiente para a leitura destes arquivos, execução do algoritmo e por fim a análise dos resultados. Considerando os dois diferentes tipos de conjuntos

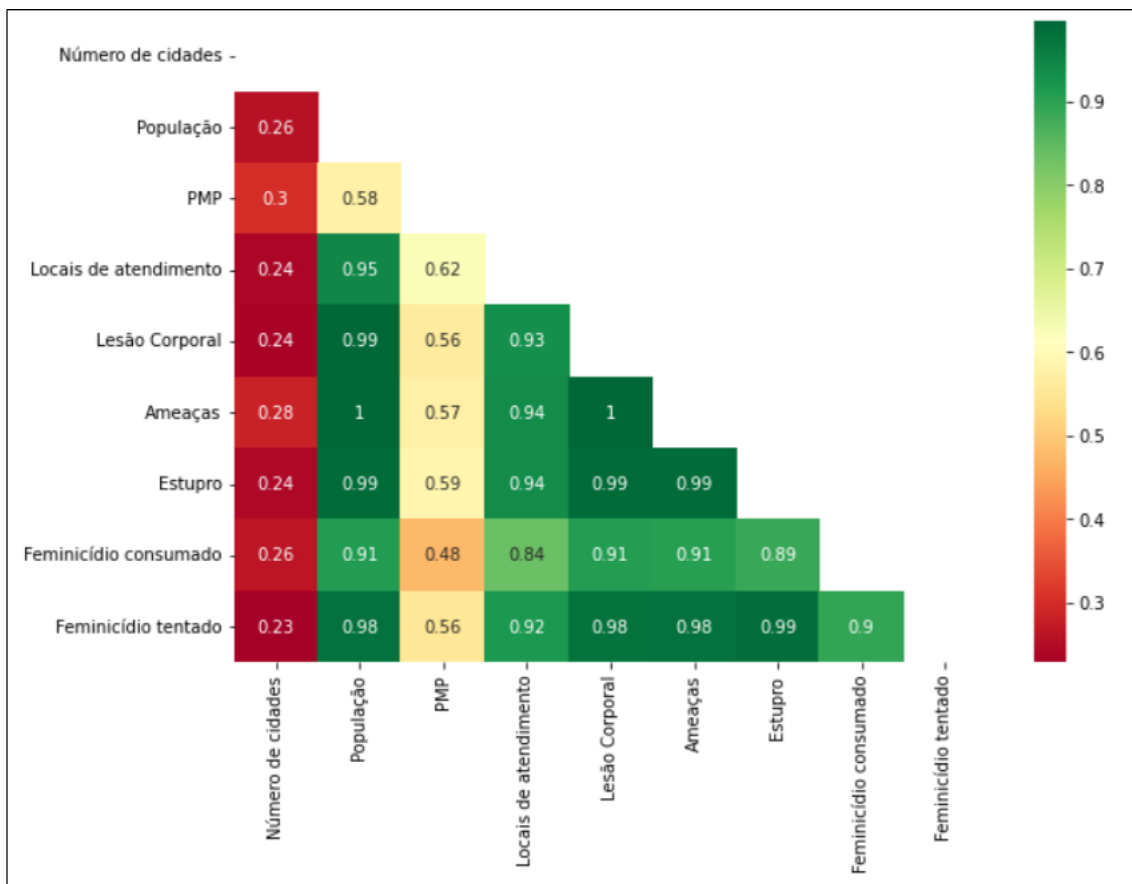


Figura 5. Correlação de Pearson dos dados do *dataset* de 2019 por microrregiões

Fonte: a autora

de dados gerados (cidade e microrregião) o modelo escolhido foi implementado em dois arquivos separados para facilitar a análise de resultados em cada contexto de dados. A seguir serão apresentados alguns dos preparos realizados no ambiente para a execução do projeto.

Os dados de entrada para o treinamento do modelo são arquivos em formato *csv*. Estes arquivos contendo os dados são carregados através do uso de bibliotecas disponibilizadas pelo *python*, utilizando o *jupyter* ou *google colab* para a execução. Foi necessário realizar a instalação da biblioteca *scikit-optimize* para possibilitar o uso do *BayesSearchCV*, que utiliza o *XGBRegressor* como estimador. O *BayesSearchCV* implementa métodos que dão suporte à otimização sobre hiperparâmetros, que são aplicados para auxiliar nas pesquisas de validação cruzada.

3.4.2. Detalhes da implementação

Após a configuração do ambiente citada anteriormente, pode-se dar início a implementação do modelo, importando os *datasets* gerados do período entre 2019 e 2021. Apesar dos dados do ano de 2018 terem sido coletados para a utilização no treinamento do modelo, eles foram removidos do estudo, pois como alguns estudos e reportagens [9] indicaram, houve em 2019 uma virada no número de casos relatados. Portanto, os dados

de 2018 se tornaram dispensáveis para o estudo e para evitar que estes dados impactassem negativamente nos resultados deste projeto, o *dataset* foi removido.

Para cada um dos arquivos desenvolvidos para os modelos, um dos *datasets* de cidades e o outro de microrregiões, a mesma lógica foi implementada. A abordagem adotada para predição de novos casos, foi de treinar modelos separados para cada tipo de violência contra a mulher ao invés de agrupar todos os números de casos, tendo em vista que cada um possui sua individualidade nas correlações.

Cada modelo foi treinado utilizando *BayesSearchCV*, citado previamente, por ser o responsável por otimizar o passo a passo para explorar os hiperparâmetros mais promissores no contexto do problema. Os parâmetros passados para o modelo com seus respectivos valores foram:

- *estimator*: algoritmo *XGBRegressor*, explicado anteriormente;
- *search_spaces*: recebe um dicionário, onde cada chave é o nome do parâmetro e o valor é uma lista de possíveis entradas para aquele parâmetro. Para cada modelo é selecionado a combinação dos valores de cada parâmetro que obteve o melhor desempenho;
- *scoring*: entre as opções disponíveis [14] para o modelo, foi selecionada a *neg_root_mean_squared_error*³.

Por fim, o modelo é retornado e as informações dos seus resultados são impressas para serem avaliadas, como a melhor combinação de parâmetros e o valor do melhor desempenho do *Root Mean Squared Error* (RMSE). Os modelos treinados são utilizados para realizar uma avaliação dos resultados em uma predição para o ano de 2022. Para essa avaliação são simuladas a implantação de patrulhas Maria da Penha e novos locais de atendimento, para observar o impacto no número de casos por cidades e microrregiões.

3.5. Análise e interpretação dos resultados

Os modelos treinados com os dados do período de 2019 a 2021 são utilizados para testar predições de casos nos próximos anos. Para a realização destes testes dois novos *datasets* foram gerados para o ano de 2022, para cada um dos tipos avaliados no projeto (cidade e microrregiões). Ambos não possuem as colunas com informações do número de casos de violência contra a mulher. O primeiro *dataset* gerado, chamado de “2022 original” manteve todas as informações conforme os dados de 2021, para poder analisar o resultado da predição. O segundo, chamado de “2022 alterado”, foi realizada a adição manual para algumas cidades e microrregiões de patrulha Maria da Penha e/ou de locais de atendimento para avaliar o impacto da presença desses serviços para cada modelo.

3.5.1. Resultados por cidades

Os resultados das aplicações dos modelos sobre o ano de 2022 serão demonstrados a seguir, para algumas cidades em tabelas para demonstrar o impacto nos números de casos de violência. Para exemplificação de diferentes situações foram selecionadas quatro cidades com diferentes situações dos serviços de apoio ao decorrer do período avaliado.

³Esta métrica calcula a média dos erros do modelo ao quadrado (RMSE), ou seja os resultados com diferenças pequenas tem menos importância enquanto os resultados com diferenças maiores recebem mais peso no cálculo. Retornando o valor obtido negativo.

A primeira cidade é Aceguá, disponível na Tabela 1, que possui uma população média de 4400 habitantes, até hoje ela não possui nenhum dos serviços de apoio a mulher na sua cidade. É possível notar uma variação no número de casos relatados anualmente, mas ao adicionar um local de atendimento na cidade alguns tipos de violência tiveram um aumento significativo no número de casos. Desta forma é possível concluir que uma boa parte dos casos não são relatados e que com a presença de pelo menos um dos serviços de apoio na cidade as denúncias aumentariam. Possivelmente, esta cidade apresentaria uma redução nos anos seguintes, de casos de violência, devido ao aumento na segurança para as mulheres e o rastreamento dos agressores.

Aceguá	PMP	Locais de Atendimento	Lesão Corporal	Ameaça	Estupro	Feminicídio Consumado	Feminicídio Tentado
2019	Não	Não	6	4	0	0	0
2020	Não	Não	5	7	1	0	0
2021	Não	Não	1	6	1	0	0
2022 original	Não	Não	4	6	1	0	0
2022 alterado	Não	Sim	5	13	3	0	0

Tabela 1. Dados de violência contra a mulher da cidade de Aceguá.

Fonte: a autora

A segunda cidade avaliada é Alegrete, disponível na Tabela 2, que possui uma população média de 77,6 mil habitantes, no início do período avaliado (2019) ela já possuía pelo menos um local de atendimento à mulher na cidade. Neste caso, por já existir um dos serviços de apoio podemos verificar um impacto menor nos números de casos ao adicionar na cidade também o serviço da patrulha Maria da Penha. A patrulha faz um acompanhamento de mulheres que denunciaram agressores ou que já foram agredidas e solicitaram por este serviço que as acompanham em situações de risco. Este serviço tende a impactar positivamente nos número de casos de violência, os reduzindo conforme a brigada militar consegue acompanhar um maior número de vítimas que se cadastram e evitar que casos de violência se repitam.

Na terceira cidade avaliada, Palmeira das Missões, com seus dados disponíveis na Tabela 3, que possui uma média de 34300 habitantes, adquiriu o serviço da patrulha Maria da Penha no meio do período avaliado (2020). A partir disso, já pode-se notar um aumento significativo no número de casos relatados, assim como na primeira cidade (Aceguá, da Tabela 1) onde ambas inicialmente não possuíam nenhum serviço e ao adquirir apresentaram esse aumento. Portanto, podemos inferir que o número de subnotificações dessas cidades reduziu. Ao adicionar o outro serviço de apoio nesta cidade, de locais de atendimento, o número de casos de lesão corporal e ameaças apresentaram uma redução leve, indicando uma possível estabilidade no número de casos podendo evoluir nos anos seguintes para a redução de casos, devido a existência do sistema de apoio.

A última cidade avaliada detalhadamente é Taquara, disponível na Tabela 4, que

Alegrete	PMP	Locais de Atendimento	Lesão Corporal	Ameaça	Estupro	Feminicídio Consumado	Feminicídio Tentado
2019	Não	Sim	142	299	9	1	2
2020	Não	Sim	122	215	15	1	2
2021	Não	Sim	101	204	13	0	1
2022 original	Não	Sim	123	237	13	1	1
2022 alterado	Sim	Sim	144	328	14	1	1

Tabela 2. Dados de violência contra a mulher da cidade de Alegrete.

Fonte: a autora

Palmeira das Missões	PMP	Locais de Atendimento	Lesão Corporal	Ameaça	Estupro	Feminicídio Consumado	Feminicídio Tentado
2019	Não	Não	63	198	0	0	0
2020	Sim	Não	86	242	5	0	0
2021	Sim	Não	59	225	7	0	0
2022 original	Sim	Não	70	235	5	0	0
2022 alterado	Sim	Sim	61	222	7	0	1

Tabela 3. Dados de violência contra a mulher da cidade de Palmeira das Missões.

Fonte: a autora

assim como a cidade de Aceguá, da Tabela 1, não possuem nenhum serviço de apoio a mulher na cidade. Taquara possui uma população média de 54,6 mil habitantes e mesmo sem ter serviços disponíveis localmente específicos de suporte a mulher, o número de casos relatados eram altos. Estes possuem uma variação anual irregular que pode indicar casos de subnotificações devido a ausência de recursos. Ao adicionar ambos os serviços na cidade, patrulha Maria da Penha e locais de atendimento houve um aumento de alguns tipos de violência e estabilidade em outros, possivelmente representando a diminuição de subnotificações.

Por fim, a análise geral dos resultados obtidos dos modelos para cada tipo de violência por cidade foram bons. Os resultados em RSME de cada um desses modelos podem ser verificados na Tabela 5. Apesar do resultado com um alto número no modelo de casos de ameaça ainda é considerado um bom desempenho devido a alta variação média dos dados, onde no período avaliado o menor valor de casos relatados era zero e o maior 4082. O mesmo é aplicado para o resultado do modelo de lesão corporal por cidades, onde o menor valor de casos relatos também era zero e o maior de 3520 casos.

Portanto, apesar dos bons resultados com alto desempenho dos modelos seria

Taquara	PMP	Locais de Atendimento	Lesão Corporal	Ameaça	Estupro	Feminicídio Consumado	Feminicídio Tentado
2019	Não	Não	124	217	4	1	3
2020	Não	Não	117	171	14	1	2
2021	Não	Não	109	144	9	0	1
2022 original	Não	Não	117	182	9	0	1
2022 alterado	Sim	Sim	119	260	12	0	2

Tabela 4. Dados de violência contra a mulher da cidade de Taquara.

Fonte: a autora

Tipo de Violência	Resultado do Modelo
Ameaça	38.1382343469239
Lesão Corporal	13.242436254065376
Estupro	3.109381633673528
Feminicídio Tentado	1.7078164051953268
Feminicídio Consumado	0.44433456577234526

Tabela 5. Resultados em RSME dos modelos desenvolvidos de cidades

Fonte: a autora

possível melhorá-los ao remover cidades que deveriam possuir análises exclusivas devido ao seu tamanho e número de casos, como por exemplo Porto Alegre. Em Porto Alegre, para os modelos que não tiveram ótimos resultados, como o de ameaça e lesão corporal, é possível notar que o seu número de casos apresentam, para cada um desses tipos de violência, pelo menos 200% maior, podendo chegar a diferenças de quase 400% maior, que a cidade com o segundo maior número de casos de ameaça e lesão corporal.

3.5.2. Resultados por microrregiões

Os resultados obtidos com os modelos treinados para microrregiões, serão demonstrados para os *datasets* gerados de 2022. Assim como na análise de resultados das cidades, para as microrregiões algumas foram selecionadas para exemplificar os resultados obtidos deste modelo, da mesma forma realizada na seção anterior.

A primeira microrregião analisada é a de Carazinho, disponível na Tabela 6, que agrupa 18 cidades, com uma população de 160 mil habitantes em média. Até o ano de 2020 nenhuma das cidades desta microrregião tinha os serviços de apoio à mulher, que inicialmente indicou um aumento no número de casos relatados. Posteriormente, a média de casos estabiliza indicando, possivelmente, uma redução do número de subnotificações

mesmo com o aumento do número de locais de atendimento.

A microrregião de Santo Ângelo, a segunda analisada, que agrupa 16 cidades com uma média de 197 mil habitantes, apresentou resultados similares aos de Carazinho, como é possível ver na Tabela 7. Visto que os *datasets* de “2022 Original” contêm os mesmos dados de 2021 das regiões para testar a capacidade de predição do modelo, podemos concluir que o modelo está tendo um bom desempenho.

Carazinho	PMP	Locais de Atendimento	Lesão Corporal	Ameaça	Estupro	Feminicídio Consumado	Feminicídio Tentado
2019	Não	Não	251	618	16	1	3
2020	Sim (8)	Não	286	661	24	0	1
2021	Sim (8)	Sim (1)	256	672	22	0	0
2022 original	Sim (8)	Sim (1)	256	669	24	1	1
2022 alterado	Sim (8)	Sim (4)	239	663	30	2	1

Tabela 6. Dados de violência contra a mulher da microrregião de Carazinho.

Fonte: a autora

Santo Ângelo	PMP	Locais de Atendimento	Lesão Corporal	Ameaça	Estupro	Feminicídio Consumado	Feminicídio Tentado
2019	Sim (1)	Sim (2)	356	791	33	2	8
2020	Sim (2)	Sim (2)	331	680	36	1	4
2021	Sim (2)	Sim (2)	314	675	35	3	5
2022 original	Sim (2)	Sim (2)	324	678	35	2	5
2022 alterado	Sim (2)	Sim (6)	304	654	63	3	6

Tabela 7. Dados de violência contra a mulher da microrregião de Santo Ângelo.

Fonte: a autora

Por fim, a última microrregião analisada foi a de Cerro Largo, disponível na Tabela 8, a qual agrupa 11 cidades do Rio Grande do Sul, com um total aproximado de 66 mil habitantes. Esta não possui até hoje nenhum serviço de apoio à mulher e simulando a adição de pelo menos dois locais de atendimento e patrulha Maria da Penha em uma das cidades, pode-se notar o aumento de casos relatados. Da mesma forma que as outras microrregiões analisadas, o modelo desempenhou bem para a predição de novos casos.

Concluindo, desta forma, que ambas as abordagens para análise dos dados de violência contra a mulher tiveram bom desempenho, conforme esperado. Os resultados dos modelos treinados por microrregião em RSME, explicado anteriormente, podem ser verificados na Tabela 9.

Cerro Largo	PMP	Locais de Atendimento	Lesão Corporal	Ameaça	Estupro	Feminicídio Consumado	Feminicídio Tentado
2019	Não	Não	68	184	8	1	1
2020	Não	Não	78	161	9	0	1
2021	Não	Não	64	158	14	2	1
2022 original	Não	Não	71	170	10	1	2
2022 alterado	Sim (1)	Sim (2)	85	167	14	1	3

Tabela 8. Dados de violência contra a mulher da microrregião de Cerro Largo.

Fonte: a autora

Em comparação com os resultados individuais para cada abordagem a de microrregiões desempenhou melhor nos testes de predições, devido ao impacto causado com a inclusão das grandes cidades, como Porto Alegre, que possuem um elevado número de casos. Os modelos de microrregiões seguiram os padrões dos últimos anos apesar da notável diferença entre cada uma delas, como por exemplo a microrregião de Porto Alegre com média de 11 mil casos de ameaça anuais e a de Cerro Largo com apenas 170 casos de média aproximada nos últimos anos.

Tipo de Violência	Resultado do Modelo
Ameaça	173.58880875589014
Lesão Corporal	128.15834264485642
Estupro	14.339513711922388
Feminicídio Tentado	6.294282023146267
Feminicídio Consumado	1.6513466127228233

Tabela 9. Resultados em RSME dos modelos desenvolvidos de microrregiões

Fonte: a autora

4. Conclusão

Ao longo do desenvolvimento deste trabalho foram implementados modelos de Aprendizado de Máquina, para que auxiliassem na análise de dados de violência contra a mulher no estado do Rio Grande do Sul. Este tinha como objetivo identificar possíveis indicadores desses casos e também de subnotificações para possibilitar a otimização dos recursos de segurança destinados à esta área, visando a prevenção de novos casos.

Os modelos desenvolvidos, como foi possível ver na seção “Análise e interpretação dos resultados”, apresentaram bons resultados cumprindo assim o objetivo esperado para este projeto. Foi possível prever números de casos para os anos seguintes e os seus respectivos impactos em caso de criação e implantação de novos serviços de apoio à mulher. Além disso, conseguiu-se notar o aumento do número de casos relatos

em cidades e microrregiões com a presença destes serviços indicando assim, possíveis locais que apresentam altos números de subnotificações. Como foi demonstrado no caso da cidade de Aceguá, apresentada na Tabela 1, que até hoje não possui nenhum dos serviços de atendimento, ao adicionar pelo menos um deles notou-se um aumento de aproximadamente 100% no número total de casos relatados.

Durante as pesquisas realizadas para o desenvolvimento deste projeto, foi possível notar o baixo número de pesquisas na área da computação relacionados a este tema. Podendo indicar, desta forma, uma problemática tendo em vista que a tecnologia pode auxiliar a traçar estratégias de segurança para a mulher no estado. O alto número de casos relatados de violência contra a mulher, dos dados coletados e também depoimentos dessas vítimas, servem como estímulo para seguir o desenvolvimento desta pesquisa.

Esta melhoria pode ser realizada, em trabalhos futuros, através de buscas com o objetivo de aprimorar ainda mais os dados obtidos com informações sobre as cidades e suas microrregiões. Portanto, podem obter melhores resultados na identificação de casos subnotificados, que ainda representam boa parcela dos casos reais de acordo com pesquisas [5]. Este aprimoramento dos dados pode ser feito em conjunto com ONGs e serviços existentes de apoio à mulher vítima de violência. Por fim, os resultados obtidos dos modelos poderiam ser utilizados para suporte ao desenvolvimento de uma aplicação que realizasse as simulações, feitas manualmente neste projeto para a análise dos resultados, de forma automática. Desta forma, esta aplicação poderia informar quais cidades e/ou microrregiões onde a alocação de recursos resultaria em um impacto positivo de forma efetiva.

Referências

- [1] BRIGADA MILITAR DO RIO GRANDE DO SUL. Ano implementação da Patrulha Maria da Penha por cidades RS; Disponível em: <https://www.brigadamilitar.rs.gov.br/historico-da-patrolha-maria-da-penha>.
- [2] DA SILVA, D. H. M. P., DE MORAIS, A. M., AND DE MORAIS, A. M. Nossa voz: uma plataforma digital para visualização de dados sobre violência contra a mulher no brasil. In *Anais do XV Women in Information Technology* (2021), SBC, pp. 315–319.
- [3] DATACAMP PATHAK, M. XGBoos Python Tutorial; Disponível em: <https://www.datacamp.com/tutorial/xgboost-in-python>.
- [4] DE ALMEIDA TELES, M. A., AND DE MELO, M. *O que é violência contra a mulher*. Brasiliense, 2017.
- [5] DE MEDICINA UFMG, F. Casos de violência doméstica são subnotificados e realidade contrapõe dados; Disponível em: <https://www.medicina.ufmg.br/casos-de-violencia-domestica-sao-subnotificados-e-realidade-contrapoe-dados/>.
- [6] DE PESQUISA EM CIÊNCIA DE DADOS NA UNIVERSIDADE FEDERAL DO CEARÁ, I. L. L. Big Data e Inteligência Artificial em ação no combate ao crime; Disponível em <https://insightlab.ufc.br/big-data-e-inteligencia-artificial-em-acao-no-combate-ao-crime>.
- [7] ESCOLA, I. Mapa das Microrregiões do Rio Grande do Sul; Disponível em: <https://www.infoescola.com/mapas/mapa-das-microrregioes-do-rio-grande-do-sul/>.
- [8] GERAL, A. A., 1994. Convenção Interamericana para Prevenir, Punir e Erradicar a Violência contra a Mulher - Convenção de Belém do Pará; Disponível em: <http://www.oas.org/juridico/portuguese/treaties/a-61.htm>.
- [9] GOVERNO RS, S. H. A. D. D. P. P. A. M. D. E. D. R. G. D. S. *Cartilha da Mulher - Guia para a adesão de políticas de proteção às mulheres nos municípios*. Governo do Estado do RS - Secretaria de Comunicação, 2021.
- [10] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Produto Interno Bruto - PIB; Disponível em: <https://www.ibge.gov.br/explica/pib.php>.
- [11] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE, 2010. IBGE Cidades Panorama; Disponível em: <https://cidades.ibge.gov.br/brasil/rs/>.
- [12] JONER, H. Inferência preditiva geoespacial da criminalidade em porto alegre: uma abordagem de aprendizado de máquina. In *Teses e dissertações disponível no Lume - repositório digital* (2020), UFRGS.
- [13] JÚNIOR, A. B. P., AND DO NASCIMENTO RIBEIRO, F. I. Análises de dados de violência doméstica com o uso de aprendizagem de máquina: um mapeamento sistemático. In *Anais da VII Escola Regional de Computação do Ceará, Maranhão e Piauí* (2019), SBC, pp. 143–150.
- [14] LEARN, S. Metrics and scoring: quantifying the quality of predictions; Disponível em: https://scikit-learn.org/stable/modules/model_evaluation.html.

- [15] MONTENEGRO, T. C., ET AL. *Violência contra a mulher: análise comparativa de dados públicos antes e durante a pandemia de Covid-19*. Universidade Federal de Campina Grande, 2021.
- [16] PORTAL BRASILEIRO DE DADOS ABERTOS. Disponível em: <https://dados.gov.br/>.
- [17] SECRETARIA DA SEGURANÇA PÚBLICA - SSP/RS. Disponível em: <https://ssp.rs.gov.br/estatisticas>.
- [18] SECRETARIA DA SEGURANÇA PÚBLICA - SSP/RS. Indicadores da Violência Contra a Mulher - Lei Maria da Penha; Disponível em: <https://ssp.rs.gov.br/indicadores-da-violencia-contra-a-mulher>.
- [19] SECRETARIA DA SEGURANÇA PÚBLICA - SSP/RS. Observatório mulher - Locais de Atendimento; Disponível em: <https://www.ssp.rs.gov.br/observatorio-mulher>.